

H.264/AVC-Scalable Extension의 표준화 연구동향과 알고리즘 분석

김원하(경희대학교 전자정보대학), 한우진(삼성전자 DM연구소)

1. 서론

현재의 통신환경은 유무선 연동, 방송망과 통신망의 융합, IP망을 이용한 IP Convergence 등의 서비스를 가능케 하는 핵심 네트워크인 광대역 통합망(BcN : Broadband Convergence Network)이 나타나는 등, 비균일 통신망이 서로 융합되고 있으며 앞으로도 이러한 추세는 더욱 가속화될 것이다. 이와 같은 비균일 통신 환경에서 디지털 융합(Digital Convergence)의 추세에 맞추어 가자면 동영상 전송도 범용 미디어 통신(Universal Media Access, UMA)이 되도록 해야 한다. 따라서 동영상의 부호화 방법은 압축효율을 최대화 시키는 것뿐 만 아니라 다양한 단말기 및 변화하는 통신 환경에 친화적으로 대응할 수 있는 Scalable Video Codec으로 발전해야 한다.

Scalable Video Coding(SVC)의 핵심 기능은 한 번의 압축된 비트스트림에서 서로 다른 여러 종류의 해상도, 화질, 프레임율을 갖는 영상을 다양한 디바이스와 다양한 네트워크 환경에 적응적으로 복원할 수 있도록 하는 것이다. 이러한 SVC 기술에 관해서는 기존의 많은 연구가 있어 왔다. 이미 MPEG-4에서는 FGS(fine grain

scalability)^[1] 기술이 SNR scalability 기술로 채택된 바 있으며, 그 밖에도 wavelet에 기반한 수많은 기법들이 SNR, spatial, temporal scalability를 지원하면서도 고효율 압축 성능을 유지할 수 있도록 제안되었다.

현재 H.264/AVC의 확장 형태로 표준화 진행되고 있는 H.264/AVC SE^[2]에서는 H.264/AVC에 이미 정의된 기술들과 추가적인 기술들을 결합함으로써 scalability 특징을 유지하면서도 높은 압축 효율을 달성하였다. H.264/AVC SE의 가장 큰 특징은 multi-layer 구조와 계층적 B-picture 구조 그리고 FGS 기술에 있다고 할 수 있다. 계층적 B-picture 기술은 temporal scalability를 지원하기 위한 시간적 분해 구조로 close-loop 구조를 기반으로 하며, temporal scalability에 따른 분해(decomposition) 레벨을 갖는다. H.264/AVC SE의 multi-layer structure 때문에 서로 다른 layer간 중복성을 제거하는 기술인 inter-layer prediction이 여러 연구기관에 의해 연구되고 있다. Inter-layer prediction은 크게 motion vector와 texture에 대한 두 부분으로 나누어지는데, motion vector의 경우 하위 layer의 motion vector로 부터의 예측, 혹은 이로부터

의 refinement 만으로 motion vector를 표현하는 방법이 주로 사용되며, texture의 경우 하위 layer의 복원된 영상, 혹은 복원된 잔여 신호를 이용하게 된다.

이외에도, H.264/AVC SE에는 ESS(extended spatial scalability)라고 불리는 임의 배율 spatial scalability 기술, cyclic scanning 기법이 적용된 FGS(fine grain scalability), R-D curve 특성 및 scalability 특성을 향상시키는 quality layer등 다양한 기술들이 채택되어 왔으며, 현재는 최종 bit-stream의 효율적 설계를 다루는 high-level syntax나 H.264/AVC의 기존 기법에 해당하는 CAVLC를 추가하는 일 등 좀 더 다양한 분야로 연구 분야가 확대되고 있다.

II. H.264/AVC-SE의 전반적인 구조

H.264/AVC 기반 스케일러블 확장 부호화의 전체 구조의 개념은 <그림 1>과 같다. H.264/AVC의 스케일러블 확장(SE, Scalable Extension) 부호화에서 공간 scalability와 단계적인(Coarse Grain) SNR scalability는 다중 계층(multi-layer) 구조로 구현하고, 시간적 scalability는 계층적(hierarchical) 영상 예측 방법으로 구현한다. 이와 더불어 미세한 화질의 scalability는 FGS기법인 점진적(progressive)으로 양자화 간격을 조절하여 부호화하는 방법으로 구현한다.

공간 scalability를 위한 다중 계층 구조에서는 원하는 해상도의 영상을 제공하기 위하여 각 계층이 다른 해상도의 영상들을 부호화한다. 각 공간 계층에 대해 다중 계층구조로 발생하는 계층간 중복 정보를 제거하기 위하여 이전 계층의 부호화된 신호를 복원하여 예측 신호로 사용한다. 또한 이전 계층의 공간 해상도가 현재 계층과 다

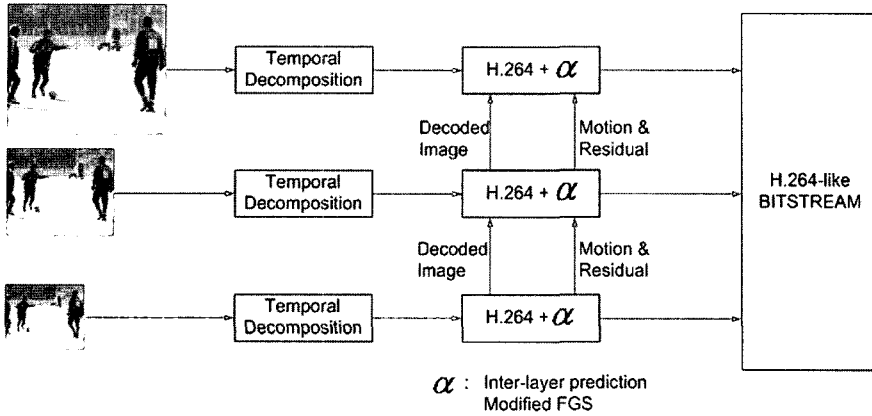
른 경우 보간법을 이용해 신호를 업샘플한다. 결과적으로, 현재 부호화 하고 있는 신호와 예측된 신호와의 중복된 정보를 매크로블록 단위로 제거한 잔여 신호를 부호화하는 기존의 부호화 방법을 따르면서 다 계층 구조에 적합한 계층 간 예측 방법이 추가된 셈이다. 계층 간 예측은 인트라 텍스처와 잔여신호, 움직임 정보에 대해 수행한다.

단계적(Coarse Grain) SNR scalability는 각 영상 신호 부호화 계층별로 원하는 화질이 발생되도록 양자화 파라미터 값을 설정하여 구현한다. 단계적 SNR 계층 사이에서의 미세한(Fine Grain) SNR은 점진적으로 양자화 간격을 변화시켜 원하는 비트율의 화질을 얻는다.

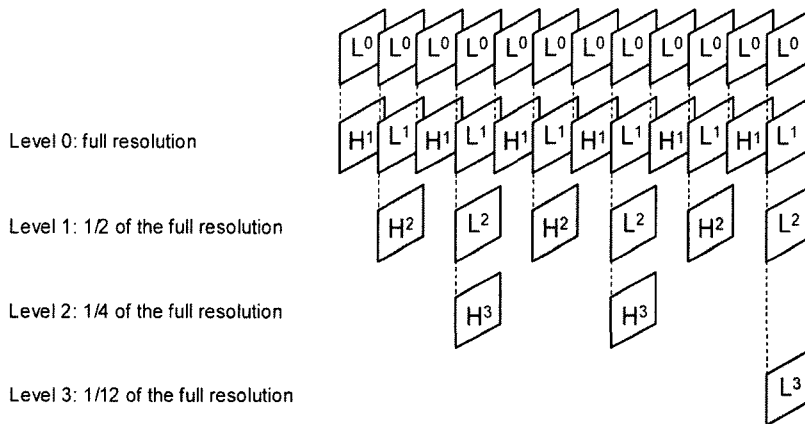
각 공간 계층에서 시간적 분해 구조를 사용하여 시간적 scalability를 구현한다. 시간적 분해 구조는 H.264/AVC에 명시된 움직임 추정 및 보상 기법을 이용하여 예측된 영상들 생성하는 계층적 B (hierarchical B picture, H-B) 구조가 사용된다.

III. 시간적 scalability (temporal scalability)

H.264/AVC SE의 시간적 scalability는 시간적 분해(decomposition)를 통해 이루어진다. <그림 2>는 12장의 picture를 시간적으로 분해하였을 때 레벨에 따른 시간적 해상도를 나타낸다. 분해 레벨에 따라 발생하는 L-frame들에 해당하는 picture들은 참조(reference) 영상이 되고 H-frame에 해당하는 picture들은 움직임 보상 예측과 움직임 보상을 수행하여 생성되는 예측(predictive) 영상들이 된다. <그림 2>에서 3레벨 시간적 분해 후 생성된 L3 프레임은 복호화 시 시간적으로 1/12의 해상도를 갖으며, H3 프레임



<그림 1> H.264/AVC 기반 스케일러블 확장 부호화의 전체 구조의 개념



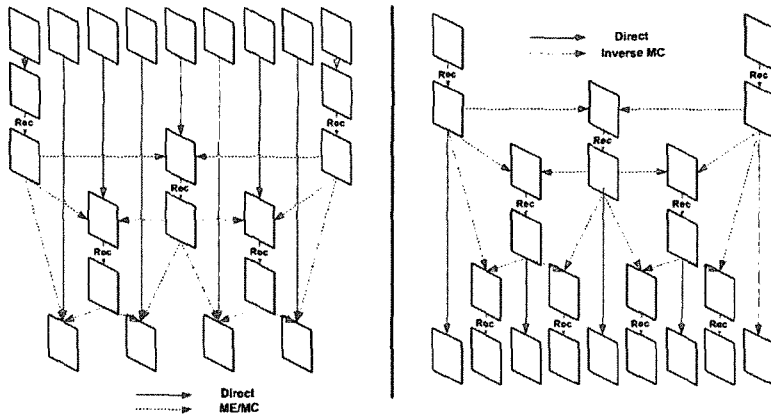
<그림 2> 시간적 scalability를 위한 분해 예

들이 추가로 복호화 되면 시간적으로 1/4 해상도를 갖는 영상 프레임들이 복호화 된다. L3 프레임과 H3, H2, H1 프레임들이 모두 복호화 되면 부호화된 모든 프레임들이 복호화 된다.^[3]

시간적 분해를 수행하기 위해서 H.264/AVC의 구조를 적용한 Hierarchical B (HB) 구조와 리프팅(lifting)기법에 근거한 Motion Compensated Temporal Filtering (MCTF)를 사용한다. II장에서 언급한 바와 같이 전체 연산량을 15% 증가시키는 이유로 2005년 10월 프랑스 니스에서 열린

제 74차 MPEG회의에서 MCTF 구조에서 update 단계를 제거하기로 결정함에 따라 복호화 측면에서는 기존의 MCTF도 Hierarchical B(HB) 구조와 동일한 구조가 되었다.

<그림 3>은 GOP 크기 8에서 HB 구조의 부호화 및 복호화 과정을 나타낸다. 음영 표시된 프레임이 실제 부호화 되는 프레임이다. <그림 4>는 실제 'FOREMAN' 영상에서 GOP 크기가 8일 때 HB 구조로 시간적 분해되는 실시 예를 나타낸다. 각 GOP 내의 영상들을 부호화/복호화



<그림 3> HB의 부호화 및 복호화 과정

하는데 있어서 기본 영상이 되는 picture를 key picture라 한다. key picture들은 이전 GOP의 key picture를 참조하여 부호화 하며 GOP내의 다른 부호화된 picture보다 우선적으로 전송한다. 복호화에서는 GOP 내의 영상들을 복호화하기 위해서 먼저 key picture를 복호화한다. <그림 4>에서 나타내듯이 Hierarchical B picture 구조는 예측 영상 생성 시 level 별로 참조 인덱스가 달라진다. level 1에서는 전 GOP의 key picture와 현재 GOP의 key picture를 참조하여 하나의 B picture를 생성한다. level 2에서는 level 1에서의 B picture를 복원한 영상과 key picture들을 참조하여 2개의 B picture를 생성한다.

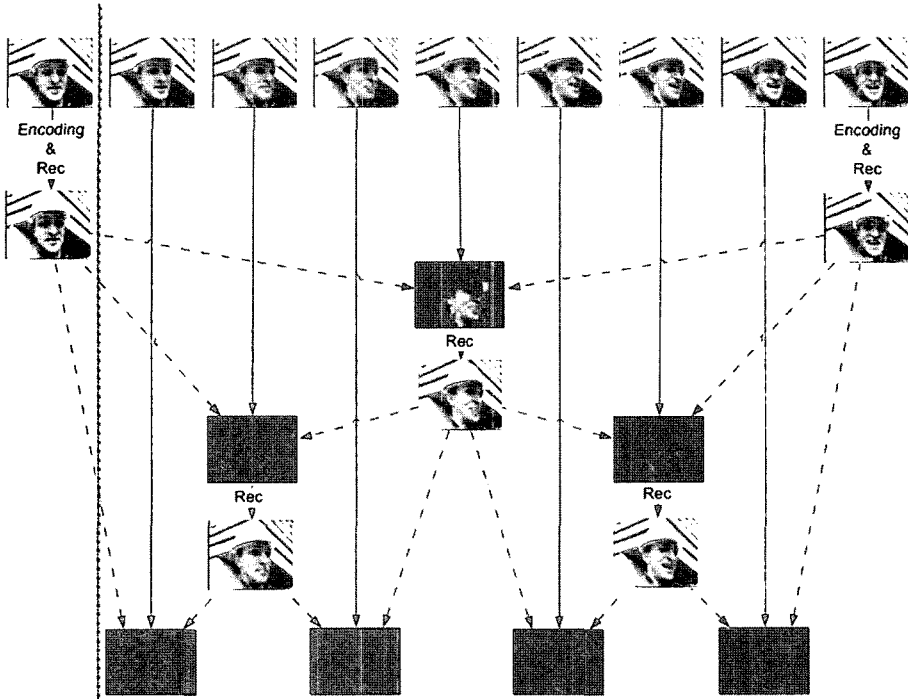
위의 HB 구조는 기존의 H.264/AVC 구조에서 사용되는 어떤 픽처라도 참조 픽처로 사용될 수 있도록 하는 메모리 관리 제어 연산(Memory Management Control Operation, MMCO)과 참조 영상에 저장되는 복원 picture 버퍼(Decoded Picture Buffer, DPB)에서 참조 picture의 참조 순서를 제어하는 RPLR (Reference Picture List Reordering) 명령에 의해 구현될 수 있다.^[9]

IV. 공간 scalability

공간 scalability를 위한 다중 계층 구조에서는 원하는 해상도의 영상을 제공하기 위하여 각 계층이 다른 해상도의 영상들을 부호화한다. 각 공간 계층에 대해 계층 간 중복된 정보를 제거하기 위하여 현재 부호화 하고 있는 계층의 공간 해상도보다 한 단계 낮은 공간 해상도의 영상신호를 현재 부호화 하고 있는 계층의 공간 해상도로 업샘플하여 예측 신호로 사용한다. 예측된 신호는 현재 부호화 하고 있는 영상신호와 매크로블록 단위로 차감되며 이때 발생한 잔여 신호를 부호화한다. 다음에서는 공간 scalability를 위한 다중 계층 구조의 계층 간 예측 방법에 대하여 분석한다.

1. 계층 간 예측(interlayer prediction)

각 계층은 계층별로 독립된 시간적 분해를 실시하고 부호화함으로써 중복된 정보를 포함하여 부호화 된다. 때문에 인트라 texture, 움직임 정보와 잔여신호등의 중복된 정보에 대해 계층 간 예측이라는 추가적인 예측 방법을 수행하여

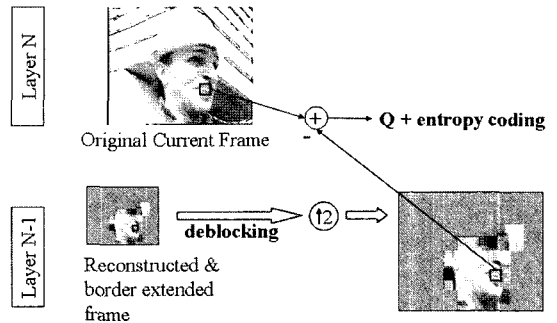


〈그림 4〉 GOP가 8일 때의 HB 구조로 시간적 분해되는 실제 예

다중 계층 구조가 갖는 부호화 효율의 저하를 막는다.

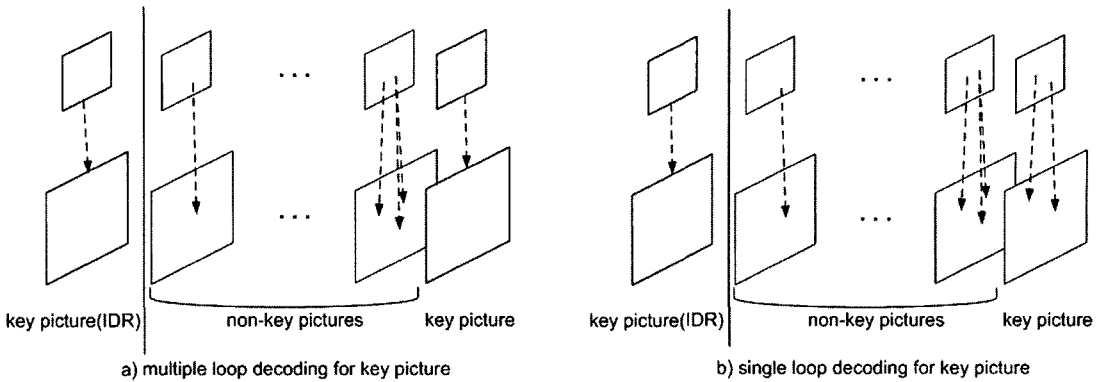
가) Inter-layer intra prediction(인트라 texture prediction)

부호화 하려는 매크로블록에 대응하는 하위 계층의 블록이 화면 내(Intra) 예측 모드로 부호화된 경우 하위 계층의 해당하는 매크로블록을 복원하고 복원된 블록을 부호화 하려는 매크로블록의 공간 해상도로 업샘플하여 예측 신호로 사용한다. 중복된 정보를 제거한 잔여 매크로블록 신호는 H.264/AVC 문법으로 부호화된다. 이러한 방법으로 부호화되는 매크로블록의 모드를 Intra Base 모드라 한다. <그림 5>는 Intra Base 모드의 부호화 과정을 도식화한다.^[4]



〈그림 5〉 Intra-BL 모드의 부호화 과정

하위 계층 픽처 내에 있는 인트라 매크로블록을 복원할 때는 매크로블록 내 혹은 인접 인트라 매크로블록 사이에서 발생하는 블록 간 blocking 효과를 제거하기 위하여 복원 후 디블록화 필터를 적용하며, 업샘플링 할 때 매크로블록 경계선에서의 transient 구간을 제거하기 위



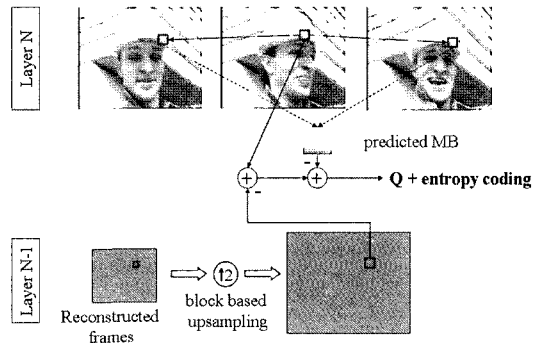
〈그림 6〉 key picture에 대한 multiple/single loop decoding 방식

하여 경계선을 각 방향 4-sample 확장한다¹⁾. 공간 해상도비가 2인 경우 업샘플은 H.264/AVC의 반 화소 업샘플 필터를 사용한다. 공간 해상도비가 2가 아닌 경우, 업샘플은 1/4 화소 H.264/AVC 보간법으로 수행된다²⁾.

복원하려는 picture가 key picture 경우 복호화 방식에는 single loop decoding과 multiple loop decoding이 있다. Single loop decoding은 하위 계층 key picture의 intra 매크로블록만 복원하여 예측하는 것이고 multiple loop decoding은 하위 계층의 key picture를 완전히 복원하여 복원된 key picture를 업샘플 하고 해당 위치에 있는 매크로블록을 예측 영상으로 사용하는 것이다. <그림 6>은 multiple/single loop decoding 방식을 나타낸다. 현재 실험 결과로 보면 multiple loop 방식이 좋은 부호화 효율을 내는 장점이 있으나 복잡도가 높은 단점도 있다¹²⁾. Single/multiple loop decoding 방식은 부호화 시 결정되며 하위 계층 key picture의 PPS(Picture Parameter Set)에서 constrained_intra_pred_flag로 제어한다.

나) inter-layer residual prediction

부호화 하려는 매크로블록에 대응하는 하위



〈그림 7〉 잔여 신호를 이용한 계층 간 예측 부호화 과정

계층의 블록이 화면 간 예측 모드로 부호화되어 잔여 신호를 포함한 경우에도 잔여 신호에 대한 계층 간 예측을 수행한다. 현재 매크로블록의 움직임 정보가 하위 계층의 대응하는 블록의 움직임 정보와 같거나 유사하다면 부호화된 하위 계층의 잔여 신호를 업샘플 하여 현재 블록의 예측 신호로 이용하였을 때도 계층 간 중복된 정보를 제거하여 부호화 효율을 높일 수 있기 때문이다. 그러나 현재 블록의 움직임 정보가 하위 계층의 해당 블록의 움직임 정보와 차이가 크다면 하위 계층의 블록을 부호화 할 때 참조한 하위 계층의 블록들과 현재 블록을 부호화하기 위해 참조하는 현재 계층의 블록들의 위치가 서로 다르다.

따라서 이 경우 계층 간 중복된 정보가 거의 없어서 계층 간 예측의 효과가 없다. 그러므로 잔여신호의 계층 간 예측은 움직임 정보에 따라서 적응적으로 수행한다⁴⁾.

<그림 7>은 현재 매크로블록의 움직임 정보가 하위 계층의 대응하는 블록의 움직임 정보와 같거나 유사할 때 잔여 신호를 이용한 계층 간 예측 과정을 도식화한다. JSVM 2.0 reference software에서는 잔여 신호의 계층 간 예측을 다음과 같이 수행 한다⁵⁾.

- ▷ 1 단계 : 하위 계층의 복원된 잔여 신호를 업샘플하여 부호화하려는 블록의 영상 신호를 차분한다.
- ▷ 2 단계 : 현재 부호화하려는 매크로블록에 하위 계층에서 업샘플된 움직임 정보로 현재 계층에서 움직임 보상하여 예측 매크로블록을 생성한다.
- ▷ 3 단계 : 1 단계에서 차분된 영상 신호에서 2 단계에서 생성한 예측 매크로블록 신호를 차분하여 차분된 신호를 부호화한다.

잔여 신호에 대한 업샘플은 앞서 언급한 인트라 texture의 경우와 같이 하위 계층과 상위 계층 간의 공간 해상도비에 따라 수행되며, 공간 해상도비가 2인 경우 업샘플 필터는 단순 쌍일차(bilinear) 필터를 사용한다. 공간 해상도비가 2가 아닌 경우, 업샘플은 1/4 화소 H.264/AVC 보간 방법을 이용한다. 잔여 신호의 블록 경계는 경계 값으로 확장되며, 업샘플 필터는 변환 블록 사이의 경계에 대해서는 적용되지 않는다⁶⁾.

다) inter-layer motion prediction

공간 해상도가 다른 계층 간에는 하위 계층이 존재 하면 현재 계층에서 움직임 추정을 사용하지 않고 하위 계층에서 획득한 움직임 정보를 업

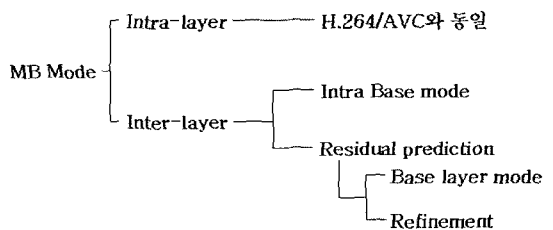
샘플하여 사용하는 base layer 모드가 있고 하위 계층이 존재여부와 관계없이 현재 계층에서의 H.264/AVC 인터 모드로 움직임 정보를 사용하는 모드가 있다. 또한 base layer 모드로 얻은 움직임 벡터에서 모든 방향으로 1/4 화소 단위로 움직임 예측을 통하여 refinement를 부호화하는 1/4 화소 refinement 모드가 있다.

움직임 정보에 대해 하위 계층이 존재하지 않으면 H.264/AVC 인터 모드를 적용하고 하위 계층이 존재하면 H.264/AVC 인터 모드, base layer 모드, 1/4 화소 refinement 모드 중에서 선택적으로 부호화한다.

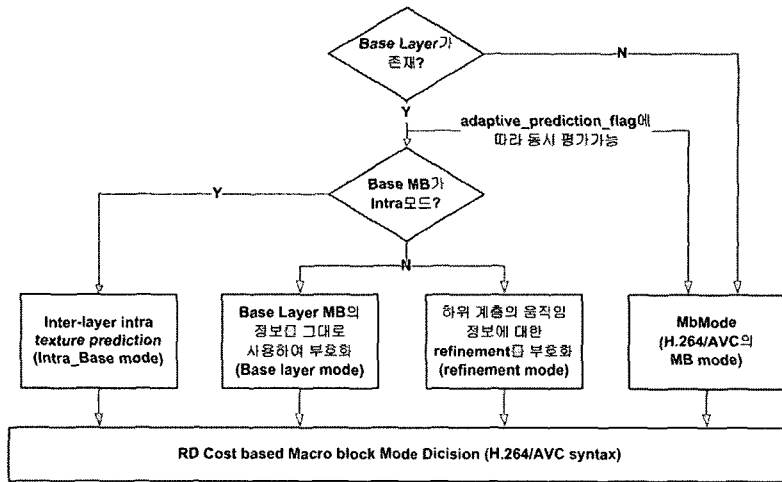
V. H.264/AVC-SE의 매크로블록 모드

매크로블록 모드를 정리하면 다음과 같다.

- ▷ Intra base 모드 : 하위 계층의 매크로블록 모드가 화면 내 예측 모드인 경우 해당 매크로블록을 공간 해상도 비율로 업샘플하여 현재 매크로블록의 예측 값으로 사용.
- ▷ Base layer 모드 : 하위 계층의 움직임 정보와 예측 정보를 그대로 사용하여 부호화. 현재 working draft^[11]에서는 residual_prediction_flag로 표시한다.
- ▷ Refinement 모드 : 공간 해상도 비율로 업샘플된 하위 계층의 움직임 정보를 이용하



(그림 8) H.264/AVC-SE의 매크로블록 모드



<그림 9> H.264/AVC SE의 매크로블록 모드의 결정 과정

여 예측하며 1/4 화소 단위로 refinement 추정하여 추가로 부호화.

▷ Intra layer : Intra, Inter, PCM, Skip 모드 등 H.264/AVC의 모든 MB 모드.

<그림 8>은 H.264/AVC-SE의 매크로블록 모드를 정리하여 도식화한 것이다.

<그림 9>와 같이 H.264/AVC SE의 매크로블록 모드는 하위 계층의 존재 여부에 따라 일차적으로 나뉜다. 하위 계층이 존재 할 경우 하위 계층의 매크로블록이 화면 내 예측 모드인지 아닌지 판별하고 하위 계층의 매크로블록 모드가 화면 내 예측 모드인 경우 Intra_Base 모드에 대한 비트율 왜곡 비용(rate distortion cost)을 평가한다. 하위 계층의 매크로블록이 화면 내 예측 모드가 아닌 경우 하위 계층의 움직임에 대한 예측 모드로 부호화 되며 이 모드는 base layer 모드와 1/4 화소 단위 refinement 모드로 나뉘어 각 모드에 대한 비트율 왜곡비용을 평가한다. 적응적 예측(adaptive prediction)이 수행될 경우 모든 모드 결정 과정에 대하여 Inter/Intra 모드에 대한 비용 평가를 실시한다. 하위 계층이 존재하지 않

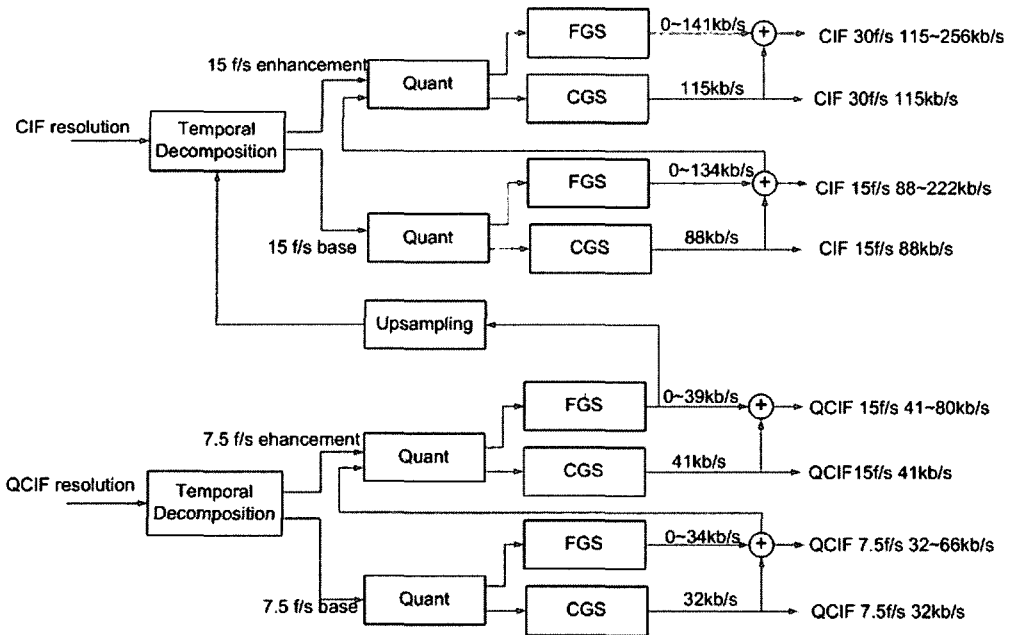
는 경우 매크로블록 모드는 H.264/AVC의 매크로블록 모드와 같다.

VI. SNR scalability

단계적(Coarse Grain) SNR scalability는 각 영상 신호 부호화 계층별로 원하는 화질이 발생되도록 양자화 파라미터 QP 값을 설정하여 비트율의 화질을 구현한다. CGS 계층은 부호화시 정의된 계층이며 공간적 시간적으로 정의된 각 계층에 대한 비트율 계층을 말한다. CGS 계층 사이에서 미세(Fine Grain) SNR의 조절은 CGS의 비트율에서 발생하는 양자화 오류를 점진적으로 보정해서 원하는 비트율의 화질을 얻는 FGS기법을 사용한다. 이에 따라 디코더에서는 모든 비트율에 대한 비트열을 복호 할 수 있다.

VII. Combined Scalability

<그림 10>은 Temporal, Spatial, SNR scalability를 결합한 4계층으로 설계된 SVC 부호화 구조



〈그림 10〉 4계층 부호화구조의 SVC

의 예이다. 각 계층마다 FGS 계층이 추가되어 미세 SNR에 대한 scalability를 구현한다. 그림과 같은 구조로는 하나의 부호화된 bit stream에서 영상의 해상도를 CIF와 QCIF로 조절할 수 있다. 시간적 scalability는 CIF 해상도에서 최대 30fps까지 2의 배수로 조절할 수 있고 QCIF의 해상도에서는 최대 15fps 2의 배수로 조절할 수 있다. 비트율은 CIF 해상도에서 256kbps내에서 임의로 조절할 수 있고 QCIF 해상도에서는 80kbps 내에서 임의로 조절할 수 있다.

Ⅷ. 결론

Scalable Video Coding은 단 하나의 동영상 압축 비트열로 모든 네트워크, 모든 기기들에 적응적으로 동작할 수 있는 부호화 방법인 스케일러블 비디오 부호화는 앞으로 방송, 감시 시스템

등 많은 멀티미디어 관련 분야에서 활용 가치가 높은 것으로 예상되고 있다. 이에 본 논문에서는 표준화 진행 중인 H.264/AVC기반 SVC의 연구 방향 및 표준화 진행 현황을 소개하고 그 기술을 분석하였다. SVC의 표준화는 2007년 1월 완성을 목표로 진행 중임으로 지속적인 기술의 분석과 개발이 필요로 되고 있다^[10].

참고문헌

- [1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 301-317, Mar. 2001.
- [2] ISO/IEC JTC1/SC29/WG11, "Report on call for evidence on scalable video coding (SVC) technology," in 65th MPEG meeting, no. M5701, Trondheim, Norway, July 2003.
- [3] J. Reichel, H. Schwarz, M. Wien, "Joint Scalable

Video Model JSVM-3”, Join Video Team of ISO/IEC MPEG&ITU-T VCEG, Doc.JVT-P202, Poznan, Poland, July. 2005.

- [4] “JSVM 2 software”, Join Video Team of ISO/IEC MPEG&ITU-T VCEG, JVT-O055, Busan, Korea, April 2005.
- [5] Julien Reichel, Heiko Schwarz, Mathias Wien Scalable, “Video Coding - Working Draft 3”, Join Video Team of ISO/IEC MPEG&ITU-T VCEG, Doc.JVT-P201, Poznan, Poland, July. 2005.
- [6] H. Schwarz, T. Hinz, D Marpe, T Wiegand, “Constrained Inter- Layer Prediction for Single-Loop Decoding in Spatial Scalability”, Fraunhofer Institute for Telecommunications- Heinrich Hertz Institute, Berlin, Germany.
- [7] Y. Bao, M.Karczewicz, J. Gidge, X. Wang, “FGS block enhancements for scalable video coding”, ISO/IEC JTC1/SC29/ WG11, MPEG/M-11428, Palma, Spain, Oct. 2004
- [8] Y. Bao, M.Karczewicz, J. Gidge, X. Wang, “Cyclical block coding for FGS”, ISO/IEC JTC1/SC29/ WG11, MPEG/M-11509, HongKong, China, Jan. 2005
- [9] Y. Bao, M.Karczewicz, J. Gidge, X. Wang, “FGS subband enhancements for scalable video coding”, ISO/IEC JTC1/SC29/ WG11, MPEG/M-11427, Palma, Spain, Oct. 2004
- [10] <http://www.chiariglione.org/mpeg/workplan.htm>

저자소개



김원하

1985년 연세대학교 전자공학과(공학사)
 1988년 Univ. of Wisconsin-Madison, 전기공학(석사)
 1996년 Motorola 인턴쉽
 1997년 Univ. of Wisconsin-Madison, 전기공학(박사)
 1997년-2000년 Los Alamos National Lab. 연구원
 2000년-2003년 명지대학교 정보통신공학과 조교수
 2003년-현 재 경희대학교 전자정보대학 부교수
 주관심분야 멀티미디어 신호처리, 영상처리



한우진

1995년 KAIST 전산학과(공학사)
 1997년 KAIST 전산학과(석사)
 2000년 KAIST 전산학과(박사)
 2002년-2003년 SL2 연구소장
 2003년-현 재 삼성전자 DM연구소 책임연구원
 주관심분야 멀티미디어신호처리, 영상압축, audio/speech 압축 기술