

영상 전화 시스템에서의 신호처리기술

金 漢 樹
高等技術研究院 情報通信研究室

I. 서론

최근 들어 종합정보통신망(ISDN: Integrated Services Digital Network)의 보급이 활성화 되면서 통신 분야에 있어서도 기존의 음성 통신 서비스 이외에 영상과 데이터를 함께 통신할 수 있는 새로운 서비스에 대한 요구가 증대되고 있다. 이미 선진 각국에서는 ISDN의 보급 확대를 위해 다양한 노력을 진행중에 있으며, 국내에서도 93년 8월부터는 현재의 시험 서비스를 상용 서비스화 할 예정이다.

ISDN의 성공적인 확산에 가장 큰 영향을 주는 것 중 하나는 기존의 전화 통신망(PSTN: Public Switched Telephone Network)에서의 통신 서비스와 차별화 되는 새로운 양질의 서비스 제공에 있으며, 그 대표적 예가 바로 동영상 전화기이다. 즉, 음성과 함께 실시간으로 동영상을 통신할 수 있는 ISDN 용 영상 전화는 기존의 PSTN 용 전화나 정지화 전화, 그리고 최근 상품화가 이루어지고 있는 AT&T나 마르코니 사의 PSTN 동영상 전화에 비해 확실히 차별화 되는 통신 단말이라고 할 수 있다.

국제 전신 전화 자문 위원회(CCITT)의 권고안 H.320으로 대표되는 시청각 통신 단말 시스템의 구조는 표1과 같다.

표 1에서의 권고안에 따르는 영상전화 시스템의 구조는 그림 1과 같다. 영상전화 시스템에서 영상 신호 압축 부분은 H.261^[1]로 권고되어 있으며 영상 신호 처리의 관점에서 볼 때 합성 또는 성분별 애널로그 영상 신호와 성분별 디지털 영상 신호간의 변환등이 포함된 전/후처리 과정, 디지털 영상 데이터 압축(compression)/신장(decompression) 과정, 그리고

표 1. 영상전화 시스템 관련 CCITT 표준안

권고안	내 용
H.320	협대역 시청각 시스템의 기술 규격
H.242	호 설정 및 해제
H.230	제어 및 표시용 신호
H.221	멀티미디어 정보의 다중화/역다중화
H.261	영상부호화
G.711	64 kbps PCM 오디오
G.722	48, 56, 64 kbps ADPCM 오디오
G.728	16kbps 오디오
I 시리즈	단말과 망 간의 인터페이스

압축된 영상 신호의 전송을 위한 가변 길이 부호화/복호화(variable length coding/decoding) 및 오류 정정 부호화(Error Correction Coding)등의 세 부분으로 나뉜다. 영상전화 시스템에서 사용되는 영상 신호 처리 기술은 CCITT에서 H.261로 표준 권고된 아래 JPEG^[2], MPEG^[3] 등 거의 모든 표준안

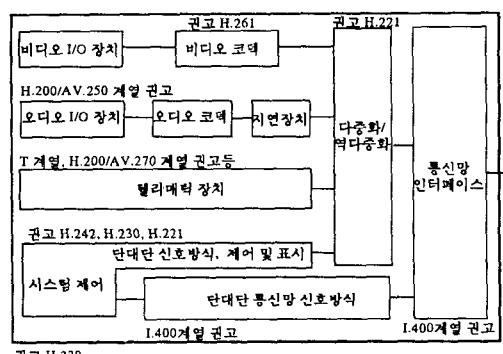


그림 1. H.320 시청각 단말의 구조

의 핵심 기술로 적용되고 있다. 본 기고에서는 이들 중 실제 영상 신호를 다루는 부분인 전/후처리 및 압축/신장 부분에 대해 논하기로 한다.

II. 영상 신호 처리 및 H.261의 동작 방식

1. 영상 신호 처리

영상 신호 처리 기술은 애널로그 신호처리 기술과 디지털 신호처리 기술 등 크게 두 가지로 나뉜다. 1920년대에 처음 개발된 애널로그 영상 신호 처리 기술은 실세계의 연속된 영상을 다루기 때문에 full-color, full-motion 영상 처리에 유용하였으며 따라서 애널로그 영상 처리 기술은 TV 등과 같은 단방향 방송 분야에 주로 사용되어 왔다. 이와 반대로 컴퓨터 그래픽스 분야에서 먼저 시작된 디지털 영상 처리는 애널로그 신호에 비해 저장이나 편집, 조작 등이 훨씬 용이하였으며 이러한 이점을 디지털 영상 신호에도 적용하려는 시도가 행해졌다. 그러나 애널로그 영상을 디지털화 하면 데이터량이 엄청나게 늘어나기 때문에 컴퓨터의 성능을 향상시키거나, 저장용량 또는 전송 용량을 늘려야 하는 단점이 있었다.

영상 데이터를 압축하는 목적은 원래의 영상이 전달하고자 하는 의미를 손상시키지 않는 범위내에서 가능한 한 적은 양의 정보로 표현하는 데 있다. 영상 데이터를 압축하는 방식은 크게 두 가지로 나뉜다. 첫번째 방법은 동일한 정보의 표현 방식을 좀 더 효과적인 방식으로 하는 것으로서, 예를 들면 원래 데이터 세트보다 적은 저장 공간을 차지하는 새로운 코드를 만드는 것을 들 수 있다. 이와 같은 압축 방식을 비손실(lossless) 부호화라 한다. 비손실 방식에서는 어떤 코드를 새로운 코드로 대치하는 형식이기 때문에 메시지가 차지하는 공간이 줄어들기는 하지만 손실은 전혀 발생하지 않는다.

두번째 방법은 손실(lossy) 부호화 방식이다. 이 때에는 데이터 세트의 일부를 선택적으로 버리거나 변경시키는 방식을 택한다. 최근의 영상 데이터 압축의 경우에는 압축률을 높이기 위해 손실 부호화와 비손실 부호화를 병행하여 사용한다. 그런데 영상 신호는 그 자체의 특성을 잘 이용하면 압축/신장 후의 화질을 좋게 유지하면서도 상대적으로 높은 압축률을 얻을 수 있다.

첫째로, 대부분의 영상은 중복도가 매우 크다. 이

러한 중복성은 한 영상 내에서 뿐만 아니라 동영상인 경우에는 영상과 영상 사이에도 존재한다. 극단적인 예를 들면 배경 부분은 전혀 변화가 없고 화면내의 물체만 움직이는 경우를 들 수 있다. 둘째로 사람의 시각 인식 기능은 색상 보다는 휘도 영상 신호에 더욱 민감하며, 화면 내용이 복잡한 부분보다는 복잡하지 않은 부분에 더 민감하다. 따라서 영상 데이터가 가지고 있는 중복성과 인간의 시각 인식 기능이 가지고 있는 특징을 잘 이용하면 실제 영상에서의 화질 저하 없이도 상당량의 데이터 감축이 가능해 진다.

2. H.261의 동작 방식

CCITT 권고안 H.261은 영상 전화나 영상 회의에 응용할 수 있는 영상 데이터 압축 기법에 대해 규정하고 있다. 이 권고안은 pX64 방식이라고도 불리는 데 그 이유는 통신 채널 상에서의 데이터 전송 속도가 64 Kbps의 p 배 ($p=1\sim32$)이기 때문이다. $p=1$ 인 경우에는 저화질의 영상 전화용 비디오 신호가 64Kbps 선로로 전송되며 $p=32$ 인 경우에는 고화질의 영상 회의용 비디오 신호가 2.048Mbps 선로로 전송된다. 아울러 H.261 권고안은 전송되어 온 비트 스트림의 구성과 그 해석 방법 등을 규정하고 있기 때문에 상이한 업체에서 만든 제품이더라도 이 권고안만족하면 상호간의 통신이 가능해 진다.

그림 2와 그림 3은 H.261에 따른 영상 신호의 부호화기와 복호화기의 블록도이다. CCITT 부호화기는 복합형 부호화기(hybrid encoder)라고도 불리는데 그 이유는 변환 부호화(transform coding)와 예측 부호화(predictive coding)가 서로 복합된 형태이기 때문이다. 아울러 현재 영상과 재생 영상간의 단순한 차이만을 예측 부호화 하는 이외에 운동 보상(motion compensation)을 옵션으로 가지고 있어 연속 영상간 이동 추적 과정을 통해 예측형 부호화기의 효율이 향상되며 결과적으로 전송 데이터의 양이 감소된다.

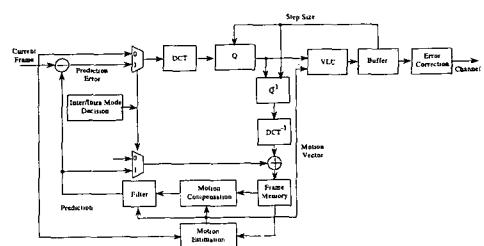


그림 2. CCITT 권고 H.261의 부호화기

III. H.261에서의 영상 처리 및 압축 기법

1. 전/후처리 과정

영상 전화 시스템의 주요 특징중의 하나는 부호화에 사용되는 디지를 영상 신호에 대해 세계 공통의 표준을 제시하고 있다는 점이다. 즉, 실제 부호화에 쓰이는 디지를 영상 신호 규격은 그림 4 처럼 CIF (Common Intermediate Format) 또는 QCIF (Quater CIF)의 형태를 한다. 따라서 영상 신호의 전/후처리 과정은 NTSC, PAL, SECAM 등과 같은 애널로그 TV 표준 신호와 CIF 또는 QCIF 디지털 비디오 신호간의 상호 변환에 해당한다. 이러한 상호 변환 과정에 포함되는 필터링, 컬러 성분 분리등은 수년전까지만 해도 애널로그 영역에서 행해졌으나 최근 VLSI의 발전에 따라 디지를 영역에서 컬러 성분 분리나 필터링 등을 용이하게 수행할 수 있게 되었다. [그림 5 참조]

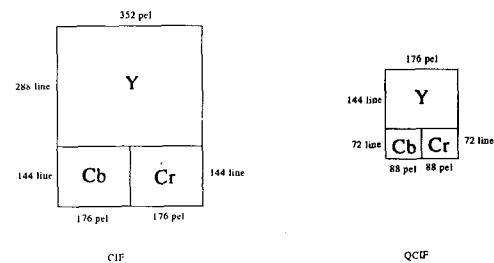


그림 4. 영상전화 시스템에서의 영상 형태

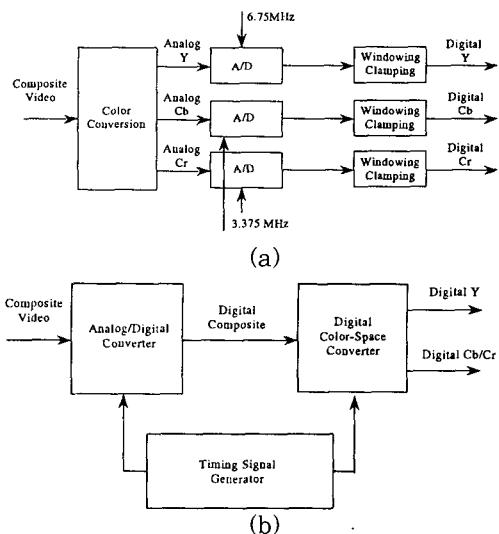


그림 5. 영상 신호의 전처리 과정

1) 시공간 표본화

H.261 영상 데이터 부호화기는 순차 주사 방식 (progressive scanning)의 29.97 Hz 영상을 대상으로 부호화를 행한다. 영상은 휘도(luminance) 영상 Y와 두 개의 색차(color difference) 영상 Cr 및 Cb로 구성되는 성분별 영상(component image)이며 상세한 규격은 CCIR 권고안 601^[4]에 따른다.

black: 16

white: 235

zero color difference: 128

peak color difference: 16 및 240

이들 값은 공칭값(nominal value)이며 실제 부호화기 자체는 1 ~ 254의 디지털 화소 값에 대해 동작을 해야 한다.

2) Digital Scaling

표본화가 완료된 디지털 성분 신호는 H.261 부호화기의 규격에 맞도록 scaling^[5] 행해져야 한다. 뿐만 아니라 수직 방향으로의 재표본화(resampling)가 이루어져야 한다.

H.261에서의 영상 포맷은 두 종류가 있는데 첫번째 포맷은 CIF(Common Intermediate Format)로서 휘도 영상은 352 화소 X 288 주사선, 색차 영상은 176 화소 X 144 주사선으로 구성되며 각 화소는 직교 배열(orthogonal arrangement)을 가지고

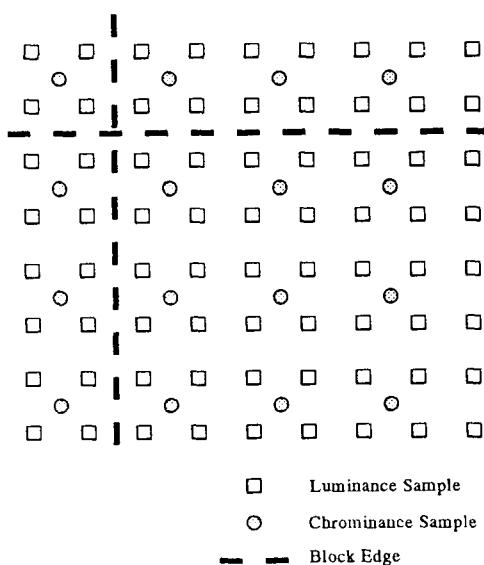


그림 6. 휘도 화소와 색차 화소의 위치 관계

있다. 아울러 색차 화소는 그림 5와 같이 색차 화소의 8X8 블록의 경계가 휘도 화소의 16X16 블록 경계와 일치하는 구조를 가진다. 이와 같은 영상을 얻기 위해서는 525 또는 625 주사선 영상의 휘도와 색차 신호를 각각 6.75 MHz 및 3.375 MHz로 표본화 하면 되며, 이들 주파수는 CCIR 601에서의 표본화 주파수 13.5 MHz 및 6.75 MHz의 1/2에 해당한다.

또 하나의 영상 포맷은 QCIF(Quater-CIF)로서 CIF 영상에 비해 화소수 및 주사선수가 각각 1/2인 영상 포맷이다. 이러한 두 가지 영상 포맷은 각 지역의 TV 표준에 관계없는 세계 공통의 규격이기 때문에 별도의 포맷 변환이 필요없게 된다. 아울러 영상 전화 시스템은 실제 전송한 영상들 사이에 있는 비전송 영상의 갯수를 0, 1, 2, 3으로 설정하여 최대 영상 전송률을 제한하는 방법과 CIF 및 QCIF를 선택하는 외부 수단을 갖추고 있어야 한다. [H.221]

2. 영상 데이터 압축

전처리 과정을 거친 디지털 영상 데이터는 H.261의 권고안에 따라 Picture, GOB(Group of Blocks), MB(Macroblock), B(Block)의 계층 구조를 가지면서 비디오 부호화기로 입력된다. 비디오 부호화기의 주된 기능은 영상 신호에 들어 있는 시공간 중복성(spatio-temporal redundancy)을 효과적으로 제거하는 것이다. 영상 전화 시스템에서는 이를 위해 두 가지 기법을 적용하고 있는데 먼저 시간적, 즉 연속된 영상 간의 중복성을 제거하기 위해 운동 보상형 예측(MCP: Motion-Compensated Prediction)을 사용하며, 시간적 중복성이 제거된 오차 영역의 데이터에 대해 이산 여현 변환 부호화(DCTC: Discrete Cosine Transform Coding)를 적용하고 있다.

데이터 압축 기법은 크게 에너지 감축(energy reduction) 기법과 에너지 집중(energy compaction) 기법으로 나뉘며, 전자의 대표적인 예는 예측형 부호화이고 후자의 대표적인 예는 변환 부호화이다. 이 두 가지 부호화 기법은 모두 영상 데이터가 시공간적으로 높은 상관성(correlation)을 가지는 성질을 이용한 것으로서 예측 부호화의 경우 전송 대상이 되는 예측 오차 신호의 에너지가 원래 신호의 에너지에 비해 감소하는 성질을 이용하며, 변환 부호화의 경우에는 원래의 영상 신호 전체에 분포되어 있는 신

호에너지가 변환을 통해 특정 부분으로 집중되는 성질을 이용한 것이다.

1) 운동 보상형 프레임간 예측

디지털 영상 데이터의 중복성 제거에 유용한 방법 중 하나가 예측을 이용한 방식이며 영상 전화 시스템에서는 특히 시간 방향, 즉 연속된 영상간의 중복성 제거를 위해 프레임간 예측을 사용한다. 통상적으로 영상 전화와 같이 화면 절환(scene change)이 없고 부호화 대상이 인물의 어깨선 이상으로 제한되는 head-and-shoulder 영상인 경우, 연속 영상간의 신호 변화는 대부분 영상 내에 포함되어 있는 물체의 운동에 기인한다. 따라서 예측에 앞서 물체의 운동 정보를 추정하여, 추정된 운동 정보를 영상간 예측에 사용하면 예측 효율을 더욱 높일 수 있게 된다. 이러한 기법을 운동 보상형 예측(MCIP: Motion-Compensated Interframe Prediction)이라 한다.

이때 운동 추정은 각 화소 단위로 추정을 행하는 화소 순환 기법(PRA: Pel Recursive Algorithm)과 블럭 단위로 추정을 행하는 블럭 정합 기법(BMA: Block Matching Algorithm)으로 대별되며, H.261에서는 앞서의 전처리 과정에서 얻어진 16X16 크기의 매크로블럭 단위로 블럭 정합 기법을 적용하여 매크로블럭별 운동 정보를 추정한다.

블럭 정합 기법은 연속된 두 장의 영상 중에서 영상 신호를 블럭 단위로 비교, 신호 유형(pattern)의 정합도에 기초하여 운동을 추정하는 것으로서, 현재 영상 블럭과 기준 영상내의 특정 탐색 영역내의 모든 블럭과 패턴 정합(pattern matching)을 행하기 때문에 탐색 영역 내로 제한된 운동 즉, 블럭의 실제 최대 운동이 탐색 영역을 벗어나지 않는 경우에 대해서는 매우 높은 정확도의 운동 정보가 찾아진다. 따라서 신호 패턴이 서로 중복되지 않고 특정 탐색 영역에 제한되어 있는 경우 매우 좋은 성능을 보인다. 이와 반대로 신호 패턴이 중복되거나 실제 운동이 매우 커서 탐색 영역을 벗어나는 경우에는 추정의 정확도도 떨어진다.

블럭별 운동 추정 방식을 그림 7에 나타내었다. 현재 입력되는 영상의 16 X 16 블럭의 중심 좌표를 (P_x, P_y)이라 하고 연속 영상 사이에서 한 블럭의 최대 이동 범위를 (mx, my)라 하자. 그러면 그림 7에서 현재 블럭의 운동을 추정하기 위해서는 먼저 기준 영상 내에 최대 이동 범위 만큼의 탐색 영역(Search Area)을 설정한 후 이 탐색 영역 내의 모든 블럭에

대해 현재 블럭과의 정합을 수행, 정합도가 최대인 블럭을 현재 블럭의 예측에 사용한다. 이때 정합도의 판단은 MAE(Mean Absolute Error) 또는 MSE(Mean Square Error)가 최소인지에 따라 결정된다. 운동 추정이 끝난 후에는 추정된 운동 벡터를 이용하여 운동 보상형 예측이 행해진다.

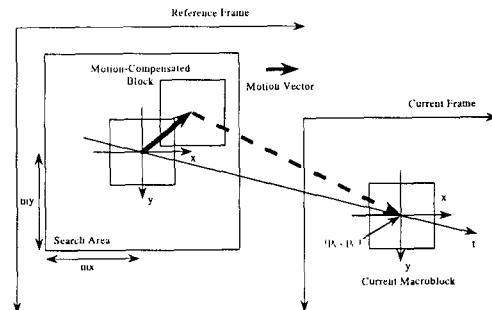


그림 7. 블럭 정합 방식의 운동 추정

이상과 같은 운동 보상 기법은 H.261 부호화기에 서는 옵션 사항이지만 복호화기는 한 매크로 블럭당 한개씩의 운동 벡터를 복호화 할 수 있어야 한다. 아울러 운동 벡터의 크기는 +/- 15 이하의 정수값을 가진다. 한편 한 매크로 블럭 내의 6개 블럭중 휘도 블럭 4개에 대해서는 매크로 블럭 단위의 운동 벡터가 그대로 적용되고 색차영상 블럭 2개에 대해서는 휘도 블럭의 운동 벡터를 1/2하고 소수점 이하를 버려 정수화 시킨 벡터가 사용된다.

초기의 블럭 정합 기법 관련 연구는 정합에 사용되는 계산량을 줄이기 위해 탐색 영역 내에서 탐색 위치를 효율적으로 선택해 내는 데 집중^[5] 되었으며 Jain & Jain, Koga, Ninomoya, Rao등의 고속 알고리즘이 이와 같은 범주에 속한다. 고속 알고리즘은 운동 보상 효과의 컴퓨터 모의 실험이나 DSP 칩 응용 하드웨어 개발에 주로 사용되었다. 그러나 최근의 하드웨어에서는 VLSI 기술의 발전에 힘입어 탐색 영역 내의 모든 블럭에 대해 정합을 행하는 전역 탐색(full search, exhaustive search, or brute-force search) 기법이 적용되고 있다. 탐색 위치의 효율적 선택뿐만 아니라 실제 정합도의 계산에 필요한 계산량을 줄이기 위한 연구도 진행된 바 있다. 이러한 방식으로는 정합도 산출시 서브샘플링을 행하여 계산량을 1/4로 줄이거나 또는 정합도에 어떤 임계치를 두어 정합도가 임계치 이하로 떨어지면 더 이상의

계산을 중단하고 다음번 후보 블럭에 대한 정합도를 계산해 나가는 방식이다.

2) 루프 필터

H.261에서는 운동 보상된 매크로블럭에 대해 각각의 8X8 블럭 단위로 2차원 공간 필터(2D spatial filter)를 적용하며, 이 필터가 프레임간 DPCM 루프 내에 있기 때문에 루프 필터(loop filter)라고 부른다. 루프 필터는 블럭 내부에서의 계수는 [1/4, 1/2, 1/4] 블럭 경계에서는 [0, 1, 0]인 separable non-recursive 필터이며 출력은 8비트 정수 값으로 반올림(round-off) 된다. 이 필터는 매크로블럭 내의 부호화 방법(mtype)에 따라 한 매크로블럭 내에서 블럭 단위로 on/off 된다.

3) 변환 부호화

영상 데이터 압축에 널리 쓰이는 이산여현변환은 다음과 같다.

$$F(u, v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{\pi(2x+1)u}{16} \cos \frac{\pi(2y+1)v}{16}$$

$$f(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos \frac{\pi(2x+1)u}{16} \cos \frac{\pi(2y+1)v}{16}$$

$$u, v, x, y = 0, 1, 2, \dots, 7$$

where

x, y = spatial coordinates in the pel domain

u, v = coordinates in the transform domain

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u, v = 0 \\ 1 & \text{otherwise} \end{cases}$$

변환을 한 후 변환 영역(transform domain)에서의 특징은 값이 큰 계수의 대부분이 저주파 영역에 집중된다는 점이며 이는 신호 에너지의 대부분이 저주파 영역에 집중됨을 의미한다. 앞서 언급한 바와 같이 인간의 시각 기능은 고주파 신호보다는 저주파 신호에 더 민감하므로 저주파 영역을 중점적으로 전송할 필요가 있다. 따라서 양자화를 거친 후 에너지가 일정 크기 이상인 계수들만을 선택적으로 부호화함으로써 전송 데이터의 양을 줄일 수 있다.

4) 양자화 및 가변 길이 부호화

양자화(quantization)란 어떤 값을 근사화 하는 것을 의미하며, 영상 데이터 압축시 손실(lossy) 부호화에 해당한다. H.261에서는 부호화 모드에 따라 원영상 블럭이나 예측 오차 블럭을 변환한 변환 계수에 대해 양자화를 행한다. 이때 양자화기의 구조는 전송 버퍼의 충만도(buffer fullness)로부터 결정된

step-size에 따라 변한다.

H.261에는 INTRA DC 계수에 대해 하나의 양자화기를 사용하며 나머지 계수들에 대해서는 31 종류 중 하나의 양자화기가 사용된다. 양자화기의 변경은 매크로블럭 단위로 행해지지만 INTRA DC에 대해서는 항상 step-size가 8이고 dead-zone이 없는 선형 양자화기(linear quantizer)가 사용된다. 나머지 양자화기들은 0 부근에 dead-zone이 있고 step-size가 2에서 62 까지의 짹수값을 가지는 선형 양자화기의 형태를 한다.

가변 길이 부호화는 양자화된 계수들을 좀 더 효율적으로 전송하기 위해 적용되는데 우선 zigzag scanning을 하여 가능한 한 많은 0이 연속되도록 한 후 (zero-run, non-zero level)을 토큰으로 하는 2D Huffman 부호화를 행한다.

IV. 맺음말

영상 전화 시스템에서의 영상 신호 처리는 영상 전/후 처리 과정과 영상 데이터의 압축/신장의 두 부분으로 크게 나눌 수 있다. 전/후 처리 과정은 영상 압축부인 H.261에서 요구하는 디지털 영상 규격과 애널로그 TV 신호 간의 상호 변환 부분에 해당한다. 가장 커다란 특징은 각국의 TV 표준 방식에 관계 없이 부호화에서는 전세계 단일 규격의 영상 데이터만을 처리한다는 점이다.

영상 데이터 압축/신장부에서는 영상 데이터의 상관성을 제거하는 역할을 한다. 이를 위해 운동 보상형 부호화 기법과 변환 부호화가 결합된 복합 부호화 방식이 적용된다. 아울러 압축된 영상 데이터를 효과적으로 전송하기 위해 비손실 부호화인 가변길이 부호화/복호화 기법도 적용된다.

이상과 같은 영상 전화 시스템용 영상 신호 처리 기술은 CCITT 권고 이후 여러 종류의 영상 부호화 표준에 핵심 알고리듬으로 사용되고 있으며 향후 그 활용 영역이 더욱 확대될 전망이다.

参考文獻

- [1] "Video codec for audio visual services at

- px64 Kbits/s." CCITT Recommendation H.261, 1990.
- [2] "JPEG digital compression and coding of continuous-tone still images," ISO-IEC/JTC1 Committee Draft 10918, 1991.
- [3] Motion Picture Expert Group, "MPEG video simulation model three (SM3)," Doc. ISO/IEC/JTC1/SC2/WG8/MPEG90, July 1990.
- [4] CCIR Recommendation 601, "Encoding Parameters of Digital Television for Studios," in: CCIR Recommendation and Reports, vol. XI, ITU, Geneva, Switzerland, 1982.
- [5] H.G. Musmann, P. Pirsch and H.J. Grallert, "Advances in picture coding," *Proc IEEE*, vol. 73, April 1985, pp. 523-548. 

筆者紹介



金漢樹

1960年 9月 19日生

1983年 2月 고려대학교 공과대학 전자공학과 (학사)

1985年 2月 한국과학기술원 전기 및 전자공학과 (석사)

1993年 2月 한국과학기술원 전기 및 전자공학과 (박사)

1991年 2月 ~ 현재 대우전자 영상연구소 선임연구원

1993年 1月 ~ 현재 고등기술연구원 정보통신연구실 선임연구원

주관심분야 : 영상 부호화, 영상 통신 시스템