

# MPEG-4 Fine Granular Scalability 기술을 이용한 차세대 비디오 스트리밍 시스템

서광덕\* · 권순각\*\*

## 1. 서론

MPEG-4 스트리밍을 인터넷이나 차세대 이동통신 분야에 적용하고자 하는 시도가 매우 활발하게 이루어지고 있다. 현재 스트리밍 기술로는 Apple의 Quicktime을 비롯한 몇몇 기술들이 있지만, RealNetworks의 RealMedia 기술과 Microsoft의 Windows Media Technology (WMT)가 주로 사용되고 있다. 인터넷은 그 탄생 배경 자체에서 알 수 있듯이 데이터 전송을 위해 만들어진 것이기 때문에, 멀티미디어 콘텐츠를 전송하기에 적합한 특성을 갖지 못하고 있다. 시간적인 연속성을 가지는 멀티미디어 데이터를 안정적으로 전송할만한 대역 폭을 보장하지 못하며 수시로 가용 대역 폭이 변화한다[3,4].

따라서 모든 스트리밍 기술은 플레이어에서 일정 시간 초기 버퍼링을 통하여 이를 보상하려고 하지만 네트워크가 정체되는 시간이 길어지면 버퍼의 데이터가 소진되고 결국에는 재생에 필요한 데이터를 받기까지 멈추게 되는 현상이 발생한다.

초기 버퍼링에 소요되는 시간과 재생 중에 일어나는 멈춤 현상을 최대한 줄이기 위해 RealMedia와 WMT에서는 콘텐츠를 인코딩할 때 여러 종류의 비트율로 부호화하는 기술을 채택

하였다. 실제로 RealMedia에서는 SureStream 이라는 기술을 그리고 WMT에서는 다중비트율 옵션을 적용하고 있다. 이들 서비스의 공통점은 네트워크 정체가 일어나는 경우에, 그에 적합하도록 낮은 비트율의 비트스트림을 전송하고 자연스럽게 화질을 낮춤으로써 끊어짐 없는 멀티미디어 데이터의 재생을 가능케 하는 것이다. 하지만 이와 같이 인코딩 과정에서 서로 다른 비트율을 가지는 복수개의 비트스트림을 한 번에 만들어 내기 위해서는, 고급 사양의 시스템이 필요하고 인코딩 시간이 늘어나는 등 전체적인 콘텐츠 제작 비용이 크게 늘어나게 된다. 또한 사용하는 비트율의 개수에 따라 저장해야 할 데이터의 양이 늘어나고 이는 저장 매체에 소요되는 비용을 증가시킨다.

이에 반해 MPEG-4의 확장성 (scalability) 부호화 기법을 이용하면 몇 종류의 스케일러블 (scalable)한 비트스트림을 지원할 수가 있다. 그러나 이 방법의 문제점은 확장성에서 지원이 가능한 계층 (layer)의 수에 한계가 있다[5]. 보통 3개의 계층 이상으로 부호화 하기에는 무리가 따른다. 이런 점 때문에 MPEG-4가 제공하는 확장성 부호화 기법이 현재의 인터넷 망에 활용하기에는 다소 비현실적이라는 분석이 제기되었고, 이를 해결하기 위해 기존의 확장성 부호화 기법보다 훨씬 더 세밀한 확장성을 제공하기 위한 FGS (fine granular scalability) 기술에 관한 표준화가 ISO/

\* LG전자 CDMA 단말연구소  
 \*\* 동의대학교 소프트웨어 공학과

IEC 산하의 MPEG (Moving Picture Experts Group)에서 진행 되었다[1]. 이 기술을 이용할 경우 1bit 단위의 세밀한 간격으로 비트율을 조절할 수 있게 된다. 뿐만 아니라, 전송하고자 하는 비트율을 압축하는 시점이 아닌 전송하는 시점에 결정할 수 있게 되므로, 여러 비트율로 콘텐츠를 전송해야 하는 경우, 하나의 압축된 비트스트림으로부터 여러 비트율의 전송용 비트스트림을 매우 손쉽게 생성하는 것이 가능해져 저장 공간의 부담 등이 최소화되는 장점이 있다[1,2]. 즉, FGS 기술의 등장으로 RealNetworks의 SureStream이나 MicroSoft WMT의 다중 비트율 옵션과 같은 효과를 하나의 비트스트림으로부터 얻어 낼 수 있기 때문에 스토리지 비용을 그만큼 줄일 수 있고, 비트율의 미세 변경이 가능하므로 가용 전송 대역폭에 최적화된 데이터의 전송을 실현할 수 있다.

## 2. MPEG-4 FGS 부호화 기법 원리

### 2.1 비트평면 부호화 기법

FGS는 기본적으로 Weiping Li가 창안한 비트평면 부호화 기법에 바탕을 둔다[1,3,6,7]. 일반적인 디지털 영상 부호화 기법에서는 run-length 부호화가 이용이 되지만, FGS는 run-length 부호화를 사용하지 않는 대신에 비트평면 부호화를 사용한다. run-length 부호화에서는 0이 아닌 수를 10진수로 표현하고 이 0이 아닌 수가 나올 때 까지 연속된 0의 개수를 한쌍의 정보로 가변길이 부호화 (VLC: variable length coding)한다. 반면에 비트평면 부호화에서는 0이 아닌 숫자의 이진수 표현을 사용한다.

DCT 기반의 부호화에서의 부호화 처리 단위는  $8 \times 8$  블록이다. DCT는 우수한 에너지 packing 특성을 지니므로 DCT block은 일반적으로 좌상단의 몇 개의 개수만 0이 아닌 값을 갖는다. 이러

한  $8 \times 8$  DCT block의 양자화 된 예가 그림 1에 나타나 있다.

비트평면 (bit-plane) 부호화에서 비트평면의 개수는 양자화된 DCT 블록 내에서 가장 큰 값을 갖는 절대값을 2진수로 표현하기 위해 필요한 최소의 비트수에 의해 결정된다. 예를 들어, 10이 가장 큰 값이라고 하면, 이 숫자를 이진수로 표현하기 위해서는 최소 4비트가 필요하고 따라서 비트평면 부호화를 위해 필요한 비트평면의 개수는 4가 된다. 비트평면 부호화를 위한 과정은 다음과 같이 정리 될 수 있다.

- i) 2차원 DCT 배열 (array)을 zig-zag 순서로 1차원 배열로 나열한다.
- ii) 가장 큰 절대치를 갖는 계수를 구하고, 이 계수를 2진수로 표현하기 위해 필요한 최소의 비트수  $N$ 을 구한다.  $N$ 이 비트평면의 개수가 된다.
- iii) 모든 DCT 계수를  $N$  비트의 2진수로 표현하여  $N$ 개의 비트평면을 형성한다.
- iv) 각 비트평면에 대해 3차원 심볼을 구성한다. 3차원 심볼은 다음과 같은 세가지 성분으로 구성된다. 즉, 1이 나오기 전의 연속된 0의 개수, 현재 비트평면 이후에 1의 값이 존재하는지의 여부를 나타내는 EOP (end of plane), 그리고 현재의 DCT 블록에서 현재 비트평면이 마지막 비트평면인지를 나

$$\begin{bmatrix} 10 & 0 & -3 & 0 & 0 & -1 & 0 & 0 \\ 6 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

그림 1. 양자화된 DCT 블록의 예

타내는 EOB의 3가지 정보로 하나의 심볼이 구성된다.

v) 모든 0이 아닌 계수의 사인 비트는 비트평면 부호의 뒷부분에 위치하여 부호화 된다.

다음은 그림 1의 DCT 블록에 대한 비트평면 부호화 과정을 보인다.

i) zigzag 순서화:

value: 10 0 6 0 0 3 0 2 2 0 0 0 2 0 0 1 0 ... 0  
sign : 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 ... 0

ii) 가장 큰 절대치를 갖는 계수를 구하고, 이 계수를 2진수로 표현하기 위해 필요한 최소의 비트수 N을 구한다: 10이 가장 큰 계수이며 10은 2진수로 1010이므로 비트평면의 개수 N=4가 된다.

iii) 네 개의 비트평면을 구성한다.

10 0 6 0 0 3 0 2 2 0 0 0 2 0 0 1 0 ... 0  
MSB: 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ... 0  
MSB-1: 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ... 0  
MSB-2: 1 0 1 0 0 1 0 1 1 0 0 0 0 1 0 0 0 ... 0  
MSB-3: 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 ... 0

iv) 3차원 심볼을 구성한다. (run, EOP, EOB)

MSB: (0,1,0)  
MSB-1: (2,1,0)  
MSB-2: (0,0,0), (1,0,0), (2,0,0), (1,0,0), (0,0,0), (3,1,0)  
MSB-3: (5,0,0), (9,1,1)

V) 3차원 symbol과 사인을 부호화 한다.

VLC(0,1,0), VLC(2,1,0), VLC(0,0,0), VLC(1,0,0), VLC(2,0,0), VLC(1,0,0), VLC(0,0,0)  
VLC(3,1,0), VLC(5,0,0), VLC(9,1,1), 00,1,0,1,0,1

비트평면 부호화에서 고려해야 할 주요한 문제는 크게 두 가지가 있다. 첫번째는 서로 다른 비트평면에서의 심볼은 서로 다른 확률분포를 보이므로 각 비트평면을 구분해야 한다. 두 번째로 각 비트평면상의 심볼들의 확률 분포가 서로 다르므로 각 비트평면에 독립적인 VLC 부호화 테이블

을 구성해야 된다는 점이다. 실제 FGS에서는 각 비트평면 마다의 PDF (probability distribution function) 특성을 고려한 독립적인 VLC 테이블을 구성하고 있다. 그런데, 보통 상위 3개의 비트평면을 제외한 나머지 하위 비트평면에서의 심볼에 대한 PDF는 거의 유사하므로 통합하여 하나의 VLC 테이블을 구성한다.

## 2.2 FGS 부호기 구조

그림 2는 FGS 향상계층을 위한 부호기 구조를 보이고 있다[8]. FGS 향상계층의 입력으로 입력 영상과 기본계층에서 재생된 신호의 차신호 (residual signal)가 들어가게 되고 이 신호는 8×8블록 별로 DCT 된다. 현재의 화면 전체에 대한 DCT 영역 차신호가 얻어지면 절대치가 가장 큰 계수를 구하여 비트평면의 개수를 정하게 된다. 이 후의 과정은 2.1절에서 설명한 비트평면 부호화 과정을 통해 FGS 향상계층의 비트스트림이 만들어 지게 된다. 그림 2에서 비트평면 쉬프트는 특정 부분의 화질을 향상 시키기 위해 그 부분에 해당하는 블록들의 비트평면을 MSB (most significant bit) 평면 쪽으로 상향 천이 해 주는 기능을 한다. 상위 MSB 평면의 경우 비트스트림 전송 시 먼저 전송되므로 채널 상황이 좋지 않은 상황에서 이 부분은 전송이 보장되게 된다. 그림 2에서 비트평면 VLC 블록에서는 엔트로피 부호화가 이루어지는데, FGS에서는 상위 네 개의 비트평면에 대해 각각 하나씩의 VLC 테이블을 이용한다. 네번째 이하의 비트평면에 대해서는 네번째 비트평면의 VLC 테이블을 그대로 적용하게 된다. Clipping 블록은 재생된 픽셀값의 범위가 0~255 범위를 넘어가게 될 경우 값을 0~255범위 안으로 값을 만들어 준다.

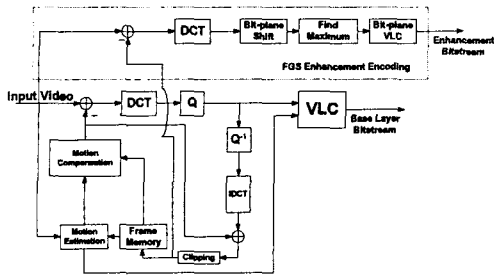


그림 2. FGS 부호기의 구조

### 2.3 FGS 복호기 구조

FGS 복호기는 그림 3에 나타나 있는 것처럼 기본계층의 비트스트림과 향상계층의 비트스트림은 서로 독립적인 복호화를 거친 후 합쳐져서 향상된 화질을 갖는 화면을 재생하게 된다. FGS 부호기 측에서 비트평면 쉬프트(shift)기능을 사용한 경우 복호기 측에서는 반대 방향으로 비트평면을 쉬프트하게 된다. 즉 부호기 측에서 비트평면을 상향시켰다면 복호기 측에서는 동일한 상승분 만큼 하향 쉬프트 해야 한다. 재생된 화면은 0~255사이의 값을 갖도록 clipping된다. 기본계층에서 재생된 그림은 다음 화면의 움직임 보상을 위해 메모리 버퍼에 저장이 되게 된다.

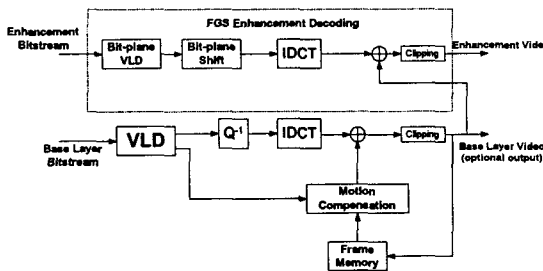


그림 3. FGS 복호기의 구조

## 3. MPEG-4 FGS 기술의 성능 분석

FGS의 부호화 효율 관점에서의 성능을 분석하기 위해 SNR 확장성과 PSNR 성능을 비교한

다. 그림 4~5는 Coastguard, Foreman 영상에 대한 FGS와 SNR 확장성의 부호화 성능을 비교하고 있다. 비트율이 증가할수록 FGS의 성능이 SNR 확장성을 앞서고 있으며, 가장 높은 비트율에서는 대략 2dB 정도의 화질 향상을 보여준다.

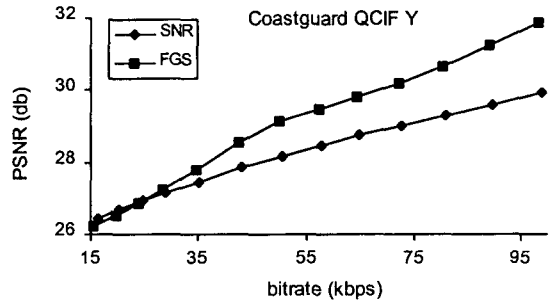


그림 4. Coastguard QCIF PSNR 비교

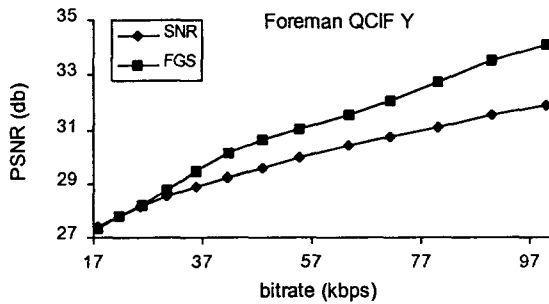


그림 5. Foreman QCIF PSNR 비교

## 4. FGS의 주관적 화질 향상을 위한 표준 기법

FGS 기술에 의해 얻을수 있는 화질의 주관적 품질을 향상 시키기 위해 두 가지 방식이 표준으로 채택 되었다. 첫번째가 주파수 가중 (frequency weighting) 기법이며, 두 번째가 선택적 향상 (selective enhancement) 기법이다.

### 4.1 주파수 가중 기법

DCT 블록에서 저주파 성분이 고주파 성분보다 화질에 더 큰 영향을 줌을 잘 알려져 있다. 따라

서 64개의 주파수 성분에 대해 서로다른 가중치를 가함으로써 저주파 성분에 보다 큰 값을 부여할 수 있다. 이렇게 하면 제한된 전송 비트율에서 저주파 성분이 우선적으로 전송되는 기회를 부여할 수 있다. 선택스에서는 주파수 가중 행렬을 정의하여 64개의 주파수 계수에 대해 적절한 가중치를 줄 수 있게 하고 있다. 그림 6은 저주파 성분에 보다 큰 가중치를 부여한 예를 보여주고 있다.

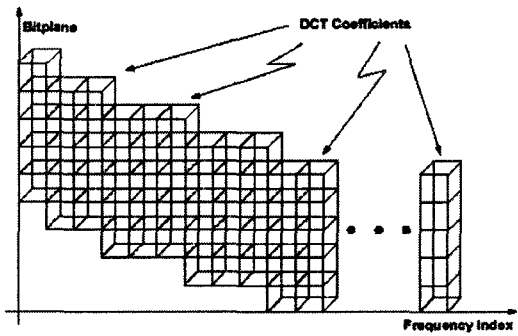


그림 6. 주파수 가중 기법을 이용한 비트평면 쉬프트 예

#### 4.2 선택적 향상 기법

선택적 향상 기법 (SE기법)은 화면에서 특정하게 화질을 강조할 필요가 있는 영역에 대해 비트 평면을 상향 쉬프트 시키기 위한 방법이다. 주로 배경보다 화면의 가운데 부분에 중요한 정보를 포함한 객체 (object)가 존재하기 때문에 영상의 가운데 부분의 화질을 향상시킬 때 유용하게 사용될 수 있다. 그림 7은 화면의 가운데 부분에 해당하는 매크로블록들의 비트평면을 MSB방향으로 2단계 상향 쉬프트한 예를 보여주고 있다. 그러나 선택적 향상 기법의 문제점은 선택적 향상이 적용된 영역 (SE region) 외부에서는 부호화 효율이 급격하게 떨어진다는 점이다. 그림 8은 Foreman QCIF 영상 한 장에 양자화 파라미터 (quantization parameter)를 28로 적용한 후 사람의 얼굴이 포함된 가운데 영역을 4 단계 상향 쉬프트했

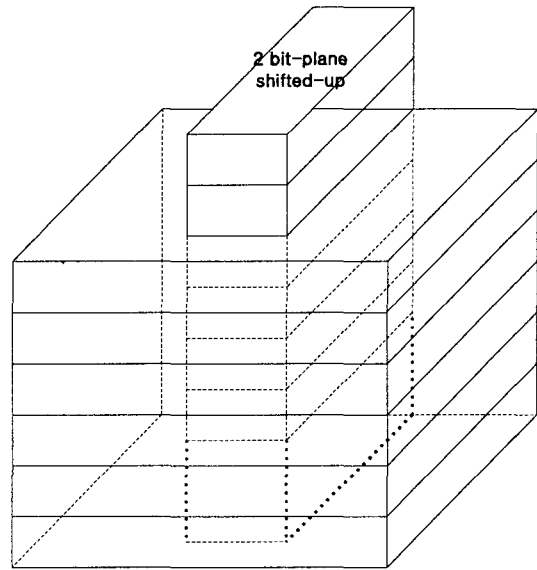


그림 7. 선택적 향상 기법을 이용한 비트평면 쉬프트 예

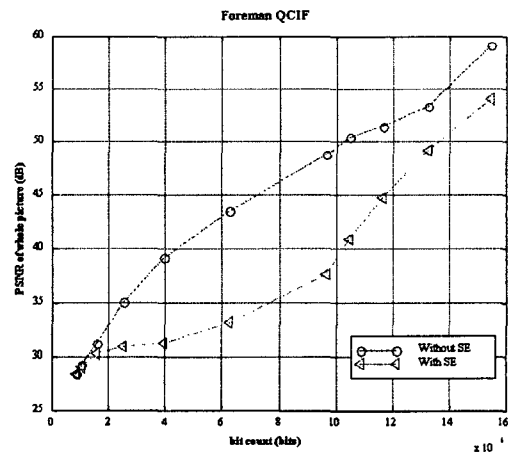


그림 8. 전체 화면에 대한 PSNR 성능 비교

을 때 전체 화질을 비교한 결과이다. 결과에서 선택적 향상 기법을 적용한 후의 전체 PSNR 화질이 적용하지 않았을 때 보다 상당히 열화됨을 알 수 있다. 그러나, 그림 9에서 보이듯이 SE 영역에서는 선택적 향상기법을 적용 했을 때의 PSNR 화질이 훨씬 좋음을 알 수 있다. 그림 10은 Foreman 영상에 대해 선택적 향상 기법을 적용하지 않았을 경우의 화질을 보여주고 그림 11는 선택적

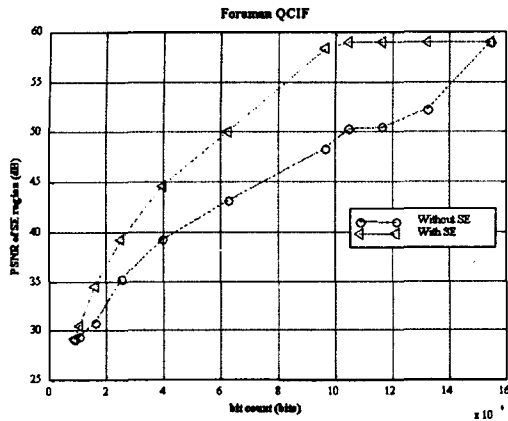


그림 9. SE영역 내부에서의 PSNR 성능 비교



그림 10. SE 기법을 적용하지 않았을 때의 화질



그림 11. SE 기법을 적용했을 때의 화질

향상 기법을 적용 했을 때의 주관적 화질을 보여 주고 있다. 그림 8~9에서 보인 객관적 화질 비교인 PSNR 비교와는 달리 선택적 향상기법을 적용한 경우가 훨씬 나은 주관적 화질을 보여주고 있다.

### 5. MPEG-4 FGS 기반의 차세대 스트리밍 시스템 기본 구조

그림 12에서는 MPEG-4 FGS 기술을 이용한 스트리밍 시스템의 전체적인 구조를 나타내었다. 시스템은 크게 MPEG-4 FGS 부호기와 실시간 스트리밍 서버, 실시간 스트리밍 클라이언트 등 3가지의 구성요소로 이루어진다. 비디오 부호기는 기본계층 (BL)과 향상계층 (EL) 등 두 개의 비트열을 생성하게 된다. 기본계층은 전송 대역폭이 최대한 보장될 수 있는 가장 낮은 비트율로 부호화되며 향상계층은 네트워크의 상황에 따라 가변적으로 부호화 된다. 한편, 앞에서 설명한 기본 계층과 향상계층 비트열은 독립된 채널을 통해 전송되므로 네트워크 상황에 따라서 클라이언트측에서는 버퍼넘침 (overflow) 또는 고갈 (under-flow)의 문제가 발생할 수 있고 지터 (jitter)등에 의한 기본계층과 향상계층의 동기화에 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 RTP (real-time transport protocol)와 RTCP (RTP

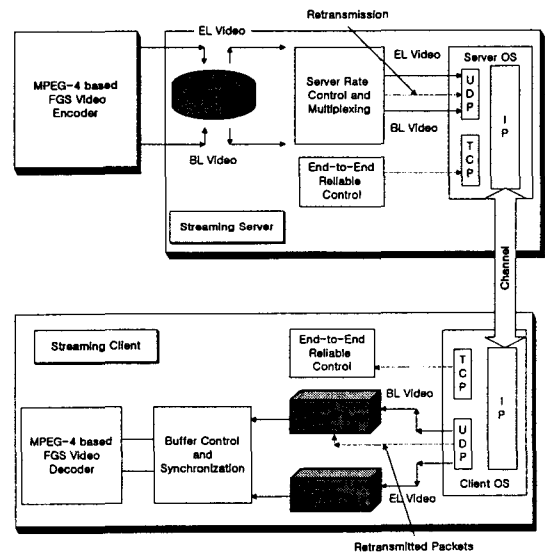


그림 12. MPEG-4 FGS 기반의 스트리밍 시스템 전체 구조

control protocol)를 사용한다[9]. RTP의 타임스탬프 (time stamp)기능에 의해 기본계층과 향상계층 간의 동기화를 맞춰주고 RTCP의 수신 패킷 정보를 이용한 네트워크 모니터링 기능을 이용하여 네트워크 상황을 송신단으로 피드백 (feedback)하여 전송측의 송신율을 조절할 수 있다. 그림 13에서는 스트리밍 서버의 구조를 자세히 나타내었다. 서버는 크게 비디오 부호기와 스트리밍 서버 등 크게 두 부분으로 나누어 진다. 부호기는 오프라인으로 미리 결정될 수 있는 네트워크의 대역폭 자원을 고려하여 부호화를 수행하게 된다. 즉, 부호기의 인터페이스 모듈을 이용 네트워크가 지원 가능한 대역폭의 범위,  $[R_{min}, R_{max}]$ 를 알아내고 이를 바탕으로 기본계층은  $R_{BL} (\leq R_{min})$ 의 비트율로 부호화 된다. 그리고 향상계층은 비트율  $(R_{max} - R_{BL})$ 을 넘지 않도록 부호화 된다.

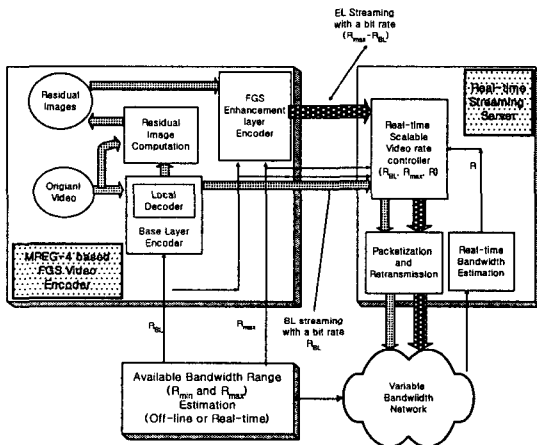


그림 13. MPEG-4 FGS 스트리밍 서버 구조

## 6. 결론

본 논문에서는 향후 All IP를 지향하는 IMT-2000등 차세대 통신망에서 비디오를 이용한 가장 유망한 응용 분야로 예상되는 스트리밍 서비스를 위한 최신 기술인 FGS에 대해 심도 있게 다루었

다. FGS는 IP 기반의 인터넷의 가장 큰 특성인 시간에 따른 대역폭 변동에 가장 능동적으로 대처할 수 있는 부호화 기법이다. 특히, 대역폭 변동에 bit단위까지 세밀하게 적응을 할 수 있고, 급격한 대역폭의 변동에도 기존의 SNR 확장성등에 비해 화질의 변화 폭을 최소화 할 수 있다. 그리고 비디오 스트리밍의 가장 큰 특성이 이미 부호화 되어 저장된 비디오를 이용한다는 점인데, 기존의 계층 부호화 (layered coding) 기법은 그 계층의 개수의 한계로 인해 확장성 제공에 한계가 있고, 최고의 화질로 저장되지 못한다. 반면에 FGS는 비트평면의 개수를 다수로 구성하여 부호화 한 후 저장하기 때문에, 최고의 화질로 저장된다. 이후 서비스시의 대역폭 변동에 따라 능동적으로 필요한 비트평면만을 전송함으로써, 네트워크의 대역폭에 적응적인 전송이 가능하고 화질도 극대화 할 수 있다.

이러한 FGS의 특성을 이용하여 이미 활용되고 있는 스트리밍 서비스 구조를 개선한다면, All IP를 지향하는 차세대 망구조에서 탁월한 성능의 스트리밍 서비스를 제공할 수 있을 것이다.

## 참고 문헌

- [1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301-317, March 2001.
- [2] Y. Tung, J. Wu, P. Hsiao, K. Huang, "An efficient streaming and decoding architecture for stored FGS video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 8, pp. 730-735, Aug. 2002.
- [3] Y. Zhang, "Streaming video over internet-issues and new development", *Proc. IEEE Workshop on Signal Processing Systems*, Taipei, Taiwan, Oct. 2000.

- [4] T. Hou, D. Wu, W. Zhu, H. Lee, T. Chiang, Y. Zhang, "An end-to-end architecture for MPEG-4 video streaming over the Internet", *Proc. Int. Conf. Image Processing* Vol. 1, Kobe, Japan, pp. 254-257, 1999.
- [5] W. Li, "Scalable video coding with fine granularity scalability", *Proc. IEEE Int. Conf. Consumer Electronics*, LA. USA, June 1999, 306-307.
- [6] W. Li, *et al.*, "Streaming video profiles based on fine granularity scalability", Contribution M5582, ISO/IEC JTC1/SC29/WG11 Meeting, Maui, Dec. 1999.
- [7] W. Li, "Fine granularity scalability using bit-plane coding of DCT coefficients", MPEG98/m4204.
- [8] ISO/IEC JTC1/SC29/WG11 N3317, FGS Verification Model Version 4.0, Noordwijkerhout, March, 2000.
- [9] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A transport protocol for real-time applications," IETF RFC 1889, Jan. 1996.



서 광 덕

- 1996년 2월 한국과학기술원 전기 및 전자공학과 (공학사)
- 1998년 2월 한국과학기술원 전기 및 전자공학과 (공학석사)
- 2001년 2월 제 7회 삼성 휴먼테크 논문대상 장려상 수상
- 2002년 2월 제 8회 삼성 휴먼테크 논문대상 은상 수상
- 2002년 8월 한국과학기술원 전자전산학과 (공학박사)
- 2003년~2004년 세계인명사전, Marquis Whos Who in Science and Engineering (7th Edition)에 등재
- 2002년 8월~현재 LG전자 CDMA 단말연구소 선임연구원 (Project Leader)
- 관심분야: 영상통신, 영상 부호화, 영상통신 시스템, 3GPP/3GPP2 표준화
- E-mail: jiance@lge.com



권 순 각

- 1990년 2월 경북대학교 전자공학과 (공학사)
- 1992년 2월 한국과학기술원 전기및전자공학과 (공학석사)
- 1998년 2월 한국과학기술원 전기및전자공학과 (공학박사)
- 2001년 3월~현재 동의대학교 소프트웨어 공학과 교수
- 관심분야: 영상부호화, 영상처리, 영상 통신
- E-mail: skkwon@dongeui.ac.kr