

특집논문-07-12-6-01

스케일러블 비디오 부호화의 개요 및 성능 분석

최해철^{a)†}, 이경일^{a)}, 강정원^{a)}, 배성준^{a)}, 유정주^{a)}

Overview and Performance Analysis of the Emerging Scalable Video Coding

Haechul Choi^{a)†}, Kyung Il Lee^{a)}, Jung Won Kang^{a)}, Seong-Jun Bae^{a)}, and Jeong-Ju Yoo^{a)}

요 약

근래의 무선망 및 인터넷의 초고속화에 따라 풍부한 멀티미디어에 대한 사용자의 요구가 증대 되고 있으며, 이동 통신의 활성화 및 DMB(Digital Media Broadcasting)와 같은 통방융합 서비스의 등장으로 여러 네트워크 망을 통해 다양한 성능을 가진 단말로의 멀티미디어 서비스가 이루어 지고 있다. 이러한 환경 변화에 따라 비트율을 줄이기 위한 비디오 압축 부호화 기술뿐만 아니라 시변의 전송 조건, 전송망 특성, 다양한 단말 성능, 사용자 선호도 등에 대한 비디오의 적응 기술에 관심이 모아졌으며, 그 결과 ISO/IEC MPEG(Moving Picture Experts Group) 및 ITU-T VCEG(Video Coding Experts Group)의 JVT(Joint Video Team)에서는 비디오 부호화의 스케일러빌리티(scalability)를 제공하기 위한 SVC(Scalable Video Coding) 표준화를 진행하였다. 본 논문에서는 공간/시간/화질적 스케일러빌리티를 지원하기 위한 SVC의 대표적 표준 기술들에 대해 설명하고, 기존의 단일 계층(single layer) 부호화 방식과 비교하여 SVC의 부가 비트 요구량 및 부호화 효율에 대해 성능 평가를 하였다.

Abstract

Seamless streaming of multimedia content via heterogeneous networks to viewers using a variety of devices has been a desire for many multimedia services, for which the multimedia contents should be adapted to usage environments such as network characteristics, terminal capabilities, and user preferences. Scalability in video coding is one of attractive features to meet dynamically changing requirements of heterogeneous networks. Currently a new scalable video coding (SVC) is standardizing in the Joint Video Team (JVT) of the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG), which will be released as Extension 3 of H.264/MPEG-4 AVC. In this paper, we introduce new technologies of SVC and evaluate performance of it especially regarding on overhead bit-rate and coding efficiency to support spatial, temporal, and quality scalability.

Keyword : Scalable Video Coding, Coding Efficiency, Performance Evaluation, H.264/MPEG-4 AVC

1. 서 론

근래의 무선망 및 인터넷의 초고속화에 따른 정지영상

및 동영상 등의 정보를 포함한 다양한 멀티미디어 서비스가 새로운 패러다임으로 활성화 되고 있으며, BcN(Broadband convergence Network) 및 통방융합 인프라의 등장으로 압축 부호화 기술만을 개발하던 시기와 달리 최근 디지털 멀티미디어의 생성, 전송 및 소비 환경은 광범위하게 급변하고 있다. 즉, 서로 다른 전송 용량 및 에러 발생율을 갖는 전송망이 혼재해 있으며, DMB(Digital Multimedia Broadcasting) 서비스와 같은 통신과 방송의 융합 서비스

a) 한국전자통신연구원 전파방송연구단 방송미디어연구그룹
ETRI, Radio&Broadcasting Research Division, Broadcasting Media Research Group

† 교신저자 : 최해철(choihc@etri.re.kr)

* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력핵심기술 개발 사업의 일환으로 수행하였음. [2005-S-103-02, 통방융합 환경에서의 유비쿼터스 콘텐츠 서비스(UCA) 기술 개발]

도 실현 되었다. 더욱이 통신 및 방송에서 이동성을 제공함에 따라 비디오를 소비하는 단말 또한 다양한 화면 크기와 연산 능력을 가진 여러 종류의 제품들이 출시되었다. 현재 멀티미디어에서의 비디오 서비스에서 널리 쓰이고 있는 비디오 부호화 기술은 고정된 영상포맷(format)과 비트율(bit-rate)로 압축 및 전송하는 단일 계층 압축 부호화(single layer video coding) 방식이다. 이는 근래의 변화된 환경에서 시변의 전송 조건, 망의 특성, 사용자 선호도, 그리고 단말의 다양한 성능 등의 조건들에 모두 적합하도록 미리 부호화 하기 어렵기 때문에 끊임 없는 품질의 멀티미디어에 대한 사용자의 욕구를 충족 시킬 수 없다. 따라서 각 네트워크의 조건과 비디오를 시청하는 단말기의 성능에 비디오의 포맷 및 비트율을 적응 시키기 위한 변환부호화(transcoding)^[1]나 비트량변환(transrating) 방식 등이 연구되어 왔다. 하지만 이 방식들은 제한된 범위에서의 영상 형식 및 비트율의 변경 폭이 좁고 높은 복잡도를 유발함으로써 인해 실시간 적응이 요구되는 서비스 환경에서 널리 쓰이지 못하고 있다.

최근 스케일러빌리티를 제공하기 위한 비디오 부호화 방식에 대한 연구가 활발히 진행되고 있다. 이 스케일러빌리티는 한 비트스트림(bit-stream)에서 임의의 부분을 제거함으로써 공간적, 시간적, 혹은 화질적 변경을 지원할 수 있기 때문에 멀티미디어의 전송 및 소비 환경 적응을 위한 매력적인 해결책 중의 하나이다. 비디오 부호화에 있어서 스케일러빌리티 기능을 위한 기존 부호화표준은 MPEG-2^[2], H.263^[3] 그리고 MPEG-4 part 2 Visual^{[4][5]}까지 계층 부호화에 기반한 여러 차례의 시도가 있었다. 하지만 이들은 하나의 계층으로 부호화하는 기존 압축 부호화 기법에 비해 부호화 효율이 낮고, 여러 종류의 스케일러빌리티를 종합적으로 지원하지 못하는 취약점이 있었다. 전송 및 소비 환경 측면에서, 과거에는 전송망이 고정된 처리율(throughput)을 가지며 패킷의 전송 성공 혹은 실패와 같은 이분법의 획일된 정책으로 일관 되었기 때문에, 비트율을 세밀히 변화 시키면서 전송망의 상황에 맞춘 서비스가 요구되지 않았으며, 사용자가 고정된 위치에서 TV 혹은 PC(personal computer)를 이용하여 비디오를 시청하였기에 공간적인 스케일러빌리티를 제공하는 이 표준들이 산업적으로 성공을 거두지 못하였다.

근래의 이동 통신의 비약적 발전은 다양한 전송망을 출현 시켰으며, 소비자의 이동성 선호에 따라 다양한 성능의 단말이 이용되고 있다. 따라서 비디오 부호화에서의 스케일러빌리티의 현실적 필요성이 점점 절실하게 되었으며, 과거 취약점을 해결하면서 실시간 비디오 전송에서의 다양한 스케일러빌리티를 지원하기 위해 현재 ISO/IEC Moving Picture Experts Group (MPEG)과 ITU-T Video Coding Experts Group (VCEG)의 Joint Video Team (JVT)은 Scalable Video Coding (SVC)^{[6]-[8]} 표준화를 H.264/MPEG-4 AVC [9]의 수정판(Amendment)으로 진행하고 있다. 여기에는 유럽, 미국, 한국, 일본, 중국 등이 적극적으로 참여 중이다.

본 논문에서는 영상 크기, 프레임율, 비트율에 대한 스케일러빌리티를 지원하는 SVC의 대표적 표준 기술들에 대해 설명하고, 기존의 단일 계층(single layer) 부호화 방식과 비교하여 SVC의 성능을 분석하였다. 특히 스케일러빌리티를 제공하기 위한 부가적인 비트 요구량 및 부호화 효율성에 대해 초점을 맞추었다.

성능 평가에서는 JSVM(Joint Scalable Video Model) 8.12^[10]를 이용하여 부호기 및 복호기로 나누어 분석한다. 이는 어떤 알고리즘들은 부호기 및 복호기에서 요구 계산량의 비중이 다르기 때문이다. SVC의 기본 계층은 H.264와 호환성을 가지므로 SVC는 H.264의 대부분의 부호화 알고리즘을 포함한다. 다만 현재 산업적으로 많이 쓰이지 않는 switching picture, temporal direct mode 등이 제외되었다. H.264의 성능 분석은 이미 수 차례 보고 되었기에, 본 논문에서는 SVC의 새로운 알고리즘에 대해 집중하였다. SVC 표준화 회의에 사용되는 테스트 영상들을 입력하여 각 알고리즘의 설정에 따른 부호화 혹은 복호화 소요 시간 차이를 측정함으로써 그 복잡도를 분석하였으며 그 때의 PSNR(Peak signal to noise ratio)를 이용하여 부호화 효율을 비교하였다.

본 논문의 2장에서는 스케일러빌리티의 개념의 응용에 대해 설명하고, 3장에서는 새로운 표준인 SVC의 스케일러빌리티 지원을 위한 대표적 기술들에 대해 소개한다. 그리고 4장은 SVC의 복잡도 분석 및 성능 평가를 기술하고, 마지막으로 5장에서 결론을 맺는다.

II. 스케일러빌리티

스케일러빌리티란 공간, 시간, 화질 측면에서 비트스트림의 한 부분을 제거하여 복호 가능한 비트스트림을 재생성할 수 있는 기능으로 정의할 수 있다. 즉 그림 1과 같이, 최상의 화질, 최고의 프레임율, 최대의 영상크기를 갖는 하나의 부호화된 글로벌(global) 비트스트림으로부터 시간, 공간, 그리고 화질 세 개의 축 상에서 임의의 지점의 복호 가능한 비트스트림(그림 1에서의 박스)을 적출(extraction)할 수 있다. 따라서 그 세 가지 차원의 조합에 따라 폭넓은 스케일러빌리티를 제공할 수 있다.

그림 2은 스케일러블 비디오 부호화를 이용한 멀티미디어

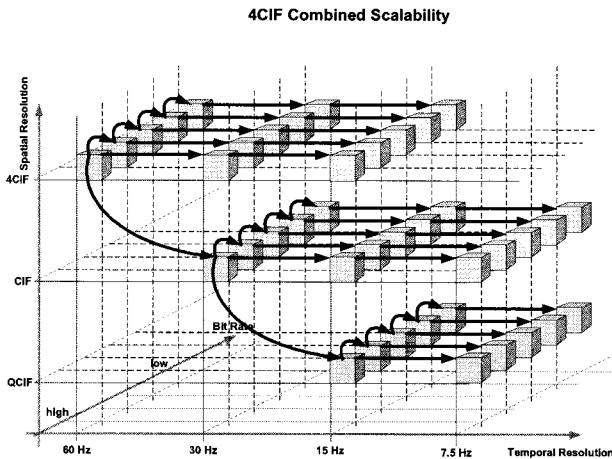


그림 1 스케일러블 비트스트림의 유효한 적출(extraction) 비트스트림의 예
Fig. 1. Extraction points supported by SVC bit-stream

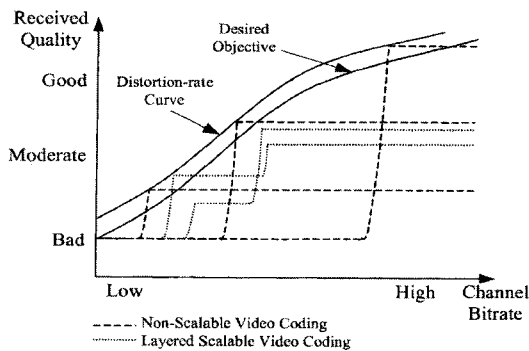


그림 2 채널 대역폭 변화에 따른 비트율 적응
Fig. 2. Bit-rate adaptation to channel bandwidth variation

어 서비스의 장점을 보여주고 있다. 스케일러빌리티를 지원하지 않는 부호화 방식은 채널의 대역폭이 증가하여도 이미 부호화된 비트스트림이 갖는 화질로만 서비스 될 수 밖에 없으나, 스케일러블 부호화 방식으로 압축된 비트스트림은 채널의 대역폭 변화에 맞춰 가변적인 화질로 전송 및 소비 될 수 있다.

기술적인 관점에서, 하나의 스케일러블 비트스트림은 두 개 혹은 그 이상의 의존적인 계층으로 구성될 수 있다. 이 경우, 스케일러블 코덱(codec)은 하나의 기본 계층(base layer)과 스케일러블 상위 계층(enhancement layer)들로 구성된다. 최상위 계층을 복호할 때, 자신 및 자신 보다 낮은 계층의 정보들이 함께 이용된다. 화질 스케일러빌리티를 예로 든다면, 일반적으로 기본 계층은 기본적인 비디오 화질을 제공하고, 연속된 상위 계층은 이전 계층들로 만들어진 비디오보다 높은 화질을 갖도록 부호화한다. 마찬가지로 시간 및 공간 해상도에서도 동일한 원리를 적용하여 스케일러빌리티를 지원한다. 이는 하나의 차원에 대한 스케일러빌리티로는 최적의 가변성과 적응성을 제공할 수 없는 응용분야가 있을 수 있으며, 시간, 공간, 화질 측면 모두의 스케일러빌리티를 필요로 하기 때문이다.

III. 스케일러블 비디오 부호화

1. 개요

MPEG-4 AVC/H.264 Amd. 3 SVC의 스케일러빌리티 종류에는 공간적 스케일러빌리티(spatial scalability), 시간적 스케일러빌리티(temporal scalability), 화질적(SNR) 스케일러빌리티, 그리고 복합 스케일러빌리티(combined scalability)가 있다. 여기서 시간적 스케일러빌리티는 Hierarchical B picture^{[1][12]} 구조로써 지원하고, 공간적인 스케일러빌리티는 과거 스케일러빌리티를 제공하는 표준^{[2]-[5]}들에서 활용되었던 상향표본화(up-sampling)를 이용한 다계층(multi-layers) 방식에 기반한다. 그리고 화질 스케일러빌리티를 위해서는 CGS(Coarse Grain Scalability)의 경우 공간 스케일러빌리티를 위한 기법과 동일한 방식에 양자화 계수(quantization pa-

parameter)만을 변경하며, MGS(Medium Grain Scalability)의 경우 양자화 에러에 대한 점진적 부호화 기법을 이용한다^{[7][13]}.

JSVN 부호기 구조는 제공하려는 스케일러빌리티의 종류 및 단계에 따라 달라진다. 예를 들어 3 개의 공간적 계층으로 구성하는 경우 SVC 부호기 구조는 그림 3와 같다. 공간적 스케일러빌리티를 위해 입력 영상의 공간 해상도를 줄여 작은 크기의 영상 시퀀스를 공간적 하위 계층에 입력한 후, 각 공간적 계층에서 Hierarchical B picture를 적용하여 시간적 스케일러빌리티를 지원하며, 움직임 정보와 잔여 텍스처 정보는 H.264 표준에 기반하여 부호화한다. 부

가적으로화질 스케일러빌리티를 제공하기 위해 각 공간적 계층에서 MGS 기법이 적용 될 수 있다. 각 공간적 계층에서의 부호화 시, 공간적 하위 계층에서 이미 부호화된 화면 내 부호화 정보, 움직임 벡터, 잔여 신호 등은 현재 공간적 계층의 예측 신호로 활용될 수 있다. 공간적 첫 번째 계층은 반드시 H.264와 완벽한 호환성을 가져야 하는 제약이 있다.

2. 시간 스케일러빌리티 관련 기술

H.264에서 이미 부호화된 모든 영상은 참조 영상으로써

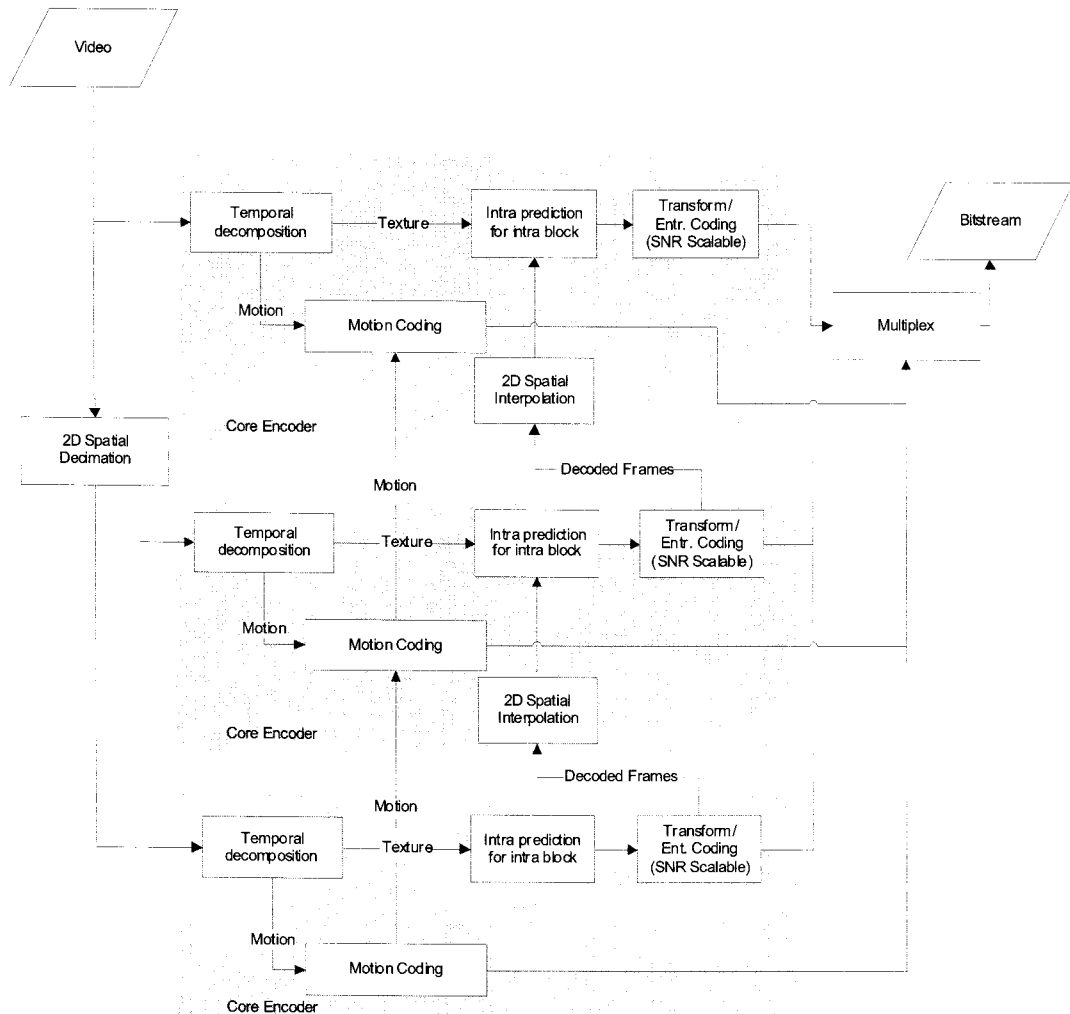


그림 3. 공간적 3 계층의 스케일러빌리티를 제공하는 스케일러블 부호기 예^[1]
 Fig. 3. An example of SVC encoder supporting 3 spatial layers^[7]

다음에 오는 영상의 움직임 보상 예측에 이용될 수 있다. Memory management control operation (MMCO) commands는 복호화된 영상을 저장할 수 있는 decoded picture buffer (DPB)를 적절하게 제어하며, DPB에 저장된 영상들은 reference picture list re-ordering (RPLP) commands를 사용하여 임의적으로 참조 영상으로써 선택이 가능하다. 이러한 H.264의 특징들은 이전 비디오 부호화 방식들이 지원하지 못했던 임의 순서의 부호화 및 참조 영상의 선택을 가능하게 한다. 이를 기반으로 B 영상의 참조를 허용하는 General B picture으로써 시간적 스케일러빌리티를 지원하는 hierarchical B picture가 제안되었다.

그림 4는 4 단계를 갖는 전형적인 hierarchical B picture 구조를 보여준다. 비디오의 첫 영상은 IDR (Instantaneous Decoder Refresh) 영상으로서 화면내 부호화된다. 이전에 부호화된 모든 영상들이 현재 부호화되는 영상보다 화면 표시 순서(display order) 상 앞서 위치할 때, 이것을 key picture(그림 4의 8번째, 16번째 프레임)라고 부른다. 시간적으로 현재 key picture 1장과 이전 key picture까지의 모든 non-key picture들을 하나의 group of pictures (GOP)라고 한다. Key picture는 임의 접근을 위해 화면내 부호화되거나, 이전 key-picture를 참조 영상으로 이용하는 P 영상으로 화면간(inter) 부호화 된다. GOP의 영상들은 그림 4와 같이 계층적으로 참조되어 다양한 프레임율을 가질 수 있다. 예를 들어 B3 영상들을 모두 보내지 않는 경우 최대 프레임율의 반으로 복호화 가능하다.

3. 공간적 스케일러빌리티 관련 기술

공간적 스케일러빌리티는 그림 3에서와 같이 작은 영상 크기의 비디오 시퀀스부터 큰 영상 크기의 시퀀스들을 각 공간적 계층에 입력하여 부호화 하는 다계층 부호화 기법에 기반한다. 이 때 각 공간적 계층은 독립적으로 부호화할 수도 있지만, 하위 계층의 부호화 결과물인 복원된 화소값, 움직임 벡터, 잔여 신호 정보, 및 구문정보(syntax elements) 등을 상위 공간적 계층에서 이용함으로써 부호화 효율을 높일 수 있다. 공간적 계층의 계층간(inter layer) 상관관계를 이용해 부호화 효율을 높이기 위한 3 가지 계층간 예측 기술(inter-layer prediction)은 다음과 같다.

■ 계층간 화면내 텍스처 참조 (Inter-layer intra texture prediction)

계층 기반 방식에서 부호화 효율을 높일 수 있는 가장 기본적인 방식은 하위 공간적 계층의 영상을 복원한 후 이를 현재 공간적 계층의 영상 크기에 맞게 보간하고 이를 참조 영상으로 이용하는 것이다. MPEG-2, MPEG-4 Visual에서는 프레임 단위로 하위 공간적 계층의 복원된 영상을 참조 영상으로 이용하는 방식을 채택하였다. 영상의 복잡도가 낮은 경우 하위 공간적 계층의 영상은 현재 공간적 계층의 영상과 높은 공간적 중복성을 갖기 때문에 이러한 방식은 부호화 효율을 극대화 시킬 수 있다. 하지만 하위 공간적 계층이 P 영상 혹은 B 영상로 부호화 된 경우, 이를

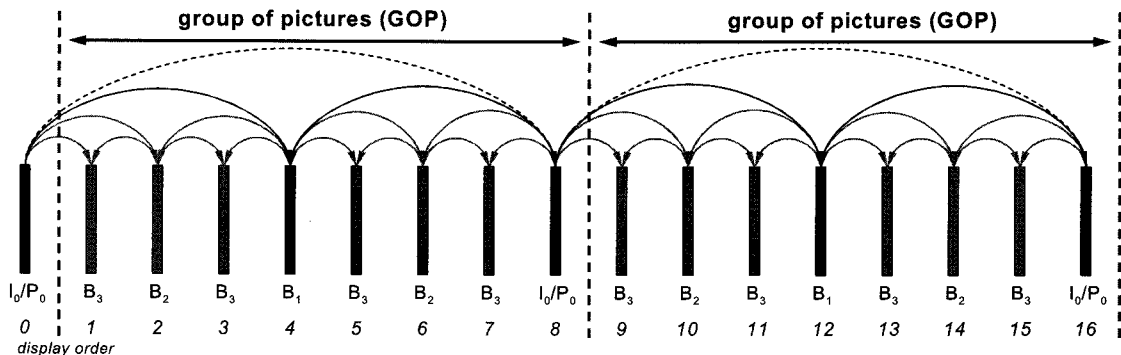


그림 4. GOP 크기가 8이며, 4개의 시간적 레벨을 갖는 hierarchical B picture 구조
 Fig. 4. Hierarchical B picture structure where GOP size is 8, which can support 4 temporal levels

복호화 하기 위해선 이 하위 공간적 계층 영상의 참조 영상들도 복호화 해야 한다. 심지어 그 참조 영상의 참조 영상도 복호화해야 하는 복잡한 문제가 발생할 수 있다. 따라서 위 방식은 기본적으로 현재 목적(target) 공간적 계층뿐만 아니라, 하위 공간적 계층에 있는 대부분의 영상을 화소 단위까지 복원해야 한다. 이는 복호기의 과도한 연산을 요구 하며, 이전 스케일러블 비디오 표준의 활용에 치명적인 약점이었다. 반면 SVC에서는 복호기의 복잡도 감소를 위해 프레임 단위가 아닌 블록 단위로 복원된 화소를 참조하며, 이때 오직 화면내 부호화된 블록에 대해서만 inter-layer intra texture prediction을 허용한다. SVC에서는 이를 Intra_BL 모드라 하며 그림 5는 Intra_BL 모드를 도시하고 있다.

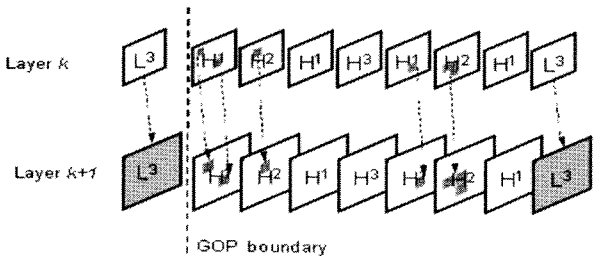


그림 5. 공간적 계층간 Intra_BL 모드
Fig. 5. Intra_BL mode for inter-layer texture prediction

공간 계층간 영상의 크기가 다른 경우 Intra_BL 모드에

서 하위 공간적 계층으로부터 예측 신호를 얻기 위해서는 하위 계층의 동일 대응 지점에 위치한 블록(co-located block)의 복원된 화소 값에 대해 상향 표본화(up-sampling)을 수행한 후 디블록킹(deblocking) 필터를 적용시킨다. 상향 표본화를 위해서 밝기 값은 16개의 위상차(phase shift)를 갖는 4-tap cubic spline 필터 중 계층간 영상 크기 비율에 따라 선택 적용하며, 컬러 값은 복잡도 감소를 위해 단순히 bi-linear 필터를 활용된다. 디블록킹 필터는 과도한 열화를 막기 위해 bS(boundary filtering strength) 값을 1로 고정시킨다. 결국 상향 표본화된 신호를 현재 블록의 참조 신호로 활용하여 현재 원 영상 블록과의 차이 값만을 부호화한다. 이 Intra_BL 모드는 SVC에서 부호화 효율을 높이는 주된 기술 중 하나이다.

■ 계층간 움직임 정보 참조 (Inter-layer motion prediction)

하위 공간 계층의 대응 위치 블록이 화면간 부호화 되었다면 상위 계층의 블록은 하위 계층으로부터 움직임 정보 및 모드 정보를 참조할 수 있다. 따라서 상위 계층에서는 움직임 정보 및 매크로블록의 분할(partitioning) 정보, 참조 영상 정보 등을 매크로블록 헤더에 부호화하지 않을 수 있다. 영상 시퀀스를 낮은 비트율로 부호화할 때 이러한 구문 요소 정보들의 비중이 상당히 높기 때문에, 구문요소에 할당된 비트량을 대부분 줄일 수 있는 이 기법은 높은 성능을 가질 수 있다.

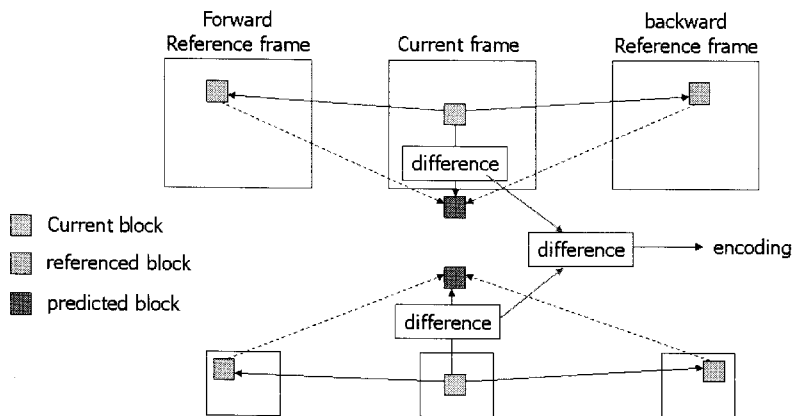


그림 6. SVC의 계층간 잔여신호 참조 과정
Fig. 6. Inter-layer residual prediction of SVC

■ 계층간 잔여 신호 참조 (Inter-layer residual prediction)

공간적 하위 계층의 양자화 계수 값이 크면, 잔여 정보를 예측하는 것은 부호화 효율에 도움이 될 수도 있고 그렇지 않을 수도 있으나, 현재 계층 블록의 양자화 계수가 34 이하인 경우 공간적 하위 계층의 잔여신호가 현재 공간적 계층의 잔여 신호를 위한 예측으로 사용될 때 부호화 효율이 증대될 수 있다^[4]. SVC는 그림 6과 같이 residual_prediction_flag를 이용하여 블록 단위로 하위 공간적 계층의 복원된 잔여 계수를 현재 공간적 계층의 잔여 계수를 부호화 할 때 예측 신호로써 적응적으로 활용한다.

■ 공간적 계층간 참조를 위한 구문 요소들

그림 7은 블록 단위의 공간적 계층간 참조를 위한 구문 요소들을 도시하고 있다. 기본 공간적 계층(base layer)을 제외한 상위 공간적 계층에서 현재 복호하려는 슬라이스(slice)의 헤더에 adaptive_prediction_flag 값이 1이면, 그 슬라이스가 블록 단위의 선별적 계층간 참조를 사용했다는 것을 의미한다. 매크로블럭 헤더의 첫 번째 구문요소는 base_mode_flag 이다. 이 값이 1이면 하위 계층의 대응 위치 블록이 화면내 혹은 화면간 부호화 되었느냐에 따라 Intra_BL 모드 혹은 BL_skip 모드로 부호화 되었음을 의미한다. Intra_BL 모드는 앞서 설명한 inter-layer intra texture prediction 이며, BL_

skip 모드는 대응 하위 계층의 매크로블록 분할 정보, 참조 영상 지시자, 움직임 벡터, 잔여 신호 모두를 공간적 계층간 영상 크기 비율에 맞게 조정 후 현재 복호하려는 블록의 구문 값으로 이용하며, 이와 관련된 더 이상의 부가 정보는 보내지 않는다. 따라서 매크로블록 헤더에 들어가는 많은 구문정보의 비트량을 획기적으로 줄일 수 있다. base_mode_flag가 0 이면 mb_type이 비트스트림에 적혀 있다. mb_type이 화면내 부호화를 의미하는 값이면, 기존 H.264 방식과 동일하게 부호화 하며, 그렇지 않은 경우 motion_prediction_flag 와 residual_prediction_flag의 값에 따라 하위 계층의 움직임 정보 및 잔여 신호 정보를 현재 복호하려는 블록의 참조 신호로 이용한다. 여기서 주목할 점은, BL_skip 모드에서는 ‘하위 공간적 계층의 움직임 벡터 x 영상 크기 비율’ 가 현재 공간적 계층의 최종 움직임 정보인 반면, motion_prediction_flag 에 의한 움직임 정보 복호에서는 하위 계층의 움직임 정보가 참조 신호로 활용 되며 추가적으로 움직임 정보 차이값(MVD; motion vector difference)이 비트스트림에 포함된다. 잔여 신호의 경우 위와 같은 방식으로 복호된다.

4. 화질 스케일러빌리티 관련 기술

SVC는 화질적 스케일러빌리티를 위해서 그림 8과 같이

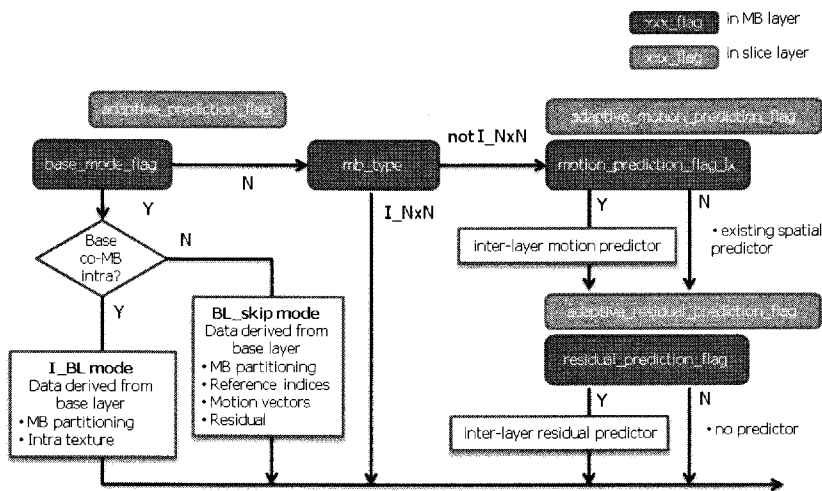


그림 7. 공간적 계층간 참조를 위한 구문 요소들의 흐름도
Fig. 7. Syntax flowchart for inter-layer prediction

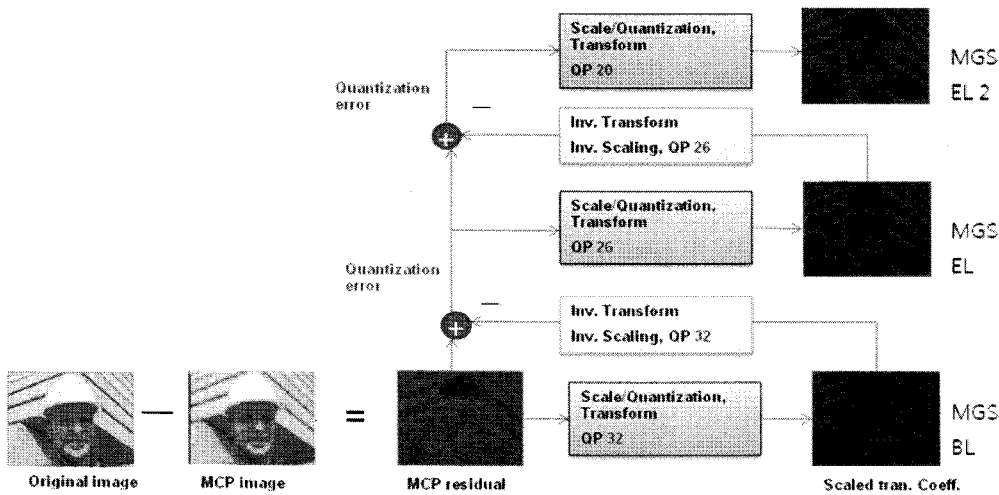


그림 8. 화질 스케일러빌리티를 위한 부호화기 구조 예
 Fig. 8. An example of SVC encoder for quality scalability

양자화에 의한 오차 신호를 하위 화질적 계층 보다 작은 양자화 계수 값으로 보정하여 화질적 계층별로 부호화 하는 방식이다. 즉 첫 번째 화질적 계층은 큰 간격으로써 등성등성하게 양자화한 후, 상위 화질적 계층으로 갈수록 양자화 간격을 세밀하게 줄여 보다 좋은 화질적 계층을 쌓는다. 기술적으로 살펴보면, 앞서 설명한 hierarchical B picture 에 의한 시간적 분할(decomposition) 이후의 신호들을 임의 양자화 계수값으로써 최소 화질을 갖는 기본 계층을 부호화 한 후, 이 기본 화질적 계층을 복호하여 원 신호와의 차이값을 구한다. 이 차이값, 즉 양자화 에러를 다시 DCT (Discrete Cosine Transform) 변환하여 기본 화질적 계층보다 작은 양자화 계수로 양자화 한다. 이때 양자화 계수값은 이전 계층보다 6 작은 값을 선택하는데, 이는 H.264의 양자화기는 양자화 계수 값 6이 줄면 양자화 간격 크기가 반 정도로 감소 (비트량은 약 두 배 증가) 되도록 설계 되었기 때문이다.

IV. SVC 성능 분석

SVC 표준에서는 단순히 스케일러빌리티를 지원하는 것 뿐만 아니라, H.264 와 비교 시 비슷한 비트율-왜곡 (Rate-Distortion) 성능을 가지는 것을 목표로 한다. 본장에서는 SVC의 다양한 스케일러빌리티 지원 방식에 따른 성

능을 분석한다. 코덱은 SVC 표준 회의 reference software 인 JSVM 8.12을 이용했으며, 사용된 PC의 사양은 펜티엄 4, 2.93 GHz Intel® Core(TM) 2 CPU X6800, 그리고 3 GB RAM 이다. 테스트 영상으로는 SVC 표준회의에서 기술 경쟁을 위해 사용되는 football, foreman, 그리고 crew 영상열을 이용하였다. 복잡도는 부호화 혹은 복호화 과정에 소비되는 시간을 1/1000 초 단위로 측정하였고, PSNR을 이용하여 부호화 효율성을 판별하였다.

1. 기준(Anchor)

다음은 성능 비교에서의 기준을 위한 비트스트림을 부호화 설정이다. 기준(anchor) 부호화 설정은 QCIF(176x144) 및 CIF(352x288) 의 두 개의 공간적 계층을 가지며, 공간적 계층간 참조는 적응적으로 허용한다. GOP는 16으로 설정하여 5 종류의 시간적 스케일러빌리티를 제공할 수 있다. 참조영상 수(number of reference frame)는 1 이며, IDR을 제외한 화면내 부호화 프레임은 포함하지 않는다. 움직임 검색은 JSVM에서 지원하는 고속 방식이 쓰였으며, 루프 필터는 항상 사용된 반면 MGS는 사용되지 않았다. 기본 계층과 상위 계층의 최대 프레임율은 각각 15Hz, 30Hz이며, 양자화 계수의 초기값은 각각 32.25, 38 이다. 그림 9 는 앵커 설정을 이용하여 부호화한 후 각 함수들을 크게

5가지로 분류하여 도시한 것이다. H.264 에서와 같이 모드 결정을 위한 R-D 최적화 관련 모듈의 소요 시간이 70%가 넘으며, 그 다음이 디블리킹 및 상향 표본화에서 쓰이는 필터 관련 모듈이 높은 소요 시간을 차지하였다.

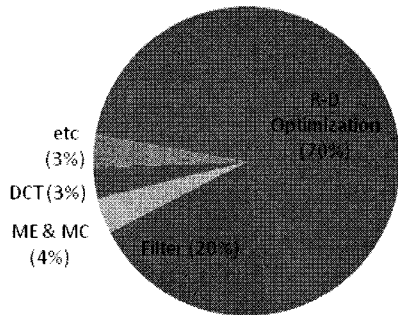


그림 9. SVC 부호기에서의 소요 시간 비율 (ME: 움직임 예측, MC: 움직임 보상)
Fig. 9. Elapsed time for categorized encoding modules of SVC

2. 공간적 계층간 참조 (inter-layer prediction)

SVC에서 이미 부호화 된 하위 공간적 계층의 정보(예, 화면내 텍스처, 움직임 정보, 블록 모드, 잔여 신호)는 적응적으로 상위 공간적 계층의 부호화에 이용될 수 있다. 만약 이를 허용하지 않는다면, 표 1 에서와 같이 비슷한 PSNR에서 비트율은 37% 정도 증가한다.

표 1. 공간적 계층간 참조의 성능 비교 (football)
Table 1. Performance of inter-layer prediction (football)

Inter-layer prediction	Time (sec.)		Bit-rate (kbps)	PSNR (dB)
	Encoder	Decoder		
Adaptive	616.562	3.312	575.842	31.415
Always used	246.656	3.781	554.074	31.091
Not allowed	431.343	3.297	763.073	31.374

계층간 참조를 항상 사용하도록 강제하는 것은 부호화 효율에 대한 큰 손실 없이 복잡도를 크게 감소 시킬 수 있다는 결과에 주시할 필요가 있다. 이는 그림 9에서와 같이 부호화 과정에서 70% 정도를 차지하는 R-D 최적화에 기반한 모드 결정 과정을 상위 계층에서는 생략할 수 있기 때문이다. 일반적으로 잔여 신호는 에리 정보이기 때문에 상관성이 없다고 알려져 있다. 따라서 이전 비디오 부호화 표준[2]-[5]

들에서는 주변 혹은 이전에 부호화된 블록의 잔여 신호를 참조 신호로써 활용하지 않았었다. 반면 SVC 에서는 공간적 하위 계층의 잔여 신호를 적응적으로 이용할 때, 부호화 효율을 가짐을 실험적으로 검증하고 표준으로 채택하였다. 표 2는 영상에 따른 잔여 신호 참조 알고리즘의 부호화 효율을 보여준다. 공간적 계층간 잔여 신호 참조를 적응적으로 이용할 때 최대 8.66%의 비트율 절감 효과를 얻을 수 있다.

표 2. 공간적 계층간 잔여 신호 참조 알고리즘의 부호화 효율
Table 2. Performance of inter-layer residual prediction

Sequence	Residual prediction	Bit-rate (Kbps)	Bit saving (%)	PSNR-Y (dB)	PSNR-U (dB)	PSNR-V (dB)
Soccer	Not Allowed	461.516		34.7287	41.7311	43.2837
	Allowed	443.106	-3.98	34.7601	41.65	43.2592
Harbour	Not Allowed	689.489		31.9134	41.8437	43.5996
	Allowed	674.682	-2.14	31.93	41.8545	43.5954
Football	Not Allowed	798.926		34.2182	39.3765	41.0282
	Allowed	729.989	-8.62	34.1951	39.3103	40.9925
Crew	Not Allowed	523.214		34.7708	39.8832	38.5016
	Allowed	477.293	-8.77	34.7554	39.7624	38.3828
City	Not Allowed	281.846		34.9142	43.0264	44.6104
	Allowed	278.209	-1.29	34.8915	43.0135	44.5811

3. 공간적 스케일러빌리티

그림 10은 초기 양자화 계수 값을 41, 38, 35, 32 로 줄여 가면서 QCIF과 CIF를 모두 지원하는 SVC(양자화 계수를 제외한 나머지 설정은 앵커와 동일), 계층간 독립적으로 부호화하여 QCIF과 CIF 비트스트림을 모두 포함하는 simulcast, 그리고 CIF 크기의 영상만을 지원하는 단계층 부호화 방식(H.2654)의 성능을 함께 보여 준다. 그리고 표 3은 공간적 스케일러빌리티를 지원하기 위한 SVC의 부호화 소요 시간을 나타낸다. SVC 부호화 방식은 기존 H.264의 시간적 중복성을 줄이는 기술 뿐만 아니라 하위 계층의 대응 동일 위치의 블록이 참조 블록으로 이용되어 공간적 중복성도 줄일 수 있다. 이는 그림 10에서와 같이 simulcast 방식 보다 높은 부호화 효율을 갖지만, 표 3처럼 블록 모드의

후보가 늘어난 만큼 추가적인 복잡도를 요구한다.

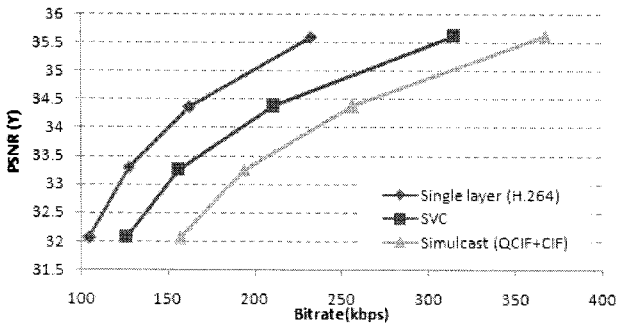


그림 10. 공간적 스케일러빌리티를 지원하는 SVC의 부호화 효율 비교 (Foreman)

Fig. 10. Coding efficiency for spatial scalability (Foreman)

표 3. SVC의 공간적 스케일러빌리티 지원 시 수행 시간 (Crew)

Table 3. Performance of spatial scalability (Crew)

Spatial scalability	Encoding time (sec.)	Bit-rate (kbps)	PSNR(dB)
CIF (H.264)	139.281	196.603	34.5
QCIF/CIF	223.769	237.056 (CIF)	34.4 (CIF)
QCIF/CIF/4CIF	5755.078	227.558 (CIF) 1494.110 (4CIF)	32.6 (CIF) 35.8(4CIF)

4. 시간적 스케일러빌리티

시간적 스케일러빌리티는 hierarchical B picture 방식에 의해 가능하며, 지원 가능한 레벨의 수는 GOP 크기에 의존한다. 예로 GOP 크기가 2n인 경우 n+1 가지의 프레임들을 제공할 수 있다. 표 4처럼 시간적 스케일러빌리티를 위한 소요 시간 및 비트율은 그 지원 레벨의 수와 큰 관계 없이 일정함을 확인 할 수 있다.

표 4. SVC의 시간적 스케일러빌리티 지원 시 성능 비교

Table 4. Performance of temporal scalability

Temporal levels	GOP size	Time (sec.)		Bit-rate (kbps)	PSNR (dB)
		Encoder	Decoder		
4	8	578.390	3.110	561.614	31.251
5	16	616.562	3.093	575.842	31.415
6	32	631.515	3.171	585.758	31.492
7	64	590.578	3.203	600.137	31.546

5. 화질적 스케일러빌리티

화질적 스케일러빌리티 지원을 위한 방식은 CGS 및 MGS 두 가지가 있다. 두 방식의 차이점은 다음과 같다. MGS는 원 영상과 화질적 하위 계층의 복원된 영상 사이의 양자화 잔여 신호를 부호화 하는 방식이다. 그러므로 화면 내 부호화가 오직 화질적 기본 계층에서만 이루어 진다. 반면 CGS는 입력 영상의 화면 크기가 모든 계층이 동일하다는 것을 제외하면 앞서 기술된 공간적 스케일러빌리티 지원을 위한 방식과 동일하게 부호화한다.

MGS는 임의 네트워크 노드에서 버려질 수 있기 때문에, SVC에서는 에러 전파(drift)를 고려한 알고리즘을 채택하였다. 움직임 예측 및 보상을 위해 참조로 사용된 영상의 화질에 따라, CGS와 MGS의 성능 비교를 그림 11에 도시하였다. 여기서 case 0는 현재 MGS 계층의 영상이 움직임 예측 및 보상 모두에 참조 신호로 이용된 경우이며, case 1은 가장 상위 MGS 계층의 영상이 움직임 예측에 사용되고 현재 MGS 계층의 영상이 움직임 보상에 이용된 경우이다. 마지막으로 case 2는 가장 상위 MGS 계층의 영상이 움직임 예측 및 보상에 이용된 경우이다. 실험 결과에서의 각 화질적 계층 사이의 양자화 계수 차이 값은 4이다.

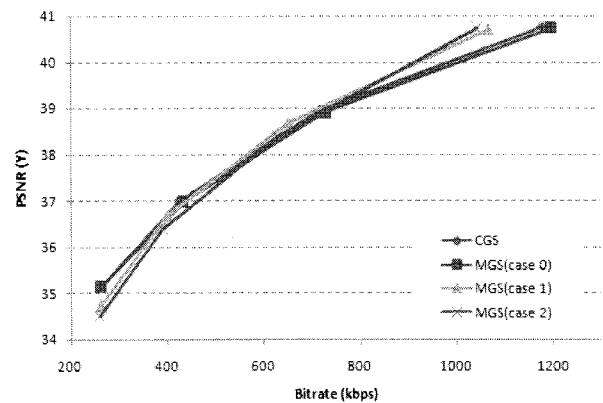


그림 11. 참조 영상의 화질에 따른 CGS와 MGS의 성능 비교 (Foreman)

Fig. 11. Coding efficiency for quality scalability (Foreman)

직관적으로 가장 상위 화질 계층의 복원 영상이 원영상과 가장 근접하다. 따라서 전송 시 패킷 손실이 없는 경우,

실험 결과와 같이 가장 상위 화질 계층의 복원 영상을 참조 신호로 활용하는 것이 부호화 효율이 가장 높다. 하지만 이 방식은 그림 12에서와 같이 패킷 손실이 발생한다면 예러 전파에 더 큰 영향을 받는 단점이 있다. 이 실험은 300장의 시퀀스에 대해 인터넷 패킷 손실 모델^[15]에 따라 모의 실험한 결과이다.

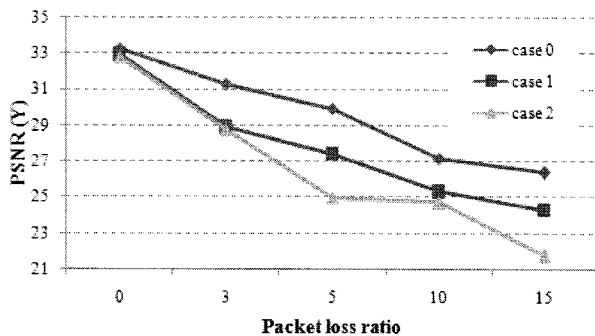


그림 12. MGS의 예러 전파 조절 방식에 따른 패킷 손실 시 화질 열화 비교
Fig. 12. Error robustness for MGS drift control methods (Foreman)

화질 스케일빌리티를 이용한 실제 응용 분야에서의 효율은 그림 13와 같다. SVC는 하나의 비트스트림으로 460kbps ~1160kbps까지 다양한 화질로 서비스 할 수 있는 반면, 단 계층 부호화 방식인 H.264는 채널의 변화와 무관하게 초기 부호화된 비트스트림의 비트율로만 전송 및 소비 될 수 있다. 그림에서와 같이 현재까지 reference S/W의 부호기 성능은 기존 H.264 방식과 비교하여 볼 때 약 20%의 부가적인 비트량을 필요로 한다.

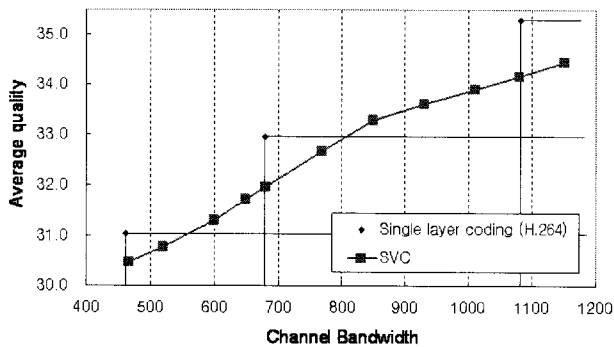


그림 13. 채널 대역폭 변화에 따른 SVC 및 H.264의 비트율 적응
Fig. 13. Bit-rate adaptation according to channel bandwidth

V. 결론

현재 ISO/IEC MPEG 과 ITU-T VCEG의 JVT에서는 멀티미디어를 다양한 망 및 단말 환경에서 서비스할 때 QoS를 보장하기 위한 목적으로 스케일러빌리티를 제공하기 위한 Scalable Video Coding에 대한 표준화를 진행하고 있으며 2007년 7월 현재 SVC phase 1에 대해 대부분 완료하였다. 본 논문에서는 스케일러빌리티의 유용성과 SVC의 특징 및 스케일러빌리티를 지원하기 위한 대표적인 기술들에 대해 소개 하였다. 또한 SVC 방식에서 다양한 스케일러빌리티를 제공할 때의 부호화 효율, 부가적인 비트량, 그리고 복잡도에 대해 비교 분석하였다. 이러한 성능 평가는 비디오 전송 및 소비와 관련된 여러 응용 분야에서 반드시 필요한 자료이며, 특히 실시간 스트리밍을 위한 SVC 코덱 개발에서 반드시 선행 되어야 할 연구이다.

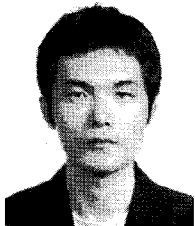
결론적으로 SVC는 이종망 환경에서 발생하는 사용자의 선호도, 소비 단말기와 네트워크의 조건 등 다양한 전달 및 소비 환경에 효과적으로 대처할 수 있으며, 스케일러빌리티를 갖는 계층들의 조합을 하나의 비트열로 부호화하여 적응적인 전달/소비를 제공함으로써 여러 소비 환경 각각에 QoS를 보장하는 콘텐츠를 서비스할 수 있다^[16]. 더욱이 동일한 생산 비용이 여러 개의 전달/소비 환경에 공유될 수 있기 때문에 비용과 효율의 생산성 측면에서도 SVC는 매우 매력적이다. 추후 스케일러블 비디오는 최근의 멀티미디어 서비스 환경에서 보다 향상된 비디오 서비스와 새로운 시장의 기회를 제공할 수 있는 중요한 기술 중 하나가 될 것이다.

참고 문헌

- [1] A. Vetro, C. Christopoulos, and H. Sun. "Video transcoding architectures and techniques: An overview", IEEE Signal Processing Mag., vol. 20, pp. 18-29, Mar. 2003.
- [2] ITU-T and ISO/IEC JTC1, "Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video," ITU-T Recommendation H.262 ISO/IEC 13818-2 (MPEG-2), 1994.
- [3] ITU-T, "Video coding for low bit rate communication," ITU-T Recommendation H.263, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

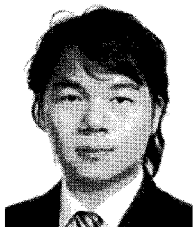
- [4] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," IEEE Trans. on Circuit System and Video Technology, vol. 11, no. 3, pp. 301-317, Mar. 2001.
- [5] ISO/IEC 14496-2, "Information technology-Coding of audio-visual objects - part 2: Visual", International Standard, second edition, December 2001.
- [6] ITU-T document, "Joint Draft 10 of SVC Amendment," Joint Video Team JVT-W201, JVT 23nd meeting, San Jose, USA, April 2007.
- [7] ITU-T document, "Proposed modifications for Joint Scalable Video Model," Joint Video Team JVT-W202, JVT 23nd meeting, San Jose, USA, April 2007.
- [8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable extension of the H.264/MPEG-4 AVC video coding standard," Joint Video Team, doc. JVT-U145, Hangzhou, China, October, 2006.
- [9] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC), Version 1: May 2003, Version 2: Jan. 2004, Version 3: Sep. 2004, Version 4: July 2005.
- [10] SVC Reference Software, [Online]. Available: CVS server (host address : garcon.iemt.rwth-aachen.de).
- [11] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," Proceedings of ICME'06, Toronto, Canada, July 2006.
- [12] H. Schwarz, D. Marpe, and T. Wiegand, "Comparison of MCTF and closed-loop hierarchical B pictures," Joint Video Team, doc. JVT-P059, Poznan, Poland, July 2005.
- [13] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granular scalable video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 3, pp. 332-344, Mar. 2001.
- [14] Hung-Chih Lin, Wen-Hsiao Peng, and Hsueh-Ming Hang, "Low-complexity Macroblock Mode Decision Algorithm for Combined CGS and Temporal Scalability," Joint Video Team JVT-W029, JVT 23rd meeting, San Jose, USA, April 2007.
- [15] Y. Guo, H.Q. Li and Y.K. Wang, "SVC/AVC loss simulator," ISO/IEC JTC 1/SC 29/WG 11, JVT-Q069, Oct. 2005.
- [16] Haechul Choi, Jung Won Kang, and Jae-Gon Kim, "Dynamic and Interoperable Adaptation of SVC for QoS-Enabled Streaming," IEEE Transaction on Consumer Electronics, Vol. 53, Issue 2, pp. 384-389, May 2007.

저 자 소 개



최 해 철

- 1997년 2월 : 경북대학교 전자공학과(학사)
- 1999년 2월 : KAIST전기 및 전자공학과(석사)
- 2004년 8월 : KAIST전기 및 전자공학과(박사)
- 2004년 9월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 선임연구원
- 2007년 3월 ~ 현재 : 과학기술연합대학원대학교 이동통신 및 디지털방송공학겸임 조교수
- 주관심분야 : 영상통신, 영상분할, 비디오 부호화, Scalable Video Coding, MPEG Video, ITU-T VCEG



이 경 일

- 2004년 2월 : 경희대학교 컴퓨터공학과(학사)
- 2006년 2월 : 경희대학교 컴퓨터공학과(석사)
- 2006년 3월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 연구원
- 주관심분야 : 비디오 부호화, H.264/AVC, Scalable Video Coding, MPEG Video, ITU-T VCEG, 비디오 스트리밍, 스케일러블 스트리밍

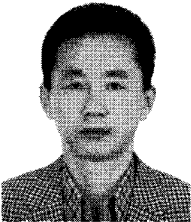
 저 자 소 개

**강 정 원**

- 1993년 2월 : 한국항공대학교 항공전자공학과 (학사)
- 1995년 2월 : 한국항공대학교 항공전자공학과 (석사)
- 2003년 8월 : Georgia Institute of Technology ECE (박사)
- 2003년 10월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 선임연구원
- 2007년 3월 ~ 현재 : 과학기술연합대학원대학교 이동통신 및 디지털방송공학겸임 조교수
- 주관심분야 : 비디오신호처리, 비디오 부호화, Scalable Video Coding, 비디오 적응 변환

**배 성 준**

- 1997년 2월 : 고려대학교 전자공학과(학사)
- 1999년 2월 : KAIST전기 및 전자공학과(석사)
- 2004년 8월 : KAIST전기 및 전자공학과(박사)
- 2004년 8월 ~ 2005년 10월 : 하나로텔레콤 기술전략팀
- 2005년 10월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 선임연구원
- 주관심분야 : 비디오 부호화, Scalable Video Coding, MPEG Video, ITU-T VCEG, 비디오스트리밍, 스케일러블스트리밍

**유 정 주**

- 1982년 2월 : 광운대학교 전자통신공학과(학사)
- 1984년 2월 : 광운대학교 전자통신공학과(석사)
- 2001년 7월 : 영국 LANCASTER 대학교 컴퓨터공학과(박사)
- 1984년 3월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 방통융합미디어연구팀장
- 주관심분야 : 영상통신, 비디오 부호화, Scalable Video Coding, MPEG Video, ITU-T VCEG, 비디오스트리밍, 스케일러블스트리밍