A Structure of JPEG XT Encoder considering Effect of Quantization Error

Osamu WATANABE†
Takushoku University
Dept. of Electronics & Computer Systems
†Email: owatanab@es.takushoku-u.ac.jp

Ryo SUZUKI
Tokyo Metropolitan University
Faculty of Info. and Commun. Systems

Hitoshi KIYA
Tokyo Metropolitan University
Faculty of Info. and Commun. Systems
Email: kiya@tmu.ac.jp

***Abstract*—The coding performance of the normative encoder of the JPEG XT in profile is analyzed and the problem on the encoder is summarized in this paper. It is pointed out that there is a mismatch in the handling of the quantization error between the normative encoder and the standard decoder. To avoid this problem, an improved structure has been proposed with consideration of the mismatch. The experimental results for the quality of the reconstructed HDR images demonstrate that the proposed structure may give a better coding performance than the normative encoder.**

I. Introduction

A high dynamic range (HDR) image, which has over eight bits as its bit-depth for each color channel, is now becoming popular in digital photography applications such as digital still cameras, digital video cameras, and displays. Meanwhile, some methods to compress HDR images have been reported [1]–[3]. Since the legacy JPEG [4], which is known as the most widely used digital photography format, only supports eight bits per color channel, the legacy JPEG cannot be used to compress HDR images. Even though newer standards such as JPEG XR [5] and JPEG 2000 [6] can be applied for the HDR image coding, they have not been widely used in the image applications due to the lack of backward compatibility with the legacy JPEG.

To resolve this problem, a new standard referred to as JPEG XT [7], [8] is primarily designed to provide coded data containing high dynamic range and wide color gamut content while simultaneously providing eight-bits-per-pixel low dynamic range (LDR) images. JPEG XT coded data contains two layers to reconstruct HDR images: the base-layer for tone-mapped LDR version of an HDR image and an enhancement layer, which is generated by the base-layer and the original HDR image. The legacy JPEG decoder can understand the base-layer of JPEG XT coded data because it has the same codestream syntax as the legacy JPEG. The concept of this two-layer coding is shown in Fig.1.

The JPEG XT standard currently consists of nine parts. Part 1 specifies the base technology such as the core JPEG , which is widely used in image coding applications. Part 2 specifies a backward compatible extension of JPEG towards HDR images using the legacy JPEG encoding for its metadata. Part 3 defines an extensible box-based file format for all subsequent JPEG's standards and future extensions of JPEG will be based on it. This box-based format is compatible with JFIF, and thus can be read by all existing JFIF implementations. Part 4 defines the conformance testing, and Part 5 provides the reference software of the JPEG XT. Part 6 defines extensions of the JPEG standard for backward compatible coding of integer samples between 9 and 16 bits per color channel. Part 7 specifies the coding of HDR images with floating point samples. This is a super-set of both Parts 2 and 3 and offers additional coding tools addressing needs of low-complexity and/or hardware implementations. Part 8 defines lossless coding mechanisms for integer and floating point samples. This is an extension of Parts 6 and 7. Part 9 allows the coding of transparency information. It defines the lossy and lossless representation of alpha channels.

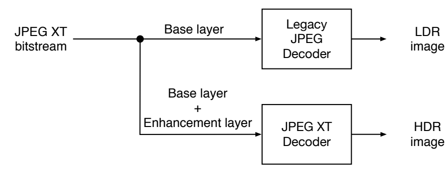


Fig. 1. Concept of backward compatibility to legacy JPEG.

Although JPEG XT has three different profiles (A, B and C), only profile A is discussed in this paper. Several works on the performance evaluation of JPEG XT images have provided the results on the objective and subjective quality measurements [9], [10]. It was reported that the performance of Profile A exhibits earlier saturation on objective image quality vs. bit-rate. This means it does not work properly with a certain coding condition. Although those reports have shown the coding performance of three profiles of JPEG XT, no analysis on this problem have been provided.

In this paper, the problem in the coding performance of Profile A is analyzed and an improved structure of the JPEG XT encoder is proposed. First, basic details of the JPEG XT are summarized, and the mechanism of the early saturation on the image quality in Profile A is analyzed. Then, the proposed structure of the Profile A encoder with consideration of quantization error is described. Finally, the experimental results are provided, and positive improvements in the coding performance of Profile A are demonstrated.

II. JPEG XT codec (Profile A)

A. Encoder

Figure 2 shows a block diagram of the normative JPEG XT encoder of Profile A. The input HDR image is first tone-mapped from the original HDR image RGB_{HDR} to the LDR image RGB_{LDR}, which is coded by the legacy JPEG encoder as a base layer. At the end of encoding process, the

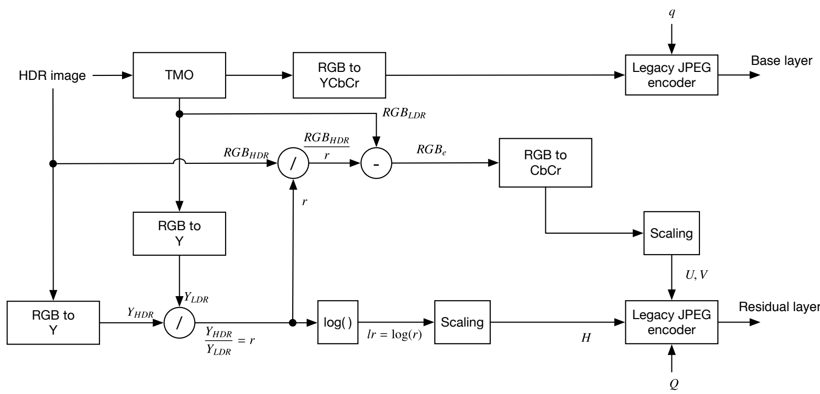


Fig. 2. Block diagram of normative JPEG XT encoder (Profile A). "TMO" is an acronym of Tone Mapping Operator.

codestreams encoded by the legacy JPEG encoder of the base layer and enhancement layer are composed into the JPEG XT codestream.

Let H denotes the luminance component to be coded as the residual data, and Y_{HDR} and Y_{LDR} denote the luminance of RGB_{HDR} and RGB_{LDR}. The ratio of HDR to LDR r:

$$r = \frac{Y_{HDR}}{Y_{LDR}} \tag{1}$$

is used to obtain H, which is given by

$$H = \frac{lr - \min(lr)}{\max(lr) - \min(lr)} \cdot 255, lr = \log(r). \tag{2}$$

Let U and V denote the chroma components to be coded as the residual data. U and V are obtained by the following equations.

$$U = \frac{Cb - \min(Cb)}{\max(Cb) - \min(Cb)} \cdot 255 \tag{3}$$

$$V = \frac{Cr - \min(Cr)}{\max(Cr) - \min(Cr)} \cdot 255, \tag{4}$$

where Cb and Cr are generated by from RGB to CbCr conversion of the subtracted RGB image given by

$$RGB_e = \frac{RGB_{HDR}}{r} - RGB_{LDR}. \tag{5}$$

The three components (H, U, and V) are encoded as a YCbCr image by the legacy JPEG encoder and the values of $\max(lr), \min(lr), \max(Cb), \min(Cb), \max(Cr)$, and $\min(Cr)$ are recorded as the header information. The encoded residual data is referred to as a residual layer. The phrase "residual layer" is used to describe an enhancement layer in the following part of this paper. A residual layer is multiplexed into a base layer using the APP_{11} marker of the standard JPEG file format.

There are two parameters for the quantization of a tone-mapped-HDR image and the residual data. For tone-mapped HDR images, parameter q controls the quality of decoded tone-mapped HDR images. That is, this q is exactly equal to the quality factor of the legacy JPEG. For the residual data, parameter Q is able to control the quality of decoded HDR images. The Q is transformed into a constant in a certain manner, and the constant is multiplied by the quantization matrix. For both q and Q, the larger the value, the better the quality of decoded images, and the range of values is from 1 to 100.

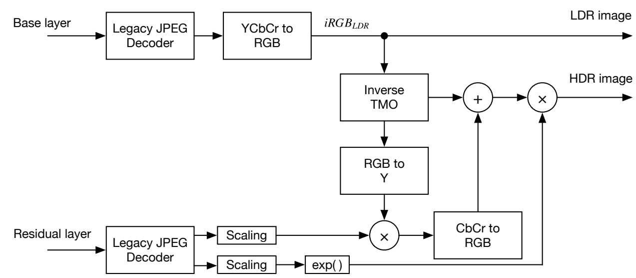


Fig. 3. Block diagram of the standard compliant JPEG XT decoder (Profile A).

B. Decoder

The standard compliant decoder of Profile A is illustrated in Fig. 3. First, a JPEG XT codestream is parsed, and then a base layer and a residual layer are obtained. A base layer is decoded into a tone-mapped HDR image. Base layers can be decoded by the legacy JPEG decoder. A decoded tone-mapped HDR image is inversely tone-mapped for the HDR reconstruction described as follows.

A luminance component and chroma components of a decoded residual layer are used to reconstruct the original HDR image with the inversely tone-mapped image.

III. Proposed structure

A. Problem on normative encoder

Let us consider the relationship between the encoding process in the normative encoder and the decoding process in the standard compliant decoder.

In the normative encoder, chroma components U and V for the residual coding are generated by Eq. 5. RGB_e in Eq. 5 does not include any quantization error because RGB_{LDR} does not either. However, at the standard decoder depicted in Fig. 3, the reconstruction process uses the $iRGB_{LDR}$, which is given by

$$iRGB_{LDR} = RGB_{LDR} + Err(q), \tag{6}$$

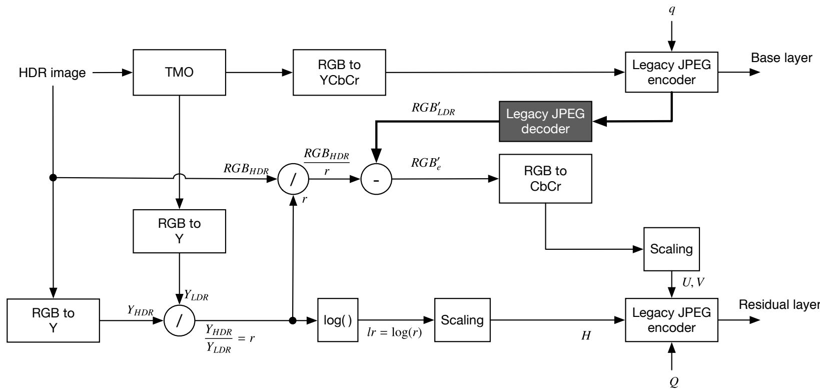


Fig. 4. Block diagram of proposed structure: the dark shaded block is newly introduced.

where $Err(q)$ is the quantization error generated at the normative encoder.

This $Err(q)$ obviously affects the reconstruction of not only the LDR image but also the HDR image because the power of $Err(q)$ is relatively larger than that of the quantization error with the residual data and the $Err(q)$ propagates through the reconstruction process of the HDR image. In other words, there is a mismatch of the quantization error between the normative encoder and the standard decoder.

Due to this property, the quality of the reconstructed HDR will be degraded and cannot be controlled by quality parameters q and Q.

B. Proposed structure considering quantization error

To avoid the problem described above, a new structure of the encoder is proposed. The JPEG XT standard just specifies the procedure of the decoding process shown in Fig. 3. The encoder may take any structure that can provide the codestream decodable with the decoder specified in the standard. In other words, users can change the structure of the normative encoder shown in Fig. 2 to fit their demands.

The block diagram of the proposed structure is illustrated in Fig. 4. In the proposed structure, the legacy JPEG decoder shown by the dark shaded block is newly introduced and the subtraction operation is modified. RGB'_e, which is converted into chroma components U and V, is generated by

$$RGB'_e = \frac{RGB_{HDR}}{r} - RGB'_{LDR}, \quad (7)$$

where RGB'_{LDR} is the output of the dark shaded legacy JPEG decoder. Thus, the U and V, which are the chroma components to be encoded into the residual data, are generated using RGB'_{LDR}, which includes the quantization error $Err(q)$. This means that the mismatch of the quantization error between the encoder and the standard decoder described in the previous subsection is removed and the quality of the decoded image will be controllable by parameters q and Q.

IV. EXPERIMENTAL RESULTS

To confirm the effectiveness of the proposed structure, the experimental results on the reconstructed HDR image quality in terms of PSNR and FSIM [11] are provided. The image quality is evaluated from two aspects: the quality of the reconstructed HDR images in tone-mapped domain and that in the HDR domain. The tone-mapping operator used for the experiments is the reinhard-global [12]. For the experiments, an HDR image 'Montreal_float.hdr' (RGBE, 32 bits-floating point for each color component) is encoded by the reference software available from the JPEG committee [13]. Note that the modified source code is used for the structure of the proposed encoder.

Figure 5 shows the PSNR results for the quality evaluation of the ton-mapped domain. Figures 5(a) and (b) are the results of the fixed q and Q, respectively. The quality of the reconstructed image with the proposed structure (Red bar) monotonically increases with Q , while that with the normative encoder (Blue bar) in Fig. 5(a) shows early saturation. For the fixed Q, it is confirmed that the quality with the proposed structure is better than that with the normative encoder at lower quality from Fig. 5(b). For the HDR domain, the results on FSIM are shown in Fig. 6. From these figures, the proposed structure can generate better quality than the normative encoder at the lower q values in the fixed Q case and at the higher Q values in the fixed q case.

Another comparison between the normative encoder and the proposed structure in terms of the file sizes of the encoded codestreams is shown in Fig. 7. It is confirmed that the difference between the proposed structure and the normative encoder is small.

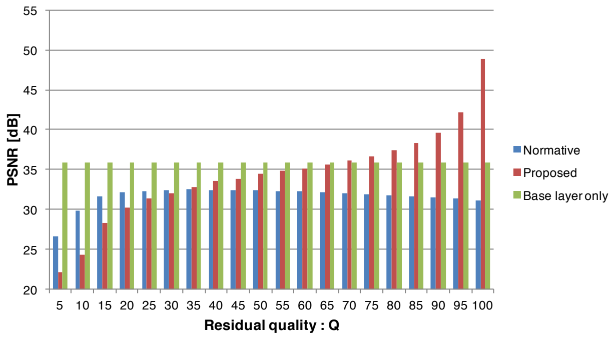
V. CONCLUSION

The coding performance of the JPEG XT profile A encoder was analyzed and the problem on the encoder was summarized in this paper. To avoid the problem, an improved structure has been proposed. The experimental results for the quality of the reconstructed HDR images were provided and confirmed that the proposed structure may give a better coding performance than the normative encoder.

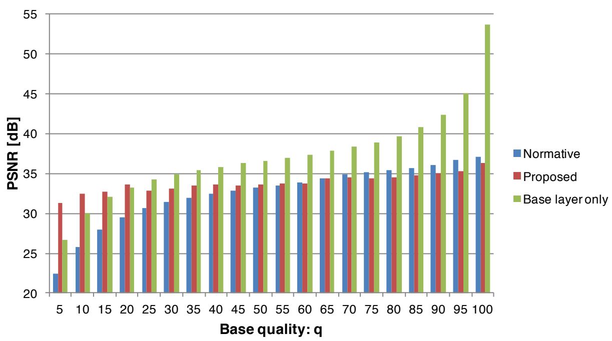
REFERENCES

[1] M. Iwahashi and H. Kiya, "Two layer lossless coding of HDR images," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 1340–1344.

[2] M. Iwahashi, T. Yoshida, N. Mokhtar, and H. Kiya, "Bit-depth scalable lossless coding for high dynamic range images," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, 2015.



(a) Fixed $q = 40$ for Base layer



(b) Fixed $Q = 40$ for residual layer

Fig. 5. PSNR comparison in tone-mapped LDR domain: Bars in blue, red and green stand for the results of the normative JPEG XT (base+residual), proposed (base+residual) and base layer only

[3] M. Iwahashi and H. Kiya, "Efficient lossless bit depth scalable coding for HDR images," in *APSIPA ASC, 2012 Asia-Pacific*, Dec 2012, pp. 1–4.

[4] "Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines," International Standard ISO/IEC IS-10918-1, Feb. 1994.

[5] "Information technology — JPEG XR image coding system – Part 1: System architecture," ISO/IEC TR-29199-1, Jul. 2011.

[6] "Information technology — JPEG 2000 image coding system – Part 1: Core coding system," International Standard ISO/IEC IS-15444-1, Dec. 2000.

[7] T. Richter, "On the standardization of the JPEG XT image compression," in *Picture Coding Symposium (PCS), 2013*, Dec 2013, pp. 37–40.

[8] "Information technology — Scalable compression and coding of continuous-tone still images – Part 1: Scalable compression and coding of continuous-tone still images," ISO/IEC 18477-1, Jun. 2015.

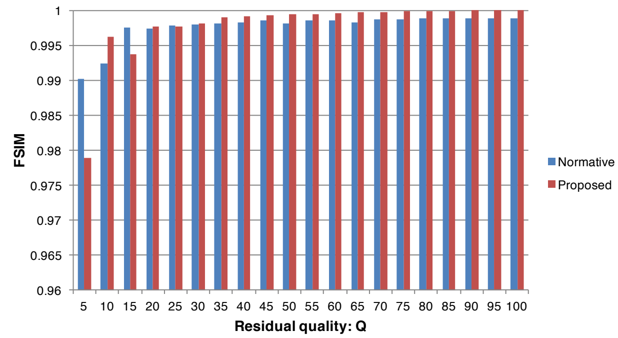
[9] C. Mantel, S. Ferchiu, and S. Forchhammer, "Comparing subjective and objective quality assessment of HDR images compressed with JPEG-XT," in *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, Sept 2014, pp. 1–6.

[10] A. Pinheiro, K. Fliegel, P. Korshunov, L. Krasula, M. Bernardo, M. Pereira, and T. Ebrahimi, "Performance evaluation of the emerging JPEG XT image compression standard," in *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, Sept 2014, pp. 1–6.

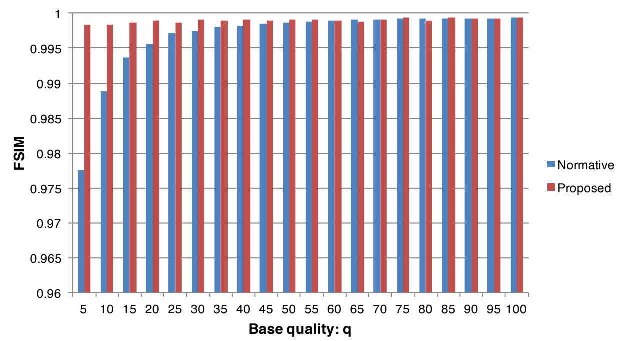
[11] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment." *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–86, Aug. 2011.

[12] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.

[13] "Text of CD ISO/IEC 18477-5 (Reference Software)," ISO/IEC JTC 1/SC 29/WG 1 N69019, Jun. 2015.

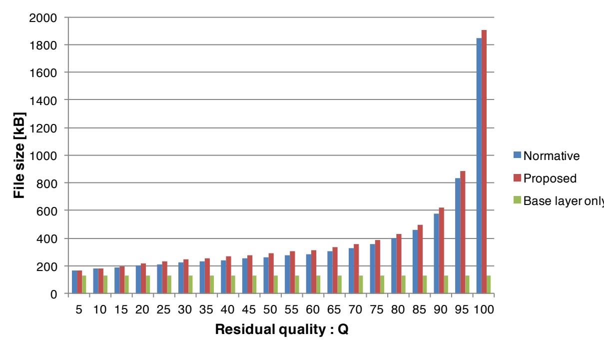


(a) Fixed $q = 40$ for Base layer

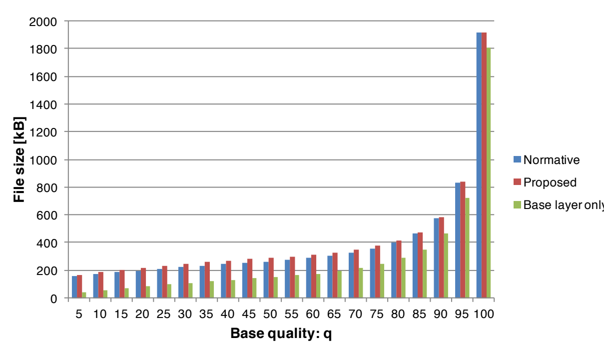


(b) Fixed $Q = 40$ for residual layer

Fig. 6. FSIM comparison in HDR domain: Bars in blue and red stand for the results of the normative JPEG XT (HDR), proposed (HDR)



(a) Fixed $q = 40$ for Base layer



(b) Fixed $Q = 40$ for residual layer

Fig. 7. File size comparison: Bars in blue, red and green stand for the results of the normative JPEG XT (base+residual), proposed (base+residual) and base layer only