Performance Evaluation of JPEG XT Standard for High Dynamic Range Image Coding

Seungcheol Choi, Oh-Jin Kwon, Dukhyun Jang, and Seokrim Choi

***Abstract*—This paper evaluates the JPEG XT profiles (Profile A, B, and C) proposed for an international standard of backward compatible JPEG image coding for high dynamic range (HDR) images. JPEG XT consists of two layers: the base layer containing tone-mapped low dynamic range (LDR) image data and the residual layer containing residual data for reconstructing the HDR image. Both layers are required to be coded by the legacy JPEG standard for the backward compatibility. This paper analyzes the compression performance based on the PSNR. Five HDR images, which are used in the verification test by JPEG working group, were selected as an experimental dataset. The photographic tone mapping operator was used for generating LDR image of the base layer. Experimental results demonstrate that Profile C can give good PSNRs and JPEG XT can encode HDR images at the bit rate around 1.0 to 1.4 bits per pixel for achieving PSNR values above 40 dB. The JPEG XT can reduce the file size more than 16 times to store the floating point HDR sample data than lossless OpenEXR while insuring the quality.**

***Index Terms*—performance evaluation, JPEG XT, high dynamic range, backward compatible image coding**

I. Introduction

RECENTLY, digital cameras providing high dynamic range (HDR) images whose color is represented by 12 or 14 bits per color channel are popular. However, most display devices and the de facto international image coding standard, the legacy JPEG (ISO/IEC 10918), still remain in low dynamic range (LDR) images whose color is represented by 8 bits per color channel.

Even though JPEG XR (ISO/IEC 15444) and JPEG 2000 (ISO/IEC 29199) were proposed for the HDR image coding standard, their usage has been limited in the industry mainly due to the lack of backward compatibility with the legacy JPEG. To resolve this problem, the JPEG working group (WG) is undertaking a new standardization of backward compatible JPEG image coding for HDR images named JPEG XT [1].

To satisfy the essential requirement of the backward compatibility with the legacy JPEG decoder, the JPEG XT consists of two layers: the base layer and the residual layer as shown in Fig. 1. The base layer is the backward compatible part and contains an encoded codestream of tone-mapped LDR version of the HDR image. A legacy JPEG decoder may decode the base layer by neglecting the residual layer. The residual layer provides a residual image data and transformation information to reconstruct the HDR image from the tone-mapped LDR image. Both encoded codestreams are completely compatible to the legacy JPEG decoder.

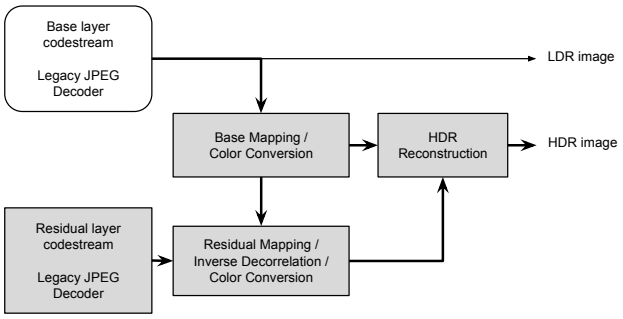


Fig. 1. Overview of the JPEG XT decoding process

The JPEG XT standard currently consists of nine parts. Part 1 defines the base coding technology, which is the legacy JPEG (ISO/IEC 10918) standard. Part 2 specifies a backward compatible extension of JPEG. Part 3 defines an extensible box-based file format, which is compatible to the JPEG File Interchange Format (JFIF, ISO/IEC 10918-5). Part 4 and 5 will define the conformance testing and provide the reference software of the JPEG XT standard. Part 6 defines extensions of the JPEG standard for coding of integer samples between 8 and 16 bits precision. Part 7 specifies the coding of HDR images of floating point samples. Part 8 defines lossless coding mechanisms combined with Part 6 and Part 7. Part 9 finally allows the lossy and lossless representation of alpha channels enabling the coding of transparency information and arbitrarily shaped images.

To-date, three profiles have been proposed for JPEG XT Part 7 [2]. This paper evaluates three profiles proposed for Part 7 floating point coding using a JPEG XT demo implementation [3].

Simplified functional blocks of three JPEG XT Part 7 profiles are depicted in Fig. 2. JPEG XT file contains codestreams of the base and residual layer. The application marker distinguishes them in the legacy JPEG decoder. White blocks are the common functional blocks for three profiles, and grey blocks illustrate the differences among the three profiles. Roughly, JPEG XT uses the Weber-Fechner law. The intensity of a sensation is proportional to the

Manuscript received March 24, 2015. This work was supported by the ICT Standardization program of MSIP (The Ministry of Science, ICT & Future Planning).

Seungcheol Choi is with the Department of Electronics Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-Gu, Seoul 143-747, Korea (corresponding author to provide phone: 82-2-3408-3828; fax: 82-2-3408-4329; e-mail:choisc@sju.ac.kr).

O. Kwon, D. Jang, and Seokrim Choi are with the Department of Electronics Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-Gu, Seoul 143-747, Korea (e-mail: {ojkwon@sejong., jangdh77@sju., schoi@sejong.}ac.kr.

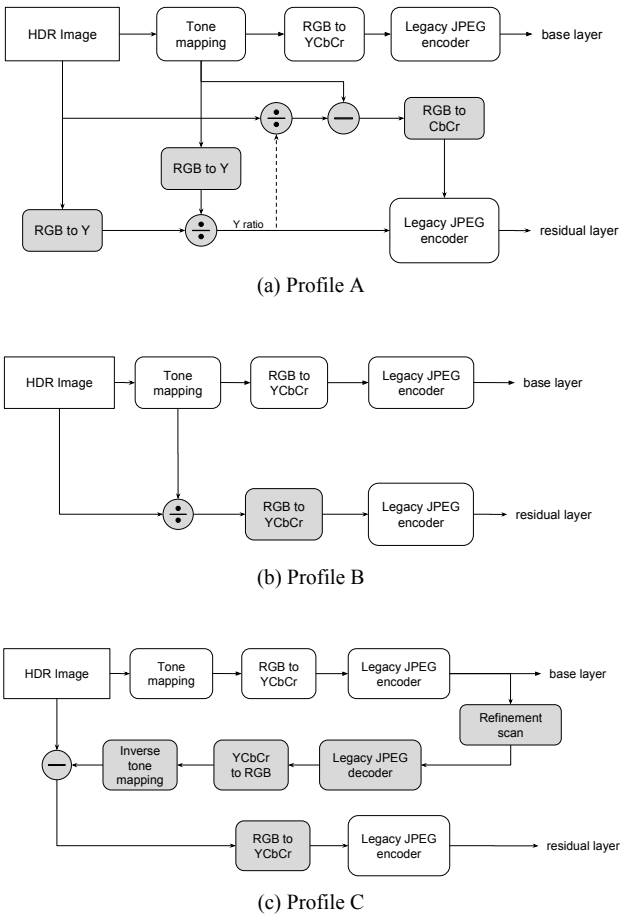


Fig. 2. Simplified encoding processes of JPEG XT profiles

logarithm of the intensity of the stimulus causing it. An HDR image is represented as a sum of base and residual image in the logarithmic domain or a product in the linear domain.

For all three profiles, the input HDR image is first tone-mapped from the original HDR image to the LDR image, which is coded by legacy JPEG encoder as a base layer. At the end of encoding process, the codestreams encoded by legacy JPEG encoder of the base layer and residual layer are composed into the JPEG XT codestream.

In Profile A, the HDR image is first converted from RGB to YCbCr color space then the residual data is produced by division and subtraction operations in the YCbCr space. The Y component of the residual image is the ratio of the HDR to LDR and the CbCr components are color residuals.

Profile B provides a tool that allows the use of HDR imaging in a simpler way. The base layer is either an exposed and clamped or tone-mapped version of the original HDR image. The HDR residual layer contains the fractional part of the tone-mapped LDR image divided by the original HDR image for each RGB channel.

Profile C represents the HDR image as a sum of the predicted HDR image and the final residual errors. As shown in Fig. 2(c), the refinement scan is performed to increase the bit precision up to 12 bits in the DCT domain.

The rest of the paper is organized as follows. Section II addresses the methodology of this evaluation and describes experimental setups. Section III presents the experimental results on three profiles of JPEG XT. Finally, Section IV concludes the paper.

II. Experimental Setup

A. Tone Mapping Operator

Generally, the tone mapping operator (TMO) is used to display the HDR image or video on the conventional LDR display device. To support backward compatibility with the legacy JPEG, the JPEG XT codestream delivers a tone-mapped LDR image of the original HDR image as a base layer. Various TMOs have been proposed in the literature. In JPEG XT, the TMO is open to users' choice, which means no default TMO is provided in the JPEG XT specification.

For this paper's evaluation, "Reinhard-02" by Reinhard *et al*. [4] is chosen as a TMO due to that JPEG WG selected the Reinhard-02 for the verification test of the JPEG XT standard [5].

B. Quality Metric

Objective quality metrics evaluate the image coding performance. These metrics can be classified into two categories according to the characteristics. One of which is based on the human visual system model such as VDP [5], HDR-VDP [6], PDM [7], and Sarnoff VDM [8]. The other is based on the statistical and structural measurement such as PSNR and SSIM [9]. However, the former metrics, due to their high computing complexity, are not as popular as PSNR and SSIM.

In this paper, a modified PSNR is used for coding performance comparisons. When the original and the coded image are denoted by $x(m,n)$ and $\hat{x}(m,n)$, $1 \le m \le M$, $1 \le n \le N$, respectively, where M and N are the vertical and the horizontal image size, respectively. The used PSNR is calculated as follows:

$$PSNR = 20\log_{10}(DR) - 10\log_{10}(MSE) \qquad (1)$$

with

$$MSE = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} | x(m,n) - \hat{x}(m,n) |^2 \qquad (2)$$

and

$$DR = \max\{x(m,n)\} - \min\{x(m,n)\} . \qquad (3)$$

C. Dataset

Five HDR images recommended by the JPEG WG are used for this paper's evaluation [10]. They are shown in Fig. 3. The dataset is composed of one indoor and four outdoor images not showing a discrete histogram in uncompressed PFM [11] file format. As shown in Table I, the *DR*, which

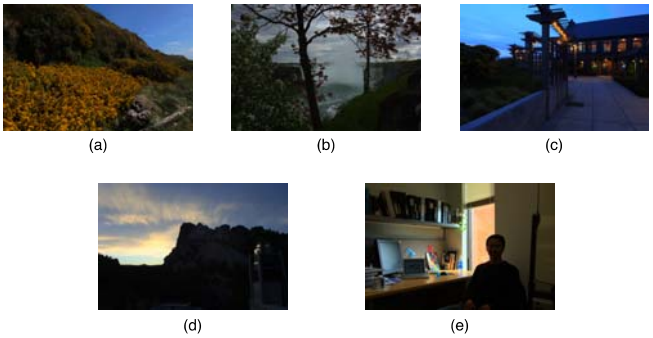


Fig. 3. Sample images: (a) BloomingGorse2, (b) CanadianFalls, (c) McKeesPub, (d) MtRushmore2, and (e) WillyDesk. Note that these images were tone-mapped using Reinhard-02 TMO.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
SELECTED HDR IMAGES

Name	Original pixel value		
	min	max	*DR*
BloomingGorse2	-0.0263	1.2051	1.2314
CanadianFalls	0.0001	1.6250	1.6249
McKeesPub	0.0000	7.1719	7.1719
MtRushmore2	0.0001	13.000	12.999
WillyDesk	-0.1533	17.5313	17.6846

expresses the range of the value within an image in RGB color space, was selected from 1.2314 to 17.6846.

D. Encoding Parameter

When a TMO is selected, the performance of three profiles of the JPET XT Part 7 depends on two quality values: one for the base layer legacy JPEG (denoted by q in this paper) and the other for the residual layer legacy JPEG (denoted by Q in this paper). 50 quality levels for q and Q each are employed for this paper's test.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 4 shows how Q levels affect the coding performance. The PSNR values in Fig. 4(a) and Fig. 4(b) were obtained by fixing q at 80 and 90, respectively, and increasing Q from 2 to 100 by 2. The plots show the PSNR versus bit rates for the all combination of q and Q. The result may be summarized as follows:

1) For all profiles, a bit rate of the JPEG XT codestream has been mainly determined by q.
2) Profile A demonstrates that the PSNR value does not increase monotonically with the bit rate for all images.
3) Profile B is highly dependent on sample images. Even for high values of q, it does not assign sufficient bits for the high DR images.
4) Profile C gives comparatively better PSNRs and exhibits lower dependency on sample images than the other profiles.
5) Profile C demonstrates unstable performances at the lower Q as depicted in Fig. 4(c).

Fig. 5 shows the mean PSNR performance on three profiles. This result may be summarized as follows:

1) Profile C shows the best PSNR performance among three profiles. At the higher bit rate, Profile C, which can use up to 12-bit for the residual layer, shows an advantage.
2) For all three profiles, the PSNR increases rapidly at the bit rate below 2 bits/pixel. Details are shown in Fig. 4(c).
3) Profile A and B exhibit earlier saturation. This means they do not work properly with the corresponding q and Q combination.

It is also noted that Fig. 4 and Fig. 5 show that the compression performance of the JPEG XT profiles is affected by not only the q and Q, but also the *DR* and content type (e.g., simple scene, complex scene, indoor, outdoor).

To estimate the coding performance of the JPEG XT three profiles, the bit rate compared to lossless image coding schemes is shown in Table II. The OpenEXR [12] is one of

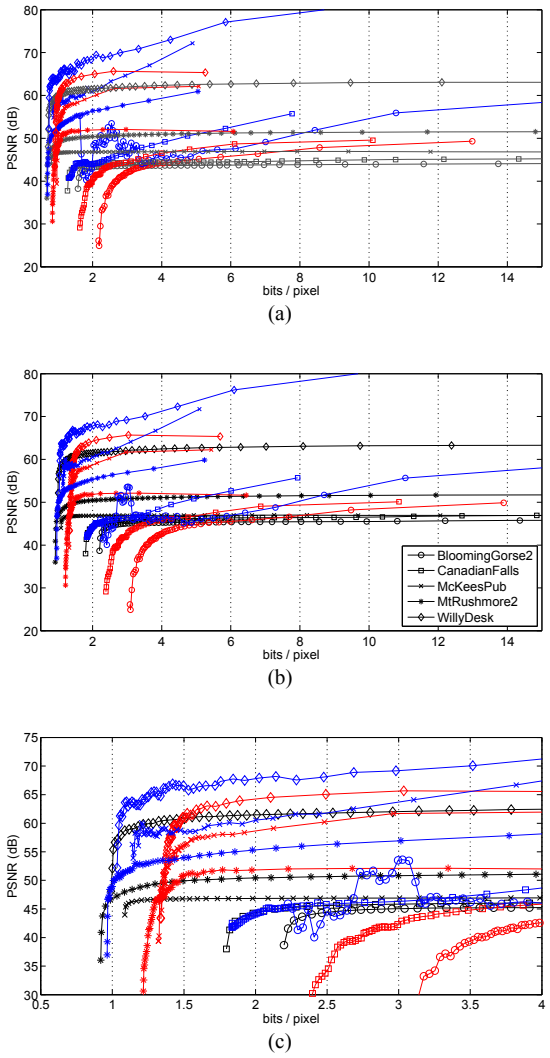


Fig. 4. Coding performance of each profile for different quality values of the residual layer. The quality value of the base layer was fixed at (a) 80 and (b) 90. Plot (c) focuses on the low bits/pixel of the plot (b). Colors represent each profile: Profile A (dark-gray), Profile B (red), and Profile C (blue).

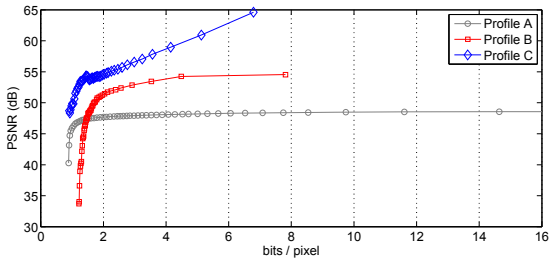


Fig. 5. Mean coding performance of each profile for different quality values of the base and residual layer.

TABLE II
COMPARISON OF COMPRESSION RATIO (BITS/PIXEL)

Name	OpenEXR	JPEG XT		
	Wavelet	Profile A	Profile B	Profile C
BloomingGorse2	20.4	1.5[a]	2.33[b]	1.87[c]
CanadianFalls	19.1	1.25[a]	1.69[b]	1.21[c]
McKeesPub	14.7	1.04	0.83	0.79
MtRushmore2	14.4	0.99	0.74	0.66
WillyDesk	15.5	2.04[a]	1.13[b]	0.75[c]
Aeverage	**16.82**	**1.364**	**1.344**	***1.056***

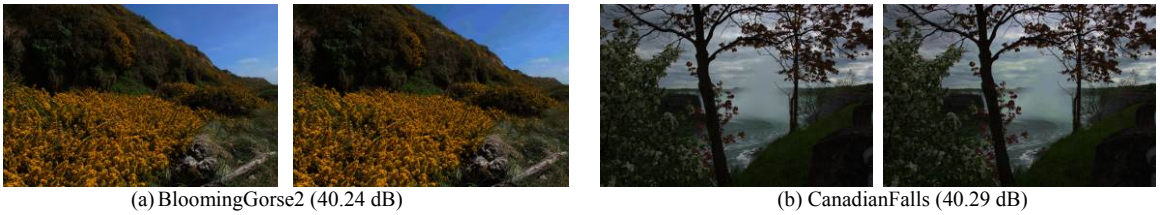
[a,b,c] Corresponding images are shown in Fig. 6, 7, and 8, respectively.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(a) BloomingGorse2 (42 dB) (b) CanadianFalls (42.39 dB)

(c) WillyDesk (61.50 dB)

Fig. 6. Subjective comparison of the Profile A: original HDR image (left) and decoded HDR image (right). All HDR images were tone-mapped for presentation.



(a) BloomingGorse2 (40.24 dB) (b) CanadianFalls (40.29 dB)

(c) WillyDesk (61.46 dB)

Fig. 7. Subjective comparison of the Profile B: original HDR image (left) and decoded HDR image (right). All HDR images were tone-mapped for presentation.

(a) BloomingGorse2 (46.68 dB) (b) CanadianFalls (42.73 dB)

(c) WillyDesk (63.24 dB)

Fig. 8. Subjective comparison of the Profile C: original HDR image (left) and decoded HDR image (right). All HDR images were tone-mapped for presentation.

the most popular HDR image formats. It offers several different image coding methods. In this evaluation, the wavelet lossless compression which provides the best compression ratio is selected for the compression. The bit rates of the JPEG XT thee profiles were selected so that they meet the following conditions:

1) The q value remains above 70.
2) The PSNR should be at least 40 dB.
3) The Q value is selected to minimize the visible distortions.

For the subjective comparison, the reconstructed HDR image and the original HDR image are shown in Fig. 6, 7, and 8. The HDR images were tone-mapped for the printed representation. It is shown that all profiles can provide high quality HDR images with bitrate between 1.0 and 1.4 bits/pixel, while 16.82 bits/pixel are needed on average for

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the OpenEXR wavelet compression.

IV. Conclusion

This paper gives the results of study to evaluate the compression performance of three profiles proposed for the JPEG XT part 7. The Profile C gives the best overall PSNR result of coding performance. On the other hand, the Profile A and B have shown that they reach early saturations with the increasing q and Q combinations and their results tend to be dependent on sample images. It is also shown that all the three JPEG XT profiles may achieve the PSNR over 40 dB in the range of 1.0 to 1.4 bits/pixel on average, which corresponds to a compression ratio of about 16 times compared to the lossless OpenEXR wavelet compression.

Experimental results also indicate that, as far as the PSNR performance is concerned, Profile A and B need to be improved to the performance of Profile C. Furthermore, the unstable performances of the Profile C at low Q values still need to be resolved.

The future works to analyze the influence of the TMOs on the compression performance and to select the most suitable quality metric and HDR dataset for objective evaluations will be needed.

References

[1] ISO/IEC SC29 WG1: "Proposal for a New Work Item-JPEG Extensions," JPEG document, WG1N6164, 2012.
[2] T. Richter, A. Artusi, and M. Agostinelli, "Text of ISO/IEC DIS1 18477-7," JPEG document, WG1N6839, Strasbourg, France, Oct. 2014.
[3] T. Richter, "JPEG XT Demo software for all profiles (ISO Version)," JPEG document, WG1N6883, 2014.
[4] E. Reinhard, M. Stark, P. Shirley, and J. Fewerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3. Pp. 267-276, July 2002.
[5] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, ed., MIT Press, 1993, pp. 179-206.
[6] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Visible difference predicator for high dynamic range images," in *2004 IEEE Int. Conf. System, Man and Cybernetics*, vol. 3, pp. 2763-2769.
[7] S. Winkler, *Digital Video Quality: Vision Models and Metrics,* John Wiley & Sons, Ltd, West Sussex, England, 2005.
[8] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed. Singapore: World Scientific, 1995, pp. 245–283.
[9] Z. Wang and A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.,* vol. 13, no. 4, pp. 600–612, Apr. 2004.
[10] T. Richter, "Minutes from th XT Interim Meeting in Brussels, December 3-5," JPEG document, WG1N6586, 2013.
[11] *PFM Portable Float Map Image Format*, [Online]. Available: http://www.pauldebevec.com/Research/HDR/PFM/
[12] *Technical Introduction to OpenEXR*, Industrial Light & Magic. [Online]. Available: http://www.openexr.com/TechnicalIntroduction.pdf

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)