HVS-BASED FREQUENCY WEIGHTING FOR FGS VIDEO CODING

Seunghwan Kim and Yo-Sung Ho

Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-dong Puk-gu, Kwangju, 500-712, Korea
{kshkim, hoyo}@kjist.ac.kr

ABSTRACT

In this paper, we propose a new scalable video coding algorithm using the property of the human visual system (HVS). Since HVS can be modelled by a nonlinear sensitivity function, we can use a frequency weighting matrix derived from the modulation transfer function (MTF) to enhance visual quality of reconstructed images. In order to apply frequency weighting in bit-plane coding at the enhancement layer, we convert this frequency weighting matrix into a frequency shift matrix. By applying HVS-based frequency weighting, we can obtain improved visual quality and fine scalability at different bit rates. We also define a new metric, the just noticeable difference (JNDE), to estimate enhanced image quality in terms of HVS. In this metric, we apply the Weber's law to determine the probability of noticeable errors.

Keywords: FGS, Frequency Weighting, HVS, JNDE

1. INTRODUCTION

Video coding has traditionally been to optimize video quality at different bit rates. In the conventional communication system, we encode the source data under two conditions: (1) the encoder should know the channel capacity and compress the input data to produce a coded bitstream less than the channel capacity, (2) the decoder should decode all the bits received from the channel to reconstruct the source information. However, in the packet-based transmission environments, the encoder no longer knows the channel capacity; therefore, it does not know at which bit rate the video quality should be optimized. The decoder does not process all the bits received from the channel because it should consider available resources at the user terminal [1].

Recently, the MPEG-4 group has proposed a functionality, called fine granular scalability (FGS), to provide efficient video streaming over Internet [2,3]. For FGS, the encoder estimates the channel capacity before encoding and compresses the base layer into less than the channel capacity. Therefore, transmission of the base layer bitstream is always guaranteed. In the enhancement layer of FGS, the residual signal is compressed by bit-plane coding.

In this paper, we propose a frequency weighting method for FGS based on the human visual system (HVS). In order to enhance visual quality of reconstructed images, we assign different weighting to each DCT coefficient corresponding to the human visual sensitivity. We convert the frequency weighting matrix to a frequency shift matrix to implement HVS-based frequency weighting in the bit-plane coding for FGS. We also define a new error metric, the just noticeable difference error (JNDE), to measure the reconstructed image quality based on HVS. By JNDE, we can effectively measure perceptual visual quality of images from the point of HVS.

2. FINE GRANULAR SCALABILITY

FGS has recently been introduced in respond to an increasing demand for a video coding method for video streaming over Internet. Fig. 1 shows the encoder structure of the two-layer FGS system.

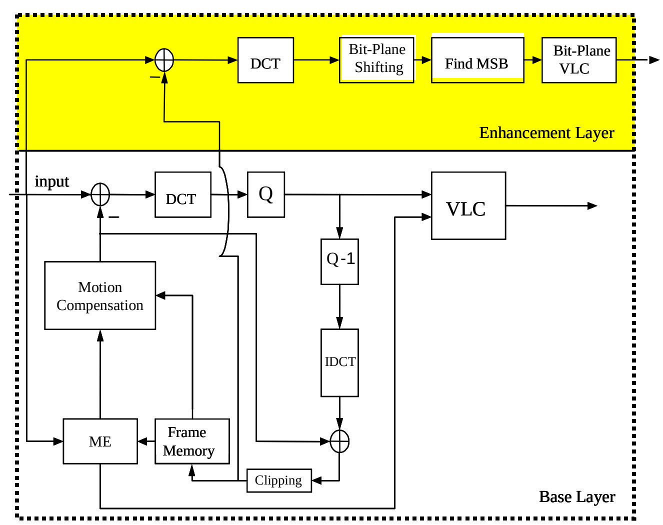


Fig. 1. FGS Encoder

This work was supported in part by K-JIST, in part by KOSEF through UFON, and in part by MOE through BK 21.

As shown in Fig. 1, the FGS encoder consists of two parts: based layer and enhancement layer. In the base layer, the basic information of the input signal is coded in the same way as the traditional block-based coding method. In the enhancement layer, the residual signal that is not coded in the base layer is divided into 8×8 blocks and each block is DCT transformed. All the 64 DCT coefficients are represented by binary numbers. These binary values form several bit-planes, where each bit-plane of the block is defined as an array of 64 bits, taken from absolute values of the DCT coefficients at the same significant position. Finally, each bit-plane is entropy coded by (RUN,EOP) symbols to produce the output bitstream [1].

3. FREQUENCY WEIGHTING BASED ON HVS

Our visual system has a different sensitivity at each frequency component. In general, human eyes are more sensitive to low frequencies than to high frequencies. Therefore, low frequency coefficients are visually more important than high frequency coefficients. In order to have better image quality at a given bit rate, we should transmit low frequency components faithfully [4]. To improve visual quality of images, we can exploit the modulation transfer function (MTF) that represents the importance of each frequency component in terms of HVS. MTF can be described by

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

where *f* is the radial frequency in cycles/degree of the visual angle, and *a*, *b*, *c* and *d* are constants.

Using the convolution-multiplication property of the DCT for a sampling density of 64 pels/degree, we can develop an 8×8 weighing matrix representing the HVS sensitivity [5]. The 8×8 DCT coefficients are multiplied by the corresponding elements of this matrix, reflecting their HVS weighting. Fig. 2 shows the 8×8 frequency weighting matrix that represents the importance of 64 DCT coefficients in terms of HVS.

0.4942	1.0000	0.7203	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	0.0329	0.0311	0.0215	0.0133	0.0075	0.0040	0.0020
0.0374	0.0174	0.0142	0.0143.	0.0067	0.0040	0.0022	0.0011
0.0160	0.0075	0.0063	0.0047	0.0032	0.0020	0.0011	0.0006

Fig. 2. Frequency Weighting Matrix

In order to provide HVS weighting, every DCT coefficient value should be multiplied by the corresponding element of the frequency weighting matrix. Therefore, the frequency weighted DCT coefficient value is

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

where *C(i, j, k)* is the DCT coefficient value of the *i*-th component in the *j*-th block of the *k*-th macroblock, *C'(i, j, k)* is the frequency weighted coefficient value by *fw(i)* that is the value of frequency weighing of the *i*-th DCT coefficient in each block.

Shifting up a DCT coefficient by one bit has the same effect as multiplying the DCT coefficient by two. In general, if we shift a DCT coefficient by n bits up in the bit-plane representation, it corresponds to multiplying the DCT coefficient by 2^n. Therefore, in order to emphasize some DCT coefficients, we can shift up those coefficients by corresponding factors.

From this point, we can make some transformation from the frequency weighting matrix to the frequency shift matrix. In order to make the transformation, we select the maximum shift factor $maxn(fw)$ which represents the number of bits to be shifted up at the most important DCT coefficient. In the frequency weighting matrix, the weighting value is normalized by one. Therefore, the frequency weighting matrix should be multiplied by $2^{maxn(fw)}$, because the maximum value one in the frequency weighting matrix is should be changed by $2^{maxn(fw)}$ which represents to shift the most important DCT coefficient by $maxn(fw)$ bits. Fig. 3 shows the scaled frequency weighting matrix by $maxn(fw) = 4$.

7.9072	16.000	11.236	6.1024	2.9696	1.3584	0.5984	0.2560
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	1.3020	0.6800	0.3936	0.2128	0.1072	0.0512
1.3584	0.6272	0.4976	0.3440	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

Fig. 3. Scaled Frequency Weighting Matrix

After scaling the frequency weighting matrix, the matrix is transformed to the frequency shift matrix which can be directly applied to the bit-plane coding by Eq. (3).

$$nfw(i) = \lfloor \log_2 \lceil 2^{maxn(fw)} \cdot fw(i) \rceil \rfloor \quad (3)$$

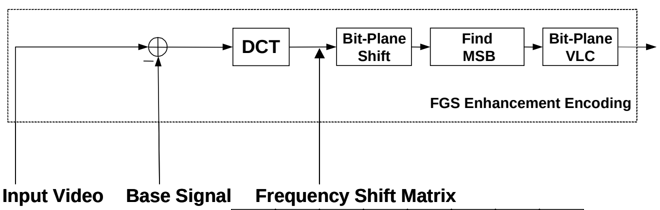
where $nfw(i)$ is a shift factor at the *i*-th DCT coefficient, and $2^{maxn(fw)} \cdot fw(i)$ represents the scaled frequency weighting. Fig. 4 shows the frequency shift matrix from the frequency weighting matrix in Fig. 3.

In Fig. 4, the shift factor of the DC coefficient is three, which is smaller than the maximum shift factor four. However, the DC coefficient means the average value of the image; therefore, we use the maximum shift factor for the DC coefficient.

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

Fig. 4. Frequency Shift Matrix

Fig. 5 explains the operations of the proposed frequency weighting method, implemented with the frequency shift matrix before bit-plane coding.



4	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

Fig. 5. Proposed Frequency Weighting Method

The maximum shift factor ,$maxn(fw)$, depends on the bitrate of the base layer. The higher bitrate of the base layer, the smaller maximum shift factor is required. In other words, the statistical characteristics of the residual signal would be similar to a random noise as we increase the bitrate of the base layer.

4. PERCEPTUAL VISUAL QUALITY

We define a new error metric, JNDE, to measure the reconstructed image quality based on HVS. JNDE only counts noticeable errors by human eyes, and represents whether some errors are noticeable or not. The just noticeable probability is induced by the Weber's law. The Weber's law is

$$\frac{\Delta I}{I} = \alpha \tag{4}$$

where I represents the initial stimulus intensity, ΔI represents the difference threshold, and α is the ratio of noticeable difference over the initial intensity. It shows the size of the difference threshold appeared to be related to the initial stimulus magnitude. From the Weber's Law, we have noticeable errors in two aspects. One is the relation between the original image and the reconstructed image, and the other is the relation among neighboring pixels. Fig. 6 illustrates these two relations.

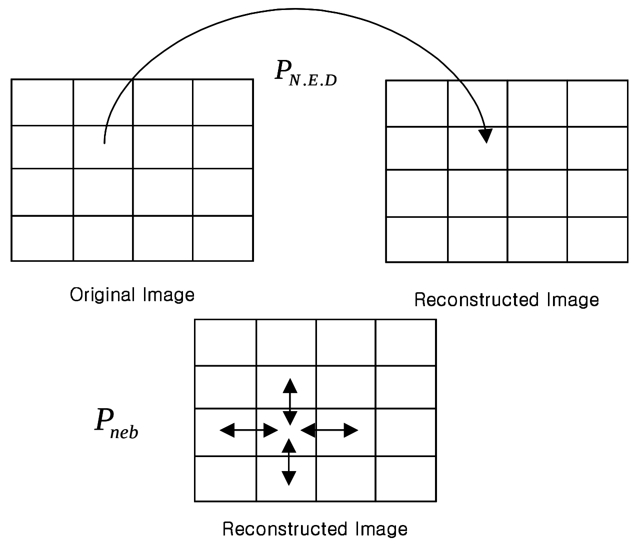


Fig. 6. Probability of JNDE

In Fig. 6, $P_{N.E.D}$ is a probability of noticeable errors considering the relation between the original pixel value and its reconstruction pixel value, and P_{neb} is a probability of noticeable error considering neighboring pixels. From these two probability models, we can calculate the total probability of noticeable errors.

In Eq. (5), we expand the Weber's law into the relation between the original pixel value and its reconstructed pixel value.

$$\frac{\Delta I}{I} = \frac{D}{P_k} \geq \alpha \tag{5}$$

where *D* is the difference between the original pixel value and its reconstructed image pixel value, and p_k is the original pixel value.

If the original image has a uniform distribution, the probability where the original pixel value is lower than the maximum threshold value $p_{ths}$is represented by

$$P(p_k \leq p_{ths}) = \frac{D/\alpha + 1}{2^n} \tag{6}$$

where *n* represents the number of bits per pixel. We also consider that the original pixel value and its error are independent. For a given error, the probability of just noticeable error is described by

$$P_{N.E.D} = P(p_k \leq p_{ths}) = \frac{D/\alpha + 1}{2^n} \tag{7}$$

Based on the number of noticeable errors in the neighboring pixels, we can classify the noticeable errors into one of the four different cases. These noticeable errors are defined by P_{neb}.

If all the four neighboring pixels are noticeable, its probability is

$$_4C_4 \cdot (P_e)^4 \cdot (1 - P_e)^o \quad (8)$$

where P_e represents the probability of noticeable errors between the given pixel and one of the neighboring pixels,

If three out of four neighboring pixels are noticeable, the probability is.

$$\frac{3}{4} \cdot_4 C_3 \cdot (P_e)^3 \cdot (1 - P_e)^1 \quad (9)$$

where 3/4 is a weighting factor for noticeable errors because only three neighboring pixels are noticeable.

If the number of the noticeable pixel errors is two, its probability is

$$\frac{2}{4} \cdot_4 C_2 \cdot (P_e)^2 \cdot (1 - P_e)^2 \quad (10)$$

The case of one noticeable pixel errors is also calculated by the same way of above equations.

From above equations, we can derive the P_{neb} by adding all the four possibilities. Therefore, P_{neb} is

$$P_{neb} = \sum_{1 \le k \le 4} \frac{k}{4} \cdot_4 C_k \cdot (P_e)^k \cdot (1 - P_e)^{4-k} \quad (11)$$

Consequently, the probability, P_D of the total noticeable errors at a given difference D, is calculated by

$$P_D = P_{neb} \cdot P_{N.E.D} \quad (12)$$

Therefore, the average number of noticeable errors is $P_D \times$ (the number of errors at *D*).

The above equations can explain why the frequency weighted method provides better image quality than no frequency weighted method. We calculate the number of noticeable errors between original and reconstructed images.

Table 1 shows the number of pixels at a given error. ”W” represents the frequency weighting case and ”N” represents no frequency weighting case. JND(W) is the noticeable pixel error calculated by multiplying the noticeable probability. We use $\alpha = 0.02$ to calculate the probability of noticeable error in Table 1.

In the frequency weighting case, most errors are small. For example, a frequency weighted image has 1377 more pixels than no frequency weighted image at difference error 1. In terms of HVS, 274 out of 1377 pixels are regarded as noticeable. From JND(W-N), we note that the frequency weighted image has much less noticeable errors than no frequency weighting. Consequently, when we apply the frequency weighting method, most errors are unnoticeable. Therefore, we can obtain improved perceptual image quality in terms of HVS.

Table 1. Number of Pixels with Difference D

W	N	W−N	D	JND (W)	JND (N)	JND (W−N)
13,341	12,150	1,191	0	13,341	12,150	1,191
23,197	21,010	1,377	1	4626	4347	274
17,211	17,222	−11	2	6790	6795	−5
12,386	12,890	−504	3	7306	7603	−298
8,898	9,443	−545	4	6986	7414	−428
6,457	6,885	−428	5	6330	6751	−421
4,747	5,055	−300	6	4747	5055	−308
3,529	3,743	−214	7	3529	3743	−214
2,654	2,831	−159	8	2654	2813	−159
2,011	2,128	−117	9	2011	2128	−117
1,535	1,628	−93	10	1535	1628	−93
1,178	1,250	−72	11	1178	1250	−72
909	966	−57	12	909	966	−57
701	744	−43	13	701	744	−43
545	580	−35	14	545	580	−35
425	449	−24	15	425	449	−24
332	347	−15	16	332	347	−15
260	269	−9	17	260	269	−9
204	209	−5	18	204	209	−5
161	164	−3	19	161	164	−3

5. EXPERIMENTAL RESULTS

With the frequency weighting method based on the sensitivity function of the human eye, we have three advantages over no frequency weighting: (1) constant image quality (2) improved visual quality, and (3) fine granular scalability.

In order to evaluate the performance of the proposed algorithm, we use FOREMAN sequence of the CIF format. We also use the frequency weighting matrix with the maximum shift factor of four and change the coding bit rate of the base layer adaptively.

Table 2 lists PSNR values and variances of 297 frames at different bit rates. In Table 2, ”Average” is the mean value of PSNR and ”Variance” represents the variance of PSNR values of images.

From Table 2, we can observe that the frequency weighting method provides relatively constant image quality. The frequency weighting method is more efficient in low bit rates of the base layer coding. If we encode the base layer at high bit rates, most low frequency components are coded in the base layer. Therefore, the characteristics of the residual signal is similar to a random noise. Table 3 shows that the frequency weighting method provides fine granular scalability.

In Table 3, we have fixed the bit rate of the base layer coding at 46.67 kbits/s. ”FW0” represents no frequency weighting method, and ”FW4” represents the frequency

Table 2. PSNR Averages and Variances

Bit Rate of the Base Layer (kbits/s)		41.35	46.82	51.99	97.7	176.8
No Frequency Weighting	Total Bit Rate	116.15	187.40	187.40	353.5	558.1
	Average	33.18	36.30	34.18	39.64	44.33
	Variance	2.43	4.40	3.12	5.95	4.60
Frequency Weighting	Total Bit Rate	131.92	165.20	168.75	284.8	401.1
	Average	32.10	34.45	32.93	37.00	39.74
	Variance	0.42	1.38	1.45	2.57	3.03

Table 3. Bit Rates (kbits/s) for Enhancement Layers

Coded Bit Plane	FW0	FW1	FW2	FW4
Base	46.67	46.67	46.67	46.67
Base+E1	65.47	65.37	64.16	57.57
Base+E2	187.40	145.54	131.35	64.26
Base+E3	455.68	360.00	239.29	165.20
Base+E4	882.77	712.09	481.32	289.05

weighting method where we use the frequency weighting matrix with maximum frequency shift of four. From Table 3, we note that the frequency weighting method provides fine granular scalability.

Fig. 7 shows the 93-th frame of the FOREMAN sequence. Fig. 7(a) is the reconstructed image with no frequency weighting that is coded at 187.40 kbits/s and has 31.92 dB in PSNR. Fig. 7(b) is the reconstructed image with frequency weighting that is coded at 187.40 kbits/s and has 31.32 dB in PSNR.

Fig. 8 shows the 104-th frame of FOREMAN sequence. Fig. 8(a) is the reconstructed image with no frequency weighting that is coded at 187.40 kbits/s and has 39.16 dB in PSNR. Fig. 8(b) is the reconstructed image with frequency weighting that is coded at 165.20 kbits/s and has 34.93 dB in PSNR.

From Fig. 7 and Fig. 8, we observe that perceptual quality of reconstructed images with frequency weighting is more acceptable than those without frequency weighting.

6. CONCLUSIONS

In this paper, we have proposed an HVS-based fine granular scalable video coding algorithm. According to the human visual sensitivity, we assign frequency weighting to each DCT coefficient. In order to apply the frequency weighting method to the bit-plane coding in the enhancement layer, we convert the frequency weighting matrix to the frequency shift matrix. Applying the frequency shift matrix in the enhancement layer, we have obtained constant and improved image quality over frames and fine granular scalability. We have also defined a new error metric, JNDE, to measure perceptual visual quality of reconstructed images. In this metric, we ignore unnoticeable errors to the human eye based on the Weber's law; therefore, we can effectively measure perceptual visual quality of images in terms of HVS.

(a) No Frequency Weighting (b) Frequency Weighting

Fig. 7. Results for the 93-th Frame

(a) No Frequency Weighting (b) Frequency Weighting

Fig. 8. Results for the 104-th Frame

7. REFERENCES

[1] W. Li, "Overview of Fine Granular Scalability in MPEG-4 Video Standard," *IEEE Trans. Circuit and System for Video Technology*, vol. 11, no 3, pp. 301-317, March 2001.

[2] 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. Multimeida*, vol. 3 no. 1, pp. 53-68, March 2001.

[3] M. Van der Schaar and H. Radah, "A Hybrid Temporal-SNR Fine Granular Scalability," *IEEE Trans.Circuit and System for Video Technology*, vol. 11, no 3, pp. 318-331, March 2001.

[4] F. Ling, W. Li and H. Sun, "Bit-plane Coding of DCT Coefficients for Image and Video Compression," *Proc. SPIC Visual Communication and Image Processing*, pp. 25-27, Jan. 1999.

[5] K. Rao and P. Yip, *Discrete Cosine Transform*, Academic Press. INC, New York, 1990.