

## Video 장면전환 중 디졸브 검출에 관한 연구

박성준<sup>†</sup>, 송문호<sup>†</sup>, 곽대호<sup>†</sup>, 김운경<sup>†</sup>, 정민교<sup>‡</sup>

<sup>†</sup> 고려대학교 전기전자전파 공학부 통신신호처리 연구실

<sup>‡</sup> 한국통신 멀티미디어 연구소 영상미디어 연구실

전화 : (02) 927-0560 / 팩스 : (02)927-0544

## Automatic Detection of Dissolving Scene Change in Video

Sung-Jun Park, S. Moon-Ho Song, Woon-Kyung M. Kim, and Min-Gyo Chung

School of Electrical Engineering, Korea University  
Multimedia Tech. Research Laboratory, Korea Telecom

e-mail : sjpark@poisson.korea.ac.kr

### Abstract

For efficient storage and retrieval of large video data sets, automatic video scene change detection is a necessary tool. Video scene changes fall into two categories, namely fast and gradual scene changes. The gradual scene change effects include, dissolves, wipes, fades, etc. Although currently existing algorithms are able to detect fast scene changes quite accurately, the detection of gradual scene changes continue to remain a difficult problem. In this paper, among various gradual scene changes, we focus on dissolves. The algorithm uses a subset of the entire video, namely the sequence of DC images, for improvement of detection velocity

### I. 서론

멀티미디어와 네트워크의 향상으로 인하여 디지털 비디오는 미디어 중 가장 중요하고 유용한 멀티미디어 매체로서의 역할을 하게 되었다. 그러나 대용량과 계산의 복잡성으로 인하여 디지털 비디오는 쉽게 다루기 힘든 매체로 인식되고 있는 것 또한 사실이다. 그러므로 이러한 비디오의 대용량성을 효율적으로 다루기 위해서는 적절한 기준을 통하여 효과적으로 저장하고 검

색할 수 있는 색인화 과정을 필요로 하는데, 이는 빠르고 정확한 자동 비디오 분석기의 발달을 요구한다.

자동 비디오 분석기는 보통 여러 개의 부분적인 알고리즘으로 나뉘게 되는데 부분 알고리즘은 다시 급격한 장면 전환과 점진적인 장면 전환을 검출하는 알고리즘으로 나뉘어 진다. 점진적인 장면전환의 형태를 자세히 살펴보면 다음과 같은 특징으로 또 나눌 수 있다. 첫 번째로 이전장면에서 다음 장면으로 선형적으로 바뀌는 디졸브, 두 번째로 비오는 날 자동차의 와이퍼가 빗물을 훔쳐내듯이 바뀌는 와이프, 세 번째로 영화의 첫 시작 부분과 같은 곳에서 점점 밝아지거나 반대로 점점 어두워지는 페이드 인/아웃이 있다.

최근의 비디오 분석기 논문을 살펴보면 장면 전환을 검출하는 방식에 있어서 크게 두 그룹으로 나뉜다. 하나는 복원된 상태(비 압축 영역)의 raw image에서 접근하는 방식으로 histogram, edge 등의 영상의 특성을 사용한다. 또 하나는 압축 영역에서 접근하는 방식으로서 DC 영상[1], 혹은 DC영상과 약간의 AC 계수를 사용하는 방식, 움직임 벡터, Macroblock 데이터의 양을 이용한다.

이러한 장면 변환 검출 알고리즘에 관하여 Yeo[1], Nakajima[2] 등의 논문에서 볼 수 있듯이 몇몇 알고리즘이 나와있는데, 전반적으로 급격한 장면전환에서는 비교적 잘 검출되지만, 점진적인 장면전환 검출시 검출율이 떨어지고 점진적인 장면전환이 아닌 구간을 점진적인 장면전환으로 오인하여 검출하는 경우 또한 나

타난다. 또한 계산의 복잡성과 대용량으로 인하여 검출 속도에서도 그 문제성이 제기된다. 그러므로 점진적인 장면전환의 경우에 있어서 검출율이 높고 잘못 검출하는 경우를 줄일 수 있는 알고리즘의 필요성이 대두된다.

본 논문에서는 DC 영상만을 추출하여 빠르고 정확성이 높은 점진적인 장면전환 검출 알고리즘을 제안한다. 본 논문은 다음과 같은 구성을 따른다. 2장에서는 점진적 장면전환 검출 알고리즘을 설명하고 3장은 그에 대한 실험 결과를 알아본다. 4장에서는 간단한 결론을 고찰할 것이다.

## II. 점진적 장면전환 검출 알고리즘

실시간으로 장면전환 검출 장면전환을 정확히 검출할 수 있는 효과적인 알고리즘 설계를 위해 전체 영상에서 필수적인 부분만을 추출하여 재정렬 할 필요가 있다. 즉 전체 영상의 정보를 대부분을 가지고 있으면서도 전체영상의 비해 그 크기가 훨씬 작아 알고리즘의 빠른 구현을 실현하는 것이 가능해야 한다. 이를 만족하기 위해서 영상의 DC 성분을 추출하여 형성된 Thumbnail sequence가 가장 적합한 경우임을 보일 것이다.

대부분의 디지털 비디오는 8X8 블록 단위로 DCT 코딩되어 있는데 DCT 코딩에서 각각의 블록은 하나의 DC계수와 63개의 AC계수로 형성된다. Yeo의 논문 [1]을 참조하면 DC성분은 블록의 평균 픽셀값의 8배와 같으므로 각각의 프레임에서 DC성분을 취한다면 전체 데이터의 64배를 줄일 수 있다. 또한 영상의 특성상 낮은 주파수 성분에 대부분의 정보를 가지고 있으므로 DC성분만을 취하더라도 정보의 손실량은 크게 줄지 않는다. 이렇게 각각의 프레임에서 DCT의 DC 성분인 (0,0)성분만을 추출하여 시간에 관해 집합을 형성한 것을 thumbnail sequence라고 본 논문에서 정의한다. thumbnail sequence형성은 복잡한 역 DCT 과정 없이 가능하다.

DC 추출과정을 이용하여 만든 thumbnail sequence는 원영상의 대부분의 정보를 가지고 있으므로 장면전환의 특성 역시 가지고 있다. thumbnail내에서 장면전환은 각각의 종류에 따라 다른 형태로 나타나는데 그림 1을 참조하여 살펴보기로 한다. 그림 1은 4개의 장면과 3개의 장면전환으로 이루어진 thumbnail sequence이다. 각각의 장면은 다른 그림과 색으로 이루어져 있으므로 각각의 장면마다 픽셀값이 크게 달라지게 된다. 그러므로 픽셀 값의 변화가 발생하는 곳에서 장면 전환을 찾을 수 있다. 장면 1에서 본다면, 급격한 장면전환은 수직수평방향으로 직각이고 불연속적

인 직사각형을 만나게 된다. (thumbnail sequence는 3차원의 직육면체 형태이므로) 또한 점진적인 장면 전환에서 와이프의 경우에는 장면전환은 한 프레임이 아닌 여러 프레임에 걸쳐 나타나게 되는데 한 장면이 다른 장면에 의해 밀리면서 나타나므로 thumbnail sequence의 장면 2에서 보면 수직 혹은 수평에서 비스듬하고 불연속적인 직사각형 혹은 상황에 따라 비스듬하고 불연속적인 곡면을 만나게 된다. 디졸브의 경우에는 장면 3과 장면 4 사이에서 발생하는데, 장면 3이 픽셀 값이 선형적으로 감소하여 사라지고 반대로 장면 4는 선형적으로 증가하여 나타나는 특징이 있다.

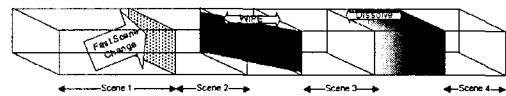


그림 1 Thumbnail 시퀀스에서의 장면 전환 발생 형태

이와 같은 장면전환들의 특성을 이용하면 간단한 수학적 방식으로 장면전환이 검출 가능하다. thumbnail sequence에서 급격한 장면전환과 와이프 검출 알고리즘은 본 연구소에서 제안한 논문[3]에 실려 있다.

디졸브 검출 알고리즘을 자세히 살펴보면 다음과 같다. 디졸브가 시간 0에서 T까지 T시간만큼 지속 될 때, 장면 3은 T시간동안 선형적으로 감소하고 장면 4는 T시간동안 선형적으로 증가한다. 만약 움직임이 없는 이상적인 경우의 디졸브를 고려하면 장면 3의 픽셀 값은  $f_1(x, y)$ 으로 장면 4는  $f_2(x, y)$ 로, 시간축에 상관 없는 2차원의 함수로 나타낼 수 있다. 그러나 실제 영상에서는 움직임이 나타나므로 장면 3과 장면 4는 시간의 영향이 나타나는 3차원 함수가 된다 그러므로 장면 3과 장면 4를

$$\begin{aligned} f_1(x, y, t) &= f_1(x, y) + \delta_1(x, y, t), \quad 0 \leq t \leq T \\ f_2(x, y, t) &= f_2(x, y) + \delta_2(x, y, t), \quad 0 \leq t \leq T \end{aligned} \quad (1)$$

의 3차원 함수로 생각하고, 디졸브 구간에 관한 수식적 모델을 세워 정리하면 식(2)의 형태로 나타난다[4].

$$\begin{aligned} g(x, y, t) &= (1 - \frac{t}{T})f_1(x, y, t) + \frac{t}{T}f_2(x, y, t), \\ &= \frac{t}{T}(f_2(x, y) - f_1(x, y)) + f_1(x, y) + v(x, y, t), \\ &= a(x, y)t + b(x, y) + v(x, y, t), \quad 0 \leq t \leq T \end{aligned} \quad (2)$$

$$\text{where } v(x, y, t) = \delta_1(x, y, t) + \frac{t}{T}(\delta_2(x, y, t)f_2(x, y) - \delta_1(x, y, t)f_1(x, y))$$

식(2)를 살펴보면 기울기  $a(x, y) = \frac{f_2(x, y) - f_1(x, y)}{T}$ , 절편  $b(x, y) = f_1(x, y)$  그리고 움직임으로 인한 오차  $v(x, y, t)$ 로 1차함수의 선형적인 모델이 가능하게 된다.

$$\begin{bmatrix} g(x, y, t-L) \\ g(x, y, t-L+1) \\ \vdots \\ g(x, y, t) \\ g(x, y, t+1) \\ \vdots \\ g(x, y, t+L) \end{bmatrix} = \begin{bmatrix} t-L & 1 \\ t-L+1 & 1 \\ \vdots & \vdots \\ t & 1 \\ \vdots & \vdots \\ t+L & 1 \end{bmatrix} \begin{bmatrix} a(x, y) \\ b(x, y) \end{bmatrix} + \begin{bmatrix} v(x, y, t-L) \\ v(x, y, t-L+1) \\ \vdots \\ v(x, y, t) \\ \vdots \\ v(x, y, t+L) \end{bmatrix} \quad (3)$$

이를 행렬의 형태로 정리하면

$$g(x, y, t) = Hz(x, y) + v(x, y, t) \quad (4)$$

가 되어

$$\|v(x, y, t)\|^2 = \|g(x, y, t) - Hz(x, y, t)\|^2 \quad (5)$$

를 최소화하는  $z(x, y) = \begin{bmatrix} a(x, y) \\ b(x, y) \end{bmatrix}$ 를 구한다. 일반적으로  $z$ 의 해는

$$z(x, y) = (H^T H)^{-1} H^T g(x, y, t) \quad (6)$$

가 된다[5]. 여기서  $T$ 는 행렬을 transpose 한 것이다. 위와 같은 식(6)을 각각의 프레임마다 구할 수 있으므로  $z(x, y)$ 는 시간축에 관한 3차원 함수  $z(x, y, t)$ 로 놓을 수 있다. 이제 디졸브를 검출하기 위하여 각각의 기울기 값  $a(x, y, t)$ 을 절대값으로 바꾸어 (절대 기울기) 프레임단위로 총 합을 구한다. 이는 디졸브 구간에서 기울기가 존재하므로 디졸브가 아닌 구간에 비해 절대 기울기의 총 합이 크게 나타난다는 특성을 이용한 것이다. 그러나 실제 영상에서는 카메라 움직임이나 영상 객체의 움직임으로 인하여 에러 요인이 발생한다. 이로 인하여 디졸브가 아닌 구간에서 기울기가 존재하게 되므로 단지 기울기 측정만으로는 정확한 검출을 기대하기 어렵다. 그러므로 이러한 에러 요인을 제거할 수 있는 또 다른 요소가 필요하게 된다.

본 논문에서는 이러한 요소로서 기울기 오차를 이용한다. 기울기 오차 측정은 식(5)에서 나온 오차를 크기  $\sqrt{2L+1}$ 로 normalize하는  $l_2$  norm을 이용한다.

$$e(x, y, t) = \frac{\|v(x, y, t)\|^2}{\sqrt{2L+1}} \quad (7)$$

위 식들을 이용하여 구한 절대 기울기 요소  $|a(x, y, t)|$ 와 오차요소  $e(x, y, t)$ 를 이용하여 디졸브 구간의 기울기를 측정하고 에러 요인을 제거한다. 이를 구현하는 방법은 다음과 같다. 절대 기울기 요소에 오차 요소를 나누어준 후, 프레임 별로 총 합을 구함으로써 각 프레임을 대표하는 기울기를 정한다. 이를 수식적으로 구현하기 위해서 식 (8)을 이용한다.

$$s(t) = \sum_{x=0}^N \sum_{y=0}^M \frac{|a(x, y, t)|}{e(x, y, t)} \quad (8)$$

여기에서 thumbnail의 한 프레임은  $N \times M$ 의 2차원 행렬로 이루어져 있다고 하고 각 프레임의 오차보정 기울기를  $s(t)$ 라고 놓는다.

식(8)에서 구한 오차보정 기울기  $s(t)$ 은 디졸브 구간에서 그렇지 않은 구간에 비해 큰 값을 가지게 될 것

이다. 그러므로 오차보정 기울기 값이 특정 임계치보다 크게 나오면 이를 디졸브 구간으로 선언한다.

### III. 실험결과

제안된 알고리즘은 디졸브를 갖고있는 몇 개의 영상 클립에서 테스트되었다. DC계수만 추출하여 형성한 thumbnail의 성능을 알아보기 위하여 전체영상과 thumbnail 영상의 비교를 그림 2에 나타내었다.

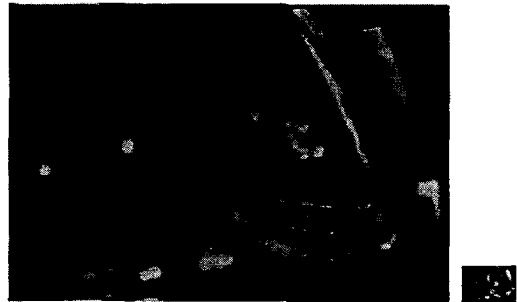


그림 2 전체 영상과 DC값만을 추출한 thumbnail 영상

그림 2에서 보듯이 전체 영상에 비해 크기는 상당히 작지만 객체의 형태나 픽셀 값 등 전체영상의 정보를 대부분 가지고 있음을 알 수 있다.

2장에서 제안한 알고리즘을 테스트하기 위하여 sunrise라는 디졸브가 포함되며 영상의 움직임이 적은 비디오 클립에 실험을 적용해 보았다. 실제 디졸브가 일어나는 프레임 number는 다음 표 1과 같다.

디졸브	구간	길이
1	15-42	28
2	68-82	15
3	117-129	13

표 1 Sunrise 영상에서 실제 디졸브가 일어나는 프레임 number와 구간

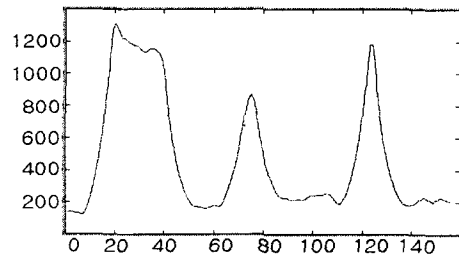


그림 3 디졸브 검출 알고리즘을 Sunrise 영상에 적용하여 얻은 결과

그림 3은 2장에서 제안한 디졸브 검출 알고리즘을

처럼 디졸브 구간에서 다른 구간에 비해 큰 피크 값이 나타남을 알 수 있다. 특히 첫 번째 디졸브와 같이 디졸브 구간이 긴 부분에서는 긴 plateau와 같은 부분이 나타난다.

다음은 디졸브를 포함하고 카메라 혹은 비디오 내의 객체의 움직임이 존재하는 mytoday라는 비디오 클립에 대한 알고리즘에 대한 적용 결과이다. 디졸브 위치는 다음 표 2와 같다.

디졸브	구간	길이
1	104-114	11
2	168-175	8
3	241-256	16
4	321-336	16

표 2 Mytoday 영상에서 실제 디졸브가 일어나는 프레임 number와 구간

다음은 역시 제안한 디졸브 검출 알고리즘을 이용하여 디졸브를 검출하였다. 그림 4에서는 기울기 오차 특성의 성능을 테스트하기 위하여 두 가지 그래프를 나타내었다. 그림 4에서 점선은 기울기 오차 요소를 나누지 않고 단순히 기울기 합만을 더한 경우이고 실선은 본 논문에서 제안한 방식과 같이 기울기 오차 요소를 나누어 합을 더한 경우이다.

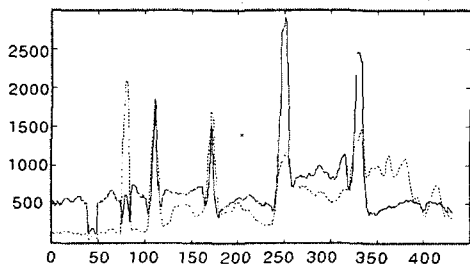


그림 4 디졸브 검출 알고리즘을 Mytoday 영상에 적용한 결과

표 2를 참조하면 총 4개의 디졸브가 존재하는데 실선의 경우에는 총 4개의 peak가 존재하고 구간도 역시 표 2의 디졸브 구간과 일치함을 알 수 있다. 그러나 점선의 경우에는 80프레임 근처에서 불필요한 peak가 존재하고 세 번째, 네 번째 디졸브에서 peak가 크지 않으며, 또한 디졸브 구간이 없는 350 프레임 이후에도 약간의 leakage가 발생함을 알 수 있다. 실제로 mytoday 비디오 클립에서는 80프레임에서 급격한 장면전환이 나타나고 350프레임 이후에는 사람의 움직임이 크게 차지하고 있어 기울기 측정 요소가 크게 나타났다. 이상과 같이 살펴보았을 때 디졸브를 검출하기 위해서는 기울기 오차 요소가 필수적임을 알 수 있다.

#### IV. 결론

본 논문에서는 멀티미디어 영상자료의 체계적인 정리와 비디오 데이터베이스 구축을 위하여 사용할 수 있는 장면 전환 검출 알고리즘에 대하여 논하였다. 본 논문에서 제안한 알고리즘은 raw image 형태의 비디오나 Motion JPEG 이나 MPEG등 여러 디지털 비디오 표준에 다양하게 이용되어 질 수 있다. 현재 점진적인 장면 전환 검출에 관한 적절한 알고리즘이 없는 상태이므로 본 논문에서 제안한 디졸브 검출 알고리즘은 비디오 데이터베이스 구축을 위해 좋은 참조가 될 것으로 기대된다.

이와 같이 기존의 알고리즘을 구현, 분석하고 새로운 알고리즘을 개발하였으며 이의 실험과 성능평가를 하였다.

#### 감사의 글

본 연구는 한국통신의 정보통신 기초 연구비(과제 번호 98-18)에 의하여 지원되었음.

#### 참고문헌

- [1] B. L. Yeo, B. Miu, "Rapid scene analysis on compressed video", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5. no. 6, pp.533-544, 1995
- [2] Y. Nakajima, K. Ujihara, A. Yoneyama, "Universal scene change detection on MPEG-coed data domain", in *Proc. SPIE Visual Comm. and Image Proc.*, pp. 992-1003, 1997
- [3] H. Kim, S. Park, J. Lee, W.M Kim, and S. M. Song, "Processing of partial video data for detection of wipes", in *Proc. SPIE Storage and Retrieval for Image and Video Databases VII*, vol. 3656, 1999
- [4] S. M. Song, T. H. Kwon, W. M. Kim, H. Kim, and B. D. Rhee, "On detection of gradual scene changes for parsing of video data", in *Proc. SPIE Storage and Retrieval for Image and Video Databases VI*, vol. 3312, San Jose, CA, pp. 404-413, 1998.
- [5] Jerry M. Mendel "Lessons in Estimation Theory for Signal Processing, Communications, and Control", Prentice Hall, International Editions, pp. 27-31, 1995.