

# 낮은 비트율 동영상 부호기의 율-왜곡 특성

정희원 황재정\*, 지석상\*\*, 허영\*\*\*

## Rate-Distortion Characteristics in Low Bit-rate Video Coder

Hwang, Jae-Jeong\*, Jee, Seok-Sang\*\*, and Huh, Young\*\*\* *Regular Members*

### 요약

전송과정에서 발생하는 왜곡량에 따라 전송률의 하한이 결정되는 율-왜곡 이론은 시각적으로 민감한 부분을 왜곡없이 부호화하여 전송하는 영상 시스템에서 기본이 되는 중요한 요소이다. 율-왜곡 이론은 정보량의 개념으로부터 시작되어 원 신호의 확률분포와 왜곡의 측정기준에 의해 결정되는데, 이 논문에서는 가우시안과 라플라시안 분포함수에 걸쳐 오차기준과 자승 오차기준을 적용하여 율-왜곡 함수를 각각 구하였다. 나아가서 저전송률 부호기로 개발된 H.263 부호기에 이 함수를 적용하여 분석하였다. 비교를 위해 자승 오차기준에 의한 이론치와 실제 측정치를 제시하였다. H.263 부호기는 엔트로피 부호화, 부호화 블록패턴 등 다양한 기법을 사용하여 율-왜곡 함수에 의한 이론치보다 주어진 MSE에서 정규화 비트율이 최대 0.55만큼 더 낮은 전송률을 얻을 수 있었다.

### ABSTRACT

Rate-distortion theory means that the lower limit of transmission rate is determined according to the distortion quantity. This theory is fundamentally important factor in the visual system without visually sensitive distortion. Rate-distortion theory originating in the notion of information quantity is derived using probability distribution of original signal and measurement of distortion. In this paper, we develop the R(D) function with the distribution function of Gaussian and Laplacian, using the absolute error criteria and the square error criteria. Then we applied these functions to H.263 video coder at very low bit-rates. The H.263 coder presents lower bit-rates than the theoretical bounds by the rate-distortion theory by means of various techniques such as coded block pattern and entropy coding.

### 1. 서론

저전송률 비디오 부호기에서 흔히 사용하는 영상 포맷은 CIF 또는 QCIF이다. QCIF는 176×144화소 해상도를 가지며 색성분을 포함하여 각각 8비트로 양자화하고 초당 30프레임을 전송한다고 가정하면 전체 소요 전송률은 176×144×24×30=18.3Mbits/sec이다. 한편 사논의 이론적인 채널용량은  $C = W \log_2(1 + SNR)$ 이며<sup>[1]</sup>, W를 3KHz로 하고 SNR=30dB의 환경에서 PSTN상 얻을 수 있는 최대 전송률은 28.8Kbits/sec이다. 따라서 QCIF 영상시퀀

스의 정보량은 전송 가능한 PSTN 용량보다 634배가 된다. 즉 이 만큼의 압축 전송이 요구된다.

H.263은 이 요구를 수용하기 위해 다양한 기술을 이용한다<sup>[2,3]</sup>. 우선 대부분의 동영상 부호기와 마찬가지로 DCT에 의한 공간적 상관도 제거와 프레임 간 예측에 의한 시간적 상관도 제거 기법을 이용한 혼합부호기이다. QCIF 영상의 정보량을 줄이기 위해 보다 적은 수의 프레임을 전송할 수 있다. 줄어든 프레임율은 바로 정보량의 압축과 비례하지만 시간 해상도가 줄어들므로 적절한 방법이라 말할 수는 없다. 공간해상도를 줄이는 것도 같은 의미에서 바람직하지 않다. 따라서 양자화를 거칠게 하는

\* 군산대학교 전자정보공학부(hwang@kunsan.ac.kr),

\*\* 전북대학교 전자공학과,

\*\*\* 한국전기연구소 영상응용그룹(yhuh@keri.re.kr)

논문번호 : 000008-1128, 접수일자 : 2000년 11월 28일

\* 본 논문은 첨단정보기술연구센터(AITrc)를 통하여 한국과학재단으로부터 지원을 받았다.

것만이 효율적인데, 양자화 간격 감소와 함께 오차가 증가하므로 비트율과 오차간의 관계가 중요한 문제로 대두된다<sup>[4,5,6]</sup>.

이 논문에서는 울-왜곡 이론[7,8]을 발전시켜 가우시안 및 라플라시안 분포함수의 경우에 적절한 함수를 유도하고자 한다. 왜곡측정 기준으로서 절대치 오차 기준과 자승 오차 기준을 살펴보았다. 유도된 울-왜곡 함수를 저전송률 부호기로서 개발된 H.263 부호기에 적용하여 부호기의 성능을 울-왜곡 함수의 측면에서 분석하였다.

## II. 울-왜곡 함수

채널용량이 제한된 경우 울과 왜곡의 관계는 반비례적이다. 더 많은 정보를 전송하면 더 좋은 화질을 얻게 된다. 예를 들어 10bit 정확도의 양자화를 선택하면 3bit 정확도의 경우보다 수신된 화질은 훨씬 원 영상에 가깝다. 그러나 더 많은 정보량을 전송해야 하므로 채널용량의 한계에 부딪히게 된다. 샤논의 울 하한 기준에 의하면 왜곡이 제로인 경우 울은 무한대로 되는데 이것은 연속함수에서 적용되며, 이산함수의 경우에는 원 영상의 정보량이 유한하게 결정되어 있으므로 왜곡이 제로이더라도 정보량(엔트로피)이 상한선으로 된다. 즉,  $R(0) = H(x)$ 이다. 영상부호화에서는 일반적으로  $R(x) \leq H(x)$ 로 부호화 하여도 시각적 감도의 비선형성으로 인해 수용 가능하다. 문제는 얼마나  $R(x)$ 를 적게 하여도 좋을 것인가 하는 점이다. 이에 대한 해답은 여러 변수가 도입되어야 하지만 여기서는 입력값과 출력값과의 상호정보량의 관점에서 울-왜곡 관계식을 유도한다.

사상  $x = X$ 에 대한 자기정보량  $i(x)$ 는

$$i(x) = -\log p(x) \quad (1)$$

로 주어진다. 이진 정보에 대해 대수의 밑은 2가 되지만 여기서는 일반적인 경우를 다룬다. 평균정보량을 우리는 엔트로피라 한다. 즉,

$$H(x) = -\sum_{x=1}^{N-1} p(x) \log p(x) \quad (2)$$

영상부호기에서  $x$ 는 입력값을 의미하며 출력값  $y$ 에 대해서도 식(2)와 유사한 엔트로피 식으로 표현된다.

변수  $x$ 와  $y$ 의 결합분포확률을  $p(x, y)$ 라 하면 조

건분포확률은 다음과 같이 표현된다.

$$p(x, y) = p(x, y) / p(y) \quad (3)$$

대칭성을 고려하여  $p(x, y)$ 와  $p(y, x)$ 는 동일한 것으로 간주한다. 통신채널에서 울-왜곡 관계를 최소화하는 것은 정보량을 최대로 전송하는 조건에서 구해진다. 즉  $x$ 가 전송되어 왜곡이 없이 수신된기를 원하는데 여기서  $x$ 와  $y$ 의 상호정보량은 입력값의 자기정보량과 조건부 정보량의 차이로 정의된다. 즉,

$$\begin{aligned} i(x:y) &= i(x) - i(x|y) \\ &= -\log p(x) - (-\log p(x|y)) \\ &= \log [p(x, y) / p(x)] \end{aligned} \quad (4)$$

이것을 결합분포  $p(x, y)$ 에 대해 평균하면 다음과 같은 평균상호정보량을 얻는다.

$$I(x, y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)} \quad (5)$$

정보량의 측면에서  $I(x, y) \geq 0$ 이며 대칭성의 측면에서  $I(x, y) = I(y, x)$ 의 특성을 갖는다. 또한 전송신호가 왜곡없이 수신됨을 가정하면 식(4)에서 상호정보량은 제로로 되어 최소값을 갖는다. 따라서 울-왜곡 함수  $R(D)$ 는 평균 상호정보량의 최소치로 정의된다. 즉,

$$R(D) = \inf I_{p(x, y)}, \quad p(y|x) \in p_D \quad (6)$$

여기서,  $p(y|x)$ 는 입력이  $x$ 일 때 출력이  $y$ 로 되는 조건부 확률이며 임의의 왜곡을 수반한다.  $p_D$ 는 이러한 왜곡이  $D$ 가 되는 조건부 확률의 집합을 의미한다. 영상정보 처리에서 입출력은 대표적으로 양자화 과정에서 얻어지며 균일 또는 비균일 분포의 왜곡을 발생시킨다. 왜곡의 정량화는 입력값과 출력값의 차이를 기준으로 하는데 일반적으로 절대오차 기준과 자승오차 기준을 이용한다. 어떤 기준을 사용하는가에 따라 원 신호의 정보량과 전달된 정보량의 차이는 달라진다. 오차 평가 기준에 의해 적합한 울-왜곡 관계식이 얻어지며 또한 입력신호의 확률분포에 따라 달라진다. 위 식에서 상호정보량의 최소치가 울-왜곡 함수이며 상호정보량은 조건부 확률에 의해 정해지기 때문이다. 오차 기준은 절대오차와 자승오차 기준이 이용 가능하며 아래에서 각각의 오차 기준에 따라 울-왜곡 함수를 전개한다.

1. 절대오차 기준

상호정보량이 최소로 되는 울-왜곡 함수를 사논은 절대치 오차 기준  $\rho(x-y)=|(x-y)|$ 에 의해 다음과 같은 하한(lower bound)을 도출하였다.

$$R_L(D) = H(p) - \log(2eD) \tag{7}$$

여기서  $H(p)$ 는 확률분포함수에 따라 주어지는 원 신호의 정보량이며  $e$ 는 자연로그의 밑(2.7182)을 의미한다. 분포함수로서 일반적으로 이용되는 가우시안과 라플라시안의 경우에 울-왜곡 기준을 아래에 유도한다.

(1) 가우시안 분포함수의 경우

가우시안 분포함수는 제로평균을 가정하여 다음과 같이 주어진다.

$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \left[ \frac{-x^2}{2\sigma^2} \right] \tag{8}$$

이것을 식(2)의 정보량 식에 대입하면,

$$H(p) = \frac{1}{2} \log(2\pi e \sigma^2) \tag{9}$$

으로 유도된다. 식(9)를 식(7)에 대입하면 하한식은 다음과 같다.

$$R_L(D) = \frac{1}{2} \log \frac{\pi\sigma^2}{2eD^2}, 0 \leq D \leq \sqrt{\frac{\pi\sigma^2}{2e}} \tag{10}$$

(2) 라플라시안 분포함수의 경우

라플라시안 분포함수는 다음과 같이 주어진다.

$$p(x) = \frac{\alpha}{2} \exp[-\alpha|x|], \quad \alpha = \frac{\sqrt{2}}{\sigma} \tag{11}$$

이것을 식(2)의 정보량 식에 대입하면

$$H(p) = \log(2e/\alpha) \tag{12}$$

으로 유도된다. 식(12)를 식(7)에 대입하면 하한식은 다음과 같다.

$$R_L(D) = -\log(\alpha D), \quad 0 \leq D \leq \frac{1}{\alpha} \tag{13}$$

2. 자승오차 기준

자승오차 기준  $\rho(x-y) = (x-y)^2$ 에 의해 유도

된 사논의 하한은 다음과 같다.

$$R_L(D) = H(p) - \frac{1}{2} \log(2\pi eD) \tag{14}$$

이것을 이용하여 가우시안과 라플라시안 분포의 경우에 울-왜곡 하한식을 유도한다.

(1) 가우시안 분포함수의 경우

식(9)에서 얻은 정보량을 식(14)에 대입하면

$$R_L(D) = \frac{1}{2} \log \frac{\sigma^2}{D}, \quad 0 \leq D \leq \sigma^2 \tag{15}$$

으로 유도된다. 이것이 영상부호기에서 일반적으로 쓰이는 울-왜곡 함수이며 왜곡이 신호에너지보다 적은 범위에서 가장 적은 비트율을 전송할 수 있음을 의미한다. 왜곡이 신호에너지보다 크면 음의 비트율을 얻는데 현실적으로 무의미하므로 이 경우 음은 제로가 된다.

(2) 라플라시안 분포함수의 경우

식(12)에서 얻은 정보량을 식(14)에 대입하면

$$R_L(D) = \frac{1}{2} \log \frac{e\sigma^2}{\pi D}, \quad 0 \leq D \leq \frac{e\sigma^2}{\pi} \tag{16}$$

으로 유도된다. 이것을 식(15)와 비교하면 비율  $e/\pi$  만이 다르게 된다. 울이 제로가 되는 시점도 이 배율만큼 앞당겨짐을 알 수 있다. 따라서, 라플라시안 분포의 경우 왜곡은 신호에너지보다 이 비율만큼 적은 범위에서 정의된다.

영상부호기에서는 일반적으로 자승오차 기준을 선호하므로 다음 절에서는 저전송률 부호기로서 개발된 H.263 부호기의 결과를 식(15)와 식(16)의 울-왜곡 기준에 따라 비교 분석한다.

III. H.263 부호기의 구조

ITU-T Rec. H.263은 전화망이나 무선망과 같은 초저속 통신망을 이용한 오디오/비디오 서비스에서 동영상 정보를 압축 부호화하기 위한 표준이다. 초저속 통신망에서의 멀티미디어 서비스를 위한 터미널 제어 표준인 H.324 또는 LAN 상에서의 서비스를 위한 H.323 등에서 동영상 부호화에 이용되며 이에 대응하는 음성 부호기는 G.723이 대표적으로 이용된다. 또한 MPEG-4의 전단계 부호기로서 역할을 하며 H.261이나 MPEG-1과 같은 파형부호화 방식으로 64Kbps 이하의 저전송률 서비스가 가능하다.

MPEG-4에서도 H.263은 VLBV(Very Low Bitrate Video)코어로서 이용된다. 64Kbps 이하의 전송률을 실현하기 위해 다양한 알고리즘을 채택하였으며 전형적인 파형부호기의 향상된 버전으로 인식된다.

그림 1에서 보이는 바와 같이 H.263 부호기/복호기는 일반적인 동영상부호기와 유사한 구조를 갖고 있다.

공간적 중복도를 줄이기 위해 DCT 기반의 변환을 하며 시간적 중복도를 줄이기 위해 움직임 예측을 포함한다. 양자화기는 버퍼상태를 이용하여 제어되며 변환계수와 각종 헤더 정보를 가변장 부호화한다. 이렇게 유사한 구조를 가지면서 저전송률을 실현할 수 있는 이유는 세부적인 부호화 과정에 있다.

입력 영상의 포맷은 16CIF, 4CIF, CIF, QCIF, Sub-QCIF 등을 채택하는데 일반적으로 QCIF(176×144)을 이용한다. 부호기는 반드시 둘 다 지원해

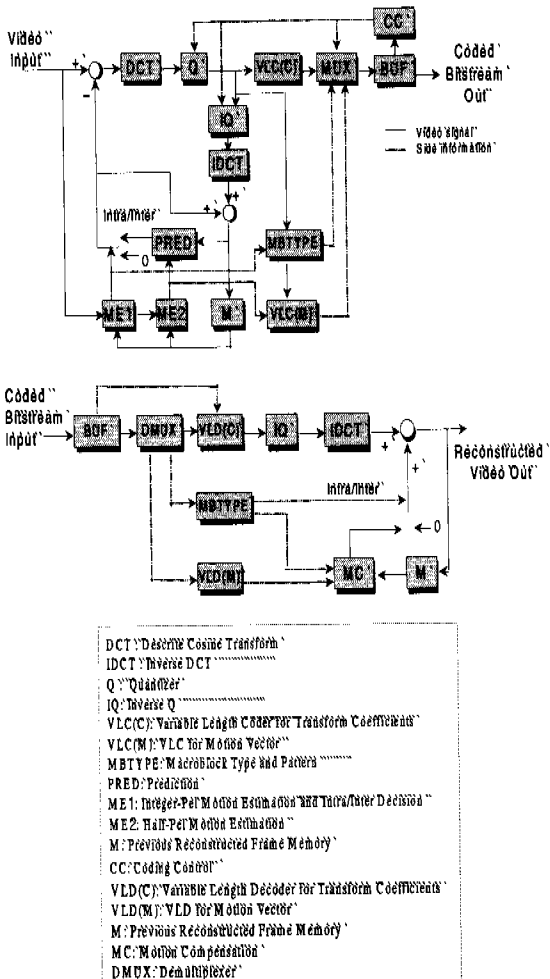


그림 1. H.263 부호기/복호기

야 한다. 비디오 스트림은 픽처층, GOB층, 매크로 블록층, 블록층등 다양한 네 개의 층으로 이루어진다. 부호화의 기본이 되는 매크로 블록층에서 다른 동영상 부호기보다 향상된 다양한 모드를 사용한다.

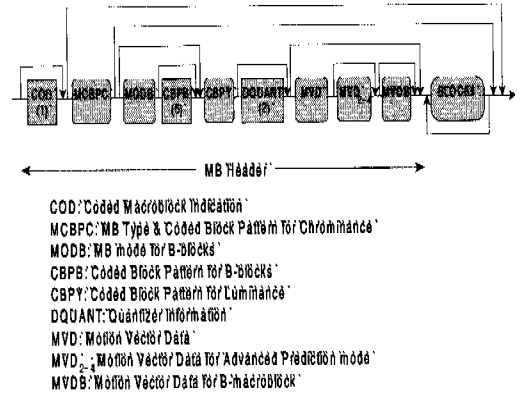


그림 2. 매크로 블록층의 데이터 구조

그림 2에 보이는 매크로 블록층의 구조를 살펴보면 움직임 벡터를 최대 4개까지 사용하여 움직임 보상의 정확도를 높인다. 특히 매크로 블록이 부호화되었는지 아닌지를 의미하는 COD는 매우 유용한 특징이다. COD 비트가 0이면 MB는 부호화됨을, COD가 1이면 부호화되지 않음을 의미한다. 부호기에서 이 MB는 모든 블록의 MV가 0이고 계수 데이터가 없는 INTER MB로 취급하며 Advanced prediction 모드에서 COD가 1이면 중첩된 MC를 수행한다.

이처럼 MB 전체에서 0이 아닌 계수가 없으면 부호화되지 않음을 의미하며 단 1비트로 전송되기 때문에 효율은 극대화된다. 실제 영상전환 등에서 이용되는 비교적 단순한 영상에서 COD 1비트는 자주 나타난다.

1	25	1	16	1	1	1	1	1	1	1
1	1	1	1	21	18	1	1	1	1	1
1	1	1	19	71	35	17	1	1	1	1
1	1	1	20	45	65	57	1	1	1	1
1	1	1	1	112	40	49	1	1	1	1
1	1	34	20	47	86	115	1	1	1	1
1	1	1	1	26	17	1	1	15	1	1
1	1	1	1	20	1	1	1	1	1	1
1	1	13	13	1	1	1	1	1	1	1

그림 3. 매크로블록 단위로 할당된 비트수 예

그림 3에서 QCIF 픽처에 대해 매크로 블록당 할당된 비트수를 보여주는데 99개의 MB 중 36개를 제외한 73개의 MB는 단 1비트만을 할당, 즉 COD 비트만을 전송하여 압축율을 높이고 있다. H.261과 같이 6비트의 부호화 블록패턴(CBP)을 사용하는 것보다 5비트만큼 효율을 향상시킨다.

#### IV. 실험 결과 및 고찰

울-왜곡 특성을 H.263부호기에서 분석하기 위해 QCIF(176×144화소)영상 시퀀스에 대해 64kbps에서 8kbps까지 변화시키면서 왜곡 특성을 조사하였다. 원래 영상은 30프레임/초로 되어 있으나 2개씩 건너 뛰어 10프레임/초 시퀀스로 실험하였다.

분산량은 실험 영상의 정보량을 결정하는 중요한 변수인데, 그림 4와 같이 실험 영상의 분산량은 프레임에 따라 큰 변화가 없다. 다만, foreman과 suzie 영상의 경우 장면이 급히 변화하면서 분산량이 변함을 보이는데 이 부분이 코딩 효율을 떨어뜨리는 주 원인으로 나타난다.

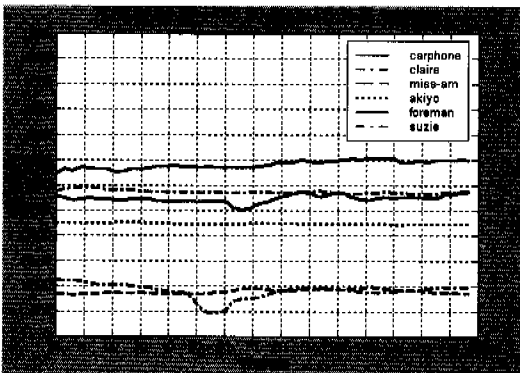


그림 4. 실험용 시퀀스의 프레임별 분산량

왜곡 측정 기준은 PSNR을 이용하며 MSE는 자승오차기준을 의미하여 다음과 같이 표현된다.

$$PSNR = 10 \log_{10} (255^2 / MSE) \quad (17)$$

$$MSE = \frac{1}{NM} \sum_n \sum_m [x(n, m) - \hat{x}(n, m)]^2 \quad (18)$$

식(17)에 의해 측정된 각 시퀀스(carphone, foreman, miss\_america, akiyo, claire, suzie)에 대한 PSNR이 그림 5에 제시된다. 울이 감소함에 따라 왜곡이 증가하는데 가우시안 또는 라플라시안 분포를 이용하여 각각을 비교한다.

왜곡량은 식(14)과 같이 자승오차를 사용했으므로 식 (15)와 식 (16)의 자승오차기준에 의한 울-왜곡 특성을 실제 측정 결과와 비교한다. 가우시안 분포의 경우, 왜곡이 신호에너지와 같으면 울은 제로가 된다. 신호에너지 이상의 왜곡을 가지면서 정보를 전송할 수 없음을 의미한다. 라플라시안의 경우 비율  $e/\pi$  만큼 줄어들므로 정보량이 전송될 수 있는 범위가 그만큼 감소한다.

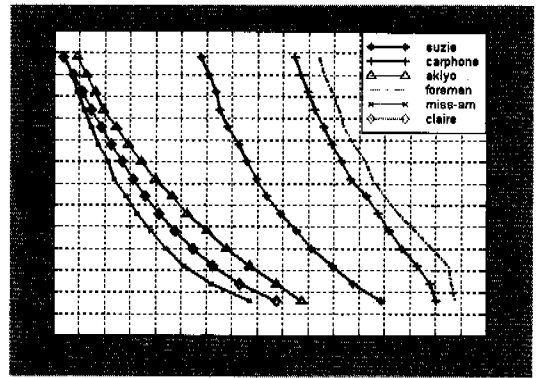


그림 5. 샘플 시퀀스의 비트율에 따른 왜곡량

그림 6에 가우시안의 경우 울-왜곡 특성을 실제 실험 결과와 비교하였다. 신호에너지(분산)는 왜곡이 최대인 경우 (8kbps)에 대해 정규화하여 이것을 식 (15)와 같이 자승오차 기준을 적용한 가우시안 분포 함수의 경우 울-왜곡 함수의 분산으로 하였다. 라플라시안의 경우 식 (16)을 이용하여 그림 7의 자승오차 기준으로 표기하여 비교하는데, 전송 가능한 최저 비트율은 상대적으로 감소하여 약 12kbps가 된다. 이는 이론치에서 음의 비트율로 떨어지는 부분을 제외하기 때문이다. 즉, 매우 낮은 비트율이 되면 왜곡량이 증대하여 다양한 부호화 기법에 의한 효과는 사라짐을 설명한다.

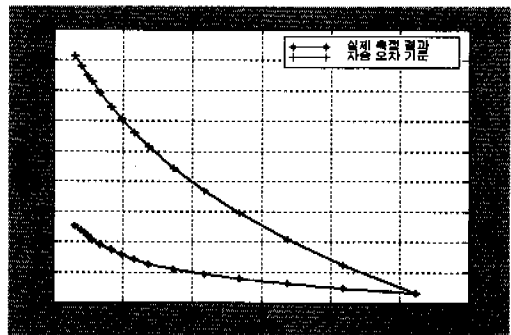


그림 6. 가우시안 분포함수의 울-왜곡 특성과 실제 측정치 (suzie.qcif)

그림 6과 그림 7에서 suzie.qcif 영상에 대해 측정된 결과 H.263 부호기에 의해 허용 가능한 전송률은 이론치보다 모두 낮게 나타났다. 즉, 같은 MSE를 가지고 더 낮은 비트율로 동영상 정보를 전송할 수 있음을 의미한다. MSE가 적을수록 차이는 커지고 반대로 MSE가 증가하며 전송률이 낮아지는 전형적인 윌-왜곡 관계를 설명한다. 이론치보다 낮은 전송률을 보이는 것은 H.263에서 채택한 다양한 기법에 의해, 예를 들어 COD를 이용해 단 1비트로 전송하는 것 등, 실제 전송률은 정보데이터 자체를 다루는 것보다 훨씬 낮아짐을 알 수 있다.

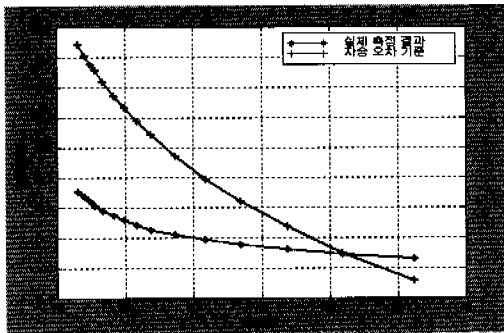


그림 7. 라플라시안 분포함수의 윌-왜곡 특성과 실제 측정치 (suzie.qcif)

### V. 결론

영상 정보를 압축하여 부호화하는데 정보량(엔트로피)의 측면에서 함수화한 것이 윌-왜곡 함수이다. 윌이 증가하려면 왜곡이 적어야 하며, 반대로 왜곡이 증가하면 주어진 채널 대역폭을 충분히 사용하지 못하고 채널용량이 감소하며 즉, 비트율이 감소한다는 것은 지극히 자연스런 현상이다. 그러나, 신호의 확률 분포에 따라 정보량이 다르게 발생하고 왜곡의 측정기준이 무엇이나에 따라 함수는 다르게 결정되어야 한다. 이 논문에서는 가우시안과 라플라시안 확률분포에 의해 입력영상의 정보량이 결정되고 절대치 오차기준과 지승오차 기준을 적용하여 각각의 경우 윌-왜곡 함수를 도출하였다. 모든 경우 윌은 왜곡의 증가와 함께 음의 방향으로 감소하며 왜곡의 대수함수로 결정된다. 다만, 윌이 음이 아닌 조건에서 왜곡의 최대치가 얼마나 되는지에 대해 신호에너지가 되거나 또는 그 이하가 됨을 제시하였다.

윌-왜곡 특성을 실제 널리 사용되는 H.263부호기에서 얻은 결과와 비교 분석하였다. 주어진 데이터

자체를 부호화하는 것이 아니라, 시공간적으로 상관도를 제거하고 엔트로피 코딩에 의해 부호화함은 물론 비트스트림의 계층구조를 채택하고 다양한 헤더정보를 이용하여 효율적인 부호화를 하기 때문에, 결과로 얻은 전송률은 윌-왜곡함수에 의한 이론치보다 정규화 비트율에서 최대 0.55만큼 낮게 부호화될 수 있음을 보였다. 이때 특정의 전송률에서 결과 영상의 화질이 얼마나 우수한가하는 문제는 여기서 다루지 않았다. 매크로블록별로 극대화한 H.263부호기 자체의 성능을 비교하는 것이 이 논문의 주안점이다.

### 참고 문헌

- [1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Int. Conv. Rec.*, Vol. 7(Part 4), pp. 142-163, 1959.
- [2] ITU-T, Draft Recommendation H.263 (Video coding for low bitrate communication), 9 Dec.1995.
- [3] ITU-T SG15 WP15/1, Video codec test model, TMN5, 31 Jan. 1995.
- [4] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. IT-27, pp. 519-521, Mar. 1960.
- [5] P. F. Panter and W. Dite, "Quantization distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44-48, Jan. 1951.
- [6] P. F. Swaszek and J. B. Thomas, "Design of quantizers from histograms," *IEEE Trans. on Commun.*, vol. COM-32, pp. 240-245, Mar. 1984.
- [7] T. Berger, *Rate distortion theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
- [8] L. D. Davisson, "Rate-distortion theory and application," *Proc. IEEE*, vol. 60, pp. 800-808, July 1972.

