

# 인간 시각의 주파수 특성을 이용한 미세 계위적 동영상 부호화 방법

김 승 환, 호 요 성, \*마 평 수

광주과학기술원, \*한국전자통신연구원

전화 : 062-970-2258 / 핸드폰 : 011-9052-5739

## A FGS Video Coding Algorithm Using HVS-based Frequency Weighting

Seung-Hwan Kim, Yo-Sung Ho, \*Pyeong-Soo Mah

Kwangju Institute of Science and Technology (K-JIST)

\*Electronics and Telecommunications Research Institute (ETRI)

E-mail : {kshkim, hoyo}@kjist.ac.kr, pmah@etri.re.kr

### Abstract

In this paper, we propose a scalable video coding algorithm using frequency weighting in the DCT domain. In order to provide FGS (fine granular scalability), we use a frequency weighting matrix based on HVS. We also define a new error metric, JNDE (just noticeable difference error), to measure reconstructed image quality based on HVS.

### I. 서론

일반적으로 비디오 부호화의 목적은 주어진 비트율에서 비디오 화질을 최적화하려는 것이다. 기존의 통신 시스템 부호화기는 전송채널의 용량을 알고 있다고 가정하며, 부호화기는 수신된 모든 비트들을 복호할 수 있다는 가정하에 부호화를 수행한다[1]. 그러나 부호화기는 시간에 따라 변하는 채널 용량을 정확히 예측할 수 없으며, 부호화기 또한 다른 운영체제들과 자원을 공유하는 경우에 수신된 비트를 모두 복호할 수 없다.

최근 인터넷 스트리밍을 위해 제안된 FGS (fine granular scalability) 방식은 부호화를 수행하기 전에 미리 전송채널의 용량을 예측하여 최소채널 용량보다 작은 비트율로 기본계층을 부호화하므로 항상 기본계층 데이터의 전송은 보장된다. 또한 원 영상과 기본계층 영상의 차이 값인 잔여 영상은 DCT 변환 영역에서

비트평면 단위로 부호화되어 전송된다[1].

FGS 방식은 비트평면 부호화를 이용하므로 전송 채널과 시스템의 복잡성을 고려하여 일부 비트평면까지만 복호할 수 있다. 게다가, 비트평면이 DCT 계수별로 부호화되므로, 인간의 시각에 중요한 주파수 성분을 강조하기 위해 각 계수별로 해당하는 가중값을 고려하여 비트평면을 천이시킬 수 있다.

본 논문에서는 HVS (human visual system) 관점에서 각 DCT 계수의 중요성에 따른 비트평면 천이를 통해 인간 시각에 맞는 화질 향상을 얻는다. 또한 기존의 화질측정 단위인 PSNR (peak signal-to-noise ratio) 대신에, 인간의 시각적 특성에 맞는 새로운 화질 측정 단위 JNDE (just noticeable difference error)를 정의하고, 이를 이용하여 주파수 가중을 이용한 화질 향상을 HVS 관점에서 분석한다.

### II. FGS 기본 구조

#### 2.1 FGS 상위계층 부호화

그림 1에 보인 것처럼, FGS의 기본계층은 기존의 블록기반 부호화 방법과 동일하게 부호화된다. 사각형 점선으로 표시되어 있는 상위계층에서는 잔여 영상을 부호화한다. 잔여영상 정보는 8x8 블록별로 나뉘어져서 DCT 변환되므로, 블록당 64 개의 DCT 계수를 가진다. 이때 각 DCT 계수들은 십진수에서 이진수로 바

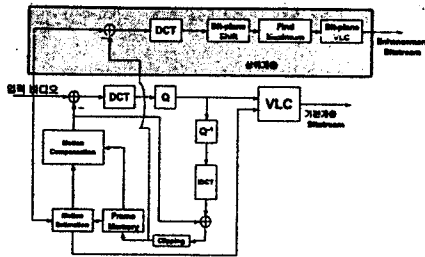


그림 1. FGS 부호화기

뛰어 비트평면을 구성하고, 각 비트평면은 가변길이 부호화되어 전송된다[1].

2.2 FGS 주파수 가중

각 DCT 계수들은 화질에 서로 다른 영향을 미친다. 특히, 저주파 성분이 고주파 성분보다 인간의 시각에 더 민감하므로, 저주파 계수의 정밀도는 고주파 계수의 정밀도보다 더 중요하다. 이러한 관점에서, 저주파 계수들을 위로 천이시켜 상대적으로 고주파 계수들보다 우선적으로 전송하면, 일부 비트스트림만을 처리하는 경우에 화질의 향상을 얻을 수 있다.

그림 2는 저주파 계수부터 차례로 천이 계수를 내림순(4, 3, 2, 1, 0)으로 적용한 예를 나타내고 있다.

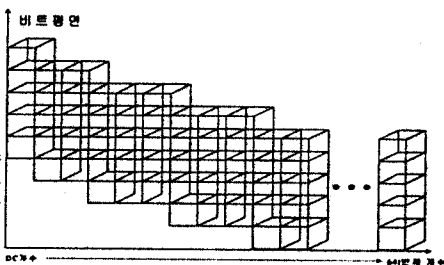


그림 2. 주파수 천이

III. HVS 기반 주파수 가중

3.1 주파수 가중 행렬

HVS는 비선형 주파수 함수인 MTF (modulation transfer function)를 식 (1)과 같이 정의하고 있다.

$$H(f) = a(b+cf) \exp(-cf) \quad (1)$$

여기서  $f$ 는 cycles/degree를 나타내고,  $a, b, c, a$ 는 상

수값을 나타낸다. 이러한 상수값을 조절하여 각 주파수의 가중치를 보여주는 MTF의 모양을 결정한다. 특히, 8x8 블록의 DCT 계수들의 주파수 가중 행렬은 실험을 통하여 제시되었다[2]. 그림 3에서 각 위치에 해당하는 가중값은 실제 DCT 계수들의 곱셈에 대한 가중치를 의미한다.

0.4942	1.0000	0.7023	0.3814	0.1856	0.0849	0.0374	0.0160
1.0000	0.4549	0.3085	0.1706	0.0845	0.0392	0.0174	0.0075
0.7023	0.3085	0.2139	0.1244	0.0645	0.0311	0.0142	0.0063
0.3814	0.1706	0.1244	0.0771	0.0425	0.0215	0.0103	0.0047
0.1856	0.0845	0.0645	0.0425	0.0246	0.0133	0.0067	0.0032
0.0849	0.0392	0.0311	0.0215	0.0133	0.0075	0.0040	0.0020
0.0374	0.0174	0.0142	0.0103	0.0067	0.0040	0.0022	0.0011
0.0160	0.0075	0.0063	0.0047	0.0032	0.0020	0.0011	0.0006

그림 3. 주파수 가중 행렬

3.2 주파수 가중을 이용한 비트평면 천이

비트평면을 천이하는 것은 천이값에 대해 2의 지수배로 그 계수값을 증가시키는 의미가 있다[4]. 즉, 어떤 계수를  $n$  비트 더 보내기 위해 요구되는 가중값은  $2^n$ 이 된다. 따라서 가중 행렬이 비트평면 부호화와 결합하기 위해 먼저 각 가중값을 2에 대한 지수 형태로 나타내야 한다.

$$C'(i, j, k) = 2^{n_m(i)} \cdot C(i, j, k) \quad (2)$$

$$2^{-n_m(i)} = C(i, j, k) / C'(i, j, k) \quad (3)$$

$C(i, j, k)$ 는  $k$  번째 매크로블록의  $j$  번째 블록내  $i$  번째 DCT 계수,  $C'(i, j, k)$ 는 주파수 가중을 고려한 DCT 계수,  $n_m(i)$ 는 주파수  $i$ 에 대한 비트평면 천이 계수를 나타낸다. 그림 3에 있는 가중 값들이 2에 대한 지수 형태로 변환하기 위해 그림 3에 있는 가중 값에  $2^{-n_m(i)}$ 를 곱하여 크기를 조정한다. 그림 4의 경우는 최고 주파수 천이가 4인 경우로 그림 3의 가중 값들이  $2^4$  배만큼 증가되었다.

7.9072	16.000	11.236	6.1024	2.9696	1.3584	0.5984	0.256
16.000	7.2784	4.9360	2.7296	1.3520	0.6272	0.2784	0.1200
11.236	4.9360	3.4224	1.9904	1.0320	0.4976	0.2272	0.1008
6.1024	2.7296	1.9904	1.2336	0.6800	0.3440	0.2080	0.0752
2.9696	1.3520	1.0320	0.6800	0.3936	0.2128	0.1072	0.0512
1.3584	0.6272	0.4976	0.344	0.2128	0.1200	0.0640	0.0320
0.5984	0.2784	0.2272	0.1648	0.1072	0.0640	0.0352	0.0176
0.2560	0.1200	0.1008	0.0752	0.0512	0.0320	0.0176	0.0096

그림 4. 크기 조정된 주파수 가중 행렬

$$n_m(i) = \lceil \log_2 [ C'(i, j, k) / C(i, j, k) ] \rceil \quad (4)$$

그림 4의 크기가 조정된 주파수 가중 행렬의 값을 정수로 올리고, 밑이 2인 로그값을 구한 후에 내림을 취하면, 그림 5의 천이 주파수 가중 행렬을 얻는다.

3	4	3	2	1	0	0	0
4	3	2	1	1	0	0	0
3	2	2	1	1	0	0	0
2	2	1	1	0	0	0	0
1	1	1	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

그림 5. 천이 주파수 가중 행렬

#### IV. 실험결과 및 분석

그림 5에서 얻어진 주파수 가중을 이용하여 각 DCT 계수들이 각각 가중되어 부호화된다. 그림 6의 경우에 297 장의 그림을 1849 kbit로 부호화했고, 그림 7의 경우는 같은 그림을 최고 천이계수 2를 적용하여 주파수 가중을 하여 1296 kbit로 부호화했다. 단, DC 계수는 최고 천이계수 "2"를 사용했다. 왜냐하면, DC 값은 인간 시각에는 민감하지 않지만 실제 영상의 평균값을 의미하고 가장 많은 영상에너지를 가지고 있기 때문이다. 실험결과는 주파수 가중을 적용한 그림 7을 부호화한 비트수가 그림 6보다 적지만 더 좋은 화질을 보여주고 있다.

본 논문에서는 적절한 화질 평가를 위해서 원영상과 복원영상 사이의 오차 값에만 의존하는 PSNR 보다는 인간시각에 오차로 판별되는 경우만 사실상의 오차로 간주하고자 한다. 따라서, 인간시각에 구별되는 오차만 고려하는 관점에서 해석하기 위해 Weber's law에 근거하는 새로운 화질평가 방법을 제시한다.

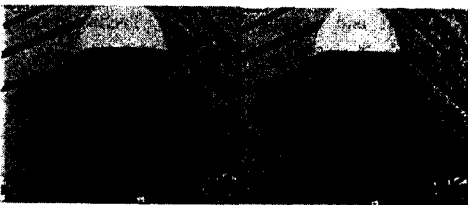


그림 6. 비주파수 가중 그림 7. 주파수 가중

영상의 밝기 성분을 8 비트로 나타냈을 때 256 단계의 밝기로 표현할 수 있다. 그리고 각각의 밝기에 따라 오차에 대한 인간시각은 서로 다른 민감도를 가

지고 있다. 즉, 원영상이 아주 밝은 흰색이었을 경우가 어두웠을 경우에 비해서 보다 같은 오차값에 더 민감하다.

$$\frac{\Delta I}{I} = \alpha \quad (5)$$

식 (5)는 Weber's law를 나타낸다. 여기서 I는 초기에 주어진 자극의 값(영상에서는 원래 주어진 화소값)에 해당한다.  $\Delta I$ 는 초기 자극에 대한 변화 값을 의미한다.  $\alpha$ 는 인간이 감지 할 수 있는 초기값과 변화된 값의 경계 비율을 의미한다. 우리는 Weber's law를 원영상과 복원된 영상과의 관계에 적용한다. 즉 원영상을 초기 주어진 자극(I)으로 간주하고 원영상과 복원된 영상과의 오차 ( $\Delta I$ )를 변화된 자극의 값으로 간주하면 주어진 오차값에 대한 인간시각의 민감도는 원영상의 화소값에 의존한다. 그리고 원영상의 화소값은 잡음과 같이 예측이 불가능하므로 [0, 255] 사이에 고르게 분포한다고 할 수 있다.

$$\frac{\Delta I}{I} = \frac{D}{P_A} \geq \alpha \quad (6)$$

$$P_A \leq \frac{D}{\alpha} = P_{A-D} \quad (7)$$

식 (6)은 원영상의 화소값  $P_A$ 에 대하여 오차 D가 주어졌을 때 인간의 시각에 오차로 간주되는 경계 비율( $\alpha$ ) 이상인 경우 즉 인간시각에 오차로 간주되는 범위를 나타낸다. 그리고 식 (7)에서는 식 (6)에서 주어진 오차값 D에 대해서 오차로 간주되기 위한 원영상의 화소값  $P_A$ 의 최대 경계값( $P_{A-D}$ )이  $\frac{D}{\alpha}$ 임을 나타낸다. (단  $\alpha$ : bits/화소, 2<sup>-8</sup>: 전체 화소값 단계)

$$P(P_A \leq P_{A-D}) = \sum_{n=0}^{P_{A-D}} \frac{1}{2^n} = \frac{D}{\alpha} + 1 \quad (8)$$

또한 원영상의 화소값  $P_A$ 가 [0, 255] 사이에 균일하게 분포 되어있고 각각의 확률이 1/2<sup>n</sup>으로 균일한 이산 확률변수 가정하에 원영상의 화소값이  $P_{A-D}$ 보다 작을 확률이 식 (8)과 같이 나타난다. 그리고 원영상의 화소값에 관계없이 오차가 발생하므로 오차 D와 원영상의 화소값이 서로 독립이며,  $P_{N,E,D}$  (오차가 D일때 JNDE (인간시각에 오차로서 간주)될 확률)는 식(9)와 같이 나타낼 수 있다.

$$P_{N,E,D} = P(P_A \leq P_{A-D} | D) = P(P_A \leq \frac{D}{\alpha}) = \frac{(D/\alpha) + 1}{2^8} \quad (9)$$

따라서 JNDE 관점에서의 오차 개수는 식 (10)과 같이

실제 발생한 오차의 개수에 그 오차값 D에서 오차로서 간주될 수 있는 확률값을 곱한 개수와 같다.

$$JND(d1-d2) = P_{N,K,D} \cdot (d1-d2) \quad (10)$$

위에서 얻어진 수식을 바탕으로 주파수 가중과 비주파수 가중의 거의 같은 PSNR를 갖는 두 영상의 경우 주파수 가중의 영상이 더 나은 화질을 보여주는 이유를 설명한다. 먼저 그림 8에서 보여지는 바와 같이, 원영상에서 주파수 가중된 영상과의 차이를 오차 1, 비주파수 가중된 영상과의 차이를 오차 2라고 하였다.

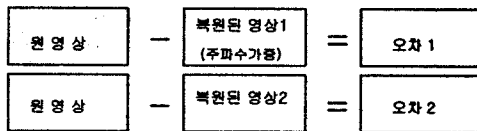


그림 8. 원영상과의 오차

표 1에서는 그림 8에서 주어진 실험을 바탕으로 주파수 가중을 수행한 경우와 비주파수 가중 경우에 대해 주어진 오차 D에 대한 수치를 나타낸다. fw는 주파수 가중, No fw는 비주파수 가중을 나타내고, a=0.02를 적용하였다. 또한 각 오차 D에서 주파수 가중 경우와 비주파수 가중 경우의 오차 개수를 뺀 값을 d1-d2로 나타내었다. 따라서 d1-d2값이 양수인 경우는 주파수 가중의 오차가 많은 경우이고, 음수인 경우에 비주파수 가중의 오차 개수가 많음을 의미한다. 그리고 JNDE를 통하여 인간시각에 오차로 간주되는 화소의 개수를 다시 계산한 결과를 JND( )로 표시하

표 1. 각 오차에 대한 화소개수

fw	No fw	d1-d2	D	JND(fw)	JND(No fw)	JND(d1-d2)
13,341	12,150	1191	C	13,341	12,150	1191
23,197	21,820	1377	1	4,621	4,347	274
17,211	17,222	-11	2	6,790	6,795	-5
12,386	12,960	-574	3	7,306	7,603	-297
8,866	9,443	-577	4	6,866	7,414	-548
6,457	6,885	-428	E	6,330	6,751	-421
4,747	5,055	-308	6	4,747	5,055	-308
3,529	3,743	-214	7	3,529	3,743	-214
2,654	2,813	-159	8	2,654	2,813	-159
2,011	2,128	-117	9	2,011	2,128	-117
1,535	1,628	-93	10	1,535	1,628	-93
1,178	1,250	-72	11	1,178	1,250	-72
909	966	-57	12	909	966	-57
701	744	-43	13	701	744	-43
545	580	-35	14	545	580	-35

였다. 표 1에서 알 수 있듯이, JNDE를 적용하지 않았을 때 오차 1인 경우에 대해서 주파수 가중의 경우가 비주파수 가중의 경우보다 1377개나 더 많은 오차 개수가 JNDE를 적용해서 계산했을 때 274개로 그 차이가 현저히 줄었다.

### V. 결론

주파수 가중을 적용했을 때 적은 비트로 부호화했지만 오히려 더 나은 화질을 얻을 수가 있었다. 이는 저주파 계수가 고주파 계수에 비해 상대적으로 인간시각에 더 민감하다는 것이 증명된 것이다. 또한 64개의 계수들에 대한 중요도를 모두 고려해서 얻은 결과이므로 사실상 HVS 관점에서는 최적화된 부호화라고 볼 수 있다. 또한 표 1에서 보는 바와 같이 JNDE를 적용하여 이미지의 화질을 비교했을 때 PSNR과 단순히 오차에 대한 화소의 개수를 고려한 경우보다 매우 효과적으로 화질을 비교할 수 있었다. 이는 시각에 감지되지 않은 부분을 제거했기 때문에 얻을 수 있는 결과이다. 앞으로 JNDE는 HVS 관점에서의 화질 비교 및 동일 HVS 화질에서 주파수 가중을 이용했을 때 주파수 가중을 하지 않았을 때와 부호화 효율을 객관적으로 비교할 수 있는 중요한 화질 평가의 기준으로 사용될 수 있다.

### 감사의 글

본 연구는 광주과학기술원(K-JIST) 초 고속광네트워크연구센터(UFON)를 통한 한국과학기술단 우수연구센터(ERC) 지원금에 의한 것입니다.

### 참고 문헌

- [1] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," IEEE Trans. Circuit and System for Video Technology, vol. 11, no. 3, pp. 301-317, March 2001.
- [2] K. Rao and P. Yip, Discrete Cosine Transform, Academic Press, INC, New York, 1990.
- [3] K. Rao and P. Yip, "Human Visual Weighted Progressive Image Transmission," IEEE Trans. Comm. vol. 38, no. 7, pp. 1040-1044, July 1990.
- [4] H. Radah, M. Van der Schaar and Y. Chen, "The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming Over IP," IEEE Trans. Multimedia, vol. 3, no. 1, pp. 53-68, March 2001.