

# 장면의 유사도 패턴 비교를 이용한 내용기반 동영상 분할 알고리즘

원인수<sup>†</sup>, 조주희<sup>†</sup>, 나상일<sup>†</sup>, 진주경<sup>†</sup>, 정재협<sup>†</sup>, 정동석<sup>\*\*</sup>

## 요 약

본 논문은 내용기반 동영상 분할을 위한 장면의 유사도 패턴 비교 방법을 제안한다. 동영상 장면 전환의 종류는 크게 급진적 전환과 디졸브(dissolve), 페이드인(fade-in), 페이드아웃(fade-out), 와이프 전환(wipe transition)을 포함하는 점진적 전환 형태로 나눌 수 있다. 제안하는 방법은 모든 종류의 장면 전환 검출 문제를 단지 발생 유무의 문제로 간단 정의하고, 장면 전환 종류는 별도로 구분하지 않는다. 장면 전환을 검출하기 위해서는 프레임간의 유사도를 정의해야 한다. 본 논문에서는 장면 내 유사도(within similarity)와 장면 간 유사도(between similarity)를 정의하며 두 유사도의 통계적 패턴 비교를 통하여 최종적으로 장면 전환을 검출하게 된다. 장면 내 유사도와 장면 간 유사도의 비율을 구하는 방법을 통해 플래시라이트나 영상 내 물체 움직임에 대한 거짓 양성 검출을 별도의 후처리 과정 없이도 방지할 수 있음을 확인하였다. 프레임의 특징 값으로는 컬러 히스토그램과 프레임 내 평균 화소값을 이용하였다. TREC-2001, TREC-2002 동영상 셋을 포함한 실험 셋에서 성능을 평가한 결과 제안하는 알고리즘의 경우 총 91.84%의 재현율(recall)과 86.43%의 정확도(precision)의 성능을 보임을 확인할 수 있었다.

## Content based Video Segmentation Algorithm using Comparison of Pattern Similarity

In-su Won<sup>†</sup>, Ju-hee Cho<sup>†</sup>, Sang-il Na<sup>†</sup>, Ju-kyong Jin<sup>†</sup>,  
Jae-hyup Jeong<sup>†</sup>, Dong-seok Jeong<sup>\*\*</sup>

## ABSTRACT

In this paper, we propose the comparison method of pattern similarity for video segmentation algorithm. The shot boundary type is categorized as 2 types, abrupt change and gradual change. The representative examples of gradual change are dissolve, fade-in, fade-out or wipe transition. The proposed method consider the problem to detect shot boundary as 2-class problem. We concentrated if the shot boundary event happens or not. It is essential to define similarity between frames for shot boundary detection. We proposed 2 similarity measures, within similarity and between similarity. The within similarity is defined by feature comparison between frames belong to same shot. The between similarity is defined by feature comparison between frames belong to different scene. Finally we calculated the statistical patterns comparison between the within similarity and between similarity. Because this measure is robust to flash light or object movement, our proposed algorithm make contribution towards reducing false positive rate. We employed color histogram and mean of sub-block on frame image as frame feature. We performed the experimental evaluation with video dataset including set of TREC-2001 and TREC-2002. The proposed algorithm shows the performance, 91.84% recall and 86.43% precision in experimental circumstance.

**Key words:** Content based Video Segmentation(내용기반 동영상분할), Pattern Similarity(패턴 유사도), Video Retrieval(동영상 검색), Shot Boundary Detection(장면 전환 검출)

※ 교신저자(Corresponding Author): 정동석, 주소: 인천광역시 남구 용현동 253(402-751), 전화: 032)860-7415, FAX: 032-868-3654, E-mail: dsjeong@inha.ac.kr  
접수일: 2011년 6월 21일, 수정일: 2011년 8월 26일  
완료일: 2011년 9월 7일

<sup>†</sup> 인하대학교 전자공학과  
정희원(E-mail: woninsu@inha.edu)

준회원(E-mail: tama212@naver.com)  
준회원(E-mail: leptons@naver.com)  
준회원(E-mail: jukyong77@hotmail.com)  
준회원(E-mail: great\_jhyup@hanmail.net)

<sup>\*\*</sup> 정희원, 인하대학교 IT공대학장  
※ 이 논문은 정부(교육과학기술부)의 재원으로 한국연구재단의 중점연구소 지원사업으로 수행된 연구임(2011-0018394)

## 1. 서론

오늘날 컴퓨터와 통신 기술의 급속한 발달은 인터넷을 통한 멀티미디어 데이터의 활발한 공유를 가속시키고 있다. 사용자들은 장소와 시간에 따른 제약 없이 다양한 동영상을 접할 수 있으며 직접 동영상 콘텐츠를 만들어 내기도 한다. 전 세계적으로 방대한 양의 동영상이 만들어지고 또한 여러 사람들에게 공유됨에 따라 동영상 데이터베이스 내에서 동영상의 효율적인 관리 방법이 제안되고 있다. 이 중에서도 장면 분할에 의한 동영상 관리 방법은 동영상의 내용을 기반으로 하는 색인(indexing), 브라우징(browsing) 및 검색(retrieval)을 위한 가장 기본적인 과정이다. 동영상 장면 전환의 종류는 크게 카메라 컷을 기준으로 한 급진적 변환(abrupt change)과 디졸브(dissolve), 페이드인(fade-in), 페이드아웃(fade-out) 와이프 전환(wipe transition) 등을 포함하는 점진적 변환(gradual change)로 구분할 수 있다. 동영상 분할을 위한 장면 전환 검출에 대한 대부분의 연구는 시간적으로 인접한 프레임 영상간의 정량적 유사도를 구하고 유사도 값의 국부 최소값을 구하는 방법을 이용한다.

본 논문은 동일 장면 내 프레임들의 유사도가 장면 간의 프레임 유사도보다 높다는 기본적인 가정을 통하여 장면 간의 유사도(between similarity)와 장면 내 유사도(within similarity)를 정의하고 두 유사도간의 비율을 비교하여 장면 전환을 검출한다. 이를 구하기 위해 현재의 영상 프레임을 기준 프레임으로 정의하고 과거 혹은 미래의 단방향으로 인접한 프레임 간의 유사도를 구해 프레임 유사도 패턴을 생성하여 장면 전환을 검출하였다.

논문의 구성은 다음과 같다. 2장에서는 동영상 장면 전환 검출 방법에 대한 기존 연구에 대하여 서술하며 3장에서는 제안하는 알고리즘에 대하여 설명하였다. 4장에서는 알고리즘 성능을 검증하며 5장에서 결론과 향후 과제에 대해 논한다.

## 2. 기존 연구 내용

동영상을 내용적으로 분할하기 위해 수많은 장면 전환 기술에 대한 연구가 수행되었다. 대부분의 장면 전환 알고리즘의 경우 동영상 프레임간의 특징을 비

교하여 유사도를 구하게 된다. 프레임 영상간의 유사도는 압축된 스트림 자체에서 특징 정보를 이용하여 정의하는 방법과 디코딩 이후의 스트림으로부터 특징 정보를 이용하여 정의하는 방법으로 나눌 수 있다. 압축된 스트림에서는 DCT 계수[1,2], 매크로블록의 종류[3] 혹은 움직임 벡터[4]에 기반을 두는 방법 등이 있다. 압축 스트림 자체에서 장면 전환을 검출하는 방법은 처리 속도 면에서 이득을 얻지만 압축 방식에 따라 고려해야하는 변수가 다르기 때문에 모든 압축 스트림에 대해 공통된 방법을 적용시키기 어렵다는 제약 사항이 있다. 반면 디코딩 이후의 스트림을 이용하는 방법은 화소 값의 차이[5-7], 히스토그램 차이[8,9], 프레임의 평균 밝기 값[10]의 차이에 기반을 두는 방법이 제안되었다. 화소 값 기반 방법은 비교적 계산량이 적고 급진적 장면 전환에 대해 좋은 검출 성능을 보이지만 영상 프레임 내 물체의 움직임이나 카메라 움직임, 혹은 플래시라이트 등에 민감하기 때문에 높은 거짓 양성율(false positive rate)을 보이는 특성을 보인다[11]. 반면 히스토그램 기반의 방법은 위에서 열거한 방해요소에 대해 비교적 강인한 특성을 보이지만 계산 량이 많기 때문에 속도 면에서는 성능이 떨어진다는 특징이 있다. 디졸브와 같은 점진적 변환에서는 장면 전환 시간이 긴 경우 현재 프레임 영상에서 바로 인접한 이웃 프레임만을 비교하는 것만으로는 장면 전환 검출에 한계를 나타낸다. 따라서 많은 연구에서 디졸브 전환을 수식적으로 모델링하고 그에 알맞은 검출 방법들[10,12]을 제안하였다. Su[4]의 경우 기존 장면 전환 모델링의 선형적 한계에 대해 언급하고 전환 구간 내 프레임 간 화소값 증감의 단조성에 기반하여 디졸브 전환을 모델링하였는데 윈도우 내에 존재하는 화소를 행동 패턴에 따라 지지자, 회색분자, 반대자 3개의 카테고리 분류하였다. 또한 장면 전환을 검출하기 위한 통계학적 방법도 연구되었는데 Hanjalic[13]의 경우, 장면 전환의 발생 여부를 검출하는 문제로 간단히 정의하고 급진적 장면 전환과 점진적 장면 전환을 구하기 위한 통계적 접근 방법을 제안하였다. 검출을 위해 프레임 유사도에 대한 장면 간 확률 밀도 함수(inter-shot distribution function)와 장면 내 확률 밀도(intra-shot distribution function)를 정의하였고 장면 길이에 따른 사전 정보(priori information)를 사용하였다. 장면 전환을 검출하기 위한 다양한 프레

임 비교 방법도 제안되었다. 현재 프레임과 이웃한 프레임들을 비교하기 위해 패턴을 생성하는 방법들이 제안되고 있다. Ngo[14]는 시공간 슬라이스(spatio-temporal slice)를 이용한 시각적 패턴을 만들고 장면을 내용적으로 분할하는 방법을 사용하였다.

이렇듯 많은 연구가 진행되었지만, 급진적인 변환에 대한 검출 성능에 비해 점진적 변환에 대한 좋은 성능이 나오진 못하고 있다.

### 3. 장면 전환 검출 알고리즘

제안하는 동영상 장면 전환 검출 알고리즘은 각 프레임 영상으로부터 특징 값을 추출하고 이를 이용하여 장면 간 유사도와 장면 내 유사도 패턴을 생성한다. 생성된 패턴을 비교하여 최종적으로 장면 전환을 검출해 내는 방법에 대해 서술한다.

#### 3.1 특징 값 추출

동영상의 장면 전환을 검출하기 위해서는 연속하는 프레임과 프레임간의 비교를 통해 시각적 유사성을 측정할 필요가 있다. 제안하는 알고리즘에서는 영상 프레임으로부터 HSV 컬러 히스토그램과 전체 프레임 영상을  $N \times N$ 개의 크기를 갖는 부분 블록 영상으로 분할하고 해당 부분 블록 내 화소의 평균값을 특징 값으로 추출하였다. HSV 컬러 히스토그램은 영상의 밝기 변화나 카메라 움직임에 대해 비교적 안정적인 성능을 보이는 장점이 있다. 하지만 히스토그램 자체는 영상의 공간적 정보를 포함하지 않고 점진적 변화에 대해서 변별력을 보이지 못하는 특성을 보인다. 이를 보완하기 위해서 밝기 영상의 부분 블록 내 화소 평균값을 특징으로 함으로써 공간정보의 변화를 감지할 수 있도록 하였다. H, S, V 채널별로 히스토그램을 추출하였고, 연속하는 프레임 간의 유사도를  $FS_C$ 로 정의하고, 식 (1)과 같이 히스토그램 교차(histogram intersection) 방법을 사용하여 계산하였다.

$$FS_C(i, j) = \frac{1}{C} \left( \sum_{k=1}^{bin} \min(H_i(k), H_j(k)) \right) \quad (1)$$

$H_i, H_j$ 는  $i$ 번째 프레임과  $j$ 번째 프레임에서 추출한 히스토그램을 의미하며,  $C$ 는 정규화 상수이다. 반면 영상  $f(x, y)$ 로부터  $W \times H'$  크기를 갖는  $(m, n)$ 의 부

분 블록의 특징  $B_i(m, n)$ 와 이를 이용해 부분 블록의 유사도  $FS_B$ 를 식(2)(3)(4)과 같이 정의한다.  $S$ 는 부분 블록 크기를 의미하고, HSV 각 채널의 히스토그램 도수(bin)는 8을 사용하였다.

$$B_i(m, n) = \frac{1}{S} \sum_{x=1}^{W'} \sum_{y=1}^{H'} f(Wm+x, Hn+y) \quad (2)$$

$$FS_B(i, j) = \frac{1}{N^2} \sum_{m=1}^N \sum_{n=1}^N \delta(B_i(m, n), B_j(m, n), |i-j|) \quad (3)$$

$$\delta(a, b, \Delta t) = \begin{cases} 1 & \text{if } |a-b| \geq TH(\Delta t) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$f(x, y)$ 는 프레임 내  $x, y$  위치에서의 화소 값을 나타낸다.  $\Delta t$ 는 비교하는 프레임간의 시간적 거리를 의미하며 부분 블록을 비교하는  $TH(\Delta t)$ 는  $\Delta t$ 에 대하여 각각 적용되는 임계값을 의미한다. 임계값은 실험치를 통해 결정하였다. 임계값을 결정하기 위한 실험은 TREC-2001 동영상 데이터에서 10개 동영상을 가지고 실험한 결과를 사용하였으며, 장면 전환 시작점을 기준으로 인접한 프레임으로부터 최대 30장 이후 프레임까지 비교하였다. 그림 1은  $\Delta t$  별로 측정된 평균 부분 블록 차이 곡선을 보이고 있다. 유사도 커브에서 디졸브, 페이드 인/아웃, 와이프 전환을 포함한 모든 장면 전환의 경우가 장면 내 부분 블록 차이 평균 값보다 크며  $\Delta t$ 가 5 이하일 때는 점진적 변화의 경우 평균 유사도 차이가 크게 두드러지지 않음을 확인할 수 있다.  $\Delta t$  별로 임계값을 정하기 위해 3가지 점진적 장면 전환에 유사도 커브의 평균을 구하여 이를 임계값으로 정하되 최소 임계값 10을 만족하도록 정하였다. 전체 프레임 간 유사도 값은

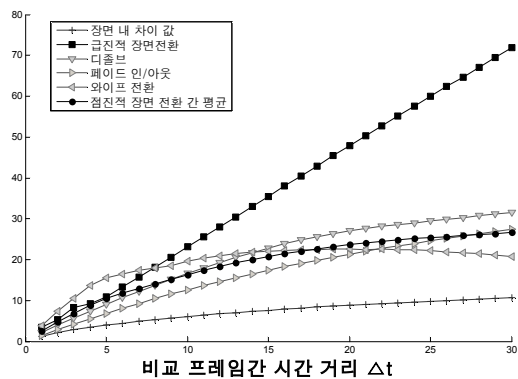


그림 1.  $TH(\Delta t)$  정의. 비교 프레임 간 거리 대비 장면 전환 종류 별 부분 블록 차이 값의 평균

식 (5)와 같이 정의한다.

$$S(i, j) = w_1 FS_C(i, j) + w_2 FS_B(i, j) \quad (5)$$

$w_1, w_2$ 는 가중치 상수로  $w_1 + w_2 = 1$  조건을 만족하도록 하며, 실험에서는 각 0.5를 사용하였다.

### 3.2 장면 전환 모델

동영상의 장면 전환을 검출하기 위한 모델링으로서 시간  $T$ 에 걸친 장면  $X$ 와  $Y$ 간의 전환은 다음 식 (6)과 같이 표현될 수 있다.  $t$ 가 0이거나  $T$ 일 때 장면 전환의 경계를 의미한다.

$$\left(\frac{T-t}{T}\right)X + \frac{t}{T}Y \quad \text{단, } 0 \leq t \leq T \quad (6)$$

하지만 위와 같은 모델링은 장면  $X$ 에서  $Y$ 로의 전환이 선형적으로 발생함과 동시에 장면  $X$ 와  $Y$ 에서의 카메라 움직임이 없거나 아주 적다는 가정을 포함하고 있다. 이에 본 논문에서는 동일한 장면 내 프레임 간 유사도와 서로 다른 장면 간의 프레임 간 유사도의 관계를 다음과 같이 정의하였다.

(1)  $N$ 만큼의 길이를 갖는 장면  $X$ 를 이루고 있는 영상 프레임 시퀀스가  $X = \{x_1, x_2, x_3, \dots, x_N\}$  라면 장면 내 프레임 간 유사도  $S_W$ 를 다음과 같이 정의한다.

$$S_W^{ij} = S(x_i, x_j), \forall x_i \in \vec{X}, \forall x_j \in \vec{X} \quad (7)$$

(2) 서로 다른 장면  $X, Y$ 를 이루고 있는 각각의 영상 프레임 시퀀스  $X = \{x_1, x_2, x_3, \dots, x_N\}$ 와  $Y = \{y_1, y_2, y_3, \dots, y_M\}$ 에 대하여 서로 다른 장면 간 유사도  $S_B$ 는 다음과 같이 정의한다.

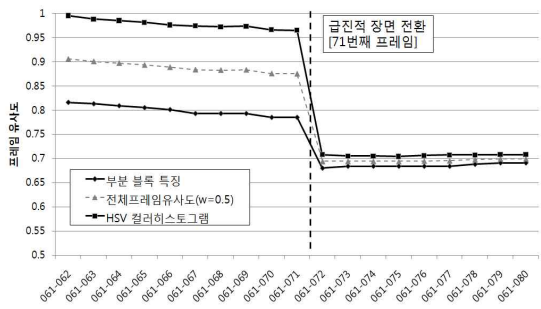
$$S_B^{XY} = S(x_i, y_j), \forall x_i \in \vec{X}, \forall y_j \in \vec{Y} \quad (8)$$

(3) 장면 전환점을 기준으로 장면 간 유사도  $S_B$ 는 장면 내 유사도  $S_W$ 보다 작거나 같다.

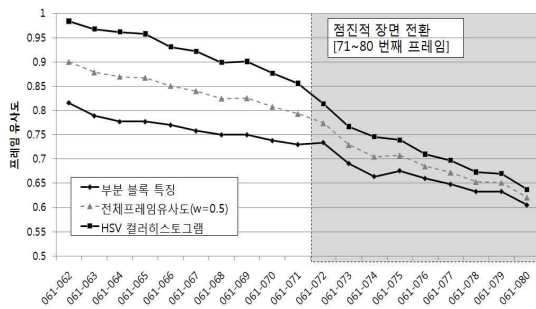
$$\begin{aligned} \max(S_B^{XY}) &\leq \min(S_W^{ij}) \\ \max(S_B^{XY}) &\leq \min(S_W^{ij}) \end{aligned} \quad (9)$$

그림 2는 각 장면 전환 종류 시퀀스 내에서의 프레임 간 거리  $\Delta t$ 에 따른 유사도를 보이고 있으며 기준 프레임  $F_S$ 를 기준으로 미래 프레임과 비교한 결과이다. 가로 축은 기준 프레임과 비교 프레임의 프레임 번호 쌍을 의미한다. 그림 2(a)는 61번째 프레임을 기준으로 71번째 프레임에서 급진적 장면 전환이 발생한 예이다. 그림 2(b)는 점진적 변화의 경우로 기준 프레임을 61번째 프레임으로 하였을 때 71번째 프레임부터 점진적 장면 전환이 발생하는 동영상 시퀀스에 대한 유사도 그림이다. 장면 전환 지점을 기준으로 프레임 간 유사도가 비교적 크게 감소하며 그 값이 동일 장면 내에서 비교한 유사도 값보다는 작다는 것을 확인할 수 있다.

그림 3은 플래시라이트를 포함하는 시퀀스에 대한 급진적 장면 전환의 예시이다. 기준 프레임이 18번째 프레임인 경우로 30번째 프레임까지는 연속적으로 플래시라이트가 발생하고 30번째 프레임에서 급진적 장면 전환이 발생하는 동영상 시퀀스이다. 플래시라이트는 짧게는 한 프레임, 길게는 다섯 프레임에 걸쳐 순간적으로 발생한다. HSV 히스토그램 유사도 곡선은 그림과 같이 변동이 심한 특징을 보인다. 따라서 히스토그램 유사도만을 이용하면 플래시라이트와 급진적 장면 전환 검출에 어려움이 있다.



(a) 급진적 변화 (기준프레임 61th, 장면전환 71th)



(b) 점진적 변화

그림 2. 기준 프레임과 연속한 이웃 프레임 간 시간 거리 대비 유사도. (a) 급진적 변화 (b) 점진적 변화

하지만 부분 블록 특징 유사도는 플래시라이트 발생에 큰 변화가 없는 것을 보인다. 급진적 장면 전환이 발생하는 30번째 프레임을 기준으로 이후 전체 프레임 유사도는 장면 전환 이전 유사도에 비해 낮음을 확인할 수 있다.

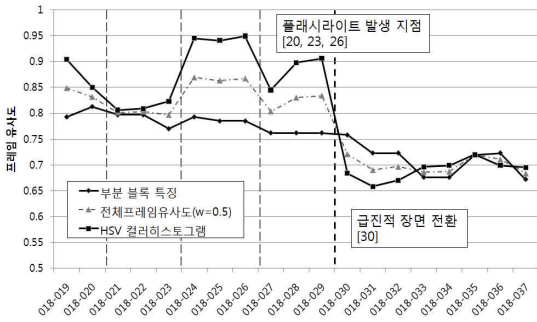


그림 3. Seed 프레임과 연속한 이웃 프레임 간 시간 거리 대비 유사도. 플래시라이트와 급진적 변화가 존재하는 동영상 시퀀스

### 3.3 패턴 생성 방법 및 장면 전환 검출 알고리즘

본 논문에서는 장면과 장면간의 전환이 발생하는 구간을 검출하기 위하여 장면 내 유사도와 장면 간 유사도의 비율을 구하고 그 비율이 국부 최소값을 이루는 지점을 장면 전환 지점으로 검출한다. 이 방법은 장면 내에서 플래시라이트나 물체의 움직임과 같은 일시적이면서 급격한 시각적 변동이 발생하더라도 장면 내 유사도와 장면 간 유사도의 비율을 비교하여 거짓 양성을 방지할 수 있는 효과를 얻는다. 이는 별도의 후처리를 요구하지 않으면서도 플래시라이트와 같은 현상에 강인한 특성을 보이게 된다. 제안하는 알고리즘에서는 기준 프레임을 기준으로 과거 혹은 미래의 단방향으로 인접한 이웃 프레임들의 시각적 특징을 비교하도록 하였다. 두 프레임 간격으로  $L$ 개의 이웃하는 프레임을 비교하여  $L$  차원을 갖는 유사도 값 벡터를 생성하고  $T_s$  개의 유사도 벡터를 통해 그림 4와 같이  $p \times q$  크기를 갖는 패턴  $P$ 를 생성하였다.

패턴의 상단 영역은 장면 내 비교 영역  $R_U$ 로 정의하며 패턴의 하단 영역을 장면 외 비교 영역  $R_L$ 로 정의하였다.  $i$ 번째 프레임에서 생성한 패턴  $P_i$ 로부터 장면 내 비교 영역의 유사도 값의 합  $V_W^i$  과 장면 간 비교 영역의 유사도 값의 합  $V_B^i$  을 구하고 두 값의 비율  $TR(i)$ 을 구하도록 한다.

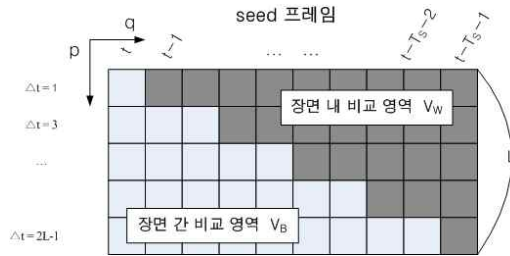


그림 4. 패턴 생성 방법

$$V_W^i = \sum_{(p,q) \subset R_U} P_i(p,q) \tag{10}$$

$$V_B^i = \sum_{(p,q) \subset R_L} P_i(p,q) \tag{11}$$

$$TR(i) = \frac{V_B^i}{V_W^i} \tag{12}$$

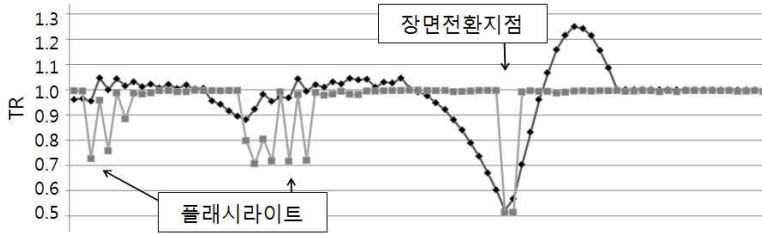
최종적으로  $TR$ 값이 임계값  $TH_{SBD}$  미만인 경우 장면 전환의 후보 영역으로 정의하고 후보 구간 내에서  $TR$ 의 국부 최소값을 찾아 장면 전환 지점으로 검출하도록 한다. 그림 5는 제안하는 알고리즘을 통하여 실제 동영상 시퀀스에 대하여 패턴을 생성하고  $TR$  값을 보인 예시이다. 그림 5와 같이 플래시라이트가 발생한 경우 장면 내 유사도 특징 변화가 불규칙한 패턴을 보이는데,  $V_W$ 의 값이 상대적으로 낮아지며  $V_B$  값은 일정 크기 이상을 유지하는 특성을 보인다. 이는  $TR$ 값을 일정 값만큼 유지 하는 효과로 이어진다. Yeol[3]의 방법과 같이 후처리를 위하여 프레임 유사도 시퀀스에서 윈도우 내에서 존재하는 2개 이상의 국부 최소값을 비교하여 플래시라이트인지 여부를 판별할 필요가 없어 효율적이다. 제안하는 방법을 통해 장면 전환에 대한 거짓양성율을 줄이는 효과를 얻을 수 있음을 확인하였다.

### 3.4 디졸브 검출 성능 향상 방법

점진적 장면 전환에서 특히 디졸브는 카메라 움직임이나 물체 이동과 혼동할 여지가 크기 때문에 정확한 검출이 어렵다[4]. 또한 장시간에 걸친 디졸브 전환은 시간적으로 거리가 가까운 영상 프레임 간 유사도가 높은 특성을 가지기 때문에 3.3절에서 설명한 패턴 생성 방법만으로는 성공적인 디졸브 검출에 한계를 보인다. 따라서 본 논문에서는 영상 프레임 시퀀스로부터 영상 프레임의 전체 평균 밝기 값을 이용하여 윈도우 내에서의 분산과 분산의 1차 미분 값을



(a) 패턴 P



프레임

(b) TR

그림 5. 장면 전환과 플래시라이트와 같은 공격에 대한 패턴 P와 TR (pxq=6×12)

이용 디졸브 검출 성능을 향상하고자 하였다. NW 크기를 갖는 윈도우 내에서의 분산과 1차 미분값을 이용 1차 미분 값이 영교차하는 구간을 디졸브 전환의 후보 영역으로 정의하였다. 잡음을 제거하기 위해 구간평균방법(moving average)을 수행하며, 1차 미분 값의 영교차 지점에서의 음의 최소 극값과 양의 최대 극값이 각각 일정값 이상을 만족함과 동시에 구간 내 TR값의 최소값이  $TH_{SBD-high}$  보다 작은 경우 실제 장면 전환으로서 검출하도록 하였다. 그림 6은 프레임 별 밝기 분산의 1차 미분 값의 시퀀스를 보여준다.

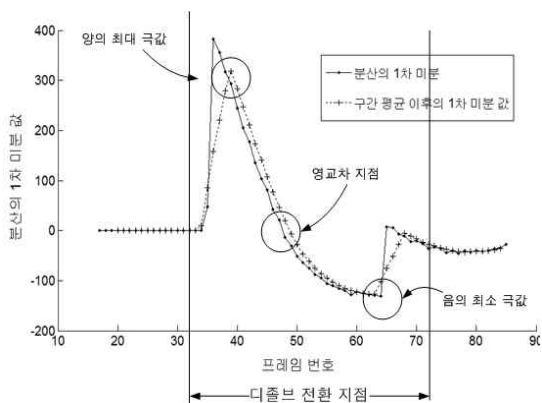


그림 6. 밝기 분산의 1차 미분을 이용한 디졸브 검출

## 4. 실험 결과 및 분석

### 4.1 실험 데이터 구성

제안하는 알고리즘의 성능을 평가하기 위하여 영화, 쇼, 뉴스, 다큐멘터리, 스포츠 등 여러 장르의 동영상을 29개 수집하고 장면 전환이 일어나는 구간을 포함하는 49개의 시퀀스를 추출하여 실험 동영상 구성하였다. 또한 TREC-2001 실험 동영상 7개와 TREC-2002 실험 동영상 12개를 포함하여 실험 영상을 구성하였다. 다음 표 1은 실험 동영상 셋에 대한 장면 전환의 세부 내용을 보여준다. 장면 검출 성능을 평가하기 위한 성능 평가 기준은 재현율(recall)과 정확도(precision) 그리고 F1 평가 기준을 사용하도록 하였고 이들의 정의는 다음과 같다.

$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP} \quad (13)$$

$$F1 = \frac{2 \times Precision \times Recall}{Recall + Precision} \quad (14)$$

TP는 진양성(true positive) FN은 거짓 음성 오류(false negative)를 의미하며 FP는 거짓 양성 오류(false positive)를 의미한다. F1 기준의 경우 재현율과 정확도가 동시에 높은 경우 값이 커지기 때문에

표 1. 실험 데이터 셋의 구성

| 동영상 셋     |         | 전체 길이 (프레임 수) | 급진적 장면 전환 | 점진적 장면 전환 |
|-----------|---------|---------------|-----------|-----------|
| TREC-2001 |         | 273536        | 1008      | 223       |
| TREC-2002 |         | 301741        | 920       | 238       |
| total     |         | 575277        | 1928      | 461       |
| 동영상       | 프레임 크기  | 전체 길이 (프레임 수) | 급진적 장면 전환 | 점진적 장면 전환 |
| seq01     | 640×352 | 8994          | 60        | 10        |
| seq02     | 720×384 | 6773          | 48        | 19        |
| seq03     | 624×352 | 9302          | 122       | 2         |
| seq04     | 448×252 | 8643          | 43        | 2         |
| seq05     | 448×326 | 8643          | 43        | 2         |
| seq06     | 448×280 | 8642          | 72        | 20        |
| seq07     | 720×480 | 7201          | 101       | 1         |
| seq08     | 704×396 | 40138         | 457       | 16        |
| seq09     | 704×396 | 12590         | 98        | 22        |
| seq10     | 640×352 | 14398         | 78        | 25        |
| seq11     | 640×352 | 9900          | 84        | 12        |
| seq12     | 640×352 | 11691         | 69        | 14        |
| seq13     | 448×336 | 5942          | 14        | 27        |
| seq14     | 480×272 | 13488         | 68        | 9         |
| seq15     | 704×396 | 12055         | 50        | 18        |
| seq16     | 544×480 | 8513          | 53        | 11        |
| seq17     | 328×240 | 8646          | 78        | 4         |
| seq18     | 544×480 | 4321          | 11        | 5         |
| total     |         | 199880        | 1549      | 219       |

장면 전환 검출 알고리즘의 성능을 평가하기 위한 평가 기준으로 많이 사용되고 있다[3].

4.2 패턴 사이즈별 성능 비교

패턴 사이즈 별 성능 평가를 위해 *TR* 값만을 이용하여 장면 전환을 검출하도록 하였다. 그림 7은 패턴 사이즈  $p \times q$  에 대한 알고리즘 성능 평가 결과를 나타낸 것이다.  $TH_{SBD}$  에 대한 ROC 곡선을 그렸을 때,  $p \times q$  값이  $4 \times 8$ 과  $6 \times 12$  일 때 검출 성능이 포화되는

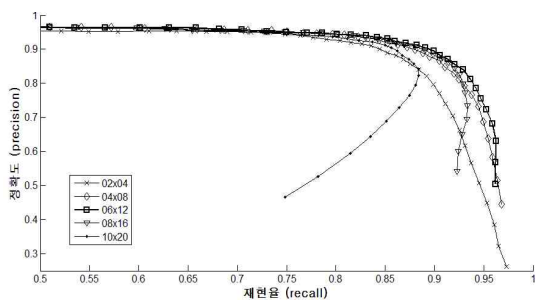


그림 7. ROC 곡선. 프레임 비교 패턴 크기( $p \times q$ )에 따른 성능 평가

현상을 확인할 수 있다. 이후 알고리즘 검증에 관한 실험에서는  $4 \times 8$  크기를 갖는 패턴을 이용하여 장면 전환을 검출하도록 하였다.  $4 \times 8$  크기의 패턴에서  $TH_{SBD}$  값이 0.81 일 때 재현율 88.02%, 정확도 89.96%의 성능을 보였다.

4.3 장면 전환 검출 성능 검증

장면 전환 검출 성능을 평가하기 위하여 동영상 별 검출 성능을 측정하였다. 전반적으로 좋은 성능을 보이고 있으나 점진적 장면 전환에서 장면 간의 유사도가 매우 높거나 장면 전환의 길이가 매우 긴 경우 검출 성능이 상대적으로 떨어지는 것으로 나타났다. 또한 영상 프레임 전체를 가릴 정도의 가까운 거리에서의 물체 움직임은 거짓 양성 검출로 이어진다는 문제가 있었다. 이는 임계값 설정이 동영상 데이터에 따라 민감성과 강인성에 대한 한계가 주된 원인으로 앞으로 해결해야할 과제로서 추후 계속 연구되어야 할 것이다.

표 2는 실험 동영상 셋 별 재현율과 정확도 그리고 F1 평가 기준 결과를 보여주고 있다. 표 3은 동영상

표 2. 제안한 방법의 장면 전환 검출 성능

|           | Recall | Precision | F1    |
|-----------|--------|-----------|-------|
| Dataset 1 | 92.13  | 83.23     | 87.45 |
| TREC-2001 | 93.10  | 91.60     | 92.34 |
| TREC-2002 | 90.06  | 86.26     | 88.12 |

표 3. 제안한 방법의 동영상 별 장면 전환 검출 결과

| 동영상       | 급진적 장면 전환 |       | 점진적 장면 전환 |       | 양성 오류 횟수 |
|-----------|-----------|-------|-----------|-------|----------|
|           | 전체 횟수     | 검출 횟수 | 전체 횟수     | 검출 횟수 |          |
| seq01     | 60        | 60    | 10        | 9     | 25       |
| seq02     | 48        | 46    | 19        | 12    | 3        |
| seq03     | 122       | 117   | 2         | 1     | 11       |
| seq04     | 43        | 43    | 2         | 2     | 36       |
| seq05     | 43        | 42    | 2         | 1     | 15       |
| seq06     | 72        | 68    | 20        | 11    | 18       |
| seq07     | 101       | 101   | 1         | 1     | 5        |
| seq08     | 457       | 428   | 16        | 7     | 35       |
| seq09     | 98        | 89    | 22        | 19    | 20       |
| seq10     | 78        | 75    | 25        | 19    | 3        |
| seq11     | 84        | 82    | 12        | 10    | 13       |
| seq12     | 69        | 65    | 14        | 10    | 41       |
| seq13     | 14        | 13    | 27        | 13    | 5        |
| seq14     | 68        | 67    | 9         | 7     | 20       |
| seq15     | 50        | 46    | 18        | 14    | 16       |
| seq16     | 53        | 52    | 11        | 11    | 34       |
| seq17     | 78        | 69    | 4         | 4     | 15       |
| seq18     | 11        | 11    | 5         | 3     | 5        |
| TREC-2001 | 1008      | 994   | 223       | 152   | 105      |
| TREC-2002 | 920       | 858   | 238       | 185   | 166      |

별 검출 성능 평가 결과를 나타내며 전체 동영상 셋에 대하여 F1 평가 기준에서 가장 좋은 성능을 보인 임계값  $TH_{SBD}$ 에 대한 실험 결과를 서술하였다.

4.4 기존 알고리즘과의 성능 비교

그림 8은 시공간 슬라이스 방법[14], 적응적 임계값을 이용한 방법[15], 그리고 제안하는 알고리즘의 F1 성능 평가 기준을 비교한 그래프이다. TREC-001 동영상 셋을 사용하였으며 그림 8과 같이 제안하는 알고리즘에서 전반적으로 높은 F1 값을 보임을

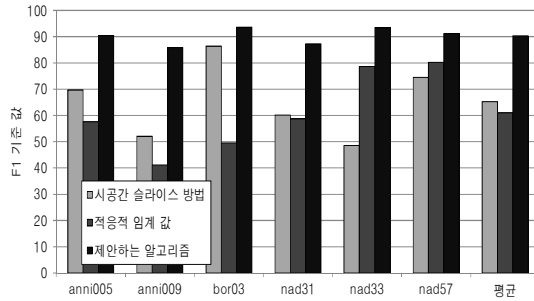


그림 8. 알고리즘 성능 비교

확인하였다.

6. 결론 및 향후 과제

본 논문에서는 동영상의 장면 전환을 검출하기 위하여 장면 간 유사도와 장면 내 유사도를 정의하고 프레임 영상 간 비교 과정에서 생성된 패턴을 사용하여 장면 전환을 검출하는 방법을 제안하였다. 비교적 잡음에 강한 히스토그램 기반의 특징을 이용함과 동시에 영상의 공간적 정보를 이용하기 위한 부분 블록 내 화소 평균값을 이용하여 히스토그램이 갖는 한계를 보완하고자 하였다. 실제 실험 환경에서 높은 검출 성능을 보임을 확인할 수 있었다. 특히 제안하는 알고리즘은 현재 프레임과 이웃한 영상 프레임들과의 유사도 측정 시 미래/과거 단방향으로의 비교 작업만 수행하였고 비교적 이웃 영상 프레임과의 비교 횟수가 적기 때문에 알고리즘 구현에 있어서 보다 효율적이다. 하지만 실시간 처리를 위해서는 처리 속도의 향상이 좀 더 요구되며 이를 위한 연구를 동반할 필요가 있을 것이다. 제안하는 알고리즘은 장면 간 유사도와 장면 내 유사도를 정의하여 플래시라이트나 특수 효과를 위한 프레임 삽입과 같은 상황에 대해서도 비교적 강건하게 장면 전환을 검출해낼 수 있다. 따라서 제안하는 알고리즘은 입력되는 동영상 스트림의 내용을 기반으로 장면을 분류하고 내용을 요약하는 어플리케이션에 적용할 수 있을 것이며 더 나아가 동영상의 내용 기반 검색에서 동영상 내용을 서술하기 위한 식별자 설계를 위한 요소 기술로 적용할 수 있을 것이라 기대한다.

참 고 문 헌

[1] B. L. Yeo and B. Liu, "Rapid Scene Analysis



- on Compressed Video," *IEEE Trans. on Circuit and System for Video Technology*, Vol.5, pp. 533-544, 1995.
- [2] H. J. Zhang, C. Y. Low, and S. W. Smoliar, "Video Parsing and Browsing using Compressed data," *Multimedia Tools and Applications*, Vol.1, pp. 89-111, 1995.
- [3] J. Cao and A. Cai, "A Robust Shot Transition Detection Method Based on Support Vector Machine in Compressed Domain," *Journal of LSEVIER, Pattern Recognition Letter*, Vol. 28, pp. 1534-1540, 2007.
- [4] C. W. Su, H. Y. M. Liao, H. R. Tyan, and L. H. Chen, "A Motion Tolerant Dissolve Detection Algorithm," *IEEE Trans. on Multimedia*, Vol.7, pp. 1106-1113, 2005.
- [5] H. J. Zhang, A. Kanakanhall, and S. W. Smoliar, "Automatic Partitioning of Full-Motion Video," *Video and Image Processing on Multimedia System*, Vol.1, pp. 10-28, 1993.
- [6] A. Nagasaka and Y. Tanaka "Automatic Video Indexing and Full-Video Search for Object Appearances," Proceeding of IFIP 2nd Working Conference. Visual Databases System, Vol.2, pp. 113-127, 1992.
- [7] C. L. Huang and B. Y. Liao, "A Robust Scene-Change Detection Method for Video Segmentation," *IEEE Trans. on Circuits and System for Video Technology*, Vol.11, pp. 1281-1288, 2001.
- [8] Z. Rasheed and M. Shah, "Video Categorization using Semantics and Semiotics," *Video mining*, pp. 185-217, 2003.
- [9] 고경철, 이양원, "변형된  $\chi^2$ -테스트와 자동 임계치-결정 알고리즘을 이용한 장면전환 검출 기법," 전자공학회논문지, 제42권, 제4호, pp. 51-58, 2005.
- [10] B. Truong, C. Dorai, and S. Venkatesh, "New Enhancements to Cut, Fade, and Dissolve Detection Processes in Video Segmentation," Proceedings of the 8th ACM International Conference on Multimedia, pp. 219-227, 2000.
- [11] 고경철, 최영일, 이양원, "플래시라이트에 강건한 장면전환 검출 알고리즘," 전자공학회논문지, 제43권, 제6호, pp. 83-91 2006년.
- [12] R. Lienhart, "Reliable Dissolve Detection," Proceeding of SPIE in Storage and Retrieval for Media Databases, Vol.4315, pp. 219-230, 2001.
- [13] A. Hanjalic and H. Zhang "Optimal Shot Boundary Detection Based on Robust Statistical Models," Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Vol.2, pp. 710-714, 1999.
- [14] C. W. Ngo, T. C. Pong, and H. J. Zhang. "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing," *IEEE Trans. on Image Processing*, Vol.12, pp. 341-355, 2003.
- [15] 원인수, 이준우, 임대규, 정동석, "개선된 적응적 임계값 결정 알고리즘과 후처리 기법을 적용한 동영상 분할 방법," 멀티미디어학회논문지, Vol. 13, pp. 663-673, 2010.



**원 인 수**

2006년 인하대학교 전자공학과  
학사  
2009년 인하대학교 전자공학과  
석사  
2009년~현재 인하대학교 전자공  
학과 박사과정

관심분야 : 영상 처리, 내용기반 멀티미디어검색, 동영상  
복사 검출, 동영상 분할



**진 주 경**

2003년 인하대학교 전자공학과  
학사  
2005년 인하대학교 전자공학과  
석사  
2011년 인하대학교 전자공학과  
박사  
2011년~현재 인하대학교 포스트닥

관심분야 : 영상처리, 멀티미디어 신호처리, 패턴인식, 내  
용 기반 검색



**조 주 희**

2009년 인하대학교 전자공학과  
학사  
2011년 인하대학교 전자공학과  
석사  
관심분야 : 영상 처리, 컴퓨터 비  
전, 동영상 압축 기술, 내  
용 기반 검색



**정 재 협**

2009년 인하대학교 전자공학과  
학사  
2011년 인하대학교 전자공학과  
석사  
2011년~현재 인하대학교 전자공  
학과 박사과정

관심분야 : 패턴인식, 동영상 복사 검출



**나 상 일**

2002년 인하대학교 전자공학과  
학사  
2004년 인하대학교 전자공학과  
석사  
2010년 인하대학교 전자공학과  
박사  
2008년~현재 ETRI 연구원

관심분야 : 영상처리, 컴퓨터 비전, 패턴인식, 내용 기반  
검색



**정 등 석**

1977년 서울대학교 전기공학과  
학사  
1985년 Virginia Tech 전자공학  
과 공학 석사  
1988년 Virginia Tech 전자공학  
과 공학 박사

1988년~현재 인하대학교 전자공학부 교수  
1990년~1994년 전자공학회 논문지 편집위원  
1990년~1994년 통신학회 논문지 편집위원  
2000년~2004년 정보전자공동연구소 소장  
2010년~현재 인하대학교 IT공대학장  
관심분야 : 영상처리, 컴퓨터 비전, 패턴인식, 내용기반  
멀티미디어검색