Reprinted from
Journal of Imaging Science and Technology® 64(5): 050409-1–050409-15, 2020.

Glossiness-aware Image Coding in JPEG Framework

Midori Tanaka▲
College of Liberal Arts and Sciences, Chiba University, Chiba, Japan
E-mail: midori@chiba-u.jp

Tomoyuki Takanashi and Takahiko Horiuchi▲
Graduate School of Engineering, Chiba University, Chiba, Japan

Abstract. *In images, the representation of glossiness, translucency, and roughness of material objects (Shitsukan) is essential for realistic image reproduction. To date, image coding has been developed considering various indices of the quality of the encoded image, for example, the peak signal-to-noise ratio. Consequently, image coding methods that preserve subjective impressions of qualities such as Shitsukan have not been studied. In this study, the authors focus on the property of glossiness and propose a method of glossiness-aware image coding. Their purpose is to develop an encoding algorithm that produces images that can be decoded by standard JPEG decoders, which are commonly used worldwide. The proposed method consists of three procedures: block classification, glossiness enhancement, and non-glossiness information reduction. In block classification, the types of glossiness in a target image are classified using block units. In glossiness enhancement, the glossiness in each type of block is emphasized to reduce the amount of degradation of glossiness during JPEG encoding. The third procedure, non-glossiness information reduction, further compresses the information while maintaining the glossiness by reducing the information in each block that does not represent the glossiness in the image. To test the effectiveness of the proposed method, the authors conducted a subjective evaluation experiment using paired comparison of images coded by the proposed method and JPEG images with the same data size. The glossiness was found to be better preserved in images coded by the proposed method than in the JPEG images.* © *2020 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2020.64.5.050409]

1. INTRODUCTION

Image encoding is essential for reducing the memory required for storage and data transmission time. Many methods for compressing image information while preserving the quality of decoded images have been developed. JPEG is a widely used international standard for encoding images. The group that developed the JPEG format also developed JPEG 2000, which is based on wavelet transform [1]. Google developed WebP, which supports the use of the alpha channel [2], and Bellard proposed the Better Portable Graphics (BPG) format [3], which is based on the High Efficiency Video Coding (HEVC) technique [4]. The Moving Picture Experts Group proposed the High Efficiency Image File format (HEIF) [5], which is based on HEVC.

Further efforts have been made to improve the encoding performance of the JPEG method. Mozilla developed the JPEG encoder mozjpeg to further compress data while maintaining compatibility with JPEG [6]. It is based on libjpeg-turbo [7], which is derived from the JPEG library libjpeg [8], and its development is ongoing. Alakuijala et al. proposed the JPEG encoder Guetzli [9], which was optimized using the image quality evaluation algorithm Butteraugli [10]. However, the subjective quality of images encoded by these methods is severely degraded by blur at low bit rates and visual artifacts such as ringing and blocking. The reason is that they were designed based on an objective evaluation index such as the peak signal-to-noise ratio (PSNR) as the main measure of the quality of the decoded image.

In recent years, qualities of material objects such as glossiness, translucency, and roughness (Shitsukan) have attracted attention in various fields such as neuroscience, psychology, and imaging engineering [11, 12]. Digital image processing generally does not preserve the Shitsukan of actual objects in reproduced images [13]. Therefore, editing techniques that enable appropriate rendering of the Shitsukan of objects in images have been studied, for example, Shitsukan enhancement [14, 15]. An encoding method that preserves the subjective qualities of an object becomes essential for the adequate representation of Shitsukan in rendered images. However, no encoding method has been proposed so far in which the Shitsukan is positively preserved. In this study, we construct a new image encoding method that can preserve the Shitsukan property of glossiness. Our purpose is to develop an encoder that produces images that can be decoded using a standard JPEG decoder because JPEG is commonly used worldwide.

The remainder of this article is structured as follows. Section 2 provides an overview of previous studies related to the problem. In Section 3, we propose an encoding method that preserves glossiness by adding processing steps to the standard JPEG encoder. To verify the effectiveness of the proposed method, in Section 4, we show the results of subjective evaluation experiments. Section 5 presents conclusions and directions for future work.

▲ IS&T Members.

Received Mar. 26, 2020; accepted for publication Aug. 6, 2020; published online Sep. 11, 2020. Associate Editor: Mathieu Hebert.
1062-3701/2020/64(5)/050409/15/$25.00

2. RELATED WORKS

Image coding methods such as JPEG2000 [1], BPG [3], and HEIF [4] have been proposed to improve the compression efficiency of JPEG encoding and add new functions. Their encoding efficiency could potentially exceed that of the JPEG standard, but they are supported by fewer software packages and browsers than JPEG. In addition, encoding and decoding can be performed more rapidly by the JPEG method than by the improved methods because the encoding and decoding processes are simple. Therefore, JPEG is still frequently used as an image coding method.

Therefore, several methods for improving the coding performance within the JPEG framework have been proposed. The main steps in the JPEG process that can be changed are chroma subsampling and discrete cosine transform (DCT) coefficient quantization. In particular, DCT coefficient quantization has a greater effect on the coding result than the other steps. Thus, methods that improve the compression efficiency by modifying the quantization process have been proposed. Wu et al. [16] proposed a method for estimating the optimal quantization table that maximizes the rate/distortion (RD) ratio of JPEG images using a recursive algorithm. Many methods have been introduced to estimate the optimal quantization table by solving the RD optimization problem [17–21]. Crouse et al. [19] proposed a method for estimating a binary threshold parameter of the same size as the quantization table in RD optimization. Ratnakar et al. [21] proposed a method for estimating a threshold table of the same size as the quantization table in RD optimization. With improvements in computer performance, a method for estimating the optimal quantization table using a metaheuristic algorithm has been proposed [22]. Wang et al. [23] estimated the optimal quantization table using a genetic algorithm and confirmed that coded images with a higher compression ratio and superior image quality relative to JPEG images could be obtained. Tuba et al. [24] proposed a method for estimating the optimal quantization table using the guided fireworks algorithm. However, coding methods that actively evaluate the subjective qualities in the framework of JPEG have not been adequately studied.

In recent years, techniques for generating images that maintain their subjective qualities using generative adversarial networks [25] have been proposed. In these techniques, images are encoded to obtain a color distribution similar to that of the input image by optimizing the distribution of the reproduced image without considering the PSNR. Rippel et al. [26] reported a coding method that yielded better results than the HEVC-based BPG system by adopting the resolution pyramid and MS-SSIM-based optimization. Their architecture consisted of an autoencoder featuring pyramidal analysis, an adaptive coding module, and regularization of the expected code length. They also supported their approach with adversarial training specialized for use in a compression setting. Consequently, they could produce visually pleasing reconstructions at very low bit rates. Using the same optimization scheme, Agustsson et al. [27] proposed a system

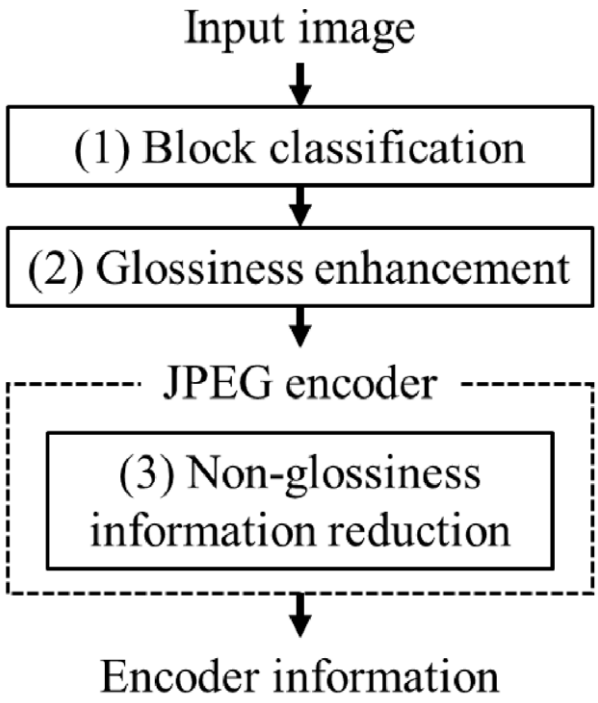


Figure 1. Processing steps in the proposed method.

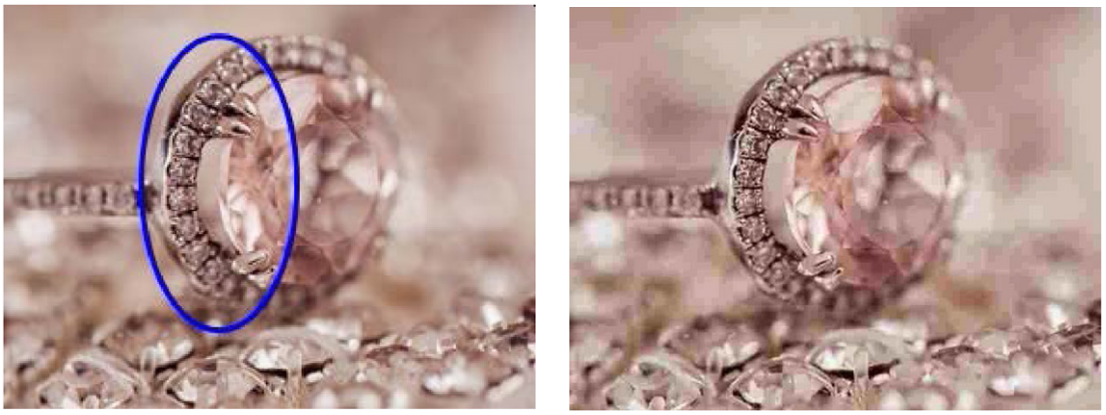
that integrated the decoding domain using an automatic encoder with a domain generated from a user-defined label domain. Their model synthesized details that it could not afford to store; it yielded visually pleasing results at bit rates at which previous methods failed and showed strong artifacts. However, it is well known that viewers' perceptual impressions can change even though the appearance seems natural because these methods are optimized according to whether a restored image is subjectively natural. Therefore, these approaches cannot be applied to represent Shitsukan. Kudo et al. [28] proposed an improved version of these techniques that optimizes the coding features obtained from the encoder and explicitly correlates them to the restored images by introducing mutual information maximization as a type of regularization. However, it does not address the encoding of the Shitsukan of material objects.

3. PROPOSED IMAGE CODING METHOD

In this section, we propose an encoding method that preserves glossiness by adding processing steps to the standard JPEG encoder.

3.1 *Outline of the Method*

The procedure of the proposed method is shown in Figure 1. Our method uses an encoder that adds processing steps that preserve Shitsukan to the standard JPEG encoder. It consists of three processes: (1) block classification, (2) glossiness enhancement, and (3) non-glossiness information reduction. To implement the third procedure, we set the general JPEG encoder as the target for application of our method, which can meet the following two conditions:



(a) Original image [29] including the Block A information (inside blue circle) (BMP, 819,054 bytes)

(b) JPEG image with glossiness loss (JPEG, 8,747 bytes)

Figure 2. Example of JPEG encoding of an image including Block A information.

Condition 1: The target encoder quantizes the DCT coefficients by a general procedure.

Condition 2: The target encoder can set a progressive mode.

These two conditions are necessary for efficient compression by non-glossiness information reduction. In JPEG processing, Huffman coding is generally performed after the DCT coefficients are obtained. By storing the DCT coefficients in the progressive mode in the order of the frequency band, the DCT coefficient can be made continuous with 0 and the compression rate can be increased by Huffman coding.

3.2 *Block Classification*

In block classification, the input JPEG image is divided into blocks of 8×8 pixels, and each block is classified according to the type of glossiness information as follows:

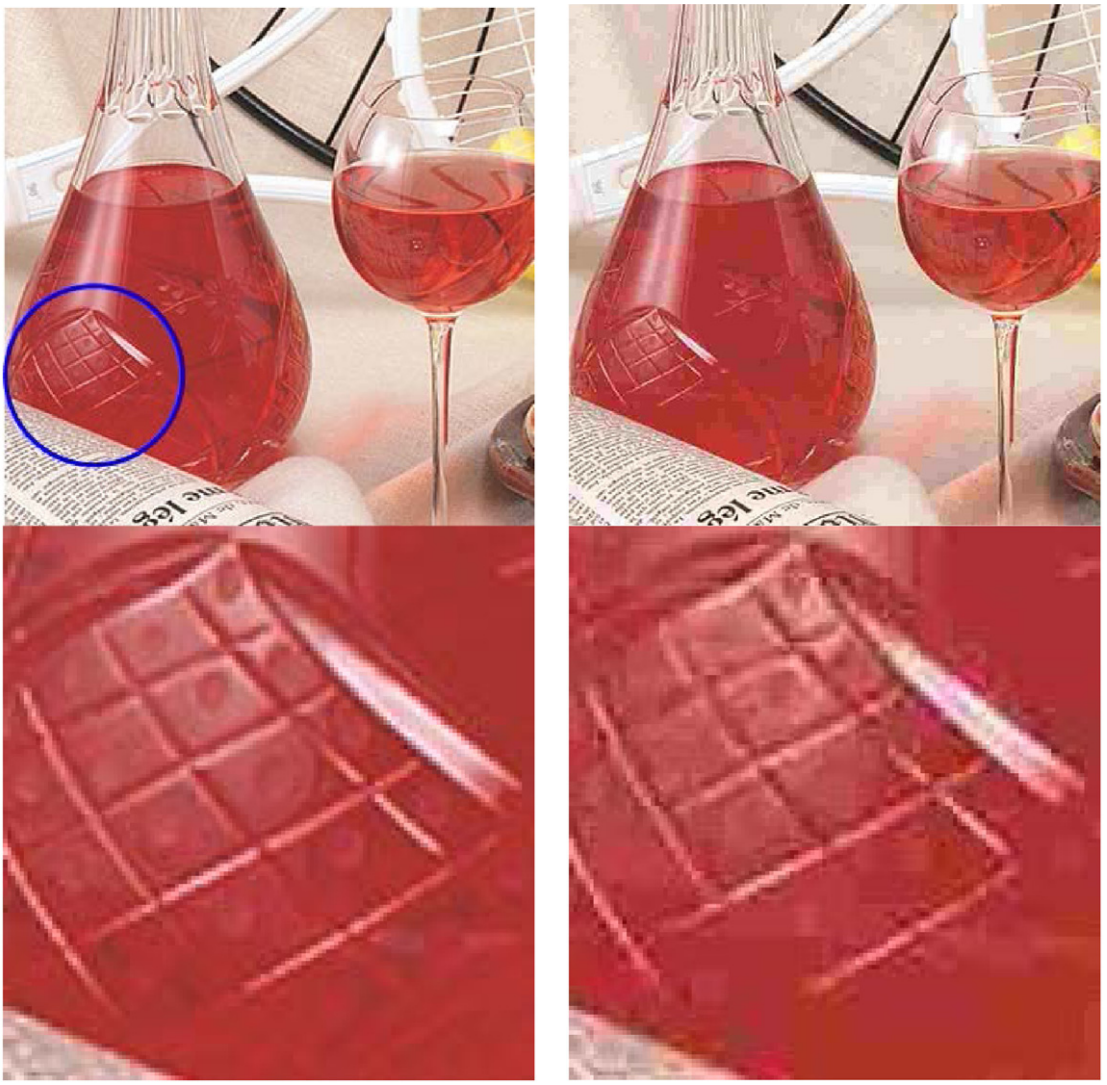
- Block A includes areas of specular reflectance with a complex shape on the object surface.
- Block B includes glossiness caused by intense highlights.
- Block C includes glossiness associated with dull highlights.
- Block D does not contain any of the types of glossiness in Blocks A–C.

Block A includes brilliant glossiness (e.g., glitter or sparkle) at high spatial frequencies, where bright shiny areas are generated by reflected light and the dark area becomes shadowed. Hence, the contrast between bright and dark areas is higher when the incident light is stronger. Because the superficial intensities of the incident and the reflected light increase with increasing contrast, high glossiness is also perceived with increasing contrast. By contrast, the glossiness is lost when the contrast decreases. Blur in a digital image generally results from a quantization error in processing during the standard JPEG encoding. As shown in Figure 2, this blur causes the contrast to decrease, causing a loss of glossiness in Block A. Therefore, it is necessary to preserve the image information so as to prevent this loss. Thus, detection of Block A is critical.

Block B includes specular highlights, the clarity and brilliance of which affect the glossiness. Objects with high specular reflectance appear very glossy. The background reflections of the surrounding highlights and the distinctness of the image also become more robust. Therefore, the perception of glossiness is stronger when the incident light and the reflected light, and thus the highlights, are more intense. In Block B, blur and noise such as mosquito noise appear in highlight areas after standard JPEG encoding as shown in Figure 3. These cause loss of glossiness owing to decreased highlight intensity and blur at the edges of the highlight. It is necessary to detect Block B so that the image information is preserved and the loss of glossiness is prevented.

In Block C, specular highlights are diffused on an object surface, and the smooth, dull gloss resulting from this highlight diffusion affects glossiness. Block C does not contain intense reflected light as Block B does, but it has unique gloss depending on the illumination. Hence, glossiness can be perceived when the glossiness of highlights is adequately reproduced. As shown in Figure 4, blur and noise also appear in the highlight after standard JPEG encoding. This result indicates that the loss of glossiness is generated by roughness at the edge of the highlight and decrease in the highlight intensity. It is necessary to detect Block C to prevent this loss.

Further, as the compression rate increases, blocking noise occurs as does in all blocks and affects the perception of glossiness. In this study, we develop a detection algorithm for Blocks A–C or Block D by combining edge detection and



(a) Original image (ISO-SCID 300) including Block B information (inside blue circle) (BMP, 1,058,454 bytes)

(b) JPEG image with loss of glossiness (JPEG, 13,995 bytes)

Figure 3. Example of JPEG encoding of an image including Block B information.

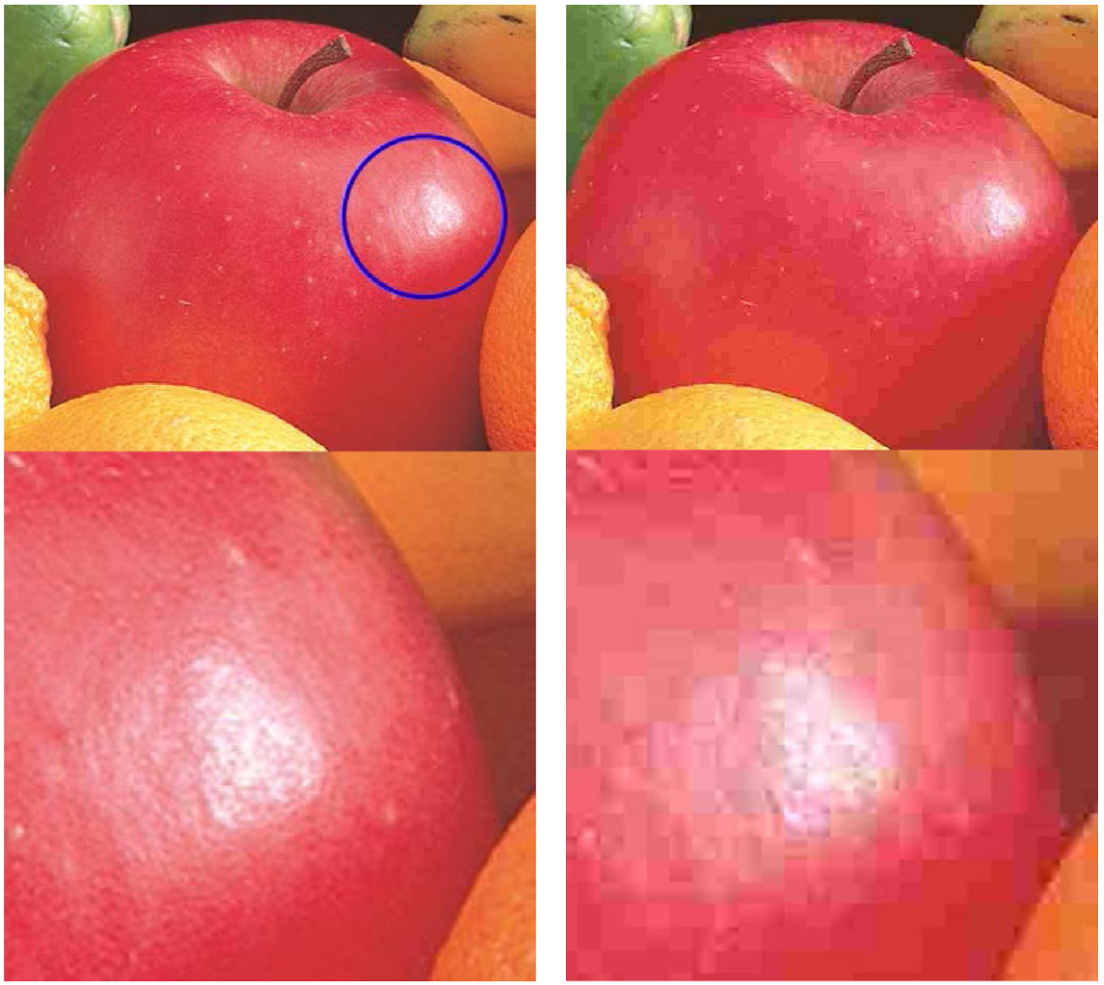
subband decomposition. Here, each block can only belong to one Block (A, B, C, or D) and cannot belong to more than one.

3.2.1 *Subband Decomposition*

Boyadzhiev et al. [15] showed that a subband in the spatial frequency of an image could represent specific Shitsukan information and indicated how this information could be edited by increasing and decreasing these subband signals. In their work, an image with a specific frequency band (called a subband) was obtained by calculating the difference of the smoothed image using sequential kernel sizes after applying an edge-preserving smoothing filter while doubling the kernel size for the luminance component of the original image. As shown in Table I, $2^3 = 8$ types of subband images $\{SB_{\mathrm{HHP}}, SB_{\mathrm{HHN}}, SB_{\mathrm{HLP}}, SB_{\mathrm{HLN}}, SB_{\mathrm{LHP}}, SB_{\mathrm{LHN}}, SB_{\mathrm{LLP}}, SB_{\mathrm{LLN}}\}$ are obtained by dividing the image with a specific frequency band into high-frequency and low-frequency images and then dividing it further according to the amplitude and sign. Table II summarizes the image characteristics for each standard, and Figure 5 shows the result of subband decomposition of the image in Fig. 3(a).

3.2.2 *Subband Block Extraction*

By using the subband decomposition explained in Section 3.2.1, the blocks with glossy features are detected by calculating the subband component of each block of 8×8 pixels. First, for each subband component, the absolute averaged values $\{SB_{\mathrm{HHP}}(i), SB_{\mathrm{HHN}}(i), SB_{\mathrm{HLP}}(i), SB_{\mathrm{HLN}}(i), SB_{\mathrm{LHP}}(i), SB_{\mathrm{LHN}}(i), SB_{\mathrm{LLP}}(i), SB_{\mathrm{LLN}}(i)\}$ of each pixel in block i are



(a) Original image (ISO-SCID 300) including Block C information (inside blue circle) (BMP, 909,054 bytes)

(b) JPEG image with loss of glossiness (JPEG, 6,488 bytes)

Figure 4. Example of JPEG encoding of an image including Block C information.

calculated. In addition, the information for each subband is binarized using Otsu's discriminant analysis method [30]. Then, we can obtain $\{SB^B{}_{\mathrm{HHP}}(i)$, $SB^B{}_{\mathrm{HHN}}(i)$, $SB^B{}_{\mathrm{HLP}}(i)$, $SB^B{}_{\mathrm{HLN}}(i)$, $SB^B{}_{\mathrm{LHP}}(i)$, $SB^B{}_{\mathrm{LHN}}(i)$, $SB^B{}_{\mathrm{LLP}}(i)$, $SB^B{}_{\mathrm{LLN}}(i)\}$. Here, this component is set as 1 if each subband component is sufficient; otherwise, it is set as 0. Figure 6 shows the results of binarization of each block for the subband images in Fig. 5.

3.2.3 *Edge Block Extraction*

The reproduction of edges is also essential to preserve glossiness. In our proposed method, edge components, in addition to subband components, are used for block decomposition.

As in a subband block extraction, the edge intensity of the luminance component of the original image is extracted in a block unit of 8×8 pixels. First, a Laplacian filter is applied to the luminance component of the original image, and the edge extraction image E is obtained. Second, for each

Table I. Eight types of subband.

Subband name	Components of each subband
HHP	High frequency, high amplitude, positive sign
HHN	High frequency, high amplitude, negative sign
HLP	High frequency, low amplitude, positive sign
HLN	High frequency, low amplitude, negative sign
LHP	Low frequency, high amplitude, positive sign
LHN	Low frequency, high amplitude, negative sign
LLP	Low frequency, low amplitude, positive sign
LLN	Low frequency, low amplitude, negative sign

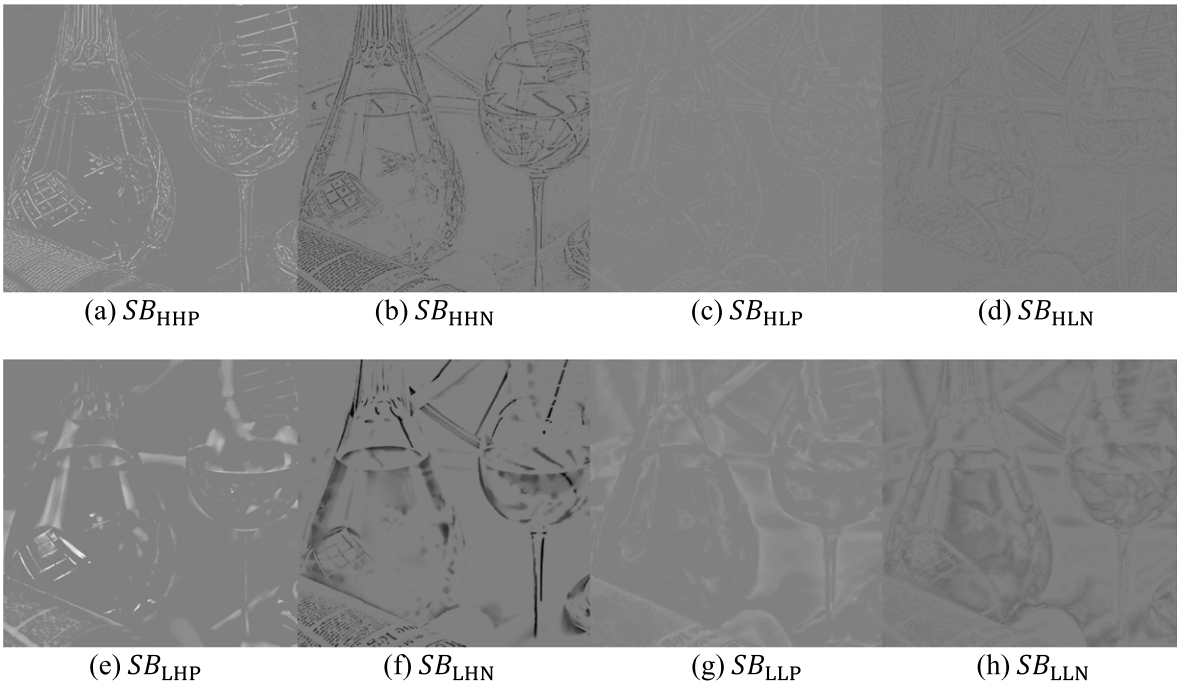


Figure 5. Subband decomposition of image in Fig. 3(a).

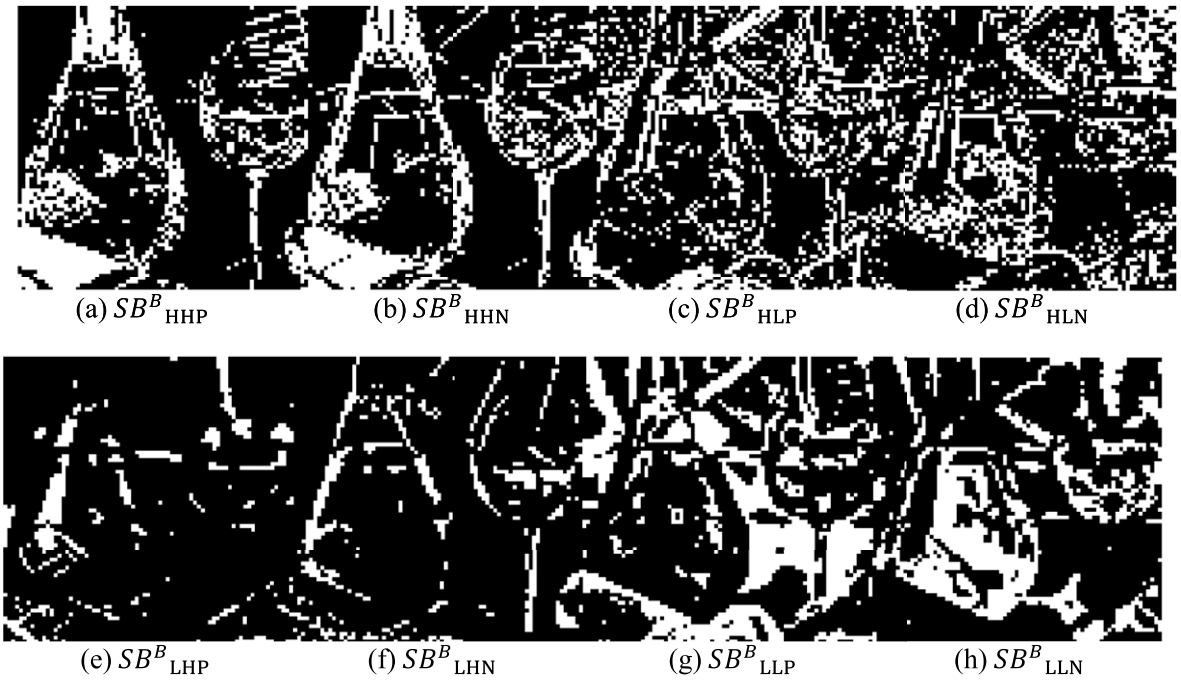
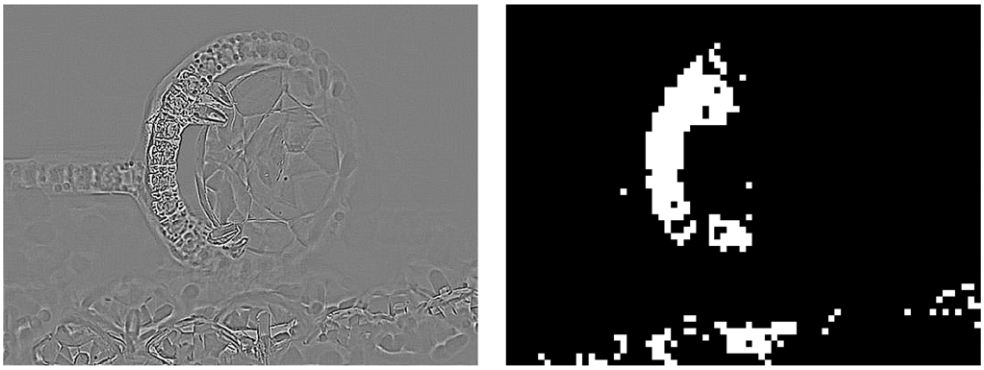


Figure 6. Binary images of each block in Fig. 5. White indicates high subband components.

edge component, the absolute averaged value of each pixel in block i, $EB(i)$, is calculated. Then, all the blocks are binarized using Otsu's discriminant analysis method, and we can obtain $EB^B(i)$, which is 1 if the edge component of block i is large and 0 otherwise. Figure 7 shows the results of edge block extraction and the extracted edge component of the image in Fig. 2(a).

3.2.4 *Classification*

On the basis of the subband blocks and edge blocks, all blocks are classified as belonging to Blocks A, B, C, or D.



(a) Edge extraction result (b) Binary image of extracted edge block

Figure 7. Results of extraction of edge component of image in Fig. 2(a). White indicates a high edge component.

Table II. Image characteristics represented by frequency, amplitude, and sign.

Frequency	High	Small features such as roughness and wrinkles
	Low	Large-scale features such as global highlight and shadows
Amplitude	High	High-contrast features such as intense highlights and shadows
	Low	Low-contrast features such as weak highlights and shadows
Sign	Positive	Bright features such as highlights
	Negative	Dark features such as shadows

Block A: This block has detailed specular reflectance with a complex shape on the object surface.

When block i meets the following conditions, the target block is classified as Block A:

- The block includes rich edge information $[EB^B(i) = 1]$.
- The block includes the HHP subband component, the HHN subband component, or both $[SB^B{}_{\mathrm{HHP}}(i) = 1$ OR $SB^B{}_{\mathrm{HHN}}(i) = 1]$.
- The block has few LHP and LHN subband components $[SB^B{}_{\mathrm{LHP}}(i) = 0$ AND $SB^B{}_{\mathrm{LHN}}(i) = 0]$.

Here, "OR" and "AND" represent logical operators. Block A has a complex shape and generally includes considerable edge information. In addition, many of the HHP and HHN subband components are included because Block A has a strong specular highlight associated with the high contrast generated by a small highlight and shadow. However, it does not include large-scale highlights and shadows, which are represented by the LHP and LHN subband components. Figure 8 shows an example of the results of Block A classification. We confirmed that gloss information was detected around a complex shape on a decorative ring.

Block B: This block has glossiness caused by an intense highlight.

When block i meets the following conditions, the target block is classified as Block B:

- The block includes sufficient HHP and LHP subband components $[SB^B{}_{\mathrm{HHP}}(i) = 1$ AND $SB^B{}_{\mathrm{LHP}}(i) = 1]$.

Block B includes strong glossy highlights. Specularly reflected light from the object surface corresponds to a surrounding high-frequency environment illuminated by light with a large low-frequency component. Under these conditions, the highlights become strong, and they include both the HHP and LHP subband components. Figure 9 shows the results of Block B classification. A strong glossy area with a highlight peak around the surface of the decanter was extracted.

Block C: This block has glossiness associated with a dull highlight.

When block i meets the following conditions, the target block is classified as Block C:

- The block includes sufficient LHP subband components $[SB^B{}_{\mathrm{LHP}}(i) = 1]$.
- The block has few HHP subband components $[SB^B{}_{\mathrm{HHP}}(i) = 0]$.

Because Block C includes a smooth highlight that is diffused into the area surrounding the highlight edge, it includes a large LHP subband component. However, the information representing the reflection of the surroundings on the object surface (i.e., Block B) is not included. Therefore, the high-frequency HHP subband component is not included. Figure 10 shows the results of Block C classification, which demonstrates that the glossiness in the dull highlights on the surfaces of the decanter and wine glass was detected.

Block D: This block has other features.

All blocks that are not identified as Blocks A, B, and C are classified as Block D. Figure 11 shows examples of each type of block detected in Figs. 2(a), 3(a), and 4(a).

(a) Original image (b) Block A

Figure 8. Classification result of Block A.

(a) Original image (b) Block B

Figure 9. Classification result of Block B.

3.3 *Glossiness Enhancement*

To prevent the loss of glossiness during JPEG encoding, we apply adaptive glossiness enhancement to each block type as follows.

Block A: Blur owing to quantization during JPEG encoding reduces the detailed contrast associated with the complex shape, and glossiness decreases. Therefore, for Block A, a sharpening filter is used.

Blocks B and C: When highlights become less distinct during JPEG encoding, glossiness decreases. Hence, glossiness enhancement of highlights is applied to Blocks B and C. To enhance the highlights, we emphasize the LHP subband component using a band-sifting method [15]. The multiplication coefficient of the LHP component of the subband is experimentally set as 1.7. Here, pixel values above 255 are set as the maximum, and all higher values are clipped. Figure 12 shows examples of the results of glossiness enhancement for each block in Figs. 2(a), 3(a), and 4(a). These examples demonstrate the effects of the proposed glossiness enhancement for many types of glossiness.

3.4 *Non-Glossiness Information Reduction*

In the proposed method, we improve the compression ratio by reducing information that is not related to the perception of glossiness during JPEG encoding. In particular, for Block C, which has no high-frequency glossiness component (HHP subband component), and Block D, which has no glossiness information, a DCT component with a small absolute value

(a) Original image (b) Block C

Figure 10. Classification result of Block C.

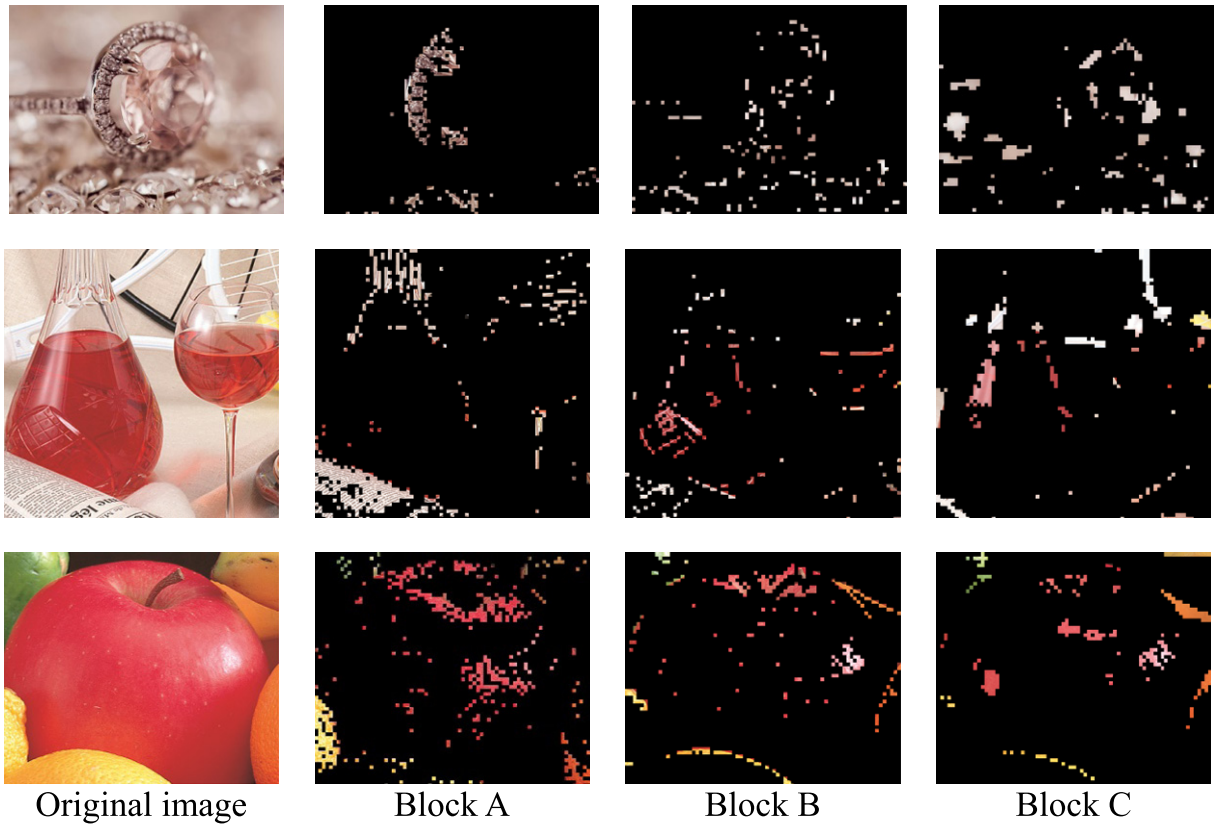


Figure 11. Classification results of each block.

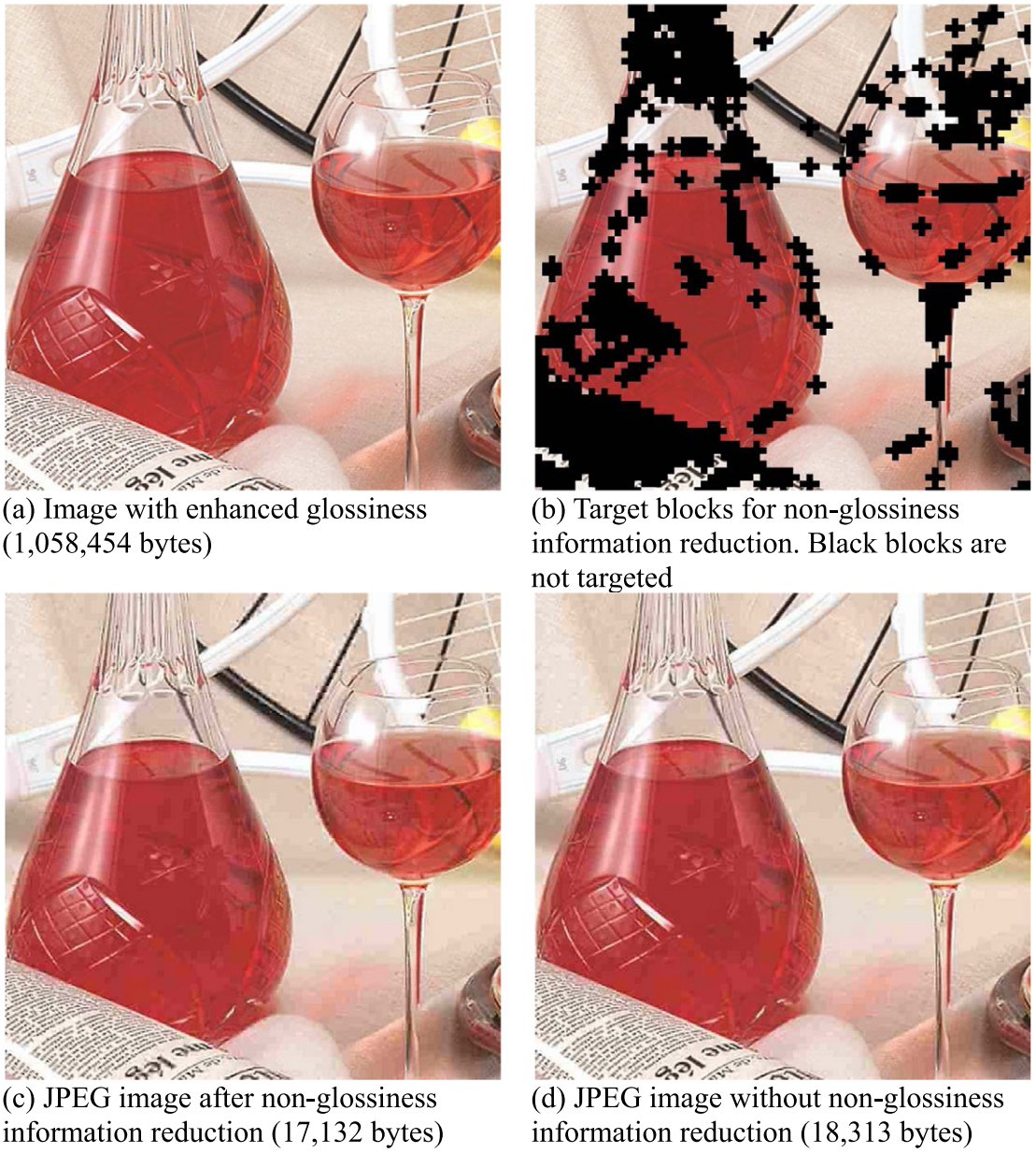
of 0 is assigned based on the ranking method described below. Block C is typically targeted for information reduction because in natural images, the DCT coefficient for low frequencies is significant and that for high frequencies is low.

(a) Original image (b) Image with enhanced glossiness

Figure 12. Glossiness enhancement result.

However, to reduce block distortion, Blocks C and D located at a 4-neighboring position of Blocks A and B are not targeted for non-glossiness information reduction.

In Blocks C and D, only the three components with the highest absolute values of the DC component and the DCT coefficient were used for quantization, and the other 60 DCT coefficients were set as 0. This value was experimentally determined on the basis of a subjective evaluation of the appropriate number of coefficients to retain. Figure 13 shows the effects of non-glossiness information reduction of the image in Fig. 3(a). Figs. 13(a) and 13(b) show the image in Fig. 3(a) after non-glossiness information reduction and the blocks targeted for non-glossiness information reduction, respectively. The black blocks were not targeted for information reduction, and we can confirm that most of them contained information associated with glossiness. Figs. 13(c) and 13(d) show the results of JPEG encoding with/without non-glossiness information reduction. Although the glossiness of the JPEG images is very similar, the data size of the image with non-glossiness information reduction is approximately 7% smaller.



(a) Image with enhanced glossiness (1,058,454 bytes)

(b) Target blocks for non-glossiness information reduction. Black blocks are not targeted

(c) JPEG image after non-glossiness information reduction (17,132 bytes)

(d) JPEG image without non-glossiness information reduction (18,313 bytes)

Figure 13. Examples of images with/without non-glossiness information reduction.

4. EXPERIMENTS

To verify the effectiveness of the proposed method, we conducted subjective evaluation experiments.

4.1 *Experimental Stimuli*

In the subjective evaluation, observers evaluated the glossiness of objects in three types of images displayed on a monitor, namely, the original, JPEG, and proposed images, using Thurstone's paired comparison method. All possible combinations of the three types of images were subjectively evaluated. The data size of each image processed by the proposed method was adjusted to match the size of the corresponding JPEG image. In our experiment, we used a sharpening filter to process Block A described in Section 3.3 as shown in Figure 14. Here, we used a fixed k value of -0.2.

We prepared five test images with different glossiness as the original images as shown in Figure 15. Two current JPEG encoders, libjpeg and mozjpeg, were used. We generated four types of encoded images with different compression ratios for each test image by changing the quality factors. Therefore, the test images of 40 patterns ($5 \times 2 \times 4$) were used as the experimental stimuli for subjective evaluation. Fig. 15 shows examples of the experimental stimuli for each test image. These images were encoded using libjpeg, and the

k	k	k
k	$1.0\text{-}8k$	k
k	k	k

Figure 14. Sharpening filter used in our experiment.

corresponding image processed by the proposed method has the same data size. We also estimate the PSNR in addition to the data size of each image. The PSNRs of the images obtained using the proposed method are slightly lower than those of the JPEG images. This result indicates that the JPEG encoder yields better image quality than the proposed method in terms of the PSNR.

4.2 *Experimental Procedure*

A monitor (EIZO ColorEdge CG277) with an Adobe RGB color gamut was used to display the experimental stimuli. The maximum luminance and color temperature were set as 270 cd/m^2 and 6500 K, respectively. The monitor was placed in a dark room, and its distance from the observer was 80 cm. The experimental environment is shown in Figure 16.

Observers were shown pairs of experimental stimuli, and they indicated via a two-alternative forced choice based on Thurstone's method which one was perceived as being glossier. As shown in Fig. 15, the stimuli contained artifacts such as blocking, but the observers were instructed to evaluate only the gloss. Observers had unlimited time to respond. The long side of each pair of displayed image sizes covered a viewing angle of 10°. Between evaluations, the stimuli were replaced by a checkered pattern for 1 s to reset the memory and the afterimage of the color and structure information. The experimental stimuli were shown to the observers in a randomly chosen order. Because the experimental stimuli could be combined in three possible ways, each observer evaluated 120 pairs of stimuli. To avoid burdening the observers, we did not flip the pairs of stimuli. Nine men and one woman participated in the experiment. Their average age was 23.7 years, and the standard deviation was 0.95 years. All the observers were confirmed to have normal color vision with correction: oculi uterque 20/20.

4.3 *Experimental Results*

In Figures 17(a) and 17(b), the experimental results for the images encoded using libjpeg and mozjpeg, respectively, are summarized as Z-score data. A higher Z-score indicates higher perceived glossiness. The observers perceived higher

Figure 15. Examples of experimental stimuli. JPEG images were encoded using libjpeg.

glossiness in the images obtained by the proposed method than by those obtained using either JPEG encoder. This result indicates that the proposed method might preserve glossiness. The images obtained by applying the proposed method to test images 2 and 5 were rated as having higher glossiness than the original images. To faithfully encode the glossiness of the original image, it might be necessary to adaptively determine the parameters of each image although these parameters were fixed in this experiment. By contrast, for test image 4, glossiness was lost compared with one of the original images even though the glossiness was better than that in the JPEG image. In an introspection survey, the observers could not perceive a significant difference in the glossy highlights because they focused mainly on the highlight areas with saturated pixels. The degradation caused by encoding another area of the highlight might affect the subjective evaluations.

Table III presents the number of observers who reported that the images processed by the proposed method were glossier than the JPEG images. Because 10 observers participated in the experiment, the average score was 10.00 when all observers reported that the four images of different quality coded by the proposed method were glossier than the

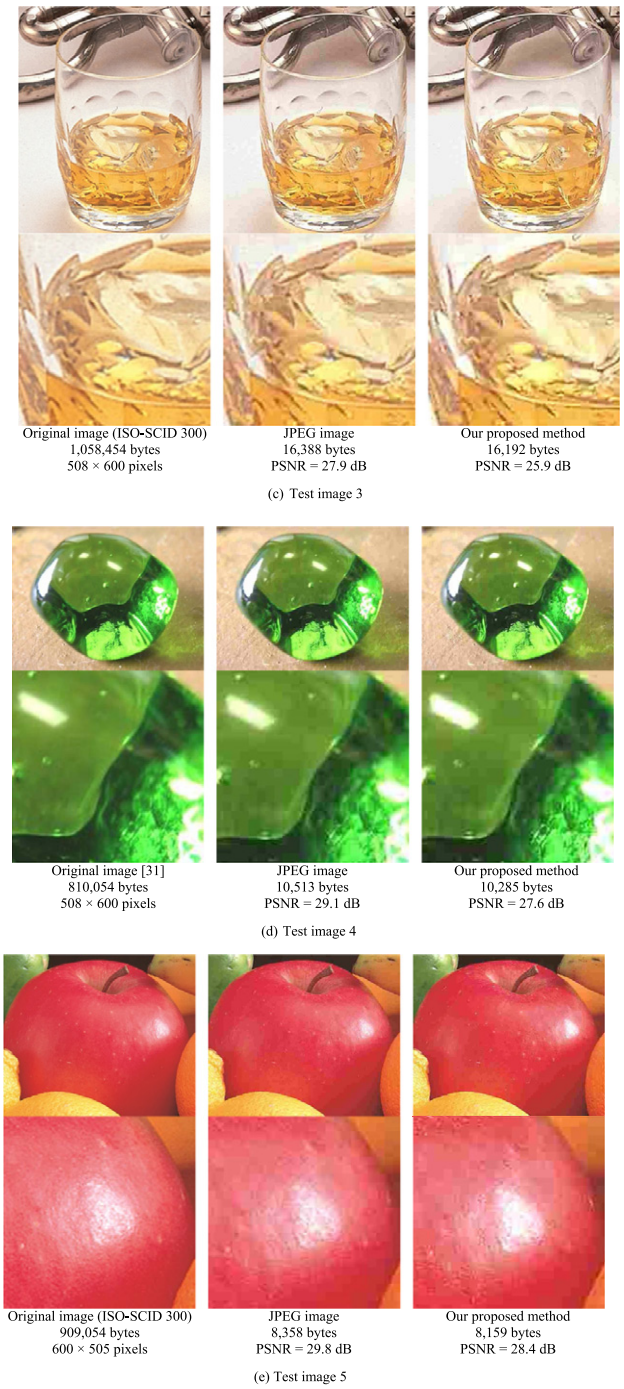


Figure 15. (*continued*)

JPEG images. As shown in the table, for all test images, the average number of observers was higher than 5.00, and the effectiveness of the proposed method was confirmed.

5. CONCLUSIONS

We proposed a new image encoding method to preserve the glossiness of input images. The proposed method is constructed within the framework of JPEG encoding; thus, the images can be restored by the JPEG decoder. It consists of three procedures: block classification, glossiness enhancement, and non-glossiness information reduction. To verify the effectiveness of the proposed method, we conducted a subjective evaluation experiment using paired comparison of images with the same data size that were decoded by the proposed method and two JPEG encoders. For test images with various gloss used in the experiment, the results showed that the glossiness of images obtained by the proposed method was higher than that of the JPEG images. In addition, we confirmed that our method could preserve the glossiness of the original image.

Some test images obtained by the proposed method were rated as having higher glossiness than that of the original images. The proposed algorithm has a few constant parameters. However, the optimal parameter values depend on the input image. To faithfully reproduce the glossiness of the original image, it is necessary to adaptively determine the parameters. Verification with more test images will be a subject of future work. Furthermore, we will also examine an image encoding method for preserving other Shitsukan features.

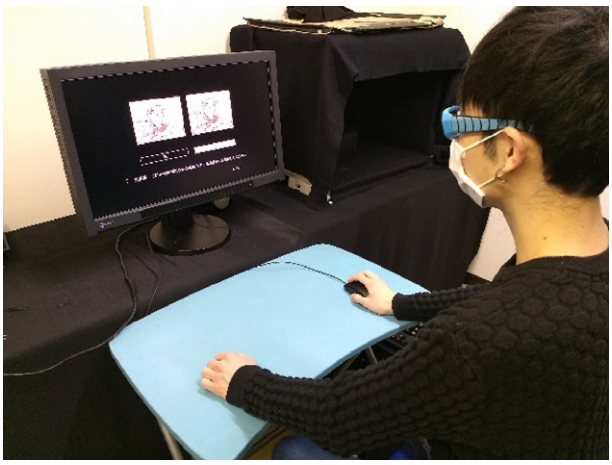


Figure 16. Experimental environment. The experiment was conducted in a dark room.

Table III. Number of observers who reported that the images processed by the proposed method were glossier than the JPEG images.

Test image	libjpeg		mozjpeg	
	Avg.	Std.	Avg.	Std.
1	7.75	0.96	6.75	0.96
2	7.25	1.71	9.00	0.00
3	6.75	0.96	7.00	0.82
4	5.00	1.83	5.50	1.00
5	8.00	0.82	7.25	0.50

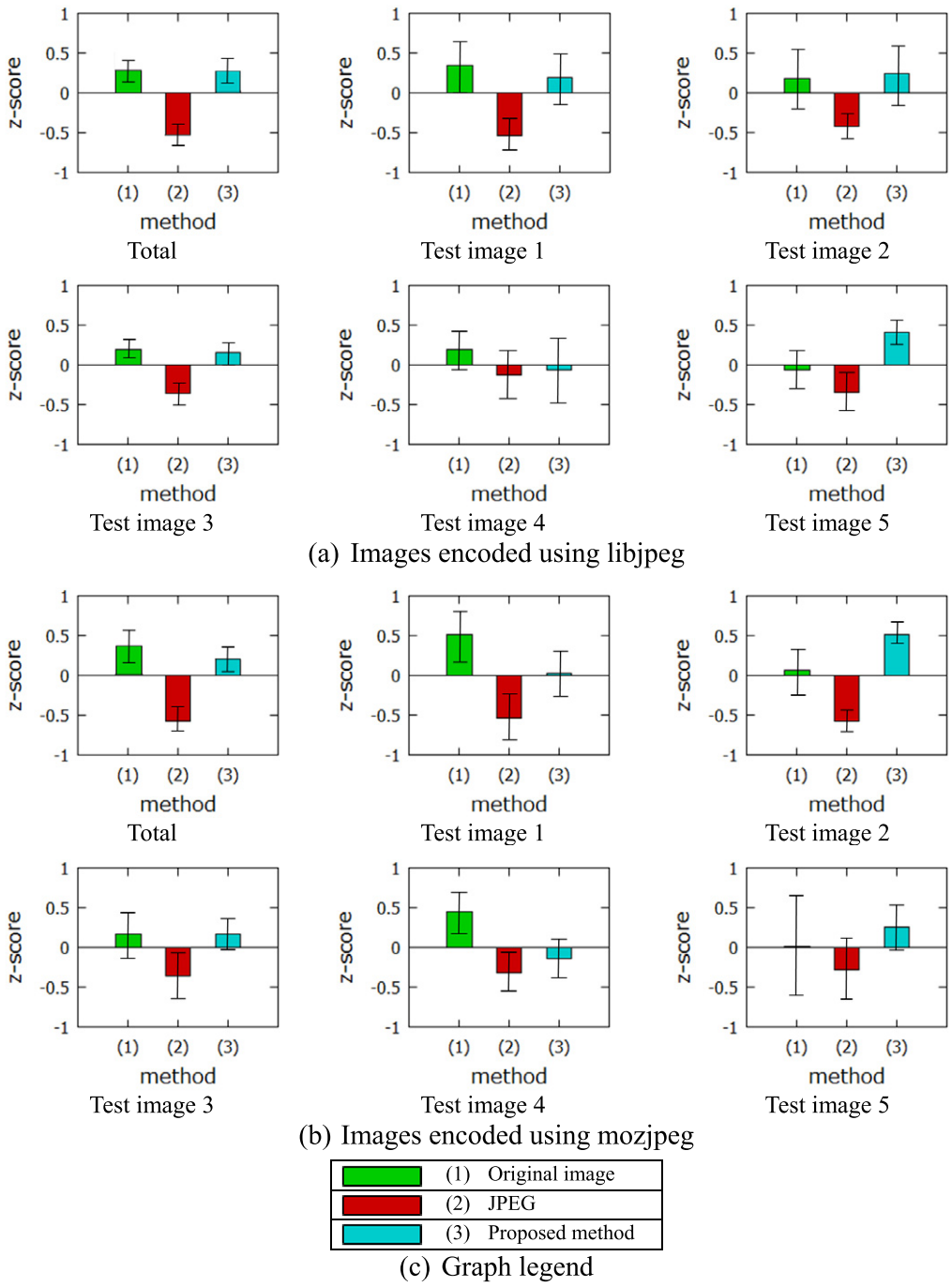


Figure 17. Experimental results (Z-score). Error bars show 95% confidence intervals. (a) Images encoded using libjpeg. (b) Images encoded using mozjpeg. (c) Graph legend.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas (No. 15H05926) from MEXT, Japan.

REFERENCES

[1] ITU-T T.800, "Information technology – JPEG 2000 image coding system: Core coding system", 2002.

[2] "WebP", Retrieved March 8, 2020, from https://developers.google.com/speed/webp/.

[3] ITU-T H.265, "High efficiency video coding", 2013.

[4] F. Bellard,"BPG Image format", Retrieved March 8, 2020, from https://bellard.org/bpg/.

[5] ISO/IEC 23008-12, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 12: Image File Format, 2017.

[6] "Mozilla/mozjpeg", Retrieved March 8, 2020, from https://github.com/mozilla/mozjpeg.

[7] "libjpeg-turbo", Retrieved March 8, 2020, https://libjpeg-turbo.org/.

[8] "Independent JPEG Group", Retrieved March 8, 2020, from http://www.ijg.org/.

[9] J. Alakuijala, R. Obryk, O. Stoliarchuk, Z. Szabadka, L. Vandevenne, and J. Wassenberg, "Guetzli: Perceptually Guided JPEG Encoder," arXiv:1703.04421v1, 2017.

[10] "Google/Butteraugli", Retrieved March 8, 2020, from https://github.com/google/butteraugli.

[11] R. W. Fleming, S. Nishida, and K. R. Gegenfurtner, "Perception of material properties (part I)," Vis. Res. **109**, 123–236 (2015).

[12] R. W. Fleming, S. Nishida, and K. R. Gegenfurtner, "Perception of material properties (part II)," Vis. Res. **115**, 157–302 (2015).

[13] M. Tanaka, R. Arai, and T. Horiuchi, "PuRet: material appearance enhancement considering pupil and retina behaviors," J. Imag. Sci. Technol. **61**, 040401 (2017).

[14] M. Tanaka and T. Horiuchi, "Investigating perceptual qualities of static surface appearance using real materials and displayed images," Vis. Res. **115**, 246–258 (2015).

[15] I. Boyadzheiev, K. Bala, S. Paris, and E. Adelson, "Band-sifting decomposition for image based material editing," ACM Trans. Graph. **34** (2015).

[16] S.W. Wu and A. Gersho, "Rate-constrained picture-adaptive quantization for JPEG baseline coders," *IEEE Int'l. Conf. on Acoustics, Speech, and Signal Processing* (IEEE, Piscataway, NJ, 1993), vol. 5, pp. 389–392.

[17] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," IEEE Trans. Image Process. **3**, 700–704 (1994).

[18] M. Crouse and K. Ramchandran, "JPEG optimization using an entropy-constrained quantization framework," *Proc. DCC '95 Data Compression Conf.* (IEEE, Piscataway, NJ, 1995), pp. 342–351.

[19] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform image coding: entropy-constrained analysis and applications to baseline JPEG," IEEE Trans. Image Process. **6**, 285–297 (1997).

[20] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," IEEE Signal Process. Mag. **15**, 23–50 (1998).

[21] V. Ratnakar and M. Livny, "An efficient algorithm for optimizing DCT quantization," IEEE Trans. Image Process. **9**, 267–270 (2000).

[22] S. Viswajaa, B. V. Kumar, and G. R. Karpagam, "A survey on nature inspired meta-heuristics algorithms in optimizing the quantization table for the JPEG baseline algorithm," Int'l. Adv. Res. J. Sci., Eng. Technol. **2**, 114–123 (2015).

[23] H. Wang and S. Kwong, "Preferential multi-objective genetic algorithm for JPEG quantization table optimization," Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms (Special Issue), 608–613 (2005).

[24] E. Tuba, M. Tuba, D. Simian, and R. Jovanovic, "JPEG Quantization Table Optimization by Guided Fireworks Algorithm," *Combinatorial Image Analysis: 18th Int'l. Workshop, IWCIA 2017* (IEEE, Piscataway, NJ, 2017) pp. 294–307.

[25] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *NIP'14: Proc. 27th Int'l. Conf. on Neural Information Processing Systems*, Vol. 2 (ACM, NY, 2014) pp. 2672–2680.

[26] O. Rippel and L. Bourdev, "Real-time adaptive image compression," *ICML'17: Proc. 34th Int'l. Conf. on Machine Learning* (ACM, NY, 2017), vol. 70, pp. 2922–2930.

[27] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative adversarial networks for extreme learned image compression," *Proc. IEEE Int'l. Conf. on Computer Vision* (IEEE, Piscataway, NJ, 2019), pp. 221–231.

[28] S. Kudo, S. Orihashi, R. Tanida, and A. Shimizu, "GAN-based image compression using mutual information maximizing regularization," *Picture Coding Symposium* (IEEE, Piscataway, NJ, 2019), pp. 1–5.

[29] "GAHAG", Retrieved March 8, 2020, from http://gahag.net/.

[30] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Sys. Man. Cyber. **9**, 62–66 (1979).

[31] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?," J. Vis. **9**, 784 (2009).