

Scalability 기능을 가진 MPEG-4 비디오 부호화기의 구현

임성재*, 이정우*, 전형국**, 호요성*

* 광주과학기술원 정보통신공학과

** 한국전자통신연구원

전화: 062-970-2263 / 핸드폰: 019-9153-1705

Implementation of MPEG-4 Video Codec with Scalability

Seong-Jae Lim*, Jeong-Woo Lee*, Hyung-Kook Jun**, and Yo-Sung Ho*

* Kwangju Institute of Science and Technology (K-JIST)

** Electronics and Telecommunications Research Institute (ETRI)

E-mail: {sjlim, jeongwoo, hoyo}@kjist.ac.kr, hkjun@etri.re.kr

Abstract

Scalability is a desirable functionality that can provide a solution to various problems for multimedia services. A single scalable bitstream can satisfy quality requirement for several decoders of different capabilities. Since the scalable encoder is able to adjust the amount of bitstream according to the available channel bandwidth, scalable transmission is advantageous in error-prone environments. In this paper, we have implemented an MPEG-4 scalable video codec for the home server platform.

I. 서론

최근 멀티미디어 서비스에 대한 요구가 크게 늘어나면서 초저속 오디오/비주얼 부호화 방식의 구현에 많은 관심이 모아지고 있다. 특히, 정보 가전을 위한 홈서버 플랫폼 기술을 개발하기 위해 방대한 정보량을 가지는 멀티미디어 신호를 전송 채널의 대역폭이 작은 유선 통신망이나 무선 채널을 통해 효율적으로 전송하기 위해서는 MPEG-4 비디오 데이터를 압축하고 복원하는 기술의 개발이 필요하다.

최근에 완성된 MPEG-4 객체기반 동영상 부호화 방식은 인터넷이나 이동 통신, 또는 초고속정보통신망을 통한 다양한 멀티미디어 통신 서비스에 사용될 수 있으므로, MPEG-4 기술을 자체적으로 개발하여 확보하는 것이 매우 필요하다. 또한, MPEG-4 표준의 확장성과 호환성의 특징을 최대한 살리기 위해 MPEG-4 코덱을 S/W로 구현할 필요가 있다. 특히, 사용자 단말

기의 성능에 맞는 서비스를 제공하고, 무선통신 환경을 고려해 채널 오류가 많은 환경에서도 사용자가 기본적인 화질을 보장받을 수 있도록 하기 위해 계위적 (scalable) 부호화 기능을 제공하는 영상 부호화 코덱의 개발이 절실하다.

비디오 신호를 포함한 멀티미디어 서비스를 위해 송신측에서 압축된 비트열을 서버를 통해 클라이언트 측에 보낼 경우, 서버와 클라이언트의 처리 능력이 서로 다를 수 있으며, 전송 채널의 가용 대역폭도 변할 수 있다. 이러한 환경에서 데이터를 효과적으로 전송하기 위해서는 시/공간 계위 부호화를 지원하는 기술이 필요하다. 본 논문에서는 MPEG-4 비디오 계위 부호화를 지원하는 인코더/디코더 기술로 홈서버 플랫폼에서 효과적으로 사용될 수 있는 공간적 계위부호화 기술과 시간적 계위부호화 기술을 분석하고 구현한다.

II. MPEG-4 비디오 계위부호화 기술

그림 1은 일반화된 계위부호화를 지원하는 High Level Codec의 구조를 나타낸다.

계위부호화 전처리기의 입력은 비디오 VOP(video object plane)이며, MPEG-4 기본계층(base layer) 부호기의 입력 in_0은 입력 VOP를 전처리기에서 공간적으로 다운샘플링(downsampling)하여 얻는다. 기본계층으로부터 복원된 VOP는 공간적 업샘플링(upsampling)을 수행하는 중간처리기(midprocessor)의 입력이 된다. 전처리기의 다른 출력인 상위 공간계층 VOP는

MPEG-4 확장계층(enhancement layer) 부호기의 입력 in_1이 된다. 기본계층과 확장계층 비트열은 MSDDL (MPEG-4 system description language) 다중화기에 의해 다중화되고 저장되거나 전송된다. 이것은 MSDDL 역다중화기에 의해 MPEG-4 기본계층과 확장계층 복호기로 각각 비트열이 전달된다. 계위부호화 후처리기는 outp_0의 디스플레이를 위해 복호화된 기본계층의 데이터를 공간적으로 업샘플링하므로, outp_1인 확장계층과 비슷한 연산을 수행한다.

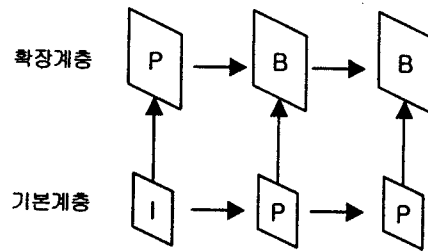


그림 2. 기본계층과 확장계층의 관계

1. 공간적 계위부호화

공간적 계위부호화는 공간 해상도의 계위부호화를 제공하고, 기본계층과 확장계층의 두 계층으로 구성되어 있다. 기본계층은 해상도가 낮거나 크기가 작은 기본적인 비트열이며, 확장계층은 추가적으로 적용되어 기본계층 프레임의 공간 해상도를 개선시킨 비트열이다.

1.1 공간적 계위부호화 과정

공간적 계위부호화의 경우, 확장계층 비트열은 영상의 해상도를 개선시키는데 사용된다.

공간적 예측은 참조계층(reference layer)에서 예측을 위한 참조 프레임(시간적으로 일치하는 VOP)을 선택하고 그 프레임의 크기를 확장계층에 맞도록 재샘플링(resampling)하는 두 부분의 작업으로 나누어지며, 이 동작은 중간처리기에서 수행된다. 재샘플링 처리는 모든 VOP에 대해서 처리되지만 한 매크로블록의 복호화를 위해서는 재샘플링된 프레임 중 그 위치에 대응하는 매크로블록 부분만이 필요하다. 이 과정을 살펴보면, 우선 기본계층 VOP는 수직방향으로 재샘플링되고 수직방향으로 재샘플링된 영상은 수평방향으로 재샘플링된다.

확장계층에서 VOP는 P-VOP 혹은 B-VOP로 부호화된다. 기본계층과 확장계층 VOP 사이의 관계는 그림 2와 같다. 기본계층 내의 I-VOP와 시간적으로 일치하는 VOP는 P-VOP로 부호화되고, 기본계층 내의 P-VOP와 시간적으로 일치하는 VOP는 B-VOP로 부호화된다. 공간적 계위부호화의 경우, 기본계층에서 복호화된 VOP가 예측동작의 참조로 사용된다. 참조계층(기본계층)내의 시간적으로 일치하는 VOP는 확장계층내의 VOP가 부호화되기 전에 부호화되어야 한다.

1.2 공간적 계위복호화 과정

그림 3에 보인 것처럼, 기본계층과 확장계층 비트열이 각각 기본계층 복호기와 확장계층 복호기의 입력으로 들어간다. 중간처리기는 입력으로 들어온 다운샘플링된 기본계층 비트열에 대하여 공간적 다운샘플링 혹은 업샘플링을 수행한다. 후처리기는 우선, 기본계층 비트열에 대해서는 공간적으로 다운샘플링 혹은 업샘플링을 수행하여 outp_0을 출력하고, 확장계층 비트열에 대해서는 해상도에 전혀 변화를 주지 않고 출력시킨다.

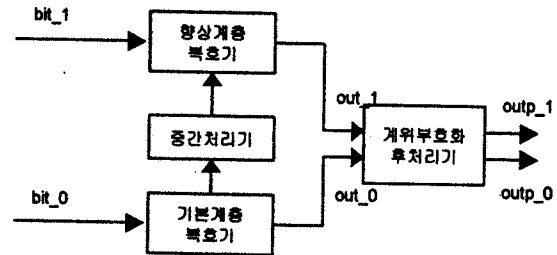


그림 3. 계위부호화를 위한 High Level 복호화 구조

공간적 계위부호화에서 확장계층 비트열은 영상의 해상도가 증가되어 있는 상태이다. 출력단에서 저해상도가 요구되면 기본계층 비트열만을 복호하고, 고해상도가 요구되면 기본계층과 확장계층 비트열을 모두 복호한다. 기본계층 비트열은 그 자체로 복호가 가능하지만, 확장계층 비트열은 반드시 기본계층 비트열로부터 예측하여 복호해야 한다.

움직임 보상 예측은 기본계층 프레임을 사용하여 예측한다. 여기서 P-VOP의 전방향 예측과 B-VOP의 역방향 예측은 부호화된 것처럼, 기본계층 프레임을 참조 프레임으로 보고 공간적 예측을 한다. B-VOP의 전방향 예측은 확장계층 VOP를 참조 프레임으로 보고 시간적 예측을 하고 B-VOP의 보간(interpolate) 예측은 이 두 공간적/시간적 예측의 조합이 된다.

2. 시간적 계위부호화

시간적 계위부호화는 공간적 계위부호화와 마찬가지로, 기본계층과 확장계층을 포함한다. 확장계층은 기본

계층의 시간적 해상도(temporal resolution)를 확장한다. 확장계층의 복호된 VOP들은 기본계층의 프레임율(framerate)을 증가하기 위해서 사용된다.

그림 4는 시간적 계위부호화를 사용한 확장계층을 위한 움직임 보상의 처리 과정이다. 예측된 표본들, $p[y][x]$ 는 확장계층과 참조계층의 프레임 저장소로부터 구성된다. 차분 데이터 표본들, $f[y][x]$ 는 복호화된 표본들 $d[y][x]$ 을 구성하기 위해 $p[y][x]$ 에 더해진다.

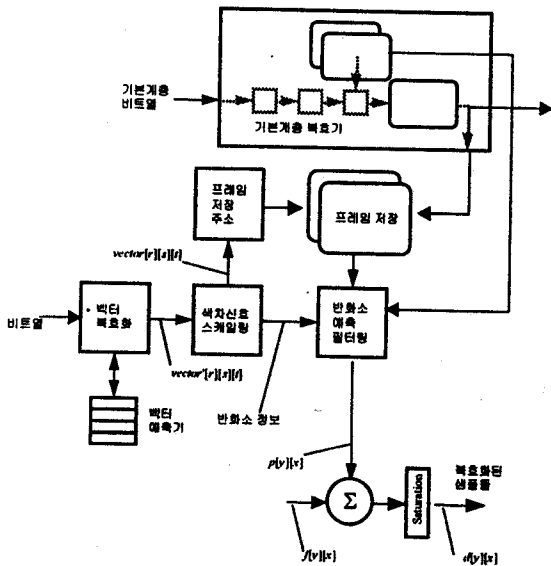


그림 4. 확장계층을 위한 움직임 보상 처리과정

2.1 배경 구성

배경 구성은 확장계층에서 객체들에 대한 배경 영역을 구성하기 위해 사용되며, 기본계층의 이전 프레임과 다음 프레임을 사용하여 수행된다. 그림 5에서 점선은 기본계층에서 이전 프레임의 선택된 객체의 모양을 나타내는데, 그 모양은 물체의 움직임에 따라 기본계층의 다음 프레임에서 깨진 선으로 나타낸다.

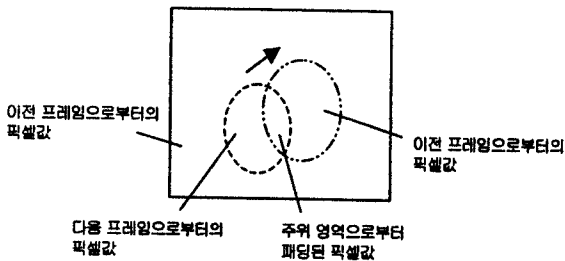


그림 5. 배경 구성

2.2 시간적 계위부호화 과정

배경 구성에 대해 bgc_mode 지역 변수의 값을 "1"로 설정하고 P_VOP 부호화를 위해 $prev_vol$ 을 $next_vol$

로 대체한다. 기본계층에 대한 작업 중, 시간적 계위부호화가 B_VOP를 사용한다면, $prev_vol$ 에 대한 VOP 정보를 동일하게 가진 복사본을 $original_prev_base_vol$ 에 저장한다. 기본계층이 아닌 시간적 계위부호화의 경우에는 UpdateVolStore() 또는 NotUpdateVolStore()를 사용해 P_VOP 부호화를 위한 VOL 대체 작업을 수행한다. 그리고 배경 구성을 위해 사용되는 기본계층의 영상을 바꾼다.

시간적 계위부호화를 위한 확장계층의 부호화는 확장계층이 L_VOP 또는 P_VOP를 사용하는 경우와 B_VOP를 사용하는 경우에 대해서 다르게 수행한다. 이 모든 작업이 끝난 후에, BGCPProcess()를 사용하여 복호기에서 배경 구성을 수행하는데 요구되는 기본 작업을 처리한다.

MoMuSys 코덱과 관련해서 시간적 계위부호화와 관련된 기본 내용은 다음과 같다.

시간적 계위부호화의 방법은 기본계층의 확장 영역에 따라 0부터 2까지로 3가지 종류가 존재하고 기본계층과 확장계층 간의 관계에 따라 다시 경우0 (case 0)부터 경우2 (case 2)의 3가지 경우로 나뉜다.

2.3 시간적 계위부호화 과정

확장계층의 VOP는 I, P, 그리고 B-VOP로 복호된다. VOP의 모양은 "video_object_layer_shape"에 의해서 결정되며, 그 값이 00이면 사각형, 01이면 임의의 모양을 의미한다. 기본계층의 B-VOP는 확장계층에 참조되지 않는다.

I-VOP의 경우는 비계위적(non-scalable)의 경우와 같다. P, B-VOP의 경우는 일반적인 경우는 비계위적 복호화 과정과 같으나, 임의의 모양인 경우에는 달라진다. 그리고 ref_select_code의 값에 따라서 예측 참조(prediction reference)가 결정된다.

III. 계위부호화기 구현 및 성능 평가

표 1은 목표 비트율을 256 kbps로 설정했을 때, 프레임율에 따른 공간적 계위부호화 지원 알고리즘의 성능을 분석하였다.

표 1. 프레임율에 따른 부호화 시간

결과 초당 프레임율	계층	부호화 시간	초당 부호화 프레임 수
30	기본/확장	10분 40초	0.23
15	기본/확장	2분 48초	0.45
10	기본/확장	1분 50초	0.45

표 2는 프레임율을 15.0 fps로 고정하고, 목표 비

트율을 24kbps에서 256 kbps까지 변경하는 경우에 실제 발생된 비트량과 평균 PSNR 값을 나타낸 것이다.

표 2. 목표 비트율에 따른 부호화 성능 비교

목표 비트율 \ 결과	계층	총 비트량	VOP당 평균 비트량	평균 PSNR
256 kbps	기본	419648	5593	41.38
	확장	666728	8887	34.78
64 kbps	기본	319848	4262	40.10
	확장	193976	2583	32.26
24 kbps	기본	120192	1600	35.74
	확장	160832	2141	31.27

Q2 비트율 제어 알고리즘을 사용하는 경우, 구현된 시간적 계위부호화 지원 알고리즘의 성능을 분석하였다. 이때, 기본계층과 확장계층의 초당 비트율과 프레임율(15프레임)을 같게 하여 부호화를 수행하였다. 표 3은 비트율 제어 사용 여부에 따른 각 초당 비트율에 대한 PSNR 값을 나타낸다. 이 값은 잡음에 대해 신호가 얼마나 강인한가를 보여준다.

표 3. 비트율 제어 사용 여부에 대한 PSNR 성능비교

비트율제어 사용 모델	초당 비트율	PSNR (기본계층)	PSNR (확장계층)
모델 사용 안함	Default	36.83	36.82
Q2 모델	24000	34.42	34.38
Q2 모델	64000	37.30	36.95
Q2 모델	256000	42.66	42.14

표 4는 Q2 비트율 제어 모델을 사용할 경우에 대해, 부호화된 총 비트량을 보인다. 총 비트량은 대략 (적용된 초당 비트율) * (디스플레이 시간)의 값을 가지면 어느 정도의 성능을 보장할 수 있다. 이 때, 디스플레이 시간은 초당 30 프레임을 기준으로 하여, 총 150프레임을 부호화 하므로 5초가 된다.

표 4. 비트율 제어 모델을 사용하는 경우, 총 부호화된 비트량

초당 비트율 (bits/sec)	발생 비트량 (기본 계층)	발생 비트량 (확장 계층)
24000	144968	68224
54000	321088	84448
256000	1286320	560152

실제 MPEG-4에서 주로 고려되는 저-비트율 환경

이라 할 수 있는 64 kbps인 경우에는 지나친 프레임 건너뛰기를 피하면서도 최대한의 프레임을 전송할 수 있음을 보인다. 즉, 사용자에게 보다 편안하고 완전한 영상 데이터를 전송할 수 있음을 의미한다.

비트율 제어 모델을 사용한 기본적인 시간적 계위 부호화의 성능 평가를 수행한 본 실험에서, 초당 비트율에 따른 PSNR을 통해 보이는 성능 및 총 비트 수에서 비트율 제어를 사용하지 않은 경우에 비해 어느 정도의 성능적 효과를 볼 수 있었다.

IV. 결론

본 논문에서는 MPEG-4에서 제공하는 프레임 기반 부호화 과정에서 가장 기본적으로 제공되는 공간적 계위 부호화와 시간적 계위부호화 기법에 대해서 분석하였고, MPEG-4 비디오 부호화 표준에 맞게 시/공간 계위 부호화 소프트웨어를 개발하였다. 개발된 공간적 계위 부호화 기술과 시간적 계위부호화 기술은 10 kbps정도 이하의 낮은 비트율에서 수 Mbps의 높은 비트율에 이르기까지 광범위한 범위의 비트율에서 적용 가능하기 때문에 정보가전 뿐만 아니라 무선 인터넷상으로 사용자의 단말기 성능과 전송 효율을 고려하여 효율적인 데이터 전송이 가능하도록 한다.

감사의 글

본 연구는 광주과학기술원(K-JIST) 초고속광네트워크 연구센터 (UFON)를 통한 한국과학재단 우수연구센터 (ERC)와 교육부 두뇌한국21(BK21) 정보기술사업단의 지원에 의한 것입니다.

참고 문헌

- [1] ISO/IEC 14496-5/FDAM1 (MPEG-4 Video Reference Software), ISO/IEC JTC1/SC29/WG11, MPEG01/N3769, Jan. 2001.
- [2] MPEG-4 Video Verification Model version 18.0, ISO/IEC JTC1/SC29/WG11, MPEG01/N3908, Jan. 2001.
- [3] Efficiency of MPEG-4 Spatial Scalability at higher bitrates, ISO/IEC JTC1/SC29/WG11/3194, 1998.
- [4] Report Of The Formal Verification Tests On MPEG-4 Temporal Scalability in Simple scalable Profile, ISO/IEC JTC1/SC29/WG11, MPEG98/N2605, Dec. 1998.
- [5] 미키 스케이치, MPEG-4의 세계, 영풍문고, 1998.