

## 명도와 에지히스토그램을 이용한 비디오분할 Video Segmentation Using Luminance and Edge Histogram

유현우 † , 장동식 † , 박진형 † , 이법섭 † , 송광섭 ‡

† 고려대학교 산업공학과

‡ 코스모정보통신

### 요 약

비디오데이터의 증가에 따른 효율적 검색, 저장, 브라우징을 위한 방법론이 필요하게 되었다. 본 논문에서는 이러한 시스템을 구축하기 위한 첫 번째 단계인 비디오 분할기법을 제안하고자 한다. 이러한 비디오 분할은 샷경계검출 혹은 장면전환검출이라고 하는데 본 논문에서는 밝기 히스토그램과 에지갯수를 이용하여 프레임간의 유사도를 구별하고 이 유사도가 일정 임계값을 넘지 못하면 장면전환이 있는 것으로 간주한다. 점진적 장면전환검출은 현재프레임과 이전의 샷경계 프레임과의 유사도를 비교하여 검출한다. 다양한 비디오 데이터에 공통적으로 적용할 수 있는 임계값을 설정하기 위해 상관관계(correlation)기법을 사용한다. 실험 결과 급진적 장면전환은 각각 90%, 98%의 정확도(precision)와 회수율(recall)을 나타내었고 점진적 장면전환은 59%, 75%의 정확도와 회수율을 나타내었다.

### 1. 서론

컴퓨터와 통신기술의 눈부신 발전에 따라서 대용량의 문자, 음성, 영상, 비디오등의 멀티미디어 정보가 폭발적으로 증가하고 있다. 이들 정보 중에서 특히 영상 및 비디오가 차지하는 비중은 다른 미디어의 그것과는 비교가 되지 않을 정도로 많다. 따라서 방대한 디지털 영상 및 비디오 데이터를 어떻게 효율적으로 관리하고 검색하느냐가 중요한 문제로 대두되고 있다. 본 논문에서는 내용에 기반한 동영상 정보를 효과적으로 관리하고 검색하는데 필요한 핵심기술인 장면전환검출(scene change detection) 혹은 shot boundary detection 알고리즘을 제안하고자 한다. 장면전환검출을 위해서는 우선 연속적인 비디오 시퀀스를 시간 축으로 분할하여, 카메라의 기록(record)과 멈춤(stop)사이를 나타내는 샷(shot)의 경계를 검출해야한다.

일반적으로 장면전환에는 샷사이의 급격한 변화가 있는 급진적 장면전환(그림1)과 디졸브(dissolve)같은 점진적 장면전환(그림2)이 있다. 비디오 시퀀스 내에 샷을 정의하기 위해서는 급진적/점진적 장면전환점을 검출해야한다. 쉽게 짐작할 수 있듯이 점진적인 변화를 갖는 경우의 장면전환 검출이 급진적인 변화를 갖는 경우보다 훨씬 어려운 문제가 된다. 또한 샷 내부에 패닝(panning), 틸팅(tilting), 주밍(zooming), 트랙킹(tracking)등의 카메라 움직임이나 플래쉬(flash)에 의한 갑작스런 조명변화가 존재하는 경우 샷이 아닌 부분에서 샷을 잘 못 검출하는 경우도 있다.

### 2. 기존연구

장면전환검출알고리즘으로 인접한 두 프레임의 휘도 및 색차 화소값들의 차이를 이용하는 방법과 영상 히스토그램의 변화를 이용하는 방법이 있다.

첫 번째 방법은 연속적인 두 프레임의 같은 위치에 있는 화소값의 절대치를 모든 화소에 대해 구하여 누적한 값이 주어진 임계치를 넘으면 장면전환으로 간주하는 방법이다[1, 2, 4]. 이 방법은 적용이 간단한 반면 영상내 잡음이나 카메라의 움직임등에 민감하다는 단점이 있다. 이러한 단점을 극복하기 위해 연속적인 두 프레임사이의 히스토그램변화를 측정하여 변화값이 일정한 값을 초과하는 경우 장면전환으로 판단하는 히스토그램 기법이 있다[2, 3, 4, 5, 7]. 히스토그램은 영상의 전체적인 특징을 표현하는 것으로 잡음, 카메라 움직임등에 비교적 강한 장점을 가진다. 그러나 이 방법도 유사한 프레임들 중에서 명도나 히스토그램의 변화가 심한 경우 장면전환점을 잘못 검출하는 경우가 있다.

장면전환을 검출하기 위해서는 최종적으로 알고리즘에 의해 얻어진 프레임간의 차이값을 임계값과 비교해야 하는데, 장면전환이 발생하는 전, 후 프레임의 유형이 매우 다양하기 때문에 실질적으로 모든 시퀀스에 공통적으로 적용되는 하나의 임계값을 설정하기가 어렵다. 또한 프레임의 통계적 성격은 그것이 포함되어 있는 샷에 따라 변하기 때문에 동일한 시퀀스에서도 전체적으로 적용되는 임계값을 설정하기 어려운 문제가 있다.



그림 1: 급진적 장면전환



그림 2: 점진적 장면전환

대부분의 급격한 장면전환은 히스토그램을 이용하여 잘 검출된다. 그러나 점진적 장면전환이나 영상내 특수효과가 있는 경우는 아직 검출율이 낮으며 이에 대한 많은 연구가 진행되고 있다[3, 4, 5, 6, 7]. 디졸브등 특수 효과에 의한 점진적 장면전환이 발생하는 경우에 연속적인 두 프레임 사이의 변화는 미소한 반면 이들 변화가 누적되어 결국 점진적 변화가 시작된 장면과 끝나는 장면은 큰 차이를 보이는 것이 보통이다.

### 3. 제안된 알고리즘

본 논문에서는 에지갯수와 명도히스토그램의 유사도를 이용하여 명도변화에 강인한 장면전환검출 알고리즘을 제안한다. 두 개의 임계값설정을 통해서 급진적/점진적 장면전환을 검출하고 모든 비디오 시퀀스에 공통적으로 적용될 수 임계값 설정을 위하여 프레임간 상관계수(correlation)를 이용한다.

#### 3.1 명도히스토그램(LuminanceHistogram)을 이용한 상관계수

$k$ 번째 프레임의 명도평균과 분산을  $m_k, \sigma_k$ 라고 할 때, 연속된  $k, k+1$  번째 프레임사이의 명도상관계수  $LIFC$ (Luminance Inter-Frame Correlation)는 식(1)과 같이 계산된다.

$$S_{LIFC}(k, k+1) = \frac{\sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (X_k[i][j] - m_k)(X_{k+1}[i][j] - m_{k+1})}{\sigma_k \sigma_{k+1}} \quad (1)$$

$$-1 \leq S_{LIFC}(k, k+1) \leq 1$$

이때,  $W$ 와  $H$ 는 각각 프레임의 가로와 세로길이,  $X_k[i][j]$ 는  $k$ 번째 프레임의  $(i, j)$ 위치에서의 명도값을 나타낸다.

#### 3.2 에지갯수(Edge Histogram)를 이용한 상관계수

명도만을 이용하여 장면전환을 검출할 경우 유사한 프레임간에도 명도변화가 크면 장면전환으로 잘못검출(False Alarm)될 수 있다. 이러한 단점을 극복하기 위해 프레임간의 에지정보도 이용하여 장면전환을 검출한다.

영상을  $3 \times 3$ 의 9개의 동일한 크기의 영역으로 나누고 각 영역당 에지를 구해서 에지갯수를 계산한다. 이 에지갯수를 이용하여 상관계수를 구한다.

$k$ 번째 프레임의 에지갯수의 평균과 분산을  $m_k, \sigma_k$ 라고 할 때, 연속된  $k, k+1$  번째 프레임사이의 에지상관계수는  $EIFC$ (Edge Inter-Frame Correlation)는 식(2)과 같이 계산된다.

$$S_{EIFC}(k, k+1) =$$

$$\frac{\sum_{i=0}^8 (X_k[i] - m_k)(X_{k+1}[i] - m_{k+1})}{\sigma_k \sigma_{k+1}} \quad (2)$$

$$-1 \leq S_{EIFC}(k, k+1) \leq 1$$

이때,  $X_k[i]$ 는  $k$ 번째 프레임  $i$ 영역의 에지갯수를 나타낸다.

### 3.3 통합상관계수(Integrated Correlation)

최종적으로는 명도히스토그램과 에지갯수를 통한 상관계수를 결합한 식(3)을 사용한다

$$S_{IFC}(k, k+1) = w_l S_{LIFC}(k, k+1) + w_e S_{EIFC}(k, k+1) \quad (3)$$

$$w_l + w_e = 1 \quad (4)$$

여기서  $w_l, w_e$ 는 각각 명도히스토그램을 이용하는 것과 에지갯수를 이용하는 것과의 가중치를 나타낸다

### 3.4 급격한/점진적인 장면전환검출

장면전환검출을 위해서는 샷의 경계를 검출해야 하는데 경계가 되는 샷을 검출하는 기본개념은 다음과 같다. 만약 현재 프레임이 이전프레임과 충분히 다를 경우(급격한 장면전환)나 이전의 샷경계와 충분히 다를 경우(점진적 장면전환) 현재프레임을 샷의 경계 프레임으로 판단한다. 충분히 다른 것을 판단하는 기준으로는 보통 인접한 각각의 프레임에서 특징(예를 들면 명도히스토그램)을 추출하여 특징간의 거리를(예를 들면 유클리디안거리) 일정 임계값과 비교하여 유사도를 판단하나 이 경우는 다양한 영상에 공통적으로 적용될 수 있는 임계값 설정이 어려우므로 본 논문에서는 프레임간의 유사도인 상관계수를 이용하여 모든 비디오에 공통적으로 적용되는 임계값을 설정한다. 제안된 알고리즘은 그림3에 간략히 표현되어있다.

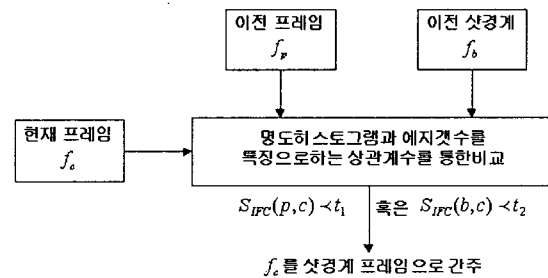


그림3: 샷경계검출 다이어그램

그림에서  $t_1$ 은 급진적 장면전환검출을 위한 임계값으로,  $t_2$ 는 점진적 장면전환검출을 위한 임계값으로 사용된다.

연속된 모든 비디오 프레임에 장면전환 검출을

위한 변화를 측정하는 것은 상당한 계산시간을 요구한다. 계산시간을 줄일 수 있는 방법은 비디오 데이터의 시간적 해상도를 줄이는 것이다. 즉, 초당 30프레임의 비디오를 초당 15프레임으로 표본화하여 장면전환을 검출하는 것이다. 실제적으로 비디오 데이터는 시간적인 중복성이 크므로 이 방법이 사용이 합리적이며 점진적 장면전환을 검출하는 데에도 유용하게 사용될 수 있다.

#### 4. 실험 및 성능분석

##### 4.1 실험환경

본 논문에서 제안한 장면전환검출방법의 성능을 평가하기 위해 Windows 98환경에서 Pentium II 400MHz PC를 사용하여 실험하였다. VC++을 이용하여 프로그램 하였고 GUI는 그림4와 같다.

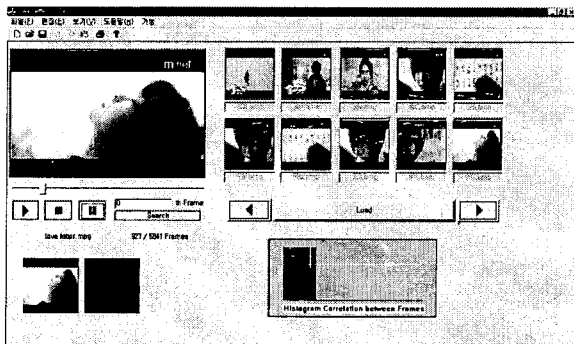


그림4. GUI(Graphic User Interface)

##### 4.2 실험결과 및 성능분석

실험에 사용된 비디오는 맥주광고와 영화, 뮤직비디오의 3개 MPEG파일을 사용하였다. 장면전환검출을 위해 명도히스토그램과 에지갯수와와 가중치  $w_l$ ,  $w_e$ 는 각각 0.5로 균등하게 주었다. 성능분석을 위해서 정확도(precision)와 회수율(recall)을 계산하였다. 비디오에서  $N_{TOTAL}$ 을 전체 프레임수,  $N_{SCD}$ 를 급격한 장면변화가 있거나 점진적 장면변화가 있는 프레임의 수,  $N_{CORRECT}$ 를 정확히 장면전환을 검출한 수,  $N_{MISSED}$ 를 장면전환을 찾지못하고 놓친갯수,  $N_{FALSE}$ 를 장면전환이 아닌 프레임을 검출한 수라 하면, 정확도와 회수율의 정의는 식(5), 식(6)과 같다.

$$\text{정확도(Precision)} = \frac{N_{CORRECT}}{N_{CORRECT} + N_{FALSE}} \quad (5)$$

$$\text{회수율(Recall)} = \frac{N_{CORRECT}}{N_{SCD}} \quad (6)$$

실험결과 급진적 장면전환(표1)에서는 정확도와 회수율이 평균 0.90과 0.98을 나타내었으며 점진적 장

면전환(표2)에서는 평균 0.59, 0.75를 나타내었다. 영화 Love letter의 경우 급진적 장면전환을 검출하지 못한(표1의  $N_{MISSED}$ ) 3개 프레임은 실제 장면변화가 일어났지만 명도와 에지 데이터가 프레임간에 유사하므로 검출하지 못하였고(그림5), 5개 프레임을 검출하지 못한 뮤직비디오 누구보다(비쥬)도 비슷한 현상을 발견할 수 있었다. 장면전환을 잘못 검출 하는 경우(표1의  $N_{FALSE}$ )는 장면변화가 없지만 외부조명이 갑자기 밝아지거나 어두어지는 경우 혹은 카메라의 움직임이 있을 경우에 일어났다(그림6). 점진적 장면전환에서 장면전환을 검출하지 못한 경우(표2의  $N_{MISSED}$ )는 장면전환 전후가 매우 유사한 경우에 발생했고(그림7), 점진적 장면전환을 잘못 검출한 경우는 그림8에서와 같이 동일한 샷에서 변화가 심할 경우 생겼다.

종류	급진적 장면전환검출					(P:R)
	$N_{TOTAL}$	$N_{SCD}$	$N_{CORRECT}$	$N_{MISSED}$	$N_{FALSE}$	
광고 Hite맥주	554	3	3	0	0	(1.0:1.0)
영화 Love letter	5841	96	93	3	23	(0.80:0.97)
뮤직비디오 누구보다 (비쥬)	8128	141	136	5	14	(0.91:0.96)

표1. 실험 비디오 데이터와 실험결과(급진적 장면전환의 경우) :  $N_{TOTAL}$ 는 전체 프레임수,  $N_{SCD}$ 는 급격한 장면전환개수,  $N_{CORRECT}$ 는 정확히 찾은 급격한 장면전환수,  $N_{MISSED}$ 는 놓친갯수,  $N_{FALSE}$ 는 잘못 검출한 수를 나타낸다.(P:R)은 정확도(Precision)와 회수율(Recall)을 나타낸다.

종류	점진적 장면전환검출					(P:R)
	$N_{TOTAL}$	$N_{SCD}$	$N_{CORRECT}$	$N_{MISSED}$	$N_{FALSE}$	
광고 Hite맥주	554	8	8	0	4	(0.67:1.0)
영화 Love letter	5841	4	2	2	2	(0.5:0.5)
뮤직비디오 누구보다 (비쥬)	8128	0	×	×	×	×

표2. 실험 비디오 데이터와 실험결과(점진적 장면전환의 경우) :  $N_{TOTAL}$ 는 전체 프레임수,  $N_{SCD}$ 는 점진적 장면전환개수,  $N_{CORRECT}$ 는 정확히 찾은 점진적장면전환수,  $N_{MISSED}$ 는 놓친갯수,  $N_{FALSE}$ 는 잘못 검출한 수를 나타낸다.(P:R)은 정확도(Precision)와 회수율(Recall)을 나타낸다.

이전에 언급한바와같이 점진적 장면전환은 전후 프레임간의 유사성 때문에 급진적 장면전환 검출보다 어렵다는 것을 알 수 있었고 아직 연구되어야 할 과제로 생각된다.

## 6. 결론 및 향후연구방향

본 논문에서는 내용기반 동영상검색에 필수적인 장면전환검출알고리즘을 제안하였다. 밝기 히스토그램과 에지갯수를 이용하여 프레임간의 유사도를 구별하고 이 유사도가 일정 임계값 내에 있으면 장면전환이 있는 것으로 간주했다. 점진적 장면전환 검출은 현재프레임과 이전의 샷경계 프레임과의 유사도를 비교하여 검출하고 상관관계(correlation) 기법을 사용하여 다양한 비디오데이터에 공통적으로 적용할 수 있는 임계값을 설정하였다. 현재는 압축형태의 MPEG파일을 완전히 디코딩(Full decoding)한 후 장면전환검출을 하였지만 향후에는 압축영역에서 직접 장면전환 검출을 할 수 있도록 함으로써 빠른 검출이 가능하도록 하는 것과 장면전환검출후의 프레임들로부터 검색에 필요한 색상, 질감, 형태 등의 특징들을 추출하고 색인 하여 검색에 사용할 수 있는 환경을 구축하는 것이 연구 과제로 남아있다.

## 참고문헌

- [1]. K. Otsuji, Y. Tonomura, and Y. Ohba, "Video browsing using brightness data," *Visual Communication and Image Processing*, vol. SPIE-1606, pp.980-989, 1991
- [2]. H. J. Zhang, A. Kankanhalli, S.W.Smoliar, "Automatic partitioning of Full Motion Video", *Multimedia Systems*, Vol.1, No.1, pp.10-28, 1993
- [3]. B.L. Yeo and B. Liu, "Rapid scene analysis on compressed video", *IEEE Transactions on Circuits and Systems for Video Technology*, pp 533-544, 1995
- [4]. G. Ahanger and T.D.C.Little, "A survey of technologies for parsing and indexing digital video", *J. Visual Comm. and Image Rep.* 7. pp. 28-43, 1996
- [5]. H.J.Zhang, J.Wu, D.Zhong, and S.W.Smoliar, "An integrated system for content-based video retrieval and browsing", *Pattern Recognition* 30, pp.643-658, 1997
- [6]. W.Xiong and J.C.-M Lee, "Efficient Scene Change Detection and Camera Motion Annotation for Video Classification", *Computer Vision and Image Understanding*, vol. 71, no. 2, pp. 166-181, 1998
- [7]. A. K. Jain, A. Vailaya, and W. Xiong, "Query by Video Clip", *Multimedia Systems: Special Issue on Video Libraries*, vol. 7, no. 5, pp. 369-384, 1999.



그림5: 급진적 장면전환을 검출하지 못한 경우(명도와 에지 데이터가 프레임간에 유사하다)



그림6: 급진적 장면전환을 잘못 검출한 경우(외부 조명변화와 카메라움직임으로 인해)



그림7: 점진적 장면전환을 검출하지 못한 경우(명도와 에지 데이터가 매우 유사하다)



그림8: 점진적 장면전환으로 잘못 검출한 경우(샷내에 변화가 심함)