

주 제

스케일러블 비디오 부호화 개요 및 표준화 동향

한국전자통신연구원 최해철, 김재곤, 홍진우

차례

I. 서 론

II. 스케일러블 비디오 부호화 개요 및 표준화 동향

III. 스케일러블 비디오 부호화 기술

IV. 결 론

I. 서 론

인터넷의 발전과 함께 정지영상 및 동영상 등의 정보를 포함한 다양한 멀티미디어 서비스에 대한 요구가 커지고 있다. 이에 따라 한정된 대역폭 내에서 증가하는 멀티미디어 서비스 수요를 처리하기 위한 여러 가지 압축 방식이 연구되어 왔으며, 특히 1988년에 디지털 동영상 정보의 부호화 및 저장에 대한 표준 규격에 대한 필요성이 대두되면서 ITU-T에서는 유/무선 통신망 환경에서 동영상 서비스를 위한 표준 규격인 H.261 [1]과 H.263 [2]을 제정해 왔고, ISO/IEC는 다양한 응용 분야에서 멀티미디어 서비스 제공을 위한 동영상 표준 규격인 MPEG-2 [3], MPEG-4 [4]의 표준화를 위해 노력해 왔다. 이러한 노력의 결과로 압축 효율이 향상된 동영상 부호화 방법이 개발되었다. 이러한 부호화 방식은 인터넷, 휴대폰, DVD, DMB 응용에서 광범위하게 사용되고 있

며, 동영상 부호화 표준 기술은 막대한 시장을 형성하고 있는 디지털 콘텐츠 산업의 기술 경쟁력을 결정짓는 핵심 기술이 되었다

하지만 압축 부호화 기술을 개발하던 시기와 달리 최근의 디지털 멀티미디어의 생성, 전송 및 소비 환경은 광범위하게 급변하고 있다. 서로 다른 특성(대역폭, 패킷 손실률)을 갖는 이종망이 혼재해 있으며, 통신과 방송의 융합화가 빠르게 진행되고 있다. 더욱이 비디오를 소비하는 단말 또한 다양한 화면 크기와 연산 능력을 가진 새로운 제품들이 쏟아지고 있다. 이러한 환경에서 고정된 영상 포맷과 비트율(bitrate)로 부호화하고 전송하는 Non-scalable 비디오는 End-to-End QoS에 대한 사용자의 요구를 효과적으로 충족시킬 수 없다. 변환부호화(transcoding)나 비트량변환(transrating)을 통하여 제한된 범위에서의 영상 포맷 및 비트율 적응이 가능하지만 큰 복잡도로 인해 실시간 적응이 요구되는 환경에서 비효율적

이다. 따라서 언제 어디서나 어떠한 단말로든 효과적으로 동영상 전송/소비하기 위해서는 앞서 기술한 고압축 특성을 가진 부호화 기술 뿐만 아니라 전송 및 소비 환경에 따라 적응적으로 변환할 수 있는 스케일러빌리티(scalability) 기능이 매우 중요하다.

동영상에 있어서 스케일러빌리티 기능을 위한 기존 연구는 MPEG-2 [3]와 MPEG-4 (part2) [4] [5] 비디오에서 계층 부호화에 기반한 스케일러빌리티의 지원을 위한 시도가 있었으나 하나의 계층으로 부호화하는 기존 압축 부호화 기법에 비해 부호화 효율이 낮고, 여러 종류의 스케일러빌리티를 종합적으로 지원하지 못한다. 이러한 문제점을 해결하면서 실시간 비디오 전송에서 QoS를 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]이라는 표준화를 MPEG-4 part 10 AVC/H.264 [7]의 수정판(Amend-ment)으로 진행하고 있으며, 여기에는 유럽, 미국, 한국이 적극적으로 참여하여 기여하고 있다.

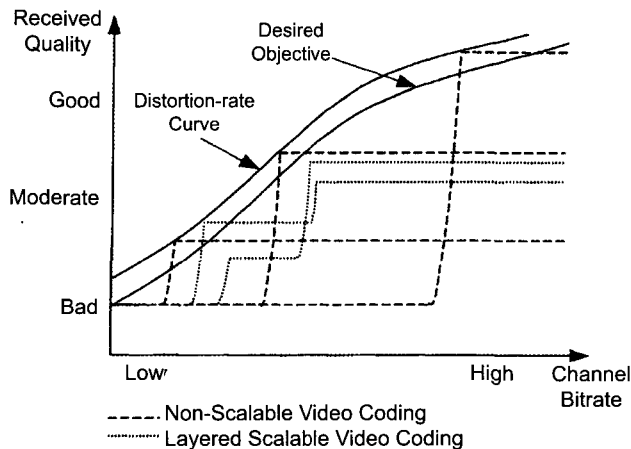
본 고에서는 스케일러블 비디오 부호화의 개요에

대해 설명하고, 현재 표준화가 진행 중인 SVC의 표준화 동향을 기술한다. 그리고 SVC에 표준으로 채택된 기술 중 시간/공간/화질적 스케일러빌리티를 제공하기 위한 대표적인 알고리즘에 대해 설명한다.

II. 스케일러블 비디오 부호화 개요 및 표준화 동향

1. 스케일러블 비디오 부호화 개요

네트워크를 통한 멀티미디어 서비스에서 비디오 부호화의 목적은 일정한 비트율에서 화질을 우회하는 것이다. 이 때 그 비트스트림(bit-stream)은 어떤 비트율에서도 복호화가 되어야 한다. Non-scalable 비디오 부호화는 하나의 비트스트림만을 생성한다. 이에 비하여 스케일러블 비디오 부호화는 하나의 비디오 시퀀스를 여러 개의 계층, 즉 기본 계층(base layer)과 상위 계층(enhancement layer)으로 나누어서 압축한다. 기본 계층이란 독립적으로 복호 가능한 비트스트림이다. Non-scalable



(그림 1) 비디오 부호화 R-D 성능 그래프

비디오 부호화는 기본 계층으로만 구성된다고 할 수 있다. 향상 계층은 기본 계층에 있는 비트스트림을 개선하기 위해서 사용되는 추가 비트스트림이다. 향상 계층은 독립적으로 복호할 수 없고 기본 계층을 참조해서 복호해야 보다 좋은 화질을 얻게 되며 전체 비트스트림을 사용하여야 최고의 화질을 얻을 수 있다.

(그림 1)에서의 비트율-왜곡 곡선(distortion-rate curve)은 서로 다른 부호화 기술을 이용할 때의 화질을 보여주고 있다. 보는 바와 같이 계층적 스케일러블 부호화 기술은 하나의 계층으로 이루어진 non-scalable 곡선을 두 계층으로 된 곡선으로 바꾸었다. 스케일러블 비디오 부호화의 목표는 R-D 곡선과 평행한 연속적인 곡선을 얻는 것이다.

SVC 표준에서 “스케일러빌리티(scalability)”란 H.264/MPEG4-AVC [7] 부호화와 비교 시 비슷한 비트율-왜곡(Rate-Distortion) 성능을 가짐과 동시에 공간, 시간, 화질 측면에서 비트스트림의 한 부분을 제거할 수 있는 기능(functionality)으로 정의되어 있다. 실제 기존 H.264와 비교 시 10% 내외의 비트율을 허용한다.

SVC는 여러 개의 비디오 계층을 하나의 비트열로 부호화하며, 각 층은 각각의 비트율, 프레임율, 영상 크기 및 화질을 가지고 있다. 즉 비디오를 공간적/시간적/화질적 차원의 임의 값을 가지는 비트열로 부호화하며, 그 세 가지 차원의 조합에 따라 폭넓은 스케일러빌리티를 제공할 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 계층과 스케일러블 상위 계층들로 구성된다. 여기서 기본 계층 및 연속되는 상위 계층의 정보가 함께 이용되어 보다 개선된 비디오 비트스트림을 만든다. 그 예로 화질 스케일러빌리티는 하나의 비트열로부터 동일한 공간 및 시간 축(dimension)을 갖지만 각각 다른 화질을 갖는 비

오 계층들을 만들어 낸다. 일반적으로 기본 계층은 기본적인 비디오 화질을 제공하고, 연속된 상위 계층은 이전 계층들로 만들어진 비디오보다 높은 화질을 갖도록 부호화한다. 마찬가지로 시간 및 공간 해상도에 서로 동일한 원리를 적용하여 스케일러빌리티를 지원한다. 하나의 차원에 대한 스케일러빌리티로는 최적의 가변성과 적응성을 제공할 수 없는 응용분야가 있을 수 있으며, 이는 시간, 공간, 화질 측면 모두의 스케일러빌리티를 필요로 한다.

2. 스케일러블 비디오 부호화 표준화 동향

공간/시간/화질적 스케일러빌리티는 가장 기본적인 스케일러블 기능으로서 MPEG-2와 MPEG-4에 모두 포함되어 있다. 근래의 스케일러빌리티 기능 연구로 MPEG-4 FGS (Fine Granularity Scalability)가 있다. MPEG-4 FGS[5]는 세밀한 간격으로 비트율을 조절할 수 있을 뿐만 아니라 전송하고자 하는 비트율은 압축하는 시점이 아닌 전송하는 시점에 결정할 수 있게 된다. 따라서 여러 비트율로 콘텐츠를 전송하는 경우에 하나의 압축된 비트스트림으로부터 여러 비트율의 전송용 비트스트림을 매우 손쉽게 생성하는 것이 가능해졌다. MPEG-4 FGS 표준화가 종료된 이후, MPEG에서는 MPEG-21 (part 13)을 시작점으로 SVC라는 소그룹(AdHoc)을 결성하였다. 2004년의 CFP(Call for Proposal) [8]에서 다음 두 가지 범위의 시나리오를 설정하고 제안 기술을 검증하였다. 첫 번째 시나리오는 넓은 범위의 스케일러빌리티로 감시시스템, 방송, 저장장치를 적용대상으로 삼고 있으며, 1.2초 범위의 임의 접근(random access)을 요구한다. 두 번째 시나리오는 보다 제한된 범위의 스케일러빌리티로 스트리밍(streaming), 이동통신 등의 분야에 관한 것

이며 임의 접근의 제한은 없다.

기본적인 SVC의 프레임워크로 다음 2가지의 기법이 제안 검증되었다.

- DCT 기반의 기법: 기존 MPEG 표준들과 유사한 접근 방식을 취하며 이들과 호환성을 갖는다.
- 웨이블릿(wavelet) 기반 기법: 최초로 비디오 압축 도메인에서 재귀하지 않는(non-recursive) 구조를 제공한다.

검증 결과, 웨이블릿 기반 기법이 근본적으로 스케일러빌리티를 포함하고는 있으나, 웨이블릿 필터의 제한으로 인해 압축 효율성 측면에서 독일 HHI에서 제안한 DCT 기반의 기법보다 낮은 성능을 보여주었다. 따라서 DCT 기반 기법이 SVC의 기본 프레임워크로 채택 되었으며, 이를 기반으로 많은 알고리즘들이 연구, 개발되고 있다.

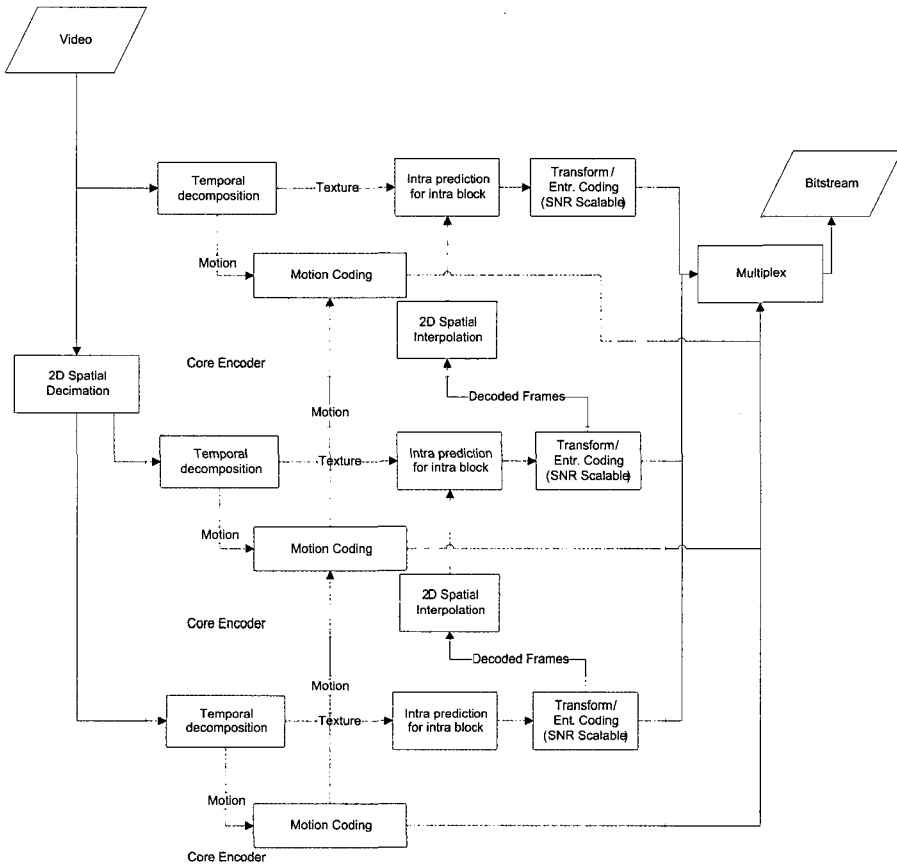
지난 2005년 1월 MPEG 홍콩 회의에서, 또 다른 비디오 국제 표준화 단체인 ITU-T의 VCEG이 SVC 표준화에 동참하면서 JVT에서 스케일러블 비디오 표준화가 수행 중이다. 2006년 1월에 Committee Draft가 발표되어 있으며, 2006년 7월에는 Final Committee Draft가 예정되어 있고, 2007년 1월에 Final Document of International Standard가 확정될 예정이다. 2006년 7월 현재 JD(Joint Draft) 6 [6]와 JSVM(Joint Scalable Video Model) 6 [9]가 발표 되었다.

III. 스케일러블 비디오 부호화 기술

SVC의 부호기에 대해 기술하고 있는 JSVM [9]은 다음 두 가지 방법을 통해 스케일러블 기능을 제공한다. 한 가지는 비트플레인(bit-plane) 기반 산술 코딩(arithmetic coding)과 같은 부호화 방법이고,

다른 하나는 기존의 여러 스케일러블 표준에서 사용되었던 계층 기반 기법이다. JSVM에서는 시간적/공간적/화질적 스케일러빌리티를 모두 제공하기 위해 상기 두 가지 방법을 모두 사용하였다. 시간적 스케일러빌리티는 MCTF(Motion Compensated Temporal Filtering) 및 Hierarchical B picturer 구조로써 지원하고, 공간적인 스케일러빌리티는 down/up 표본화(sampling)를 통해 여러 해상도를 가지는 영상들을 구성함으로써 가능하게 하며, 화질(quality 혹은 SNR) 스케일러빌리티를 위해서는 CGS의 경우 공간 스케일러빌리티를 위한 기법과 동일한 방식에 양자화 계수(quantization parameter)만을 변경하며, FGS의 경우 2-scan 방식과 cyclic 부호화 기법을 이용하여 스케일러빌리티를 지원한다.

JSVM 부호기 구조는 제공하려는 스케일러빌리티의 종류 및 단계에 따라 달라진다. 3 개의 공간적 계층으로 구성하는 경우 SVC 부호기 구조는 (그림 2)와 같다. 공간적 스케일러빌리티를 위해 입력 영상의 공간 해상도를 줄여 크기가 작아진 영상을 각 계층에 입력 시킨다. 이때 통상적으로 MPEG-4의 down-sampling 필터를 사용하지만, 이 필터에 대해서는 표준으로 규정하지 않고 있다. 그 후 공간적으로 분할된 각 계층의 입력 비디오 신호에 Hierarchical B picture (혹은 MCTF)를 적용하여 시간적 스케일러빌리티를 지원한다. 이 과정의 결과는 시간적인 저주파 영상 {L}과 고주파 영상 {H}의 집합, 잔여 텍스처 정보, 블록 단위의 움직임 정보이다. 움직임 정보와 잔여 텍스처 정보는 AVC 알고리즘을 기초로 부호화하면서, 부가적으로 화질 스케일러빌리티를 제공하기 위해 CGS(Coarse Grain Scalability) 혹은 FGS(Fine Grain Scalability) 기법을 적용한다. 또한 전체적으로 각 공간 계층 간의 정보 중복성을 줄이기 위해, 움직임 정보를 그 자체로 부호화하지 않고



(그림 2) 공간적 3 계층의 스케일러빌리티를 제공하는 스케일러블 부호기

하위 계층에 있는 대응 블록의 움직임 정보와의 차이 혹은 그 수정 값을 포함시켜 부호화하는 방식을 이용할 수 있다. 잔여 텍스처(residual texture) 정보 및 화면내(intra) 매크로블록의 예측 과정에 대해서도 동일한 개념의 방식을 AVC에서의 방식들과 함께 적용한다. 유의할 점은 SVC는 H.264/MPEG4-AVC의 수정본으로 표준화가 진행 중이기에, SVC의 공간적 기본 계층은 반드시 H.264/MPEG4-AVC와 완벽한 호환성을 가져야 한다.

다음 절에서 시간/공간/화질 스케일러빌리티 지원을 위한 특징적인 기술들에 대해 각각 기술한다.

1. 시간 스케일러빌리티 관련 기술

1.1. Hierarchical B picture

AVC에서 모든 영상은 참조 영상으로 사용할 수 있고, 다음에 오는 영상의 움직임 보상 예측에 이용할 수 있다. Memory management control operation (MMCO) commands는 복호화된 영상을 저장할 수 있는 decoded picture buffer (DPB)를 적절하게 제어한다. 그리고 DPB에 저장된 다른 영상의 움직임 보상 예측에 이용되는 참조 영상들은 reference picture list re-ordering (RPLP) commands를 사

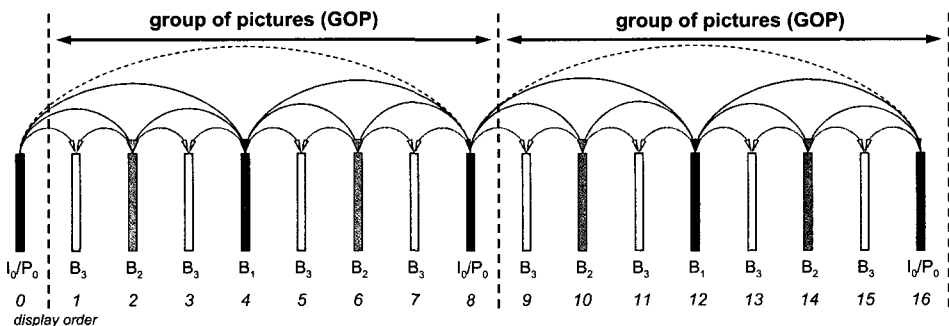
용하여 임의적으로 선택이 가능하다. 이러한 AVC의 특징들은 이전 비디오 부호화 방식들이 지원하지 못했던 임의적인 부호화 및 참조 영상의 선택을 가능하게 한다. 이를 기반으로 B 영상의 참조를 허용하는 General B picture를 정의함으로써 시간적 스케일러빌리티를 지원하는 Hierarchical B picture가 제안되었다.

(그림 3)은 4 단계를 갖는 전형적인 hierarchical B picture 구조를 보여준다. 비디오의 첫 영상은 IDR (Instantaneous Decoder Refresh) 영상으로서 화면내 부호화된다. 소위 key picture라고 불리는 IDR picture (그림 3의 검은색 표시)는 일정한 간격을 두어 부호화 된다. 이전에 부호화된 모든 영상들이 현재 부호화되는 영상보다 화면 표시 순서 (display order) 상 앞서 위치할 때, 이것을 key picture라 부른다. 시간적으로 현재 key picture와 이전 key picture 사이에 위치하고 있는 I장의 key picture와 모든 non-key picture들을 하나의 group of pictures (GOP)라고 한다. Key picture는 임의 접근을 위해 화면내 부호화되거나, 이전 key-picture를 참조 영상으로 움직임 보상 예측을 통해 P 영상으로 화면간(inter) 부호화 된다. GOP의 나머지 영상들은 (그림 3)과 같이 계층적으로 예측된다.

시간적 스케일러빌리티는 SVC 표준화 초반 MCTF에 의해 구현 되었었다. MCTF는 웨이블릿 기반의 기법으로써 시간 축으로 저주파(예: 이웃 영상들간의 평균 영상) 및 고주파 영상(예: 이웃 영상들간의 차이 영상)으로 필터링하여 시간적 스케일러빌리티는 지원하는 기법이다. 하지만, 복호기의 과도한 복잡도 문제로 때문에 현재 복호기 표준에서는 MCTF가 사라지게 되었고, 부호기에서만 non-normative로 남아있다. 이 MCTF로 인해 시간적 분할(Decomposition)을 할 때 어떤 영상을 참조하여 예측 영상을 만드느냐에 따라 Open-loop(원영상 참조)와 Closed-loop(복호된 영상 참조) 두 가지로 구분되는 기술들이 제안되었다.

2. 공간적 스케일러빌리티 및 CGS 관련 기술

공간적 스케일러빌리티는 각 영상 크기의 계층을 쌓는 피라미드 개념으로 지원한다. 이 때 각 해상도의 비디오는 하위 계층의 부호 결과물인 움직임, 텍스처 및 잔여 신호 정보를 이용 함으로써 새로운 계층으로 부호화된다. Coarse Grain SNR (CGS) 스케일러빌리티도 공간적 스케일러빌리티와 동일한 방식으로



(그림 3) GOP 크기가 8이며, 4개의 시간적 레벨을 갖는 hierarchical B picture 구조

지원한다. 단지 다른 점은 각 계층의 입력 영상 크기가 다른 경우 공간적 스케일러빌리티를 지원하는 것이며, 그렇지 않고 동일한 경우 CGS를 지원하는 것이다. 따라서 공간적 스케일러빌리티에서만 계층간 참조 시 움직임과 텍스처 정보가 이 영상 해상도의 비율에 따라 변화한다. 본 절에서는 각 공간적 또는 CGS 계층의 계층간(inter layer) 상관관계를 이용하기 위해서 제안된 다양한 계층간 예측 기술을 설명한다.

2.1. Inter-layer intra texture prediction

이전 계층의 정보를 사용하는 계층간 화면내 텍스처 예측은 Intra_Base 매크로블록(macroblock) 모드에서 사용된다. 이 모드에는 화면내 텍스처 예측은 다음 3가지로 가능하게 한다.

- Unrestricted inter-layer intra texture prediction

계층간 화면내 텍스처 예측은 제한 없이 계층의 어떠한 블록에서도 사용 할 수 있다. 이 설정에서 복호기는 목표 영상 해상도의 복구를 위해 비트스트림의 모든 하위 영상 해상도를 가지는 계층들을 복호 decode할 필요가 있다. 이 모드는 Joint Draft에서는

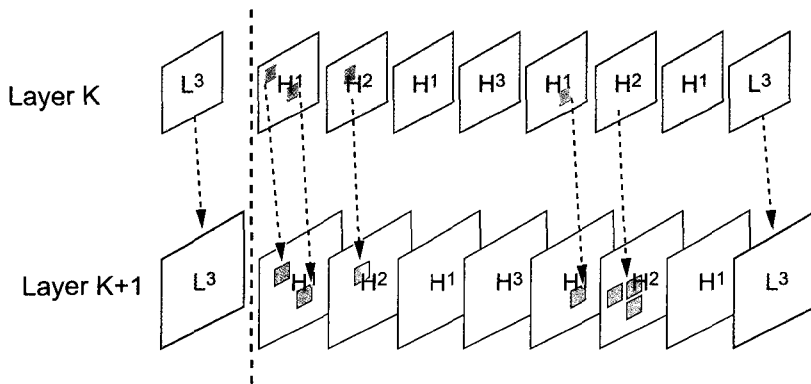
제공되지 않으나, JSVM software [9]에서는 compiler flag으로 설정 될 수 있다.

- Constrained inter-layer intra texture prediction

Non-key picture 영상들에서의 계층간 화면내 텍스처 예측은 기본 계층의 대응 위치의 블록이 화면내 부호화된 매크로블록 안에 위치할 때만 적용된다. 이 모드에서 key picture를 제외한 영상의 시간적 복구(temporal reconstruction) 과정은 실제로 복호화된 공간 또는 CGS 계층을 필요로 한다. 반면에 key picture를 위해서는 multiple decoding loop가 필요하다.

- Single-loop decoding을 위한 constrained inter-layer intra texture prediction

모든 영상들에 대해 계층간 화면내 텍스처 예측은 기본 계층의 대응 위치 블록이 화면내 부호화된 매크로블록 안에 위치할 때만 적용된다. 이 설정에서는 목표 영상 해상도를 위해 single decoding loop만이 쓰인다. Key picture를 위한 single 또는 multiple loop는 기본 계층의 key picture에서 사용되는 Picture



(그림 4) 화면내 macroblock을 위한 low-complexity 계층간 예측

Parameter Set의 constrained_intra_pred_flag에 의해 제어된다.

Intra_Base 모드에서 부호화된 고주파 매크로블록을 화면내 예측을 하기 위해서, 기본 계층 고주파 신호의 대응 위치 블록은 de-blocking 필터링 되고 궁극적으로 (그림 4)와 같이 보간(interpolation)된다. 보간 과정은 상위 계층과 기본 계층간의 실제 공간 해상도 비율에 의해 결정된다. 이 비율이 2일 경우, 보간 과정은 AVC의 half-pel 보간 필터를 사용한다. 반면에 이 비율이 1 또는 2가 아닌 non-dyadic 인 경우, 확장 공간 스케일러블리티에서의 필터[6]가 이용된다.

보간 필터를 사용하기 전에, 이전 계층의 화면내 매크로블록은 경계 확장(border extension) 작업을 하여 4-sample border로 확장된다. 경계 확장을 수행하기 전에, deblocking 필터가 모든 화면내 매크로블록의 경계 부분에 적용된다. 공간 계층이 여러 개인 경우 과도한 필터링으로 화면 경계가 열화되는 것을 막기 위해 경계 확장 과정에서의 샘플 수를 줄일 수 있다

2.2. Inter-layer motion prediction

상위 계층의 움직임 정보를 부호화할 때 이전 계층의 모드를 이용하는 두 가지 매크로블록 모드가 존재한다. 이 두 모드는 Base Layer Mode와 Quarter-pel Refinement Mode이다.

• Base Layer mode

이 경우, 상위 계층에서는 움직임 정보 및 매크로블록의 분할(partitioning) 정보를 별도로 부호화하지 않는다. 이 매크로블록 모드는 이전 계층의 대응 매크로블록의 분할정보 및 움직임/예측 정보가 그대로 이용되는 것을 의미한다. 이전 계층의 공간 해상도가 현재 계층보다 작을 때, 매크로블록 분할 정보와

움직임 벡터가 비율에 맞춰 조정된다. 만약 이전 계층의 매크로블록이 화면내 매크로블록이면, 현재 매크로블록 모드는 Intra_Base (이전 계층으로부터 예측된 화면내 매크로블록)로 표시된다. 현재 매크로블록의 분할을 위해, 이전 계층의 대응 위치 매크로블록/서브-매크로블록, 분할 지시자(partition index)와 같은 참조영상 지시자(reference index)가 사용된다. 이 때 움직임 벡터는 항상 계층과 기본 계층간의 해상도 비율을 곱하여 사용한다.

• Quarter-pel Refinement Mode

이는 이전 계층이 현재 계층보다 작은 공간 해상도를 가질 때만 쓰인다. Quarter-pel Refinement 모드는 Base Layer Mode 와 유사하다. 참조영상 지시자와 움직임 벡터 정보뿐만 아니라 매크로블록 분할 정보도 사용된다. 이 때 분할은 기본 계층 모드를 위해 일정 비율로 축소된다. 반면에 움직임 벡터의 경우, 각 quarter-sample된 움직임 벡터 보정(refinement) 값(+1 혹은 -1)이 추가적으로 전송되고, 이 보정된 값이 더해진다.

만약 위에 언급된 두 방법을 사용하지 않으면, 대응 참조영상 지시자, 움직임벡터, 매크로블록 모드는 AVC 규격(syntax)에 따라서 부호화된다.

2.3. Inter-layer residual prediction

한 계층에서 다음 계층으로 전달되는 움직임 정보가 변하면, 잔여 정보를 예측하는 것에 도움이 될 수도 있고 되지 않을 수도 있다. 만약 현재 계층 블록의 움직임 정보가 이전 계층의 대응 위치 블록의 움직임 벡터와 동일하거나 거의 유사하면, 이전 계층의 잔여 신호가 현재 계층의 잔여 신호를 위한 예측으로 사용할 때 부호화 효율이 증대될 수 있다. 그래서 현재 잔여 신호와 이전 계층의 복호된 잔여 신호의 차가 부호화된다. 그러나 움직임 벡터가 유사하지 않거나, 매우

다른 경우에는 잔여 신호의 예측이 부호화 효율을 저하시킬 수 있다. 그러므로 잔여 신호의 예측을 다음의 flag로 선택하여 사용한다.

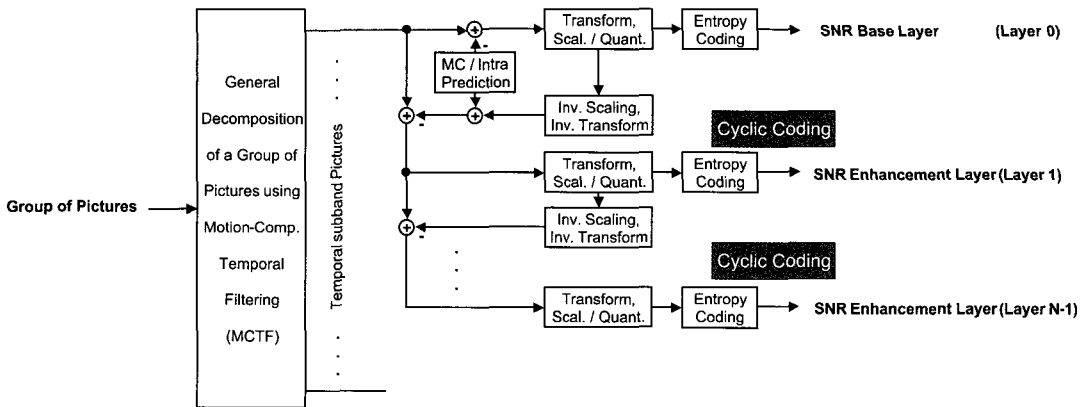
- Residual_prediction_flag

만약 이전 계층이 현재 계층의 절반인 공간 해상도를 가질 경우, 잔여 신호는 예측 신호로 사용되기 전에 separable bi-linear 필터를 사용하여 상향 표본화(up-sample)된다. 계층간 공간 해상도 비율이 1 또는 2가 아닐 경우, 보간 작업은 AVC의 quarter-pel을 기반으로 하여 작동된다. 보간 필터는 변환 블록 간의 경계를 건너서 적용되지 않는다.

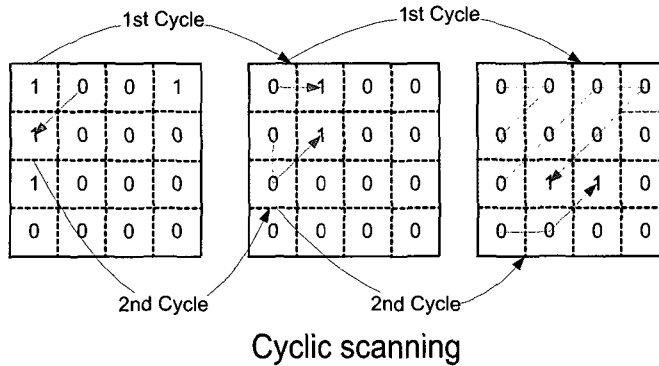
3. 화질 스케일러빌리티 관련 기술

JSVM은 화질적 스케일러빌리티를 위해서 (그림 5)와 같이 양자화에 의한 오차 신호를 이전 계층보다 적은 양자화 계수 값으로 보정하여 부호화하는 개념이다. 자세히 살펴 보면, 우선 시간적 분할(decomposition) 이후의 신호들을 기본 화질로 부호화한다. Hierarchical B picture 구조를 이용하는 경우, I, P, B 영상을 DCT 변환 후 산술 부호화 한다. 이 후

복호된 기본 화질의 신호와 원 신호와의 차이, 즉 양자화 에러를 다시 DCT 변환하여 기본 계층보다 적은 양자화 계수(Quantization parameter)로 양자화 한다. 이때 QP는 이전 계층보다 6 적은 값을 선택하는데, 이는 H.264에서 QP 값이 6이 변하면 양자화 step size가 두 배 변하게 되는 데서 기인한다.(비트량도 대략 두 배 증가) 이렇게 양자화 에러를 단계적으로 보정 부호화 하는 부호화 방식을 Progressive Refinement 라고 하며, 부호화된 슬라이스를 Progressive Refinement (PR) slice라 한다. PR 슬라이스는 2 scan 기법 및 cyclic 부호화 기법을 이용하여 부호화 함으로써 전체 프레임에서 각 블록별로 변환한 후에 주파수가 낮은 순서대로 전송을 하게 된다. 따라서 복호화 단에서 주파수별로 정보를 전송받으면 중간에 데이터가 끊어 지더라도 복호가 가능하다. 즉 임의의 부분을 제거하여도 복호에 영향을 미치지 않도록 FGS를 지원한다. (그림 6)은 cyclic 부호화 방식의 예이다. Cyclic 부호화 방식의 목적은 한 블록 내의 정보를 전체 비트스트림으로 분산시켜 FGS 패킷의 임의의 부분이 제거 되었더라도 전체 영상의 화질 차이가 적도록 하는 것이다. 이를 위해 사이클(cyclic)이라는 것을 정의하였다. 각 4x4 블록에



(그림 5) 화질 스케일러빌리티를 위한 부호화기 구조



(그림 6) Cyclic 부호화 방식의 예

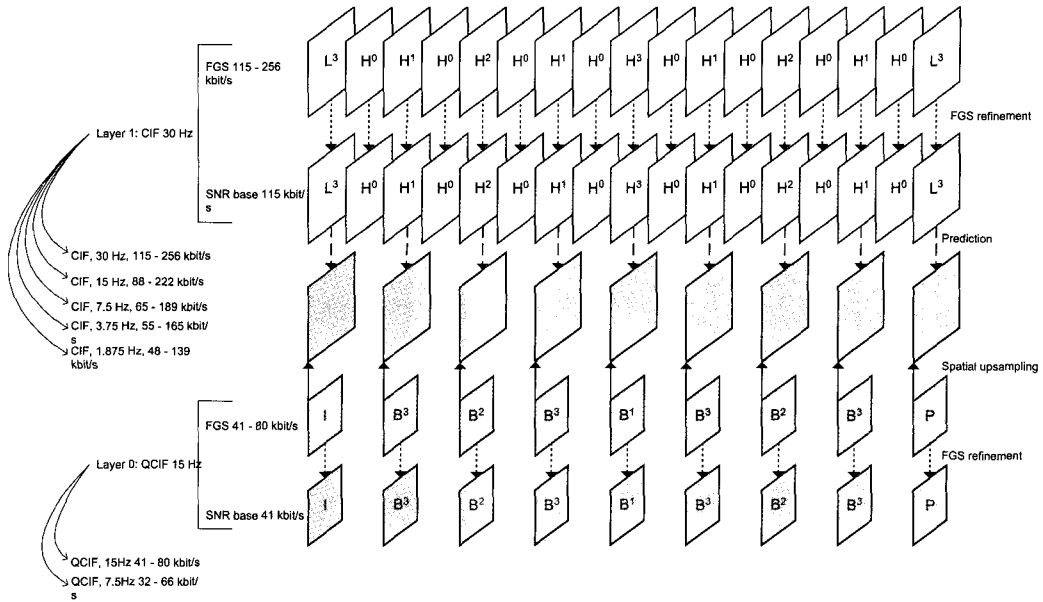
서 1개의 non-significant 비트가 나오기 까지만 부호화 한 후, 다음 블록을 부호화한다. 이를 영상의 끝까지 진행하는 것을 하나의 사이클이라고 한다. (그림 6)에서는 2개의 사이클을 보여 준다. 우선 첫 번째 사이클을 살펴보면, 첫 번째 블록에서는 첫 번째 위치에 non-significant 값이 존재하므로 1개의 값만 부호화 되며, 2번째 블록에서는 2개의 값이 부호화 되고, 3번째 블록에서는 9개의 값이 부호화 된다. 이 사이클을 모든 블록 내의 모든 값이 부호화 될 때까지 반복하게 된다.

4. 복합 스케일러빌리티 (Combined Scalability)

SVC에서는 앞 절에서 설명한 공간/시간/화질 스케일러빌리티는 서로 유기적으로 사용될 수 있다. 예를 들면 비트스트림 CIF@30fps로부터 비트스트림 QCIF@30fps와 비트스트림 CIF@15fps를 추출할 수 있다. (그림 7)은 복합 스케일러빌리티의 한 예를 나타낸다. 공간 해상도 기본 계층(QCIF)가 AVC를 사용하여 15Hz로 부호화 되었다. 이때, 시간적 스케일러빌리티를 제공하기 위해 Hierarchical B picture의 구조가 사용된다. 공간 상위 계층 (CIF)는

4개의 시간적 분할 단계를 가지는 MCTF(혹은 Hierarchical B picture)를 사용하여 30Hz로 부호화되었다. 각 공간 계층은 FGS를 사용하여 SNR 스케일러빌리티를 제공한다.

비트율 측면을 보면, QCIF 계층(계층 0)는 최대 80kbit/s로 부호화된다. SNR 스케일러빌리티를 이용하여 QCIF 15Hz 비트스트림은 41kbit/s에서 80kbit/s 사이의 비트율로 추출되어 전송될 수 있다. 더 낮은 프레임율을 제공하기 위해 B3 프레임들을 추출하지 않을 수 있다. QCIF 7.5 Hz 비트스트림은 SNR 스케일러빌리티와 함께 사용되어 32kbit/s에서 66kbit/s사이의 비트율로 추출되어 전송될 수 있다. 계층 0를 포함하는 CIF 계층(계층 1)는 최대 256kbit/s로 부호화된다. SNR 스케일러빌리티를 이용하여 CIF 30Hz 비트스트림은 115kbit/s에서 256kbit/s사이의 비트율로 추출되어 전송될 수 있다. 시간적 스케일러빌리티를 위해 {H} 0, {{H} 0, {H} 1}, {{H} 0, {H} 1, {H} 2}, 와 {{H} 0, {H} 1, {H} 2, {H} 3}를 제거하여, 15, 7.5, 3.75와 1.875Hz의 프레임율을 제공할 수 있다. 각 프레임율에 대해 SNR 스케일러빌리티를 사용하여 (그림 7)과 같은 비트율을 제공할 수 있다.



(그림 7) Combined Scalability의 예

IV. 결론

현재 ISO/IEC MPEG 과 ITU-T VCEG의 JVT 에서는 멀티미디어를 네트워크를 통해 전송할 때 QoS를 보장하기 위한 목적으로 스케일러빌리티를 제공하기 위한 Scalable Video Coding에 대한 표준화를 진행하고 있다. 따라서 본 고에서는 SVC를 소개하고, 이의 표준화 동향과 SVC 부호기의 기술적 구조, 공간적/시간적/화질적 스케일러빌리티 지원을 위한 대표적인 기술들에 대해 설명하였다.

결론적으로 SVC는 이중망 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 허용함으로써 여러 소비 환경 각각에 QoS를 보장하는 콘텐츠를 제공할 수 있다. 더욱이 동일한 생산 비용이 여

러 개의 전달/소비 환경에 공유될 수 있기 때문에 비용과 효율의 생산성 측면에서도 SVC는 매우 매력적이다. 따라서 스케일러블 비디오는 최근의 멀티미디어 서비스 환경에서 보다 향상된 비디오 서비스와 새로운 시장의 기회를 제공할 수 있을 것이다.

[참고 문헌]

- [1] ITU-T Recommendation H.261, "Video codec for audiovisual services at p x 64 kbit/s"
- [2] ITU-T, "Video Coding for Low Bitrate Communication,:" ITU-T Recommendation H.263, Version 1: Nov. 1995, Version 2: Jan. 1998.
- [3] 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.

- [4] ISO/IEC 14496-2, “ Information technology-Coding of audio-visual objects - part 2: Visual”, International Standard, second edition, December 2001.
- [5] 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.
- [6] ITU-T document, “Draft of Scalable Video Coding ? Working Draft 6”, Joint Video Team JVT-S201, JVT 19th meeting, Geneva, Switzerland, April, 2006
- [7] ITU-T Recommendation H.264 (2003), “Advanced Video Coding for generic audiovisual services” | ISO/IEC 14496-10:2003, “Information technology-Coding of audio-visual objects - part 10: Advanced video coding”
- [8] MPEG documents, “Registered Responses to the Call for Proposals on Scalable Video Coding,” ISO/IEC JTC1/SC29/WG11 MPEG2004/M10569.
- [9] ITU-T document, “Joint Scalable Video Model JSVM-6”, Joint Video Team JVT-S202, JVT 19th meeting, Geneva, Switzerland, April, 2006



최해철

1997년 경북대학교 전자공학과(학사)
 1999년 KAIST전기 및 전자공학과(석사)
 2004년 KAIST전기 및 전자공학과(박사)
 2004년 ~ 현재 ETRI 전파방송연구단 방송미디어 연구그룹 선임연구원

관심분야 : 영상통신, 영상분할, 비디오 부호화, Scalable Video Coding, MPEG Video, ITU-T VCEG



김재곤

1990년 경북대학교 전자공학과(학사)
 1992년 KAIST전기 및 전자공학과(석사)
 2005년 KAIST전기 및 전자공학과(박사)
 2001년 ~ 2002년 뉴욕 콜롬비아대학교 연구원
 1992년 ~ 현재 ETRI 디지털방송연구단 방송미디어 연구그룹 선임연구원/방송융합미디어연구팀장

관심분야 : 영상통신, 비디오신호처리, 디지털방송, 멀티미디어 프레임워크, TV-Anytime/MPEG-7/MPEG-21



홍진우

1982년 광운대학교 전자공학과 (공학사)
 1984년 광운대학교 전자공학과 (공학석사)
 1993년 광운대학교 전자공학과 (공학박사)
 1984년 ~ 현재 ETRI 방송미디어 연구그룹 책임연구원/그룹장

관심분야 : 오디오/음성 신호처리, MPEG-21 멀티미디어프레임워크, 방송서비스기술