

논문97-2-2-12

# 비트율 계위 비디오 부호기

이범열\*, 임성호\*, 민병의\*, 황승구\*, 황재정\*\*

## Bitrate Scalable Video Coder

Beom-Ryeol Lee\*, Sung-Ho Im\*, Byung-Eui Min\*, Soung-Ku Hwang, and Jae-Jeong Hwang\*\*

### 요 약

본 논문에서는 H.263비디오 압축 표준을 기반으로 한 비디오 부호기로서 두개의 계층을 갖는 비트율 계위 비디오 부호기를 제안한다. 제안하는 부호기의 기본계층에서는 H.263표준에서 제안하고 있는 기본 부호화 알고리즘을 수용하여 비디오 데이터의 고 압축 부호화 효율을 얻는다. 상위계층에서는 부호화의 고급 기법이라 할 수 있는 인간 시각 시스템 정보를 근거로 양자화를 설계하고 적용한다. 상위계층에서는 기본계층의 이산 역현 변환 계수에 보다 세밀한 양자화를 가하여 부호화하게 되며, 이때 필요한 비트스트림의 구조를 설계한다. 복호화기에서는 두 계층에서 수신한 데이터를 역양자화 및 역변환을 수행한 후 더하여 원 영상을 복원하게 된다. 실험 결과를 통하여 비트율 약30kbps 정도에서 계층을 적용하지 않은 부호기와 비교하여 거의 비슷한 화질을 얻을 수 있었으며, 상위계층의 비트스트림을 전송하기 위한 부가 정보를 약 프레임당0.5kbits이하를 유지하도록 하였다.

### Abstract

We propose a H.263-based video coder with two-layer scalability. The base layer is coded by using the default H.263 coding algorithms to achieve high compressed video data and the enhanced layer is coded by enhanced coding such as HVS-based quantization updating. The enhanced layer contains only coded refinement data for the DCT coefficients of the base layer. Bitstream syntax and semantics for enhancement layer are designed and quantizer design using the HVS is proposed. Data from the two layers are combined after inverse quantization and inverse DCT processes in the decoder. We show with experimental results that the proposed layered codec achieves comparable picture quality with non-layered codec at bitrates of 30 kbps or less. Overhead information for the bitstream layer can be limited to less than 0.5 kbits/frame.

### I. 서 론

계위 비디오 부호화(Scalable video coding)는 다양한

시간적/공간적 해상도 및 화질을 지원하는 비디오 응용에 적합한 방법으로 다자간 비디오 영상 회의, ATM(Asynchronous Transfer Mode) 망에서의 비디오 데이터 전송 및 고휘상도 비디오 전송 시 표준 텔레비전 신호 수신 등에 이용할 수 있으며, 여러 발생 가능성이 있는 유/무선 비디오 전송 및 복원에 적합한 방법이다. 이와 같은 비디오 부호화 방식은 모든 계층을 동시전송(Simulcast)하거나, 각 계층별로 대역폭을 할당하여 전송하게 되며, 시스템의 성능에 따라 다양한 전송방식으로 구현된다[1].

특히 MPEG-2표준에 정의된 프로파일에는 신호대잡음비(SNR; Signal-to-Noise) 계위 및 공간적 계위 등의 프

\* 한국전자통신연구원 멀티미디어연구부  
Multimedia Technology Dept. Electronics and Telecommunications Research Institute

\*\* 군산대학교 전파공학과  
Dept. of Radiocommunication Eng., Kunsan National University

\* 이 논문은 일본 1996년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

로파일을 포함하고 있다[2]. 이러한 비트스트림 계위에 있어서 기본계층에서는 비디오의 중요 정보를 낮은 비트율로 포함하고 있으며 상위계층에서는 기본계층의 비트 스트림 데이터를 복원할 때 고화질을 유지할 수 있도록 보다 세밀한 정보를 전송하게 된다.

비트율 계위 비디오 부호화에서는 동일한 공간 해상도에 서로 다른 화질을 갖는 둘 또는 그 이상의 비디오 계층을 구성하게 된다. 상위계층에서는 기본계층의 DCT (Discrete Cosine Transform) 계수에 대하여 발생하는 양자화 오차에 대한 보다 세밀한 양자화를 수행한다. 이들 계층별 비디오 데이터는 수신단에서 블록 단위로 DCT 영역 또는 공간 영역에서 더해져 기본계층 데이터를 보완하여 고화질의 영상을 얻게 된다[3]. 첫번째 비트스트림으로부터 구한 기본계층의 비디오는 그 자체가 비계위(Nonscalable) 복원 과정을 거치거나 공간적 또는 시간적 계위 복호화 과정을 거친 것이다. 두번째 비트스트림으로부터 구하게 되는 상위계층의 비디오는 주로 DCT 계수와 비교적 적은 부가 정보(Overhead)로 구성된다. 그러므로 부호기의 전체 비트량은 소량 증가하게 된다. 여기서 상위계층의 복원 데이터만을 가지고는 전체 영상을 복원할 수 없으며, 기본 또는 하위 계층의 복원과 병행하여 처리하게 되고, 결국 모든 계층의 DCT 계수를 더한 후에 비계위 복호화 과정과 동일한 과정을 거친다. 여기서 부호화 방법에 따라 복호기에서 각 계층의 개별적 복호화 과정이 완료된 후에 블록별 공간 영역 데이터를 합하여 최종적인 복원 영상을 구할 수도 있다.

관련 분야의 연구로는 MPEG-2의 비디오 표준이 발표되면서 계위 비디오 복호기에 대한 알고리즘이 채택되었고[2], 64kbps의 채널을 고려한 H.261 비디오 표준을 기반으로 ATM(Asynchronous Transfer Mode)망에서 SNR 계위 비디오 부호기에 대하여 제안되어 관련 분야에 많은 기여가 되었다[3]. 또한 SNR 계위 비디오 부호기에 최적의 임계치를 구하는 문제를 다룬 바 있고[4], MPEG2 비디오 패킷 손실을 개선하기 위한 방법으로 계위 비디오를 제안한 바 있다[5]. 최근에는 멀티미디어 단말 표준에도 대역폭 및 터미널의 처리 능력에 따라 적응성을 부여한 비디오 부호기 구조로써 계위 비디오 부호기를 제안하고 있다[6].

본 논문에서는 H.263 표준을 기반으로 비트율 계위 비디오 부호기의 구조 및 상위계층 비디오 스트림의 구성을 제안한다. 또한 상위계층의 부호화를 위해서 인간의 시각 모델을 반영한 양자화기를 설계한다. 제안한 부호기로부터 비트율에 따른 다양한 화질의 비디오를 생성하게 되며, 비디오 스트림의 복원시 시간적으로 개선된 화질을 보장하도록 하였다. 제안한 부호기를 사용하여 H.323 멀티미디어 터미널 표준을 지원하는 영상 회의 시스템에서 다양한 비트율의 비디오 전송에 응용하도록 하였다. 또한 이동 채널을 통한 휴대형 멀티미디어 장치에도 적절한 에러 비율에 따른 비디오 전송에 이용할 수 있도록 하고 있다. 제2장에서

서 계층화된 비트율 계위 비디오의 부호화기의 구조에 대하여 기술하고, 이어 제3장에서는 H.263 비디오 부호기 표준에 기반한 상위계층의 비디오 스트림의 선택 구조를 설명한다. 상위계층의 부호기에 대하여 인간의 시각 모델을 근거로 한 양자화기의 설계에 관하여 제4장에서 기술하며, 실험 결과에 대하여 제5장에서 비디오 시퀀스 별로 비교해 본다. 끝으로 제6장에서는 결론과 함께 향후 개발 및 응용 방향에 대하여 기술한다.

## II. 두 계층의 비트율 계위 부호기

제안하는 비트율 계위 비디오 부호기에서는 공간적으로나 시간적으로 동일한 해상도의 비디오 데이터를 서로 다른 화질을 갖도록 독립적으로 부호화하여 동시에 두개의 부호화된 비트열을 전송하게 된다. 기본계층의 부호화 과정은 계위를 적용하지 않는 부호화 과정과 완전히 동일한 과정을 거치게 된다. 이때 기본계층의 부호화는 복호화 과정에서 필연적으로 양자화 오차를 발생시킨다. 부호화 과정에서 각 블록에 대한 양자화는 해당 블록에 대한 양자화 오차를 수반하게 되는데 상위계층의 부호기에서 이때의 양자화 오차에 대하여 다시 한번 양자화 과정을 거치도록 하여 이 정보를 추가로 전송하게 된다.

그림1에는 제안한 비트율 계위 부호기의 전체 부호화 과정에 대하여 나타내었다. 기본계층의 부호화 과정은 H.263 부호화 표준을 따르도록 했다. 이 표준 알고리즘은 64 kbps 이하의 부호화 율에 적합하게 고려된 것으로 움직임 벡터를 응용한 DCT/DPCM(Differential Pulse Code Modulation) 복합 부호화 방식이라 할 수 있으며, DCT 변환 및 양자화 과정을 통하여 공간적인 상관성을 바탕으로 공간적 중복성(Redundancy)을 줄이고, 인트라 프레임의 예측을 통하여 시간적 상관성을 줄여 정의된 특정 포맷의 코드로 부호화하여 전송하는 방식이다[7].

일반적으로 비트율 계위 부호기 구현 시 움직임 보상을 위한 예측 루프(Prediction loop)에 상위계층의 비디오 데이터에 대한 역 양자화 결과를 패환(Feedback)할 수 있거나 하지 않을 수도 있다. 그림1의 경우 역 DCT 루틴에 상위계층의 역 양자화 과정을 패환하지 않는 방식으로 설계되었다. 전체 부호기는 두 계층의 비디오 스트림을 출력하게 되는데 기본계층에서는 H.263 비디오 부호화 표준을 사용하여 저 화질의 데이터 스트림을 생성하며, 상위계층에서는 기본계층에서 발생하는 양자화 오차에 대하여 다시 한번 세밀한 양자화를 수행하여 전체적으로 고화질이 유지되도록 하고 있다. 여기서 양자화 오차 값을 움직임 보상을 위한 패환 루프에 추가하지 않았다. 기본계층 부호기에서 움직임 예측 및 보상을 위하여 상위계층의 출력을 복원하여 패환을 가하게 되면 부호화한 영상과 수신단의 복호화 한 영상간에 드리프트(Drift) 현상이 발생하게 된다[5]. 이러한 현상은 부호기에 상위계층의 출력을 패환하지

않음으로써 피할 수 있으며, 부호화기의 상위계층에서도 기본계층의 부호기에서 발생하는 잉여 예측 오차(Residual prediction error)만을 부호화하면 되므로 전체 부호기가 더욱 단순해지는 이점이 있다.

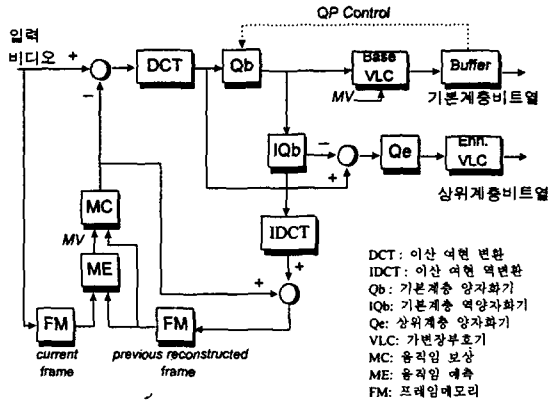


그림 1. 비트율 계위 비디오 부호기  
Fig. 1. Bitrate scalable video encoder

그림 2는 비트율 계위 복호기를 보이고 있다. 복호화기에서도 기본계층 비디오를 복원하기 위하여 H.263비디오 복원 알고리즘을 그대로 사용한다[7]. 상위계층의 복호기에서는 기본계층 비디오의 화질을 개선하기 위한 추가 정보를 포함하고 있으므로 하위 계층과 상위계층의 비디오 스트림에 대하여 각각 독립적으로 블록에 대하여 복호화 과정으로 역 양자화 및 역DCT를 수행한 후 블록 단위의 계층별 공간 영역 데이터에 대하여 합하게 되면 전체 부호기의 고화질 비디오 시퀀스를 얻게 된다. 그림1에서 제시한 바와 같이 제안한 부호기의 특징으로는 비디오 스트림에 대하여 비트율 계위를 부가하기 위해서 기본계층에서는 H.263 비디오 부호화 표준을 따른 것이고, 이때 기본계층에서 발생하는 양자화 오차에 대하여 인간 시각 모델을 근거한 양자화를 설계하여 추가적인 양자화를 수행하게 한 점이다. 기본계층 부호기에서는 기본H.263 부호기를 사용하게 되는 효율적인 부호화를 위하여 부호화 알고리즘에서 제안하고 있는 확장된 옵션을 사용할 수 있다. 상위계층에서는 기본계층의 왜곡(Distortion)에 대하여 재양자화(Requantization)하기 위하여 영상에 밝기 및 공간 주파수에 대한 시각적 감도를 고려하여 가중치(Weighting value)를 취하게 된다. 여기서 가중치에 대한 기준은 DCT 함수에 대한 공간주파수와 기본 주파수의 관점에서 8x8 블록에 대한 시각적 감도 행렬을 계산하여 이용하게 되는데 이 값은 기본계층의 양자화 오차에 대하여 각 영상의 왜곡에 대한 평가 지수로도 활용될 수 있다. 즉, 정량적인 부호화 오차에 대하여 시각적인 오차율과의 차이점이 시각적 감도와 관련하여 차이점을 보이게 되며 이러한 특성을 양자화기 설계에 반영하고자 하는 것이다. 상위계층의

양자화기 설계에 관하여는 다음 장에서 자세하게 기술하기로 한다.

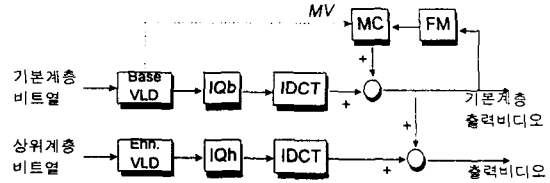


그림 2. 비트율 계위 비디오 복호기  
Fig. 2. Bitrate scalable video decoder

그림 2의 복호화기에서 보는 바와 같이 두개의 분리된 역 양자화기와 역 변환 과정이 수행된다. 이 과정은 부호기에서 인터프레임간의 예측에 기본계층만을 사용하여 부호화를 수행하였기 때문에 필요한 것이다. 부호화기의 부호화 방법에 맞는 복호화기를 위해서 상위계층의 복원 결과는 복호화기에서 수행되는 예측에 사용되어서는 안된다. 비디오 데이터 전송상의 네트워크 전송량 폭주나 수신 시스템의 처리 능력 등의 문제로 전송된 상위계층의 비디오 패킷에 대하여 손실이 발생할 경우 수신단의 복호화기에서 영상의 출력은 일시적으로 화질 저하를 겪게 된다. 즉, 기본계층에서 복원되는 비디오 화질만을 얻게 된다. 상위계층의 비디오 데이터가 없다 해도 단일 계층 비디오만으로, 예를 들어, 영상회의를 진행하기 위한 최소한의 적당한 화질을 유지할 수 있는 것이다. 이때 기본계층 부호기에서는 비트율을 증가시킬 수 있으며, 상위계층의 비디오 정보는 버리게 된다. 마찬가지로 복호화기에서도 상위계층의 비디오 채널로부터 전송되는 비디오 데이터는 무시하게 된다.

제안하는 비트율 계위 비디오 부호기의 구조상 하드웨어의 복잡도는 기본계층의 H.263기반 부호기 구조에 양자화 오차를 얻는 차분기 및 상위 계층 양자화기와 그 가변장 부호기 모듈이 첨가하는 정도로 구현 가능하다. 비트율 계위 복호기에서는 H.263 복호기 구조에 상위 계층의 가변장 복호기와 영양자화기에 역변환기 및 덧셈기가 추가되어야 한다.

### III. 상위계층의 구문 구조

복수의 계층을 갖는 비디오 부호기에서 상위계층의 계수는 기본계층의 양자화 오차이므로 상위계층의 비디오 데이터의 부호화 입력은 현재 부호화하려는 매크로 블록과 예측된 이전 매크로 블록간에 차분 값을 이용할 수 있다. 이 값은 기본계층 부호기의 양자화 왜곡 값을 상위계층 부호기의 입력으로 하는 것과 같다[3]. 즉, 상위계층의 변환 계수는 기본계층의 잉여 양자화 왜곡이라고 할 수 있다. 상위계층 부호기에서는 기본계층 부호기의 양자화

왜곡에 대하여 블록별 지그재그 스캔을 수행하고 가변 임계치(Threshold)를 두어 부호화하게 되는데 이는 기본계층 부호기에서 H.263비디오 압축 표준에서 수행하는 것과 같은 방식을 이용한다. MPEG-2표준에서는 SNR계위를 지원하기 위한 부호화 스트림 구조를 정의[2]하고 있는데 이 방식을 참조하여 본 논문에서 제안하고 있는 비트율 계위 부호기에 대해서도 H.263부호화 표준에서 정의[7]된 스트림 코드를 기반으로하여 상위 비디오 부호기에 대해서 비디오 스트림 구조의 구문을 설계하였다.

상위계층 부호기의 스트림 구조에 대한 구문은 비디오 시퀀스, 픽처, 블록그룹(GOB:Group of blocks), 매크로 블록, 블록의 다섯 가지 부호화 레이어(Coding layer)로 나누고, 그림3과 같은 구문 구조를 제안하였다. 비디오 시퀀스 부호화 레이어에서는 시퀀스 시작을 나타내는 코드(22bits), 부호기에서 사용하는 비트율 계위 종류를 나타내기 위해서 계위 모드 정보(2bits)를 포함하여 전체 부호기가 몇 단계의 계위를 두고 있는가 하는 식별자(3bits)를 포함하고, 다음 부호화 레이어인 픽처 부호화 레이어를 포함하도록 하였다. 그림3의 각 부호화 레이어 리스트에서 next\_start\_code()는 각 부호화 레이어에서 스텝비트(Stuffing bits)를 제외한 해당 레이어의 시작 코드를 찾는 과정을 의미하고, bytealigned()는 다음 코드를 검출하기 위한 시작 비트가 바이트 단위로 맞추어 최상위 비트임을 나타낸다. 비디오 시퀀스 부호화 레이어는 기본계층과 상위계층의 부호기에서 사용된다. 픽처 부호화 레이어와 GOB부호화 레이어는 기본계층의 부호기에서 적용되는 H.263 표준 부호기에서와 동일한 구문을 사용하며 GOB헤더는 부호화 효율을 고려하여 사용되지 않을 수도 있다. 매크로 블록 부호화 레이어에서는 비트율 계위로 부호화 되었음을 매크로 블록 타입 코드에서 나타내도록 하였다. 상위계층 부호기에서는 예측에 관한 정보는 고려되지 않는다. 왜냐하면 잉여 계수는 시간적 정보를 갖지 않기 때문이며, 따라서 움직임 벡터에 대한 정보는 상위계층 부호기의 매크로 블록에는 포함되지 않는다. 매크로 블록 부호화 레이어의 패턴 코드(Pattern code)에 따라 블록 부호화 레이어에서 블록별로 부호화가 되었는지 안되었는지 하는 정보를 나타내게 된다. 상위계층 부호기에서 기본계층의 부호기에서와 마찬가지로 PB 프레임 모드를 결정했을 경우에도 추가로 부호화된 블록 정보를 포함하게 된다. 블록 부호화 레이어에서는 매크로 블록 부호화 레이어에서 지정된 패턴 코드에 따라 한 매크로 블록의 끝까지 부호화된다. 블록 부호화 레이어에서는 H.263 부호화 표준에서 정의된 방식의 포맷에 따라 DCT계수를 부호화하기 위하여 마지막 0이 아닌 지 여부를 결정하는 정보(Last), 계속된 0비트의 갯수(Run) 및 0이 아닌 계수의 실제 값(Level)을 조합하여 만들어진 3차원 가변장 부호(Variable-length code)로 부호화된다.

```

Video_Sequence(){
sequence_start_code : 22bits
scalable_mode : 2
layered_id : 3
do {
picture()
} while (next_bits() == picture_start_code)
next_start_code()
}

Picture() {
picture_start_code : 22bits
temporal_reference : 8
picture_type : 13
picture_quantizer : 5
continuous_presence_multipoint : 1
if (continue_presence_multipoint)
picture_sub_bitstream_indicator : 2
if (picture_type == PB_frame) {
temporal_reference_for_B_frame : 3
B_picture_quantizer : 2
}
extra_insertion : 1
bytealigned()
do {
gob()
} while (next_bits == gob_start_code)
next_start_code()
}

gob() {
if (next_bits() == gob_start_code) {
gob_start_code : 17bits
group_number : 5
if (continuous_presence_multipoint)
gob_sub_bitstream_indicator : 2
gob_frame_id : 2
gob_quantizer : 5
}
do {
macroblock()
} while (!end_of_block)
next_start_code()
}

macroblock() {
code for macroblock_type 4
coded_block_pattern_for_chrominance : 1-9 bits
if (picture_type == PB_frame) {
macroblock_mode_for_B_frame : 2
if (macroblock_mode_for_B_frame == '1')
coded_block_pattern_for_B_block : 6
}
coded_block_pattern_for_luminance : 2-6
if (macroblock_type == 4)
differential_quantizer : 2
for (i=0; i<block_count; i++)
block(i)
if (picture_type == PB_frame)
for (i=0; i<block_count; i++)
block(i)
}

block() {
if (pattern_code[i]) {
if (pattern_type != PB_frame)
intra_dc : 5bits
else
event_code_of_first_DCT_coeff : 3-13 or 19
while (!end_of_block)
event_code_of_subsequent_DCT_coeff : 3-13 or 19
}
}
}
    
```

그림 3. 상위계층의 비디오 스트림에 대한 구문 구조  
 Fig. 3. Video stream syntax for enhancement layer

### IV. 상위계층의 양자화기 설계

기본계층의 양자화기에서 높은 비디오 데이터 압축율을 얻기 위해 비교적 큰 양자화 간격으로 양자화를 시도한 것에 대하여 상위계층에서는 기본계층에서 발생하는 양자화 왜곡에 대하여 더욱 세밀한 양자화를 수행하기 위해 보다 작은 양자화 간격을 이용하게 된다. 그림 4에서는 계위를 사용하지 않는 부호기의 양자화 간격과 계위를 사용한 부호기에서 기본 계위 부호기와 상위 계위 부호기에서 양자화 간격을 비교하고 있다. 그림에서  $A$ 는 양자화의 대상이 되는 데이터의 최대폭, 즉 다이내믹 레인지(Dynamic range)이고,  $N_s$ ,  $N_b$  그리고  $N_e$ 는 각각 전체 시스템의 양자화 레벨 수, 기본계층의 양자화 레벨 수 및 상위계층의 양자화 레벨수를 나타낸다. 이때 선형 양자화기를 고려할 때 전체 시스템의 최대 양자화 왜곡은  $A/2N_s$ 이다. 마찬가지로 기본계층의 최대 양자화 왜곡은  $A/2N_b$ 이 된다. 여기서  $N_b \leq N_s$ 이다.

이때 전체 시스템이 갖는 최대 양자화 왜곡은 다음과 같이 표시 할 수 있다.

$$\frac{A}{2N_s} = \frac{A}{2N_b} \cdot \frac{N_b}{N_s} \quad (1)$$

여기서  $N_b/N_s$ 는 상위계층 부호기에서의 왜곡 감쇄비이다. 상위계층의 최대 양자화 왜곡은 다음과 같다.

$$\frac{A}{2N_e} = \frac{A}{2N_b} - \frac{A}{2N_s} \quad (2)$$

여기서 상위계층 부호기를 위한 양자화 레벨수는 다음과 같이 구할 수 있다.

$$N_e = \frac{A}{2\left(\frac{A}{N_b} - \frac{A}{N_s}\right)} = \frac{N_s N_b}{N_s - N_b} \quad (3)$$

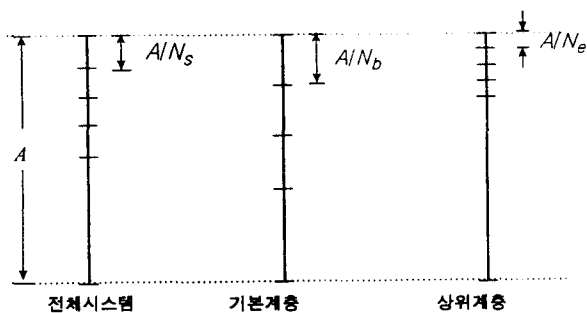


그림 4. 계층별 양자화 간격  
Fig. 4. Quantization step size at each layer

식(3)에서 상위계층의 양자화 레벨수  $N_e$ 에 대하여 시각적 중복성(Irrelevancy)을 감쇄시키기 위해서 다음과 같은 가중치를 고려할 수 있다. 즉,

$$N_e = W_h \frac{N_s N_b}{N_s - N_b} \quad (4)$$

여기서  $W_h \leq 1$ 이며,  $W_h$ 는 인간의 시각적 감도 함수에 따른 정규화한 가중 요소(Weighting factor)이다[8].

인간의 시각 모델은 시각 특성에 대한 분석으로 다양한 전개가 가능할 수 있지만 기본적으로 영상의 밝기에 대한 시각적 감도, 공간 주파수 및 시간적 주파수에 대한 감도, 그리고 색상에 대한 감도 등을 고려할 수 있다. 그러므로 모델링의 특성에 따라  $W_h$ 는 다양하게 정의 될 수 있고, 본 논문에서는 기본계층 부호기에서 발생하는 잉여 양자화 왜곡에 대하여 공간 주파수에 따른 감도를 구하여 가중치로 사용한다.

먼저 시각적 특성을 살펴보면 영상의 밝기에 대한 대비(Contrast ratio) 효과로서 배경에 따라 빛의 세기(Intensity)에 대하여 다른 감도를 갖게 된다. 두가지 밝기에 대해 시각적으로 구별을 할 수 있는 최소의 차이는 JND(Just Noticeable Difference)값으로 알려진 바 있다[9]. 이 값은 값 자체의 차이에 의해서도 달라지지만 밝기(Intensity)의 레벨에 대해서도 변화한다. 같은 밝기의 배경에 대하여 배경과는 밝기 차이를 보이지만 같은 밝기의 영역의 크기가 다를 때 시각적으로 느끼게 되는 감도 차이를 비교해 볼 수 있다. 배경의 밝기가 일정하게  $L$ 값을 유지하고, 배경의 중심에  $\Delta L$ 만큼 밝기가 증가한 서로 다른 영역에 대하여 밝기  $L + \Delta L$ 을  $JND(L)$ 로 정의한다. 이때  $JND(L)/L$ 는 웨버의 비(Weber's ratio)라고 한다[10]. 일반적으로 이 값은 일정 영역의 밝기(Intensity)에 대해서는 일정한 값을 유지하는 것으로 대비(Contrast ratio)라고도 한다.

웨버의 비에 대하여 일정한 특성들이 Shen 등에 의해서 실험[11]되었고 다음과 같은 결과가 있다. 배경의 밝기  $L$ 가 낮은 계조도(Gray level) 값을 갖는 영역(Gray level: 0~50정도)에서는 모니터에서 가까운 거리일 때 시각적 감도가 높다. 그러나  $JND$ 값은 배경의 밝기가 약 50 이상의 레벨에 대해서는 바라다 보는 거리에 따라 변하지 않는다. 조명을 비춘 상태의 영향에 대해서는 어두운 공간에서 낮은 계조도 영역에 대해서 시각적 감도는 더 높게 나타나고, 높은 계조도 영역에 대해서는 별 영향이 없는 것으로 나타난다. 모니터의 밝기에 대한 영향을 살펴보면 보다 밝은 상태에서 시각적 감도는 높고, 보다 낮은 밝기에서는 감도가 약한 것으로 나타났다. 웨버의 비에서 중간 정도의 밝기에서는 감도가 일정하고, 보다 높거나 낮은 밝기에서는 감도가 증가한다는 것과 비교하면 이 실험 결과에서  $JND$ 값은 계조도가 50~234정도에서 상당히 일정하게 나타나고 있다. 계조도235에서 255에서는 약간의 증

가를 보였다. 즉, 모니터 밝기의 변화 범위(Dynamic range)는 계조도가 최대가 되더라도 그렇게 많이 확장되지 않는다는 것과 밝기가 높은 상태에서의 시각적 감도는 크지 않다는 것을 나타낸다.

이상과 같은 인간의 시각 시스템 특성에 관한 연구 결과에서 밝기의 대비에 대한 효과를 고려할 때 블록의 영상 데이터에 대해서 낮은 밝기에서는 양자화 오차에 대한 시각적 감도가 낮고, 중간 정도의 밝기에 대해서는 감도를 유지하므로 이러한 밝기의 대비에 대한 시각적 모델을 근거로 블록 영상의 낮은 레벨(0-50)의 밝기(Luminance)와 매우 높은 레벨(235-255)에 대해서는 양자화 오차가 없는 것으로 간주하고 부호화되도록 본 논문의 상위 계층 부호기의 양자화기 설계에 반영하였다.

블록 영상이 갖은 공간 주파수에 대하여 시각적 감도는 특정 영역의 공간 주파수 영역에서 높은 감도를 보인다. 시각의 공간에 대한 반응은 충분히 높고 또한 낮은 공간 주파수에 대하여 무시할 정도의 값으로 떨어진다. 여기서 공간주파수(Spatial frequency)란 영상을 바라보는 눈의 시선이 접한 각도당 한 사이클의 사인파가 변하는 밝기 패턴의 수로 나타낸다[12, 13]. 따라서 공간 주파수에 대한 시각적 감도는 시스템의 물리적 특성에 따라 저역 통과 응답(Low-pass response)으로 모델링할 수 있고, 측면 금지(Lateral inhibition)의 특성을 고려하여 고역 통과 응답으로 모델링 한다[10]. 공간 주파수의 함수로서 블록내 영상의 밝기(Luminance)에 대한 감도측정이 실험되었으며, 특히 Mannos 등에 의해서 MTF(Modulation Transfer Function)의 응답이 공간 주파수 8 cycles/degree 정도에서 최고치를 나타냄을 보였다[14]. 공간 주파수는 화면으로부터 눈의 거리와 영상의 해상도에 밀접한 관계가 있다. 예를 들면 샘플링 주파수가 64 pels/degree이고 해상도가 512x512인 영상에 대하여 8x8블록의 가장 낮은 공간 주파수는 약 4 cycles/degree이다. 여기서 공간 주파수 개념으로 블록 효과(Blocking effect)가 두드러지게 눈에 띄는 현상을 설명할 수 있다. 이처럼 블록 영상에 대한 공간 주파수와 이것에 대한 시각적 감도를 구하여 정량적인 양자화 오차에 대하여 갖는 감도에 따라 가중치를 둘 수 있다[8,15]. 이것은 상위계층의 양자화 직전에 양자화 왜곡에 대한 감도율을 조정하는 가중치 요소로 작용되어 복원 영상의 화질을 개선하는 방향으로 영향을 미치게 된다.

### V. 실험 및 결과

제안한 비트율 계위 비디오 부호기의 시험을 위해서 H.263 비디오 부호기 표준을 기반으로 기본계층 비디오 부호기와 상위계층 비디오 부호기를 구성하였고, 기본계층 비디오 부호기에서의 움직임 예측 및 보상, 8x8 DCT변환과 같은 대부분의 부호화 알고리즘은 비디오 부호기 시험 모델을 따랐다[16].

시험 영상 데이터로는 Carphone, Miss America, Trevor 등 QCIF(Quadrature Common Intermediate Format) 크기의 YUV 데이터를 이용하였다.

본 논문의 주된 목적은 기존의 비디오 부호기 표준을 이용하여 비트율 계위 부호기를 구현하는 데 있었고, H.263 비디오 부호기 표준의 부호화 레이어를 기반으로 한 구분 구조 설계와 상위계층 비디오 부호기를 위한 인여 양자화기 설계에 초점을 맞추게 되었다. 따라서 본 논문에서 제안하는 부호기의 성능을 MPEG-2 등 기존 계위 비디오 부호기 관련 분야의 연구와 직접 비교하기에는 부호기의 목표 적용분야, 부호기가 지향하는 목표 비트율 범위, 그리고 입력 영상의 기본 해상도 등의 상이점 때문에 곤란한 점이 있으나 부호기 구조상 전송 효율 면에서는 MPEG-2의 SNR 계위 부호기와 유사하도록 구현하였으며, 계위를 사용하지 않는 부호기와 동일한 비트율에서 같은 화질을 유지하도록 설계하였다.

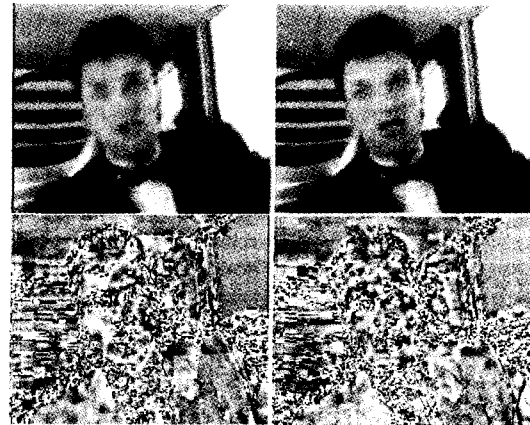
그림 5에서는 시험 영상의 실험 결과 영상을 보이고 있는데 각각 정해진 비트율(16, 20, 32, 64kbps)에 대한 기본계층 부호기의 복원 영상과 상위계층 부호기의 입력되는 양자화 오차 영상이다. 특별히 Carphone 시험 비디오 시퀀스의 147번째 프레임에 대하여 처리결과를 보이고 있다. 기본계층 부호기의 복원 영상의 화질은 예를 들어 영상 화의와 응용을 진행하기 위한 최소한의 화질은 유지되어야 하므로 비트율 면에서 약 16kbps 정도로 정하였다. 그림 5에서 보는 바와 같이 기본계층 부호기의 양자화 결과에 대한 오차를 보면 전적으로 랜덤 잡음이라기 보다는 원 영상에 대한 상관성을 상당히 포함하고 있는 것으로 볼 수 있다. 따라서 변환 및 양자화를 통해 시각적으로 민감한 오차를 발생시키며, 양자화에서 이 점이 고려되어야 함을 알 수 있다. 그림 6은 실험 영상에 대한 각 비트율별로 프레임 율을 도시하였으며, 비트율의 증가와 더불어 화질 개선과 프레임당 비트량이 증가하는 것은 당연한 결과로 볼 수 있다. 인트라모드 영상에서 매우 높은 비트율이 요구되며, 인터모드 영상에 대해서 비교적 작은 값의 범위에서 변화되고 있음을 볼 수 있다. 평균 비트율이 낮아 질수록 기본계층의 코딩 효율이 매우 낮아 짐을 볼 수 있다. H.263 비디오 부호기 표준에서는 비선형 양자화 행렬을 사용하지 않고 있으며 대신에 양자화 변수값을 조정하여 양자화기의 간격을 조정하고 비트율 제어를 수행할 수 있게 되어 있다[16]. 그러나 이 양자화 변수값의 변화 폭은 1에서 31로 제한되어 있어 이 변수만을 이용한 비트율 제어에는 한계가 있다. 그림 7에서는 각각 평균 비트율을 만족시키기 위한 비트율 제어 알고리즘으로 기본계층 부호기의 양자화 변수(QP: Quantization parameter)의 변화 모습을 볼 수 있다. 그림 7에서와 같이 높은 비트율에서의 비트율 제어에는 문제가 없으나 비트율이 낮아 질수록 양자화 변수 값의 변화만으로 평균 비트율을 만족시키지 못한다는 것을 알 수 있다. 특히 평균 비트율 16kbps 정도에서 양자화 변수 값의 포화 상태에 이르고 있음을 볼 수

있다. 그림 8에서는 각각의 평균 비트율에 대한 영상의 밝기(Luminance) 성분의 PSNR(Peak Signal-to-Noise Ratio)변화를 프레임별로 보여주고 있다. 기본계층 부호기의 PSNR은 평균 비트율 16kbps정도에서 28-30dB정도를 보이고 있다. 평균 비트율 별로는 약 5-6dB이상의 차이를 보이고 있다.

평균 비트율을 유지하기 위해서 부호화 버퍼 용량을 검사하여 QP값을 변수로 한 비트율 제어를 수행하게 되는데 평균 비트율이 16kbps인 경우를 살펴보면, QP값의 제약(1~31)으로 인하여 평균 비트율을 만족시키기까지 일정시간이 소요된다. 즉, 비디오 프레임을 부호화 하는동안 목표 비트율인 16kbps를 유지하려는 방향으로 계속해서 QP변화를 시도하지만 QP값은 그림 7에서와 같이 그 최대값(31)이 지속되게 되고, 충분한 프레임수의 부호화 후에 목표로 하는 평균비트율에 이르게 될 때 QP값은 정상범위로 회복하게 된다. 이때의 PSNR값은 그림8의 경우와 같이 얻게 되지만 이 값은 평균 비트율을 넘어선 값이 된다.

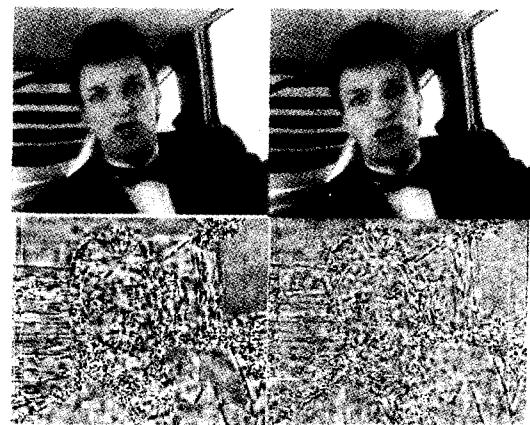
실험에서는 전체 평균 비트율을 약 32kbps정도를 유지하도록 하였고, 이 정도의 비트율을 기본계층 부호기와 상위계층 부호기에서 유지할 수 있도록 하였다. 비트율 계위 비디오 부호기에서는 그림8에서와 같이 평균 비트율의 변화에 따라 다양한 비디오 화질을 얻을 수 있을 뿐만 아니라 제한된 비트율로 비디오 데이터를 부호화하여 전송할 수 있다. 비트율 계위 비디오 부호기에서 비트율은 각 계층에 대해서 별도로 제어하게 된다. 상위계층 부호기는 보다 작은 양자화 간격을 이용하여 비트율을 높게 된다. 기본계층 부호기에서의 H.263 비디오 부호기 표준에 기반한 부호기에서는 양자화 변수 값의 두 배로 DCT변환 계수를 나누어 양자화 하게 되는데 이때 시각적 감도가 낮은 고주파수 영역에 대하여 큰 값을 생성하게 되므로, 기본계층의 양자화 오차에 대하여 비트율 제어 및 영상의 주관적 화질 개선의 관점에서 관련된MTF에 따라 가중치를 두어 양자화 한다. 그림 9는 각각의 부호기 계층에 대한 비트율을 비교하고 있다. 전체 비트율은 32kbps로 고정하였을 때의 기본계층 부호기의 프레임율을 보였고, 상위계층 부호기의 부호화 결과에 대한 프레임율의 변화를 도시하였다. 이때 비트율 계위 비디오 부호기에서 복원된 영상의 화질은 계위를 사용치 않는 비디오 부호기가 갖는 평균 비트율의 경우와 비슷하게 유지하고 있으며, 상위계층 부호기에 대한 부호화를 위한 부가정보가 비트율을 높이는 효과로 작용하게 된다.

이 추가 정보는 프레임 당 0.5kbps정도 이하가 유지되도록 하였다. 비트율 계위 비디오 부호기에서는 이러한 부가정보의 크기를 줄이는 문제도 전체 비트율 증가와 평균 비트율에 대한 복원 영상의 화질에 영향을 주게 되므로 결과적으로 중요한 문제가 된다고 할 수 있다.



(a) 16kbps/sec  
0.432kbts/frame  
28.60dB

(b) 20kbts/sec  
1.072kbts/frame  
30.27dB



(c) 32kbts/sec  
2.76kbts/frame  
33.87dB

(d) 64kbts/sec  
6.224kbts/frame  
37.42dB

그림 5. 시험 영상의 실험 결과 (Car phone 프레임#147; 평균비트율, 프레임당 비트량, Luminance 성분의 SNR)  
Fig. 5. Experimental results for test video sequence Car phone (Frame#1 47: Bitrates, Bits per frame, and SNR of luminance component)

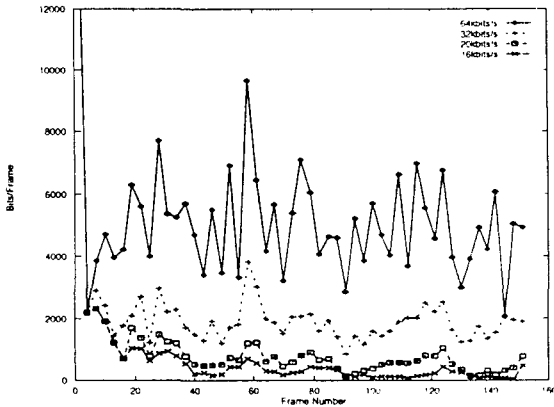


그림 6. 평균 비트율에 따른 프레임당 비트량 변화 (Car phone 영상, 프레임율: 10Hz)  
 Fig. 6. Difference of bits per frame at various average bitrates (Car phone video sequence, Framerate; 10Hz)

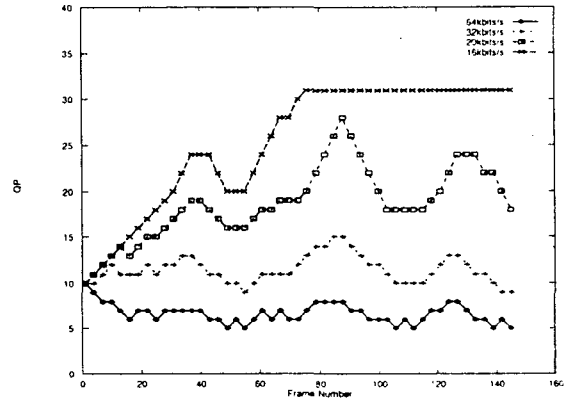


그림 7. 평균 비트율에 따른 프레임당 양자화변수값 (QP) 변화 (Car phone 영상, 프레임율: 10Hz)  
 Fig. 7. Difference of quantization parameters at various average bitrates (Car phone video sequence, Framerate: 10Hz)

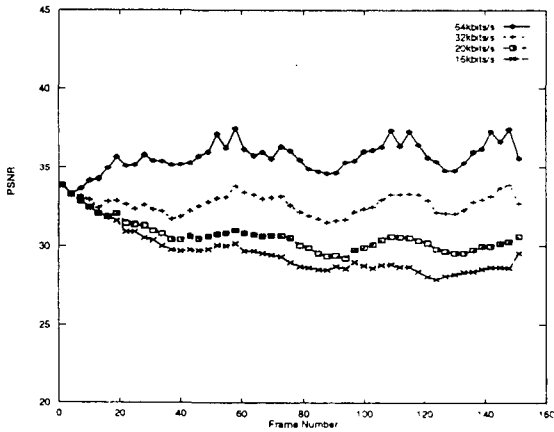


그림 8. 평균 비트율에 따른 프레임당 Luminance 성분의 Peak SNR 변화 (Car phone 영상, 프레임율: 10Hz)  
 Fig. 8. Difference of PSNR of luminance component at various average bitrates (Car phone video sequence, Framerate; 10Hz)

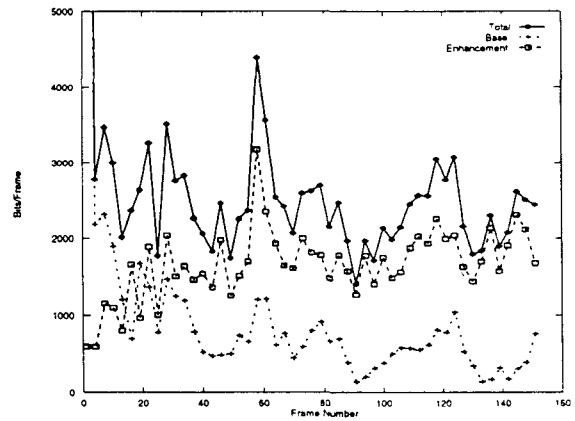


그림 9. 각 계층별 실험 결과에 대한 비트율 변화 (Car phone 영상, 프레임율: 10Hz, 기본계층의 평균비트율: 약 20 kbits/sec, 전체부호기의 평균비트율: 약 32 kbits/sec)  
 Fig. 9. Difference of bits per frame at each layer (Car phone sequence; Framerate: 10Hz; Average bitrates of the base layer: 20 kbits/sec, Average bitrates of the total encoder: 32 kbits/sec)

## VI. 결론

본 논문에서는 비트율 계위 비디오 부호기를 H.263 비디오 부호화 표준을 기반으로 구현하였다. 여기서 기본계층의 비디오 부호화에는 H.263 부호화 알고리즘을 따라 수행하도록 하는 비트율 계위 비디오 부호기 및 복호기의 구조를 제안하였고, 상위계층 부호기를 위한 비트스트림 구문을 정의하였으며, 상위계층 부호기의 입력이 되는 기

본계층의 부호화 결과 발생되는 양자화 오차에 대한 시각적 감도에 따른 양자화기를 설계하고 구현하였다.

전체 부호기의 구조를 단순화하기 위해 상위계층 부호기에 대한 패환을 기본계층 비디오 부호기의 움직임 예측을 위한 패환 루프에 포함 시키지 않았고, 복호기에서도 각 계층의 비디오 스트림에 대하여 독립적인 해당 역양자화 및 역변환을 수행한 후 공간 영역 데이터 레벨에서 더해지도록 하여 최종 고화질의 비디오를 복원할 수 있도록 하였다. 기본계층 비디오 부호기에서는 영상회의 등의 응



용에 사용할 수 있는 최소한의 비디오 화질을 유지할 정도의 약 16kbps 정도의 평균 비트율을 할당하게 했을 경우, 이때 발생하는 양자화 오차에 대하여 상위계층 비디오 부호기에서는 블록 영상의 밝기에 대한 대비와 해당 블록의 공간 주파수에 감도를 계산하여 양자화기 설계시 양자화 오차에 대한 감도 성분만큼의 가중치로 작용하도록 설계하였다. 전체 부호화기에서 복원하는 영상의 화질은 평균 비트율을 고정하였을 때 계위를 사용하지 않은 비디오 부호기와 비슷한 화질을 유지하도록 하였으며 이때 비트율 계위 비디오 부호기가 갖는 추가 정보는 각 프레임당 평균 0.5kbps를 넘지 않도록 하였다.

제안한 비트율 계위 비디오 부호기를 이용하면 시스템의 처리 속도나 비디오 스트림을 전송하는 채널의 대역폭 및 네트워크 트래픽 증가 등으로 상위계층의 비디오 스트림의 전송이 어려운 경우에도 기본계층의 비디오 복원만으로 끊김없는 영상회의 서비스를 진행할 수 있는 H.323 기반 멀티미디어 터미널이 가능하다.

## 참고 문헌

- [1] K. R. Rao and J. J. Hwang, *Techniques and standards for image, video, and audio coding*, pp. 307-316, Prentice-Hall, pp. 307-316, 1996.
- [2] ISO/IEC 13818-2 | ITU-T H.262, *Generic coding of moving pictures and associated audio information: Video, First Edition*, pp. 87-110, May 15, 1996.
- [3] M. Ghanbari, "An adapted H.261 two-layer video codec for ATM networks," *IEEE Trans. on Comm.*, vol. 40, pp. 1481-1490, Sept. 1992.
- [4] J. De Lameillieure, "Rate-distortion optimal thresholding in SNR scalability based on 2D dynamic programming," *SPIE Dig. Comp. Tech. and Sys. for Video Comm.*, vol. 2952, pp. 689-698, Oct. 1996.
- [5] R. Aravind, M. Reha Civanlar, and Amy R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. on C & S for Video Technology*, vol. 6, pp. 426-435, Oct. 1996.
- [6] Draft ITU-T Recommendation H.323V2, *Visual telephone system and equipment for LANs which provide a non-guaranteed quality of service, Annex B: Procedures for Layered Video Codecs*, pp. 76-83, Feb. 27, 1997.
- [7] Draft ITU-T Recommendation H.263, *Video coding for low bit-rate communication*, May 2, 1996.
- [8] 이범렬 외 2명, "시각적 Intra코딩에 의한 동영상 부호기," 제9회 신호처리합동학술대회, 제9권 2호, pp. 1301-1304, 1996.
- [9] C. H. Chou, "Adaptive transform coding of images based on removing just noticeable distortion," *SPIE Visual Com. Im. Proc.*, vol. 2501, Taipei, Taiwan, pp. 607-618, May 1995.
- [10] R. J. Clarke, *Transform coding of images, Chapter 6. Human Visual Response*, Academic Press, pp. 259-290, 1985.
- [11] D. Shen and S. Wang, "Measurements of JND property of HVS and its applications to image segmentation, coding and requantization," *SPIE Dig. Comp. Tech. Sys. for Video Com.*, Berlin, FRG, vol. 2952, pp. 113-121, Oct. 1996.
- [12] J. G. Robson, "Spatial and temporal contrast-sensitivity functions of the visual system," *J. Opt. Soc. of Am.*, vol. 56, pp. 1141-1142, 1966.
- [13] C. F. Hall and E. L. Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Trans. on Sys., Man., and Cyber.*, vol. SMC-7, pp. 161-170, Mar. 1977.
- [14] J. L. Mannos and D. J. Sakrison, "The effects of a fidelity criterion on the encoding of images," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 525-536, Jul. 1974.
- [15] B. R. Lee, et al., "H.263-based SNR scalable video codec," ICCE'97, THPM13.6, pp. 254-255, Jun. 1997.
- [16] Telenor Research, *Video Codec Test Model(TMN5)*, Jan. 31, 1995.

## 저 자 소 개



### 이 범 렬

1987년 전북대학교 전자공학과 졸업(학사)  
 1989년 전북대학교 대학원 전자공학 석사  
 현재 한국전자통신연구원 인공지능연구실 선임연구원  
 주관심분야 : 영상 통신, 무선 멀티미디어 데이터 전송



**임 성 호**

1986년 전북대학교 전자공학과 졸업(학사)  
1988년 전북대학교 대학원 전자공학 석사  
현 재 한국전자통신연구원 인공지능연구실 선임연구원  
주관심분야 : 멀티미디어 시스템, 패턴인식



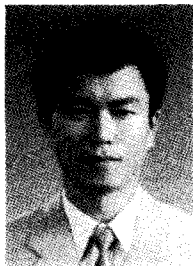
**민 병 의**

1982년 한양대학교 전기공학과 졸업(학사)  
1984년 한국과학기술원 전기 및 전자공학과 석사  
1993년 한국과학기술원 전기 및 전자공학과 박사  
현 재 한국전자통신연구원 인공지능연구실장  
주관심분야 : 휴대멀티미디어 단말, 분산가상현실, 지능형 에이전트



**황 승 구**

1979년 서울대학교 전기공학과 졸업(학사)  
1981년 서울대학교 전기공학과 졸업(석사)  
1986년 University of Florida 전기공학과(박사)  
현 재 한국전자통신연구원 멀티미디어 연구부장  
1994년 ~ 1995년 미국 SRI International(International Fellow)  
주관심분야 : 멀티미디어 시스템, HCI, 로보틱스



**황 재 정**

1983년 전북대학교 전자공학과 졸업(학사)  
1992년 전북대학교 대학원 전자공학 박사  
1983년 ~ 1987년 전주문화방송(주) 기술국  
현 재 군산대학교 전파공학과 부교수  
1990년 ~ 1991년 한독기술협력사업 파독교수  
1993년 ~ 1994년 텍사스주립대학교 객원교수  
현 재 디지털방송추진 TV분과위원  
주관심분야 : 영상통신, 방송시스템, 멀티미디어 전송