

동영상 크기 및 품질 예측에 기반한 동적 동영상 어댑테이션

김종항, 남기용⁰, 이상민, 낭종호
 서강대학교 대학원 컴퓨터학과
 {kayin, babogiyo⁰, lesmin, jhnang}@mlneptune.sogang.ac.kr

A Dynamic Video Adaptation Scheme based on the Predictions of the Size and the Quality of Encoded Video Streams

Jonghang Kim, Kiyong Nam⁰, Sangmin Lee, Jongho Nang
 Dept. of Computer Science, Sogang University

요약

Proxy를 이용한 동적 동영상 어댑테이션[1]은 이동 단말기나 현재 네트워크 상태의 특성을 고려해 동적으로 동영상을 변형할 수 있다는 장점이 있다. 하지만 기존에 제안된 동영상 어댑테이션 방법들은 품질 측정을 위해 반복적인 디코딩, 인코딩을 해야 하기 때문에 적절한 형태의 동영상을 생성하는데 많은 시간이 걸려서 지연시간이 최우선적으로 고려되는 실제 상황에서는 이용하기가 힘들다. 본 논문에서는 반복적인 디코딩, 인코딩 작업 없이 어댑테이션된 동영상을 생성하는 동적 동영상 어댑테이션 방법을 제안한다. 인코딩된 동영상의 파일의 크기와 품질에 초점을 맞추어 비디오 코덱의 특성을 분석하고, 그 결과를 테이블로 만들어 Proxy에 저장해둔다. 이동 단말기가 동영상을 요청하면, Proxy에서는 해당 코덱의 분석 결과 테이블을 참조하여 가능한 한 최고의 품질로 디코딩 및 인코딩을 하여 어댑테이션된 동영상을 전송하게 된다.

1. 서론

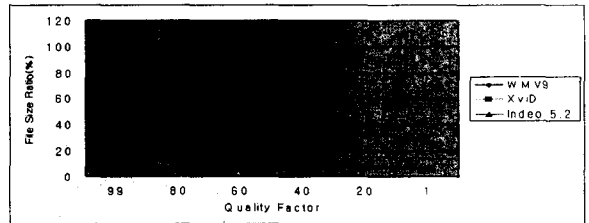
무선 환경에서는 이동 단말기의 여러 가지 한계 때문에 인터넷상의 동영상을 제공받기 위해서 동적 동영상 어댑테이션이 필요하다. 하지만, 기존의 방법들은 서버나 Proxy에서의 반복적인 디코딩, 인코딩을 통하여 분석하므로 지연 시간이 길어지는 문제점이 발생한다. Proxy상에서 동영상의 어댑테이션에 관한 기존 연구로서는 Utility Function에 기반을 두어 가로축의 비트율과 세로축의 Utility(동영상의 전체적인 품질)를 고려해 최적의 콘텐츠 버전을 찾는 방법[2]이 있다. 이 방법은 일정한 프레임 드랍에서 DCT계수의 변화에 따른 그래프를 저장해 놓고, 원하는 비트율에 따라 해당 비트율에서 가능한 한 가장 품질이 높은 점을 선택해 그 위치에 해당하는 프레임 드랍과 DCT계수를 적용하여 콘텐츠를 변형시키는 방법이다. 이 방법은 비트율의 변화에 빠르게 대응할 수 있다는 장점이 있지만 콘텐츠 버전 별로 미리 측정된 품질과 비트율 정보가 필요해 지므로 많은 인코딩과 디코딩을 거쳐야 하기에 서비스 시간이 지연되는 문제점이 발생한다.

본 논문에서는 이런 문제점을 해결하기 위하여 미리 분석한 코덱의 특성 및 콘텐츠의 특성을 이용하여 반복적인 인코딩, 디코딩 없이 양질의 품질을 보장하는 동영상을 생성하는 방법을 제안한다.

2. 동영상 크기의 예측 방법

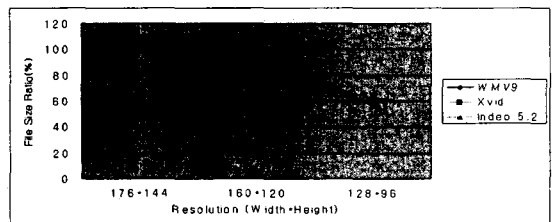
어댑테이션된 동영상 크기를 예측하기 위해서, Quality Factor와 해상도, 프레임률에 따른 인코딩 후의 동영상의 크기를 예측하는 것이 필요하다. 이들은 일정한 규칙을 가지고 있어 이를 코덱마다 테이블로 저장해두면 별도의 디코딩, 인코딩

과정 없이 어댑테이션된 동영상의 크기를 예측할 수 있다.



<그림 2-1> 코덱별 Quality Factor에 따른 파일 크기의 변화율

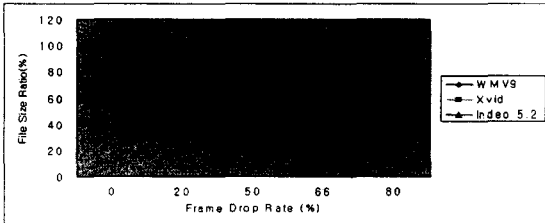
Quality Factor의 경우, 최대 Quality Factor를 적용하여 인코딩된 결과의 크기를 기준으로 볼 때, Quality Factor를 줄이면서 인코딩한 결과의 크기는 일정한 비율을 가짐을 알 수 있었다.<그림 2-1> 따라서 최대 Quality Factor로 인코딩한 파일의 크기를 기준으로 Quality Factor를 변형시키며 인코딩한 파일의 크기 비율 기록하여 두면, 그 기록을 통하여 매번 다시 인코딩하지 않아도 해상도에 따른 파일 크기의 변화를 예측할 수 있다.



<그림 2-2> 코덱 별 해상도에 따른 파일 크기의 변화율

본 연구는 한국과학재단에서 지원하는 특정기초연구사업으로 수행하였음 (과제번호 : R01-2002-000-00141-0)

해상도의 경우, 단말기가 받아들일 수 있는 최대 해상도를 적용하여 인코딩된 결과의 크기를 기준으로 볼 때, 해상도를 줄이면서 인코딩한 결과의 크기 또한 일정한 비율을 가짐을 알 수 있었다.(<그림 2-2>) 따라서 최대 해상도로 인코딩한 파일의 크기를 기준으로 해상도를 변형시키면서 인코딩한 파일의 크기의 비를 기록하여 두면, 그 기록을 통하여 매번 다시 인코딩하지 않아도 해상도에 따른 파일 크기의 변화를 예측할 수 있다.



<그림 2-3> 코덱 별 프레임 제거에 따른 파일 크기의 변화율

프레임들의 경우도 프레임 드랍을 전혀 하지 않았을 때를 기준으로 프레임 드랍을 증가시키면서 인코딩한 결과의 크기 또한 일정한 비율을 가짐을 알 수 있다.(<그림 2-3>) 따라서 프레임 드랍을 전혀 하지 않은 상태의 인코딩된 파일의 크기를 기준으로 프레임 드랍을 늘리면서 인코딩된 파일의 크기의 비를 기록하여 두면, 그 기록을 통하여 매번 다시 인코딩하지 않아도 프레임 드랍에 따른 파일 크기의 변화를 예측할 수 있다.

이렇게 Quality Factor와 해상도, 프레임률을 각각 최대로 하여 인코딩한 파일 크기와 각 값을 변화시키면서 인코딩한 파일의 크기의 비를 테이블에 기록한 뒤, <식 2-1>와 같이 어댑테이션된 파일의 크기를 예측할 수 있다.

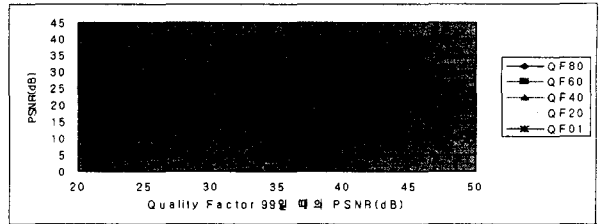
$$C^{filesize} = C^0_{filesize} * Q_{codec}(C_{QF}) * R_{codec}(C_{resolution}) * F_{codec}(C_{fps}) \quad \text{<식 2-1>}$$

$C^{filesize}$ 는 어댑테이션된 콘텐츠 C' 의 예측하려는 파일 크기를 의미한다. 우선 각 어댑테이션 기법의 시간 복잡도와 공간 복잡도를 모두 최대 품질로 인코딩한 콘텐츠 C^0 의 파일 크기 $C^0_{filesize}$ 를 측정한다. 여기에 어댑테이션 기법 별 제어 값에 따른 파일 크기의 변화율을 곱하면 콘텐츠 C' 의 파일 크기가 예측 가능하다. $Q_{codec}(x)$, $R_{codec}(x)$ 와 $F_{codec}(x)$ 는 본 논문에서 앞서 언급한 각각의 어댑테이션 기법에 따른 콘텐츠 X 의 코덱별 파일 크기 비율을 의미한다. 각각의 함수에는 콘텐츠 C' 를 만들 때 이용한 제어 값을 넣어주는데 C_{QF} , $C_{resolution}$, C_{fps} 는 각각 C' 의 Quality Factor, 해상도, 프레임률을 의미한다.

3. 동영상 품질의 예측 방법

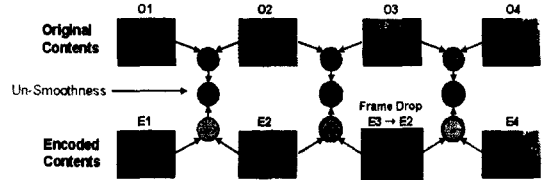
Quality Factor에 의한 공간적 품질을 측정하는 방법은 PSNR을 사용하는데, 우선 Quality Factor를 최대값으로 하여 인코딩을 한 콘텐츠 버전의 PSNR값을 측정하여 이를 기준으로 삼는다. 이후 Quality Factor를 낮추면서 PSNR을 측정하여 비교한 결과 <그림3-1>과 같이 일정한 비율을 가지는 특징이

있음을 알 수 있다.



<그림 3-1> XviD 코덱을 이용 시 Quality Factor의 조절에 따른 PSNR

프레임들은 시간적 품질을 결정하는 인자로서, 시간적 품질은 Frame-Difference[3]를 사용한다. 프레임들이 변화할 때마다 시간적 품질을 다시 구하는 것은 시간 지연이 크므로, 다음의 원리를 적용한다.



<그림 3-2> 프레임 제거에 따른 Un-Smoothness의 변화

<그림 3-2>과 같이 인코딩된 콘텐츠의 세 번째 프레임, E3가 제거된 상황을 설정해보자. 이 때 프레임 제거에 영향을 받지 않는 첫 번째 프레임과 두 번째 프레임의 Un-Smoothness는 <식 3-1>, 세 번째 프레임과 네 번째 프레임의 Un-Smoothness는 <식 3-2>와 같게 된다.

$$\begin{aligned} & |(O2 - O3) - (E2 - E3)| \\ & = |(O2 - O3) - (E2 - E2)| = |(O2 - O3)| \quad \text{<식 3-1>} \end{aligned}$$

$$\begin{aligned} & |(O3 - O4) - (E3 - E4)| = |(O3 - O4) - (E2 - E4)| \\ & = |(O3 - O4) - (O2 - O4)| = |(O3 - O2)| \quad \text{<식 3-2>} \end{aligned}$$

실제로 인코딩된 콘텐츠는 어느 정도의 손실이 발생하게 되므로 이 식을 그대로 적용하면 어느 정도의 측정오차는 발생하지만 오차의 정도는 프레임간의 차이에 비하면 굉장히 적은 양이므로 빠른 시간 안에 계산이 가능하게 된다.

공간적 품질(Spatial Quality)과 시간적 품질(Temporal Quality)의 두 가지를 조합하여 나타내는 전체적인 품질(Utility)[2]은 다음과 같이 구한다. 기존의 이미지간 품질을 측정하는 방법들은 원본과 변형본의 해상도가 같음을 전제하므로, 본 논문에서는 어댑테이션된 콘텐츠의 해상도가 단말기의 해상도에 얼마만큼 적합한지를 $Suitability_{resolution}$ 로 정의하고, 다음과 같이 공간적인 품질과 시간적인 품질을 곱해 동영상 콘텐츠의 Utility로 사용한다. $Suitability_{resolution}$ 은 어댑테이션된 콘텐츠의 전체 픽셀의 개수를 단말기가 최대 해상도일 때의 픽셀수로 나누어 구한다.

$$\begin{aligned} \text{Utility} & = Suitability_{resolution} * \text{Quality} \\ & = Suitability_{resolution} * \text{Quality}_{spatial} * \text{Quality}_{temporal} \\ & = Suitability_{resolution} * \text{Normalized PSNR} * (1 - \text{Normalized UnSmoothnes}) \end{aligned}$$

$Suitability_{resolution}$

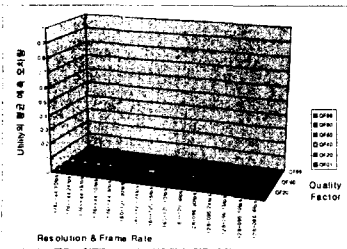
$$= \frac{\text{resolution of adpted content}}{\text{screen size of client}} = \frac{C'_{width} * C'_{height}}{Client_{width} * Client_{height}}$$

$$(C'_{width} \leq Client_{width}, C'_{height} \leq Client_{height})$$

4. 실험 및 분석

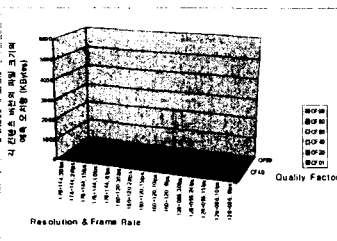
4.1. 실험 결과 비교 및 분석

본 논문에서는 콘텐츠 버전별로 파일 크기 및 품질을 예측한 값과, 실제로 콘텐츠를 만들어 파일 크기와 품질을 측정된 값을 비교하고 예측의 정확도를 평가하였다. 어댑테이션 기법별 제어 값은 Quality Factor는 99~1 사이의 범위를 6단계로 나누었으며, 해상도는 H.263에서 제시한 QCIF(176*144), SQCIF(128*96)와 MPEG-1에서 제시한 SIF(160*120) 세 단계를 사용하고, 프레임 제거율은 0%, 20%, 50%, 66%, 80%로 늘려가며 실험을 하였다. 이 경우 조합 가능한 콘텐츠의 버전의 총 개수는 90가지가 된다. 본 실험은 15초~30초 길이의 20개의 샘플 동영상에 대해 각각 90개의 버전을 실제로 만들어 측정된 값과, 한 개의 버전만을 생성하고 나머지 버전은 예측한 값을 이용하여 서로 비교해 보았다.



<그림 4-1> 콘텐츠 버전 별 파일 크기의 평균 예측 오차량

양임을 알 수 있다. 평균 오차량이 가장 큰 버전에서도 실제 파일 크기와 비교하면 4.5%의 정도의 오차율을 보여주었으며, 가장 오차율이 큰 경우는 콘텐츠의 파일 크기가 매우 작으므로 오차의 정도가 큰 문제가 되지 않는다.



<그림 4-2> 콘텐츠 버전 별 유틸리티의 평균 예측 오차량

고려하면 무시해도 좋은 수준의 예측 오차량이다.

<그림 4-1>은 20개의 샘플에 대해 콘텐츠 버전 별로 실측과 예측 파일 크기 차의 절대값을 평균낸 그래프이다. 실측 파일 크기가 90개 버전에 대해 40K Bytes~6M Bytes, 평균 852K Byte임을 감안하면 예측으로 인해 발생으로 발생하는 평균 오차량은 매우 적은

컨텐츠 버전별로 실측과 예측을 통해 얻은 유틸리티(Utility)의 절대값의 차를 낸 뒤에 20개의 샘플에 대해 평균값을 내어 비교를 해 보았다. 그 결과 <그림 4-2>와 같이 유틸리티의 평균 예측 오차량은 최대 0.038이 나타났으며, 이는 유틸리티의 범위가 0~1 사이인 점을

4.2. 기존 연구 방법과의 성능 비교

S. F. Chang이 제안한 방법[2]과 본 논문에서 제안한 방법의 비교 실험을 통해 각각의 특징을 분석하고, 본 논문의 제안 방법에 대한 타당성을 검증토록 하였다. S. F. Chang이 제안한 방법[2]은 콘텐츠의 해상도의 조절은 고려하지 않았으며, 프레임 제거율을 몇 단계로 나눈 다음 단계마다 Quality Factor를 달리한 몇 개의 버전을 인코딩을 거쳐 실제로 만든 뒤, 각각의 파일 크기 및 품질을 측정하고 그 외의 버전에 대해서는 보간법을 이용해 예측하는 방법이다.

샘플 1은 640*352의 해상도에 총 578 프레임이며 변화율 테이블과 편차가 가장 큰 샘플이다. 샘플 2는 672*352의 해상도에 총 500 프레임이며 변화율 테이블과 편차가 가장 작은 샘플의 경우이다.

<표 4-1> 크기와 품질 측정에 소요되는 시간

실험방법	실험샘플	인코딩 소요 시간(초)	품질 측정 시간(초)	전체 소요 시간(초)
S.F.Chang	샘플1	109	115	224
	샘플2	98	102	200
제안한 방법	샘플1	8	32	40
	샘플2	7	30	37

<표 4-2> 파일 크기와 품질의 평균 예측 오차

실험방법	실험샘플	파일 크기의 평균예측오차율	품질의 평균예측오차량
S.F.Chang	샘플1	14.19%	0.022
	샘플2	16.1%	0.027
제안한 방법	샘플1	16.15%	0.028
	샘플2	6.21%	0.0034

5. 결론

기존의 방법과 제안한 방법을 비교 실험한 결과, 어댑테이션 시간을 최대 91%까지 줄일 수 있었으며 이에 반해 파일 크기와 품질의 예측 오차율은 각각 평균 10%와 4% 수준으로 나타났다. 제안한 방법을 이용하면 무선 환경에서 인터넷상에 존재하는 동영상 콘텐츠의 실시간 서비스가 가능하며, Proxy를 이용한 효율적인 동적 동영상 어댑테이션 시스템 구현이 가능할 것이다.

6. 참고문헌

[1] Ingo Elsen, Frank Hartung, Uwe Horn, Markus Kampmann and Liliane Peters, "Streaming Technology in 3G Mobile Communication Systems," *IEEE Computer Magazine*, Vol. 34 Issue: 9, Sep. 2001 pp. 46-52.

[2] Yong Wang, Jae-Gon Kim and Shih-Fu Chang, "Content-Based Utility Function Prediction for Real-Time MPEG-4 Video Transcoding," *Proceedings of ICIP'2003, Barcelona, Spain, Sep., 2003*, pp. 14-17.

[3] Christoph Kuhmunch, Gerald Kuhne, Claudia Schremmer and Thomas Haenselmann, "Video-scaling Algorithm based on Human Perception for Spatio-Temporal Stimuli," *Proceedings of SPIE, Multimedia Computing and Networking*, vol. 4312, San Jose, California, USA, Jan., 2001, pp. 13-24.