

A Rate Control Scheme Compatible with a Legacy JPEG Decoder to Reduce Data Size Keeping Subjective Quality for Video Advertising on Mobile Devices

Risako Aoki^{1*}, Shun Aoki¹, Takuro Oki¹, Shingo Kobayashi¹, Ryusuke Miyamoto², Makoto Iguchi³

¹ Department of Computer Science, Graduate School of Science and Technology, Meiji University, Kawasaki, Japan.

² Department of Computer Science, School of Science and Technology, Meiji University, Kawasaki, Japan.

³ CMerTV Corporation, Japan.

* Corresponding author. Tel.: +81-44-934-7442; email: aori@cs.meiji.ac.jp

Manuscript submitted June 15, 2019; accepted August 18, 2019.

doi: 10.17706/ijcee.2019.11.3.133-142

Abstract: Video advertising has become important content for web advertising, especially in mobile devices. However, most mobile devices prohibit an auto-play function for video sequences on mobile web browsers. As a result, video coding standards such as MPEG4 and H.265 cannot be used for video advertising on mobile devices. Therefore, CSS image sprites that can show successive images are widely used to play video sequences for web advertising on mobile devices. This paper proposes a novel rate control scheme to improve the coding efficiency, based on CSS image sprites composed of successive still images coded using the JPEG standard. The proposed scheme controls data allocation to blocks according to the importance of the image regions during rate-distortion optimization of the JPEG encoding process. Subjective evaluation based on the Subjective Assessment Methodology for Video Quality (SAMVIQ) method using datasets created using DAVIS2017 and CG showed that the proposed scheme can reduce the total size of coded data while preserving the subjective image quality.

Key words: Subjective video quality, rate control, video advertising.

1. Introduction

Recently, mobile devices for consumer such as smart phones have become widely spread in the world. Specifically, about 84.0% of the people in Japan own mobile devices according to a government report in 2018 [1]. Currently, many websites are designed for mobile devices and several kinds of advertisements are embedded in such websites according to the spread of mobile devices. In addition to video streaming sites such as YouTube [2], web sites of which the primary content is text documents and static images have also come to use video advertisements because movies can contain information that is effective with only short viewing time.

Video advertisements should be automatically triggered when users access websites via web browsers. However, video sequences encoded by video coding standards using inter-frame compression, such as MPEG [3] and H.264 [4], cannot be auto-played in many embedded web browsers. Hence, video advertisement distributors adopted CSS image sprites to play video advertisements, where video sequences consist of successive still images. Such still images are encoded using the JPEG coding standard to reduce

the total data size for the video because of backward compatibility: old devices may not decode still images coded by recent coding standards [5]-[7].

In this case an inter-frame data compression, that drastically reduces the amount of data but preserves the image quality, cannot be adopted because a video sequence is simply composed of successive still images. As a result, the data size of video advertisements composed of successive still images becomes large and consumes greater bandwidth. This bandwidth increase becomes a significant problem for video advertisement distributors because it increases their payments to network service providers.

To reduce the data size of video advertisements composed of successive still images, this paper proposes a rate control scheme that effectively reduces the size of coded data while retaining their subjective quality. The aim of the proposed scheme is to reduce the allocation of data according to regions that are not subjectively important in video sequences. The proposed scheme controls the bit allocation of still images composing a video sequence encoded by the JPEG standard. This is achieved by using the rate-distortion optimization that occurs in the encoding process according to the subjective importance of image regions; smaller data is allocated to subjectively unimportant regions.

To determine important and unimportant regions in an input image sequence, the following characteristics peculiar to video advertising are used:

- text regions used as captions,
- regions according to objects to be highlighted,
- image blur used for emphasizing a subject in a video, and
- motion of target regions.

To evaluate the coding performance of the proposed scheme, subjective image quality is measured by several observers.

2. JPEG Encoding

This section explains the encoding scheme of the JPEG standard.

2.1. Overview of the Encoding Flow

Fig. 1 shows an overview of the encoding process. In the JPEG encoding scheme, input images are first divided into small blocks composed of 8×8 pixels. Next, a frequency domain image is generated by applying the Discrete Cosine Transform (DCT) [8] to divided blocks. Once a frequency domain image is obtained, quantization is applied to reduce the data on high-frequency components that are not perceptible by the human visual system. Finally, entropy coding is applied to compress the obtained DCT coefficients effectively using information theory

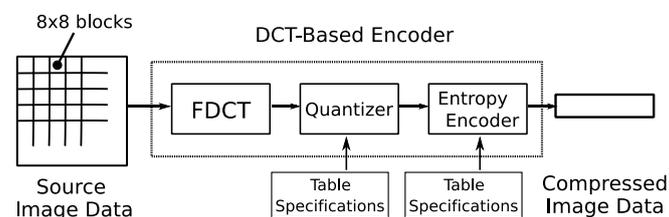


Fig. 1. JPEG encoding flow.

2.2. Progressive Coding (Fig. 2)

Generally, the coded data size becomes smaller as the entropy coding of the zero components of the AC coefficients of the DCT continues. To increase the probability of continuous zero components emerging, the well-known Zigzag scan is adopted in the JPEG standard. In addition to the Zigzag scan, the JPEG standard

has another option for encoding DCT coefficients: the progressive scan. The progressive scan lines up DCT coefficients from lower to higher frequency components in an input image. The MozJPEG [9] achieved more efficient coding than existing schemes using the progressive scan and novel rate-distortion optimization.

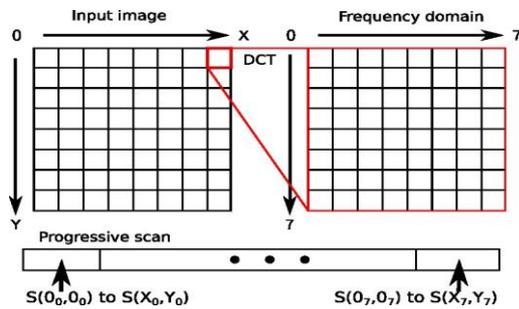


Fig. 2. Progressive coding.

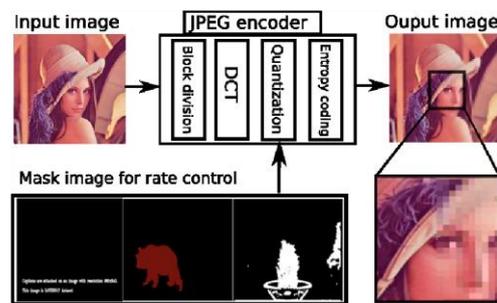


Fig. 3. JPEG encoding flow with proposed rate control.

3. Rate Control Scheme According To Importance of Target Regions

The proposed scheme reduces the data allocated to a region according to the importance of the region. The minimum size of regions for data reduction is 8×8 pixels, which is determined by the block size of the DCT adopted in the JPEG standard. To reduce the data size appropriately, an additional constraint is applied to the process for rate-distortion optimization [10], which determines the optimal amount of data allocated to the region considering distortion of the region. The rest of this section describes an overview of the JPEG coding scheme with the proposed rate control and how to determine the importance of regions.

3.1. Overview of the JPEG Coding Scheme with the Proposed Rate Control

Fig. 3 shows an overview of the JPEG coding scheme with the proposed rate control. Here, an input image is divided into rectangular regions with size of 8×8 pixels. After the division, discrete cosine transforms (DCT) are applied to convert the input image into the frequency domain. These operations are performed in the same way as the normal JPEG coding flow. Generally, quantization is applied after DCT to reduce the data size appropriately. The proposed scheme changes the cost value at rate distortion optimization according to the importance of image regions. For this purpose, a mask image representing the importance of image regions is used. After the rate control during quantization, the obtained coefficients are re-ordered in a progressive manner, and after that, entropy coding is applied to generate a coded bitstream.

3.2. Rate Control According to the Importance of Image Regions

In this paper, the following four kinds of rate-control schemes are proposed:

- 1) text regions used as captions,
- 2) regions according to objects to be highlighted,
- 3) image blur used for emphasizing a subject in a video, and
- 4) motion of target regions.

3.2.1. Rate control considering text regions used as captions

Video advertisements often include text regions that are used for explanation of target items or captions for silent environment. However, the coding quality of text regions in videos tends to be worse if no special coding and rate control is adopted. To solve this problem, the proposed scheme tries to allocate much more data to text regions than in the normal JPEG coding scheme. The amount of data is controlled by mask images according to each frame of an input image sequence, as explained above.

3.2.2. Rate control considering the importance of objects

Usually, an advertisement video sequence consists of important objects and insignificant regions such as the background. In this case, the image quality corresponding to important objects should be better than the image quality in other regions. For this purpose, the proposed scheme allocates more data to important objects represented by mask images than to other regions.

3.2.3. Rate control considering image blur

Generally, video sequences do not include image blur to show the details of scenes from near-to-far. However, sometimes, intentional image blur is included to direct the attention of viewers to certain regions; this technique is often used for artistic photographs. In this case, the data allocated to them may be reduced without degradation because the image quality of blurred regions is not good. The proposed scheme reduces the data size allocated to blurred regions to improve the total efficiency of video coding.

3.2.4. Rate control considering image motion

An existing scheme showed effective bit allocation according to the motions of regions in a video sequence. It could reduce the data size without degrading the subjective quality of a video sequence by decreasing the data size for regions where the speed was quite high [11]. These results were obtained considering that the human visual system is more sensitive to slower motions than high speed ones. The proposed scheme reduces the data allocated to regions in which the speed is quite high in the same way as in the previous scheme. The motion of target regions is evaluated by the different locations of objects in two successive frames.

4. Evaluation

To validate the proposed rate control scheme, a subjective quality based on the SAMVIQ method [14] of image sequences was evaluated using two different display sizes. Peak signal-to-noise ratio (PSNR) [12] and the file size were measured, but the widely used structural similarity (SSIM) [13] display was not adopted in this evaluation.

4.1. Subjective Evaluation

For this paper, the proposed scheme was evaluated using subjective image quality in addition to PSNR and SSIM; the SAMVIQ method [14] was adopted for subjective evaluation. The SAMVIQ method allows observers to freely access and watch several videos using an application for evaluation. Therefore, it is expected to provide stable results for quality comparison. Observers could change video sequences to be evaluated by pushing the numerical buttons below the displayed image. Observers rated the subjective quality of the video sequences and changed them as they preferred. The image quality was rated using a slide bar on the left side. Here, 50 meant the same quality and lower and higher values corresponded to worse and better quality, respectively. The lower and upper limits were 0 and 100, respectively.

As observers for subjective evaluation, 14 people aged between 20 and 27 years attended the experiment. The subjective experiments were conducted using a MacBook Pro 13, with a 13 in display, and an iPhone 8 Plus to investigate the effect of the proposed scheme on mobile devices.

4.2. Evaluation Results

This subsection shows results of subjective evaluation using the SAMVIQ method.

4.2.1. Rate control considering text regions used as captions

To validate the rate control scheme for text regions, several video sequences with text regions were generated using the DAVIS2017 dataset. The left image of Fig. 4 shows an example of the dataset with text

regions. A mask image (right image in Fig. 4) was provided for rate control.



Fig. 4. An example image and a mask image with text regions used as captions.

Table 1 shows the image quality in the experiment; higher values correspond to better quality. In this evaluation, overall quality of images was changed, in addition to the rate-control, during rate-distortion optimization.

Table 1. Image Quality Controlled by the Proposed Rate Control Scheme

	Image Quality of target regions	Image Quality of other regions
A	4	4
B	4	3
C	3	3
D	3	2
E	3	1

Fig. 5 shows the subjective evaluation results. These results confirmed that image degradation by reduction of the data size was lower on the laptop PC than on the mobile device. If the overall quality factor was smaller, the difference in the subjective image quality was greater. In this case, Setting D showed good results; the file size was reduced but the image quality was still good.

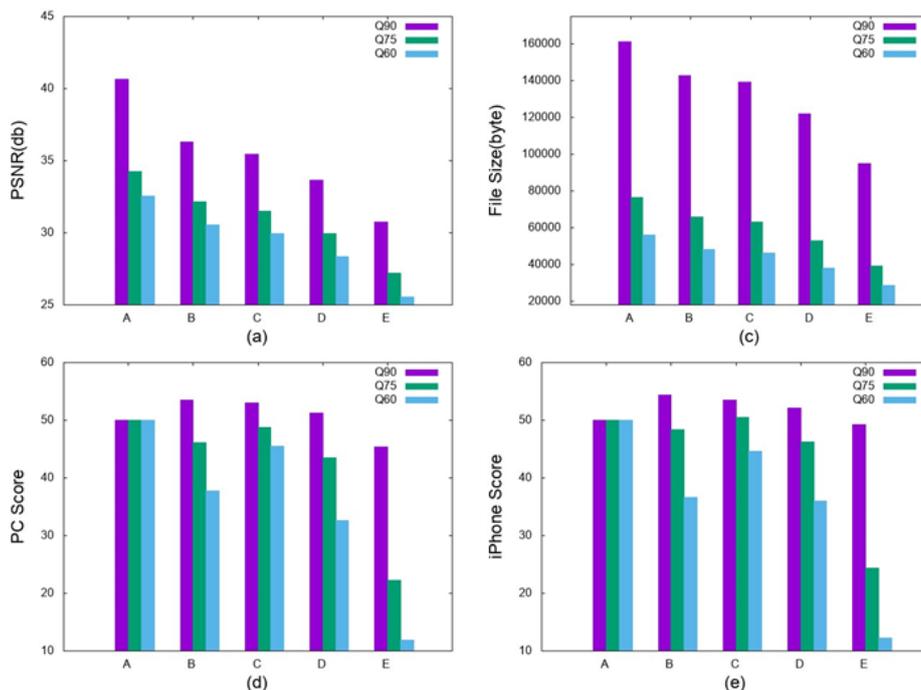


Fig. 5. Results of subjective evaluation for rate control considering text regions.

4.2.2. Rate control considering the importance of objects

For this experiment, mask images (right side in Fig. 6) were used to control the data allocation by the proposed scheme; more data for important objects and less data for other regions. Image quality was changed in the same way as the rate control considering text regions.

Fig. 7 shows the results of subjective evaluation. By these results, it was confirmed that image degradation by reduction of the data size became clearer on the laptop PC than on the mobile device. In addition, the subjective quality of the image sequences with more data for important objects became better than image sequences with a moderate bitrate for all regions. These results show that image quality of important objects is significant for subjective evaluation.



Fig. 6. An example of image sequences with mask images used for evaluation.

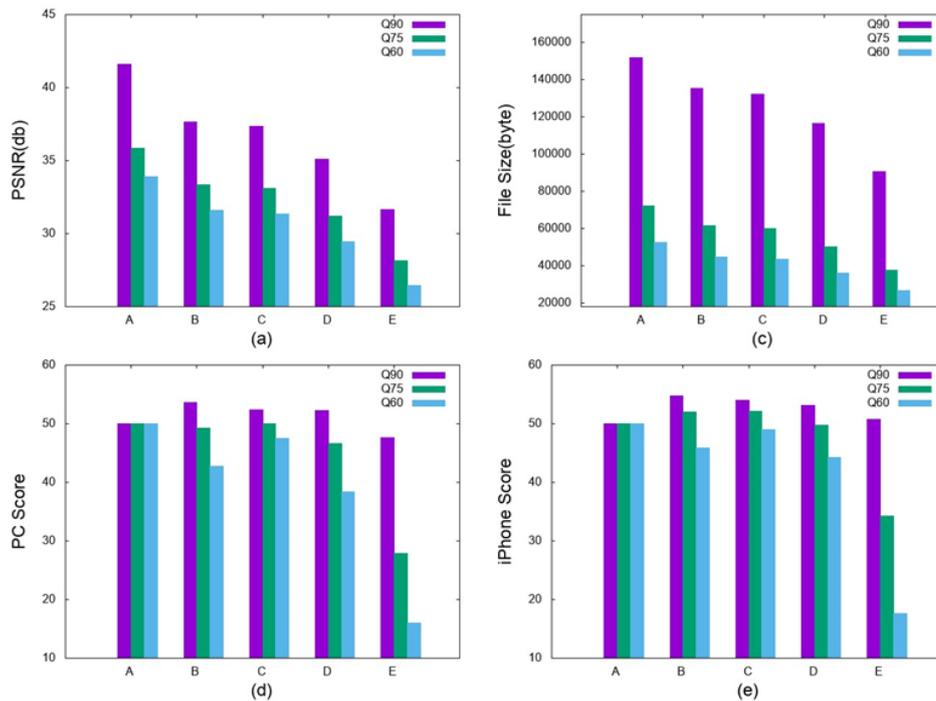


Fig. 7. Results of subjective evaluation for rate control considering important objects.

4.2.3. Rate control considering image blur

For this evaluation, image sequences with out of focus blur were generated by computer graphics using the Unreal Engine 4 [15] (Fig. 8). The aim of the proposed scheme was to reduce the data amount allocated to blurred regions. The image quality was changed in the same way as the rate control considering text regions.

Fig. 9 shows the results of subjective evaluation. These results show that the subjective quality was good

when the overall image quality was high. On the other hand, the subjective quality became much worse if the overall image quality was lower. In addition, the subjective quality of Setting C was better than of Setting B, even though the file size of C was smaller than B. As a result, rate control considering image blur did not work very well.

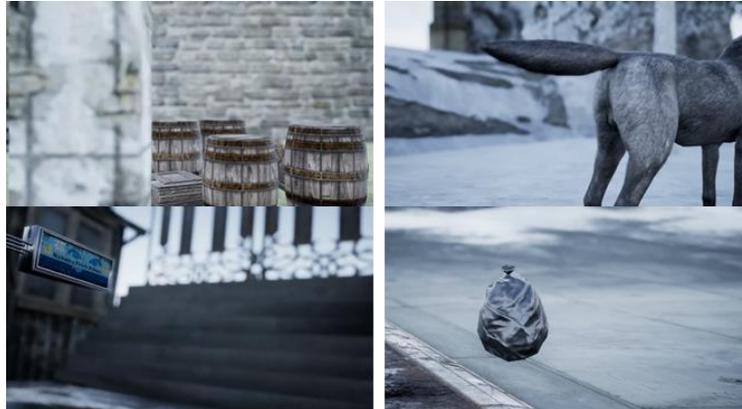


Fig. 8. Examples of the dataset used for evaluation of rate control considering image blur.

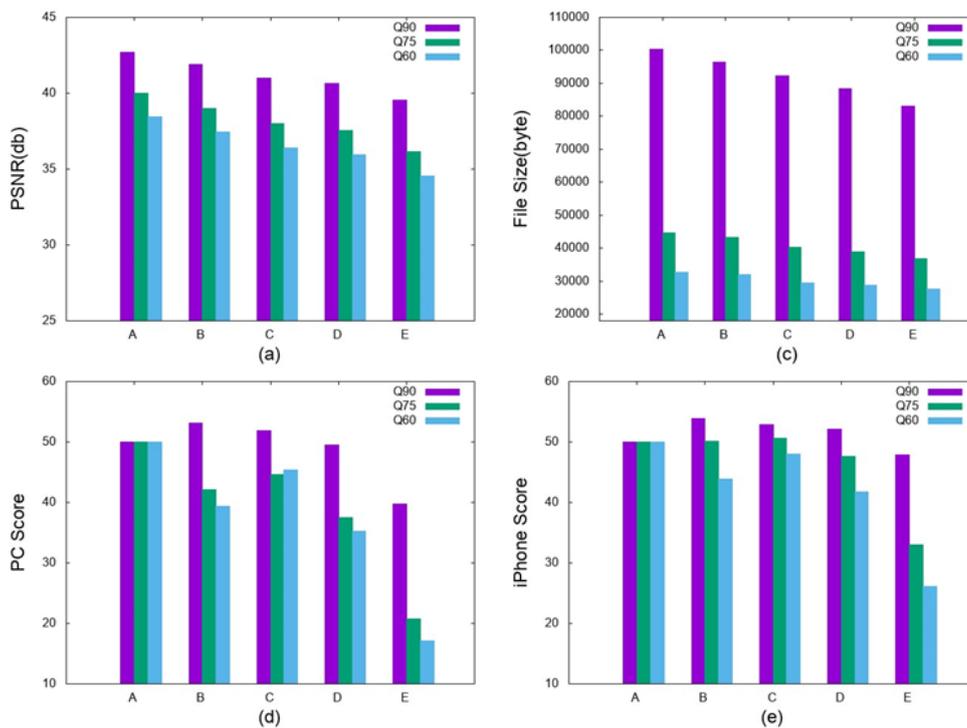


Fig. 9. Results of subjective evaluation for rate control considering image blur.

4.2.4. Rate control considering image motion

For this evaluation, the same image sequences with object masks from the DAVIS2017 dataset that were used for evaluation of rate control considering the text regions were used. However, different data allocation was applied during the encoding process; lower bitrates were used for regions with faster motions and higher bitrates were used for regions with slower motions according to Table 1.

Fig. 10 shows the results of subjective evaluation. By these results, Setting B could reduce the file size without subjective degradation of image quality. Setting D could reduce the file size drastically, but the

subjective quality also became worse. However, the image quality using Setting D was not as bad on the mobile device.

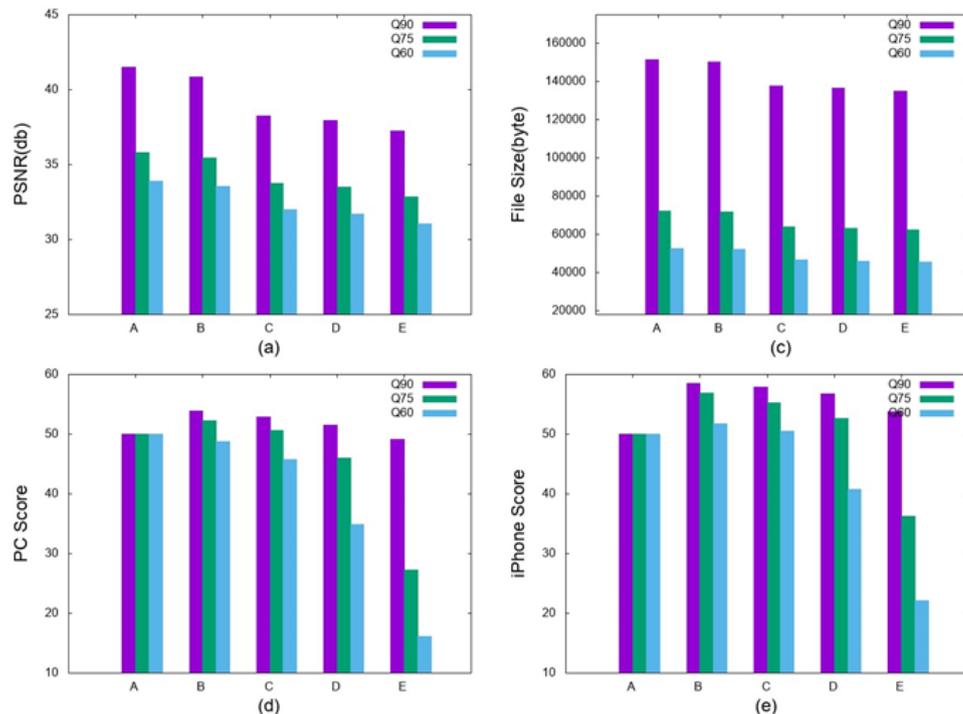


Fig. 10. Results of subjective evaluation for rate control considering image blur.

5. Conclusions

In this paper, a novel rate control scheme was proposed that is compatible with a legacy JPEG decoder required for video advertisement on mobile devices. The proposed scheme controlled the amount of data allocated to blocks composed of 8×8 pixels according to the importance of regions during the rate-distortion optimization process of JPEG encoding. Four types of schemes were proposed in this paper; rate control was performed considering text regions, important objects, image blur, and image motion.

To validate the proposed scheme, subjective evaluation based on the SAMVIQ method was performed. Evaluation of rate control considering text regions, important objects, and image motion was done using the DAVIS2017-based dataset. The CG-based dataset created using Unreal Engine 4 was used for rate control considering image blur. The experimental results showed that the proposed scheme worked well in many cases; however, rate control considering image blur worked in only limited conditions. In particular, the proposed schemes reduced file size without degradation of the subjective image quality on a mobile device. In the future, the authors will try to adopt several optimization schemes [16,17] to the proposed scheme.

References

- [1] Ministry of Internal Affairs and Communications. (2018). Results of communication usage investigation. Retrieved from http://www.soumu.go.jp/johotsusintokei/statistics/data/180525_1.pdf
- [2] YouTube. Retrieved from <http://www.youtube.com>
- [3] Sikora, T. (1997, October). MPEG digital video-coding standards. *IEEE Signal Processing Magazine*, 14, 82–100.
- [4] Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003, July). Overview of the h.264/avc video coding standard. *IEEE Trans. Cir. and Sys. for Video Technol.*, 13(7), 560–576.

- [5] JPEG2000. Retrieved from <https://jpeg.org/jpeg2000/index.html>
- [6] JPEG XR. Retrieved from <https://jpeg.org/jpegxr/index.html>
- [7] WebP. Retrieved from <https://developers.google.com/speed/webp/>
- [8] Gilbert Strang. (1999, March). The Discrete Cosine Transform. *SIAM Rev.*, 41(1), 135–147.
- [9] Mozilla JPEG encoder Project. Retrieved from <https://github.com/mozilla/mozjpeg>
- [10] Ratnakar, M., & Livny, V. (1995). RD-OPT: An efficient algorithm for optimizing DCT quantization tables. *Proceedings of Data Compression Conference* (pp. 332–341).
- [11] Miyamoto, R., Sugita, H., Hayashi, Y., Tsutsui, H., Masuzaki, T., Onoye, T., & Nakamura, Y. (2005, May). High quality motion jpeg2000 coding scheme based on the human visual system. *Proceedings of 2005 IEEE International Symposium on Circuits and Systems: Vol. 3.* (pp. 2096–2099).
- [12] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- [13] Hor, A., & Ziou, D. (2010, August). Image quality metrics: Psnr vs. ssim. *Proceedings of ICPR* (pp. 2366–2369).
- [14] Kozamernik, F., Steinmann, V., Sunna, P., & Wyckens, E. (2005, April). Samviq— A new ebu methodology for video quality evaluations in multimedia. *SMPTE Motion Imaging Journal*, 114(4), 152–160.
- [15] Unreal Engine 4. Retrieved from <https://www.unrealengine.com/ja/what-is-unreal-engine-4>
- [16] Alakuijala, J., Obryk, R., Szabadka, Z., & Wassenberg, J. (2017). Users prefer Guetzli JPEG over same-sized libjpeg. CoRR, abs/1703.04416.
- [17] Richter, T. (2016). JPEG on STEROIDS: Common optimization techniques for JPEG image compression. *Proceedings of IEEE International Conference on Image Processing* (pp. 61–65).



Risako Aoki received the B.S. degree and the M.S. degree in computer science from Meiji University, Tokyo, Japan, in 2015 and 2017, respectively. She is currently a Ph.D candidate in the Graduate School of Science and Technology, Meiji University. She is also a research associate in the School of Science and Technology, Meiji University.



Shun Aoki received the B.S. degree and the M.S. degree in computer science from Meiji University, Tokyo, Japan, in 2017 and 2019, respectively. He is currently working at CRI Middleware Co., Ltd.



Takuro Oki received the B.S. degree and the M.S. degree in computer science from Meiji University, Tokyo, Japan, in 2015 and 2017, respectively. He is currently a Ph.D candidate in the Graduate School of Science and Technology, Meiji University. He is also a research associate in the School of Science and Technology, Meiji University.



Shingo Kobayashi received the B.S. degree and the M.S. degree in computer science from Meiji University, Tokyo, Japan, in 2015 and 2017, respectively. He is currently a Ph.D candidate in the Graduate School of Science and Technology, Meiji University. He is also a research associate in the School of Science and Technology, Meiji University.



Ryusuke Miyamoto received the B.E. degree in industrial chemistry, the M.Sc degree in communications and computer engineering, and the Ph.D. degree in communications and computer engineering from Kyoto University, Kyoto, Japan, in 1998, 2001, and 2007, respectively. He is currently a senior assistant professor in the School of Science and Technology, Meiji University. He is a member of IEEE, IEICE, IIEEJ, IPSJ, and SICE.



Makoto Iguchi received the B.E. degree in electrical and electronic engineering and the M.Sc degree in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2000, 2002 respectively. He is currently the director / CTO at CMerTV Inc.