Frequency Weighting and Selective Enhancement for MPEG-4 Scalable Video Coding

Seung-Hwan Kim and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Korea
{kshkim,hoyo}@gist.ac.kr

Abstract. In MPEG-4 scalable video coding, only a small portion of input data is coded in the base layer, and most signal components remain in enhancement layers. In this paper, we propose a new frequency weighting method to send more sensitive frequency coefficients faithfully with respect to the human visual system (HVS). In order to implement the frequency weighting method by bit-plane coding, we obtain a frequency shift matrix from the HVS-based frequency weighting matrix. We also propose a fast selective enhancement method using coding information, such as motion vectors and residual image blocks. By applying the proposed ideas, we have improved visual quality of reconstructed images. In order to measure subjective image quality appropriately, we define a new error metric, called as the just noticeable difference error (JNDE), based on the Weber's law.

Keywords: FGS, Frequency weighting, Selective enhancement, JNDE

1 Introduction

Recently, several scalable video coding schemes have been proposed for various transmission networks. One of them is the MPEG-4 fine granular scalability (FGS) scheme [1]. The FGS framework has a good balance between coding efficiency and scalability while maintaining a flexible and simple video coding structure. When compared with other error resilient streaming solutions, FGS has also demonstrated good error resilience attributes under packet losses. Moreover, FGS has recently been adopted by the MPEG-4 standard as the core coding method for video streaming applications.

Since the first version of the MPEG-4 FGS standard, there have been several improvements introduced to the FGS framework [2]. First, a very simple residual computation approach was proposed. This approach provides the same or better performance than the performance of more elaborate residual computation methods. Second, an adaptive quantization approach was proposed, and it results in two FGS-based video coding tools: frequency weighting and selective enhancement. Third, a hybrid-FGS scalability structure was also proposed. This structure enables us signal-to-noise ratio (SNR) scalable, temporal scalable, or both temporal-SNR scalable video coding and streaming [2].

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3333, pp. 159–166, 2004.

Figure 1 shows the encoder structure of the two-layer FGS system. In Fig. 1, the encoder estimates the channel capacity before encoding, and compresses the base layer using coding bits less than the channel capacity. Therefore, transmission of the base layer bitstream is always guaranteed. In the base layer, the main information of the input signal is coded in the same way as the traditional block-based coding scheme. In the enhancement layer, the residual data that is not coded in the base layer is divided into non-overlapping 8×8 blocks and each block is DCT transformed. All the 64 DCT coefficients in each block are zigzag-scanned and represented by binary numbers. These binary values form several bit-planes and entropy-coded to produce the output bitstream [1,3].

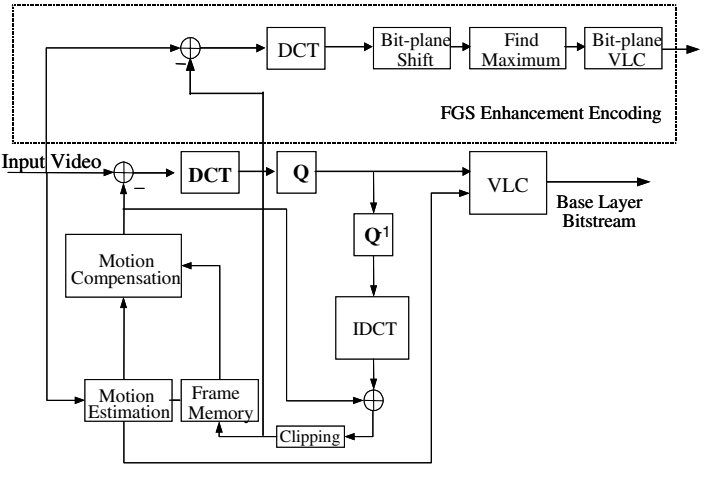

Fig. 1. FGS Encoder

Several advantages of FGS come at the expense of video quality reduction. FGS scarifies up to 2-3 dB in SNR, compared to nonscalable video coding scheme [4]. In order to overcome the performance degradation, we propose a new method of frequency weighting and selective enhancement for FGS. In the proposed frequency weighting method, we design a new frequency shifting matrix based on the human visual sensitivity function. In the selective enhancement method, the encoder decides visually important macroblocks (MB) automatically using the motion vector and position information of MB. We also define a new error metric to measure subjective image quality.

The paper is organized as follows. In Section 2, we describe a frequency weighting method based on the human visual system (HVS) and its implementation by bit-plane coding. In Section 3, we explain a fast selective enhancement method using coding information, such as the motion vector and the position information of each MB. In Section 4, we propose a new error metric to estimate the subjective image quality. After experimental results are presented in Section 5, we conclude this paper in Section 6.

2 HVS-Based Frequency Weighting

In general, human eyes are more sensitive to low frequency components than to high frequencies [5]. In order 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

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

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][6]. Each 8×8 DCT coefficient is multiplied by the corresponding element of the frequency weighting matrix, reflecting their importance on HVS. Fig. 2 shows a typical frequency weighting matrix [5].

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-based frequency weighting, we multiply each DCT coefficient by its corresponding element of the frequency weighting matrix. Therefore, the frequency weighted DCT coefficient is described by

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

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

We also convert the frequency weighting matrix to the frequency shift matrix. In order to make an appropriate mapping, we select the maximum shift factor *maxn(fw)* that represents the number of bits to be shifted up at the most important DCT coefficient. In the frequency weighting matrix, weighting values are normalized by one. Therefore, the frequency weighting matrix should be multiplied by $2^{maxn(fw)}$. After scaling the frequency weighting matrix, we transform it to the frequency shift matrix. As a result, the frequency shift matrix is obtained by

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

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. 3. Frequency Shift Matrix

where $n_{fw}(i)$ is a shift factor at the i-th DCT coefficient and $2^{maxn(fw)} \cdot fw(i)$ is the scaled frequency weighting. Figure 3 shows the frequency shift matrix with $maxn(fw)$=3.

Figure 4 represents the proposed frequency weighting process in the FGS enhancement layer, where we choose four for the maximum shift factor for the DC component.

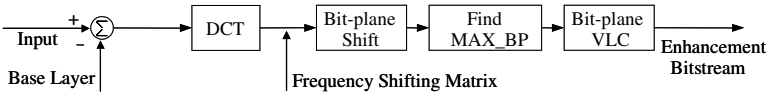

Fig. 4. Frequency Weighting for Enhancement Layer

3 Fast Selective Enhancement

In this section, we propose a fast *selective enhancement* (SE) algorithm using coding information, such as the motion vector and the location of MB, which can easily be extracted during the encoding process. Using this information, we can estimate the importance of each MB by

$$SE = P(x, y) \times ABS(mv_x) + ABX(mv_y) \tag{4}$$

where SE is the importance of the given MB, P(x ,y) is the position of the MB, $ABS(MV)$ is the absolute value of the motion vector. However, if we use only the coding information, we may miss some visually important MBs. Generally, if an MB is surrounded by visually important MBs, we can regard the MB as a visually important MB. Therefore, we apply lowpass filtering to SE values in each MB, as illustrated in Fig. 5

In Fig. 5, Vu, Hl, Hr, and Vd represent SE values of the surrounding MBs. Lowpass filtering is performed by

$$SE = (2SE + (Vu + Vd + Hl + Hr))/6 \tag{5}$$

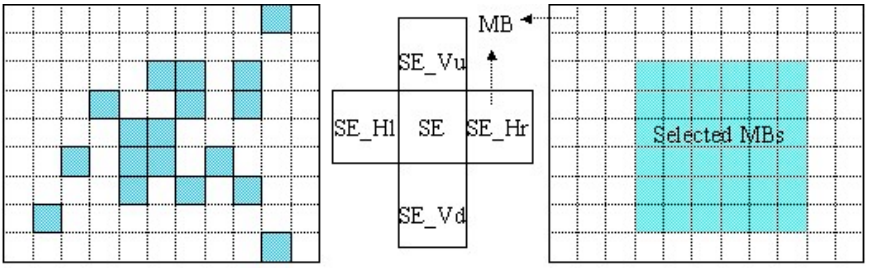

Fig. 5. Selective Enhancement Method

4 Perceptual Visual Quality

In this section, we define a new error metric to measure the subjective image quality based on the human visual system (HVS).

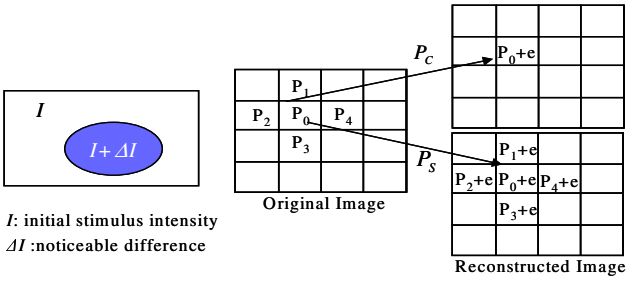

Fig. 6. Weber's Law and JNDE

According to the Weber's law, illustrated in Fig. 6, the minimum noticeable difference is proportional to the background intensity [7].

$$\frac{\Delta I}{I} = \alpha \tag{6}$$

In order to find the noticeable probability from the effect of the original pixel value, we change the Weber's law as follows [6]

$$\frac{\Delta I}{I} = \frac{D}{P} \geq \alpha \tag{7}$$

where p is the original pixel value and D is the difference between the original and its reconstructed values at a given pixel position. If the original image has a uniform distribution, the probability that the original pixel value is lower than the maximum threshold value p_{ths} is represented by

$$P_C = P(p \leq p_{ths}) = \frac{D/\alpha + 1}{2^n} \tag{8}$$

where n represents the number of bits assigned to each pixel. The noticeable probability P_S from the effect of the surrounding pixel values is

$$P_S = \sum_{k=1}^{4} k/4 \cdot {}_4C_k (P_e)^k \cdot (1 - P_e)^{4-k} \tag{9}$$

where P_e represents the noticeable probability between the given error pixel and one of the neighboring pixels. *k/4* is the weighting factor for the number of k surrounding noticeable errors. As a result, the total noticeable probability P_{JNDE} of the given difference D is [6]

$$P_{JNDE} = P_C \cdot P_S \tag{10}$$

Until now, we introduce the just noticeable difference error (JNDE) using the Weber's law. We can also represent the peak signal-to-noise ratio (PSNR) by

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \tag{11}$$

where *MSE* represents the mean square error between the original and reconstructed images. In other words, *MSE* represents the average noise power. Applying the Parseval's theorem, *MSE* can be represented in the frequency domain.

$$MSE = \sum_{k=0}^{M-1} \sum_{v=0}^{N-1} \sum_{i=0}^{L-1} \alpha \cdot F_o(i, v, k) - F_r(i, v, k)^2 \tag{12}$$

where α represents a scaling factor between the frequency and spatial domains, and $F_o(i, v, k)$ and $F_r(i, v, k)$ represent DCT coefficients in the original and reconstructed images, respectively. In Eq. (12), M is the number of MBs in the frame, and N is the number of blocks in each MB. L denotes the number of pixels in the block. Consequently, we define a new error metric P_{HVS} by

$$P_{HVS} = \sum_{k=0}^{M-1} \sum_{v=0}^{N-1} \sum_{i=0}^{L-1} \alpha \cdot \frac{F_o^2(i, v, k)}{fw(i) \cdot \{F_o(i, v, k) - F_r(i, v, k)\}^2} \tag{13}$$

where $fw(i)$ the weighting factor obtained from the frequency weighing matrix in Fig. 2.

5 Experimental Results

In order to evaluate the performance of the proposed algorithm, we use the FOREMAN sequence, whose resolution is 352 × 288 pixels (CIF). Table 1 lists bit rates for the enhancement layers, where *FW0*, *FW1*, *FW2*, and *FW4* represent the maximum shift factor=0, 1, 2, and 4, respectively. Table 1 indicates that the frequency weighing method provides finer scalability than no frequency weighting method.

Table 1. Bit Rates for Enhancement Layers

Coded Bit-Plane	FW0	FW1	FW2	FW4
Base(kbit/s)	373	373	373	373
Base+E1	523	522	513	460
Base+E2+E2	1499	1164	1050	754
Base+E1+E2+E3	3645	2880	1914	1321
Base+E1+E2+E3+E4	7061	5696	3850	2312

Figure 7 shows the 6^{th} frame of the FOREMAN sequence. Fig. 7(a) is the reconstructed image with no frequency weighting, coded at 187.4 kbps. Fig. 7(b) is the reconstructed image with frequency weighting, coded at 165.2 kbps. From Fig. 7, we observe that perceptual quality of reconstructed images with frequency weighting is more acceptable than those without frequency weighting.

(a) (b)

Fig. 7. Comparison of Subjective Image Quality

Table 2. Number of Noticeable Errors

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	4,626	4,347	274
17,211	17,222	-11	2	6,790	6,795	-5
12,386	12,890	-504	3	7,306	7,603	-298
8,898	9,443	-545	4	6,986	7,414	-428
6,457	6,885	-428	5	6,330	6,751	-421
4,747	5,055	-300	6	4,747	5,055	-308
3,529	3,743	-214	7	3,529	3,743	-214

Table 2 lists the number of pixels at a given error (D) in both the frequency weighing case (W) and no frequency weighting case (N). We use α=0.02 to calculate the probability of noticeable error ($JND\ (W, N)$). $JND\ (W)$ is obtained by multiply W with P_{JND}, which is calculated by Eq. (10). In the frequency

weighting case, most errors are concentrated in the small error (D): therefore, we can obtain perceptually improved image quality in terms of HVS.

6 Conclusions

In this paper, we have proposed an HVS-based frequency weighting and a fast selective enhancement methods. In the proposed frequency weighting method, we assign frequency weighting to each DCT coefficient according to the human visual sensitivity function. We also convert the frequency weighting matrix to the frequency shift matrix to apply the frequency weighting method to the bit-plane coding. In the proposed selective enhancement method, we only use the coding information obtained in the encoding process. With the proposed ideas, we have obtained perceptually improved image quality. We have also defined a new error metric to measure perceptual visual quality of reconstructed images, both in the time and frequency domains.

Acknowledgements. This work was supported in part by Gwangju Institute of Science and Technology (GIST), in part by the Ministry of Information and Communication (MIC) through the Realistic Broadcasting Research Center (RBRC) at GIST, and in part by the Min-istry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

1. Li, W.: Overview of Fine Granular Scalability in MPEG-4 Video Standard. IEEE Trans. on Circuit and System for Video Technology (2001) 301–317
2. Radah, H., Van der Schaar, M., and Chen, Y.: The MPEG-4 Fine Grained Scalable Video Coding Method for Multimedia Streaming over IP. IEEE Trans. Multimedia (2001) 53–68
3. Van der Schaar, M. and Radah, H.: A Hybrid Temporal SNR Fine Granular Scalability. IEEE Trans. Circuit and System for video Technology (2001) 318–331
4. Ling, F., Li, W., and Sun, H.: Bit-Plane Coding of DCT Coefficients for Image and Video Compression. Proc. SPIE, Visual Communication and Image Processing (1999) 25–27
5. Rao, K. and Yip, P.: Discrete Cosine Transform. Academic Press, New York, (1990)
6. Kim, S.H. and Ho, Y.S.: HVS-Based Frequency Weighting for Fine Granular Scalability. Proc. Information and Communication Technologies (2003) 127–131
7. Anil, K.J.: Fundamentals of Digital Image Processing. Prentice-Hall, (1989) 51