

# 압축 영상에서 히스토그램 인터섹션을 이용한 점진적인 장면 전환의 구간 검출

論 文

54D-11-5

## Detection of Gradual Shot Conversion Duration using Histogram Intersection in Compressed Video

權 哲 鉉<sup>\*</sup> · 韓 科 辰<sup>\*</sup> · 李 明 鎬<sup>\*\*</sup> · 朴 相 暎<sup>\*\*\*</sup>

(Chul-Hyun Kwon · Doo-Jin Han · Myoung-Ho Lee · Sang-Hui Park)

**Abstract** - An algorithm detecting abrupt and gradual shot boundaries is proposed in this paper. The conventional methods detect abrupt shot boundaries well, but do not show good performance on gradual shot boundaries. The proposed method is based on the fact that the difference of the characteristic between frames is large when the shot conversion occurs. And the proposed method detects abrupt and gradual shot boundaries with one algorithm. Moreover, it detects not only position where gradual shot conversion occurs, but also the exact duration where gradual shot conversion occurs.

**Key Words** - Abrupt Shot Boundary, Gradual Shot Boundary, Duration

### 1. 서 론

현대 사회는 정보화 사회이다. 따라서 필요한 정보를 적절하게 이용할 수 있다면 보다 큰 이득을 얻을 수 있다. 그래서 정보를 효율적으로 관리할 수 있는 디지털 도서관의 필요성이 대두된 것이다. 현재는 인터넷에서 텍스트를 이용하여 정보를 검색하는 검색 엔진들을 사용하고 있다. 그러나 요즘 수요가 많은 비디오 데이터베이스와 같은 멀티미디어에서는 단순한 텍스트 검색보다는 이미지의 특성과 영상의 정보를 이용한 검색이 더 효율적이다. 이것을 위해서는 비디오를 자동으로 분석하는 기능이 필요한데, 이것은 비디오 분할과 비디오 색인으로 나뉜다. 비디오 분할 단계에서 비디오 스트림은 연속적인 카메라 이동으로 얻어진 비디오 데이터 단위인 샷(shot)과 같은 의미를 가지는 연속된 여러 개의 샷이 모여서 이루어진 의미적인 데이터인 씬(scene)으로 분할된다. 이렇게 변화되는 샷을 찾는 것을 장면 전환 검출이라고 하며, 비디오 인덱싱(indexing), 장면 브라우징 등 비디오 내용 기반 검색을 위한 기본이 되는 핵심 기술이다. 요즘은 MPEG으로 압축된 영상에서 직접 장면 전환을 검출하는 방법에 초점을 맞추고 있다[1-3].

본 논문에서는 압축된 영상에서 급격한 장면 전환과 점진적인 장면 전환을 동시에 효과적으로 찾기 위해서 먼저 서로 다른 타입의 MPEG 비디오로부터 DC 영상 시퀀스라 불리는 공간적으로 축소된 영상 시퀀스를 구한다. DC 시퀀스

를 픽셀의 밝기를 이용하여 같은 샷이라고 예측되는 부분들로 분해한다. 분해된 각각의 부분들은 히스토그램을 이용해 같은 샷이라고 분류되면 하나로 병합한 후, 부분들의 지속성을 판별한다. 지속성을 고려해 급격한 장면 전환과 점진적인 장면 전환을 구별하였다. 기존의 다른 논문들은 장면 전환이 일어나는 부분만 찾았지만, 본 논문에서는 히스토그램 인터섹션(intersection)을 이용해 점진적인 장면 전환이 일어나는 구간까지 정확하게 찾는다.

### 2. 본 론

#### 2.1 DC 영상 시퀀스

이 논문의 구현을 위해서 먼저 서로 다른 타입의 MPEG 비디오로부터 DC 영상 시퀀스라 불리는 공간적으로 축소된 영상 시퀀스를 구성해야 한다. 이 DC 영상은 원영상의 64배로 축소된 영상들이다. 이 영상들이 큰 비율로 축소되었지만 여전히 영상 분석을 위한 중요한 정보들은 담고 있다. I 프레임에서 DC 영상을 얻는 것이 간단한 과정이지만, P 프레임이나 B 프레임의 경우는 간단하게 얻어지지 않는다. 우리는 MPEG 비디오의 I 프레임, P 프레임, B 프레임에서 DC 영상 시퀀스를 Yeo와 Liu's 알고리즘을 이용하여 구한다[1].

#### 2.2 영상의 픽셀 유사도를 이용한 분해

프레임  $F_1$ ,  $F_2$ 가 있다. 이때 프레임의 컬러 픽셀은  $x = [r \ g \ b]^T$ 로 나타나지만 RGB는 실제 영상을 표현하는데 효과적이지 않다. 그래서 색상을 나타내는 공간으로 RGB를 직접적으로 쓰지 않고 식 (1)과 같이 RGB로부터 얻어지는 보색축을 이용해 얻은  $x = [rg \ gy \ wb]^T$ 을 이용한다.

† 교신저자, 正 會 員 : 延世大學 電氣電子工學科 博士課程  
E-mail : wall@naver.com

\* 正 會 員 : 延世大學 電氣電子工學科 博士課程

\*\* 正 會 員 : 延世大學 電氣電子工學科 教授

\*\*\* 正 會 員 : 延世大學 電氣電子工學科 教授

接受日字 : 2005年 8月 10日

最終完了 : 2005年 9月 28日

$$\begin{aligned}rg &= R - G \\by &= 2B - R - G \\wb &= R + G + B\end{aligned}\tag{1}$$

두 프레임의 유사도를 알기 위해 두 프레임 간의 이미지 인터섹션을 다음과 같이 구한다.

$$I(F^1, F^2) = \frac{1}{n} \left\| \{ (x_i^1, x_i^2) \square d_p(x_i^1, x_i^2) \langle \theta_p \} \right\| \tag{2}$$

$d_p$ 는 두 프레임에서 동일한 좌표에 위치한 픽셀 값의 거리이고, 다음과 같이 유클리디언 거리로 구한다.

$$d_p(x^1, x^2) = \sqrt{(rg^1 - rg^2)^2 + (by^1 - by^2)^2 + (wb^1 - wb^2)^2} \tag{3}$$

$d_p$  값이 상위 문턱값  $\theta_p$  이하면 두 픽셀이 유사하다고 본다. 그래서 두 프레임 사이의 유사한 픽셀 영역이 어떤 값 이상이면 두 프레임의 유사도가 큰 것으로 보고 같은 장면이라고 추정한다.

이와 같이 영상 분해는 영상의 유사도를 이용한다. 연속적인 프레임에서 영상의 유사도가 적은 프레임을 찾아서 영상을 부분으로 분해한다. 하나의 부분인  $F^m$  을 구성하는  $f_1, f_2, \dots, f_m$  가 있다고 하자. 다음 프레임  $f_{m+1}$  이 그  $F^m$  에 속하는지 알아보기 위해서는  $F^m$  에 속하는 모든 프레임과의 유사도를 계산해야 하지만 계산량을 줄이기 위해 그 중 하나의 프레임과 유사도를 계산한다. 점진적인 장면 전환이 일어나는 부분에서는 이웃한 프레임들의 유사도가 비교적 크다. 그래서 이웃한 장면과 같은 장면으로 분류되기 때문에 점진적인 장면 전환을 찾지 못하는 경우가 발생한다. 따라서 우리는 그 프레임을 부분의 첫 번째 프레임인  $f_1$  으로 선택했다. 이와 같이 첫 번째 프레임과 유사도를 계산하면 점진적인 장면 전환이 일어나는 부분을 하나의 부분으로 분해할 수 있다. 그래서 다음 식을 만족하면  $f_{m+1}$  을  $F^m$  의 일부로 그렇지 않으면 다음 부분으로 분해된다[4].

$$I(f_1, f_{m+1}) > \theta_c \tag{4}$$

위의 영상 분해 기법을 이용하면 DC 입력 시퀀스는 유사한 프레임으로 구성된 부분으로 나누어지게 된다. 영상이 정적이면 부분의 길이도 길어지고, 동적이면 부분의 길이도 짧아진다.

### 2.3 분해된 영상의 병합

영상의 픽셀 유사도를 이용해 분해된 영상은 하나의 샷이 정확히 하나의 부분으로 분해되지 않는다. 일반적으로 하나의 장면은 카메라의 이동과 물체의 움직임 때문에 같은 장면임에도 여러 개의 부분으로 분해된다. 또한 그 움직임들이 빠를 경우 같은 장면임에도 많은 부분들로 분해된다. 따라서 이러한 부분들을 병합해야 한다. 그래서 움직임에 강한 특성을 나타내는 컬러 히스토그램 인터섹션[5]을 이용해 분해된 부분들을 병합한다.

영상을 분해할 때와 같이 색상을 나타내는 색상 공간으로 식 (1)과 같이 RGB로부터 얻어지는 보색 축을 이용한다. 영상의 색상 히스토그램을 구하기 위해서 세 축의 구간을

나눈다. 세기축인 wb가 광원으로부터의 거리 및 밝기의 변화 등에 민감하기 때문에 영향을 줄이기 위해 구간을 넓게 하였다. 그래서 본 논문에서는 rg, by는 32개의 구간으로 나누었으며 wb는 16개의 구간으로 나누어 히스토그램을 구한다. 히스토그램이 구해지면 두 영상의 색상 히스토그램은 히스토그램 인터섹션으로 비교된다. n개의 그룹을 가진 두 영상 I, M이 주어졌을 때 히스토그램 인터섹션은 식 (5)와 같이 정의된다.

$$\sum_{j=1}^n \min(I_j, M_j) \tag{5}$$

위 식의 결과는 두 영상 내에서 같은 색상 값을 가지는 화소수이다. 이 결과는 영상의 크기에 따라 값이 변하므로 식 (6)과 같이 정규화 시켜서 사용한다.

$$H(I, M) = \frac{\sum_{j=1}^n \min(I_j, M_j)}{\sum_{j=1}^n M_j} \tag{6}$$

만일 두 영상이 동일하다면 결과는 1이 되고 두 영상의 색상이 전혀 다르다면 0이 된다. 본 논문에서는 히스토그램 인터섹션이 문턱값  $\theta_h$  이상이면 유사한 영상이라고 본다.

DC 입력 영상이 식 (4)을 통해 n개의 부분으로 나누어진다고 가정하면  $F^1, F^2, \dots, F^n$  과 같이 구성될 것이다. 이때  $F^1$  과  $F^2$  가 같은 부분인지를 확인하기 위해서 다음의 방법을 이용한다.  $F^1$  에 k개의 프레임이 존재하고,  $F^2$  에 p개의 프레임이 존재한다면,  $F^1$  과  $F^2$  가 같은 장면인지 알아보기 위해서 각 부분에서 k 프레임을 추출해 히스토그램을 비교한다. k 프레임은 각 부분에서 하나가 아닌 두 개를 선택한다.  $F^m$  가 분해된 한 부분이라면, 각각  $F^m$  의 시작 프레임 근처에서 k 프레임  $f_m^1$ , 마지막 프레임 근처에서 k 프레임  $f_m^2$  을 정한다. 그래서  $F^{m-1}$  과 같은 장면인가를 비교할 때  $f_m^1, F^{m+1}$  과 같은 장면인가를 비교할 때  $f_m^2$  을 각각 k 프레임으로 이용한다. 이것은 병합하는 과정에서 한 부분의 구간이 클 경우 k 프레임에 따라서 같은 장면임에서 병합되지 못하는 경우를 발생하기 때문에 이를 막기 위한 것이다. 이 단계까지 수행하면 대부분의 장면 전환이 검출된다.

### 2.4. 장면 전환의 정제

영상의 분해와 병합 과정을 거치면 장면 전환이 일어난다고 짐작되는 곳을 대부분 찾는다. 그러나 어떤 프레임에서 급격한 장면 전환이 일어나는지, 어느 부분이 점진적인 장면 전환이 일어나는 구간(duration)인지는 알 수가 없다. 이번 단계에서는 이 문제를 해결할 것이다.

일반적으로 점진적인 장면 전환은 짧은 구간에서 발생한다. 하지만 지속성이 짧은 모든 부분이 점진적인 장면 전환이 일어나는 부분으로 볼 수는 없다. 지속성이 짧은 장면일 수도 있고, 같은 장면임에도 카메라의 빠른 움직임이나 물체의 빠른 움직임으로 인해 병합되지 못한 부분일 수도 있다. 그래서 일단은 영상의 병합까지 거친 결과에서 지속성이 임의의 값  $\theta_d$  (본 논문에서는 15로 두었다)보다 짧은 구

간은 점진적인 장면 전환이 일어날 수 있는 후보군으로 둔다. 그림 1에서 보면 B, E, F에서는 급격한 장면 전환이 일어난 곳이고 C에서 D까지는 점진적인 장면 전환이 일어날 수 있는 후보군으로 둔다.

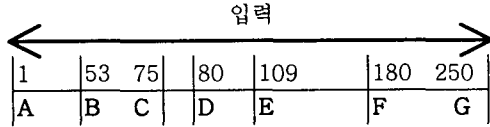


그림 1 2단계를 거치면서 부분으로 나누어진 입력영상

Fig. 1 Video Fragment Deposition after 2-Step

점진적인 장면 전환이 일어나는 구간을 찾는 과정은 다음과 같다.

$F^m$ 이  $\theta_d$ 보다 짧은 구간을 가진 부분이라면  $F^m$ 의 시작 프레임은  $F^{m-1}$ 의 마지막 프레임과  $F^m$ 의 마지막 프레임은  $F^{m+1}$ 의 시작 프레임과 식 (2)을 통해 유사도를 구한다. 일반적으로 점진적인 장면 전환은 서서히 일어나기 때문에 인접한 프레임간의 유사도는  $\theta_c$ 보다 크다. 따라서 두 곳의 유사도가 모두  $\theta_c$ 보다 크면,  $F^m$ 는 점진적인 장면 전환이 일어나는 구간이다. 두 곳의 유사도가  $\theta_c$  이하이면,  $F^m$ 은 지속성이 짧은 장면이며, 단지 한 곳만  $\theta_c$ 보다 크면 그 부분을 같은 장면으로 병합한다. 긴 시간 동안 점진적인 장면 전환이 일어나면 한 부분이 아닌 여러 부분으로 분해된다. 이를 해결하기 위해, 점진적인 장면 전환이 일어난다고 검출된 부분이 인접했을 때, 그 부분들을 하나의 점진적인 장면 전환이 일어나는 부분으로 병합한다. 따라서 점진적인 장면 전환이 일어나는 구간이 일반적인 경우보다 길지라도 검출이 가능하다.

2.5. 정확한 점진적인 장면 전환 구간 검출

이렇게 정제 과정까지 거치면 대부분의 장면 전환이 검출된다. 그러나 점진적인 장면 전환이 일어나는 곳은 정확하게 찾지만, 점진적인 장면 전환이 시작하는 프레임과 끝나는 프레임은 정확하게 찾지 못한다. 특히, 아주 서서히 변화하는 dissolve나 fade 같은 점진적인 장면 전환은 일부분만을 검출한다. Dissolve나 fade 같은 점진적인 장면 전환이 일어나는 전 구간을 정확하게 검출하기 위해서 점진적인 장면 전환이 일어나는 구간에서는 인접한 프레임 간의 히스토그램 인터섹션이 일정하다고 가정한다[6].  $F^m$ 이 정제 과정을 거친  $f^j, f^{j+1}, \dots, f^k$ 로 구성된 점진적인 장면 전환이 일어나는 구간이라고 하자. 먼저 다음과 같이 히스토그램 인터섹션의 평균을 찾는다.

$$H_{avr} = \frac{\sum_{i=j}^{k-1} H(f^i, f^{i+1})}{k-j} \quad (7)$$

정확한 점진적인 장면 전환 구간을 찾기 위해서는 두 방향을 고려해야 한다. 우선 프레임  $f^j$  이전 방향을 고려해야 하는데, 식(8)을 만족하면 그 프레임은  $F^m$ 의 부분이 된다.

식 (8)을 만족시키지 못하면 그 프레임은  $F^m$ 의 이전 장면으로 분류되고  $f^j$ 이전 프레임 방향의 과정은 멈춘다. 두 번째가  $f^k$ 이후 프레임 방향이다. 이것 역시  $f^k$ 이후 방향으로 식 (9)를 이용해 같은 방법으로 점진적인 장면 전환 구간을 찾는다.

$$\min(H(f^m, f^{m-1}), H_{avr}) / \max(H(f^m, f^{m-1}), H_{avr}) > \theta_k$$

$$m = j, j-1, j-2, \dots \quad (8)$$

$$\min(H(f^n, f^{n+1}), H_{avr}) / \max(H(f^n, f^{n+1}), H_{avr}) > \theta_k$$

$$n = k, k+1, k+2, \dots \quad (9)$$

이 과정을 통해 점진적인 장면 전환이 일어나는 곳 뿐만 아니라 그 구간까지 검출이 가능해진다.

3. 결 과

실험 데이터는 한국 MBN방송, 영국 BBC방송, 미국 CNN방송으로 급격한 장면 전환과 점진적인 장면 전환이 모두 포함된 총 90분 길이의 320 × 240 MPEG-2 영상을 사용하였다.

우리가 제안한 방법으로 실험할 때 최적의 결과를 얻기 위해서 다음과 같은 문턱값을 적용하였다. 우선 분해 과정에서 사용된 문턱값  $\theta_j$ 는 50,  $\theta_c$ 는 0.93으로 하였다. 병합 과정에서의 문턱값  $\theta_h$ 은 0.83을, 정제 과정에서의  $\theta_s$ 는 15를 그리고 점진적인 장면 전환의 구간을 찾는 데 필요한  $\theta_k$ 는 0.95로 하였다. 본 논문에서 제시한 방법의 실험 결과는 급격한 장면 전환과 점진적인 장면 전환의 검출에 대한 리콜(recall)과 정확도(precision)로 나타낸다. 그리고 검출된 점진적인 장면 전환 구간이 실제 장면 전환이 일어나는 구간의 어느 정도까지 검출하는 지를 나타낸다.

표 1은 실험 데이터에서 본 논문이 제안한 방법을 이용해 급격한 장면 전환과 점진적인 장면 전환을 검출한 결과이다. 실험 데이터에는 744개의 급격한 장면 전환과 219개의 점진적인 장면 전환이 존재했다. 급격한 장면 전환의 경우, 리콜은 0.98, 정확도는 0.96의 결과를 보였다. 점진적인 장면 전환의 경우 대부분 전환이 일어나는 구간까지 잘 검출하지만, 아주 미세한 변화로 장면이 전환할 때는 그 부분에서 장면이 전환한다는 것은 찾지만 점진적인 장면 전환이 아닌 그 구간 중 한 프레임에서 급격한 장면 전환이 일어난 것으로 검출했다. 총 219개의 점진적인 장면 전환 중에서, 이와 같이 점진적인 장면 전환이 급격한 장면 전환이 일어난 것으로 검출되는 경우가 11.7%였다. 표 1에서는 점진적인 장면 전환이 일어나는 구간은 찾지 못하지만 장면 전환은 검출했으므로 장면 전환을 찾은 것으로 인정했다. 그래서 점진적인 장면 전환의 경우, 리콜이 0.92, 정확도가 0.9이었다. 표 2는 본 논문이 제안한 방법을 이용해 점진적인 장면 전환이 일어나는 구간을 얼마나 정확하게 찾는지 나타낸 결과이다. 이 결과는 실제로 점진적인 장면 전환이 일어나는 구간을 미리 조사해서 실험에서 검출된 점진적인 장면 전환이 일어나는 구간이 얼마나 정확한 지를 나타낸다. 단지 장면 전환의 정제 과정까지만 실행했을 때는 실제 구

간의 68%만을 검출하였다. 그러나 일차적으로 검출한 구간의 히스토그램 인터섹션의 평균을 이용한 결과는 실제 점진적인 장면 전환 구간의 88%까지 검출이 가능했다.

표 1 장면 전환 검출 결과

Table 1 Shot Change Detection Results

	Total	Hit	Missed	False Alarm	Recall	Prec.
급격한 장면전환	744	731	13	29	0.98	0.96
점진적인 장면전환	219	201	18	23	0.92	0.9

표 2 점진적인 장면 전환 구간 검출 결과

Table 2 Duration of Gradual Shot Change Detection Results

	장면전환 정제 과정만 거친 결과(%)	히스토그램 인터섹션을 이용해 구한 최종 결과(%)
검출된 점진적인 장면전환 구간	68	88

#### 4. 결 론

본 논문에서는 압축된 비디오 영상에서 하나의 알고리즘으로 급격한 장면 전환과 점진적인 장면 전환을 동시에 찾을 뿐만 아니라 점진적인 장면 전환 구간을 정확하게 찾는 알고리즘을 보여주었다. 제안한 방법에서는 일단 급격한 장면 전환과 점진적인 장면 전환이 일어나는 곳을 찾았다. 그리고 찾은 점진적인 장면 전환 부분의 히스토그램 인터섹션을 이용해 점진적인 장면 전환이 일어나는 구간을 보다 정확하게 찾을 수 있었다.

#### 참 고 문 헌

- [1] B. Yeo and B. Liu, "Rapid scene analysis on compressed video", IEEE Trans. Circuits Syst. Video Tech., vol 5, pp. 533-544, 1995,
- [2] Y. Nakajima, K. Ujihara, and A. Yoneyama, data domain", Visual Commun. and Image Processing, vol. 3024, pp. 992-1003, 1997.
- [3] Oliver Bao and Ling Guan, "Scene Change Detection using DC Coefficients", IEEE ICIP, pp. 421-424, 2002.
- [4] S. Satoh, "New video analysis based on identical shot detection", IEEE Multimedia and Expo, Vol. 1, 26-29 Aug. 2002.
- [5] M. J. Swain and D. H. Ballard, "Color Indexing", International Journal of Computer Vision, vol. 7, pp. 11-32, 1991.
- [6] W. A. C. Fernando, C. N. Canagrajah, D. R. Bull, "Fade-in and fade-out detection in video sequences using histograms" IEEE International Symposium on circuits and Systems. vol. 4, pp 709-712, 2000.

## 저 자 소 개



### 권철현 (權哲鉉)

1975년 1월 9일생. 1997년 연세대 전기공학과 졸업. 2000년 동 대학원 전기전자공학과 졸업(석사). 2000년~현재 동 대학원 전기전자공학과 박사과정



### 한두진 (韓杜辰)

1976년 10월 25일생. 1999년 연세대 전기전자공학과 졸업. 2001년 동 대학원 전기전자공학과 졸업(석사). 2001년~현재 동 대학원 전기전자공학과 박사과정



### 이명호 (李明鎬)

1946년 6월 11일생. 1972년 연세대 전기공학과 졸업. 1974년 동 대학원 전기공학과 졸업(석사). 1978년 동 대학원 전기공학과 졸업(박사). 1980년~현재 연세대학교 전기전자공학과 교수



### 박상희 (朴相暻)

1939년 8월 25일생. 1962년 연세대 전기공학과 졸업. 1971년 동 대학원 전기공학과 졸업(석사). 1970년~2004 연세대학교 전기전자공학과 교수