

# 압축된 MPEG 비디오 시퀀스에서 B 프레임의 특징을 이용한 빠른 장면전환 검출 알고리즘

김 중 헌, 김 신 형, 박 두 영, 장 중 환  
배재대학교 정보통신공학과

## A Fast Scene Change Detection Algorithm Using Feature Of B Frame in Compressed MPEG Video Sequence

Joong-Heon Kim, Shin-Hyung Kim, Doo-Yeong Park, Jong-Whan Jang  
Department of Information & Communication Engineering, PaiChai University

jhkim@icelab.paichai.ac.kr, zeros@icelab.paichai.ac.kr, dwpark@mail.paichai.ac.kr, jangjw@mail.paichai.ac.kr

### 요 약

비디오 데이터의 효율적인 저장, 관리를 위해서는 장면전환 검출을 통한 비디오 분할 기술에 대한 연구가 필요하다. 본 논문에서는 MPEG 압축 비디오 상의 B(Bidirectional) 프레임의 특성을 복호화 과정을 거치지 않고 직접 추출하여 I(Intra), P(Predictive), B(Bidirectional) 프레임에 제안받지 않고 장면전환을 검출해 내는 방법을 제안한다. 장면전환 검출을 위해 복호화 하지 않고 필요한 데이터만을 추출해 내어 B 프레임의 특징만을 이용해 검출하므로 빠르면서도 정확한 장면전환을 검출한다. 또한 카메라 움직임이나 빛의 변화 같은 잡음에 강건한 방법을 제안한다.

### 1. 서 론

현대 사회의 가장 큰 특징 중에 하나는 다양한 멀티미디어 데이터를 사용하고 있다. 멀티미디어 환경의 특징은 텍스트 환경과 비교하여 데이터의 양이 방대하므로 대역이 제한된 통신 네트워크를 정보전달 매체로 이용하는 경우, 빠른 시간 내에 정확히 필요한 정보를 검색, 처리 할 수 있는 방법의 개발이 매우 중요하며 이에 대한 많은 연구가 수행되고 있다[1,2]. 멀티미디어 데이터를 복수의 표현 미디어라고 할 때, 여러 가

지 구성요소 가운데에서 동영상이 멀티미디어 데이터의 핵심이라고 말할 수 있으며[3]. 이러한 동영상에 대한 압축/복원 기술의 국제적인 표준가운데에서 우리에게 가장 잘 알려진 것이 바로 MPEG일 것이다.

이러한 동영상 압축 기술의 국제적인 표준화로 많은 동영상들이 MPEG 기술을 이용하여 압축/저장되고 있으며, 그 수가 기하급수적으로 늘어나고 있다. 그리고, 디지털 저장 매체에 저장된 동영상들의 수가 늘어남에 따라서 다시, 이 정보들을 좀더 효율적으로 이용하기 위하여 체계적으로 분류, 저장할 필요성이 생겨나게 되었다. 이런 필요하에 나타나게 된 것이 근래에 대두되기 시작한 MPEG7이다.

MPEG7은 content-based storage and retrieval을 위한 국제표준으로써, 이를 위하여 동영상 데이터 색인화를 위한 기술이 필수적으로 포함되게 될 것이다. 동영상 데이터 색인화를 위해서는 우선 동영상들을 내용 위주로 나누어야만 한다. 이를 위해서 우선적으로 필요한 기술이 바로 장면 전환 검출 알고리즘이라고 할 수 있다. 현재 장면 전환 검출 알고리즘에 대한 연구가 많이 진행되고 있으며, 많은 연구 결과들이 발표되고 있기는 하지만, 실제로 이용되기에는 아직 부족한 점이 많이 있다. 그 이유는 각 알고리즘들이 영상의 특성에 따라

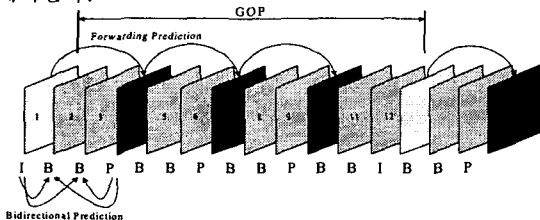
서 성능이 많이 좌우되고 있기 때문이다. 기존의 장면 전환 검출 알고리즘은 크게 나누어 보면, 압축 도메인에서의 알고리즘[4]과 비압축 도메인에서의 알고리즘[5]으로 나누어 질 수 있다. 이 중에서 계산량을 줄이고자 하는 목적 때문에 압축 도메인에서의 알고리즘에 대한 연구가 더욱 많이 이루어지고 있다. 기존의 비압축 도메인에서의 장면전환 검출 알고리즘은 복호화에 의한 연산량이 많고 압축 도메인에서의 장면전환 검출 알고리즘은 카메라의 움직임이나 빛의 변화에 의한 오검출이 많이 발생한다. 본 논문에서는 압축 영역에서 직접 B 프레임의 특징값을 이용하여 연산량을 감소시키며 정확한 장면전환 검출을 하고 카메라의 움직임이나 빛의 변화에 의한 오검출을 감소시키는 알고리즘은 제안한다.

본 논문의 구성은 제 2장에서 B 프레임이 가지고 있는 특징에 대하여 설명하였고 제 3장에서는 압축 영역에서 B 프레임의 특징을 이용한 장면전환 검출 알고리즘을 제안하였다. 그리고 4장에서는 제안한 알고리즘은 적용한 실험치 결과를 기술하며 제 5장에서는 결론 및 향후 과제를 제시하였다.

## 2. B 프레임의 특징

### 2.1 MPEG 비트열의 구조

MPEG에서는 동영상의 시간적인 중복성을 제거하기 위하여 Motion Compensation 방법과 공간적인 중복성을 제거하기 위하여 DCT(Discrete Cosine Transform) 압축방식을 사용한다. 장면전환 검출은 프레임간 컷이나 점진적인 변환(Gradual Transition)이 존재하는지를 판단하는 것이므로 프레임 단위에 대한 이해가 필요하다. I 프레임은 디코딩 시 기준이 되는 프레임인데 P, B 프레임으로 인해 발생하는 에러를 줄이기 위해 DCT 양자화를 통해 공간 중복성만을 제거하고 P, B 프레임은 참조 프레임과의 차영상과 모션벡터로 시간 중복성을 제거한다.



[그림 1] I, P, B 프레임 화면의 예측방법

### 2.2 B 프레임의 특징

B 프레임 안에서 매크로 블록의 종류는 크게 4종류로 나눌 수 있다. Forward 타입, Backward 타입, Intra

타입, 그리고 마지막으로 Forward와 Backward가 모두 존재하는 Interpolation 타입이 존재한다. 각각의 타입은 그 매크로 블록의 특징을 잘 설명해 주고 있기 때문에 각 타입이 결정되는 것을 살펴보면 다음과 같다. B 프레임이 기준프레임 사이에 두 개의 열을 이루고 있으며 그 중 첫 번째 B 프레임의 모습이다

$$sadf = \sum_{i=0}^{15} \sum_{j=0}^{15} |MB[i, j]_N^k - MB[i + \Delta x, j + \Delta y]_{N-1}^k| \text{ for } -S \leq \Delta x, \Delta y \leq S \quad (1)$$

$$FMV(x, y)_N^k = \Delta x, \Delta y \text{ if } \text{Min}(sadf)_N^k \quad (2)$$

식(2)에서  $sadf$ 가 최소값일 때 변위인  $\Delta x, \Delta y$ 가  $N$ 번째 영상의  $k$ 번째 매크로 블록의 forward 모션벡터인  $FMV(x, y)_N^k$ 가 된다. 모션벡터 검색 영역  $2S$  내에서  $\Delta x, \Delta y$ 를 변화시키며  $sadf$ 가 최소값이 되는 변위  $\Delta x, \Delta y$ 를 찾는다. 식(2)의  $MB[i, j]_N^k$ 는 B 프레임으로 인코딩 될  $N$ 번째 프레임의  $k$ 번째 매크로 블록이며  $16 \times 16$ 의 크기를 갖는다. 식(2)의  $MB[i, j]_{N-1}^k$ 는 forward 방향으로 참조하는 영상의  $k$ 번째 매크로 블록이다.

$$sadb = \sum_{i=0}^{15} \sum_{j=0}^{15} |MB[i, j]_N^k - MB[i + \Delta x, j + \Delta y]_{N+2}^k| \text{ for } -S \leq \Delta x, \Delta y \leq S \quad (3)$$

$$BMV(x, y)_N^k = \Delta x, \Delta y \text{ if } \text{Min}(sadb)_N^k \quad (4)$$

식(3), (4)에서 보는 것과 같이 backward 모션벡터도 forward 모션벡터와 같은 방법에 의하여 구하게 된다. 단 식(3)에서 보는 것과 같이 backward 방향의 매크로 블록으로  $sadb$ 를 구한다.

위에서 구한 forward/backward 모션벡터를 이용하여 참조영상에서 영상 데이터를 가지고 온 후 입력영상과의 분산값을 이용하여 참조 방향을 결정한다.

$$fMB_N^k = MB_{N-1}^k \rightarrow MB_{N-1}^{k+FMV(x,y)_N^k} \quad (5)$$

$$bMB_N^k = MB_{N+2}^k \rightarrow MB_{N+2}^{k+BMV(x,y)_N^k} \quad (6)$$

$$iMB_N^k = \frac{fMB_N^k + bMB_N^k}{2} \quad (7)$$

식(5)의  $fMB_N^k$ 는 forward 방향의  $k$ 번째 매크로 블록인  $MB_{N-1}^k$ 를  $FMV(x, y)_N^k$ 만큼 이동시킨 영상 데이터이다. 식(6)의  $bMB_N^k$ 는  $MB_{N+2}^k$ 를  $BMV(x, y)_N^k$ 만큼 이동시킨 매크로 블록의 데이터이고 식(7)의  $iMB_N^k$ 는 forward/backward 두 방향 영상의 평균이다

$$Varf = \frac{1}{256} \sum_{i=0}^{15} \sum_{j=0}^{15} (MB[i, j]_N^k - fMB[i, j]_N^k)^2 \quad (8)$$

$$Varb = \frac{1}{256} \sum_{i=0}^{15} \sum_{j=0}^{15} (MB[i, j]_N^k - bMB[i, j]_N^k)^2 \quad (9)$$

$$Vari = \frac{1}{256} \sum_{i=0}^{15} \sum_{j=0}^{15} (MB[i, j]_N^k - iMB[i, j]_N^k)^2 \quad (10)$$

식(8), (9), (10)에서 각 참조 영상과 입력 영상과의 차의 variance를 구하여 최소값이 되는 영상을 참조하게

된다. 그러나 차의 variance가 입력 영상의 variance 보다 차이가 크면  $k$  번째 매크로 블록은 모션벡터가 존재하지 않는  $intra$  타입으로 인코딩 된다.

$$M = \frac{1}{256} \sum_{i=0}^{15} \sum_{j=0}^{15} MB[i, j]_N^k \quad (11)$$

$$Vard = \text{Min}(Varf, Varb, Vari) \quad (12)$$

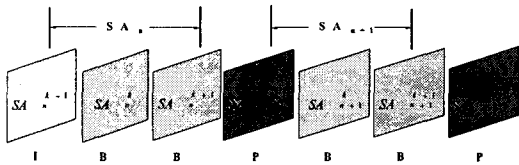
$$Varc = \frac{1}{256} \sum_{i=0}^{15} \sum_{j=0}^{15} (MB[i, j]_N^k - M)^2 \quad (13)$$

$$Var = Vard - Varc \begin{cases} Intra, & \text{if } Var > 0 \\ MV, & \text{otherwise} \end{cases} \quad (14)$$

식(11)의  $M$ 은 입력영상의  $k$  번째 매크로 블록의 평균이고 식(12)의  $Vard$ 는 식(8), (9), (10)에서 구한 variance의 최소값이 된다. 또한 식(13)의  $Varc$ 는 입력영상인  $N$  번째 프레임의  $k$  번째 매크로 블록의 분산값이 된다. 식(14)에서 보듯이  $Vard$ 가 크면 인트라 타입이 되고 입력영상과 참고 영상과의 차의 variance가 입력영상의 variance 보다 적게 되면 모션 벡터를 가지는 타입이 된다.

### 3. 압축 영역에서 B 프레임의 특징을 이용한 장면전환 검출 알고리즘

일반적으로 MPEG 시퀀스에서는 두 개 이상의 B 프레임이 열을 이루며 B 프레임들은 forward/backward의 방향성을 가지게 된다. 만약 B 프레임의 매크로 블록들이 한쪽 방향으로만 참조하려는 경향이 강하다면 참조하지 않은 영상에서 장면전환이 발생했을 확률이 높게 된다. 그러므로 B 프레임들의 방향성을 살펴보면 장면전환된 프레임을 검색할 수 있다.



[그림 2] SA(Search Area)

[그림 2]에서 보면 기준 프레임부터 다음 기준 프레임 전까지를  $SA_n$ 으로 정의하였다. 장면전환 검출을 위하여  $SA_n$ 의 B 프레임 시작프레임인  $SA_n^k$ 에서

$$D(SA_n^k) = FMVN(SA_n^k) - BMVN(SA_n^k) \quad (15)$$

$$Cut(SA_n^k) = \begin{cases} 1, & \text{if } D(SA_n^k) \geq Th_f \\ 0, & \text{if } -D(SA_n^k) \geq Th_b \\ -1, & \text{otherwise} \end{cases} \quad (16)$$

식(15)의  $FMVN(SA_n^k)$ 은  $SA_n^k$ 번째 프레임의 forward타입의 매크로 블록의 개수를 구하는 함수이고  $BMVN(SA_n^k)$

는 backward타입의 매크로 블록의 개수를 구하는 함수이다.  $D(SA_n^k)$ 는 forward타입과 backward타입의 차가 된다. 식(16)의  $Th_f$ 는 양의 정수로서 forward 방향성을 결정하는 임계값이고  $Th_b$  역시 양의 정수로서 backward 방향성을 결정하는 임계값이다.  $Cut(SA_n^k)$ 가 1인 경우는  $SA_n^k$ 가 forward 방향의 영상을 참조한 경우이고 0인 경우는 backward 방향의 영상을 참조한 경우, -1인 경우는 장면전환이 없는 영상이 된다. 장면전환 검출을 위하여  $SA_n$ 영역 안의 모든 B 프레임에 대하여  $Cut(SA_n^k)$ 를 구한다. 본 논문에서는 장면전환이 발생하는 경우를 3가지로 정의하였다. 첫 번째로  $SA_n$ 영역안의 기준 프레임에서 장면전환이 발생한 경우이고 두 번째는  $SA_n$ 의 B 프레임열의 시작에서 장면전환이 발생한 경우, 마지막으로  $SA_n$ 의 B 프레임 열의 중간에서 장면전환이 발생한 경우이다. 첫 번째의 경우, 즉 [그림 2]의  $SA_{n+1}^k$ 인 P 프레임에서 장면전환이 발생하게 되면 이전 영역  $SA_n$ 안의 B 프레임들은 다음 영역의 기준 프레임에서 장면전환이 발생하였으므로 같은 영역안의 기준 프레임을 참조하는 매크로 블록의 개수가 많으므로 B 프레임들의  $Cut(SA_n^k)$  값은 0이 되고  $SA_{n+1}^k$ 의 B 프레임들은 한쪽으로 치우친 방향성을 갖지 않으므로  $Cut(SA_{n+1}^k)$  값은 -1이 된다. 이때  $K$ 는  $SA$  영역안의 B 프레임의 개수가 된다. [그림 2]의  $SA_{n+1}^k$ 인 B 프레임열의 시작에서 장면전환이 발생하게 되면  $SA_{n+1}^k$ 안의 B 프레임들은 같은 영역의 기준 프레임을 참조하지 못하고 다음 영역에 나타나는 기준프레임을 참조하게 된다. 즉 forward 방향성을 가지는 매크로 블록의 개수가 많게 된다. 그러므로  $Cut(SA_{n+1}^k)$ 의 값은 1이 된다. 마지막으로  $SA$ 의 B 프레임 열의 중간에서 장면전환이 발생하면 장면전환이 발생한 프레임보다 이전에 나타난 B 프레임들은 같은 영역의 기준프레임을 참조하게 되고 장면전환일 발생한 프레임 이후에 나타나는 B 프레임들은 다음 영역의 기준프레임을 참조하게 된다. 예를 들어 [그림 2]의  $SA_{n+1}^k$ 에서 장면전환이 발생하면  $SA_{n+1}^k$  이전에 나타난  $SA_{n+1}^k$  프레임은 같은 영역의 기준프레임인  $SA_{n+1}^k$ 를 참조하는 매크로 블록의 수가 많으므로  $Cut(SA_{n+1}^k)$  값이 0이 되고  $SA_{n+1}^k$ 을 포함하여 다음에 나타난 B 프레임은 다음 영역의 기준프레임을 참조하는 매크로 블록 타입이 증가하므로  $Cut(SA_{n+1}^k)$ 의 값은 1이 된다.

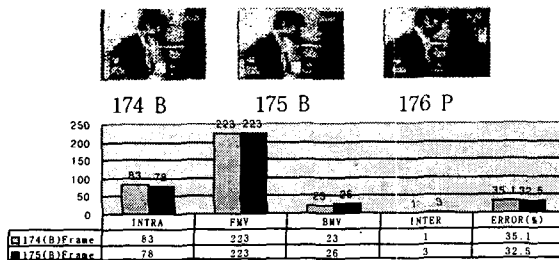
$$Df(SA_n^k) = \begin{cases} SA_n^{Anc}, & \text{if } Cut(SA_{n-1}^k) \equiv 0 \\ SA_n^{Bn}, & \text{if } Cut(SA_n^k) \equiv 1 \\ SA_n^{Bn}, & \text{if } \begin{cases} Cut(SA_n^k) \equiv 0, (SA_n^{Bn} \leq SA_n^k < SA_n^{Anc}) \\ Cut(SA_n^k) \equiv 1, (SA_n^{Anc} < SA_n^k < SA_n^{Bn}) \end{cases} \end{cases} \quad (17)$$

식(17)에서 탐색 영역을  $SA_n$ ,  $SA_n$ 영역의 기준 프레임을

$SA_n^{Anc}$ , B 프레임의 시작 프레임을  $SA_n^{Sn}$ , B프레임의 중간에 존재하는 B 프레임을  $SA_n^{Bn}$  이라 정의하여 장면 전환 프레임의 검출을 위와 같이 정의 하였다. 앞에서도 언급하였지만 B 프레임 특성상 방향성을 가지고 인코딩 된다. 그러므로 B 프레임에서 양쪽 방향으로 모두 참조를 하는 Interpolation 타입이나 어떤 방향으로도 참조할 수 없는 Intra 타입은 장면전환 검출에 있어서 어려로 정의 할 수 있다.

$$\frac{IMVN(SA_n^k) + IMBN(SA_n^k)}{FMVN(SA_n^k) + BMVN(SA_n^k)} \geq Th_{error} \quad (18)$$

즉 어려율이  $Th_{error}$  이상이면 장면전환 검출이 아닌 오검출로 판단 할 수 있다. 식(18)의  $IMVN(SA_n^k)$  은  $SA_n^k$  번째 프레임의 Interploation 타입의 개수이고  $IMBN(SA_n^k)$  는 Intra 타입의 개수이다.  $Th_{error}$  는 오검출을 판단하는 임계값이다.



[그림 3] 오검출 및 매크로 블록 타입별 개수

[그림 3]에서 보면 176번째 프레임이 오검출이나 식(18)에 의하여 오검출로 판단되어 장면전환으로 검색하지 않는다.

#### 4. 실험 및 결과

실험은 같은 동영상에 대하여, 기존에 많이 사용하는 히스토그램 비교법과 제안한 방법인 B 프레임의 특성값을 사용했을 때의 결과 값인 장면전환 검출 수, 오검출 수, 미검출 수를 비교 평가하고, 제안한 알고리즘에 대한 성능을 분석하였다.

| 동영상  | Video #1 |    | Video #2 |     | Video #3 |     |
|------|----------|----|----------|-----|----------|-----|
| 컷횟수  | 10       |    | 150      |     | 364      |     |
| 검출방법 | 기존       | 제안 | 기존       | 제안  | 기존       | 제안  |
| 총횟수  | 21       | 10 | 263      | 151 | 427      | 363 |
| 컷검출  | 10       | 10 | 147      | 148 | 359      | 362 |
| 오검출  | 11       | 0  | 116      | 1   | 68       | 1   |
| 미검출  | 0        | 0  | 3        | 2   | 5        | 2   |

[표 1] 기존 알고리즘과의 검출횟수 비교

실험에 데이터로 사용된 MPEG데이터는 직접 사람의 눈

에 의해 정확한 컷의 위치를 발견하여 실제 장면전환 개수로 했다.  $Th_b$ 와  $Th_s$ 는 실험에 의하여 매크로 블록의 개수에 70%로 사용하였다.  $Th_{error}$  역시 실험에 의해 30%로 사용하였다

[표 1]의 컷 횟수는 시각적으로 장면전환점의 수를 조사한 값이다. 총 검출횟수는 방법에 따라 실험하여 컷 검출된 수를 나타낸다. 기존의 알고리즘보다 제안한 알고리즘을 적용한 경우 오검출수가 많이 줄어들게 되는 것을 볼 수 있다. 또한 기존알고리즘은 장면전환을 검색하기 위하여 디코딩 과정 시 많은 연산을 필요로 하지만 제안한 알고리즘은 압축된 bitstream에서 특징값만 추출하게 되므로 처리 속도를 향상 시킬 수 있었다.

#### 5. 결론 및 향후 과제

MPEG 비디오 시퀀스에서 압축 영역에서의 컷 검출 관련 연구는 아직 뚜렷하게 대표할 만한 것이 없다. 본 논문에서는 MPEG 비디오 시퀀스에서 B 프레임의 특징만을 이용하여 비교하는 방법으로 컷 검출을 시행하여 기존의 알고리즘 보다 처리 시간의 단축과 정확한 검출 및 복잡하지 않은 검출과정으로 우수한 결과를 도출하였다.

향후 연구 방향으로는 압축 데이터 상에서 점진적인 장면전환을 효과적으로 검출해 내는 방법에 관한 연구가 진행되어야 하며 또한 좀더 정확한 검출을 위하여 정확한 임계값에 관한 연구가 필요하다.

#### 참고 문헌

- [1] P.Aigrain, J. J Zhang and D. Petkovic, "Content-based representation and Retrieval of Visual Media : a State-of-the-Art Review, *Multimedia Tools and Applications*, Vol13, pp. 179-202, 1996
- [2] R. Lienhart, S. Pfeiffer and W. Effelsberg, "Video Abstracting", *ACM Communication*, Vol. 40, No. 12, pp. 55-62, 1997
- [3] G. Medioni, R. Nevatia and I. Cohen, "Event detection and Analysis from Video Streams", *DARPA98*, pp. 63-72, 1998
- [4] B. Shaharay, "Scene Chang Detection and Content-Based Sampling of Video Sequences," *Proc. of SPIE*, Vol. 2419, pp. 2-13, Feb. 1995.
- [5] Ke Shen and Edward J. Delp, " A Fast Algorithm for Video Parsing Using Compressed Sequence." *ICIP 95*, pp. 252-255, 1995