

MPEG-II 영상 및 음성 압축 알고리즘

윤 대 희/연세대학교 공과대학 전자공학과 교수

최 윤 식/연세대학교 공과대학 전기공학과 교수

□ 차 례 □

I. 서 론

II. MPEG-II 영상 압축 부호화

III. MPEG II 오디오 부호화

IV. 결 론

I. 서 론

디지털 영상 및 음성이 통신, 컴퓨터, 가전등의 여러 분야로 이용 범위가 확대되면서, 상호간의 호환성 문제가 등장하게 되었다. 이러한 호환성 문제를 해결하기 위해서 ISO(International Organization for Standardization)에서는 CD(Compact Disk) ROM 수준의 Mbit/s 급의 디지털 저장 매체(digital storage media, DSM)용 영상 및 관련 음성 부호화 표준안(MPEG-I) 개발을 시작하였다. 그러한 표준화의 일환으로 세계 표준화 기구 산하에 MPEG(Motion Picture Experts Group)이 탄생하게 되었으며, 1988년 부터 MPEG-I 표준안을 만들기 위해서 본격적인 활동을 시작하였다. MPEG 은 현재 ISO-IEC/JTCl/SC29/WG11에 속하여 있다. MPEG-I 은 이미 1991년 11월에 Committee Draft 가 완성되어 표준화가 완료되었고, 현재는 3Mbit/s 급으로 부터 10Mbit/s 이상의 HDTV 급까지의 해상도를 포함하는 MPEG-II의 표준화가 93년 11월 서울회의에서 Committee Draft가 확

정되어 완료단계에 와 있는 실정이다. 따라서 본 논문에서는 MPEG-II의 영상 및 오디오 압축 기법에 관하여 기술하되, 가급적 MPEG-I 과의 차이점에 중점을 두어 기술하고자 한다.

II. MPEG-II 영상 압축 부호화

MPEG 영상 압축 부호화 기법은, 여러가지 기록 매체에 공통적으로 적용되어질 수 있어야 하기 때문에 일반적 부호화 기법(Generic Coding)이다. 따라서 부호화기 자체의 설계나 구현방법을 명시하기 보다는, 설계자의 의도와 적용 분야에 따라, 부호화기의 구성, 성능, 기능을 달리 설계할 수 있다. 즉, MPEG-II 의 규격은 여러 형태의 영상 규격을 포용하도록 아래와 같이 정의되어 있다.

2.1. 입출력 영상 규격

MPEG-II의 표준 기법은 격행주사(Interlaced Scan)의 CCIR601 (4:2:2, 4:2:0, 4:4:4 규격)

영상을 비롯하여, EDTV, HDTV, 그리고 필름영상 같은 Progressive Scan 영상등을 지원하도록 한다.

2.2. 해상도 규격

Studio 규격인 CCIR601 을 입력으로 했을때, 3-5 Mbit/s 에서는 VHS 비디오 급의 화질을, 8-10 Mbit/s 에서는 CCIR601에 근접한 화질을, 10 Mbit/s 이상에서는 HDTV 급의 화질을 제공할 수 있어야 한다.

2.3. Scalability와 Compatibility

MPEG-I 에 비하여, 두드러지는 MPEG-II 의 특성 중의 하나로, 다른 해상도를 가지는 응용 분야 간의 변환을 고려한 성능인데, 즉, Encoded 된

임의의 해상도의 Bit Stream을 선택적으로 취하여 다른 해상도의 영상을 얻을 수 있도록 한 Resolution Scalability나, 혹은 시간적으로 취하여 다른 시간해상도를 얻는 Temporal Scalability 가 그러한 기본적 Scalability의 정의이다. MPEG-II 에서는, 이러한 Scalability를 구현하기 위하여, 크게 Frequency Domain Scalability 구조와 Spatial Scalability 구조를 제안하고 있다.

2.3.1. Frequency Domain Scalability

Frequency Domain Scalability는 Frequency Domain(DCT 계수)에서의 Scalability 구조를 삽입하여 다양한 해상도의 영상을 얻고자 하는 것이다. 따라서, 이 구조는 DCT 계수를 한개나 두개의 채널에 나누어, 각 채널의 결합에 의해 낮거나 높은

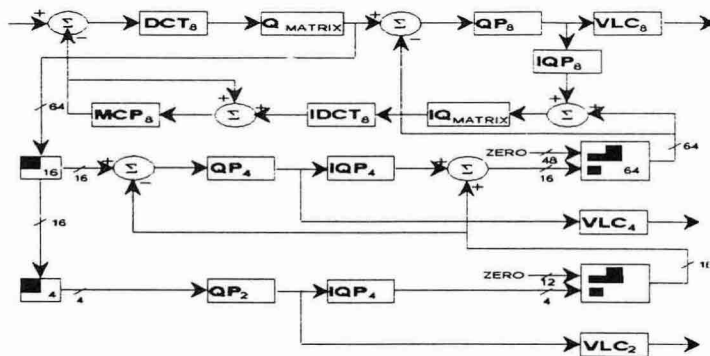


그림 1. Frequency Domain Encoder

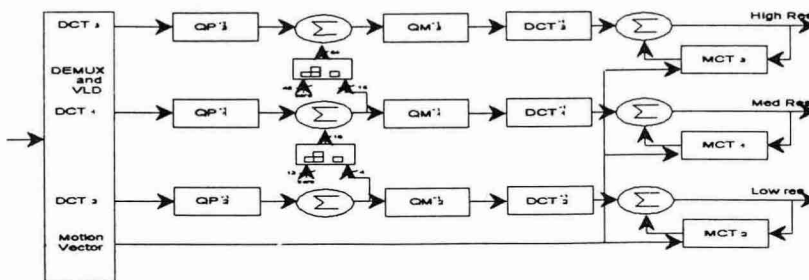


그림 2. Frequency Domain Scalable Decoder

해상도의 영상을 표현하고자 하는 것이다. 즉, 이러한 구조는 DCT 계수를 계층적 피라미드나 Subband 구조에 의해 구현할 수 있는 여지를 준다. 표현을 위한 영상 채널 layer 숫자는 임의로 정할 수 있으나, 대개, 아래의 3 개의 해상도를 구현하기 위한, 3-layer 구조가 보편이다. (그림 1, 2)

1. CCIR601 704 × 480 해상도
2. SIF 352 × 240 해상도
3. QSIF 176 × 120 해상도

2.3.2. 호환성(Compatibility)과 Spatial Scalability

호환성이란, 각각 다른 규격을 갖는 표준화 기법들을 사용하여 영상신호를 부호화하고 복호화할 수 있을때 두 표준기 간의 관계로 정의될 수 있고, 이러한 호환성을 보장하는 Scalability 방법은 화소단위나 시간단위에서의 Spatial Scalability에 의하여서만 가능하다. 이러한 Spatial Scaling/compatible Coding은 다층(Multi-layer) 구조에 의하여 이루어 지는데, 각 층에 적용되는 Coding 방식이 서로 종속적으로 이루어 지는, 소위 loose coupling 방식으로 각 층이 연계되어 있다. 그림 3은 이러한 구조를 잘 나타내 주고 있다. 또 다른 개념의 Scalability는 SNR Scalability라는 방식인데, 이 것은 같은 해상도를 갖는 영상층 간의 호환성을 가지도록 고안된 방식이다. 따라서, SNR Scalability는 두개의 다른 화질을 가진 매체를 전송하기에 적합한 구조를 가진다고 말할 수 있는

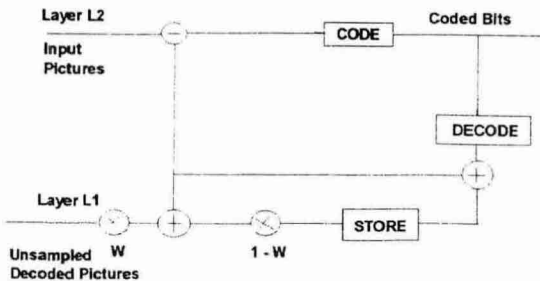


그림 3. 층간의 loose coupling 구조

데, 그 구조는 두 층간의 비트율 차이가 상위 층에서의 더 좋은 Coding 결과를 제공하도록 (다시 말하면, 높은 SNR 값을 제공하도록) 설계되었다. 이 방식에서 하위 층은 MPEG-II의 기본 형에 의해 Coding 되고, 상위 층은 SNR Scalable 조건 하에서 Coding된다. Spatial Scalability 구조 하에서의 SNR Scalability의 구현은, 양자화기의 개선에 의해 가능해 지는데, 이의 구조는 그림4, 5 와 같다.

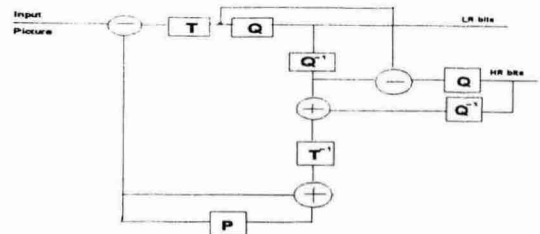


그림 4. 양자화기 개선 및 Feedback 구조를 가진 SNR Scalable Encoder

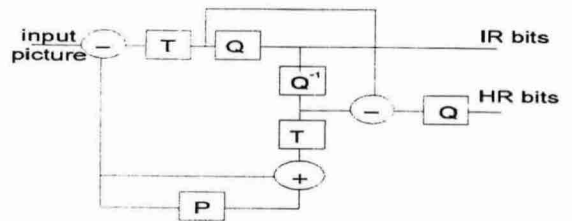


그림 5. 양자화기 개선 구조만 가진 SNR Scalable Encoder

2.4. 압축 부호화 기법

MPEG-II의 부호화 기법은, MPEG-I 과 마찬가지로 다음과 같은 기본적인 부호화에 의해 이루어 진다. 즉, 시간적 중복성(Temporal Redundancy)은 움직임 추정/보상에 근거를 둔 DPCM으로 제거하고, 공간적 중복성(Spatial Redundancy)은 DCT와 같은 변환부호화와 양자화에 의하여, 또한 통계적 중복성은 Entropy Coding에 의하여 제거하는 것이 그것이다. 그러나 이 중에서 특히 시간적 중복성 제거를 위한 움직임 추정 및

보상 방법에 있어서, MPEG-II 는 MPEG-I에 비하여 훨씬 복잡한 구조를 요구하는데, 이는 MPEG-II의 경우 Studio 구격인 CCIR601 수준의 화질을 요구함으로 보다 정확한 움직임 벡터의 추정과 보상이 요구되기 때문이다. 따라서 이 장에서는, MPEG-II 에서 특별히 다루는 움직임 추정 방법에 대하여만 기술하고자 한다.

2.4.1. 움직임 추정 방식

MPEG-II 에서 여러 실험을 거쳐, 효율면에서 최종적으로 고려하고자 하는 움직임 추정 방식은 두가지로 압축되었는데, 개선된 SVMC(Modified Single Vector Motion Compensation)-Prime 방식과 Dual Prime 방식이 그것이다.

• 개선된 SVMC-Prime 방식

개선된 SVMC-Prime 방식은 Macro-block 당 하나의 움직임 벡터를 사용하기 때문에 데이터 양을 많이 줄일 수 있고, Field-based 추정방식 중, 특히 움직임이 큰 경우에 매우 효과적인 방식으로 알려져 있다. 다만, 추정은 두가지의 기본 Type에 의해 구하여 지므로 Dual Prime에 비하여 상대적으로 부가 계산을 필요로 한다. 이들 두가지 type은 Near field 와 Same-parity field type 인데 이들 type 들은 모두 동방향 추정에만 사용되어진다. 이들 두가지 type의 구체적 추정방법은 그림 6과 같다.

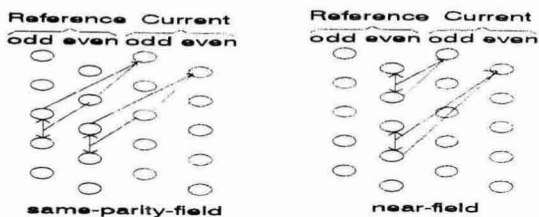


그림 6. 개선된 SVMC-Prime 추정방식
(a) Near-field type (b) Same-parity field type

• Dual Prime 방식

Dual Prime 방식은 Dual Field 추정 방식을 개선하여 움직임 벡터의 비트수를 줄인 방식이다. 즉 매크로블록 당 하나의 필드 움직임 벡터와 그에 따른 차이 벡터만을 전송한다. 이러한 Dual Prime 추정방식은 수직 방향의 해상도를 향상시킬 뿐 아니라 시,공간적 필터의 역할도 하게 된다. 구체적 추정 방법은 그림 7과 같다.

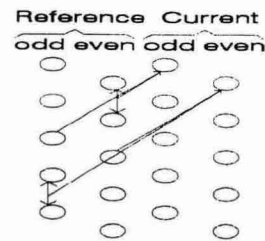


그림 7. Dual Prime 추정 방식

2.4.2. 기타 부호화 기법

변환 부호화와 양자화로 대변되는 공간적 중복성(Spatial Redundancy) 제거나, Huffman Coding에 의한 통계적 중복성(Statistical Redundancy)의 제거 방식은, 원칙적으로 MPEG-I 과 대동소이 하고, Scene Change 등의 변환이 있을 때의 Buffer 상태의 효율적 제어를 위해 제시된, 저지연(Low-delay) 기법등이 제시되었으나, 본 논문에서는 지면상 자세한 기술을 생략하고 상기 기술한 중요 차이점만을 기술하였다.

III. MPEG II 오디오 부호화

MPEG-I 오디오 부호화 방식은 표준안으로 약 128 kbit/s에서 CD수준의 음질을 얻을 수 있는 MUSICAM(Masking-pattern adapted Universal Sub-band Integrated Coding And Multiplexing) 방식과 더욱 낮은 비트율에서 응용될 수 있는 ASPEC(Adaptive Spectral Perceptual Entrophy Coding) 방식

을 채택하였다[3][5]. MPEG-I 표준안은 비트율에 따라 세가지 계층으로 나누어 지는데 계층 1, 2는 MUSICAM 방식을 기본으로 하는 서브밴드 부호화 방식으로 128~192 kbit/s의 비트율을 갖는다. 계층 3은 ASPEC 방식과 MUSICAM 방식을 결합한 형태로 96 kbit/s 이하의 응용 분야에 사용될 수 있다.

MUSICAM 방식은 청각 특성을 이용한 서브밴드 부호화 방식으로서 각 서브밴드에서 지각적인 잡음을 최소화하도록 음을 부호화하여 주관적으로 원음과 동일한 복원음을 얻을 수 있는 방식이다. MUSICAM 시스템의 기본 구조는 그림 8과 같다. 부호화 단에서는 입력된 오디오 신호를 서브밴드 필터링한 후, APCM을 사용하여 통계적인 중복성(redundancy)을 제거한다. 동시에 FFT를 사용한 심리음향 모델의 적용으로 지각적인 중복성을 제거하여 부호화 효율을 높인다. 마스킹 곡선과 입력 신호 레벨로부터 얻어진 각 서브밴드의 비트 할당 정보를 이용하여 압축된 비트열(bitstream)을 만든다.

입력 신호는 32개의 가중 중첩 가산(weighted overlap-add) 방식 등간격 필터뱅크를 통과하여 서브밴드 샘플로 바뀌어진다. 이 때, 심리음향 모델에서는 부가적인 FFT를 이용하여 마스킹 임계값을 얻어 양자화에 쓰이는 비트 할당 정보를 제

공한다. 즉, 필터 뱅크의 출력값과 마스킹 임계값을 가지고 신호대마스킹 비(SMR : Signal to Mask Ratio)를 구하여 주관적으로 양자화 잡음이 신호에 의해 마스킹될 수 있도록 비트 할당을 한다. 잡음을 완전히 마스킹할 수 없을 때는 주관적 잡음을 최소화하도록 각 서브밴드에 비트를 할당한다. 양자화된 서브밴드 샘플과 부가 정보를 가지고 비트열을 만든다.

계층 1에서는 각각의 서브밴드 신호 12샘플씩을 한 프레임으로 묶어 총 384샘플에 대해서 비트 할당을 한다. 계층 2에서는 좀더 복잡해진 부호화기로서 더 큰 압축율을 얻을 수 있다. 샘플과 부가 정보를 부호화할 때 사용 비트를 줄이기 위하여 여러가지 방법들이 사용된다. 1152샘플(32 서브밴드*36샘플), 즉 세 프레임을 한 단위로 비트 할당을 하며 약 128 kbit/s의 비트율을 사용하고자 할 때 적용된다. 계층 3은 서브밴드 샘플을 MDCT하여 더욱 세밀한 주파수 해석을 하게 된다. 또한, 허프만 부호화를 사용하여 압축율을 향상시키며 프리에코(pre-echo)에 대한 제어부가 첨가된다.

복호화기에서는 압축된 비트열을 풀어 각 서브밴드 샘플들을 복원시키고 합성 필터를 통과하여 부호화된 신호의 PCM 샘플(복원 신호)을 얻는다. 복호화기는 부호화기에 비해 간단하며 심리음향에 관한 정보가 필요하지 않다. 또한 각 계층의 복호화기는 이전 계층의 비트열을 복호화할 수 있어야 한다.

MPEG-II 오디오 부호화 방식은 방송용 등의 고품질 오디오를 위해 다채널 특성을 제공한다. MPEG-I 오디오 부호화 방식과 MPEG-II 오디오 부호화 방식의 차이점은 MPEG-I 오디오 부호화 방식의 경우 최대 두 채널(좌, 우)의 스테레오 오디오 신호만을 부호화 대상으로 하는 반면, MPEG-II 오디오 부호화 방식의 경우는 현장감을 필요로 하는 방송용 오디오를 위해 두채널 이상의 다채널 오디오 신호를 부호화한다. MPEG-II 오디오

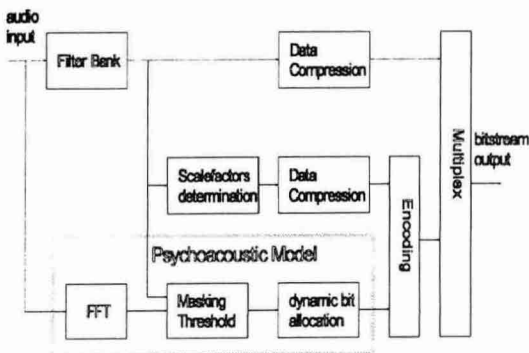


그림 8. MUSICAM 오디오 부호화기의 구조

오의 각 채널을 부호화하는 압축 방식에 있어서 는 기본적으로 MPEG-1의 계층 2, 3 오디오 압축 방식이 이용되며 부가적으로 각 채널의 특성을 이용한 부호화 방식이 사용될 수 있다. 그림 2는 다채널 오디오 부호화기의 기본 구조이다.

기존의 두채널 오디오 신호는 음상 정위(Sound Image Localization)가 불안정하여 현장감을 충실히 재현하지 못하였다. 따라서 MPEG-II에서는 이러한 현장감을 살리기 위해서 좌, 우 두채널의 기본 스테레오 신호에 중앙 채널, 서라운드 채널 및 저주파 효과 채널을 첨가하는 SMPTE의 5.1채널(L, R, C, Ls, Rs, LFE)의 부호화 알고리즘을 표준화하였다. 따라서 영화 사운드와의 호환성을 제공하여 가정에서도 대형 화면에서의 음장감을 느낄 수 있게 된다. 또한 이들 채널 외에도 방송용 오디오가 제공해야 할 음성 다중 채널, 부가적인 오디오 채널에 대해서도 표준화 작업이 진행 중이다.

IV. 결 론

디지털 저장 매체(DSM)에 효율적으로 데이터를 저장시키고자 하는 노력에서 비디오 및 오디오 분야에서 MPEG 표준화 활동이 시작되었다. 그러나 최근에 와서는 그 효용 범위가 점차 넓어져 방송 및 통신 분야에까지 확대되고 있으며, 특히 MPEG II 방식은 방송 및 컴퓨터 업계에까지 널리 확장 보급될 것으로 보인다.

현재까지는 국내 연구기관 및 기업체가 초기 단계에서부터 MPEG 방식에 참여하지 못하여 실질적인 방식 제안을 하지 못하였으나 앞으로는 MPEG IV 및 다른 표준화 연구에 국내 연구기관들이 초기에서부터 참여하여 그들 연구기관들의 방식이 표준화될 수 있도록 노력하여야겠다.

참 고 문 헌

1. ISO/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s-CD 11172"
2. ISO/IEC JTC1/SC29/WG11 No.603 "Generic Coding of Moving Pictures and Associated Audio-CD 13818"
3. 남재열, 호요성, "영상신호 압축을 위한 MPEG 표준화 동향" 전자공학회지 Vol.20, No.10, pp. 1068-1081, 1993년 10월.

筆者紹介



▲윤 대 희

- 1982년 : 연세대학교 전자공학과 공학박사(신호처리)
- 현재 : 연세대학교 전자공학과 교수



▲최 윤 식

- 1990년 12월 : 미국 Purdue Univ. 공학박사(영상처리)
- 현재 : 연세대학교 전기공학과 교수