

MPEG 압축 영역에서 B 프레임의 특징을 이용한 점진적 장면전환 검출에 관한 연구

김중헌, 박두영, 장중환
배재대학교 정보통신공학과

e-mail : joongheon@mail.pcu.ac.kr , dypark@mail.pcu.ac.kr , jangjw@mail.pcu.ac.kr

A Study of The Gradual Scene Change Detection Using The Features of B Frame in Compressed MPEG Videos

Joong-Heon Kim, Doo-Yeong Park, Jong-Whan Jang
Dept. of Information & Communication Engineering, PaiChai University

요 약

비디오 데이터의 효율적인 저장, 관리를 위해서는 장면전환 검출을 통한 비디오 분할 기술에 대한 연구가 필요하다. 기존의 장면전환 검출 알고리즘은 복호화에 의한 연산량 및 프레임들간의 비교에 의한 연산량이 많아 속도에 제한을 받았다. 본 논문에서는 MPEG 압축 비디오에서 효과적인 장면전환 검출을 위해 MPEG 압축도메인에서 B 프레임들의 매크로블록타입의 통계적 특성변화를 이용하여 점진적 장면전환검출을 빠르고 정확하게 할 수 있는 알고리즘을 제안한다.

1. 서론

장면전환 지점 검출(scene change detection)은 동영상 데이터의 내용기반 검색을 위해 필요한 기초 작업으로써 동영상 데이터로부터 장면들을 구분하기 위해 컷(cut)이라 불리는 전환 지점들을 자동으로 찾아 내는 작업을 말하며 이를 바탕으로 동영상 데이터에 대한 색인과 검색이 이루어지게 되므로 효과적인 검출 방안이 반드시 필요하다. 비디오는 여러 장면으로 구성되어 있기 때문에 각 장면이 바뀌는 부분을 검출함으로써 비디오 분류를 효과적으로 할 수 있다[1]-[5].

장면전환 종류에 따라 급격한 장면전환, 점진적인 장면전환 등으로 구분할 수 있다. 급격한 장면전환은 한 프레임만에 순간적으로 장면이 변하는 것이고, 점진적인 장면전환은 여러 프레임에 걸쳐 서서히 장면이 전환되는 것이다.

장면전환 검출에 있어 가장 일반적으로 널리 알려진 방법은 그레이레벨(gray-level) 히스토그램 차이를 비교하여 정해진 임계값을 초과하는 경우 이를 컷

로 검출하는 방법이다[2]. 히스토그램 기반의 방법들은 구현이 쉬우며 적은 연산량으로도 효과적인 검출을 하는 장점을 갖는 반면 장면내의 명암의 변화에 민감하기 때문에 컷이 아님에도 이를 컷으로 검출해 내는 오검출(fault cut)이 자주 발생하는 문제점을 갖는다. 이러한 문제점을 보완하기 위해 움직임 벡터(motion vector)를 이용하거나[3], 예지정보[4], 화소 값들간의 차이[5]등을 이용하는 방법들이 제안된 바 있다. 그러나 이러한 기법들을 사용함으로써 급진적 변화는 비교적 쉽게 검출할 수 있으나 여러가지 영상 처리 효과를 통해 한 장면에서 다른 장면으로 서서히 변화하는 점진적 장면전환을 검출하는 것은 아직 미흡한 실정이다.

MPEG 으로 압축된 동영상은 모션 예측이라는 과정을 통하여 압축률을 높이는 데 이 과정에서 I 또는 P 프레임, P 또는 P 프레임 사이에 존재하는 B 프레임들의 매크로블록들은 참조할 영상을 가리키게 되고 이것은 매크로블록 타입 정보에 나타나게 된다.

본 논문에서는 이러한 특징을 이용하여 점진적 장면전환 시 모션 예측하는 프레임들에 나타나는 매크로블록 타입의 변화를 이용하여 점진적 장면전환을

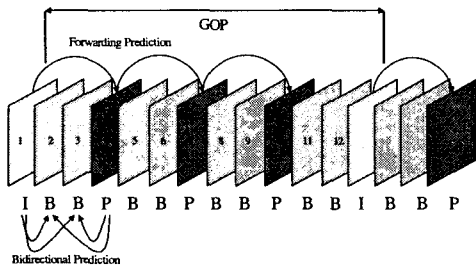
이 논문은 2002년도 두뇌한국 21 사업에 의하여 지원되었음

효과적으로 검출할 수 있는 알고리즘을 제안한다.

본 논문의 구성은 먼저 2 장과 3 장에선 MPEG 동영상의 구조 및 점진적인 장면전환의 특성을 소개하고 4 장에서는 모션 예측하는 프레임들에서 나타나는 특징을 측정하여 점진적인 장면전환을 검출하는 알고리즘을 기술한다. 5 장에서는 제안하는 기법을 동영상 데이터를 대상으로 실험한 결과를 보이며 마지막으로 6 장에서 결론과 향후 과제를 제시한다.

2. MPEG 동영상의 구조

MPEG 에서는 동영상의 시공간적인 중복성을 제거하기 위하여 MC-DCT(Motion Compensation-Discrete Cosine Transform)압축 방식을 이용한다. 간단히 MPEG 구조를 살펴보면 가장 상위 레벨에는 비디오 시퀀스(Video Sequence)가 있고, 그 아래 동영상 편집을 고려한 GOP(Group Of Picture)가 있다. GOP 는 프레임들의 집합을 나타내며 각각 I(Intra Picture), P(Predictive Picture), B(Bidirectionally Predictive Picture)로 구분되어 있고 각 프레임마다 다른 방식으로 부호화된다. 프레임은 16×16 크기의 매크로 블록을 기준으로 움직임 벡터 정보를 포함한다. 매크로 블록은 DCT 변환 단위인 8×8 블록으로 구성된다. 한 매크로 블록은 밝기 정보를 포함하고 있는 4 개의 Y 블록과 색 정보를 포함하고 있는 2 개의 Cb, Cr 블록을 가지고 있다. MC-DCT 압축 방식에 따라 압축된 MPEG 비트열은 각 계층적인 구조들은 나타내는 헤더정보, 각종 파라미터, DCT 계수, 움직임 벡터 등이 VLC(Variable Length Code)된 결과이다. 장면전환 검출은 프레임간의 컷이나 점진적인 변환(Gradual Transition)이 존재하는지를 판단하는 것이므로 프레임 단위에 대한 이해가 필요하다. 프레임은 I, P, B 프레임 세 종류로 되어 있는데 각각 다른 부호화 기술이 적용된다. I 프레임은 디코딩시 기준이 되는 프레임인데, P, B 프레임으로 인해 발생하는 에러를 줄이기 위해서 DCT 양자화를 통해 공간중복성만은 제거한다. P, B 프레임은 참조 프레임과의 차영상(Residual Image)과 움직임 벡터(Motion Vector)로 시간 중복성을 제거한다.



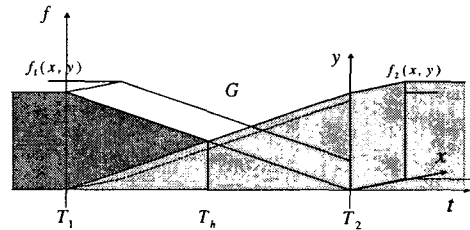
[그림 1] I, P, B 프레임 화면의 예측방법

3. 점진적 장면전환의 특성

점진적인 장면전환은 한 장면에서 다른 장면으로

다양한 효과를 사용하여 서서히 변하는 것으로 인접 프레임간의 특징 비교만을 가지고는 이를 발견해 내기가 어렵다. 일반적으로 점진적인 전환은 약 30 프레임 정도에 걸쳐 일어나는 것으로 알려져 있으며 상황에 따라 이보다 적거나 또는 많은 프레임수로 빠르게 진행되기도 한다. 점진적인 전환 효과의 종류로는 빈 화면에서 서서히 장면이 삽입되는 페이드인(fade-in), 반대로 빈 화면으로 서서히 사라지는 페이드아웃(fade-out), 두 장면이 천천히 오버랩(overlap)되면서 바뀌는 디졸브(dissolve), 한 장면이 다른 장면을 밀어내는 식인 와이프(wipe) 등이 있다.

본 논문에서는 점진적인 전환 효과들 중 디졸브 효과에서 점진적인 장면전환을 검출하는 부분을 다루기로 한다. 디졸브 효과들은 전환 효과가 일어나는 동안 서로 다른 장면의 일부가 동시에 화면에 나타나는 특성을 갖는다. 먼저 일반적인 디졸브 효과의 장면전환 형태를 살펴보면 [그림 2]과 같다.



[그림 2] 동영상 내의 디졸브 계열 장면전환의 예

[그림 2]은 디졸브 계열의 점진적 장면전환이 포함된 경우를 보여준다. 구간 $t < T_1$ 과 $t > T_2$ 는 각각 다른 일반적인 장면이며 구간 $G(T_1 \leq t \leq T_2)$ 는 점진적인 전환이 일어나는 구간으로 가정한다.

점진적인 장면전환에 대한 수학적 모델은 정의하기 위해 구간 $t < T_1$ 에 있는 프레임을 $f_1(x, y, t)$, 구간 $t > T_2$ 에 있는 프레임을 $f_2(x, y, t)$ 라 가정할 때 디졸브는 식(1)과 같이 표현할 수 있다. 이때 $\alpha(t)$ 는 감소함수이고 $\alpha(T_1)=1$ 과 $\alpha(T_2)=0$ 의 값을 갖으며 $\beta(t)$ 는 증가함수로서 각각 $\beta(T_1)=0$ 과 $\beta(T_2)=1$ 의 값을 갖는다. 또한 디졸브의 전환 조건으로 $\alpha(t)+\beta(t)=1$ 의 조건을 가정한다.

$$f(x, y, t) = \alpha(t)f_1(x, y, t) + \beta(t)f_2(x, y, t), (T_1 \leq t \leq T_2) \quad (1)$$

4. 제안한 알고리즘

[그림 2]에서 $t < T_1$ 구간에서의 $f_1(x, y, t)$ 는 같은 scene에 존재하는 프레임들의 집합이므로 t 의 증가에 따른 프레임들간의 변화가 적다. 그러므로 B 프레임들의 매크로 블록의 타입은 forward, backward, intra 그

리고 양방향 모두를 참조하는 타입이 고루 분포하게 된다. 그러나 $G(T_1 \leq t \leq T_2)$ 구간에서는 식(1)처럼 $\alpha(t)$ 와 $\beta(t)$ 의 영향으로 독립적인 두 영상이 합성되어 나타나므로 G 구간의 B 프레임들의 매크로블록 타입은 일반적인 scene 과는 다른 분포를 가진다. 즉, G 구간의 B 프레임들은 $f_1(x, y, t)$ 와 $f_2(x, y, t)$ 의 서로 다른 두 scene 사이에 존재하기 때문에 두 scene 을 참조하는 특성이 모두 나타나고 매크로블록 타입에 따라서 증가하고나 감소하는 특징을 나타낸다. 그러므로 먼저 G 구간의 B 프레임들에서 $f_1(x, y, t)$ 를 참조하는 매크로블록 타입의 비율을 $MB_{f_1}(t)$ 라 하고 $f_2(x, y, t)$ 를 참조하는 매크로블록 타입의 비율을 $MB_{f_2}(t)$, 식(1)의 $f(x, y, t)$ 의 매크로 블록 타입의 분포를 $MB_f(t)$ 라고 하면 G 구간의 영상의 매크로블록 분포인 $MB_f(t)$ 는 $MB_{f_1}(t)$ 와 $MB_{f_2}(t)$ 의 조합으로 나타낼 수 있다.

식(1)를 살펴보면 $f_1(x, y, t)$ 영상은 G 구간에서 $\alpha(t)$ 의 비율로 감소하므로 $f_1(x, y, t)$ 를 참조하는 매크로블록의 분포를 나타내는 $MB_{f_1}(t)$ 도 참조할 영상의 감소로 인하여 감소하게 되고 $f_2(x, y, t)$ 는 $\beta(t)$ 의 비율로 증가하므로 $f_2(x, y, t)$ 를 참조하는 매크로블록 분포를 나타내는 $MB_{f_2}(t)$ 또한 증가하는 분포를 가진다.

[그림 1]에서 $T_h = \frac{T_1 + T_2}{2}$ 이고 $f_1(x, y, t)$ 과 $f_2(x, y, t)$ 는 독립적이라고 가정하면 $T_1 < T_h$ 구간에선 $\alpha(t) > \beta(t)$ 이므로 $T_1 < T_h$ 구간에선 $f_1(x, y, t)$ 영상이 $f_2(x, y, t)$ 보다 많은 양을 차지하고 있다. 그러므로 $T_1 < T_h$ 구간에 있는 B 프레임들은 $f_1(x, y, t)$ 을 참조하는 매크로 블록의 개수가 $f_2(x, y, t)$ 을 참조하는 매크로 블록의 개수보다 많을 것이다. 또한 $T_h < T_2$ 구간에선 반대로 $f_2(x, y, t)$ 영상의 비율이 $f_1(x, y, t)$ 영상의 비율보다 많으므로 $f_2(x, y, t)$ 을 참조하는 매크로 블록의 개수가 $f_1(x, y, t)$ 을 참조하는 매크로 블록의 개수보다 많게 된다.

위에서 설명한 것처럼 매크로블록 타입의 분포는 T_h 를 기준으로 하여 G 구간에서 반대의 분포를 갖는 특징을 알 수 있다.

본 논문에서는 $MB_{f_1}(t)$ 은 $f_1(x, y, t)$ 를 참조하는 매크로블록의 개수이므로 B 프레임에서 이전영상을 참조하는 매크로 블록의 타입인 forward 타입의 개수라고 할 수 있고 $MB_{f_2}(t)$ 는 $f_2(x, y, t)$ 를 참조하는 매크로 블록의 개수 이므로 backward 방향의 영상을 참조하는 타입은 backward 타입이라고 할 수 있으며

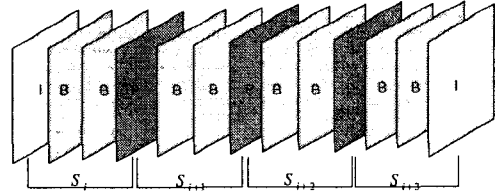
로 다음과 같이 정의 할 수 있다.

$$MB_{f_1}(t) = \sum_{i=0}^N fmv(B_i) \quad (2)$$

$$MB_{f_2}(t) = \sum_{i=0}^N bmv(B_i) \quad (3)$$

식(2), (3)의 N 는 $G(T_1 \leq t \leq T_2)$ 구간내의 B 프레임의 개수이고, B_i 는 G 구간내의 i 번째 B 프레임이다. $fmv(B_i)$ 는 B 프레임내의 forward 타입의 분포를 구하는 함수, $bmv(B_i)$ 는 backward 타입의 분포를 구하는 함수이다.

본 논문에서는 B 프레임에 매크로 블록 타입의 변화율을 좀더 정확히 구하기 위해 GOP 단위의 MPEG Sequence 를 S 크기로 분할하였다. I 또는 P 부터 다음 I 또는 P 프레임까지를 S 구간으로 분할 하여 구간 내에 존재하는 B 프레임들의 매크로블록 타입의 특징을 구하였다.



[그림 3] S 로 분할한 MPEG sequence

먼저 S 내의 첫 번째 B 프레임을 B_1 , 두 번째 B 프레임을 B_2 라고 정의하면 S 구간내의 $MB_{f_1}(S)$ 와 $MB_{f_2}(S)$ 를 다음과 같이 구한다.

$$MB_{f_1}(S) = fmv(B_1) + fmv(B_2) \quad (4)$$

$$MB_{f_2}(S) = bmv(B_1) + bmv(B_2) \quad (5)$$

식(4)의 $MB_{f_1}(S)$ 는 S 구간내의 $MB_{f_1}(t)$ 가 되고 식(5)의 $MB_{f_2}(S)$ 는 S 구간내의 $MB_{f_2}(t)$ 가 된다.

점진적 장면전환 구간을 검출하기 위해 $MB_{f_1}(S_i)$ 와 $MB_{f_1}(S_{i+1})$ 의 차를 구하여 $MB_{f_1}(t)$ 의 변화율을 구하고 $MB_{f_2}(S_i)$ 와 $MB_{f_2}(S_{i+1})$ 의 차를 구하여 $MB_{f_2}(t)$ 의 변화율을 측정하여 점진적인 장면전환을 검출한다. 즉, $MB_{f_1}(t)$ 는 $T_1 < T_h$ 구간에서는

$MB_{f_2}(t)$ 보다 큰 값을 가지며 감소하게 되고, $T_h < T_2$ 에서 $MB_{f_2}(t)$ 는 $MB_{f_1}(t)$ 보다 큰 값을 가지며 증가하게 된다. 즉, $MB_{f_1}(t)$ 값이 감소함과 동시에 $MB_{f_2}(t)$ 의 값이 증가하는 구간이 점진적인 장면전환 구간이 된다.

5. 실험 결과

본 연구에서 제안하는 점진적인 장면전환 검출 기법을 시험하기 위한 프로토타입은 WINDOWS 2000 환경에서 Visual C++ 6.0 과 연구실에서 제작한 MPEG Library 를 이용하여 MPEG-1 데이터를 대상으로 실험하였다. 각 데이터는 30 frames/sec 의 프레임율과 352 × 240 해상도를 갖는다. 일반적으로 점진적인 장면전환이 일어나는 횟수는 급진적 장면전환에 비해 매우 적은 편이다. 따라서 본 논문에서는 점진적 장면전환이 삽입되어 있는 광고와 드라마를 통하여 실험하였다. 실험은 같은 동영상에 대하여, 기존에 많이 사용하는 분산을 이용한 점진적 장면전환 검출 방법과 제안한 방법인 B 프레임의 특성값을 사용했을 때의 결과 값인 장면전환 검출 수, 오검출 수, 미검출 수와 장면전환 검출 시 소요되는 시간을 비교 평가하고 제안한 알고리즘에 대한 성능을 분석하였다.

Video Data	광고 (1850 frame)	
점진적 장면전환 수	7	
검출방법	기존 방법	제안한 방법
검출횟수	8	7
오검출	1	1
미검출	0	1
소요시간	2분 14초	1분 3초

[표 1] 광고 데이터에 대한 실험 결과

Video Data	드라마 1 (10900 frame)	
점진적 장면전환 수	15	
검출방법	기존 방법	제안한 방법
검출횟수	16	14
오검출	2	1
미검출	1	2
소요시간	9분 34초	6분 58초

[표 2] 드라마 데이터에 대한 실험 결과

Video Data	드라마 2 (30922 frame)	
점진적 장면전환 수	52	
검출방법	기존 방법	제안한 방법
검출횟수	55	50
오검출	4	1
미검출	1	3
소요시간	25분 21초	18분 57초

[표 3] 드라마 데이터에 대한 실험 결과

위의 표에서 보는 바와 같이 제안한 알고리즘은 기

존의 분산을 이용한 검출방법과 비슷한 성능을 나타냈고 장면전환을 검출하는데 있어서 소요되는 시간이 단축됨을 볼 수 있다. 제대로 검출하지 못한 부분의 동영상상을 직접 확인해 보면 대부분 빠른 카메라 워크가 동반되거나 모핑과 같은 특수효과와 같이 전환 효과가 연이어 또는 함께 사용되는 경우가 대부분 이었다.

6. 결론 및 향후과제

일반적으로 점진적인 장면전환은 급진적 장면전환에 비해 발생 빈도가 상대적으로 낮기 때문에 점진적인 전환 검출만을 위하여 특별히 복잡한 알고리즘을 사용하는 것은 비효율적이라 할 수 있다. 그러나 기존의 알고리즘의 대부분은 점진적인 장면전환 구간을 검출하기 위해 많은 연산이 필요하였다.

본 논문은 B 프레임의 매크로블록타입의 변화만을 가지고 점진적 장면전환 구간을 검출할 수 있는 기법을 제안하였다. 제안한 기법은 점진적인 장면전환을 검출하기 위한 부가적인 연산이 적기 때문에 구현이 쉽고 빠른 처리가 가능한 장점을 갖는다. 제안된 방법은 일반적인 PC 에서도 실시간으로 동작할 수 있어 내용기반 정보검색 시스템에 필요한 장면분할에 효율적으로 활용할 수 있을 것으로 기대된다

앞으로는 카메라 움직임에 의한 오검출 및 미검출을 줄이고 디졸브 계열의 구간뿐만 아니라 다른 종류의 점진적인 장면전환 구간을 검출하기 위한 연구가 필요하다.

참고문헌

- [1] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full motion search for object appearance," Proc. of 2nd Working Conf. on Visual Database System, pp. 113-127, Sep. 1991.
- [2] W.A.C. Fernando, C.N. Canagarajah, and D.R. Bull, "Fade-in and fade-out detection in video sequences using histograms," The 2000 IEEE International Symposium on Circuits and Systems, Proc. of ISCAS 2000 Geneva, Vol. 4, pp. 709-712, 2000.
- [3] Xinying Wang and Zhengke Weng, "Scene abrupt change detection," Electrical and Computer Engineering, 2000 Canadian Conference on, Vol. 2, pp. 880-883, 2000.
- [4] A. Hampapur, R. Jain, and T. Weymouth, "Production model based digital video segmentation," Multimedia Tools and Application, Vol. 1, No. 1, pp. 9-46, Mar. 1995.
- [5] B. Shaharay, "Scene change detection and content-based sampling of video sequences," Proc. of SPIE, Vol. 2419, pp. 2-13, Feb. 1995.
- [6] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," Proc. of ACM Multimedia System, Vol. 1, No. 1, pp. 10-28, 1993.
- [7] B. L. Yeo and B. Lie, "Rapid scene analysis on compressed video," IEEE Trans. On Circuits and Systems for Video Technology, Vol. 5, No. 6, pp. 533-544, Dec. 1995.
- [8] H. Zhang, C. Y. Low, and S. W. Smoliar, "Video parsing and browsing using compressed data," Multimedia Tools Application, Vol. 1, pp.89-111, 1995.