
Article

Visual Quality Balancing between Enhanced High-resolution and Ordinary JPEG-decoded Images in Steganographic Image Compression System

Isao FURUKAWA* and Junji SUZUKI**

(Received Nov. 5, 2021)

Abstract

The steganographic image compression system provides a coding method that is compatible with standard JPEG, but allows only users who know that the information of higher frequency components is embedded to obtain enhanced high-resolution images at the same time. Until now, the main focus has been on improving the quality of high-resolution images by embedding as many codes as possible in lower frequency DCT coefficients. However, this conventional method leads to greater degradation in JPEG-decoded low-resolution images, and it is not appropriate to prioritize high-resolution images over low-resolution images because of the inherent duality of providable resolution images by the system. Therefore, in this study, we employ Watson's model to evaluate the distortion caused by embedding enhancement codes to lower frequency, and extend the perceptual model to higher frequency by extrapolating DCT frequency sensitivity values to calculate visual distortion of higher frequency DCT coefficients when they are discarded due to a constraint of payload. In addition, we propose a visual quality balancing (VQB) algorithm for low- and high-resolution images by successively comparing their perceptual distances during enhancement code embedding. Finally, we demonstrate the effectiveness of the proposed algorithm through simulation experiments on standard test images with 4K resolution.

Key Words: steganography, image enhancement, JPEG compression, Watson's model, JND**1. INTRODUCTION**

In image communications, a variety of resolution images are used depending on various uses of images, and the coding methods to achieve this goal are called multi-resolution image coding. Such encoding systems are also referred to scalable coding, and JPEG2000 [1] is an example of international standards for multi-resolution still images. On the other hand, there is a steganographic technique that embeds irrelevant information in the image and conveys it as a secret message. We

have investigated a steganographic image coding system in which the secret information is replaced by enhancement codes of high-resolution frequency components of the image and only the user who knows the secret can recover the high-resolution image from the embedded codes [2]–[4].

The conventional steganographic image coding system focuses on improving the quality of high-resolution images by embedding as many enhancement codes created from higher frequency quantized DCT coefficients as possible within the payload derived from lower frequency quantized

* Department of Computer Science, Faculty of Applied Information Science, Hiroshima Institute of Technology

** School of Information Science and Technology, Aichi Prefectural University (retired)

DCT coefficients. However, since the system can provide both enhanced high-resolution image and ordinary JPEG-decoded low-resolution image, it is also important to determine the number of enhancement codes so that the visual quality of both resolution images is balanced. For this purpose, it is necessary to develop a method to adaptively determine the maximum embeddable enhancement code length in order to balance both distortions in terms of a human visual system by comparing their visual impacts caused by embedding codes to lower frequency DCT coefficients and discarding higher frequency DCT coefficients. In this paper, we review such problems of the conventional method in the first place, and intend to solve them using Watson's visual model together with its extension to higher frequency DCT coefficients, and also examine the effectiveness of the proposed method on standard 4K images.

This paper is organized as follows. The next section reviews the basic configuration of the steganographic image coding system with enhancement code embedding, and shows the requirements to balance the visual quality of enhanced high-resolution images and JPEG-decoded low-resolution images. The procedure to evaluate the visual distortion using Watson's model is also described in the section. Section 3 shows an example of the distortion when the coding scheme is applied to a simple one-dimensional/four-point DCT/IDCT system. Moreover, actual 4K images are used as input to the conventional system, and several comments are made about these results. Section 4 discusses a relationship between the visual distortion of lower frequency DCT coefficients and the enhancement code length, and extends the frequency sensitivity to higher DCT coefficients in order to estimate the distortion caused by discarding these coefficients. Section 5 proposes a method to control the amount of enhancement codes by successively comparing the visual distortions from lower and higher frequency DCT coefficients, and verify the effectiveness of the method using actual 4K images. Section 6 gives conclusions and future work.

2. BASIC SYSTEM CONFIGURATION

2.1 CODING SCHEME

Figure 1 shows a basic configuration of the steganographic image coding system in which the enhancement codes obtained from higher frequency DCT coefficients are embedded in lower frequency DCT coefficients, so that both enhanced high-resolution and JPEG-decoded low-resolution images can be obtained in the same decoder. After applying 16×16 DCT to an input image, 16×16 except 8×8 quantized DCT coefficients are run-length Huffman encoded to create enhancement codes, and these codes are embedded in the LSBs of lower frequency quantized DCT coefficients. Then, the payload, which corresponds to the upper limit of the embeddable code length, is imposed as a constraint in order not to excessively degrade the quality of the recovered image. The subsequent encoding procedure after the enhancement code embedding is exactly the same as the standard JPEG algorithm. On the other hand, at the decoder, ordinary users can apply the JPEG decoding algorithm to obtain a low-resolution image with half the vertical and horizontal resolutions of the input image. In addition, special users who know that the secret information is embedded in the LSBs of lower frequency quantized DCT coefficients can extract the enhancement codes and perform run-length Huffman decoding on them in order to recover the high-resolution image with the same resolution as the input image.

The conventional coding scheme [3] embeds as many enhancement codes as possible within the payload of lower frequency DCT coefficients so as to make the quality of enhanced image high enough. Thus, in this case, the quality of enhanced high-resolution image could be higher than that of JPEG-decoded low-resolution image. Now that the quality of high-resolution image is inversely related to that of low-resolution image under the constraint of the embeddable enhancement code length, there is a tradeoff between aforementioned two resolution images. The most straightforward way to reduce the quality degradation of JPEG-decoded low-resolution image is to keep the payload relatively low. However, uniformly reducing the bit depth in all DCT blocks

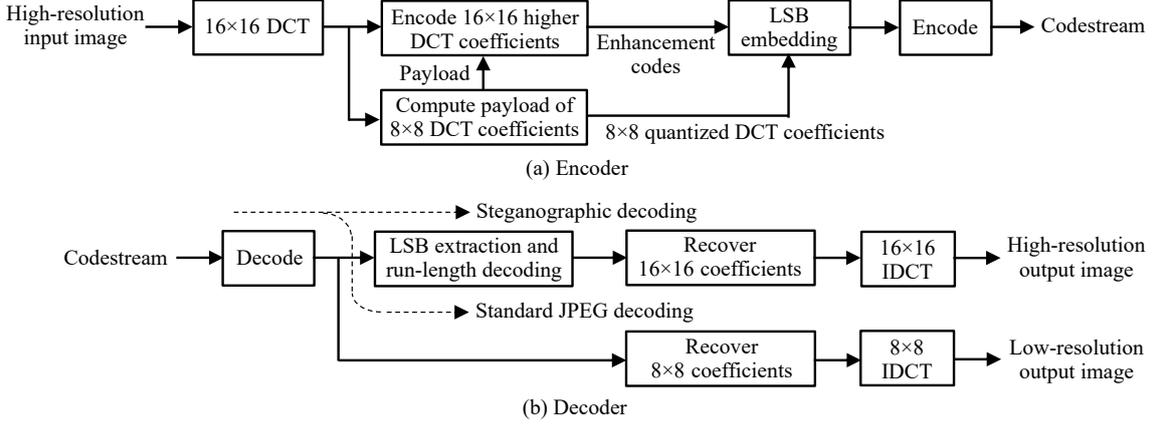


Fig. 1 Steganographic encoder/decoder configuration with enhancement code embedding and extraction.

also has an effect of reducing the number of non-zero higher frequency components, so that it is necessary to develop an enhancement code decision technique that balances the visual distortions in both frequency domains. In addition, since the distortion occurring in lower frequency is mainly blockiness (i.e., square-shaped distortion) while the distortion occurring in higher frequency is mainly blurring (i.e., indistinct contour), it is difficult to balance the impacts of the distortions by simply reducing the bit depth uniformly when the two kinds of distortions are mixed in a single image.

When capturing the distortions that occur in higher and lower frequencies, it is important to take their visual effects into account. Figure 2 shows a conceptual relationship between the enhancement code length CE obtained from higher frequency quantized DCT coefficients and the visual distortions D of both frequencies by the enhancement code embedding. As CE increases, D in lower frequency increases, while D in higher

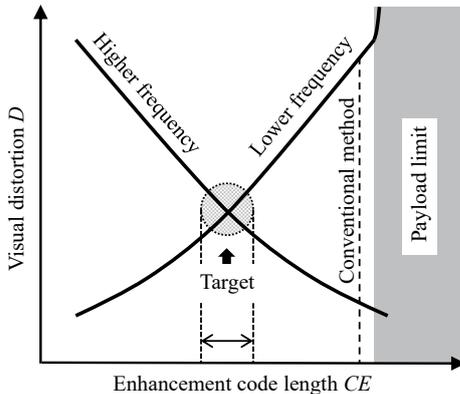


Fig. 2 Enhancement code length vs. visual distortion.

frequency decreases, so that the tradeoff relationship exists between them. The visual distortions with the conventional scheme are indicated by the dashed line just left to the payload limit. However, this results in an imbalance between the quality of these two resolution images, and the decoded images are apparently inappropriate for the purpose of the coding system. Therefore, a method needs to be developed to restrict the enhancement code length to the target region located in the center of the figure, where the visual quality of both images is balanced.

2. 2 EVALUATION OF VISUAL DISTORTION

In this study, we use Watson's model [5] for evaluations of visual distortion. The procedure for incorporating this model into the steganographic coding for calculating the visual distortion is as follows. Let $g_e(i, j)$ and $S_e(i, j)$ be the quantized DCT coefficients after embedding the enhancement codes and their inverse quantized DCT coefficients, respectively. In addition, let $S_o(i, j)$ be a reference DCT coefficients for the distortion evaluation. Then, Watson's model is used to evaluate the perceptual distance with the next three-step calculations.

$$t_L(i, j) = t_b(i, j) \left(\frac{S_o(0, 0)}{S_{o,0}} \right)^{0.649}, \quad (1)$$

$$t(i, j) = \max \left\{ t_L(i, j), |S_o(i, j)|^{0.7} t_L(i, j)^{0.3} \right\}, \quad (2)$$

$$D_{\text{wat}}(S_o, S_e) = \left(\sum_{j=0}^7 \sum_{i=0}^7 |d(i, j)|^4 \right)^{1/4}, \quad (3)$$

where $d(i, j)$ is the error of DCT coefficients divided by the contrast sensitivity threshold $t(i, j)$:

$$d(i, j) = \frac{S_e(i, j) - S_o(i, j)}{t(i, j)}. \quad (4)$$

As a result, the individual errors of Equation (4) are pooled into a single perceptual distance $D_{\text{wat}}(S_o, S_e)$ of Equation (3).

We consider two DCT coefficients $S_o(i, j)$ as the reference for calculating the perceptual distance: DCT coefficients $S_b(i, j)$ before quantization, or DCT coefficients $S_q(i, j)$ after quantization but before bit-embedding. The parameters are the quality factors $QF1$ and $QF2$ used to quantize lower and higher frequency DCT coefficients, respectively, and the bit depth n_{LSB} up to which the enhancement code bits can be embedded. The lower frequency quantized DCT coefficients are obtained by the quality factor $QF1$, and the payload PL is determined by the bit depth n_{LSB} , i.e., PL is the sum of the number of embeddable bits for non-zero higher frequency quantized DCT coefficients. Moreover, the enhancement code length CE is determined by the quality factor $QF2$ as well as the inequality $CE \leq PL$: the enhancement code length must be equal to or less than the payload. Substituting each value in the above three-step calculations, we can obtain the visual distortion of lower frequency DCT coefficients in one DCT block caused by the enhancement code embedding based on Watson's model, i.e., the pooled perceptual distance $D_{\text{wat}}(S_o, S_e)$, where S_o can be either S_b or S_q according to which reference DCT coefficients are used.

The procedure for evaluating the perceptual distance described above is shown in Figure 3, where 8×8 DCT coefficients picked from 16×16

DCT coefficients are lower frequency coefficients, and DCT coefficients excluding them are higher frequency coefficients. The perceptual distance is calculated for lower frequency coefficients, because the DCT frequency sensitivity table is given only to 8×8 DCT components. The higher frequency coefficients are used to create the enhancement codes of length CE with the constraint of the payload PL determined from the lower frequency coefficients.

3. EXAMPLES OF DISTORTION

As explained above, there is a tradeoff between the distortions of lower and higher frequency DCT coefficients with respect to the enhancement code length. The optimization of this coding system depends on the image content and visual characteristics, and the natural question is how to allocate the distortion to the lower and higher frequency DCT coefficients respectively. As a preliminary study, we will first consider a case of simple one-dimensional/four-point DCT/IDCT system, and then demonstrate the above tradeoff when the conventional coding scheme is applied to actual 4K images.

3.1 EXAMPLE FOR 1-D/4-POINT DCT/IDCT SYSTEM

Consider 1-D/4-point DCT/IDCT system, where DCT coefficients are denoted by $S(0), \dots, S(3)$, and the reconstructed signals after IDCT are denoted by $s(0), \dots, s(3)$. If we apply this system to the steganographic image coding system, $S(0)$

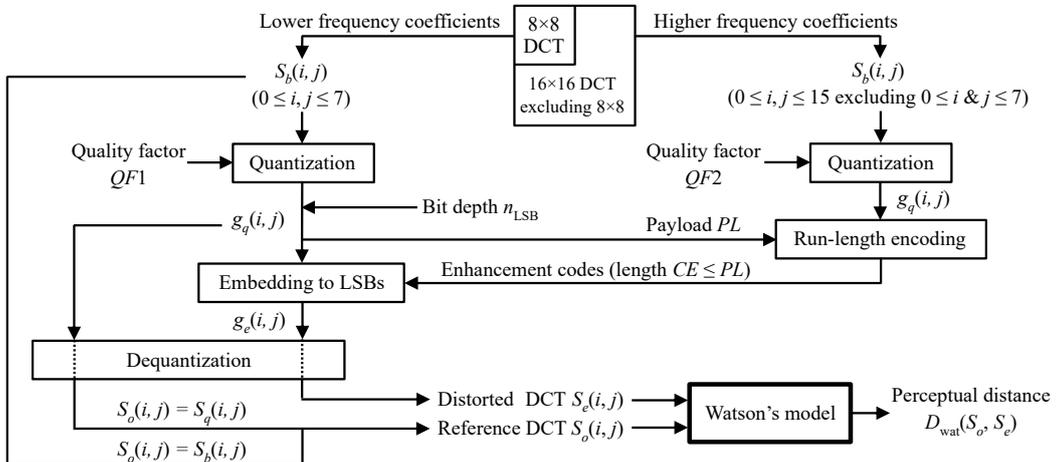


Fig. 3 Procedure for evaluating perceptual distance of 8×8 DCT coefficients with Watson's model.

corresponds to DC coefficient, $S(1)$ corresponds to a unique lower frequency DCT coefficients to be embedded, and $S(2)$ and $S(3)$ correspond to higher frequency DCT coefficients to be run-length Huffman encoded. According to the rules of this coding system, no bit-embedding is applied to $S(0)$, and it retains the original value. The bit-embedding is applied only to $S(1)$, resulting in distortion of the lower frequency components. Since $S(2)$ and $S(3)$ are higher frequency components, they are either present or absent; when $S(3)$ in addition to $S(2)$ is unlikely to be present, the distortion caused by the bit-embedding to lower frequency components is small. The tradeoff is that when $S(2)$ and $S(3)$ are both likely to exist (blurring will be small), the distortion of lower frequency components is large (blockiness will be large).

We will evaluate this relationship using a simple model for the frequency components $S(i)$. Assume that the DCT coefficients decrease exponentially with frequency. This is a reasonable assumption to make when dealing with actual images. For example,

$$S(i) = e^{-i}, \quad (5)$$

where e is the base of natural logarithm and $i = 0, \dots, 3$. Furthermore, the distortion caused by the bit-embedding in lower frequency DCT coefficients $S(1)$ is denoted by a . Let m be an appropriate positive number, and assume that the existence of $S(2)$ and $S(3)$ is determined according to the following rules.

- If $0 \leq a \leq 1/2^m$ (the lower frequency distortion is relatively small), then $S(2)$ exists and $S(3)$ is missing.
- If $1/2^m < a$ (the lower frequency distortion is relatively large), then both $S(2)$ and $S(3)$ exist.

Applying this model to the 1-D/4-point DCT/IDCT system, the reconstructed signals $s(0), \dots, s(3)$ can be expressed as

$$\begin{pmatrix} s(0) \\ s(1) \\ s(2) \\ s(3) \end{pmatrix} = \frac{1}{4} \begin{pmatrix} \sqrt{2} & \sqrt{2+\sqrt{2}} & \sqrt{2} & \sqrt{2-\sqrt{2}} \\ \sqrt{2} & \sqrt{2-\sqrt{2}} & -\sqrt{2} & -\sqrt{2+\sqrt{2}} \\ \sqrt{2} & -\sqrt{2-\sqrt{2}} & -\sqrt{2} & \sqrt{2+\sqrt{2}} \\ \sqrt{2} & -\sqrt{2+\sqrt{2}} & \sqrt{2} & -\sqrt{2-\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 1/e+a \\ 1/e^2 \\ 1/e^3 \cdot h_3(m) \end{pmatrix}. \quad (6)$$

In other words, the lower frequency distortion a is added to $S(1)$, and the existence of $S(3)$ is determined by the following function $h_3(m)$.

$$h_3(m) = \begin{cases} 0, & \text{if } 0 \leq a \leq 1/2^m \\ 1, & \text{if } 1/2^m < a \end{cases} \quad (7)$$

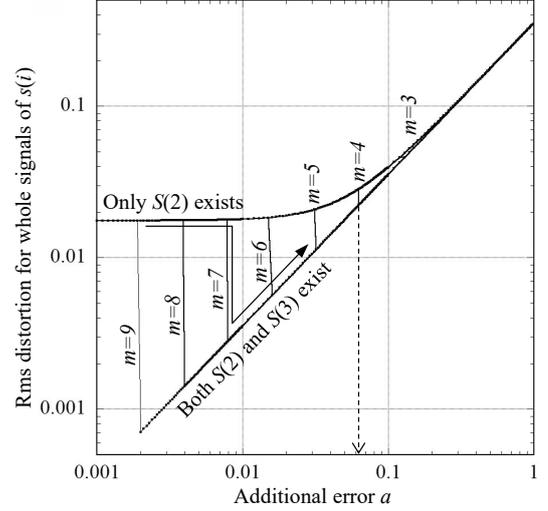


Fig. 4 Distortion of recovered signal in 1-D/4-point DCT/IDCT system.

Let $s_e(x)$ be the reconstructed signal after IDCT assuming that no enhancement codes are embedded in lower frequency DCT coefficients, and also that both the higher frequency DCT coefficients $S(2)$ and $S(3)$ exist (i.e., no encoding process is applied), and let $s_c(x)$ be the reconstructed signal obtained by IDCT after the bit-embedding. Since Watson's model for 2-D DCT coefficients is not available here, the error between them is evaluated by the root mean square (rms) for whole signals of $s(i)$.

The threshold for lower frequency distortion a , which determines the existence of the higher frequency coefficient $S(3)$, is varied for integer values of $m = 1, 2, \dots, 9$. Figure 4 shows the evaluation results of the rms of distortions by varying m . The rms error has discontinuities at $a = 1/2^m$ because only the higher frequency coefficient $S(2)$ exists for smaller a , while both $S(2)$ and $S(3)$ exist for larger a . For example, the transition of the rms error with increasing the additional error a for $m = 7$ is shown in the figure as a solid line with an arrow. For $m = 4$, $a = 1/2^4 = 0.0625$, which corresponds to the case where about 17% of the bit-embedding distortion is included for $S(1) = 1/e \approx 0.368$. For $m \leq 4$, assuming that the higher frequency coefficient $S(3)$ exists at larger a , the errors are asymptotic for the case where only $S(2)$ exists and the case where both $S(2)$ and $S(3)$ exist. Therefore, it is obvious that the error increases at a constant rate with a regardless of whether $S(3)$ is present or not, when the bit-embedding causes an error of 20% or more relative to the magnitude of $S(1)$.

3. 2 EXAMPLE FOR 4K IMAGES

Next, we show simulation results when the conventional steganographic coding scheme is applied to actual 4K images. In this case, the bit depth n_{LSB} changes from 0 up to 6 depending on the magnitude of the DCT coefficients. Two 4K images of u01_Books and u04_Kimono [7] are used as input. The simulation compares and evaluates the reconstructed images in case that the quality factor $QF2 = 95$ or 70 and the bit depth $n_{\text{LSB}} = 6$, and also in case that $QF2 = 70$ and $n_{\text{LSB}} = 2$ or 1.

Table 1 summarizes the parameter combinations of $QF2$ and n_{LSB} corresponding to the four conditions (a)-(d) for the simulation. The quality factor for lower frequency DCT coefficient quantization is always fixed to $QF1 = 95$. Condition (a) corresponds to the situation that yields maximum SNR in the simulation experiments [4]. Condition (b) is for creating more non-zero quantized AC coefficients with emphasis on higher frequency components when the same bit depth as condition (a) is used. Conditions (c) and (d) correspond to the case where the bit depth is reduced to 2 and 1, respectively, when the quality factors, $QF1$ and $QF2$, are the same as in condition (a).

Table 1 Simulation parameters for 4K images.

Conditions		$QF1$	$QF2$	n_{LSB}
(a)	Maximum SNR	95	70	6
(b)	More non-zero AC coefficients		95	6
(c)	Bit depth is reduced to 2 in (a)		70	2
(d)	Bit depth is reduced to 1 in (a)		70	1

Table 2 shows the statistics of the amount of embedded enhancement codes and the SNR of the enhanced high-resolution images based on conditions (a)-(d) for the two 4K images. In this table, RH code stands for run-length Huffman code, i.e., the enhancement code. The rate of RH means the ratio of the amount of RH codes to that of the JPEG codestream. For the various statistical values of enhancement codes per block, the number of blocks containing no enhancement codes are excluded from the calculation. The quality of high-resolution images is evaluated by SNR in the table, and the perceptual distance by Watson's model for low-resolution images will be evaluated in 4.1.

Table 2 Simulation results for 4K images

(a) u01_Books

Conditions		(a)	(b)	(c)	(d)
RH codes (bits)	Mean	45.83	71.48	30.37	19.17
	Max	218	216	116	62
	Min	4	4	4	4
Total RH (bytes)		185,600	289,505	122,992	77,619
Rate of RH (%)		9.47	14.62	6.27	3.95
SNR (dB)		29.24	28.07	28.82	28.29

(b) u04_Kimono

Conditions		(a)	(b)	(c)	(d)
RH codes (bits)	Mean	27.49	44.88	18.57	12.52
	Max	226	225	116	63
	Min	4	4	4	4
Total RH (bytes)		111,335	181,768	75,200	50,688
Rate of RH (%)		7.29	11.83	4.92	3.32
SNR (dB)		31.05	30.06	30.71	30.04

Figure 5 shows the high-resolution images obtained by using the parameter sets of conditions (a)-(d) for the 4K image u01_Books. The upper left coordinate of the cropped partial images is (1500, 1000), and the number of horizontal and vertical pixels is (240, 180). Some comments on the result are as follows.

- Image (a) shows the highest quality based on SNR among various parameter settings in the conventional method that uses the same parameters for all DCT blocks.
- Image (b) has more non-zero higher frequency quantized DCT coefficients than image (a). In this case, the payload of lower frequency quantized DCT coefficients is the same as in condition (a), so the allowable enhancement code length reaches the payload at an early stage during higher frequency coefficient scanning, and the other non-zero coefficients after the truncation point are discarded. As a result, only relatively lower components in the higher frequency domain are preserved, and the image appearance of (b) in terms of resolution is inferior to that of (a).
- In images (c) and (d), the payload is smaller than images (a) and (b) due to the smaller bit depth constraint. Therefore, the allowable enhancement code length is also reduced, and the image looks like it has undergone low-pass filtering, but instead, the blockiness is alleviated in comparison with the image (a). The effect of this distortion tradeoff is, of course, more emphasized in the image (d) than in the image (c).

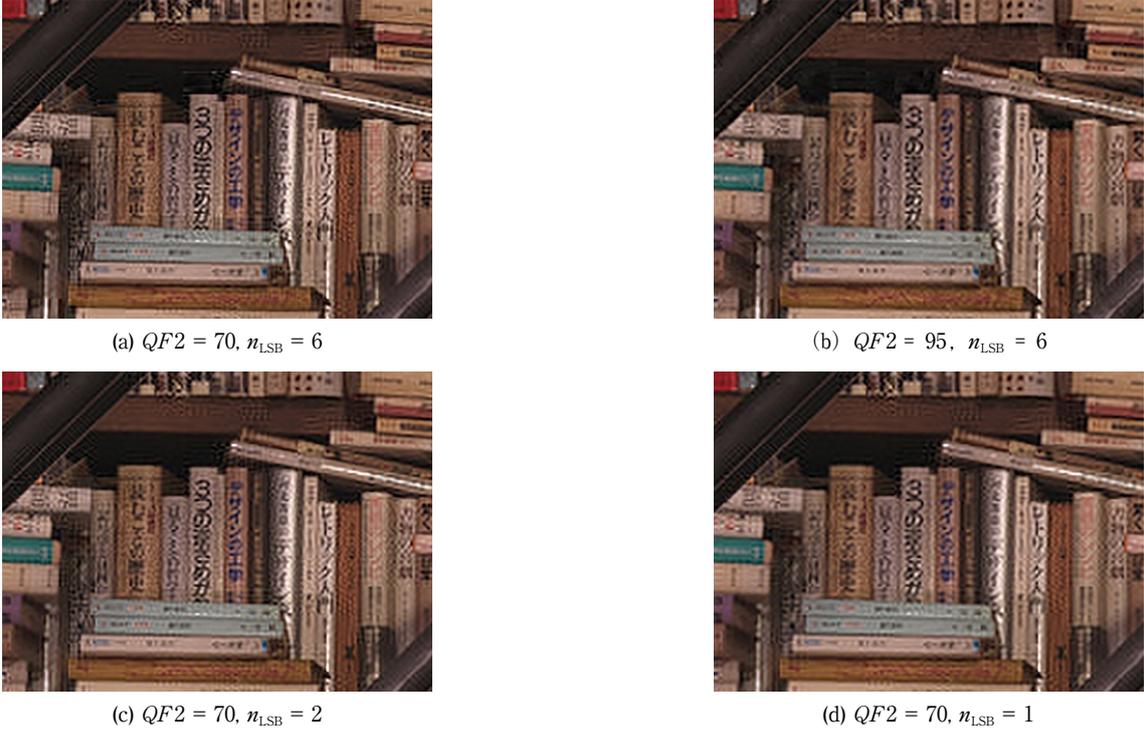


Fig. 5 Enhanced high-resolution images with varying $QF2$ and n_{LSB} for u01_Books ($QF1 = 95$).

• Even though the number of enhancement codes embedded in the image (b) is larger than in the images (a) and (c), it does not contribute to SNR improvement. In other words, excessively raising the quantization quality factor for higher frequency DCT coefficients can have negative effects on the visual quality of the reconstructed image.

4. EXTENDING FREQUENCY SENSITIVITY

4.1 EMBEDDED CODES VS. PERCEPTUAL DISTANCES

In this section, the conceptual relationship between the enhancement code length and the visual distortion of the lower frequency DCT coefficients shown in Fig. 2 is examined for actual 4K images, u01_Books and u04_Kimono. The bit depth is set to $n_{\text{LSB}} = 6$ and the combination of quantization quality factors is $(QF1, QF2) = (95, 70)$ and $(70, 70)$. We also consider two reference signals for calculating the perceptual distance as explained in 2.2: the DCT coefficients after quantization and before bit-embedding, as well as the DCT coefficients before quantization. For DCT blocks where higher frequency coefficients are not all zeros and the payload is greater than 4,

the enhancement codes corresponding to the DCT coefficients are embedded. On the other hand, for blocks in relatively flat regions of the image where the higher frequency DCT coefficients are all zeros, only a 4-bit EOB code is embedded if the payload is 4 or more, and no codes are embedded if the payload is less than 4. This embedding of a single EOB code cannot be avoided due to the termination of run-length Huffman codes, and it should be considered separately from other cases of the bit-embedding.

Figure 6 shows, for DCT blocks where code embedding has been performed, the histogram of the occurrence frequency of the perceptual distance values for DCT blocks that have only EOB as an enhancement code. Figure 7 shows, as examples for the image u01_Books and the quality factor $(QF1, QF2) = (95, 70)$, the 3-D histogram for the occurrence frequency of the combination of enhancement code length and the perceptual distance for blocks whose enhancement codes are not EOB code only. The contour plot shows the same occurrence frequency of in the form of bird's-eye view. When the reference signal for evaluating the perceptual distance is $S_o = S_b$ both the quantization distortion and the bit-embedding distortion exist, resulting in a distribution with a larger perceptual

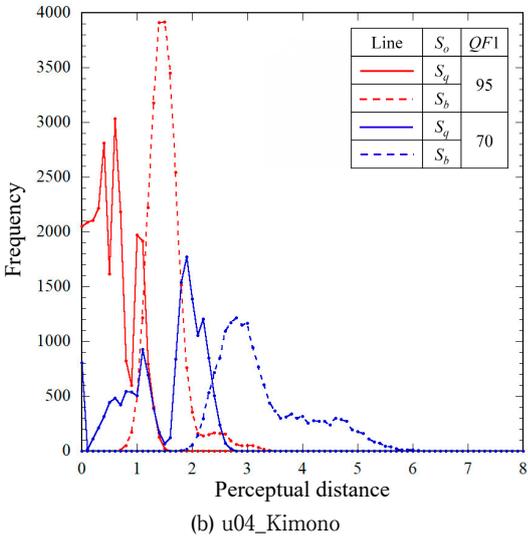
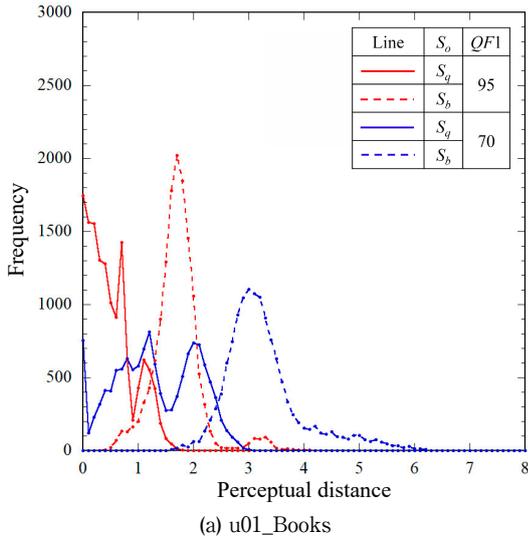


Fig. 6 Histogram of perceptual distance for DCT blocks with only EOB code.

distance than when the reference signal is $S_o = S_q$. This shows that the perceptual distance is almost directly promotional to the enhancement code length. In addition, although the payload is smaller in the case of $QF1 = 70$ than in $QF1 = 95$, which is omitted here, the distribution moves toward the larger value of perceptual distance even though the enhancement code length becomes smaller accordingly.

4.2 APPROXIMATION OF FREQUENCY SENSITIVITY

The frequency sensitivity table [5, 6] for 8×8 DCT coefficients are obtained by expressing the threshold as a quadratic function of the frequency both in their logarithmic form. The perceptual

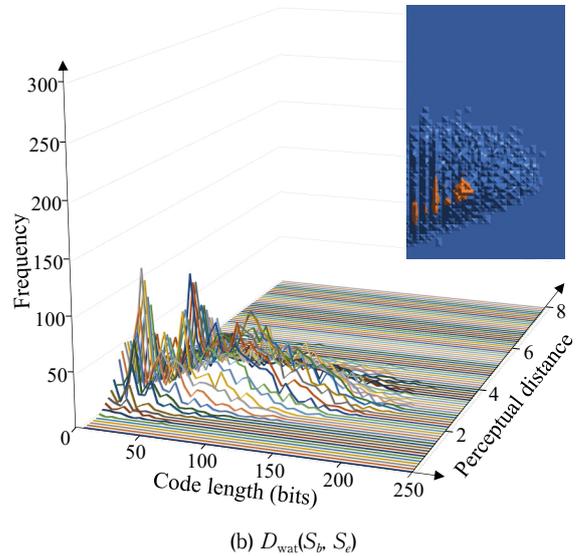
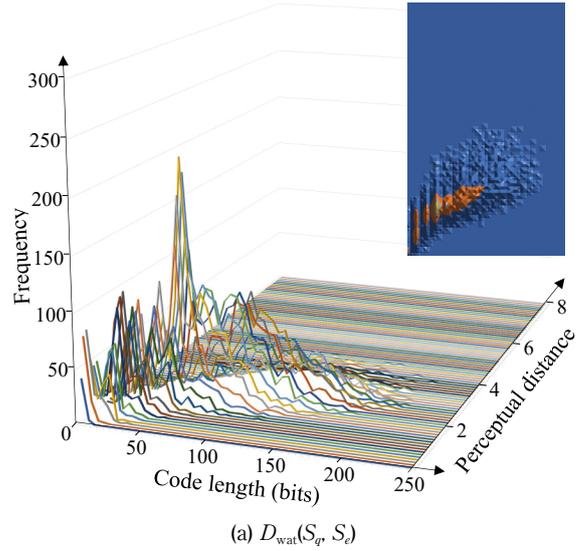


Fig. 7 Histogram and its contour of perceptual distance for u01_Books using conventional method with $(QF1, QF2) = (95, 70)$. (Except DCT blocks having only EOB code)

distance can be quantitatively expressed using Watson's model from the frequency thresholds of 8×8 DCT coefficients, though similar thresholds for higher frequency coefficients are required, since our image compression scheme deals with 16×16 DCT coefficients. Therefore, it is necessary to compare the visual distortion caused by discarding higher frequency DCT coefficients with that caused by embedding enhancement codes in lower frequency DCT coefficients, and consider to what extent it is optimal to embed the codes within the payload from the view point of overall image quality. Of course, the visual distortion caused by bit-embedding in lower frequency and coefficient-discarding in higher

frequency is different: the former is blockiness and the latter is blurring. The enhanced image in the decoder contains these different kinds of distortions at the same time, so that a subjective image quality evaluation is essential to determine the appropriate balancing between them.

The frequency sensitivity thresholds for 8×8 DCT coefficients described above are based on an approximation model using quadratic frequencies, and since the various parameters in the model are determined from subjective evaluation experiments in the literature, they can be regarded as approximations rather than exact values. Therefore, we extend the sensitivity thresholds for 8×8 DCT coefficients into those for 16×16 DCT coefficients by extrapolating them with least-square approximation.

Assume that the frequency sensitivity threshold z_{xy} is given for integer coordinates (x, y) , $0 \leq x, y \leq 7$, on the two-dimensional plane corresponding to the frequency components of 8×8 DCT. We further assume that the threshold z_{xy} is approximated using a function of the distance from the origin of the 2-D plane, when they are expressed in the logarithmic form:

$$\log z_{xy} = a + b \log \sqrt{x^2 + y^2} + c (\log \sqrt{x^2 + y^2})^2, \quad (8)$$

where \log is the natural logarithm, and a , b , and c are arbitrary constants. These constants can be determined by minimizing the sum of squared differences between the actual logarithm of z_{xy} and the approximated value in Equation (8), and we obtain $a = 0.16924$, $b = -0.52246$, and $c = 0.744195$. Note that the approximation procedure is not applied to z_{00} , because it is a special value for DC component and is excluded from the bit-embedding. Thus, we can obtain the frequency sensitivity thresholds for DCT coefficients of 16×16 size for integer values of $0 \leq x, y \leq 15$. The rms error between 8×8 frequency sensitivity threshold matrix [5] and the result of Equation (8) is 1.27.

5. BALANCING OF VISUAL DISTORTION

5.1 ENHANCEMENT CODE CREATION

Now that we have an approximation of the DCT frequency sensitivity threshold for 16×16 size, we

next consider the procedure for balancing the visual distortion caused by the steganographic image coding scheme. In other words, we compare the visual distortion caused by discarding a certain higher frequency DCT coefficient with that caused by embedding the enhancement code in the LSBs of lower frequency DCT coefficients, and consider how to determine the allowable enhancement code length that balances total visual distortion.

Assume that the i -th non-zero coefficient is $S[i]$ through scanning higher frequency quantized DCT coefficients, and the enhancement code length corresponding to that non-zero coefficient is $CL[i]$. Moreover, suppose that the visual distortion by discarding the non-zero coefficient is $S[i]/z_{xy}$. Here, z_{xy} is the DCT frequency sensitivity threshold given by Equation (8), and it can be considered as a JND of the visual distortion. Furthermore, let $D_{\text{wat}}[i]$ be the perceptual distance caused by embedding the codes of length $CL[i]$ in the LSBs of 8×8 quantized DCT coefficients. Since the run-length Huffman coding of higher frequency DCT coefficients determines entropy codes every time a non-zero AC coefficient appears, if we were to encode $S[i]$ and embed the corresponding codes into lower frequency quantized DCT coefficients, then this procedure creates additional perceptual distance of $D_{\text{wat}}[i] - D_{\text{wat}}[i-1]$, where $D_{\text{wat}}[i-1]$ is the perceptual distance caused by the bit-embedding for one previous non-zero coefficient $S[i-1]$. On the other hand, if $S[i]$ are left uncoded and discarded, then the lower frequency distortion remains unchanged, but instead additional higher frequency visual distortion of $S[i]/z_{xy}$ is generated.

For determining the enhancement code length, we consider that the bit-embedding is allowed to continue, if the visual distortion caused by embedding enhancement codes is smaller than that caused by discarding those non-zero AC coefficients. Therefore, the balancing condition for visual distortions from lower and higher frequencies is that if the enhancement codes of length $CL[i]$ can be embedded within the payload PL , then the procedure of enhancement code creation and bit-embedding continues as long as the following inequality is satisfied,

$$D_{\text{wat}}[i] - D_{\text{wat}}[i-1] \leq V_C \cdot S[i]/z_{xy}, \quad (9)$$

where V_C is a constant to adjust the impacts on visual quality when both blockiness and blurring exist. The flowchart of this visual quality balancing (VQB) algorithm is shown in Figure 8. If the distortion associated with the bit-embedding in lower frequency DCT coefficients is greater than that with discarding higher frequency DCT coefficients (i.e., $D_{\text{wat}}[i] - D_{\text{wat}}[i-1] > V_C \cdot S[i]/z_{xy}$), or if the amount of enhancement codes exceeds the payload (i.e., $PL < \sum CL[i]$), then this process is ended, and followed by the code termination process [3].

5. 2 PERFORMANCE EVALUATION FOR 4K IMAGES

First, we obtain a 3-D histogram for the occurrence frequency of the combination of enhancement code length and perceptual distance for DCT blocks similar to Figure 7 by applying the VQB algorithm described above. The histogram for blocks containing only EOB code is omitted here because it is almost the same as the results in Figure 6. The reason is that the balancing algorithm is not applied to blocks containing only EOB code, and also the number of such blocks is almost the same before and after the code termination process. Figure 9 shows the 3-D histogram for DCT blocks whose enhancement code is not a single EOB for image u01_Books, and the contour plot of the frequency diagram. Comparing this result with that of Figure 7, we can see that the enhancement

code length is significantly reduced, and the visual distance in the lower frequency is also reduced accordingly. Although omitted here, the visual quality balancing has a small impact on the results for $QF1 = 70$, indicating that the enhancement code length constraint incorporated to balance the visual distortion has a greater effect in proportion to the size of the payload.

Next, the SNR values of enhanced high-resolution and JPEG-decoded low-resolution images against V_C values are shown in Figure 10 in order to examine the interaction effects when blockiness and blurring are mixed. As V_C becomes large, more high-frequency coefficients are retained because more visual distortion in higher frequency is taken into account. This leads to the resolution enhancement effect, but on the other hand, the blockiness increases as a result of the distortion tradeoff. In other words, as V_C increases, the SNR of the enhanced high-resolution image tends to improve, while that of the JPEG-decoded low-resolution image tends to decrease. The results in the figure show that the tradeoff effect saturates at about $V_C = 5$.

Figure 11 shows the average enhancement code length per block against V_C values. Although the smallest enhancement code length is 4 bits (i.e., EOB code) in a block if the bit-embedding occurs, there are cases where the average is smaller than 4 because the blocks with no enhancement codes are also counted as the number of DCT blocks. It can be seen that the enhancement code length increases steeply in the range of V_C from 1 to about 4, and tends to saturate at about $V_C = 5$. This confirms the SNR results in Figure 10.

The saturated SNR values in Figure 10 are shown in Table 3 (indicated by VQB) in comparison with those obtained by the conventional steganographic compression system (indicated by Conv.). For enhanced high-resolution images, the SNR decreases by about 1 dB when the quality factor is $(QF1, QF2) = (95, 70)$, but it scarcely changes when $(QF1, QF2) = (70, 70)$. This is due to the fact that the relationship between the enhancement code length and the perceptual distance almost unchanged by the VQB algorithm in the latter case. On the other hand, for JPEG-

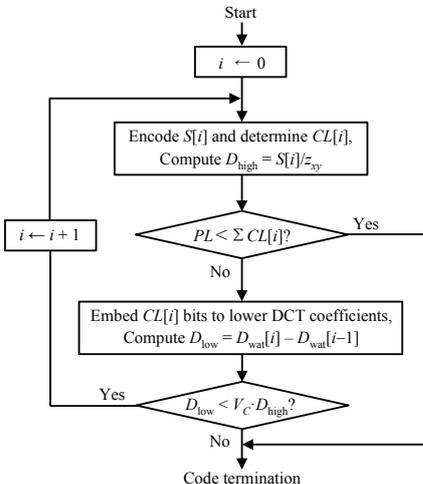


Fig. 8 Visual quality balancing (VQB) algorithm.

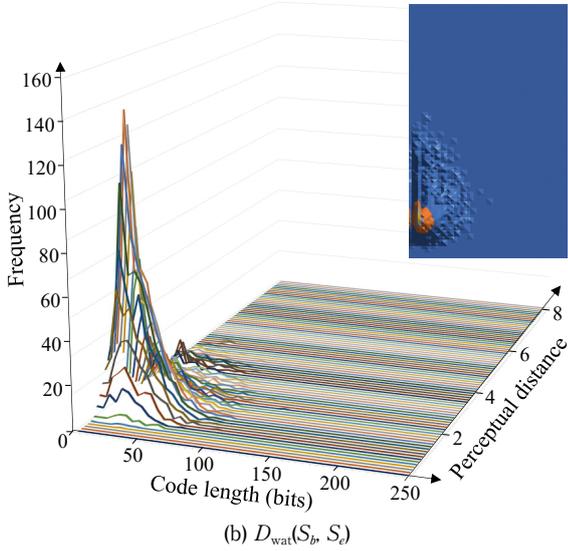
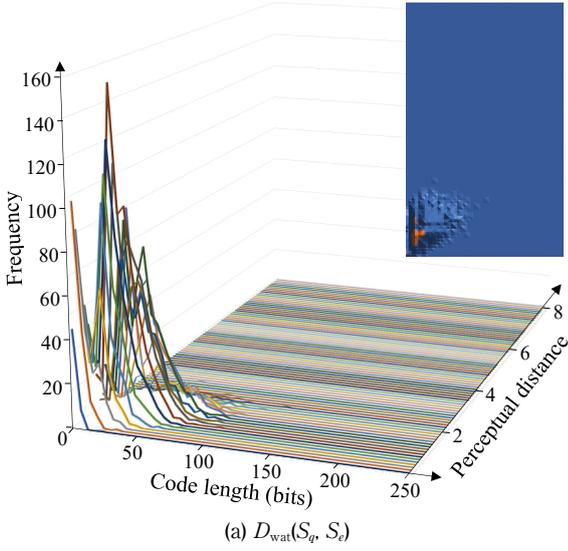


Fig. 9 Histogram and its contour of perceptual distance for u01_Books using VQB algorithm with $(QF1, QF2) = (95, 70)$. (Except DCT blocks having only EOB code)

decoded low-resolution images, the SNR is greatly improved when $(QF1, QF2) = (95, 70)$. In the proposed steganographic compression system, which determines the enhancement code length by the VQB algorithm shown in Figure 8, the objective is to improve the quality of the JPEG-decoded low-resolution image while minimizing the quality degradation of the enhanced high-resolution image. The results in Table 3 show that the goal depicted in Figure 2 has been achieved: great improvement is obtained for SNR values of JPEG-decoded low-resolution images in exchange for slight degradation in enhanced high-resolution images.

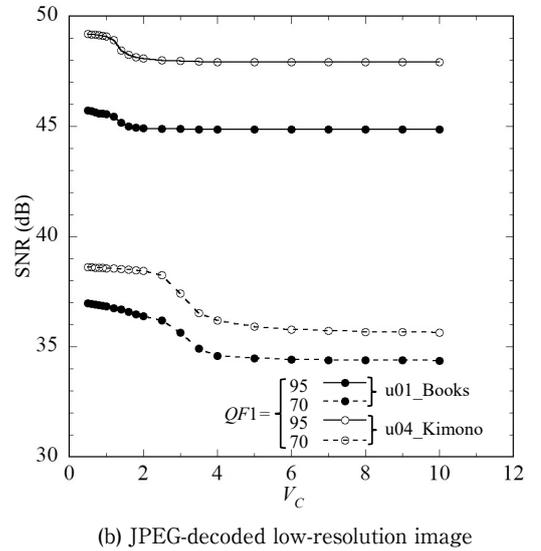
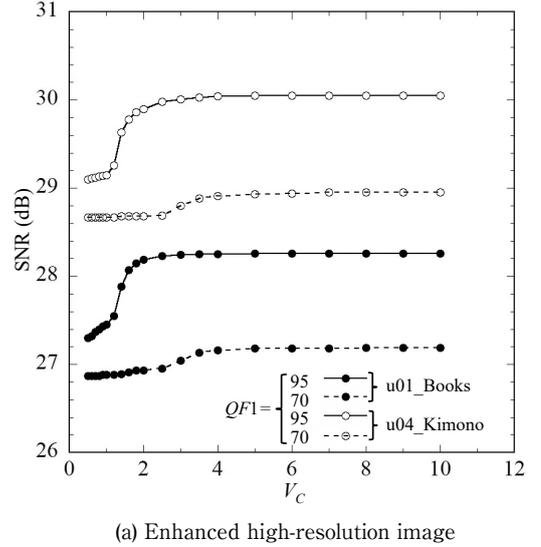


Fig. 10 SNR values of enhanced high-resolution image and JPEG-decoded low-resolution image.

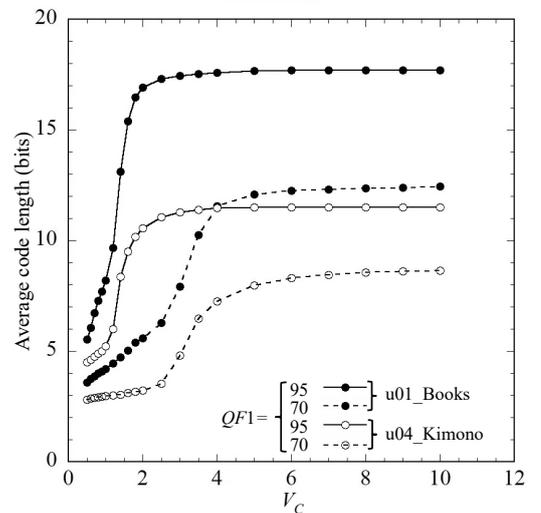


Table 3 Comparison of SNR (dB).

(a) u01_Books

Recovered image		Enhanced		JPEG-decoded	
Coding method		Conv.	VQB	Conv.	VQB
Quality factor	(95, 70)	29.24	28.26	37.76	44.86
	(70, 70)	26.97	27.19	32.60	34.37

(b) u04_Kimono

Recovered image		Enhanced		JPEG-decoded	
Coding method		Conv.	VQB	Conv.	VQB
Quality factor	(95, 70)	31.05	30.05	38.44	47.92
	(70, 70)	28.67	28.95	33.82	35.65

Figure 12 demonstrates some examples of enhanced high-resolution images and JPEG-decoded low-resolution images for u01_Books with both the conventional and the proposed VQB methods. In the simulation experiment, the parameters for compressing images are $(QF1, QF2) = (95, 70)$, $n_{LSB} = 6$, and $V_c = 10$, so that the visual quality balancing algorithm is working in the saturation region. Note that the upper left image in the figure is identical to the image (a) in Figure 5. It can be seen that the blockiness is reduced in both resolution images when the compression system brings the VQB algorithm into action. In addition, the enhanced high-resolution images with VQB are slightly low-pass filtered compared to those without VQB, and the JPEG-decoded low-resolution images achieve a more reasonable level of quality than the conventional method. In particular, in the low-resolution image, a large blockiness occurs in the shadow under the bookshelf slightly above the center of the cropped image for the conventional method, but the distortion is no longer visible in the proposed method.

Figure 13 shows the results of the enhancement code length per block and the lower frequency perceptual distance for the entire image u01_Books by comparing those for the conventional method (without VQB) and the proposed method (with VQB). The simulation parameters $(QF1, QF2)$, n_{LSB} , and V_c are the same as those in Figure 12. Each value in the figure is shown as a luminance with a maximum value of 255, and the perceptual distance is 50 times the original value. Therefore, the brighter area corresponds to larger values. From the figure, we can see that the proposed image compression method with VQB algorithm restricts the length of

enhancement codes in relatively complex texture regions of images, which in turn reduces the perceptual distance in JPEG-decoded low-resolution images.

6. CONCLUSION

In the steganographic image compression system with information embedding, there is a tradeoff between the lower and higher frequency visual distortions, and an algorithm for determining the allowable enhancement code length to balance the visual distortions was proposed, and simulation experiments were made on actual standard 4K images. As a result, compared with the conventional method of embedding as much enhancement codes as possible within the payload, which intended to improve the quality of only enhanced high-resolution image, it was confirmed that the proposed visual quality balancing (VQB) algorithm can improve the fidelity of JPEG-decoded low-resolution image in exchange for a slight decrease in the fidelity of enhanced high-resolution image.

The perceptual distances for 16×16 higher frequency DCT coefficients used here are obtained by extrapolating from those obtained by Watson's model for 8×8 lower frequency DCT coefficients, so they are only based on approximate calculations. In particular, when blockiness occurs at lower frequencies in conjunction with blurring at higher frequencies, the optimization of mutual balancing is challenging.

In this study, we adopted an algorithm that multiplies the higher frequency visual distortion by a certain constant and compares the distortion with the lower frequency perceptual distance, and continues the enhancement code embedding as long as the perceptual distance caused by bit-embedding is smaller than the visual distortion caused by discarding the corresponding AC coefficients. The optimal constant value used in the inequality that determines whether the run-length Huffman encoding and the associated bit-embedding should be continued varies from image to image, and may even vary from region to region within a single image, so that the results obtained here should be considered as a guideline for practical constructions of the

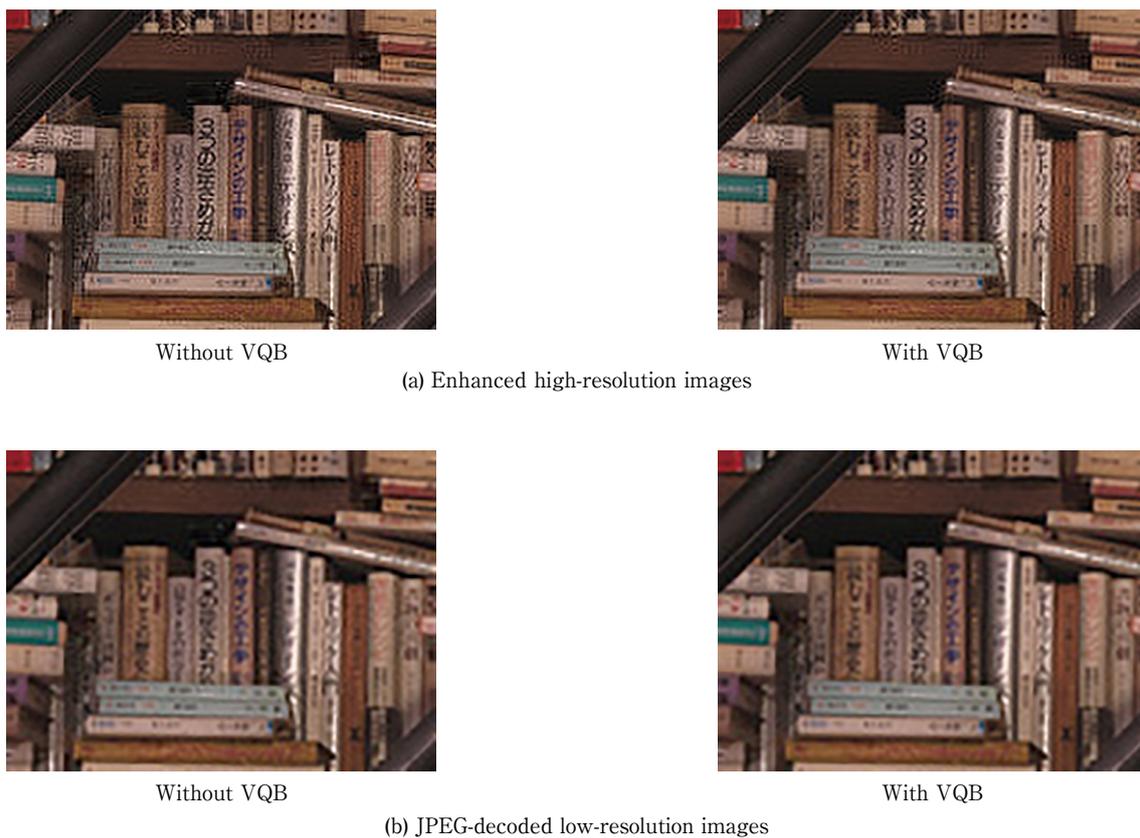


Fig. 12 Comparison of recovered images for u01_Books with and without visual quality balancing (VQB).

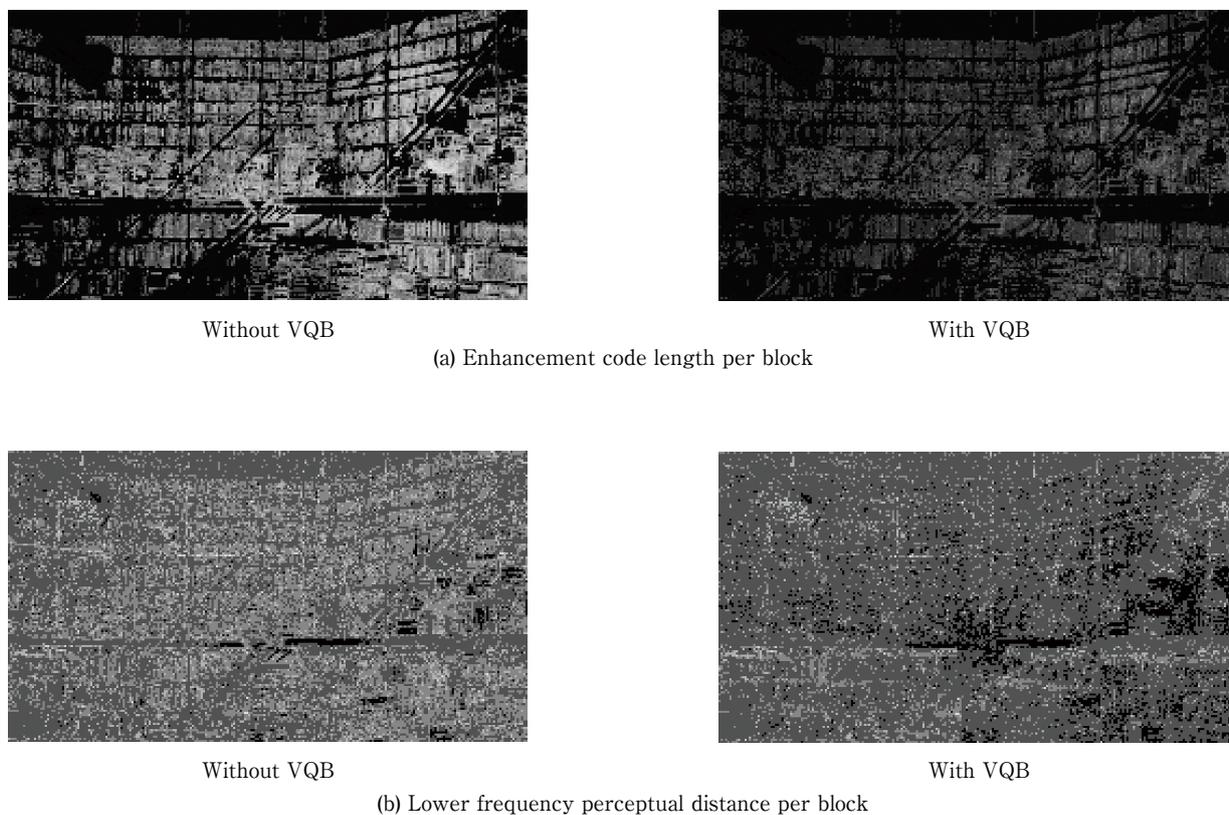


Fig. 13 Comparison of enhancement code length and lower frequency perceptual distance for u01_Books with and without visual quality balancing (VQB).

algorithm. However, it is generally confirmed that a slight decrease in SNR of enhanced high-resolution images results in a large increase in SNR of JPEG-decoded low-resolution images. It will provide a rule-of-thumb technique for visual quality balancing in the conventional steganographic image compression system.

Our future topics are to evaluate subjective image quality in detail and to develop an optimum control method for the amount of enhancement codes by balancing lower and higher frequency visual distortions with the use of diverse images and simulation parameters.

REFERENCES

- [1] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*, Springer, 2002.
- [2] I. Furukawa and J. Suzuki, "A novel approach of image enhancement using steganographic data embedding and JPEG scheme," *Bull. Hiroshima Inst. Tech. Research*, Vol.48, pp.93-103, Feb. 2014.
- [3] I. Furukawa and J. Suzuki, "JPEG-based image enhancement method with steganographic data embedding and its performance evaluation," *Bull. Hiroshima Inst. Tech. Research*, Vol. 49, pp.97-108, 2015.
- [4] I. Furukawa and J. Suzuki, "Steganographic image compression method with adaptive enhancement code embedding and its performance evaluation on 4K standard images," *Bull. Hiroshima Inst. Tech. Research*, Vol. 55, pp. 63-76, Feb. 2021.
- [5] J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd Ed., pp. 269-273, Morgan Kaufmann Publishers, 2008.
- [6] J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proceedings of the SPIE*, 1666, pp.365-374, 1992.
- [7] I. Matsuda, K. Masaoka, and H. Ikegawa, "Ultra-high definition/wide-color-gamut standard test images," *Journal of ITE*, Vol. 68, No. 8, pp.643-647, 2014.