

# Color Layout Descriptor 를 이용한 새로운 장면전환검출 방법

김수영\*, 원종운\*\*, 오상근\*, 박길흠\*

\*경북대학교 전자공학과

\*\*한국전자통신연구원 우정기술연구소

e-mail : [juliana02@palgong.knu.ac.kr](mailto:juliana02@palgong.knu.ac.kr)

## A New Scene Change Detection Method using Color Layout Descriptor

Soo-Young Kim\*, Jong-Un Won\*\*, Sang-Keun Oh\* Kil-Houm Park\*

\*Dept. of Electrical Engineering, Kyungpook National University

\*\*Postal Technology Research Center,

Electronics and Telecommunications Research Institute

### 요 약

본 논문에서는 동영상 내용 기반 검색을 위한 표준 인덱스인 color layout descriptor(CLD)에 기반한 새로운 장면전환검출 방법을 제안한다. 제안한 방법은 입력 영상을 8×8 블록으로 나누고, 그 블록을 대표하는 대표컬러를 선택하여 DCT 변환 후 추출된 DC, AC 계수에 가중치를 부여한 다음 슬라이딩 윈도우를 적용시켜 장면 전환을 검출한다. 이러한 방법은 영상의 전체적인 특성과 국부적인 특성을 동시에 반영하므로, 조명변화나 물체의 움직임에 민감했던 기존 대표적인 방법인 히스토그램 기반 방법과 에지 변화율을 이용한 방법들이 가지는 문제점을 보완할 수 있다. 또한 제안한 방법은 비디오 검색을 위한 MPEG-7 의 검색 서술자 중의 하나인 CLD 를 이용하는데, 이러한 CLD 는 컬러의 공간적 분포를 잘 설명하므로 MPEG-7 을 위한 검색 서술자에 적합한 장면전환검출 및 검색이 이루어질 수 있다. 제안한 방법을 실험한 결과 기존의 방법보다 장면전환 검출율이 높게 나타났다.

### 1. 서론

디지털 기술의 발전과 더불어 디지털 비디오 기술 또한 급속히 발전하고 있다. 디지털 비디오 기술은 거대한 양의 동영상 데이터 베이스로부터 사용자의 접근성을 개선하기 위해 중요한 논제로 대두되고 있다. 특히 영상 회의, 전자 도서관, 멀티미디어 교육 분야에서 디지털 비디오에 대한 요구가 날로 증가되고 있다. 그러나, 디지털 비디오에 대한 폭넓은 이용이 제한되는 이유는 디지털 콘텐츠를 사용하는 거대한 양의 데이터 베이스로부터 원하는 정보를 추출하기가 쉽지 않기 때문이다. 이러한 동영상에 대한 이용자의 접근성을 개선하기 위한 방법으로 동영상의 내부 구조 정보를 추출하여 동영상 데이터의 내용 목록(table of contents)을 작성함으로써 해결할 수 있다.

동영상은 그림 1 과 같은 계층적인 데이터 구조 모델로 볼 수 있다.

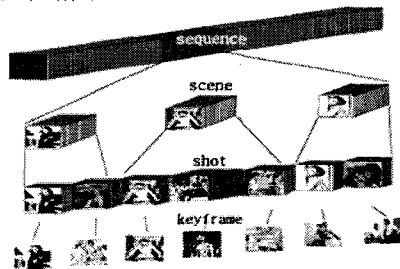


그림 1. 동영상의 계층적인 데이터 구조 모델

전체 동영상을 시퀀스라는 레벨로 분할하고, 시퀀

스는 다시 장면 단위로 분할되고, 각각의 샷(shot)을 대표할 수 있는 대표 프레임들을 추출하는 것이다. 추출된 대표 프레임들은 동영상 데이터의 내용 목록을 구성하게 된다.

내용에 기반한 동영상의 구조화 기술, 즉 동영상을 샷 단위로 나누는 것은 서로 다른 장면으로의 장면 변화를 검출하는 것으로서 장면 내에서 일반적으로 발생하는 물체의 움직임, 카메라의 움직임을 구분하는 것이다. 이와 같이 장면 내에서 발생하는 카메라 또는 물체의 움직임이 아닌 장면전환 시점을 찾는 것으로서 동영상의 내용 기반 구조화가 가능하다.

동영상의 계층적 구조화를 위한 동영상의 장면전환 검출은 동영상 데이터의 압축영역 또는 비 압축영역에서 각각 수행될 수 있는데, 일반적으로 비 압축영역에서 검출기법은 대표프레임(DC 영상)을 가지는 압축영역의 동영상에 적용 가능하다. 비 압축영역에서 장면전환검출 방법으로 각 프레임의 히스토그램 기반 방법[1], 물체의 윤곽정보를 이용한 방법[2,3] 등이 제안되었다. 히스토그램에 기반하는 방법[1]은 두 프레임간에 장면전환이 이루어 진다면 히스토그램의 변화가 비교적 크다는 성질을 이용한 방법이다. 그러나 이 방법은 움직임이 많이 존재하는 영상과 조명변화가 다양하거나 비슷한 색채로의 장면전환이 일어나는 영상에서는 장면전환검출이 정확하게 이루어지지 않는 단점이 나타난다. 한편, 물체의 윤곽 정보, 즉 에지 변화율(ECR)[2,3]을 이용하는 방법은 각 영상마다 에지를 검출하여야 하므로 처리량이 많을 뿐 아니라, 히스토그램 기반 방법에 비해 대체로 검출율이 떨어진다. 따라서 이러한 두 방법들의 문제점을 보완하고, 동영상의 계층 구조화를 위한 동영상 내용 기반 검색과 원활한 연계를 위하여 검색에 사용되는 색인을 특징으로 하여 동영상의 계층 구조화, 즉 장면 전환 검출에 이용할 새로운 방법이 필요하다.

본 논문에서는 동영상 내용 기반 검색을 위한 표준인덱스에 기반한 새로운 방법으로 color layout descriptor(CLD)를 이용한 장면전환검출 방법을 제안한다. MPEG-7의 검색 서술자 중의 하나로서, 칼라의 공간적 분포를 설명하는 CLD를 장면전환검출에 이용하면 위의 두 가지 방법, 즉 히스토그램에 기반 하는 방법과 에지 변화율을 이용하는 방법들의 장점을 가질 뿐만 아니라 실험 결과에서도 영상의 전체적인 특성과 국부적인 특성을 모두 반영할 수 있어서 조명변화나 물체의 움직임에 민감했던 기존의 방법에 비해 장면전환검출이 더욱 우수하게 나타났다.

본 논문 2장에서는 기존의 방법과 제안하는 방법을 설명하고 3장에서는 제안하는 방법의 우수성을 평가하기 위해 실험을 통해 결과를 비교하며, 끝으로 4장에서 결론을 내린다.

## 2. 장면전환검출

장면 전환 검출은 연속되는 두 개의 샷(shot)사이의 경계인 컷(cut)을 검출하는 것으로 검색이나 색인을 위한 비디오 구조화 및 해석의 첫 단계로 중요한 과정이다. 이 과정에서 주로 이용되는 기존의 방법은

히스토그램을 기반으로 하는 방법과 에지 변화율을 이용하는 방법이 있다.

### 2.1. 기존의 방법

#### 2.1.1. Histogram Differences (HD)

히스토그램 기반 방법은, 두 프레임간에 장면전환이 이루어 진다면 히스토그램의 변화가 비교적 크다는 성질을 이용한 방법이다. 밝기 레벨 히스토그램 차이[1]에 기초한 방법은 다음과 같다. 동영상에서  $m, n$  번째 프레임들  $f_m$  과  $f_n$  에서 밝기의 레벨  $i$  ( $i=1,2,\dots,B$ ,  $B$ : bin의 수)에 해당하는 히스토그램을  $H(f_m, i), H(f_n, i)$ 라 하자. 이때, 히스토그램들의 차이의 합은 다음과 같이 정의된다.

$$S(f_m, f_n) = \sum_{i=1}^B |H(f_m, i) - H(f_n, i)| \quad (1)$$

이 때, 컷은  $S(f_m, f_n)$ 이 임계치보다 클 때 정의된다. 이 방법을 컬러 히스토그램[4]에도 적용할 수 있다. 두 컬러 프레임  $I_{i-1}$ 과  $I_i$ 사이에서 컬러 히스토그램 차이  $CHD_i$ 는 다음과 같이 정의된다.

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

여기에서,  $P_i(r, g, b)$ 는  $N$ 개의 픽셀들의 프레임  $I_i$ 에서  $color(r, g, b)$ 의 픽셀들의 수이다.

#### 2.2.2 Edge Change Ratio (ECR)

히스토그램의 차이를 이용하는 방법은 영상의 전체적인 특성을 고려하기 때문에 영상의 국부적인 특성은 고려되지 못하는 단점이 나타난다. 장면이 전환되는 부분에서는 두 샷사이에서 에지의 변화가 두드러지는 현상이 나타난다. 따라서, 장면전환검출의 또 다른 방법으로서 에지 변화율(Edge Change Ratio, ECR)을 이용한다[3]. ECR은 움직임 강도를 찾는 데 적당한 특징(feature)으로 사용된다. 이러한 특징을 장면전환검출에 이용하면 영상 내에 존재하는 에지를 비교하므로 영상의 국부적인 특성을 고려할 수 있다.

각 프레임마다 Sobel 연산자를 이용하여 에지를 검출해서, 바로 전 프레임의 에지와 비교하여  $n$  번째 프레임의 에지 변화율( $ECR_n$ )을 구한다.

$$ECR_n = \max\left(\frac{X_n^{in}}{X_n}, \frac{X_n^{out}}{X_{n-1}}\right) \quad (3)$$

$X_n$ 은  $n$  번째 프레임에서 에지 픽셀의 수라 하고,  $X_n^{in}$ 은  $n$  번째의 프레임에서 새로이 에지로 나타나는 픽셀의 수이고,  $X_n^{out}$ 은  $n-1$  번째의 프레임에서 에지인 픽셀이  $n$  번째 프레임에서는 에지가 아닌 픽셀의 수이다. 이러한 에지 변화율이 임계치 보다 크다면,  $n$  번째 프레임에서 컷이 검출된다.

### 2.2. 제안하는 방법

장면전환검출을 위한 기존의 방법 중 히스토그램 차이를 이용한 방법은 영상 내 국부적인 특성은 고려하지 않으므로, 움직임이 많이 존재하는 영상과 조명

변화가 다양하거나 비슷한 색채로의 장면전환이 일어나는 영상에서는 검출이 정확하게 이루어지지 않는 단점이 나타난다. 예를 들어, 폭발장면이나 큰 물체들의 빠른 움직임 같은 것에는 장면 전환이 아닌 경우에 대해서 장면전환으로 오인할 수 있고 밝기 또는 컬러의 분포가 전체적으로 변화가 많은 동영상에 있어서는 장면 전환을 검출하지 못하는 경우가 생길 수 있다. 이러한 문제점을 보완하고자 제안된 에지 변화율(ECR)을 이용하는 방법은 국부적인 특성이 잘 고려될 수 있지만, 각 영상마다 에지를 검출하여야 하므로 연산량이 많을 뿐 아니라, 장면전환의 검출결과 면에서 히스토그램 차이를 이용하는 방법보다 좋지 않다. 본 논문에서는 이러한 두 방법들의 문제점을 보완하기 위해서, color layout descriptor(CLD)를 이용한 장면전환검출 방법을 제안한다.

CLD 는 비디오 검색을 위한 MPEG-7 의 검색 서술자 중의 하나로서, 컬러의 공간적 분포를 설명하는 서술자이다. 컬러의 공간적 분포는 빠른 영상 검색이나 내용기반 검색, 또는 시각화를 기술하는데 효과적이다[4,5]. 이러한 검색 서술자로서 중요한 역할을 하는 CLD 를 장면전환검출에서 이용하면, 일반 동영상에서의 효율적인 장면전환검출 뿐만 아니라 MPEG-7 의 검색 서술자에 적합한 장면전환검출 및 검색이 매우 용이해지리라 기대된다.

이 CLD 의 추출[6]은 그림 2 와 같은 네 개의 연속적인 과정들로 이루어진다. 우선 이미지 파티셔닝을 수행하고, 대표하는(dominant) 컬러선택, DCT 변환, 그리고 지그재그 스캔한 DCT 계수들의 비선형적 양자화를 수행하게 된다. 첫 번째 단계에서,  $W \times H$  화소 크기의 입력영상을  $8 \times 8$  개의 블록으로 나눈다. 각 블록의 크기는  $W/8 \times H/8$  이다. 두 번째 단계에서, 입력영상에 대응하는  $8 \times 8$  화소 크기의 평균 영상을 생성하기 위해서, 각 블록마다 세 개의 구성요소(Y, Cb, Cr)를 각각 대표하는 컬러를 선택한다. 이때, 대표하는 컬러는 주로 평균값(average color)을 선택한다. 세 번째 단계에서 입력영상에 대응하는 세 개의 평균영상의 구성요소(Y, Cb, Cr)를 각각  $8 \times 8$  DCT 변환한다. 이로부터 정보를 많이 포함하는 원소들을 추출하여 장면전환검출에 이용하기 위해, DCT 계수들에 대한 지그재그 스캔을 이용하여 낮은 주파수의 계수들로부터 차례로 정해진 개수의 계수를 추출한다. 끝으로 이렇게 추출된 계수를 이용하여 컷을 검출한다. 본 논문에서는 저 주파수에서부터 Y 신호에 대해서는 6 개의 계수를, Cb, Cr 은 각각 3 개씩의 계수를 이용한다. 이 계수들을 이용하여 앞 프레임과 비교하여 계수들의 차를 구하고 그 계수마다 가중치를 부여한다. 이때 가중치는 실험적으로 결정한다.

$$CLD_i = \sum_{k=0}^{N-1} (DY_i + DCb_i + DCr_i) \quad (4)$$

$$DY_i = \sum_{k=0}^5 w_k \sqrt{Y_i(k) - Y_{i-1}(k)} \quad (5)$$

$$DCb_i = \sum_{k=0}^2 w_k \sqrt{Cb_i(k) - Cb_{i-1}(k)} \quad (6)$$

$$DCr_i = \sum_{k=0}^2 w_k \sqrt{Cr_i(k) - Cr_{i-1}(k)} \quad (7)$$

여기서,  $CLD_i$ 는 i 번째 프레임의 계수를 비교한 값이고,  $DY_i, DCb_i, DCr_i$ 는 앞 프레임과의 계수 차에 가중치를 부여한 값이다.  $Y_i, Cb_i, Cr_i$ 는 i 번째 프레임의 Y, Cb, Cr 컬러 구성요소의 계수들이고,  $w_k$ 는 계수들에 부여하는 가중치이다. 이렇게 구한  $CLD_i$  값들에 대해 슬라이딩 윈도우를 적용하여 컷을 검출한다. 슬라이딩 윈도우는 어떤 시간적 범위 내에 하나의 피크값을 검출한다. 그 피크값이 컷이 된다.

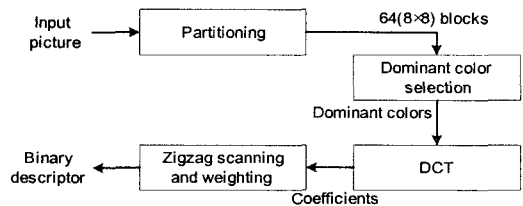


그림 2. Color layout descriptor 추출

CLD 를 사용하여 장면전환을 검출하는 방법을 이용하면  $8 \times 8$  화소 크기의 이미지를 사용하고, 정해진 소수 개의 저주파 계수들만 특징으로 이용하여 장면전환검출을 수행하므로 연산량이 매우 감소한다. 뿐만 아니라, CLD 을 이용하면 DC 계수뿐만 아니라 비중이 큰 AC 계수들도 사용하므로 영상의 전체적인 특성과 함께 국부적인 특성도 반영될 수 있으므로 영상의 전체적 특성만 고려하는 히스토그램의 차에 기반한 방법에 비해 장면전환검출이 더욱 효율적으로 이루어질 수 있다. 즉, 동영상 내부에 조명변화나 움직임이 많은 경우에 기존의 방법들인 히스토그램 차이나 에지 변화율을 이용하는 방법에서 오 검출하거나 미 검출하는 현상이 제안된 방법을 통해 개선된다.

한편, 동영상의 압축된 데이터를 디코딩 하지 않고 바로 이용하면 장면전환검출을 빠르게 할 수 있는데, 제안된 방법은 압축된 MPEG 에서 비교적 쉽게 구할 수 있는 DC 영상을 이용하므로 압축영역에서의 구현이 용이할 뿐만 아니라, 더욱 효과적인 장면전환검출이 이루어진다. 제안하는 알고리즘에 대한 블록다이어그램은 그림 3 과 같다.

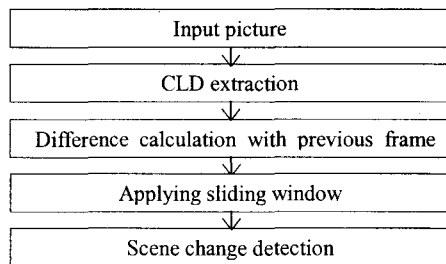


그림 3. 제안하는 알고리즘

3. 실험 및 고찰

CLD 를 장면 전환 검출에 적용하여 제안한 방법을 기존의 방법인 HD, ECR 과 비교하였다. 실험 영상의 포맷은 MPEG-2 형식으로, 그 프레임의 크기는 352 × 240 이다. Sequence1 은 색채가 다양한 시트콤을 선택하였고, Sequence2 는 움직임이 매우 많은 영상으로, 짧은 시간 내에 장면전환이 자주 일어나는 다큐멘터리이다. 그리고 Sequence3 은 조명변화가 있지만 실험 영상 중에서는 가장 변화가 작은 편인 드라마를 실험 영상으로 사용하였다. 표 1 에서는 실험에 이용한 영상의 전체 프레임 수와 실제 장면전환이 일어나는 프레임 수를 나타냈다.

표 1. 실험에 사용된 비디오 시퀀스

	Sequence1	Sequenc2	Sequence3
Display time	6m 31s	1m 41s	8m 57s
Total frame No.	11729	21844	16106
Real cut change	55cuts	108cuts	104cuts

표 2 와 3, 그리고 4 에서는 각각 영상에 대한 기존의 방법과 제안된 방법을 실험한 결과 검출된 장면전환의 수와 검출하지 못한 장면전환의 수, 그리고 검출률을 비교하였다. 검출 성능을 평가하는 방법으로 Recall 와 Precision 을 사용하였다.

$$\text{Recall} = \frac{\text{No. of detected}}{\text{No. of detected} + \text{No. of missed}} \times 100 [\%] \quad (8)$$

$$\text{Precision} = \frac{\text{No. of detected}}{\text{No. of detected} + \text{No. of false}} \times 100 [\%] \quad (9)$$

표 2. Sequence1 의 장면전환검출 결과

	HD	ECR	CLD
Detected	47	38	51
Missed	8	17	4
False	3	7	3
Recall	85.45%	69.09%	92.73%
Precision	94.00%	84.50%	94.44%

표 3. Sequence2 의 장면전환검출 결과

	HD	ECR	CLD
Detected	65	46	73
Missed	43	62	35
False	4	8	4
Recall	60.19%	42.59%	67.59%
Precision	94.20%	85.19%	94.81%

표 4. Sequence3 의 장면전환검출 결과

	HD	ECR	CLD
Detected	98	90	101
Missed	6	14	3
False	3	8	6
Recall	94.23%	86.54%	97.12%
Precision	97.03%	91.84%	94.39%

표 2,3,4 에 나타난 것처럼 제안한 방법은 기존의 방법을 대표하는 히스토그램 기반 방법과 에지 기반 방법에 비해 대체로 더욱 효과적으로 나타났다. 특히, Sequence2 의 경우에는 움직임과 조명변화가 매우 많고 단시간 내에 장면전환이 많이 일어나기 때문에 전반적으로 검출률이 떨어지지만, 제안한 방법이 기존의 방법에 비해 우수한 성능을 보였다. Sequence 3 의 경우에는 제안한 방법의 Recall 이 기존의 방법보다 우수하지만, Precision 는 조금 떨어지는 것을 알 수 있다. Sequence3 의 경우 실험 영상 중에서 가장 변화가 작은 비디오이므로, 변화가 작은 영상에 대해서는 변화가 많은 것보다 오 검출이 많이 일어나는 것을 볼 수 있다. 앞으로 이러한 부분을 보완할 수 있도록 더욱 연구해 나가야겠다.

4. 결론

본 논문에서는 동영상 내용 기반 검색을 위한 표준 인덱스인 CLD 에 기반한 새로운 장면전환검출 방법을 제안하였다. 제안한 방법은 CLD 에 기반하여 장면전환을 검출하기 때문에, 영상의 전체적인 특성과 국부적인 특성을 모두 반영할 수 있어서 조명변화나 물체의 움직임에 민감했던 기존의 방법에 비해 더욱 우수한 결과를 나타내었다. 제안된 방법은 압축영역에서 비교적 쉽게 구할 수 있는 DC 영상을 이용하므로, 압축 영역에서의 구현이 가능하다. 뿐만 아니라, CLD 는 MPEG-7 의 서술자이므로 이 방법을 이용하면 향후 MPEG-7 의 검색 서술자에 적합한 장면전환검출 및 검색이 매우 용이해지리라 기대된다.

참고문헌

- [1] Y.Tonomura, "Video handling based on structured information for hypermedia systems", in ACM Proceedings: International Conference on Multimedia Information Systems '91, 1991, pp.333-344
- [2] R.Zabih, J.Miler, and K. Mai. "A Featured-Based Algorithm for Detecting and Classifying Scene Breaks" Proc.ACM Multimedia 95, San Francisco, CA, pp.189-200, Nov.1995
- [3] "Text of 15938-3 (MPEG-7 Visual FCD)", ISO/IEC JTC1/SC29/WG 11/N4062. 2001
- [4] U.Gargi, S.Oswald, D.Kosiba, S.Devadiga, and R.Kasturi, "Evaluation of video sequence indexing and hierarchical video indexing", Proc.SPIE: Storage Retrieval Image Video Database III February 1995, 144-151
- [5] F.Idris, "Review of Image and Video Indexing Techniques", Journal of Visual Communication and Image Representation Vol.8, No.2, June 1997, pp146-166
- [6] Kasutani, E. Yamada, A.; "The MPEG-7 Color Layout Descriptor: A Compact Image Feature Description for High-speed Image/Video Segment Retrieval", Image Processing, 2001 International Conference on , Vol. 1 , 2001 , pp 674 -677 vol.1
- [7] B.S.Manjunath, Jens-Rainer Ohm, Vinod V.Vasudevan, and Akio Yamada "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, Vol.11, No.6, June