

Optical Engineering

SPIDigitalLibrary.org/oe

Frequency sensitivity for video compression

Guiwon Seo
Jonghwa Lee
Chulhee Lee

Frequency sensitivity for video compression

Guiwon Seo, Jonghwa Lee, and Chulhee Lee*

Yonsei University, Department of Electrical and Electronic Engineering, 134 Shinchon-Dong, Seodaemun-Gu, Seoul, Republic of Korea

Abstract. We investigate the frequency sensitivity of the human visual system, which reacts differently at different frequencies in video coding. Based on this observation, we used different quantization steps for different frequency components in order to explore the possibility of improving coding efficiency while maintaining perceptual video quality. In other words, small quantization steps were used for sensitive frequency components while large quantization steps were used for less sensitive frequency components. We performed subjective testing to examine the perceptual video quality of video sequences encoded by the proposed method. The experimental results showed that a reduction in bitrate is possible without causing a decrease in perceptual video quality. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.53.3.033107](https://doi.org/10.1117/1.OE.53.3.033107)]

Keywords: video compression; frequency sensitivity; HVS; coding efficiency; perceptual video quality.

Paper 131611 received Oct. 22, 2013; revised manuscript received Feb. 2, 2014; accepted for publication Feb. 19, 2014; published online Mar. 25, 2014.

1 Introduction

High definition (HD) video services have become widely available in recent years and demands for high quality video services have also been rapidly increasing. Moreover, it is expected that the storage space and transmission bandwidth capacity will lead to further increases in the production, storage, and delivery of high quality video services. For example, uncompressed HD video signals require about 1 Gbps and uncompressed ultra high definition (UHD) video signals require about 4 Gbps (UHD-4k, 3840×2160) or 15 Gbps (UHD-8k, 7680×4320). Therefore, video compression technology is essential for high quality video services. Due to this, a number of international standards have been established, such as Moving Picture Experts Group (MPEG)-2, MPEG-4, H.263, and H.264. Recently, the MPEG and the international telecommunication union telecommunication standardization sector (ITU-T) video coding experts group have been jointly developed for high efficiency video coding (HEVC) standards.

In conventional video compression methods, the goal is to minimize the mean squared error (MSE) or sum of absolute difference (SAD) metrics. These metrics have typically been used for rate-distortion (RD) optimization. However, it has been reported that MSE and SAD metrics do not accurately represent perceptual quality. Therefore, MSE has sometimes been replaced with other metrics that better reflect perceptual images or video quality.¹ For example, the just noticeable distortion (JND) model estimation was implemented on a H.264/AVC system to optimize the spatial-temporal human visual systems.² Also, in recent coding standard activities, perceptual evaluation has also been used together with peak signal-to-noise ratios (PSNRs).

There have been several attempts to measure the relationship between perceptual quality and spatial frequency. Since the discrete cosine transform (DCT) method has been widely used in most video compression standards such as H.261, H.263, H.264, MPEG1, MPEG2, and MPEG4, JND models

based on the DCT domain have been studied. In Ref. 3, distortion visibility thresholds for DCT coefficients were approximated using luminance-based models. The DCTune model uses luminance adaptation and contrast masking effects to optimize the JPEG DCT quantization matrix.^{4,5} A block classification based DCTune was proposed to improve the perceptual image coding performance in Ref. 6. A DCT-based JND model for monochrome image/video using contrast sensitivity functions was proposed in Ref. 7. In Ref. 2, quantization steps for low frequencies were allocated smaller values than for higher frequencies. These findings about frequency sensitivity have mainly focused on image coding and have used 8×8 float DCT coefficients.

There has been a research in spatial frequency sensitivity.⁸⁻¹⁰ Based on subjective experiments where specially designed patterns were used, the contrast sensitivity function of spatial frequency was defined as the sensitivity level according to the spatial frequency.⁸ All the previous experiments were performed using simulated one-dimensional (1-D) signals. However, video coding usually deals with two-dimensional (2-D) frequency sensitivities. Thus, those previous research results may not be directly applicable to image or video coding. Also, most coding methods use block transforms such as DCT, where each coefficient represents a 2-D frequency component. Our preliminary experiments showed that errors in middle frequencies did not cause less severe perceptual degradation compared to errors in lower and higher frequencies. This observation will be discussed in greater detail in Sec. 3.

In this paper, we investigate the frequency sensitivity of the human visual system in video coding through extensive subjective testing. We observed that the human visual system reacts differently at different frequencies. Therefore, we used different quantization steps for different frequency components. In other words, small quantization steps were used for sensitive frequency components while large quantization steps were used for less sensitive frequency components. Joint model (JM), the reference encoder of the H.264/AVC standard, was used for test frequency sensitivity.

*Address all correspondence to: Chulhee Lee, E-mail: chulhee@yonsei.ac.kr

2 Quantization of DCT Coefficients

The DCT is widely used in numerous applications in image and video lossy compression technologies. For example, DCT is used in video compression standards such as MPEG-2, MPEG-4, H.261, H.263, H.264, and HEVC. DCT helps to separate images into spectral subbands of differing importance. Lower frequency components are more important than higher frequency components for video quality. Moreover, in most video data, low frequency components are dominant. The forward and inverse DCT values can be defined as follows:

$$\begin{aligned}
 Y &= AXA^T, \quad A_{ij} = C_i \cos \frac{(2j+1)i\pi}{2N}, \\
 &\left(\text{if } i = 0, \quad C_i = \sqrt{\frac{1}{N}}, \quad \text{otherwise } C_i = \sqrt{\frac{2}{N}} \right) \\
 X &= A^T Y A. \tag{1}
 \end{aligned}$$

Using this definition, the 4×4 DCT can be specified as follows:

$$\begin{aligned}
 Y &= AXA^T \\
 &= \begin{pmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{pmatrix} \times \begin{pmatrix} a & b & a & c \\ a & c & -a & -b \\ a & -c & -a & b \\ a & -b & a & -c \end{pmatrix} \\
 a &= \frac{1}{2}, \quad b = \sqrt{\frac{1}{2}} \cos\left(\frac{\pi}{8}\right), \quad c = \sqrt{\frac{1}{2}} \cos\left(\frac{3\pi}{8}\right). \tag{2}
 \end{aligned}$$

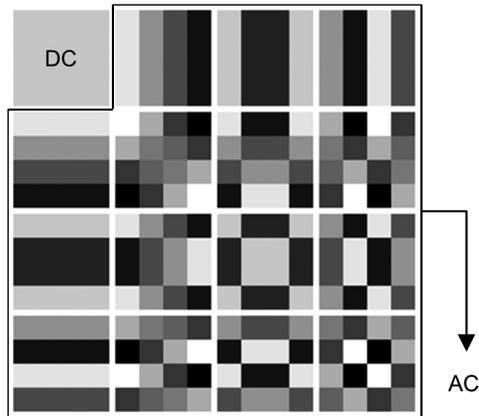


Fig. 1 4×4 DCT basis patterns.

Figure 1 shows the basis functions of the 4×4 DCT. In $N \times N$ DCT applications, there is one DC component and $N \times N - 1$ AC components. The energy of the DC component is dominant in most cases. Typically, the energies of the AC components are smaller compared with the DC components. This energy compaction property has been exploited in compression methods along with the quantization technique.

The quantization process is essential to compress video data. Quantized coefficients are usually computed as follows:

$$Z_{ij} = \text{round}(Y_{ij}/Q\text{step}), \tag{3}$$

where Y represents a DCT coefficient while Z represents a quantized DCT coefficient. $Q\text{step}$ represents a quantization step. If the quantization step is large, the quantization error increases. However, the video compression ratio improves with a large quantization step since most quantized coefficients are zero. Therefore, video quality and video bitrates can be controlled using different quantization steps. In H.264 standards, 52 different quantization step values are provided. These quantization steps are specified with different quantization parameters (QP). Table 1 shows the quantization step size according to the QP value used in the H.264 standard.¹¹

In some image and video coding standards, an optional quantization technique using a quantization matrix (q-matrix) is provided. In this basic quantization method, the same quantization step is adapted to all DCT coefficients. The q-matrix provides a full matrix for the quantization modification coefficients. Different quantization steps for different DCT coefficients can be used with the q-matrix. Figure 2 shows how the q-matrix can be used for 4×4 DCT coefficients. The q-matrix is inserted in the compressed

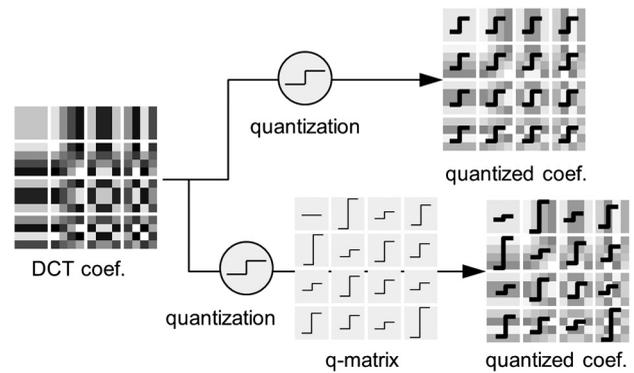


Fig. 2 Quantization with q-matrix.

Table 1 Quantization step size according to the QP value used in the H.264 standard.

QP	0	1	2	3	4	5	6	7	8	9	10	...
Qstep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	...
QP	...	18	...	24	...	30	...	36	...	42	...	51
Qstep	...	5	...	10	...	20	...	40	...	80	...	224

bit stream header. This matrix should be designed to achieve maximum perceptual quality with high compression efficiency.

3 Frequency Modeling for DCT

The 2-D DCT coefficients were represented as a 1-D vector using the zigzag scanning method. Figure 3 shows the 1-D frequency representation of the 2-D DCT coefficients.

To examine the amount of human frequency sensitivity on video coding, several quantization methods were designed and tested in the experiments. We performed a number of subjective tests to evaluate the perceptual quality of various frequency quantization models. In our subjective tests, we used three or four different models per session.

Previous research on spatial frequency sensitivity has concluded that middle range frequencies are more sensitive to human perception than low and high range frequencies.⁸ However, this research used simple sinusoidal patterns (Fig. 4) to investigate spatial frequency sensitivity and is not always directly applicable to video coding in real-world situations. Figure 5 shows three images with the same PSNR with degradations in different frequency ranges. It can be seen that the image with degradations in the middle frequencies [Fig. 5(b)] shows better perceptual quality than the images with degradations in the low

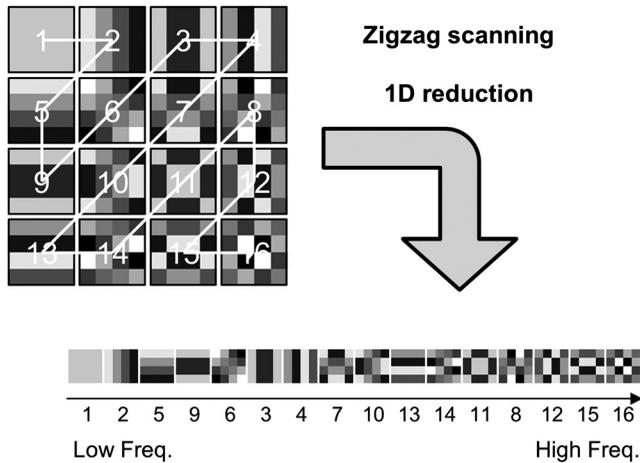


Fig. 3 One-dimensional (1-D) frequency representation with zigzag scanning.

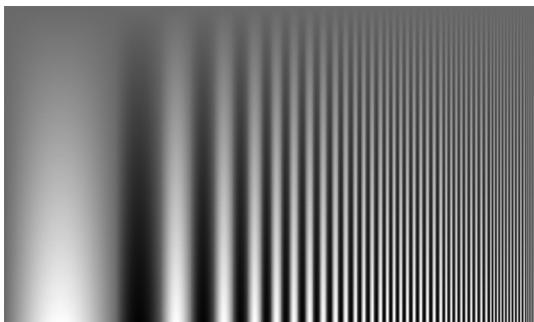


Fig. 4 Spatial frequency pattern used in Ref. 8.

or high frequencies [Figs. 5(a) and 5(c)]. Based on this observation, we used larger quantization steps for middle frequency coefficients.

In the first set, four different frequency quantization methods were designed as shown in Fig. 4. In the proposed methods, the quantization multiplier was used to adjust the quantization step as follows:

$$Q_{step_{qm}} = Q_{step_{origin}} \times \text{Quantization multiplier.} \quad (4)$$

Consequently, a large value of the quantization multiplier resulted in a large quantization step, which produced smaller compressed data at low-image quality. If the value of the quantization multiplier was 1, the original quantization step was used.

In Fig. 6(a), a trapezoid multiplier function is shown. In this model, the middle frequency components were more coarsely quantized than the lower or higher frequency components. In Fig. 6(b), a triangle multiplier function is shown. In the triangle function (triangle mode 1), the middle frequency components were also more coarsely quantized similar to the trapezoid function with a peak point. Figure 6(c) shows a linearly increasing function and Fig. 6(d) shows a linearly decreasing function. These four frequency quantization multiplier functions were calculated as follows:

$$Q_{trapezoid}(x) = \begin{cases} 4.125x + 1 & \text{for } 0 \leq x < \frac{1}{3} \\ 2.375 & \text{for } \frac{1}{3} \leq x \leq \frac{2}{3}, \\ -4.125x + 5.125 & \text{for } \frac{2}{3} < x \leq 1 \end{cases} \quad (5)$$

$$Q_{triangle_I}(x) = \begin{cases} 2.946x + 1 & \text{for } 0 \leq x < \frac{1}{2} \\ -2.946x + 3.946 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (6)$$

$$Q_{linear_increase}(x) = 1.375x + 1 \quad \text{for } 0 \leq x \leq 1, \quad (7)$$

$$Q_{linear_decrease}(x) = -1.375x + 2.375 \quad \text{for } 0 \leq x \leq 1. \quad (8)$$

The second set (set 2) of multiplier functions is shown in Fig. 7. In this set, various shapes for middle frequencies were tested. The triangle model shown in Fig. 7(a) was the same as the triangle function in Fig. 6(b) of set 1. Also, the linear increasing function shown in Fig. 7(d) was also the same as the linearly increasing function of set 1. The triangle mode 2 function shown in Fig. 7(b) was a combined model of the triangle mode 1 and the linear increasing functions. In other words, in ascending parts (low frequencies), the function used the linearly increasing function in Fig. 7(d) while the triangle mode 1 function was used in descending parts (high frequencies). Figure 7(c) shows another modified version of the triangle function. This triangle mode 3 preserved the high frequency components. The two new frequency quantization functions were calculated as follows:

$$Q_{triangle_II}(x) = \begin{cases} 1.375x + 1 & \text{for } 0 \leq x \leq \frac{2}{3} \\ -2.946x + 3.946 & \text{for } \frac{2}{3} < x \leq 1 \end{cases} \quad (9)$$

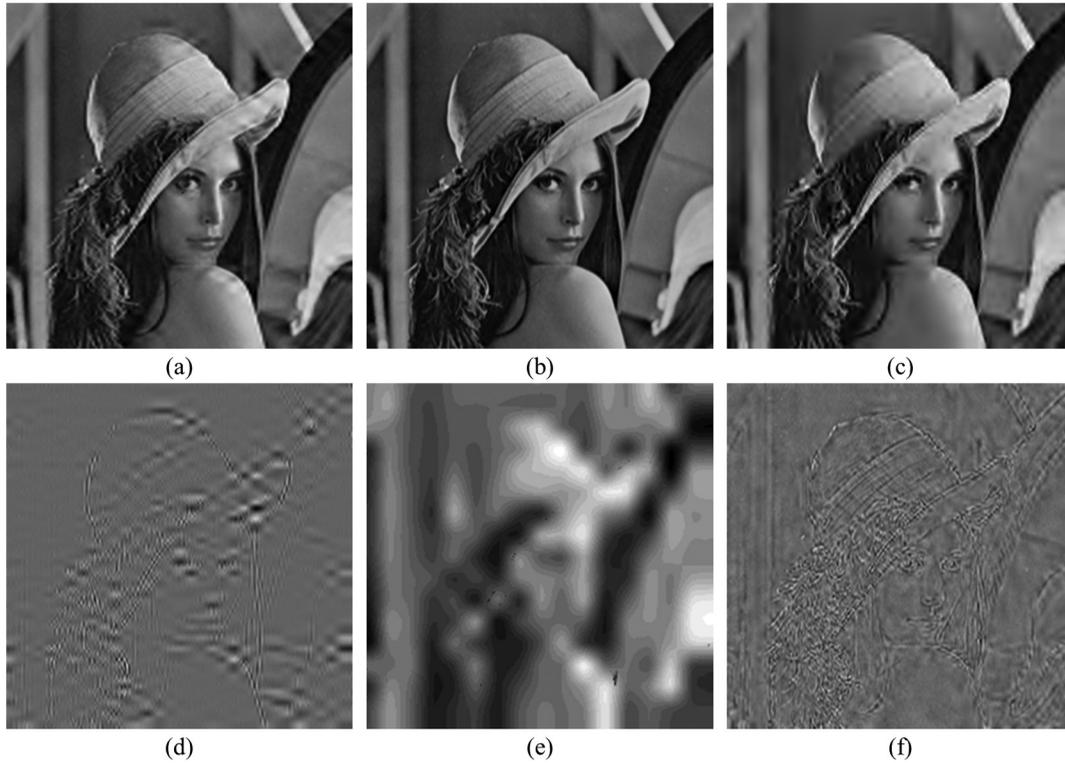


Fig. 5 Degraded Lena images with the same PSNR (29.51 dB) (a) low frequency degradations, (b) middle frequency degradations, (c) high frequency degradations, (d) difference image of low frequency degradations, (e) difference image of middle frequency degradations, and (f) difference image of high frequency degradations.

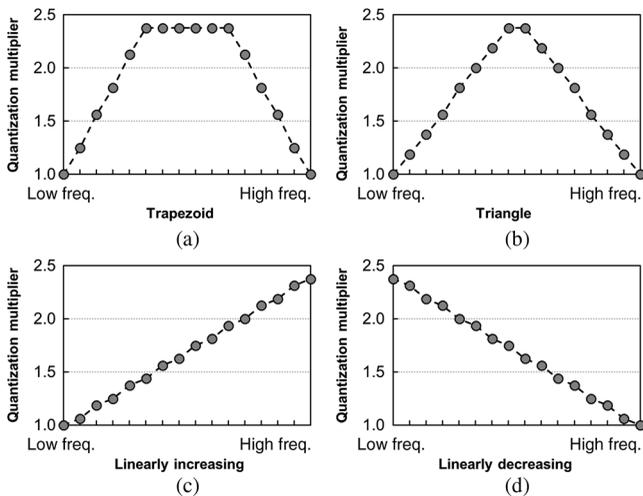


Fig. 6 Quantization multiplier functions (set 1) for H.264/AVC 4 × 4 DCT coefficients: (a) trapezoid, (b) triangle mode 1, (c) linearly increasing, and (d) linearly decreasing.

$$Q_{\text{triangle_III}}(x) = \begin{cases} 1.375x + 1 & \text{for } 0 \leq x \leq \frac{2}{3} \\ 1 & \text{for } \frac{2}{3} < x \leq 1 \end{cases} \quad (10)$$

Figure 8 shows the third set of frequency quantization functions (set 3) using three different triangle functions with different peak values. Figure 8(b) was the same as the triangle mode 1 function. These three frequency quantization multiplier functions were calculated as follows:

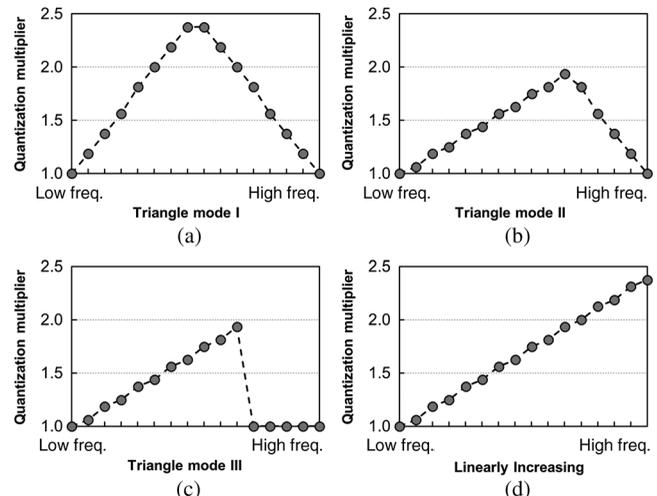


Fig. 7 Quantization multiplier functions (set 2) for H.264/AVC 4 × 4 DCT coefficients: (a) triangle mode 1, (b) triangle mode 2, (c) triangle mode 3, and (d) linearly increasing.

$$Q_{\text{triangle_low}}(x) = \begin{cases} 1.875x + 1 & \text{for } 0 \leq x < \frac{1}{2} \\ -1.875x + 2.875 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (11)$$

$$Q_{\text{triangle_med}}(x) = \begin{cases} 2.946x + 1 & \text{for } 0 \leq x < \frac{1}{2} \\ -2.946x + 3.946 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (12)$$

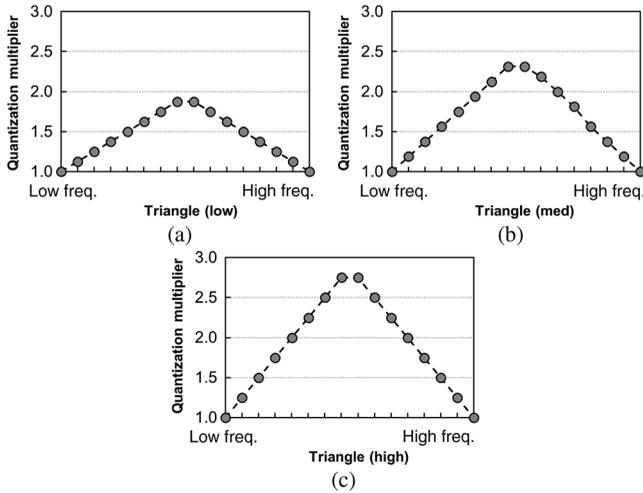


Fig. 8 Quantization multiplier functions (set 3) for H.264/AVC 4×4 DCT coefficients: (a) triangle (low), (b) triangle (med), and (c) triangle (high).

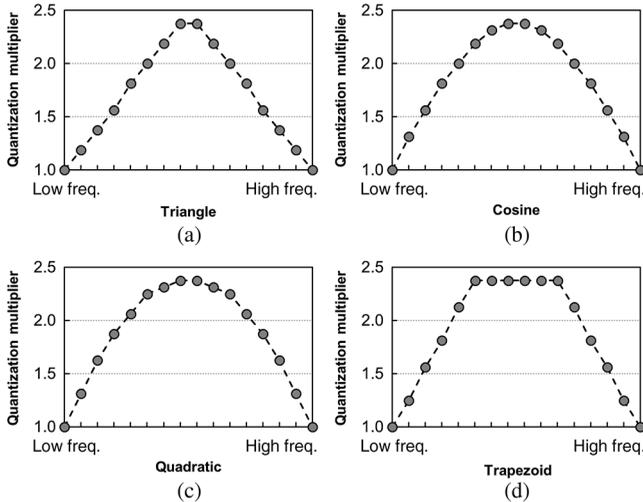


Fig. 9 Quantization multiplier functions (set 4) for H.264/AVC 4×4 DCT coefficients: (a) triangle, (b) cosine, (c) quadratic, and (d) trapezoid.

$$Q_{\text{triangle_high}}(x) = \begin{cases} 3.75x + 1 & \text{for } 0 \leq x < \frac{1}{2} \\ -3.75x + 4.75 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (13)$$

Figure 9 shows the fourth set (set 4) of the frequency quantization functions, which includes two additional

functions with coarse quantization for the middle frequencies. The triangle and trapezoid functions were identical to those of Fig. 6. The two new frequency quantization functions were calculated as follows:

$$Q_{\cos}(x) = 1.375 \cos[(x - 0.5)\pi] + 1 \quad \text{for } 0 \leq x \leq 1, \quad (14)$$

$$Q_{\text{quad}}(x) = -6.2578x^2 + 5.8667x + 1 \quad \text{for } 0 \leq x \leq 1. \quad (15)$$

A total of nine quantization multiplier functions were designed. Since a large value of the quantization multiplier function produced a large quantization step, the area of the quantization multiplier function was related to the average quantization step. Table 2 shows the areas of the nine multiplier functions. The triangle (high) quantization showed the largest area while the triangle mode 3 quantization showed the smallest area. The linearly increasing quantization had the same area as the linearly decreasing quantization. However, the linearly decreasing function produced smaller compressed data than the linearly increasing function since the energy of the low frequency components was dominant.

4 Subjective Assessments

Subjective quality assessment was performed to investigate the frequency sensitivity of the human visual system. Six subjective tests were conducted using the frequency quantization sets. In each subjective test, four QPs were selected, which reflected various levels of coding quality. Table 3 shows the test designs. In each test design, three different conditions were considered: source video sequences, QPs, and quantization methods.

In the HD test 1 experiment, nine source video sequences, four QPs, and five quantization methods were used. The nine source video sequences of full HD (1920×1080) were selected based on compression difficulties. Each source video sequence was 8-s long with 30 fps (240 frames). The default setting of H.264/AVC (JM 15.1) was used for the original quantization method. The frequency quantization set 1 was used in this design along with the original quantization method (default setting). In the test, 189 processed video sequences (PVS) were generated according to the experimental design. The number of PVSs was calculated as follows:

Table 2 Areas of the quantization multiplier functions.

	Linear (up, down)	Trapezoid	Triangle I	Triangle II	Triangle III
Area	0.6875	0.9167	0.7362	0.4692	0.3056
	Triangle (low)	Triangle (high)	Cosine	Quadratic	
Area	0.4688	0.9375	0.8754	0.8474	

Table 3 Subjective test designs for various frequency quantization sets.

Resolution	Source video	QP	Quantization methods	
HD	Test 1	9 SRCs	27, 32, 37, 42	Reference (uncompressed), original (JM 15.1), trapezoid triangle, linearly increasing, linearly decreasing
	Test 2	9 SRCs	25, 28, 31, 34	Reference (uncompressed), original (JM 15.1), triangle mode I, triangle mode II, triangle mode III, linearly increasing
VGA	Test 1	9 SRCs	27, 32, 37, 42	Reference (uncompressed), original (JM 15.1), linearly increasing, linearly decreasing, triangle
	Test 2	9 SRCs	29, 33, 37, 42	Reference (uncompressed), original (JM 15.1), triangle mode I, triangle mode II, triangle mode III
	Test 3	9 SRCs	22, 27, 32, 37	Reference (uncompressed), original (JM 15.1), triangle (low), triangle (med), triangle (high)
	Test 4	7 SRCs	22, 27, 32, 37	Reference (uncompressed), original (JM 15.1), triangle, cosine, quadratic, trapezoid

$$9\text{SRCs} \times (4\text{QPs} \times 5\text{Quan.Methods} + \text{REF}) = 189\text{PVSs.} \tag{16}$$

In the subjective test of HD test 2, the same source video sequences of HD test 1 were used. Four QPs were used in this test. In this test, denser QPs were used than the QPs of HD test 1. In HD test 2, the difference between adjacent QPs was 3 while the difference was 5 in HD test set 1. The frequency quantization set 2 was used in HD test 2 along with the original quantization method. In this design, 189 PVSs were generated.

In the subjective test of video graphics array (VGA) test 1, the frequency quantization set 1 was used with VGA (640 × 480) source videos. Each source video sequence was 12-s long with 30 fps (360 frames). Four quantization methods are selected in quantization set 1. The QPs, frame rates, and length of video clips were the same as those of HD subjective test 1. However, different resolution and source contents were used. In the subjective test of VGA test 2, the frequency quantization set 2 was used with VGA (640 × 480) source video sequences. Four quantization methods were selected in the quantization set 2. In the subjective test of VGA test 3, the frequency quantization set 3 was used with VGA (640 × 480) source video sequences. Nine source video sequences, four QPs and four quantization methods were used in this subjective test. In this test, 153 PVSs were generated. In the subjective test of VGA test

4, the frequency quantization set 4 was used with VGA (640 × 480) source video sequences. Seven source videos, four QPs, and five quantization methods were used in VGA test 4. In this design, 147 PVSs were generated.

The viewing environments were set in accordance with ITU-T and ITU-R standards.^{12,13} Lighting and display characteristics were tuned according to the standard specifications. Figure 10 shows the viewing distance setting in the subjective quality assessment in the HD tests. The distance between the display monitor and the viewers was set to 3H, where H represents the height of the display monitor. Two viewers watched the video sequences at the same time in the HD tests.

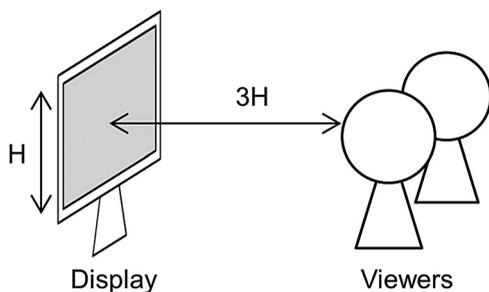


Fig. 10 Viewing distance in terms of subjective quality assessment.

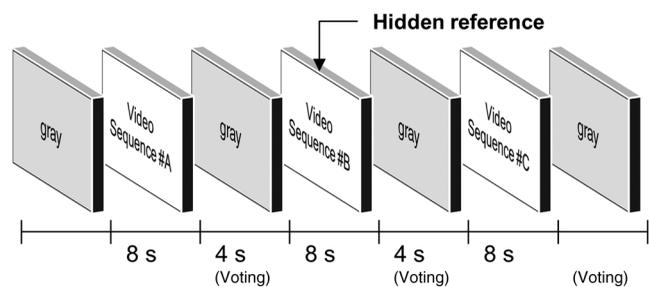


Fig. 11 The viewing order and viewing time of the ACR-HR method.



Fig. 12 Quality grades used in the ACR-HR method.

To evaluate video subjective quality, the absolute category rating–hidden reference (ACR–HR) method was used. Figure 11 shows an example of the viewing order of the ACR–HR method. In this method, each video was played once in a random order. Also, reference video sequences were hidden in the video clips. Viewers did not know which video sequence was a reference video sequence. Between the video sequences, gray videos were inserted. When gray videos were played, viewers rated the video quality.

In the quality ratings, every video sequence was rated in terms of five categories as shown in Fig. 12: excellent, good, fair, poor, and bad. Results were converted into numerical scores on a 1 to 5 scale. A single score for one video

sequence was calculated by averaging all the numerical scores of 24 viewers. Then, the difference mean opinion scores (DMOS) were calculated as follows:

$$REF[i] = \frac{1}{N} \sum_{\langle N \rangle} REF_{viewer}[i, k],$$

$$PVS[i, j] = \frac{1}{N} \sum PVS_{viewer}[i, j, k],$$

$$DMOS[i, j] = PVS[i, j] - REF[i] + 5, \tag{17}$$

where N is the number of viewers, i is the source index, j is the HRC index, and k is the viewer index.

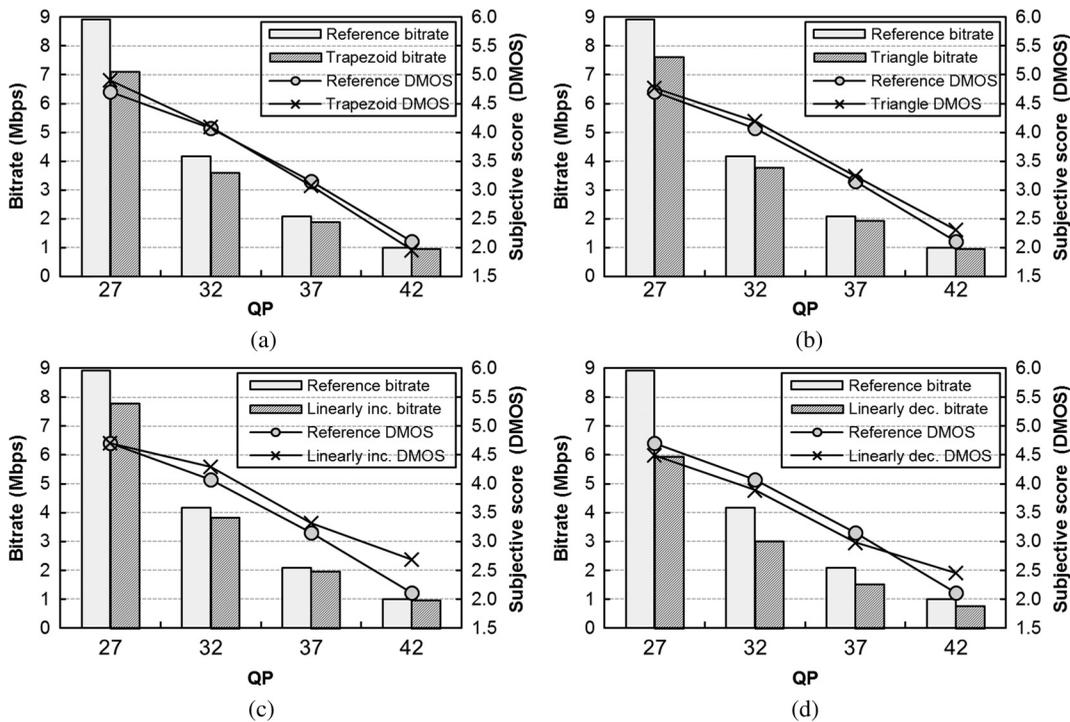


Fig. 13 Bitrate and subjective score comparison of the frequency quantization set 1 (HD): (a) trapezoid, (b) triangle, (c) linearly increasing, and (d) linearly decreasing.

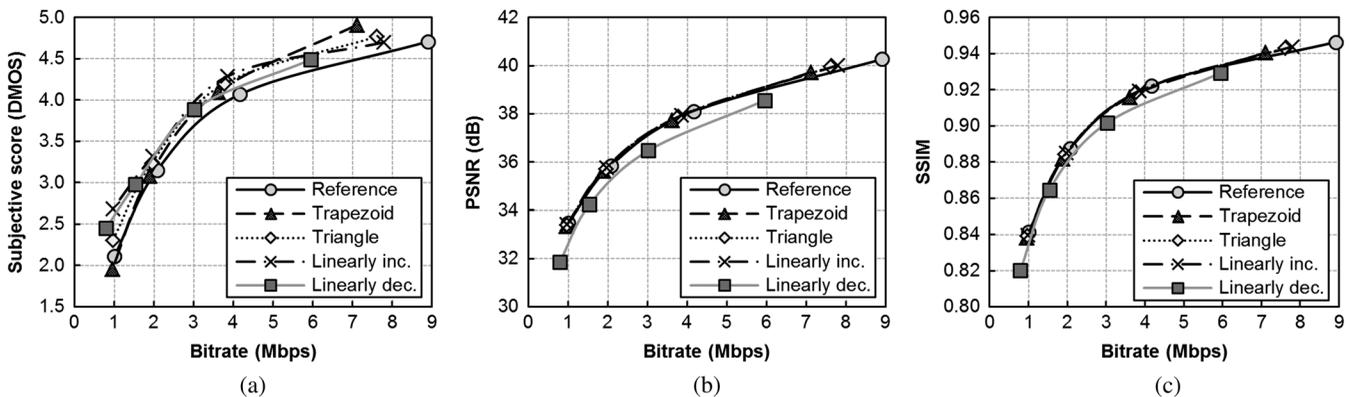


Fig. 14 Performance comparison for the frequency quantization set 1 (HD): (a) subjective scores, (b) PSNR, and (c) SSIM.

5 Experimental Results

Figure 13 shows the experimental results of the bitrates and the subjective quality rating of the frequency quantization set 1 with HD clips. In this experiment, the linearly decreasing model produced the lowest bitrates among the four quantization methods as shown in Fig. 13(d), while the linearly increasing model produced the highest bitrates as shown in Fig. 13(c). However, the linearly increasing model showed the best subjective quality while the linearly decreasing model showed the worst subjective quality. Since the low frequency components in the DCT domain had higher energy levels than the high frequency components, a large bitrate reduction of the linearly decreasing model was expected. The subjective scores (DMOS) were generally proportional

to the bitrate reduction ratio. Figure 14 shows a performance comparison in terms of the subjective scores (DMOS), PSNR, and SSIM. The Structural SIMilarity (SSIM) measures structural similarity between two images. It is known that SSIM is better correlated with the human visual system than PSNR.¹⁴ It appears that all the frequency quantization functions of set 1 produced subjective scores that were better than those of the reference method as shown Fig. 14(a). Except for the linearly decreasing model, the frequency quantization functions showed similar performance compared to the reference model in terms of PSNR and SSIM.

To investigate the coding efficiency, the bitrates of the quantization functions, which produced equivalent perceptual

Table 4 Bitrate reduction ratio for each quantization set.

SET1					
Resolution	Quantization function	Trapezoid (%)	Triangle (%)	Linearly increasing (%)	Linearly decreasing (%)
HD	DMOS	-10.37	-15.92	-21.53	-17.58
	PSNR	-2.80	-2.74	-2.39	21.10
	SSIM	-0.47	-1.23	-1.21	10.39
VGA	DMOS	-	-8.15	-8.81	15.30
	PSNR	-	-0.98	-2.43	31.37
	SSIM	-	-3.47	-3.97	17.97
SET2					
Resolution	Quantization function	Triangle I (%)	Triangle II (%)	Triangle III (%)	Linearly increasing (%)
HD	DMOS	-7.82	-4.35	-13.55	-2.11
	PSNR	-3.64	-3.40	-3.08	-3.56
	SSIM	-1.76	-2.34	-1.95	-2.54
VGA	DMOS	18.57	17.06	14.97	-
	PSNR	27.57	23.83	23.61	-
	SSIM	13.05	9.50	9.60	-
SET3					
Resolution	Quantization function	Triangle (low) (%)	Triangle (med) (%)	Triangle (high) (%)	-
VGA	DMOS	-20.45	-20.54	-20.11	-
	PSNR	-3.14	-1.89	1.58	-
	SSIM	-4.73	-4.28	-2.45	-
SET4					
Resolution	Quantization function	Triangle (%)	Cosine (%)	Quadratic (%)	Trapezoid (%)
VGA	DMOS	-33.17	-21.64	-27.77	-33.95
	PSNR	0.89	2.35	2.69	3.17
	SSIM	-4.66	-3.88	-3.56	-3.65

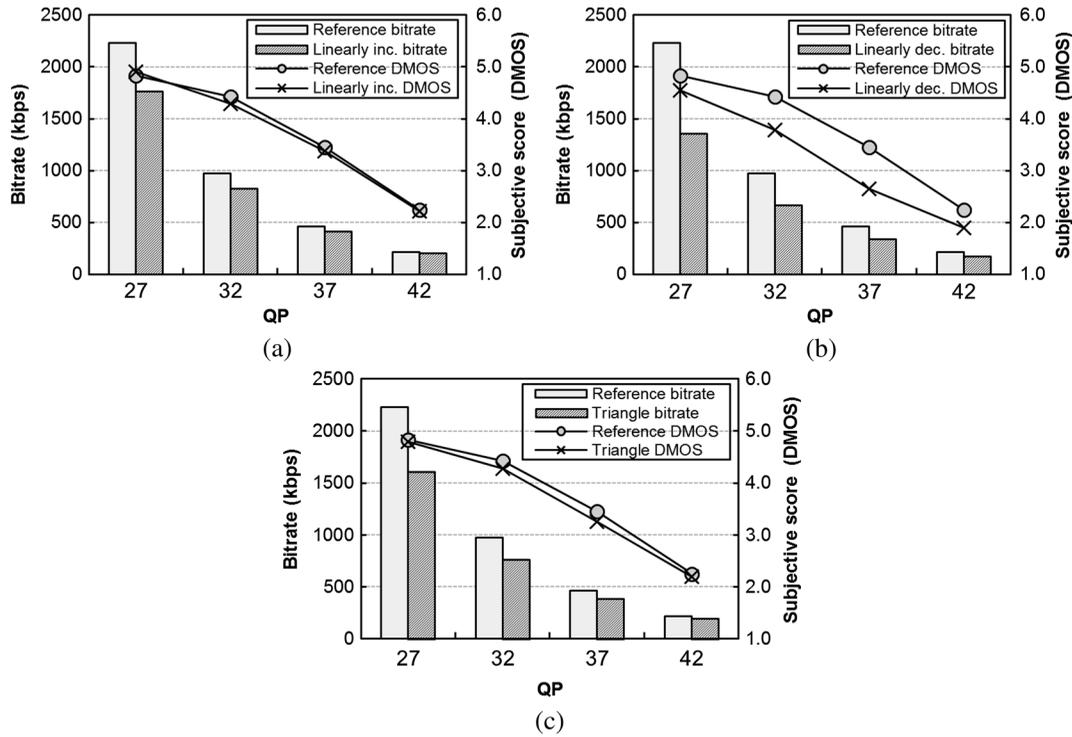


Fig. 15 Bitrate and subjective score comparison of the frequency quantization set 1 (VGA): (a) linearly increasing, (b) linearly decreasing, and (c) triangle.

quality of the reference method, were compared with those of the reference method. For instance, if a proposed quantization model had a -50% bitrate reduction ratio, only 50% of the bitrates could produce subjective quality equivalent to that of the reference method. Table 4 shows the results of bitrate reduction. Although the linearly increasing model showed the worst bitrate reduction, it showed the best efficiency improvement among the four models in terms of perceptual quality.

Figure 15 shows the experimental results of the bitrates and the subjective quality rating of the frequency quantization set 1 with VGA clips. The linearly increasing model [Fig. 15(a)] and triangle model [Fig. 15(c)] showed minor subjective quality degradations while the linearly

decreasing model [Fig. 15(b)] showed large subjective quality degradations. Also, the linearly decreasing function also produced larger bitrate reductions than it did in the HD test [Fig. 13(d)]. Obviously, applying a large quantization step to low frequency components resulted in a large bitrate reduction and a large perceptual quality degradation. Table 4 shows the bitrate reduction ratios, which produced the perceptual quality equivalent to the reference method. The linearly increasing model showed the best bitrate reduction while the linearly decreasing model showed poor performance, requiring more bits to produce the equivalent perceptual quality. Figure 16 shows a performance comparison in terms of the subjective scores, PSNR, and SSIM. The

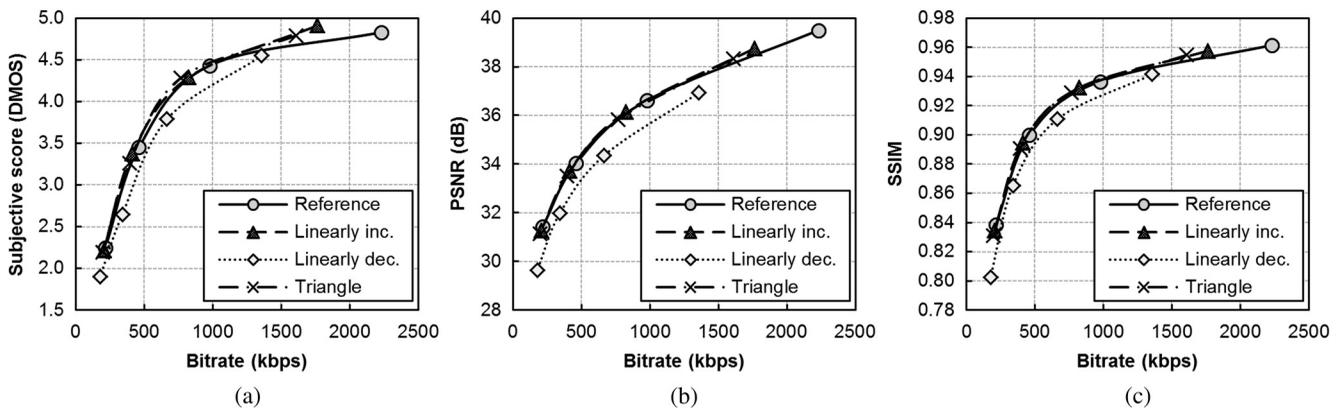


Fig. 16 Performance comparison for the frequency quantization set 1 (VGA): (a) subjective scores, (b) PSNR, and (c) SSIM.

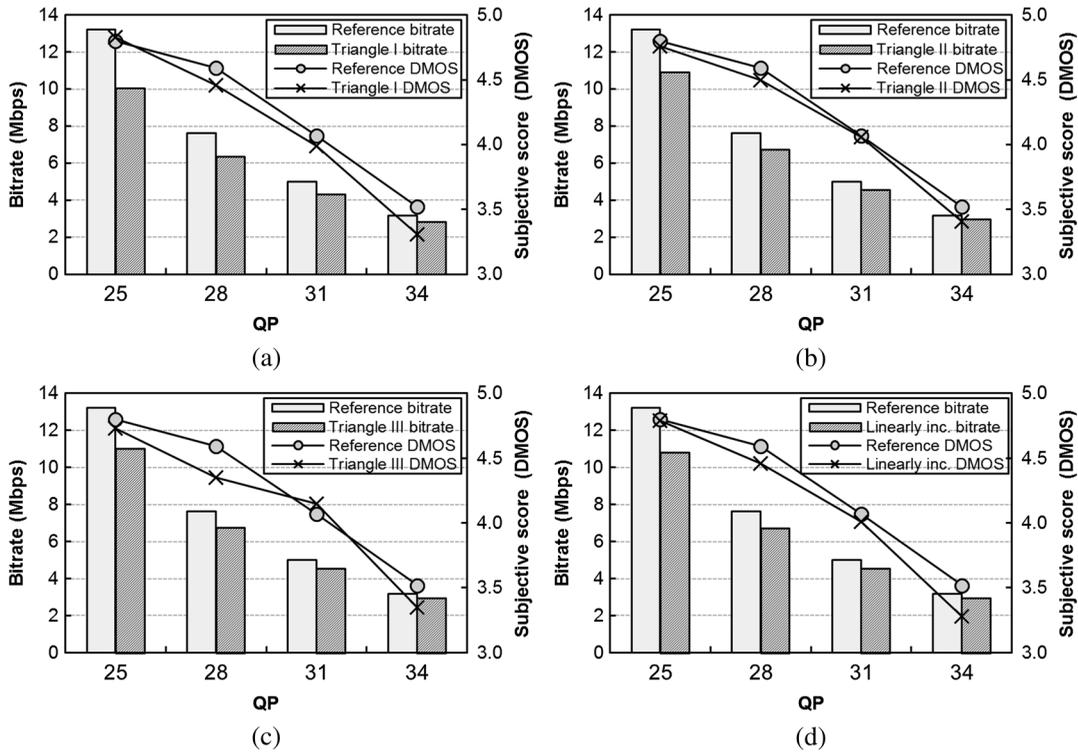


Fig. 17 Bitrate and subjective score comparison for the frequency quantization set 2 (HD): (a) triangle mode 1, (b) triangle mode 2, (c) triangle mode 3, and (d) linearly increasing.

linearly increasing and triangle models showed slightly improved performance in terms of DMOS and SSIM when compared to the reference model while the linearly decreasing model showed inferior performance.

Figure 17 shows the experimental results of the bitrates and the subjective quality rating of the frequency quantization set 2 (Fig. 7) with HD clips. In this experiment, the triangle mode 1 function produced the lowest bitrates among the four functions, as shown in Fig. 17(a), while the triangle mode 3 function produced the highest bitrates, as shown in Fig. 17(c). The triangle mode 3 function showed the worst bitrate reduction and produced

inconsistent subjective scores. Table 4 showed the bitrate reduction ratios that produced the same subjective quality as that of the reference model for the quantization set 2. The triangle mode 3 function showed the best bitrate reduction among the four models while the linearly increasing function showed poor performance, requiring more bits to produce equivalent perceptual quality. Figure 18 shows a performance comparison in terms of the subjective scores, PSNR, and SSIM.

Figure 19 shows the experimental results of the bitrates and the subjective quality of the frequency quantization set 2 with VGA source sequences. In this experiment, all

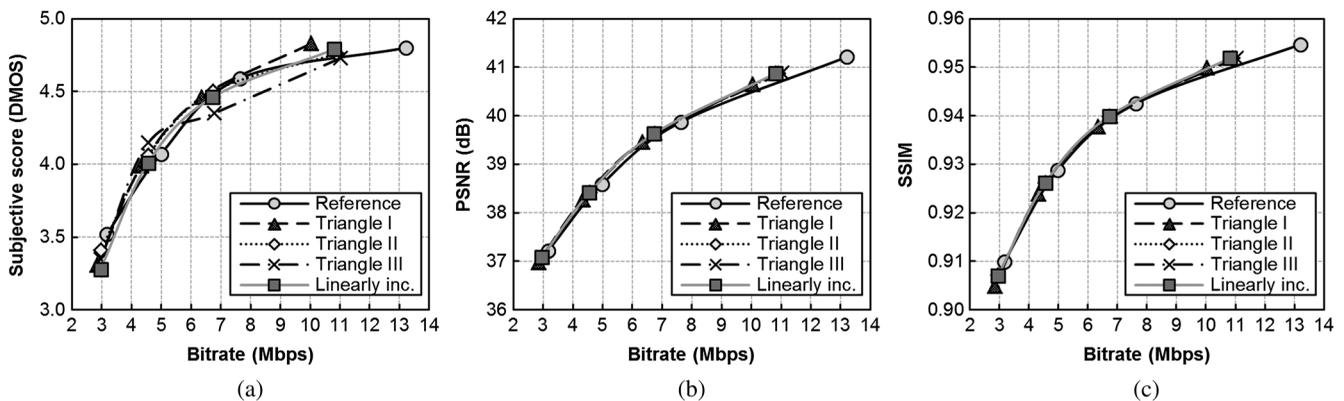


Fig. 18 Performance comparison for the frequency quantization set 2 (HD): (a) subjective scores, (b) PSNR, and (c) SSIM.

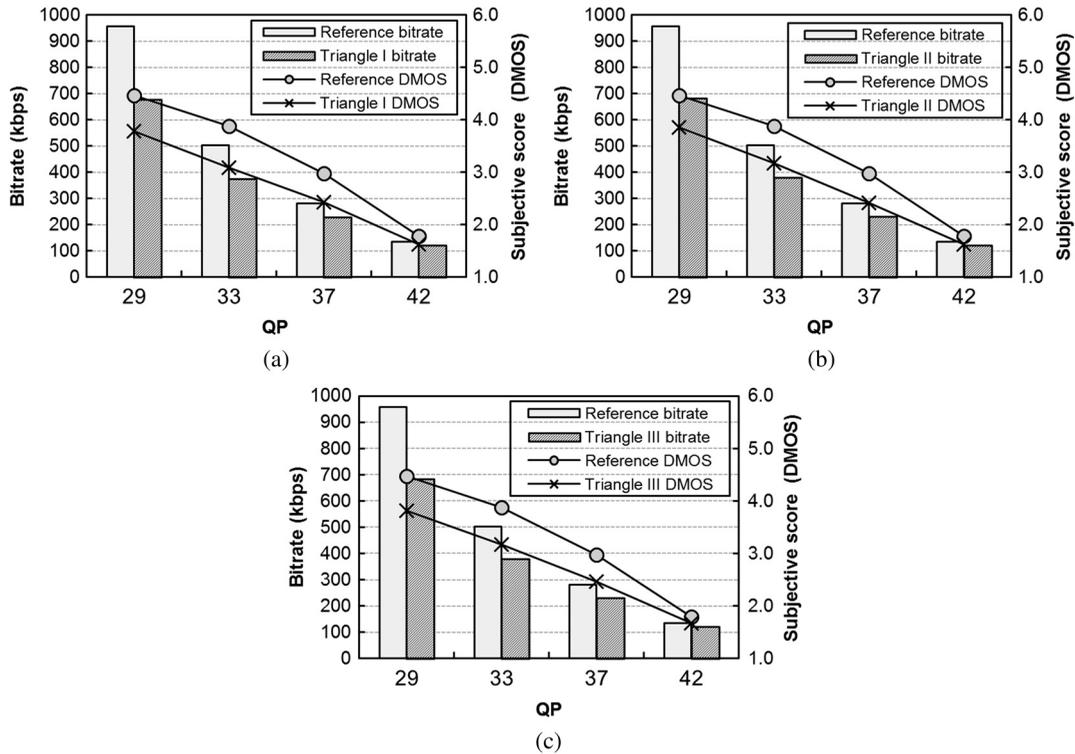


Fig. 19 Bitrate and subjective score comparison for the frequency quantization set 2 (VGA): (a) triangle mode 1, (b) triangle mode 2, and (c) triangle mode 3.

the quantization models showed large subjective score degradations as shown in Figs. 19(a)–19(c). Although the three quantization functions achieved large bitrate reductions, it appears that the subjective score degradations were larger. Consequently, the overall coding efficiency considering the bitrate reduction and the subjective scores appeared to decrease as shown in Fig. 20 and Table 4. The triangle functions showed inconsistent performance and their usefulness was rather limited.

Figure 21 shows the experimental results of the bitrates and the subjective quality of the frequency quantization set 3 with VGA source sequences. In the frequency

quantization set 3, three different triangle functions with different quantization intensities were used. The triangle-low function [Fig. 21(a)] had the smallest peak value while the triangle-high function [Fig. 21(c)] had the largest peak value. Generally, subjective scores and bitrate reductions are proportional to the peak values. Table 4 shows coding efficiency comparisons. Figure 22 shows a performance comparison in terms of the subjective scores, PSNR, and SSIM. The triangle (mid) function showed the best DMOS performance while the triangle (high) function showed the smallest DMOS improvement. In terms of PSNR and SSIM, the triangle (mid and low)

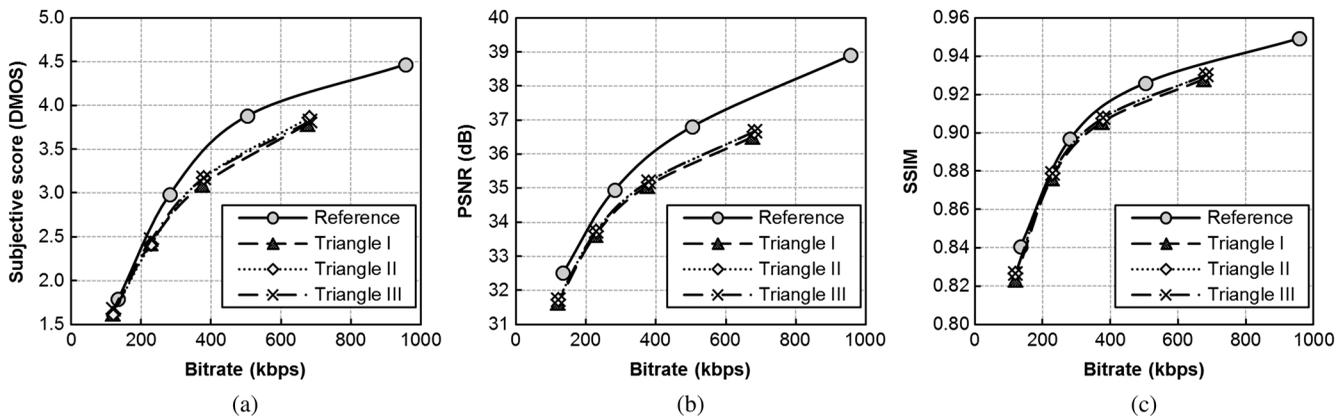


Fig. 20 Performance comparison for the frequency quantization set 2 (VGA): (a) subjective score, (b) PSNR, and (c) SSIM.

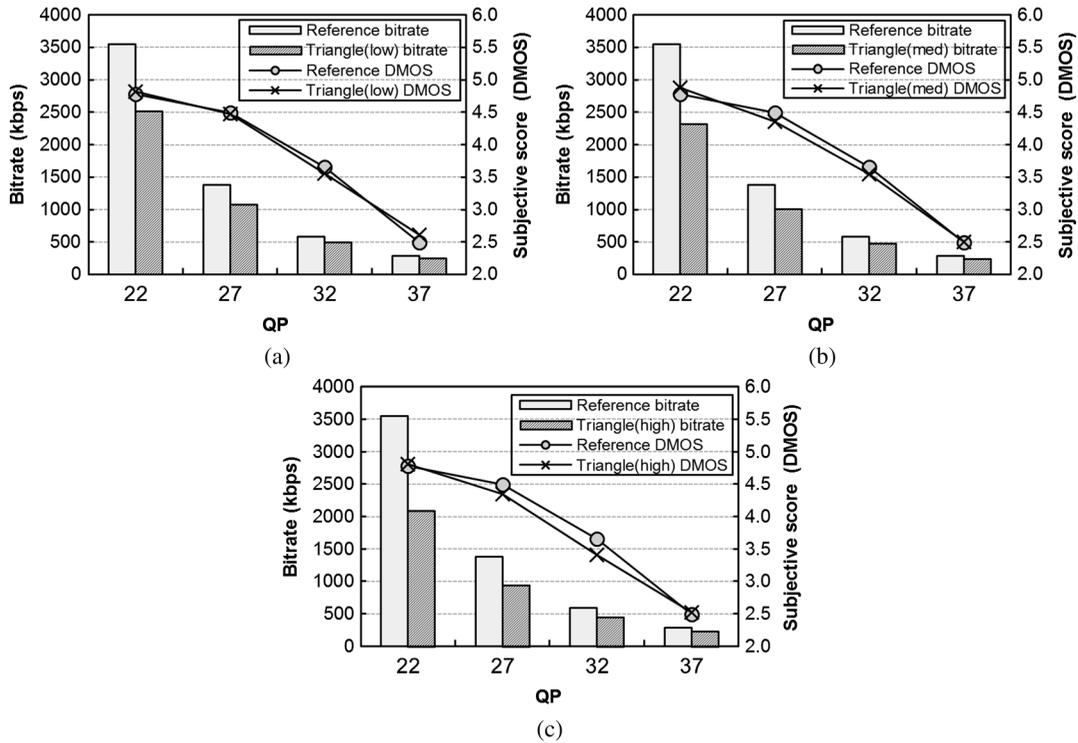


Fig. 21 Bitrate and subjective score comparison for the frequency quantization set 3 (VGA): (a) triangle (low), (b) triangle (mid), and (c) triangle (high).

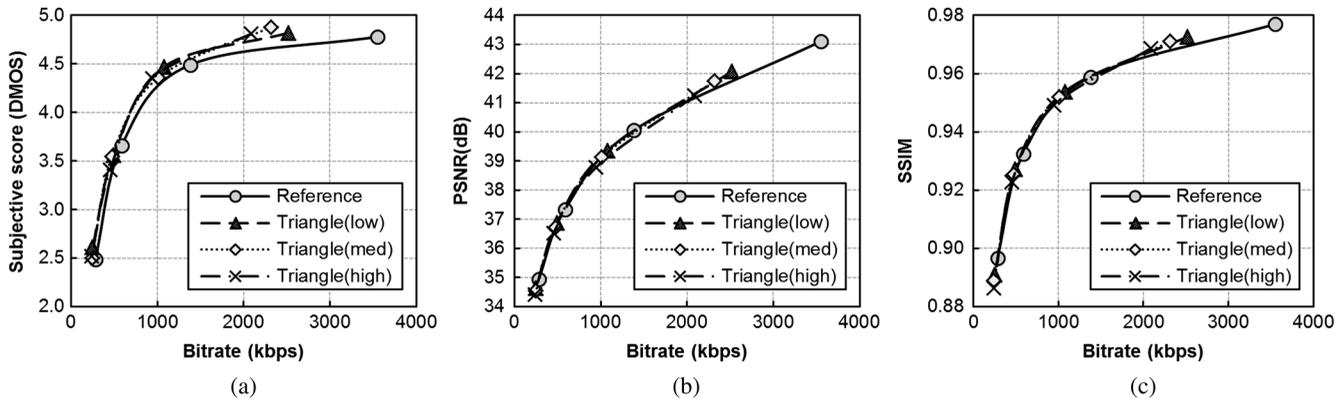


Fig. 22 Performance comparison for the frequency quantization set 3 (VGA): (a) subjective score, (b) PSNR, and (c) SSIM.

functions showed slightly improved performance for high bitrates.

Figure 23 shows the experimental results of the bitrates and the subjective quality of the frequency quantization set 4 with VGA source sequences. In the frequency quantization set 4, the triangle, cosine, quadratic, and trapezoid functions are used. These functions have similar strategies with different shapes, large quantization steps for middle frequencies, and small quantization steps for low and high frequencies. The bitrate reductions were proportional to the areas of quantization functions. Figure 24 shows

performance comparison in terms of DMOS, PSNR, and SSIM (VGA, set 4). Table 4 shows coding efficiency comparisons among the four different functions. In these subjective experiments, the trapezoid function showed the best coding efficiency while the cosine function showed the worst coding efficiency for subjective quality.

Table 5 shows the processing time comparison of the reference model (H.264) and the proposed models (3.40 GHz Intel i7-3770 CPU, 8 GB Memory). Since the proposed methods used frequency shaping functions, they did not increase the processing time. Also, since

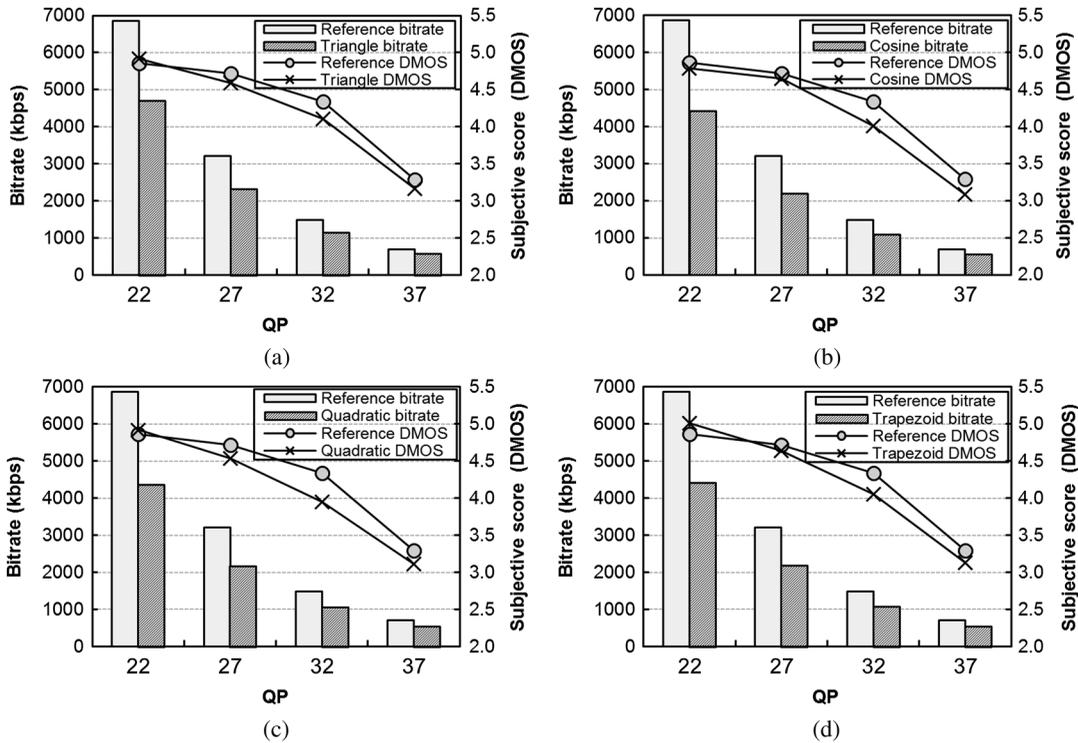


Fig. 23 Bitrate and subjective score comparison for the frequency quantization set 4 (VGA): (a) triangle, (b) cosine, (c) quadratic, and (d) trapezoid.

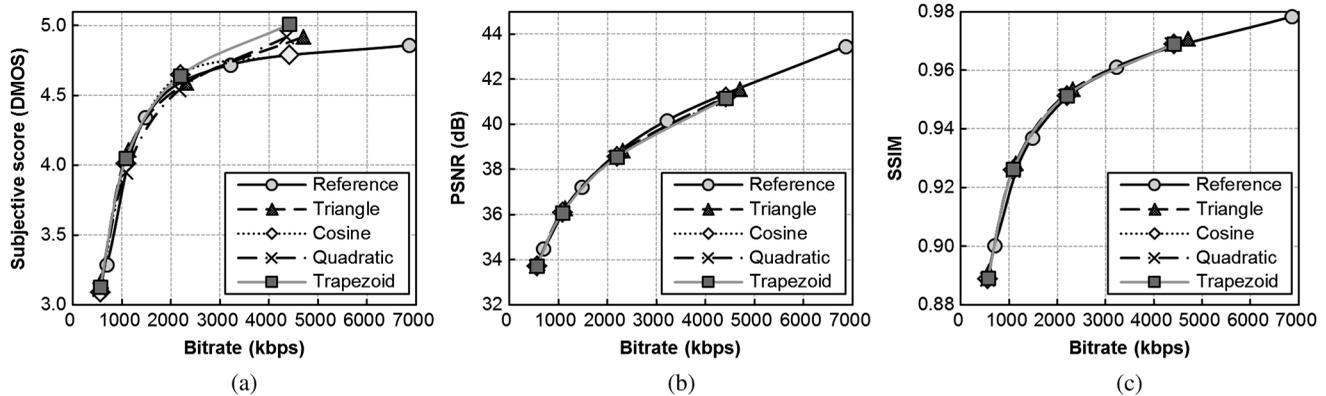


Fig. 24 Performance comparisons for the frequency quantization set 4 (VGA): (a) subjective score, (b) PSNR, and (c) SSIM.

Table 5 Processing time comparisons.

QP	VGA		HD	
	Reference model (s/frames)	Quantization model (s/frames)	Reference model (s/frames)	Quantization model (s/frames)
22	3.37	3.36	66.06	63.20
27	3.31	3.25	63.47	62.92
32	3.42	3.22	63.72	61.19
37	3.20	3.17	64.32	60.75
Ave.	3.32	3.25	64.39	62.02

the quantization matrix is already included in the H.264/AVC standard, the proposed method can be easily implemented.

6 Conclusions

In this paper, we have investigated the frequency sensitivity of the human visual system as applied to video compression standards, especially the H.264/AVC standard. Most conventional standards for video compression use the DCT method. On the other hand, those standards do not always consider the frequency sensitivity of the human visual system. In our experiments, subjective quality assessments for video quality were performed to provide a better understanding of human frequency sensitivity characteristics. In the future, these frequency characteristics may be used to improve video coding

efficiency while maintaining equivalent perceptual video quality.

Acknowledgments

This work was supported in part by the Technology Innovation Program, 10035389, funded by the Ministry of Knowledge Economy (MKE, Korea).

References

1. Z. Wang and A. C. Bovik, "Mean square error: love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, **26**(1), 98–117 (2009).
2. M. Naccari and F. Pereira, "Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment," *IEEE Trans. Circuits Syst. Video Technol.* **21**(6), 766–782 (2011).
3. A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proc. SPIE* **1666**, 365–374 (1992).
4. A. B. Watson, "DCTune: a technique for visual optimization of DCT quantization matrices for individual images," in *Society for Information Display Digest of Technical Papers*, pp. 946–949, Wiley Press, San Diego (1993).
5. I. Höntsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE Trans. Image Process.* **11**(3), 213–222 (2002).
6. X. Zhang, W. S. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortions," *Signal Process.* **85**(4), 795–808 (2005).
7. Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile from grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.* **19**(3), 337–346 (2009).
8. R. Shapley and D. M. K. Lam, *Contrast Sensitivity*, Vol. 5, MIT Press, Cambridge (1993).
9. W. J. Lovegrove et al., "Specific reading disability: differences in contrast sensitivity as a function of spatial frequency," *Science* **210**, 439–440 (1980).
10. K. Arundale, "An investigation into the variation of human contrast sensitivity with age and ocular pathology," *Br. J. Ophthalmol.* **62**(4), 213–215 (1978).
11. T. Wiegand et al., "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 560–576 (2003).
12. ITU-R Recommendation BT.500-11, *Methodology for the Subjective Assessment for the Quality of Television Pictures*, International Telecommunication Union (2002).
13. ITU-T Recommendation P. 910, *Subjective Video Quality Assessment Methods for Multimedia Applications*, International Telecommunication Union (2008).
14. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).

Guiwon Seo received the BS degree in electrical electronic engineering from Yonsei University, Seoul, Republic of Korea, where he is currently working toward the PhD degree. His research interests include image/signal processing, video compression, and video quality measurement.

Jonghwa Lee received the BS and PhD degrees in electrical and electronic engineering from Yonsei University in 2005 and 2011, respectively. He is a senior engineer at Samsung Electronics Co. Ltd., Republic of Korea. His research interests include image/signal processing, pattern recognition, and video quality measurement.

Chulhee Lee received the BS and MS degrees in electronic engineering from Seoul National University in 1984 and 1986, respectively, and a PhD degree in electrical engineering from Purdue University, West Lafayette, Indiana, in 1992. In 1996, he joined the faculty of the Department of Electrical and Computer Engineering, Yonsei University, Seoul, Republic of Korea. His research interests include image/signal processing, pattern cognition, and neural networks.