

DCT계수와 움직임 정보를 이용한 MPEG 비디오 색인

박한엽* · 최연성*

*군산대학교

Indexing using DCT Coefficients and Motion Information in MPEG Video

Park Hanyeop* · Choi Yoensung*

*Kunsan National University

E-mail : hypark@k2.kunsan.ac.kr

요 약

디지털 비디오를 자동 색인하고 검색하기 위한 기법은 여러 가지가 제안되어 있으나 대부분 압축되지 않은 비디오를 대상으로 하거나, 아니면 일단 디코딩한 후 색인한다. 본 논문에서는 압축 영역에서 직접 색인할 수 있는 기법을 제안한다.

I. 서 론

멀티미디어 기술의 발달은 디지털 비디오의 보편화를 가져오게 되었다. 디지털 비디오 정보는 대부분이 고용량이기 때문에 압축을 피할수 없고, 이에 보다 효과적이고 빠른 인덱싱 기법이 요구되어진다. 본 논문에서는 압축된 비디오 정보에서의 인덱싱시에 발생하는 많은 비용과 부하를 줄이고자 압축된 영역에서 비디오 인덱싱 기법을 제안한다. 먼저, 다음장에서 MPEG 비디오의 구조를 설명하고, DCT계수 및 움직임정보를 추정하여 비디오 인덱싱에 적용시킨다.

II. 본 론

1. MPEG 비디오의 구조

많은 표준화 작업을 통한 MPEG은 DCT부호화와 예측부호화를 통해 상당한 압축된 정보를 얻을 수 있다. 그 데이터 구조를 살펴보면 시퀀스 층, GOP층, 픽처 층, 슬라이스 층, 매크로블록 층, DCT계수를 포함하는 블록 층으로 분류되어진다.

① 시퀀스 층

비디오의 수평/수직크기, 비디오내의 매크로블록 수, 초당 매크로블록 수, 화면률, 순방향/역방향 f코드, 비트율등이 지정된다.

② GOP 층

몇 개의 프레임들을 일련의 집합으로 표현한 것으로 비디오 정보의 랜덤액세스에 이용된다.

③ 픽처 층

한 프레임에 대한 픽처타입, 움직임벡터에 대

한 플래그, 확장비트등의 공통적인 속성을 지정하는 층이다.

④ 슬라이스 층

한 프레임을 임의의 길이(MB들의 집합)로 분할한 것을 슬라이스라 하며, 이에 대한 공통적인 정보를 지정하는 층이다.

⑤ 매크로블록 층

슬라이스 층을 더욱 분할하여 매크로블록에 대한 공통적인 정보를 지정한다.

⑥ 블록 층

매크로블록 층을 8×8블록으로 더욱 세분화한 층이며, 해당되는 DCT계수를 포함한다.

MPEG에서의 예측부호화방식으로는 순방향예측, 역방향예측, 쌍방향예측이 있고 이에 대응하는 영상 타입으로는 I, P, B프레임이 있다. I프레임은 프레임간 예측을 이용하지 않는 화면으로 그 해당정보만으로 부호화되는 화면이고, P프레임은 I 혹은 P프레임로부터 순방향예측으로 부호화되는 화면이다, 마지막으로 B프레임은 MPEG의 특징인 쌍방향으로 예측되어 얻어지는 화면이다. 다음 그림 1은 MPEG비디오의 일반적인 예측 부호화의 한 예이다.

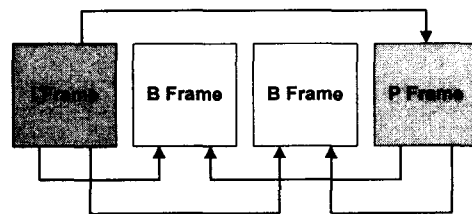


그림 1. MPEG에서의 프레임예측

그림 1은 MPEG에서의 프레임예측의 예를 보이고 있는데, 다음과 같이 정리할 수 있다.

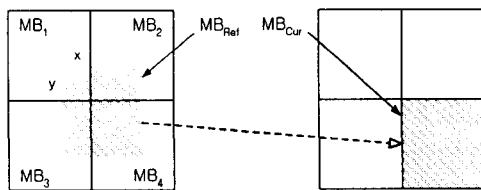
- ① I프레임으로부터 순방향예측을 통해 P프레임을 얻는다.
- ② I프레임으로부터의 순방향예측과 P프레임으로부터의 역방향예측을 통해 B프레임을 얻는다.

여기서, P나 B프레임은 다른 레퍼런스 프레임으로부터 예측이나 보간후에 발생하는 오차에러 데이터로부터 생성된 DCT계수를 가지지만, I프레임은 실제적인 픽셀정보의 DCT계수를 포함하고 있다.

MPEG 비디오 정보에서의 인덱싱은 크게 두가지 방법으로 진행될 수 있다. 첫 번째는 몇 개의 혹은 모든 비디오를 압축해제하여 인덱싱할 수 있는 특징들을 추출하여 인덱싱을 하는 것이다. 두 번째는 완전한 압축해제의 오버헤드에 따른 손실없이 압축된 정보에 포함된 부호화된 정보를 추출하여 인덱싱을 하는 것이다. 본 논문에서는 다음장부터 기술될 DCT계수의 추정과 움직임추정을 통해 비디오 인덱싱을 수행하였다.

2. DCT계수의 추정

DCT계수는 I프레임에서는 쉽게 접근할수 있지만, P와 B프레임이 예측이나 보간후 오차에러에 의해 표현되기 때문에, P와 B프레임의 DCT계수는 추정되어야 한다. P와 B프레임에서 MB의 DCT계수를 계산하기 위해, 현재 MB이 예측되어진 레퍼런스 프레임의 16×16영역의 DCT계수가 먼저 계산되어야 한다. 하지만, 이 레퍼런스 MB의 DCT계수는 이 프레임을 덮을 수 있는 네 개 MB들의 DCT계수들로부터 계산되어지는 것보다, P나 B프레임에서의 MB의 DCT계수들에 알맞은 근사치를 계산하는 것이 보다 효과적이다.[1][2]



Motion Vector = (x, y)

그림 2. MB_{Ref} 로부터 P프레임의 MB_{Cur} 예측

그림 2는 MB_{Ref} 에 의해 정의된 16×16 영역으로부터 예측되어진 P프레임에서 MB_{Cur} 을 보여주고 있다. P프레임을 인코딩하는 동안, MB_{Ref} 를 고려한 MB_{Cur} 의 오차에러만이 저장된다. MB_{Ref} 의 DC계수는 MB_1, MB_2, MB_3, MB_4 의 DC계

수들로부터 계산되어진다[1][2]. 높은비용의 계산을 피하기위해, 단지 DC계수만이 네 개의 MB들의 DC계수들의 가중된 합에 의해 추측되어진다. 여기서 가중치는 레퍼런스 MB를 덮고있는 이 MB들의 영역의 한 부분이 된다. 즉,

$$DC(MB_{Ref}) = \sum_{i=0}^4 w_i \times DC(MB_i)$$

여기서 w_i 는 그 모든 영역에 MB_i 의 빗금쳐진 지역의 영역의 비율로 주어진다.

비슷하게, B프레임에서의 MB가 두 레퍼런스 MB들에서 중간치로 결정되어진다면, 그 DC계수는 이 두 MB들의 추정된 DC계수들의 평균치로 추정되어진다.

3. 움직임정보의 추정(흐름추정)

MB는 MB가 인트라부호화된 것인지, 순방향/역방향으로 예측된 것인지, 혹은 쌍방향으로 예측된 것인지에 상관없이 각각의 프레임 형태에 따르는 0, 1, 2개의 움직임벡터들을 가질수 있다. 움직임벡터가 없는 I프레임과 쌍방향예측된 B프레임과 비교할 때, 그리고 예측방향이 다른 두 개의 B프레임을 비교할 때 문제가 발생한다. 여기서 예측방향과 프레임 형태에 독립적인 움직임벡터가 요구되며, 이에 본 논문은 각각의 움직임벡터들을 프레임 형태에 독립적으로 다음프레임에 대해 역방향으로 예측된 벡터로 표현한다.

가. I와 P프레임만을 갖는 비디오

만약 I프레임들 사이에 P프레임들만 있다면, 흐름(flow)은 두 개의 연속되는 I프레임들사이의 각 프레임들에 대해서 추정될 수 있다. P프레임에서의 MB가 I프레임의 MB에 대한 움직임벡터(x,y)에 의해 이동되었다면, I프레임의 MB가 P프레임에서의 MB에 대한 움직임벡터(-x,-y)에 의해 이동되었다고 추측하는 것이 더 논리적이다. 같은 방법으로 또다른 P프레임들이 뒤따른 P프레임의 MB들의 흐름추정(flow estimation)에 적용되어진다.

나. B프레임을 포함한 비디오

대부분의 MPEG 스트림은 연속되는 레퍼런스 프레임들사이에 B프레임을 갖는다. 두 개의 연속되는 레퍼런스 프레임들 R₁와 R₂로 놓자. 이 레퍼런스 프레임들 사이에 B프레임들을 B₁...B_n으로 나타낸다.

각 B 프레임에서 각각의 MB는 세가지 형태(순방향예측[F], 역방향예측[B], 쌍방향예측[D])중에 하나를 취할수 있기 때문에, 아홉 개의 가능한 조합이 존재한다(FF, FB, FD, BF, BB, DF, DB, DD). 흐름(flow)은 이 조합들 사이에 레퍼런스 프레임에 대한 각각의 움직임벡터를 분석함으로써 추정된다. 이는 크게 4가지 경우를 생각할 수 있

다.

① FF, FD, DF, DD의 경우

B_1 에서의 현재 MB의 순방향-예측된 움직임 벡터를 $B_1\vec{R}_i$ 로 나타내고, B_2 에서 부합되는 MB의 순방향-예측된 움직임벡터를 $B_2\vec{R}_i$ 로 나타낸다. 만약, B프레임들의 MB들 사이에 흐름을 $B_1\vec{B}_2$ 로 정의한다면, 다음과 같은 관계식을 갖는다.

$$-B_2\vec{R}_i = -B_1\vec{R}_i + B_1\vec{B}_2$$

② BB, BD, DB의 경우

R_i 에 대한 B_1 과 B_2 의 역방향-예측된 움직임 벡터를 이용하여 $B_1\vec{R}_i$ 를 얻을 수 있다. B_1 에서의 현재 MB의 역방향-예측된 움직임벡터를 $B_1\vec{R}_i$ 로 나타내고, B_2 에서 부합되는 MB의 역방향-예측된 움직임벡터를 $B_2\vec{R}_i$ 로 나타낸다. 그러면, 다음과 같은 관계식을 얻을 수 있다.

$$B_1\vec{R}_i = B_1\vec{B}_2 + B_2\vec{R}_i$$

③ BF의 경우

BF의 경우에 대해서는 스케일감소(scale down) 기법을 이용하여 R_i 에 일치하는 MB에 대한 흐름벡터(flow vector) $R_i\vec{B}_1$ 를 먼저 알아낸다. 다음에, B_2 의 순방향-예측된 움직임벡터 $B_2\vec{R}_i$ 를 이용하여 $B_1\vec{B}_2$ 를 다음과 같이 계산할 수 있다.

$$-B_2\vec{R}_i = R_i\vec{B}_1 + B_1\vec{B}_2$$

비슷하게, 스케일감소(scale down)기법을 이용하여 R_i 에 일치하는 B_2 에서의 MB에 대한 흐름벡터 $B_2\vec{R}_i$ 를 알아낸다. 그러면, B_1 의 역방향-예측된 움직임벡터 $B_1\vec{R}_i$ 를 이용하여 $B_1\vec{B}_2$ 를 다음과 같이 계산할 수 있다.

$$B_1\vec{R}_i = B_1\vec{B}_2 + B_2\vec{R}_i$$

벡터들이 스케일감소(scale down)기법을 이용하여 레퍼런스 프레임들에 대해서 추정되어왔기 때문에, 실제적인 $B_1\vec{B}_2$ 의 더욱 적합한 추정값을 얻기위해 이들 두 벡터들의 평균을 취한다.

④ FB의 경우 - 두 B프레임들 사이에 컷이 존재 정의 불가능.

4. DCT계수와 움직임에 의한 색인

MPEG으로 인코딩된 일련의 비디오 정보가 주어진다면, 다양한 질의들이 가능하도록 인덱스하는 것이 좋다. 본 논문에서, 각 장면들의 키 프레임은 공간적차원을 표현하는 DC계수들과 시간적 차원을 표현하는 흐름벡터들을 그 장면에 대한

인덱스로 이용하였다.

가. 공간적 유사성

두 프레임들 사이의 공간적 유사성은 프레임들이 휘도, 색도, 질감, 형상과 같은 비슷한 공간적 특성들을 갖는다는 것을 의미하므로, 이 유사성에 대한 실험은 그런 특성들을 표현하는 값들을 비교하는 것을 의미한다. 픽셀 영역에서는 색과 휘도 값들은 각각의 픽셀에 연관된 값들로 표현되지만, 압축된 영역에서는 64DCT 계수들은 8×8 블록의 픽셀들 모두에 대한 값들을 의미한다. DCT 계수들은 압축된 영역에서 각각의 프레임들의 공간적 정보를 의미하고 DC계수는 그 블록에 대한 평균적인 명암을 나타낸다. 본 논문에서는 각각의 프레임들을 압축해제하는 것을 원치않기 때문에, 휘도와 색도성분의 DC계수들을 이용하여 비교분석한다.

나. 시간적 유사성

많은 비디오 클립들은 상당히 비슷한 내용의 샷을 갖는다. 만약 비디오 클립형태로 질의가 일어난다면 질의 키 프레임(query key frame)에 일치하는 매크로블록들의 움직임정보를 비교하여 인덱싱할 수 있다. 그리고 가장 비슷한 움직임을 갖는 키 프레임을 찾기위해 그리고 최적으로 일치하는 키 프레임을 넘겨주기위해 공간적 정보를 비교한다. 본 논문의 초점은 잡음과 작은 변화들에 덜 민감한 움직임의 방향에 두고 있다.

5. 실험 및 검토

간단하고 효과적인 인덱싱 방법을 제공하기 위해 본 시스템은 공간과 시간적 유사성 기법을 결합하였다. 첫 번째, DC계수들을 이용하여 키 프레임(key frame)으로 인덱스하였다. 이런 DC계수들은 질의 프레임이 비디오 크립의 I 프레임이면 손쉽게 이용가능하다. 만약 B나 P프레임이면, 계수(coefficient)들은 위에서 언급했던 기법을 이용하여 추정되어진다. 질의가 단 하나의 프레임으로 구성되어진다면 처음으로 가장 비슷한 몇 개의 프레임들을 돌려줄수 있고 그것을 사용자 브라우즈 결과로 생각할 수 있다. 짧은 질의 시퀀스의 경우에서, 질의 시퀀스의 처음 몇 개의 프레임들에 대해 질의하고, 질의에 적당한 것들을 구별하기 위해 테이블에 표기한다. 이런 키 프레임들은 대표 프레임으로 다루어지고, 그들의 순위를 변경하기위해 움직임 정보들을 비교한다.

데이터베이스내의 인덱스에 매우 비슷한 질의가 시스템에 주어지면, 시스템이 타당한 매칭으로 검색하는지를 보기위해 실험되었다. 두 개의 MPEG 비디오 스트림에 대해 10개의 서로다른 질의를 수행하여 그 결과를 검토하였다. 그림 3은 각 스트림에 대한 에러율과 질의결과의 정확성을 보여주고 있다.

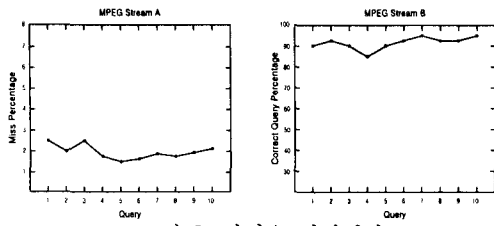


그림 3. 비디오 질의결과

video". Journal of Visual Communication and Image Representation, 7:28-43, 1996.

III. 결 론

본 논문에서는 디코딩 연산들을 수행하는 것 없이 압축된 영역에서부터 MPEG으로 압축된 비디오의 인덱싱과 검색기법을 기술하였다. 비디오는 샷들, 서브샷들, 그리고 장면들로 분해(parse)되고 키 프레임들이 선택되어 진다. 그러므로 그 특징(feature)들은 이들 키 프레임으로부터 추출되어질 수 있다. 본 연구는 I, P, B 프레임들이 동등한 것으로 간주될수 있고, 그 때문에 MPEG 인코딩에 의해 부과된 어떤 제한들을 피할 수 있는 방법을 기술하였다.

IV. 참고문헌

- [1] K. Shen and J. Delp. "A fast algorithm for video parsing using MPEG compressed sequences". In Proc. of the IEEE International Conference on Image Processing, volume 2, pages 252-255, 1995.
- [2] B.L. Yeo and B. Liu. "On the extraction of DC sequence from MPEG compressed video". In Proc. of the IEEE International Conference on Image Processing, volume 2, pages 260-263, 1995.
- [3] F. Idris and S. Panchanathan. "Indexing of compressed video sequences". In Proc. of the SPIE on Storage and Retrieval for Still Image and Video Database IV, volume 2670, pages 247-253, 1996.
- [4] E. Ardizzone, M. La Casia, V. Di Ges and C. Valenti "Content-based indexing of image and video databases by global and shape features". In Proc. of the International Conference Pattern Recognition, pages 140-144, 1996.
- [5] G. Ahanger and T.D.C. Little. "A survey of technologies for parsing and indexing digital