

DCT 기반의 계층부호화를 이용한 비디오 스트리밍 연구

한 승균, 서 덕 영
경희대학교 전자공학과
전화 : 031-201-2963

A Study on video streaming by using DCT-based scalability encoding

Seung Gyun Han, Doug Young Suh
School of Electronics & Information Electronic Engineering Major, Kyunghee University
E-mail : sghan@newmedia.kyunghee.ac.kr

Abstract

This paper suggests real-time video streaming method by using DCT-based scalability, and evaluates and analyzes the function. It is similar to using lowpass filter. That is, as following figure, this method is to split the encoded data in splitter and transmit it, and to decode the data according to the situation. This method can be applied to any video CODEC which is based on DCT. Therefore, this thesis suggests easy video streaming method by using DCT-based scalability, and shows the result of experiment. By using suggested scalability, calculations are reduced, and spacial scalability is realized. Moreover, the objective data which meet user's need according to the network condition and choose the appropriate scalability according to the capability of terminal can be extracted. And it is possible to apply any resources according to the specificity of image.

I. 서론

최근 컴퓨터와 통신 기술의 급속한 발달로 인해 인터넷을 통한 멀티미디어 데이터의 전송이 보편화되면서

정보 통신망 사용자들은 시간과 장소에 상관없이 다양한 정보에 접근할 수 있게 되었으며 이용할 수 있는 멀티미디어 정보의 양도 폭발적으로 증가하게 되었다. 하지만 정보의 량이 늘어날수록 네트워크 상황이나 단말기의 상황은 이에 따라 가질 못하게 되어 정보의 가공이 필요하게 되었다. 그 중 하나로 scalability는 인터넷 및 협대역 무선망에서의 비디오 전달, 비디오 데이터베이스 브라우징에 중요한 기능이다. 특히 현재의 인터넷에서는 전송대역에 대한 네트워크의 서비스 품질이 보증되고 있지 않으므로, 동영상을 높은 부호화 속도로 안정하게 전송하는 것이 어렵다. 또 동영상 신호를 소프트웨어로 처리하는 것이 일반화되었지만, 처리 능력이 낮은 단말에서는 수신한 부호화 데이터를 완전히 복호할 수 없는 경우도 문제이다. 보통 두 종류의 데이터, 저해상도와 고해상도의 두 종류의 데이터나 큰 영상과 작은 영상 두 종류의 데이터를 준비해 두고, 네트워크와 단말의 상태가 양호할 때는 큰 영상(고해상도)의 동영상을 복호하고, 네트워크와 단말의 상태가 악화되었을 경우는 화질을 저하시키는 것이 아니라, 작은 영상(저해상도)의 품질을 보증하는 것을 생각할 수 있다. 또한 사용자의 요구에 따라서 그 요구에 합당한 영상을 서비스 할 수 있다.

본 논문에서는 DCT 기반의 계층부호화를 이용한 비디오 스트리밍의 방법을 제안하고 그 성능을 평가 분석한다. 2장에서는 계층부호화의 종류와 그 특징을 설명

하고, 3장에서는 본 논문에서 제안하고 있는 계층부호화를 설명한다. 4장에서는 제안한 DCT 기반의 계층부호화를 이용한 방법과 기존의 방법들과의 성능을 비교, 분석하고 마지막으로 5장에서는 본 논문의 결론을 제시한다.

II. 계층부호화

2.1 MPEG-4에서의 계층부호화(Scalability)

계층부호화는 부호화시 데이터를 두 개 이상의 열(stream)로 분리, 발생시킨다. 이 때 가장 기본이 되는 비트열을 기본계층(base layer)라 하고 나머지 열을 차례로 고급계층1(enhancement layer 1), 고급계층2,...라 한다. 고급계층은 반드시 기본계층의 데이터가 있어야 디코딩이 가능하며 기본계층만으로 디코딩 했을 경우에 비해 방식에 따라 공간적, 시간적 해상도 등이 뛰어나다. MPEG-4에서는 SNR, 공간적, 시간적 계층부호화 등이 있다. SNR 방식은 기본계층과 고급계층간에 SNR (signal to noise ratio)을 달리해서 코딩하는 방식이며 공간적 계층부호화는 기본계층과 고급계층간의 공간적인 해상도를 다르게 해 코딩하는 방법이다. 그리고 시간적 계층부호화는 기본계층과 고급계층간의 시간적인 해상도를 다르게 하는 방법이다.

부호화 당시 DCT를 이용하여 고주파수 부분의 데이터를 자르는 부호화 방법이 있는데 이러한 계층부호화 인코더는 Benzle^[1], Wells^[2]와 Seeger^[3]에 의해 여러 방법이 제안되었다. 계층부호화 된 비트열 전송은 ATM의 경우 VC를 이용함으로써 가능하다. 또 인터넷을 이용하는 H.323 시스템에서도 기본계층과 고급계층을 하나의 비트열로 묶어서 보내는 방법과 분리시켜 여러 비트열로 전송하는 방법^[4] 등 계층부호화 된 비트열 전송 방법이 있다. 이 때 계층부호화 인코더의 형태는 여러 가지가 있을 수 있으며 반드시 비트열의 syntax는 규정에 부합되어야 한다. 이러한 계층부호화 인코더는 Faber^[5], Nokolaidis^[6], Ghanbari^[7] 등에 의해 여러 방법이 제안되었다.

III. 제안하는 Scalability

3.1 제안하는 계층부호화

본 논문에서는 DCT를 기반으로 하는 계층부호화를 제안하는데 이는 lowpass filter를 사용하는 것과 유사하다. 그림 1과 같이 먼저 부호화가 되어 오는 데이터를 splitter에서 분할하여 전송하고, 상황에 따라 필요한(요구한) 만큼의 데이터를 merger에서 재조립하여 복호화를 하는 방식이다. 이러한 방식은 DCT를 기반으로 하는 모든 비디오 코덱에 적용할 수 있다.

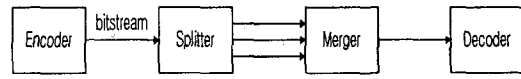


그림 1. 제안하는 계층부호화

3.2 계층의 구별

Splitter에서는 계층별로 데이터를 나누는 작업을 하는데 DCT를 기반으로 해서 기본계층과 고급계층들로 나누게 되는데, 이는 낮은 주파수에서부터 점점 고주파로 옮겨가며 계층을 구분한다. 기본계층으로는 MPEG-4의 기본적인 헤더와 움직임벡터 그리고 DC를 포함한 데이터가 된다. 그리고 고급계층1의 데이터는 DCT에서 DC를 제외한 2x2 안의 데이터가 포함되고, 고급계층2에서는 DCT에서 2x2를 제외한 4x4 안의 데이터 값을 포함하고, 고급계층3은 DCT에서 4x4를 제외한 8x8 안의 데이터가 포함된다. 그림 2는 계층별로 포함되는 데이터의 모식도를 나타낸다.

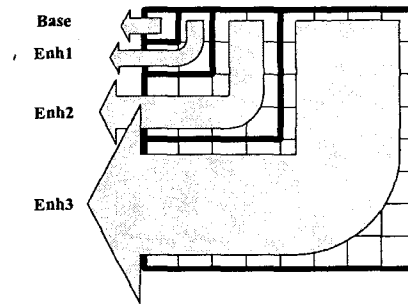


그림 2. 계층별로 포함되는 데이터의 모식도

3.3 제안하는 계층부호화 헤더 구성

DCT 기반에서 저주파에서 고주파로 옮겨가며 계층을 구분해야 한다. 이때 MPEG-4에서의 기본적인 헤더 즉, VO(video object) 헤더는 32 비트이고 VOL(video object layer)의 헤더는 134비트이다. 이는 비트열의 맨 처음에 나오는 헤더이기도 하다. 여기에 VOP(video object plane) 헤더는 53 인데 이는 PVOP일 경우이고 IVOP일 경우는 49비트이다. 그림 4는 MPEG-4의 기본적인 헤더의 구성을 보여 주고 있다.

보통 한 프레임당으로 볼 때 본 실험에서는 이러한 MPEG-4의 기본적인 헤더에 각 계층별로 헤더가 더 붙어야 한다.

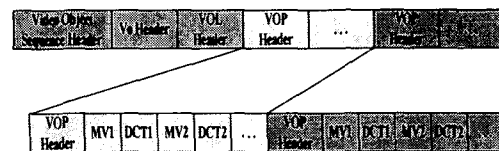


그림 3. MPEG-4의 기본적인 헤더 구성

IV. 실험방법 및 실험결과

실험에 사용한 코덱은 MPEG-4를 이용했다. MPEG-4 에서는 영상을 움직임과 복잡도의 정도에 따라 Class A~C로 구분하는데[그림 4. 참조] 본 실험에서는 Class A 계열의 Hall Monitor(Hall), Mother & daughter(MaD900), Class B 계열의 Foreman, News, Class C 계열의 Mobile & Calender(Mobile), Stefan CIF(352×288) 포맷을 입력영상으로 이용하였다. 인코딩은 단일계층의 화질을 변화 시켜 가면서 실험하였다. 실험의 조건으로는 총 프레임은 100frame, 프레임률은 30Hz이다. 그리고 데이터 분리를 사용했다.

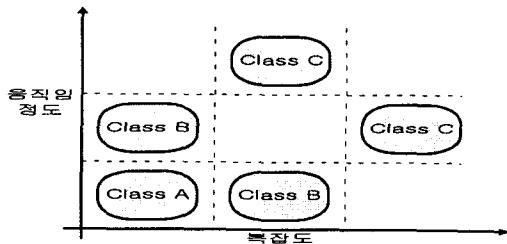


그림 4. 영상의 Class 분류

4.1 특징에 따른 영상의 분류

본 실험에 사용한 영상을 움직임과 energy compaction(EC)의 정도에 따라 다음과 같이 분류했다.
- 움직임 정도 : 전체 프레임에 대해 각 매크로블록들의 움직임벡터의 스칼라양을 합한 값으로 정의했다. 여기서

$$\text{Scalar of MV} = \sqrt{(MV_x^2 + MV_y^2)}$$

- EC : 에너지의 밀집정도. 이는 화면의 복잡도를 측정하는 척도이다.

$$EC = \frac{\sum_{i=0}^4 \sum_{j=0}^4 AC_{coefficient}}{\sum_{i=0}^7 \sum_{j=0}^7 AC_{coefficient}} \times 100\%$$

4.2 Class의 종류에 따른 비트율의 비교

비트율은 결과 그래프에서와 같이 보편적으로 Class A(MaD900)가 가장 작고, Class B(Foreman)가 그 다음으로 나타나고, Class C(Mobile)가 가장 크게 나타난다. 꼭 이렇게만 결과가 나타나는 것은 아니다. 다시 말해서 Class A ≤ Class B ≤ Class C 이지만 꼭 그렇지만은 않다.

그림 5.에서 양자화 계수가 작으면 작을수록 고급 계층1,2,3의 데이터량이 기본계층의 데이터량에 비해 상대적으로 클 수 있고, 기본계층의 데이터량은 다른 데이터량에 비해 거의 일정함을 알 수 있다. 이

는 양자화 계수가 작으면 작을수록 DCT의 고주파 성분이 많기 때문이다.

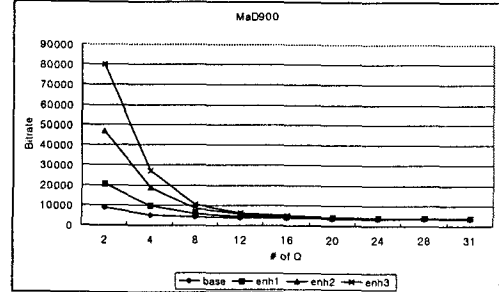


그림 5. Class A의 양자화 계수에 따른 비트율(MaD900)

4.3 Class의 종류에 따른 비트율의 상대적 비교

비트율의 상대적 비교란 총 비트율을 1로 보았을 때 각 계층이 차지하는 비율을 나타낸 것이다. 이 그래프 또한 가로축의 # of Q는 양자화 계수를 의미한다. 기본계층과 고급계층3의 데이터량이 같아지는 곳은 클래스 A와 B는 양자화 계수가 4와8사이에서, 클래스 C는 양자화 계수가 20 근처에서 그 비율이 같아지는 특성을 가진다. 그림 6.을 보면 더 명확하게 알 수 있다. 이는 영상이 복잡하면 할수록 고급계층의 데이터가 많음을 다시 확인할 수 있다.

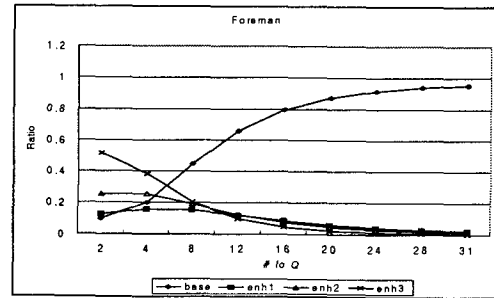


그림 6. Class B의 양자화 계수에 따른 상대적 비트율(Foreman)

4.4 계층부호화에 따른 PSNR 비교

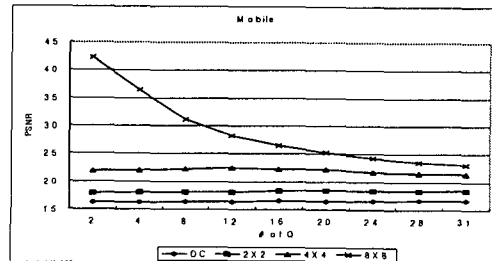


그림 7. Class C의 양자화 계수에 따른 PSNR (Mobile)

본 논문에서 제안한 계층부호화에 따라 복호화한 영상의 PSNR은 그림 7.과 같이 나타난다.

이때에도 상대적 비트율 비교에서도 그랬듯이 클래스 A와 클래스 B의 영상은 기본계층의 PSNR이 20~25dB 사이에 있고, 클래스 C는 16~17dB 사이에 있다.

4.5 네트워크 특성에 따른 PSNR

네트워크의 상황은 항상 같을 수만 있는 게 아니다. 원활히 소통이 되던 네트워크가 갑자기 많은 데이터가 폭주해서 네트워크의 혼잡이 가중되거나, 네트워크의 상황이 혼잡했다가 갑자기 소통이 원활하게 되었을 때가 있다. 본 논문에서는 상황이 좋았다가 나빠질 경우를 고급계층3에서 고급계층2로 변환시켜 복호화했고, 상황이 나빠졌다가 좋아질 경우는 고급계층2에서 고급계층3으로 변환시켜 복호화 했다. 이 때 실험 환경은 Class C의 Mobile을 사용했고, 51프레임에서 네트워크의 상황변화를 가정했다. 네트워크의 상황에 따른 PSNR의 결과는 그림 8과 같다.

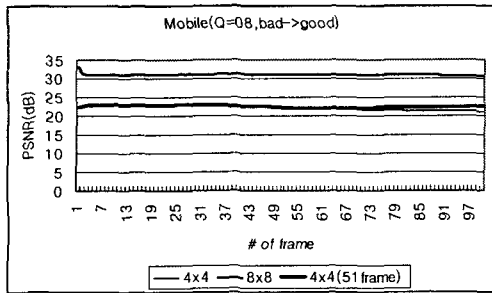


그림 8. 네트워크 특성에 따른 PSNR
(Mobile, bad→good)

그림 8은 네트워크 상황이 나쁜 쪽에서 좋은 쪽으로 바뀌었을 경우(51프레임에서, bad→good)의 실험 결과는 나타낸 그림이다. 이 결과는 단순히 수식으로 계산을 할 수 없다. 이는 고급계층2의 데이터에 고급계층3의 데이터가 유입되어 상황이 바뀌었을 경우의 프레임에서부터 본래의 고급계층2의 PSNR보다 낮았다가 점차 높아진다. 이러한 상황에서는 고급계층2에서 고급계층3으로의 PSNR의 갑작스런 변화를 볼 수 없다. 그러나 이번 실험에서는 인트라로 부호화된 메크로블록이나 refresh된 프레임을 생각하지 않았다. 만약 이러한 상황을 설정한다면 보다 더 나은 PSNR을 기대해도 좋을 것 같다. 또한 이러한 상황은 사용자의 요구에 따라 변화될 계층부호화별 PSNR이라고 생각해도 된다.

V. 결론

네트워크를 통해 멀티미디어 데이터를 전송할 경우 정보의 양이 늘어날수록 네트워크 상황이나 단말기의 상황은 이에 따라 가질 못하게 되어 정보의 가공이 필요하게 되었다. 이에 본 논문에서는 DCT 기반의 계층부호화를 이용한 쉽게 구현 가능한 비디오 스트리밍 방법에 대해 제안하고 실험한 결과를 보였다. 실시간 서비스에서 기본적 헤더와 움직임 벡터, DC를 기본계층으로, 기본계층을 제외한 2×2의 DCT 데이터를 고급계층1로, 고급계층1이하를 제외한 4×4의 DCT 데이터를 고급계층3으로, 나머지 데이터를 고급계층3으로 계층을 분리하여 재조립함으로써 쉬운 계층부호화를 실현하였다. 여기에 제안한 계층부호화를 사용하여 더 많은 계산량을 줄이고, 공간적 계층부호화를 실현시킬 수 있음을 보였다. 또한 네트워크의 상황에 따라 수신측의 요구, 단말기의 성능에 따라 알맞은 계층부호화를 선택할 수 있는 객관적인 자료를 추출 할 수 있었다. 또한 영상의 특성에 따라 자원의 활용을 적극적으로 응용할 수 있음을 보였다.

참고문헌

- [1] Ulrich Benzler, "Scalable multi-resolution video coding using a combined subband-DCT approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 6, pp. 845-854, Dec. 1998.
- [2] N. Wells, P. Tudor, "Standardization of scalable coding schemes," *Proc. IEEE ISCAS 94, Tutorials* pp121-130, May 1994.
- [3] Seger, A., "Hierarchical channel coding for Rayleigh and Rice fading," *IEEE Globecom, Communication Theory Mini Conference*, Phoenix, pp208-212, November 1997.
- [4] *Packet based multimedia communications systems*, ITU-T Standard H.323V2, Mar. 1997.
- [5] Niko Faber, Bernd Girod, John Villasenor, "Extensions of ITU-T Recommendation H.324 for error-resilient video transmission," *IEEE Commun. Mag.*, pp. 120-128, June. 1998.
- [6] T. Nikolaidis, M. Hasan, F. Marvasti, "H.263-based Two Layer Codecs for ATM Environments," in *Proc. IEEE ITC'98*, Porto Carras, Greece, vol. 3, pp. 32-36.
- [7] *Video coding for low bit rate communication*, ITU-T Standard H.263V2, " Sept. 1998.