다중 특징을 사용한 장면 전환 검출 시스템

윤성수, 정성환
창원대학교 컴퓨터공학과

Scene Change Detection System Using Multiple Features

Sung-Soo Yoon, Sung-Hwan Jung
Dept. of Computer Engineering, Changwon Nat'l University
E-mail : goldeneye@cdcs.chagnwon.ac.kr, sjung@sarim.changwon.ac.kr

요 약

대용량 동영상 데이터의 효율적인 관리와 검색을 위해서는 장면 단위의 정확한 분할이 선행되어야 한다. 본 논문에서는 동영상의 시각적인 내용에 기반한 장면 전환 검출 방법을 연구하였다. 본 논문에서는 프레임 단위의 특징과 프레임 내의 부분영역 단위의 특징을 결합한 다중 특징을 사용한 장면 전환 검출 방법을 제안한다. 실험을 통한 성능 평가에서는 기존의 방법들에 비해 Recall과 Precision에서 각각 7.7%, 10%의 향상을 보였다.

1. 서 론

초고속 통신망의 보급과 멀티미디어 관련 기술의 발달로 인하여 오디오, 동영상 등과 같은 대용량 멀티미디어 데이터가 증가하고 있다. 일반 사용자들도 멀티미디어 데이터에 손쉽게 접근하고, 사용, 생성, 가공할 수 있는 환경이 주어짐으로써 데이터들이 더욱 방대해지고 있다. 이에 따라 멀티미디어 데이터를 정보로서 관리하기 위한 효율적인 전송, 저장, 분류 및 검색 기술이 요구되어진다. 멀티미디어 데이터 중에 동영상 데이터의 경우, 대용량의 데이터를 효과적으로 검색, 관리할 수 있는 기술이 더욱 필요하다.

이러한 요구로 국제 표준화 기구인 ISO/IEC의 연합기술위원회 산하의 SC29(MPEG)에서 멀티미디어 컨텐츠의 특징들을 기술하기 위한 방법에 대하여 MPEG-7(Multimedia Content Description Interface)이라는 이름으로 표준화 작업이 진행되고 있다[1, 2].

동영상 데이터베이스를 검색하기 위한 방법은 크게 문자기반 검색 방법(Text-Based Retrieval Method)과 내용기반 검색 방법(Content-Based Retrieval Method)으로 나누어진다[3, 4].

동영상 데이터베이스에 대한 내용기반 검색을 위해서는 일반적으로 비디오 분할(Video Segmentation) 단계와 특징 추출(Feature Extraction)단계, 인덱싱(Indexing)단계 그리고 검색(Retrieval) 단계를 필요로 한다[4]. 이 중에서 가장 선행되어야 할 단계는 비디오 분할 단계인데, 연속된 동영상 데이터로부터 샷(Shot)을 검출하는 것으로, 이를 장면 전환 검출(Scene Change Detection) 또는 샷 경계 검출(Shot Boundary Detection)이라 한다[5].

본 논문에서는 내용기반 비디오 검색의 첫 단계인 장면 전환 검출을 연구한다. 효율적이고 정확한 장면 전환 검출을 하기 위해 프레임 단위의 특징과 부분영역 단위의 특징을 결합한 다중 특징을 사용한 방법을 제안한다. 본 논문에서 사용하는 특징으로는 프레임 단위의 특징으로, 컬러와 명암도 각각에 대한 히스토그램을 사용하고, 부분 영역 단위의 특징으로, DCT(Discrete Cosine Transform)변환 후, DC계수를 사용한다.

본 논문의 구성은 다음과 같다. 서론에 이어, 2장에서는 기존의 연구들을 기술하고, 3장에서는 제안한 복합 특징을 이용한 장면 전환 검출 방법에 대해 설명한다. 그리고 4장에서는 제안한 장면 전환 검출 방법에 의한 실험 및 결과에 대해 살펴보고, 마지막으로 5장에서 결론을 낸다.

2. 장면 전환 검출 방법들

동영상에서 장면 전환 검출 방법은 특징의 추출 대상에 따라 프레임 단위 장면 전환 검출 방법과 부분영역 단위 장면 전환 검출 방법, 화소 단위 장면 전환 검출 방법으로 나누어진다[6].

2.1 프레임 단위 장면 전환 검출 방법

프레임 단위 장면 전환 검출 방법은 가장 일반적으로 사용되는 방법이다. 인접한 두 프레임에 대한 특징 값을 구하고, 그 특징 값들의 차이를 계산한다. 그 값이 임계치를 초과하는 지점을 장면 전환점이라고 한다. 프레임 단위의 검출 방법으로는 히스토그램 차이 비교하는 방법[7], 에지 변화 비교하는 방법[8] 등이 있다.

2.2 부분영역 단위 장면 전환 검출 방법

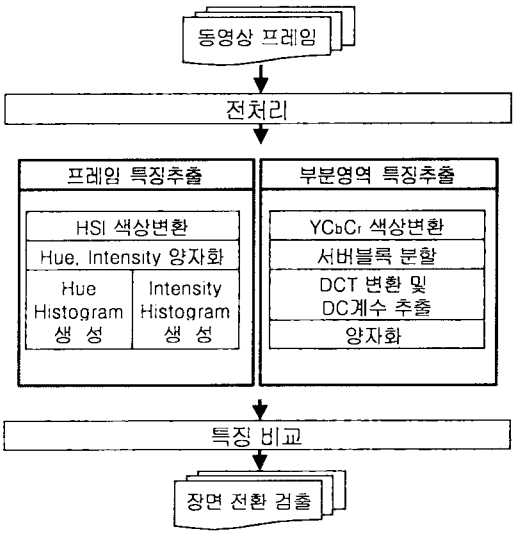
부분영역 단위의 검출 방법은 한 프레임을 몇 개의 부분 영역으로 분해한 후, 각 부분영역별로 특징 값을 구하여, 인접한 프레임간에 대응하는 영역들의 특징 값의 차이를 계산함으로써 장면 전환점을 검출하는 방법으로, 장면 전환점은 특징 값의 차이가 임계치를 초과하는 지점이 된다[9].

2.3 화소 단위 장면 전환 검출 방법

화소단위로 장면 전환을 검출하는 방법은 가장 단순한 방법이다. 인접한 프레임간에 대응하는 화소들의 특징 값 차이를 계산함으로써 장면 전환점을 검출하는 방법이다[10]. 이러한 방법은 카메라나 물체의 움직임과 잡영에 민감하여 잘못된 경계를 검출할 가능성이 크며, 특히, 점진적인 장면 전환의 경우는 거의 검출하지 못한다는 취약점을 보인다.

3. 제안한 장면 전환 검출 방법

본 논문에서 제안한 장면 전환 검출 방법은 프레임 단위의 특징과 부분영역 단위의 특징을 결합하여 사용한다. 본 논문에서 프레임 단위의 특징은 Hue 히스토그램과 Intensity 히스토그램을 이용한다. 그리고 부분영역 단위의 특징으로는 서버블록 단위로 DCT 변환을 수행한 후의 DC계수를 이용한다. 프레임 단위의 특징인 Hue 히스토그램과 Intensity 히스토그램은 프레임 전체의 색상과 밝기 정보를 나타낸다. 하지만, 이 특징들은 프레임 내의 국부적 정보와 구조적인 정보를 포함하지 못한다. 그래서, 국부적 정보를 포함하는 부분영역 단위의 특징을 사용함으로써 이러한 문제를 보안할 수 있다.



(그림 1) 제안한 장면 전환 검출 방법 구성도

(그림 1)은 본 논문에서 제안한 장면 전환 검출 방법의 전체 구성도이다. 먼저, 입력된 동영상 프레임에 대하여 전처리 과정을 수행한다. 그 다음, 특징 추출 과정에서 입력 프레임에 대하여 프레임 단위의 특징인 Hue 히스토그램과 Intensity 히스토그램을 구하고, 그리고 부분영역 단위의 특징인 DC계수를 추출한다. 마지막으로, 인접한 프레임간의 특징 비교 과정에서 프레임 단위의 특징과 부분영역 단위의 특징을 이용하여 장면 전환점을 검출함으로써 동영상은 각각의 샷으로 분할된다.

3.1 전처리

전처리 단계는 프레임 내의 잡음 제거하기 위한 단계이다. 본 논문에서는 평균 필터(Mean filter)를 적용하여 잡음을 제거한다. 평균 필터는 저역 통과 필터로서 고주파 성분을 제거하는 필터이다. 필터의 크기는 영상의 선명화를 고려하여 3×3 크기의 필터를 사용하였다. (그림 2)는 3×3 평균필터이다.

$$\frac{1}{9} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(그림 2) 3×3 평균필터

3.2 프레임 단위 특징 추출

프레임 단위 특징 추출단계에서는 프레임 내의 전반적인 정보를 포함하는 특징들인 Hue 히스토그램과 Intensity 히스토그램을 생성한다. 특징 추출 과정은 다음과 같다. 먼저, 입력 프레임을 HSI 색상 공간으로 변환한 다음, Hue성분과 Intensity성분을 양자화한다. 양자화된 Hue성분과 Intensity성분에 대하여 Hue 히스토그램과 Intensity 히스토그램을 각각 생성한다.

3.2.1. HSI 색상 공간으로 변환 및 양자화

RGB 색상공간에 HSI색상모델로 변환하는 수식은 다음과 같다[11].

$$H = COS^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\} \quad (1)$$

$$S = 1 - \frac{3[\min(R,G,B)]}{(R+G+B)} \quad (2)$$

$$I = \frac{1}{3}(R+G+B) \quad (3)$$

HSI 색상 공간으로 변환 후, Hue성분과 Intensity 성분을 각각 32레벨과 16레벨로 균등 양자화(Uniform Quantization)된다.

3.2.2 Hue 히스토그램 생성

Hue성분에 대한 정규화된 색상 히스토그램 CH는 수식(4)와 같다.

$$CH_i = \frac{1}{XY}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} H'(x,y) \quad (4)$$

$$where \quad H'(x,y)=\begin{cases}1, & H(x,y)=i\\ 0, & H(x,y)\neq i\end{cases}$$

$$i=0,\cdots,N-1$$

CH_i는 Hue 히스토그램의 i번째 빈(bin)의 값이다. X는 프레임의 가로 크기이며, Y는 세로 크기이다. N은 Hue의 레벨 수이다. 그리고 $H(x,y)$는 프레임 내의 (x,y)위치의 Hue값을 나타낸다.

3.2.3 Intensity 히스토그램 생성

Intensity성분에 대한 정규화된 명암도 히스토그램 IH는 수식(5)와 같다.

$$IH_j = \frac{1}{XY}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} I'(x,y) \quad (5)$$

$$where \quad I'(x,y)=\begin{cases}1, & I(x,y)=j\\ 0, & I(x,y)\neq j\end{cases}$$

$$j=0,\cdots,M-1$$

IH_j는 Intensity 히스토그램의 j번째 빈의 값이다. X는 프레임의 가로 크기이며, Y는 세로 크기이다. M은 Intensity의 레벨 수이다. 그리고 $I(x,y)$는 프레임 내의 (x,y)위치의 Intensity값이다.

3.3 부분영역 단위 특징 추출

부분영역 단위 특징 추출 단계에서는 프레임 내의 국부적 특성을 포함하는 특징으로, 서버블록의 DC계수를 특징 값으로 추출한다. 특징 추출 과정은 다음과 같다. 먼저, 색상 공간을 YCbCr색상 공간으로 변환한다. 그리고 프레임을 8×8 서버블록으로 분해하고, 각 서버블록에 대해 DCT를 수행한다. 계산된 DCT계수들 중 DC계수만을 추출하고, 이를 양자화하여 부분영역 단위 특징 값을 생성한다.

3.3.1 YCbCr 색상 공간으로 변환

RGB 색상 공간을 YCbCr 색상 공간으로 변형한다. YCbCr 색상 공간은 수식(6)에 의해 얻을 수 있다.

$$\begin{bmatrix}Y\\ Cb\\ Cr\end{bmatrix} = \begin{bmatrix}0.299 & 0.587 & 0.114\\ -0.169 & -0.331 & 0.500\\ 0.500 & -0.419 & -0.081\end{bmatrix}\begin{bmatrix}R\\ G\\ B\end{bmatrix} \quad (6)$$

3.2.2 서버블록별 DC 계수 추출 및 양자화

DCT(Discrete Cosine Transform)는 주파수 영역으로의 변환 함수로서 이산 여현 변환이다. DCT는 정지 영상 압축 국제 표준인 JPEG과 동영상 국제 표준인 MPEG에서 영상 압축 시, 기본 알고리즘에 채택하고 있다. DCT 변환 공식은 다음과 같다.

$$C(u,v) = a(u)a(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)\cdot \cos\left[\frac{(2i+1)u\pi}{2N}\right]\cdot\cos\left[\frac{(2j+1)v\pi}{2N}\right] \quad (7)$$

$$\forall u,v = 0,\cdots,N-1$$

$$a(w)=\begin{cases}\sqrt{\frac{1}{N}}, & \text{for } w=0\\ \sqrt{\frac{2}{N}}, & \text{for } w=1,2,\cdots,N-1\end{cases}$$

프레임을 8×8 크기의 서버블록으로 분해하여 각각의 블록에 대해 수식(7)을 적용하여 각 성분별 DC계수를 추출한다. 추출된 DC계수들을 256레벨로 균등 양자화한다.

3.4 특징 비교

특징 비교 과정은 이전 과정에서 생성한 특징들을 이용하여 장면 전환 여부를 판단하는 과정이다. 본 논문에서는 프레임 단위 특징 비교 단계, 부분 영역 단위 특징 비교 단계 그리고, 앞의 두 결과를 결합하는 장면 전환 검출 과정으로 구성된다.

3.4.1 프레임 단위 특징 비교

본 논문에서의 프레임 단위 특징 비교는 히스토그램 인터섹션(Histogram Intersection)을 사용한다. 공식은 아래의 수식(8)과 같다.

$$S_f(F^i, F^{i+1}) = \frac{\sum_{h=0}^{M-1} \min(F^i_{CH}(h), F^{i+1}_{CH}(h))}{\min(|F^i|, |F^{i+1}|)\times 2} + \frac{\sum_{b=0}^{M-1} \min(F^i_{IH}(b), F^{i+1}_{IH}(b))}{\min(|F^i|, |F^{i+1}|)\times 2} \quad (8)$$

$S_f(F^i, F^{i+1})$는 i번째 프레임 F^i와 $i+1$번째 프레임 F^{i+1}사이의 유사도이며 $S_f(F^i, F^{i+1})$는 0과 1사이의 값을 가진다. $F^i_{CH}(h)$는 F^i의 CH의 h번째 빈의 값을 의미한다. $F^i_{IH}(b)$는 F^i의 IH의 b번째 빈의 값을 의미한다.

3.4.2 부분 영역 단위 특징 비교

본 논문에서의 부분 영역 단위 특징 비교는 L1 norm을 사용한다. 공식은 아래와 수식(9)와 같다.

$$S_b(F^i, F^{i+1}) = \frac{1}{N}\sum_{n=0}^{N}(|F^i_Y(n) - F^{i+1}_Y(n)|) + \frac{1}{N}\sum_{n=0}^{N}(|F^i_{Cb}(n) - F^{i+1}_{Cb}(n)|) + \frac{1}{N}\sum_{n=0}^{N}(|F^i_{Cr}(n) - F^{i+1}_{Cr}(n)|) \quad (9)$$

$S_b(F^i, F^{i+1})$는 i번째 프레임 F^i와 $i+1$번째 프레임 F_{i+1}사이의 유사도이며, $F^i_Y(n)$와 $F^i_{Cb}(n)$, $F^i_{Cr}(n)$는 프레임 F^i의 n번째 $Y\,Cb\,Cr$ 각 성분의 양자화된 DC 계수를 의미한다.

3.4.3 장면 전환 검출

장면 전환 검출 과정에서는 앞에서 구한 프레임 단위 특징의 유사도, $S_f(F^i, F^{i+1})$와 부분 영역 단위 유사도, $S_b(F^i, F^{i+1})$를 이용하여 정확한 장면 전환점을 결정한다. 이 때, 프레임단위 유사도 $S_f(F^i, F^{i+1})$가 임계치 T_f보다 작거나 같고, 부분영역 단위 유사도 $S_b(F^i, F^{i+1})$가 임계치 T_b보다 작거나 같을 때, F_i는 장면 전환점이 된다.

4. 실험 및 결과

본 연구는 Microsoft Windows 2000 Professional 환경에서 Intel Pentium-4 CPU 1.5GHz PC에서 실험하였으며, 시스템은 Visual C++6.0 MFC를 이용하여 구현하였다.

4.1 성능 평가 척도

장면 전환 검출 성능을 평가하기 위해 식(10)과 식(11)과 같이 Recall과 Precision의 두 가지 성능 평가 척도 함수를 사용하였다.

$$Recall = \frac{C_n}{C_n + M_n} \quad (10)$$

$$Precision = \frac{C_n}{C_n + F_n} \quad (11)$$

여기서, C_n은 정확히 검출된 컷의 수를 나타내고, M_n은 찾지 못한 컷의 수, F_n은 잘못된 컷의 수를 나타낸다.

4.2 실험 결과

본 논문에서 사용된 동영상 실험 데이터는 광고, 뉴스, 영화, 스포츠로 주제별로 분류하여 시험하였다. <표 1>은 본 논문의 성능평가를 위해 사용된 동영상 데이터의 종류와 파일 수, 프레임 수, 컷의 수이다.

<표 1> 실험에 사용한 동영상 데이터

동영상 종류	파일 수	프레임 수	컷 수
광고	8	3015	132
뉴스	5	4254	63
영화	3	5391	97
스포츠	3	1884	32
합계	19	14546	324

본 논문에서 제안한 방법과 비교를 위해, 기존의 방법인 컬러 히스토그램과 명암도 히스토그램 그리고 DC계수를 이용한 방법을 Recall과 Precision을 사용하여 평가하였다.

<표 2>는 기존의 방법들과 제안한 방법의 장면 전환 검출 성능 결과표이다. <표 2>에서 보는 바와 같

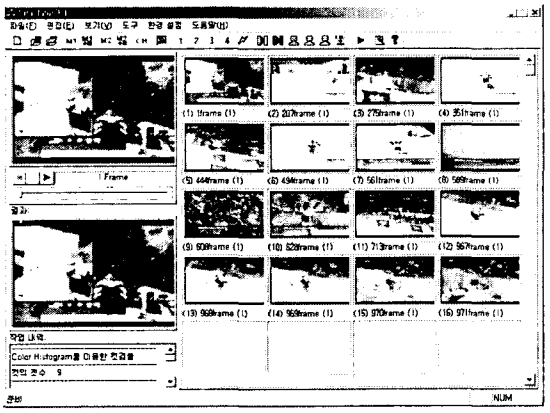
이 제안한 방법이 네 종류 동영상 데이터에 평균적으로 컬러 히스토그램에 비해서 Recall이 약 6% 향상되었고, Precision이 약 10% 향상되었다. 명암도 히스토그램에 비해서도 Recall과 Precision이 각각 약 10%, 12%씩 향상되었고, DC계수를 이용한 방법에 배해서도 Recall과 Precision이 각각 약 7%, 8%씩 향상되었다. 제안한 방법은 기존의 방법들에 비해 평균적으로 Recall은 약 8%, Precision은 약 10% 향상된 결과를 보였다.

<표 2> 제안한 장면 전환 검출방법의 결과의 비교

방법 / 종류	컬러 히스토그램		명암도 히스토그램		DC계수		제안된 통합 방법	
	R	P	R	P	R	P	R	P
광고	0.92	0.72	0.82	0.73	0.87	0.80	0.93	0.90
뉴스	0.95	0.94	0.92	0.93	0.93	0.82	0.98	0.99
영화	0.83	0.86	0.82	0.81	0.90	0.92	0.93	0.95
스포츠	0.72	0.79	0.69	0.76	0.69	0.85	0.83	0.86
평 균	86%	83%	82%	81%	85%	85%	92%	93%

본 논문에서 제안한 프레임 단위의 특징과 부분영역 단위의 특징을 결합 사용한 방법은 Recall보다 Precision에서 더 많은 향상을 보인다. 이는 제안한 방법이 잘못된 장면 전환 검출 수를 상대적으로 줄임으로써 정확한 장면 전환 검출을 할 수 있음을 나타낸다.

(그림 3)은 제안한 방법으로 장면 전환 검출을 한 예이다.



(그림 3) 제안한 방법에 의한 장면 전환 검출 결과 예

5. 결 론

본 논문에서는 프레임 단위의 특징과 부분영역 단위의 특징을 결합한 다중 특징을 이용한 장면 전환 검출 방법을 제안하였다. 프레임 단위의 특징으로는 Hue 히스토그램과 Intensity 히스토그램을 사용하였고, 부분영역 단위 특징으로는 서버블록별 DC계수를 이용하였다. 이는 프레임 내의 전반적인 특성과 국부적인 특성을 결합함으로써 장면 전환 검출의 성능을 향상하기 위함이다.

본 논문에서 제안한 방법으로 광고와 뉴스, 영화, 스포츠 등의 다양한 종류의 데이터로 장면전환 검출을 실험해 보았다.

실험 결과, 제안한 방법이 기존의 방법들에 비해 평균적으로 Recall은 약 8%, Precision은 10% 향상된 결과를 보였다.

[참고문헌]

[1] ISO/IEC JTC1/SC29/WG11 N3751. "Introduction to MPEG-7", La Baule, October 2000.

[2] ISO/IEC JTC1/SC29/WG11 N3751. "Overview of the MPEG-7 Standard" La Baule, October 2000.

[3] N. Dimitrova and M. A. Mottalab, "Content-based video retrieval by example video clip," SPIE Storage and Retrieval for Image and Video Database, Vol. 3022, pp.59-70, 1995.

[4] H. J. Zhang, C. Y. Low, S. W. Smoliar and J. H. Wu, "Video parsing retrieval and browsing: An integrated and content-based solution," Proc. ACM Multimedia '95, pp.15-24, 1995.

[5] R. M. Bolle *et al.* "Video query: research directions," IBM Journal of Research and Development, Vol.42, No.2 pp.233-252, 1998.

[6] 이미숙, 황본우, 이성환, "내용기반 비디오 검색을 이용한 장면 전환 검출 방법의 성능 분석," 한국정보과학회 학술논문 발표집(B), pp.543-546, 1997.

[7] M.J. Swain. and D.H. Ballard, "Color Indexing", International journal of computer vision No.1. pp.11-32, 1991.

[8] Hong Jiang Zhang, Stephen W. Smoliar and Atreyi Kankanhalli, "Automatic partitioning of full-motion video," Multimedia Systems, Vol.1, No.1 pp.10- 28, 1993.

[9] John S. Boreczky and Lawrence A. Rowe, "Comparison of video shot boundary detection techniques," in Storage & Retrieval for Image and Video Database IV, Proc. of SPIE 2670, pp.170-179, 1996.

[10] Ramin Zabih, Justin Miller and Kevin Mai, "A feature-based algorithm for detecting and classifying scene breaks," Proc. ACM Multimedia 95, pp.189-200, 1995.

[11] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison-Wesley publishing company, Inc., 1993.