SIP (2020), vol. 9, e16, page 1 of 15
doi:10.1017/ATSIP.2020.15

ORIGINAL PAPER

High dynamic range image compression based on visual saliency

JIN WANG,[1,2] SHENDA LI[1] AND QING ZHU[1]

With wider luminance range than conventional low dynamic range (LDR) images, high dynamic range (HDR) images are more consistent with human visual system (HVS). Recently, JPEG committee releases a new HDR image compression standard JPEG XT. It decomposes an input HDR image into base layer and extension layer. The base layer code stream provides JPEG (ISO/IEC 10918) backward compatibility, while the extension layer code stream helps to reconstruct the original HDR image. However, this method does not make full use of HVS, causing waste of bits on imperceptible regions to human eyes. In this paper, a visual saliency-based HDR image compression scheme is proposed. The saliency map of tone mapped HDR image is first extracted, then it is used to guide the encoding of extension layer. The compression quality is adaptive to the saliency of the coding region of the image. Extensive experimental results show that our method outperforms JPEG XT profile A, B, C and other state-of-the-art methods. Moreover, our proposed method offers the JPEG compatibility at the same time.

Keywords: High dynamic range, Tone mapping, Visual saliency, JPEG XT, Image compression

Received 31 May 2019; Revised 27 April 2020

I. INTRODUCTION

Most of the digital images we can see on the Internet are called low dynamic range (LDR) images. Their pixel values range from 0 to 255, and the ratio of max luminance to min luminance is 255, which may not reflect the real world exactly. High dynamic range (HDR) images provide a wider luminance range compared to LDR images, and they are closer to human visual system (HVS). Instead of conventional output devices, displaying HDR images requires high-bit-depth output devices. But for now, most existing output devices can only show LDR images. The general solution for this problem is to use a tone-mapping operator (TMO) that converts HDR to LDR images [1], thus any LDR device can display the tone mapped version of HDR images. HDR images can improve quality of experience in most of multimedia applications, such as photography, video, and 3D technology. But a primary drawback of HDR images is that memory and bandwidth requirements are significantly higher than conventional ones. So the compression method targeting for HDR images is in great need. For this reason, more and more researches have focused on this field.

[1]Faculty of Information Technology, Beijing University of Technology, 100124 Beijing, China
[2]Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing, China

Corresponding author:
Jin Wang
Email: ijinwang@bjut.edu.cn

Although some researchers are interested in HDR only compression schemes [2], the most popular image compression technique is still JPEG (ISO/IEC 10918). For most users, if they only have JPEG decoder on their devices, they will not know any information about the HDR content. As a result, an HDR image compression algorithm should provide JPEG compatibility to users only who have JPEG decoder. JPEG committee proposed JPEG XT standard (ISO/IEC 18477), which is a JPEG compatible HDR image compression scheme [3, 4]. The JPEG XT standard aims to provide higher bit depth support that can be seamlessly integrated into existing products and applications. While offering new features such as lossy or lossless representation of HDR images, JPEG XT remains backward compatible with the legacy JPEG standard. Therefore, legacy applications can reconstruct an 8-bit/sample LDR image from any JPEG XT code stream. The LDR version of the image and the original HDR image are related by a tone-mapping process that is not constrained by the standard and can be freely defined by the encoder. Although the JPEG XT standard promotes the development of HDR imaging technology and is widely accepted, still there is much room for improvements.

Khan [5] proposed a non-linear quantization algorithm aiming at reducing data volume of the extension layer. This method can significantly enhance the amount of details preserved in the extension layer, and therefore improve the encoding efficiency. It is also proved that the quantization algorithm can improve the performance on several existing two-layer encoding methods. Iwahashi *et al.* [6] used

noise bias compensation to reduce variance of the noise. They defined noise bias as the mean of the noise in the same observed pixel value from different original values, and compensated it according to an observed pixel value. This method has positive effect on bit rate saving at low bit rate lossy compression of LDR images. Pendu *et al.* [7] designed template-based inter layer prediction in order to perform the inverse tone mapping of a block without transmitting any additional parameter to the decoder. This method can improve accuracy of inverse tone-mapping model, thus obtained higher SSIM of reconstructed HDR images. Korshunov and Ebrahimi's work [8] is more interesting. They demonstrated no one tone-mapping algorithm can always stand out when compared with others. The choice of the best algorithm not only depends on the content, but also depends on the device used and other environmental parameters. They take all of those factors into consideration and optimize the performance. Fujiku *et al.* [9] first calculated a base map, which is a blurred version of the HDR image, and used it to generate the LDR image. Their method is suitable to preserve local contrast and has lower computational complexity. Choi *et al.* [10] proposed a novel method that is completely different from others. They generated the residual data in the DCT domain. The encoder predicted the DCT coefficients of the input HDR image based on LDR image, then prediction coefficients and DCT domain residual are encoded. Because inverse DCT process is not required in encoding phase, their method is rather efficient.

Although all of these methods are effective in most situations, the HVS factor is not considered. The purpose of HDR is to improve the quality of user's visual experience and make multimedia content more consistent with HVS. This means that some information in the image may be insensitive to the human eyes, but many bits have been wasted in encoding the redundant imperceptible information. Therefore, we should make a compromise between the high quality of the salient areas and the degradation of the non-salient areas. Feng and Abhayaratne [11] propose an HDR saliency estimation model and apply the visual saliency to guide HDR image compression. However, their work focuses on the saliency estimation model and existing HDR codec is used as a Blackbox. In this paper, we focus on the compression of HDR image with the guidance of visual saliency and propose a visual saliency-based HDR image compression scheme. We analyze and discover the correlation between visual saliency and residual image in HDR image compression. We use the extracted saliency map of a tone mapped HDR image to guide extension layer encoding. The correlation between visual saliency and residual image is exploited to adaptively tune the quantization parameter according to saliency of image content. We incorporate the visual saliency guidance into the HDR image codec. This will ensure much higher image quality of regions that humans are most interested, and lower quality of other unimportant regions. Our contributions are summarized into the following three points:

(1) A visual saliency-based HDR image compression scheme is proposed, in which the saliency map of a tone mapped HDR image is used to guide extension layer encoding.
(2) The correlation between visual saliency and residual image in HDR image compression is analyzed and exploited, which is modeled to adaptively tune the quantization parameter according to saliency of image content.
(3) Extensive experimental results show that our method outperforms JPEG XT profile A, B, C and other state-of-the-art methods while offering JPEG compatibility meanwhile.

The rest of the paper is organized as follows: Section II illustrates background information of this paper. Section III introduces the proposed method in details. Experimental results are given in Section IV, we first analyze the effect of different saliency models and influence of quality range. Then we compare the proposed scheme with JPEG XT profile A, B, C and other state-of-the-art methods. Finally, discussions and conclusions are given in Section V.

II. PRELIMINARY

A) JPEG XT

In JPEG XT standard part 7, a backward compatible HDR image compression scheme is proposed [3]. It is able to encode images with bit depths higher than 8 bits per channel. The input HDR image I is encoded into base layer B and residual layer R codestreams as shown in Fig. 1. Base layer B is the tone mapped version of input HDR image I, which can be compressed by the JPEG encoder, and the base layer codestream is constructed to provide JPEG backward compatibility. The residual layer codestream allows a residual layer decoder to reconstruct the original HDR image I starting from the base layer B. The coding tools of the overall JPEG XT infrastructure used to merge B and R together are then profile dependent. Since that, the two-layer JPEG compatible compression framework has received universal approval. The choice of TMO is open to users. JPEG XT does not define an encoder, but provides three schemes about how to reconstruct the HDR image, named profiles A, B, and C. The main differences between them is how to merge the base layer image and extension layer image.

Profile A reconstructs the HDR image I by multiplying a luminance scale μ with the base image B after inverse gamma correction using the first base non-linearity Φ_A.

$$I(x,y) = \mu(R_o(x,y)) \cdot [C\Phi_A(B(x,y)) + \nu(SC\Phi_A(B(x,y))) \cdot DR^{\perp}(x,y)], \quad (1)$$

where C and D are 3×3 matrices implementing color transformations, $\mu(\cdot)$ is a scalar function of the luma component of the residual layer R, and $R^{\perp}$ is the residual layer projected onto the chroma-subspace. The matrix C transforms from ITU-R BT.601 to the target colorspace in the residual layer. D is an inverse color decorrelation transformation

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

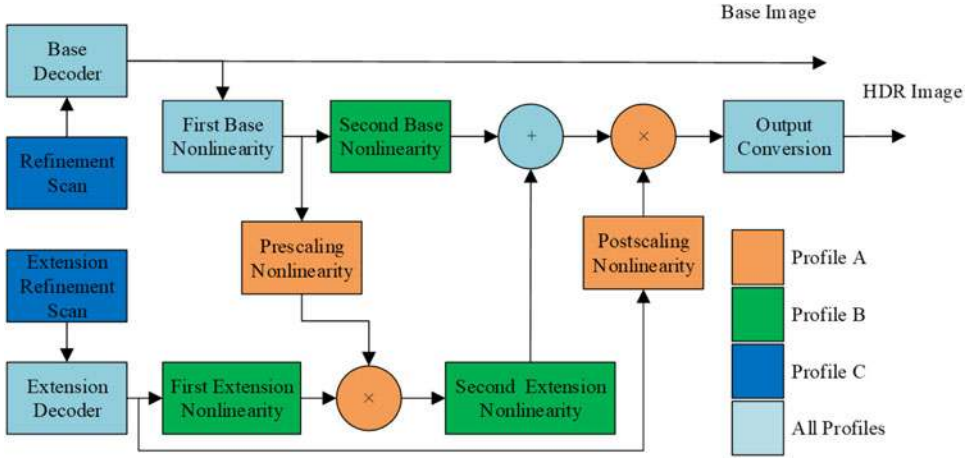

Fig. 1. Framework of JPEG XT.

from YCbCr to RGB in the residual layer to clearly separate the luminance component from the chromaticities at the encoding level. These matrices are also commonly used in the other two profiles. S is a row-vector transforming color into luminance, and $v(\cdot)$ is a scalar function taking in input luminance values.

Profile B reconstructs the HDR image I by computing the quotient that can be expressed as a difference in the logarithmic scale:

$$\begin{aligned} I(x,y)_i &= \sigma \exp(\log([C\Phi_B(B(x,y))]_i) \\ &\quad - \log(\Psi_B([DR(x,y)]_i) + \varepsilon)) \\ &= \sigma \frac{[C\Phi_B(B(x,y))]_i}{\Psi_B([DR(x,y)]_i) + \varepsilon}, \quad (i = 0,1,2), \end{aligned} \tag{2}$$

where i is the index of the RGB color channels. Φ_B and Ψ_B are two inverse gamma applied to the base and residual layers respectively. Φ_B has the objective to linearize the base layer, while Ψ_B intends to better distribute values closer to zero in the residual layer. The scalar σ is an exposure parameter that scales the luminance of the output image to optimize the split between base and residual layers.

Profile C also employs a sum to merge base and residual images, but here Φ_C not only approximates an inverse gamma transformation, but implements a global inverse tone-mapping procedure that approximates the TMO that was used to create the LDR image. The residual layer R is encoded in the logarithmic domain directly, avoiding an additional transformation. Finally, log and exp are substituted by piecewise linear approximations that are implicitly defined by reinterpreting the bit-pattern of the half-logarithmic IEEE representation of floating-point numbers as integers. The reconstruction algorithm for profile C can then be written:

$$\begin{aligned} I(x,y) = \psi \exp(&\hat{\Phi}_C(CB(x,y)) \\ &+ DR(x,y) - 2^{15}(1,1,1)^T), \end{aligned} \tag{3}$$

where $\hat{\Phi}_C(x) = \psi log(\Phi_C(x))$, in which Φ_C is the global inverse tone-mapping approximation. 2^{15} is an offset shift to make the extension image symmetric around zero. The code stream never specifies Φ_C directly, but rather includes a representation of $\hat{\Phi}_C$ in the form of a lookup-table, allowing to skip the time-consuming computation of the logarithm.

B) Visual saliency

When watching at an image, human attention may be attracted by some specific regions, while other regions will be neglected. Visual saliency characterizes the most sensitive regions to HVS. In general, there are two categories of models to estimate the visual saliency of an image: bottom-up model and top-down model. In bottom-up model, saliency is only determined by how different a stimulus is from its surroundings. Top-down model, however, takes into account the internal state of the organism at this time [12]. Due to huge difficulty and dependency of prior knowledge, top-down model is less popular than bottom-up model. Here, we introduce some methods based on bottom-up model.

All of these models can be divided into global methods and local methods, or the combination of them. Cheng *et al.* [13] proposed a global method base on histogram contrast. The saliency of a pixel is defined using its color contrast in Lab color space to all other pixels in the image. Hou *et al.* [14] proposed a global method in log-spectrum domain. They calculated spectral residual of an image, and constructed the corresponding saliency map according to the spectral residual. Instead of global method, Achanta *et al.* [15] considered local method. In their opinions, saliency is determined as the local contrast of an image region with respect to its neighborhood at various scales. They added three saliency maps together, which is obtained in different scales, and got the final saliency. Goferman *et al.* [16] combined both global and local methods. Their method considered local low-level features, i.e. contrast and color, and global features to get better result.

In recent years, saliency estimation of HDR image also has been considered [11, 17, 18, 19]. Feng and Abhayaratne

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

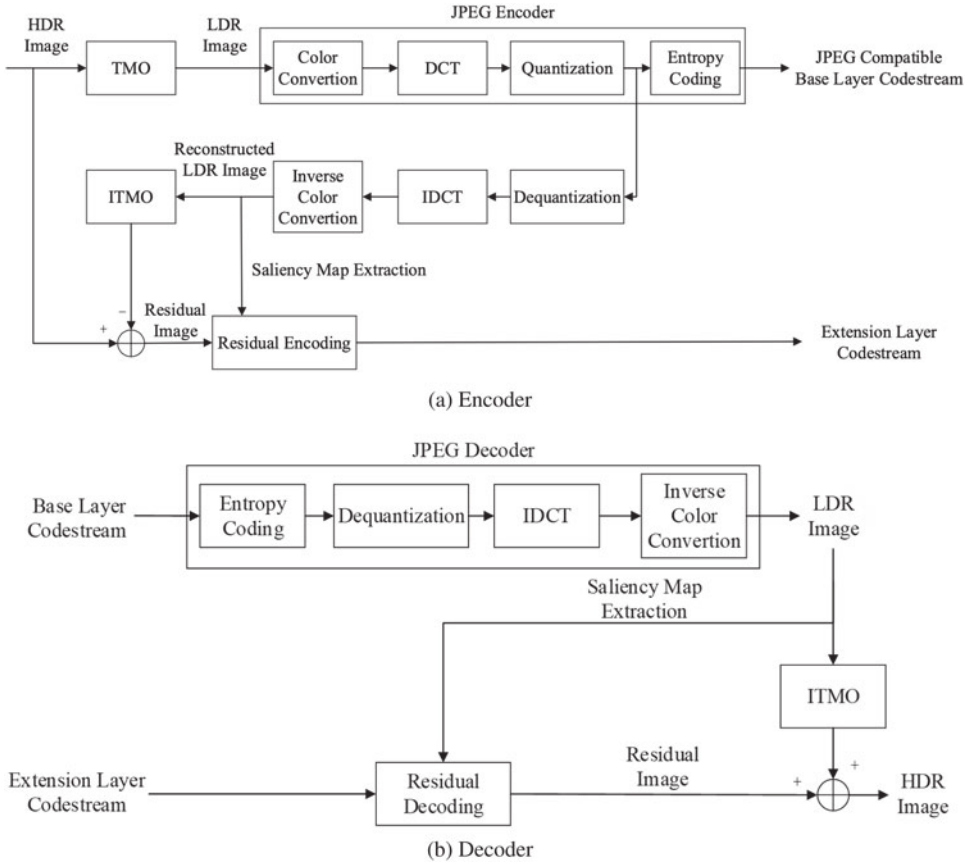

Fig. 2. Framework of our proposed method. (a) Encoder (b) Decoder.

[11] propose an HDR saliency map detection method containing three parts: tone mapping, saliency detection, and saliency map fusion. Brémond *et al.* [19] proposes a Contrast-Feature algorithm to improve the saliency computation for HDR images. Due to wider luminance range, calculating saliency map of HDR image is more difficult. However, the principle does not change, some traditional methods are also available.

III. PROPOSED METHOD

A) Framework of our scheme

In this paper, we propose a JPEG compatible HDR image compression scheme based on visual saliency. The framework of our work is as shown in Fig. 2.

In Fig. 2(a), the input HDR image is first tone mapped to get the corresponding LDR image. There are many previous works on tone mapping [20, 21, 22]. Rana *et al.* [20] and [21] propose a locally adaptive, image-matching optimal TMO which is guided by the support vector regressor (SVR) based predictor model. Rana *et al.* [22] propose a fast, parameter-free and scene-adaptable deep tone-mapping operator (DeepTMO) based on conditional generative adversarial network (cGAN). We need to point out that many existing tone-mapping methods can be applied in our compression framework, and more recent tone-mapping methods will lead higher performance. In order to focus on our compression scheme itself and make a fair comparison with existing compression methods, while considering the efficiency of the entire codec, we use a simple logarithmic function for tone mapping. Then the LDR image is encoded by JPEG (quality q) and sent to base layer codestream. The reconstructed LDR image, which is decoded by JPEG decoder, is used to encode the residual. The saliency map of reconstructed LDR is extracted by existing methods. Then all HDR values which are mapped to the same LDR value are averaged, and a look-up table is created for inverse TMO (ITMO). The ITMO of reconstructed LDR image is used to approximate the HDR image. Finally, the difference between original HDR image and approximated HDR image is called residual. In extension layer, we use another JPEG-based coding scheme (quality Q), and adaptively set different Q for different blocks. Instead of converting to YCbCr color space, we directly encoded the residual in RGB color space. As for decoding, as shown in Fig. 2(b), the base layer codestream provides JPEG compatibility, which can be decoded by JPEG decoder. Then, the saliency map of reconstructed LDR image is extracted to guide the extension layer decoding, which is just the same way as encoding. At last, the residual image and ITMO of the reconstructed LDR image are combined to obtain the final reconstructed HDR image.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

Table 1. Correlation between residual image and saliency map.

Image	q	Cheng [13]	Hou [14]	Achanta [15]	Goferman [16]	Average
Memorial	30	0.0688	0.1911	**0.3840**	0.2092	0.2631
	60	0.1989	0.1998	**0.4528**	0.2404	
	90	0.3096	0.2047	**0.4557**	0.2426	
AtriumNight	30	0.0264	0.1831	**0.5557**	0.2140	0.1755
	60	0.0385	0.1932	0.1313	**0.1935**	
	90	0.0238	0.1901	0.1496	**0.2067**	
Tree	30	**0.4993**	0.2925	0.3832	0.3483	**0.3891**
	60	**0.5254**	0.2945	0.4016	0.3303	
	90	**0.5400**	0.2932	0.4220	0.3392	
Nave	30	0.1141	0.0724	**0.2849**	0.0203	0.1428
	60	0.1845	0.0766	**0.3277**	0.0254	
	90	0.1665	0.0704	**0.3109**	0.0595	
Rosette	30	0.4373	0.1883	**0.4420**	0.2857	0.3544
	60	0.4708	0.1951	**0.5002**	0.2914	
	90	0.4710	0.1889	**0.4933**	0.2883	
BigFogMap	30	0.2062	0.1333	**0.3443**	0.2502	0.2302
	60	0.1226	0.1133	**0.3626**	0.2465	
	90	0.1473	0.1509	**0.3952**	0.2903	
Rendo6	30	0.0602	**0.2256**	0.0438	0.1364	0.1261
	60	0.0711	**0.2028**	0.0686	0.1536	
	90	0.0920	**0.2271**	0.0670	0.1651	
Rendo9	30	0.1416	**0.1985**	0.0534	0.1956	0.1762
	60	0.1581	**0.2188**	0.0437	0.1968	
	90	0.1566	0.2414	**0.2830**	0.2265	
Average		0.2179	0.1894	0.3065	0.2148	

B) Residual analysis

The residual image is defined as follows:

$$res = I - ITMO(DEC(COM(TMO(I)))), \tag{4}$$

where I denotes the original HDR image, *TMO* and *ITMO* denote the tone-mapping operator and corresponding inverse TMO, *COM* and *DEC* denote JPEG compression and decompression. Residual mainly comes from many-to-one mapping in TMO and ITMO and/or quantization error in JPEG compression and decompression.

In this section, we first analyze the correlation between residual and saliency map. Before calculating saliency map, the reconstructed LDR image is first gamma corrected with $\gamma = 1/2.4$. The saliency maps are extracted by four different methods [13, 14, 15, 16]. For those methods which can not obtain full resolution saliency map, bicubic interpolation is applied to extend the obtained saliency maps to full resolution. The saliency is normalized to 1 by equation (5):

$$Sal'(m,n) = \frac{Sal(m,n)}{\sum_m \sum_n Sal(m,n)}, \tag{5}$$

where Sal and Sal' denote saliency map before and after normalization respectively. The correlation coefficient between two matrixes is defined as:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2 (\sum_m \sum_n (B_{mn} - \bar{B})^2}}, \tag{6}$$

where A and B denote two matrices, and $\bar{A} = \sum_m \sum_n A_{mn}$, $\bar{B} = \sum_m \sum_n B_{mn}$.

We averaged correlation between saliency map and absolute value of residual in red, green, and blue channels. We changed base layer image quality q. From Table 1 we can see that for most cases, the residual images have relatively high correlation with the saliency maps. We notice for some cases the correlation coefficients between them are not very high, the reasons are from the following three aspects after our study. The first reason is the saliency estimation method. We need to point out that all these saliency estimation methods are objective approximations of saliency mechanism of HVS, thus their results are more or less inconsistent with the groundtruth. These four saliency estimation models extract saliency maps based on different principles. From the results, the Achanta's method [15] has the highest average correlation. This model calculates the difference between the pixel value of a point and its neighboring pixels on different scales, and then superimposes this difference on different scales. For smaller regions, the difference between the pixel and its surrounding pixels on any scale can be detected, and its saliency is the sum of these three different scales. Larger image areas can only be detected at larger scales. So this model always highlights small areas. In HDR images, the reconstruction error of the bright area is often the highest, that is, the residual is the largest. The scale of the bright area is generally not large and rarely appears continuously. This fits the essence of the model and therefore has the highest correlation, thus we choose this model in our scheme. The Hou's method [14] has the lowest average correlation because it considers the frequency domain characteristics. The essence of the residual error is the sum of losses in the mapping and compression process, which

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

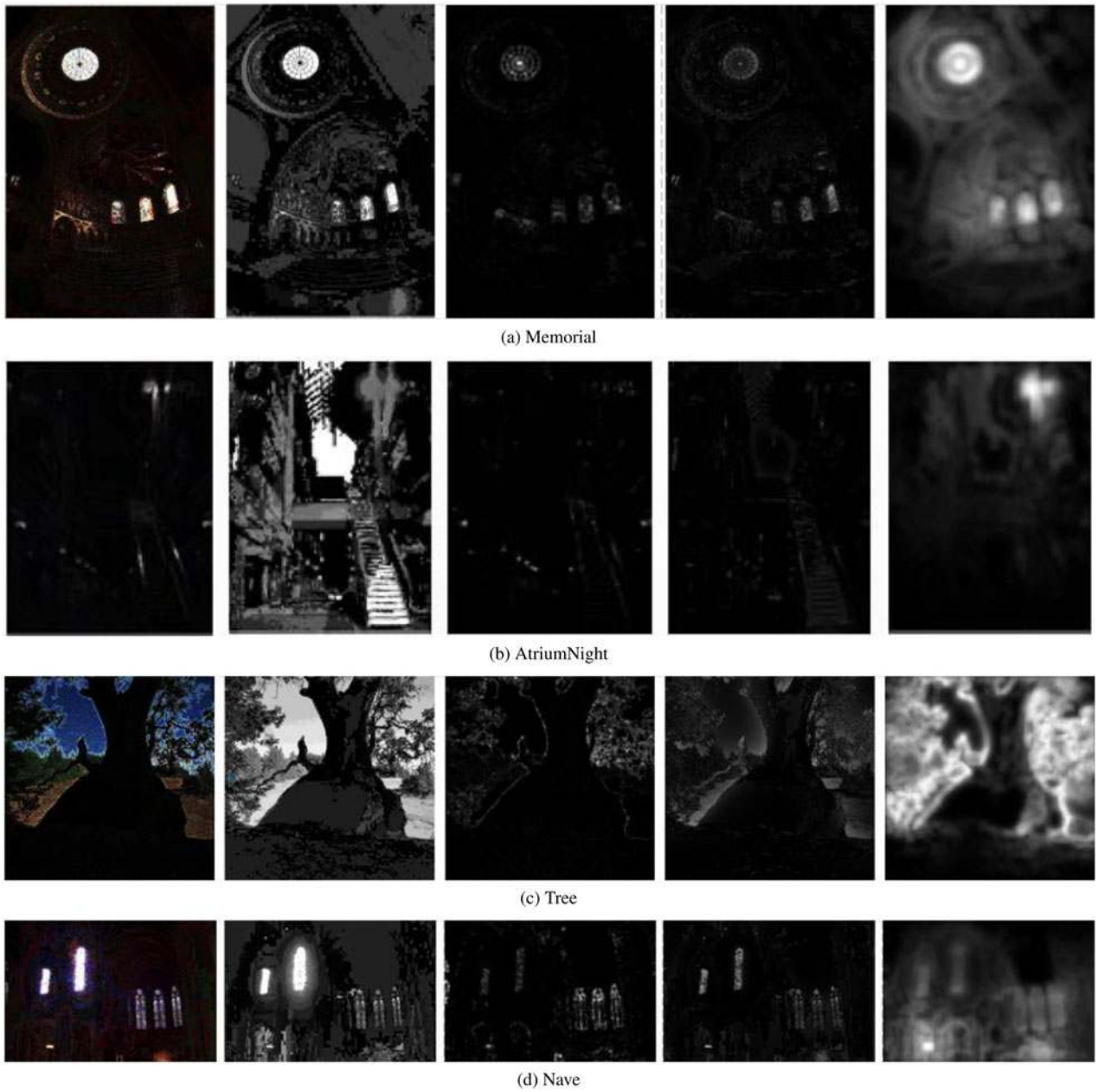

Fig. 3. Visualization of residual layers and saliency maps from different methods for test images Memorial, AtriumNight, Tree, and Nave. For each row, the first image is the residual layer, the second to the fifth images are saliency map extracted by Cheng's method [13], Hou's method [14], Achanta's method [15], and Goferman's method [16] respectively. (a) Memorial, (b) AtriumNight, (c) Tree, (d) Nave.

is not related to the frequency domain. On average, this model is not suitable to guide the compression of HDR images. The second reason is the compression quality. In general, the differences among correlation coefficients are not large when the same image is compressed with different quality, so the correlation coefficients are nearly the same for different compression quality. The exception is the correlation between the residuals of AtriumNight and the saliency maps obtained using Achanta's method [15]. They are very relevant on low-quality images. However, the correlation decreases significantly as the compression quality improves. One possible reason is that high-quality images retain more details, and these newly appeared details makes the relevant local image blocks in low-quality images no longer relevant. The third reason is the test images. There is vast difference among the content of different images, which show quite different features. We find that not all image residuals are correlated with saliency maps. For example, the correlation coefficients of Nave and Rendo6 are low, which indicates that the residuals of the two images are not concentrated in the visually significant areas but in the insensitive areas. For Tree and Rosette, the average of their correlation coefficients exceed 0.35, even up to 0.5. This shows that these image residuals are very related to saliency, that is, the areas with large residuals are also the areas that the human eye are sensitive to. For most cases, we can draw this conclusion.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

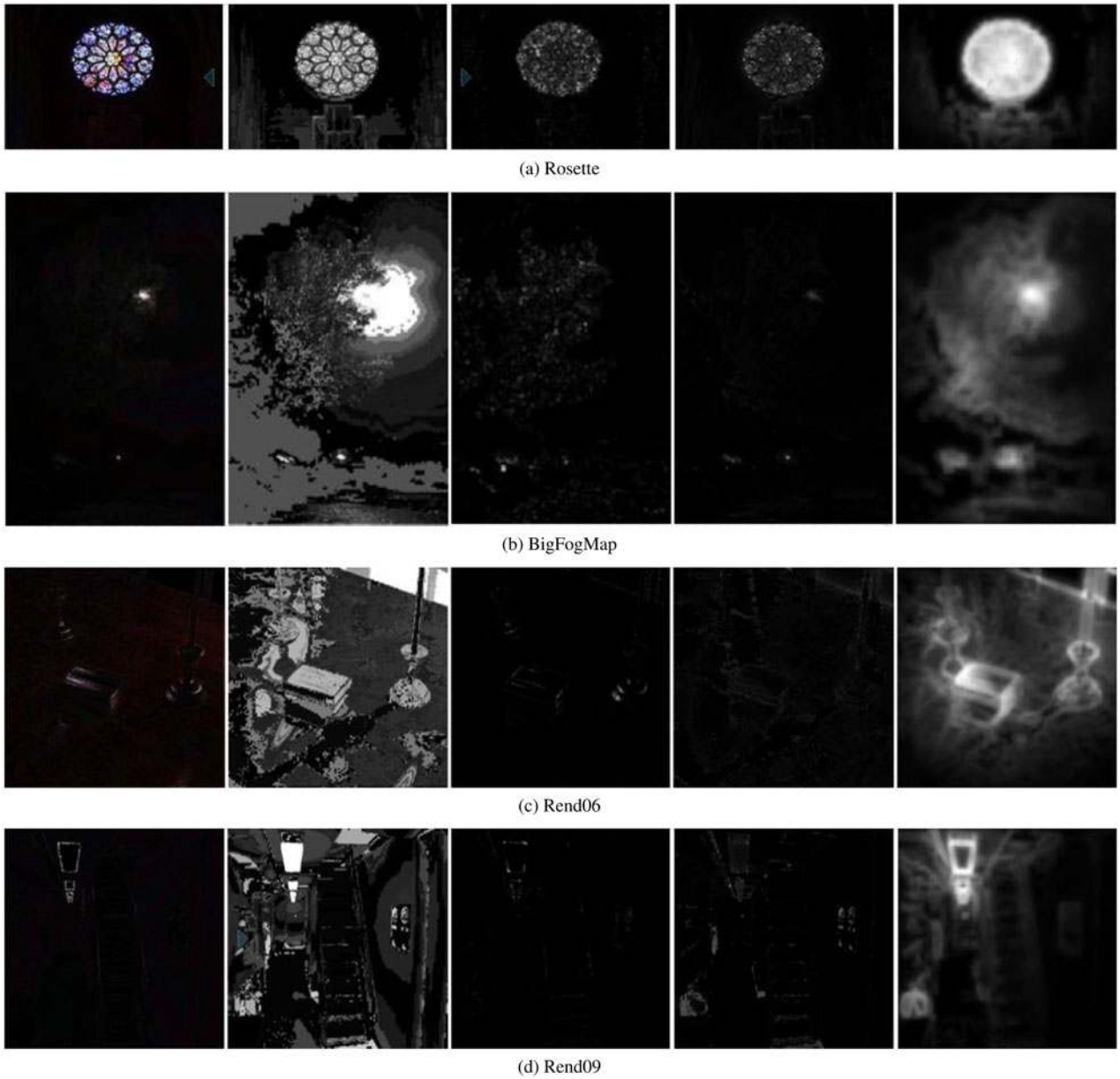

Fig. 4. Visualization of residual layers and saliency maps from different methods for test images Rosette, BigFogMap, Rendo6, and Rendo9. For each row, the first image is the residual layer, the second to the fifth images are saliency map extracted by Cheng's method [13], Hou's method [14], Achanta's method [15], and Goferman's method [16] respectively. (a) Rosette, (b) BigFogMap, (c) Rendo6, (d) Rendo9.

More intuitive results can be found in Figs 3 and 4. It is more obvious from Figs 3 and 4 that the residual layer and the saliency map are highly correlated. Specifically, the probability that the residual layer and the saliency map are "light" or "dark" at the same location is very high, which indicates that it is reasonable to use the saliency map to approximate the distribution of the residuals. For the decoder, although the distribution of the residuals cannot be obtained, it is possible to obtain a saliency map based on the reconstructed LDR image, thus obtaining the approximate distribution of the original residuals. Applying this property to compression will effectively guide the decoder to allocate the code stream more reasonably, and finally improve the overall quality of image compression.

C) Adaptive quality parameter

As we can see in Section III B), there exists correlation between residual and saliency map. As a result, we can use visual saliency to guide residual layer coding.

In JPEG coding, an image is divided into 8×8 blocks, thus we also can segment the saliency map in the same way. We first normalize the saliency map by equation (5). Then we sum up the saliency of each block, calculating the mean saliency $\bar{S}$ of all blocks. We specify the baseline quality of extension layer Q, but it is just an "average" quality of all the blocks. For the blocks whose saliency are greater than $\bar{S}$, their quality is much higher, otherwise, the quality is lower. Equation (7) shows the relationship

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

Fig. 5. Example of extension layer quality distribution matrix. (a) Reconstructed LDR image, (b) Saliency map, (c) Extension layer quality distribution.

between them:

$$dQ = round \begin{cases} -k \times \bar{S}/s & s < \bar{S} \\ 0 & s = \bar{S} \\ k \times s/\bar{S} & s > \bar{S} \end{cases}, \tag{7}$$

where s denotes the saliency of current block, and k is the parameter to control range between the maximum and minimum of the quality. dQ is "relative quality" of current block. As a result, the quality of current block is:

$$Q_{ada} = Q + dQ. \tag{8}$$

In order to avoid negative quality for the least salient region, we set minimal dQ to $-Q/2$. The relative quality calculated by (8) is adjusted to 100 if it is greater than 100. Equation (9) shows the final quality of each block.

$$Q_{ada} = \begin{cases} Q/2 & Q_{ada} < Q/2 \\ 100 & Q_{ada} > 100 \\ Q_{ada} & otherwise \end{cases}. \tag{9}$$

As a result, quantization level is adaptively selected according to saliency. Thus the rate-distortion performance of extension layer coding can be optimized. Take Fig. 5 as an example, in this case, baseline quality Q is 70 and k iso.4. Using equation (5), we can get the highest quality 77 and the lowest quality 46. The whole process is as depicted in Algorithm 1. Our proposed visual saliency-based HDR image compression and decompression are as in Algorithms 2 and 3 respectively.

Algorithm 1 Extension layer quality calculation

Input:
Baseline quality Q, coefficient k.
Output:
Extension layer quality.
Main:
1: Get saliency map;
2: Use equation (5) to normalization;
3: Calculate saliency of each block and average saliency;
4: Calculate relative quality by equation (7);
5: Calculate extension layer quality by equation (8);
6: Adjust extension layer quality by equation (9);
return Q_{ada}.

Algorithm 2 Compression algorithm

Input:
HDR image.
Output:
Base layer codestream $c1$ and extension layer codestream $c2$.
Main:
1: Get saliency map;
2: Tone mapping HDR image to LDR image;
3: Use JPEG to encode LDR image to get $c1$;
4: Calculate residual by equation (4);
5: Use Algorithm 1 to calculate extension layer quality;
6: Use extension layer quality to encode extension layer by JPEG to get $c2$;
return $c1$,$c2$.

IV. EXPERIMENTAL RESULTS

A) Experimental setup

In this section, extensive experiments are carried out to validate the performance of our compression scheme. We implement our proposed method in PC with Window 10 and MATLAB 2016b. The eight test images are from Ward's HDR image set [23], as shown in Fig. 6, with their details shown in Table 2.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

Algorithm 3 Decompression algorithm

Input:

Base layer codestream $c1$ and extension layer codestream $c2$.

Output:

LDR image and HDR image.

Main:

Decode $c1$ to get LDR image;
if JPEG user:
  return LDR image;
else HDR user:
  Inverse tone mapping LDR image to get HDR image;
  Use Algorithm 1 to calculate extension layer quality;
  Decode $c2$ to get extension image;
  Add base image and extension image;
  return HDR image;
end if

HDR visible difference predictor 2 (HDR-VDP-2) is adopted as evaluation metric as it is most consistent with human eyes [24]. As the most widely used method to estimate HDR image quality, it can predict similarity between a pair of images and calculate mean-opinion-score of test images.

B) Quality range adjusting

As mentioned in Section III B), k is the parameter to control the range of quality. In this section, we evaluate the influence of different k. Test image Memorial and Achanta's saliency detection method [15] is used. Then, we set base layer image quality q equal to extension layer quality Q. The result is shown in Fig. 7.

The result shows that: (1) Adaptive quality parameter is better than fix quality. $k = 0$ means Q is equal to all blocks. It is obvious that setting quality adaptively has positive effect. (2) At low bit rate, $k \neq 0$ has almost the same performance as $k = 0$. Because the baseline Q itself is small, the quality of non-salient region is extremely small, probably discarding all residual information. When the bit rate increases, all the blocks can obtain a suitable quality. (3) As k increases, the quality distribution become more and more uneven. In this case, when the quality is high, salient region can be nearly lossless coded. However, quality of other regions is not so satisfactory. Generally, set a moderate k can optimize the performance.

Table 2. Details of test images.

Image	Resolution	Dynamic range(log10)	Size(MB)
Memorial	512 × 768	4.8	1.27
AtriumNight	760 × 1016	4.1	2.47
Tree	928 × 906	4.4	2.81
Nave	720 × 480	6.0	1.09
Rosette	720 × 480	4.4	1.14
BigFogMap	744 × 1128	3.6	2.56
Rendo6	1024 × 1024	3.6	2.41
Rendo9	1024 × 1024	3.9	2.93

C) Bitstream balancing

In this paper, bitstream balance refers to the bitrate allocation between the base layer and the extension layer when the total bitrate is fixed. In previous section, we balance

Fig. 6. Test image set.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

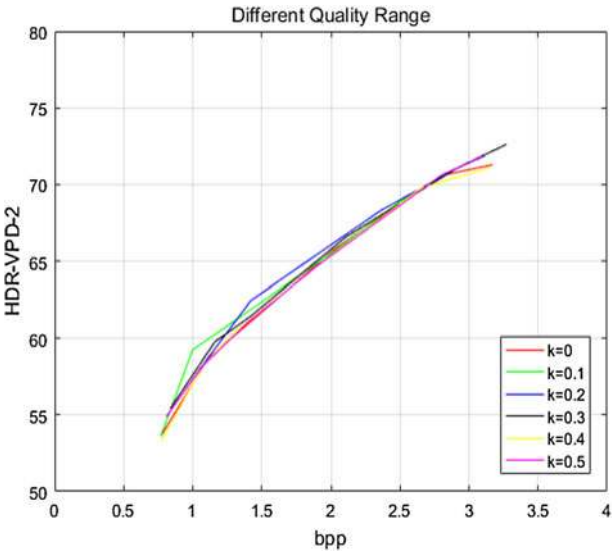

Fig. 7. Different quality ranges.

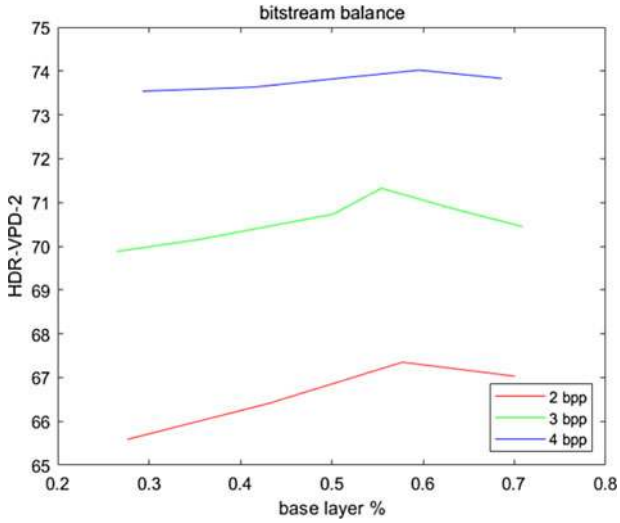

Fig. 8. Bitstream balancing.

the bitstream between salient and non-salient regions in the extension layer. In this section we discuss the allocation of bitstream between the base layer and the extension layer. The practical significance of this discussion is that when the user's storage or bandwidth are limited to a fixed bit rate C, it is important to distribute these bitstreams to the base layer and the extension layer to achieve best quality of reconstructed HDR image.

In the following experiments, we fixed the total bit rate at 2, 3, and 4 bpp respectively. As shown in Fig. 8, we can draw these conclusions from the experimental results: (1) A moderate distribution can get the best result, over-allocating on either base layer or extension layer is not always good. (2) With the increase of bitstream, the optimal point moves to the right, which means base layer tends to occupy more bitstream. (3) The gap between optimal point and other points shrinks rapidly when the bitstream is sufficient. That means bitstream balance is quite a problem at low bit rate. However, as the total bitstream is enough, allocating the bitstream between base layer and extension layer has little effect on overall performance.

D) Comparison with JPEG XT and other methods

To validate the efficiency of our proposed method, we make extensive comparison with other methods. As mentioned in Section II A), we do not take JPEG XT profile D into consideration, so we compare our method with profiles A, B, C. They are implemented by JPEG XT working group [25], and the recommended parameters are used in our experiments. Besides that, we compare our method with other state-of-the-art work, such as Choi *et al.* [10], Wei *et al.* [26], and Feyiz *et al.* [27]. HDR-VDP-2 [24] and SSIM [28] are adopted as the evaluation metric of objective quality. In our proposed method, Achanta's saliency map is applied due to its highest correlation to residual image, and the parameter is set as $k = 0.3$. The objective quality comparison results are as shown in Figs 9 and 10.

In general, our method outperforms all of other methods, except for low bit rates. One possible reason is that the baseline quality Q is not so high when the bit rate is low, which may cause severe degradation in non-salient region. As a result, the total HDR-VDP-2 of whole image is inferior to JPEG XT. However, when the bit rate increases, the balance between salient regions and non-salient ones is desirable. So our proposed method is better than other methods when the bit rate is sufficient.

We also compare the subjective quality at a fix bit rate, as shown in Figs 11, 12 and 13. For clarity of comparison, the decoded image is zoomed in specified region (100 × 100 pixels). The regions in red square are salient regions and regions in blue square are non-salient ones. As we can see, in salient regions, our method achieves best quality. While in non-salient region, our method is comparable to JPEG XT profiles A, B, and slightly inferior to profile C.

V. CONCLUSIONS

In this paper, a visual saliency-based HDR image compression scheme is proposed. It divides the input HDR image into two layers, base layer and extension layer. The base layer codestream provides JPEG backward compatibility, and any JPEG decoder can reconstruct LDR version of HDR image. Extension layer codestream helps to reconstruct the original HDR image. The saliency map of tone mapped HDR image is first extracted, then is used to guide extension layer coding. For the most salient region, we set the highest quality, and the quality of other regions is adaptively adjusted depending on its saliency.

Extensive experiments have been conducted to validate the proposed scheme. We analyze the correlation between residual image and saliency map extracted by some classical approaches. It can be easily proved that local model visual saliency will get better result. We also analyze the influence

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

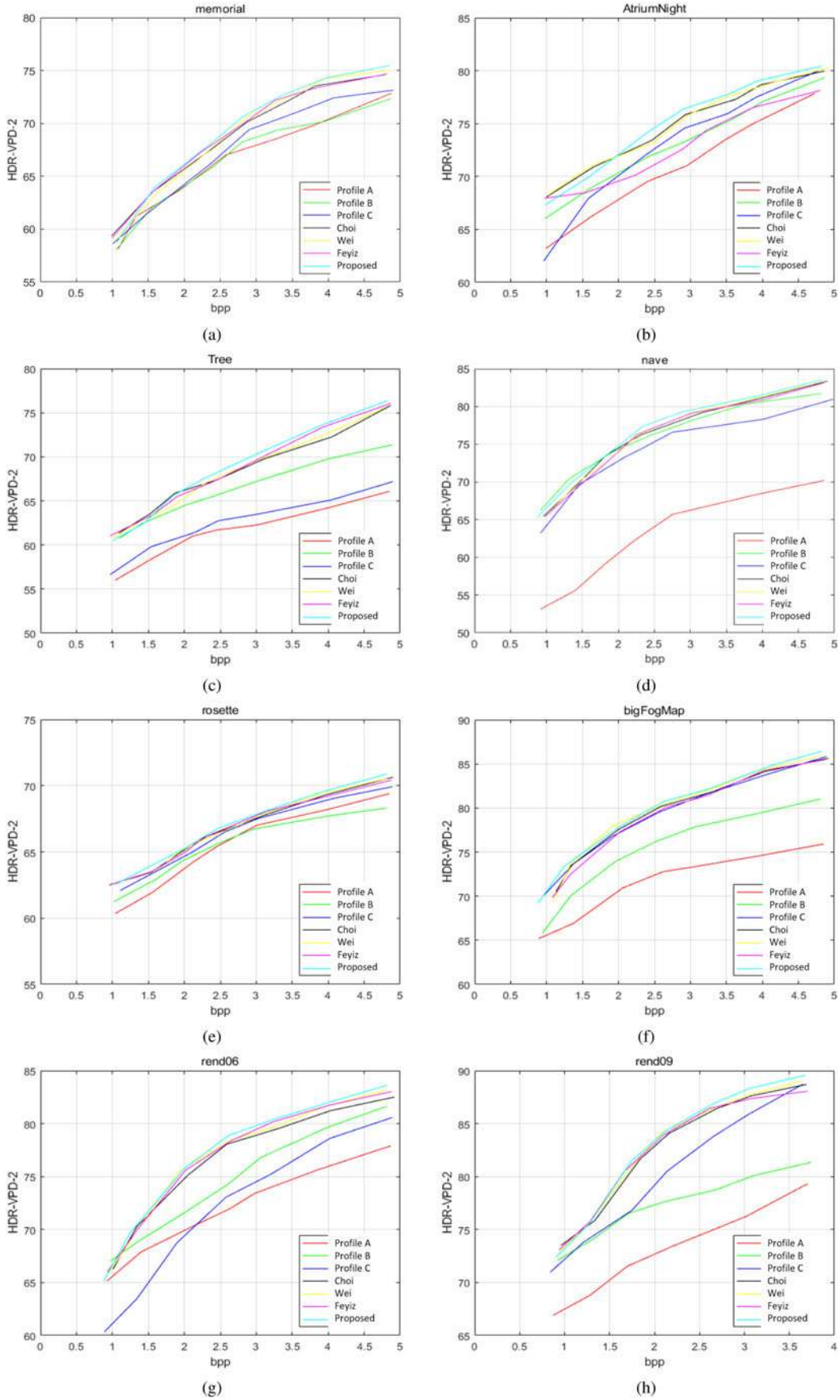

Fig. 9. HDR-VDP-2 results comparison with different methods.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

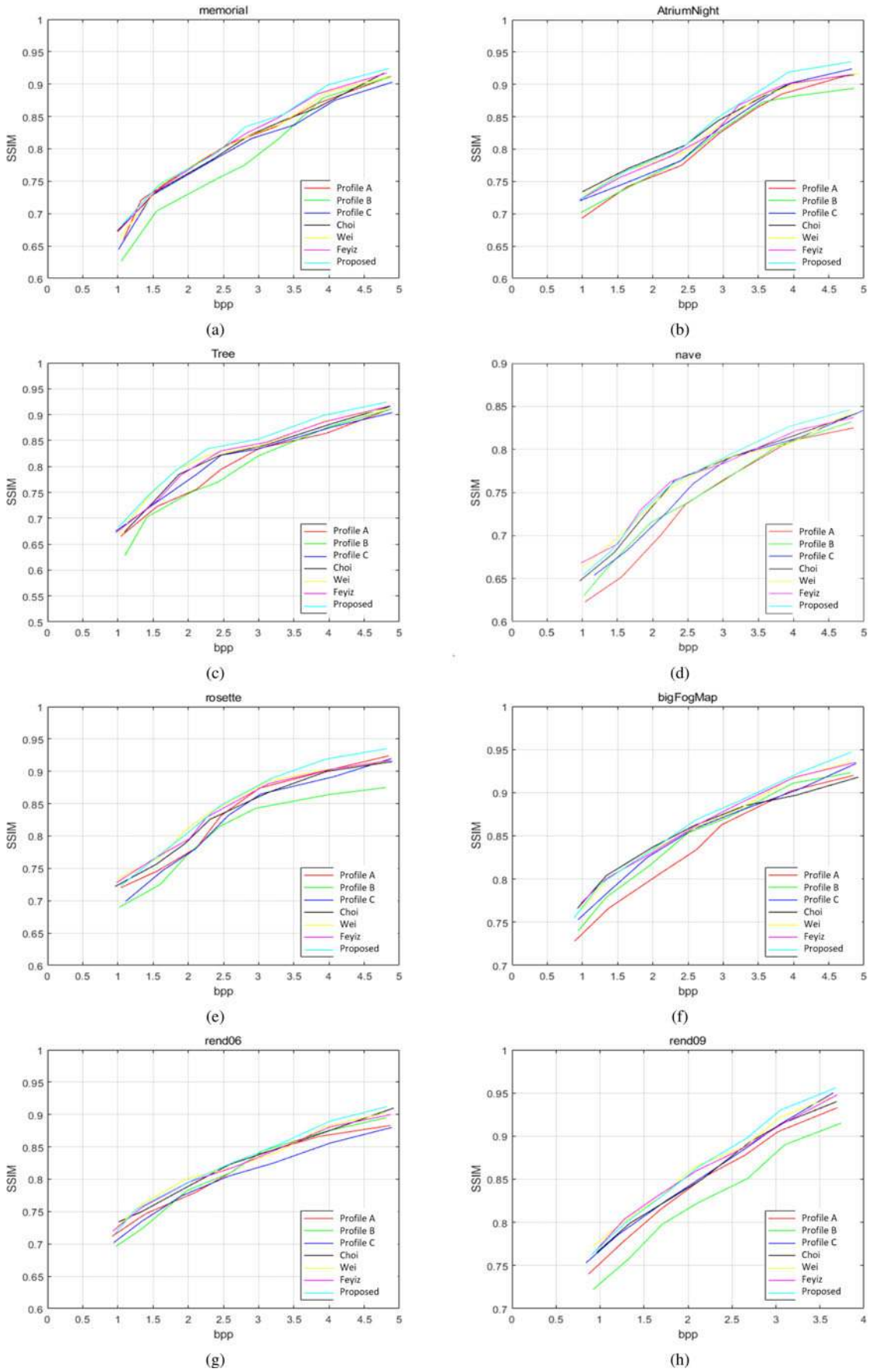

Fig. 10. SSIM results comparison with different methods.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

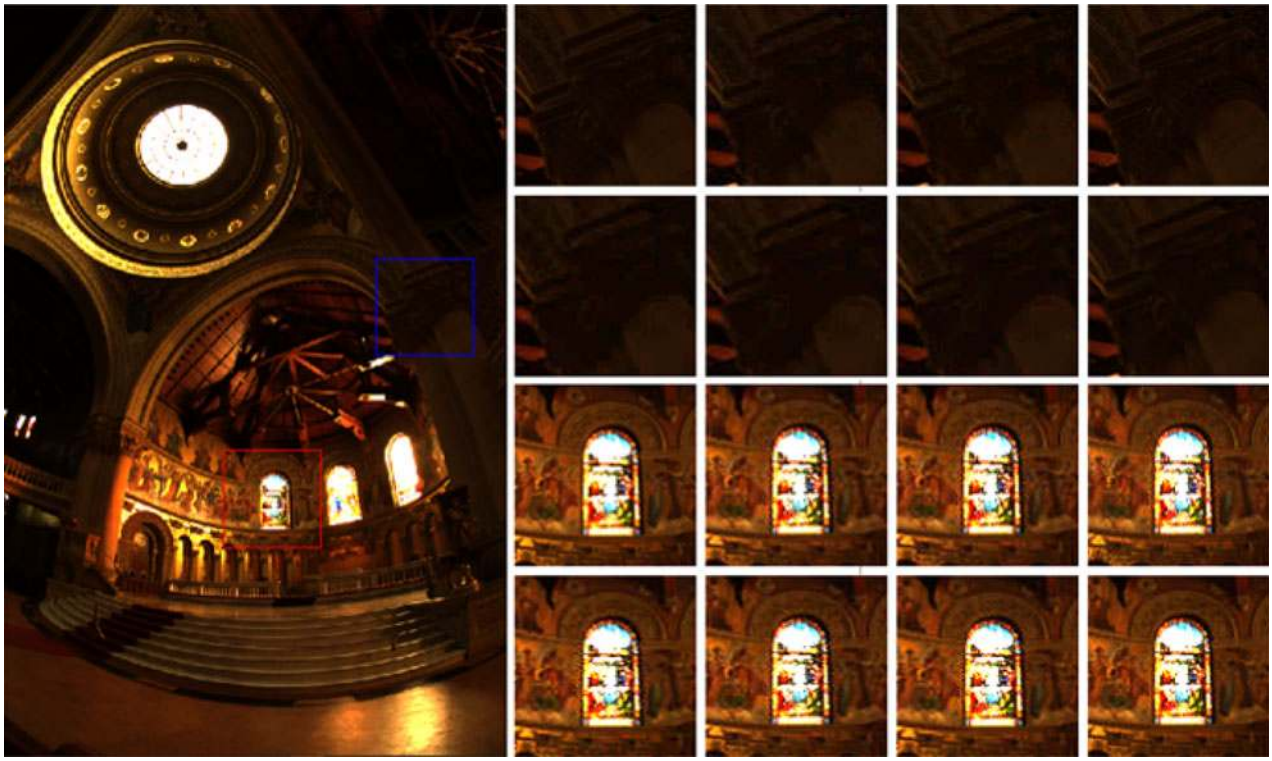

Fig. 11. Visual quality comparison of memorial with different methods at 3.2 bpp. (On the left is the original image, top two rows on the right are non-salient regions of reconstructed image, and two bottom rows on the right are salient ones. From left to right, the first row is original image, official implementation of JPEG XT profile A,B,C, the second row is Choi's method10 [10], Wei's method21 [21], Feyiz's method22 [22], and our proposed method, respectively.)

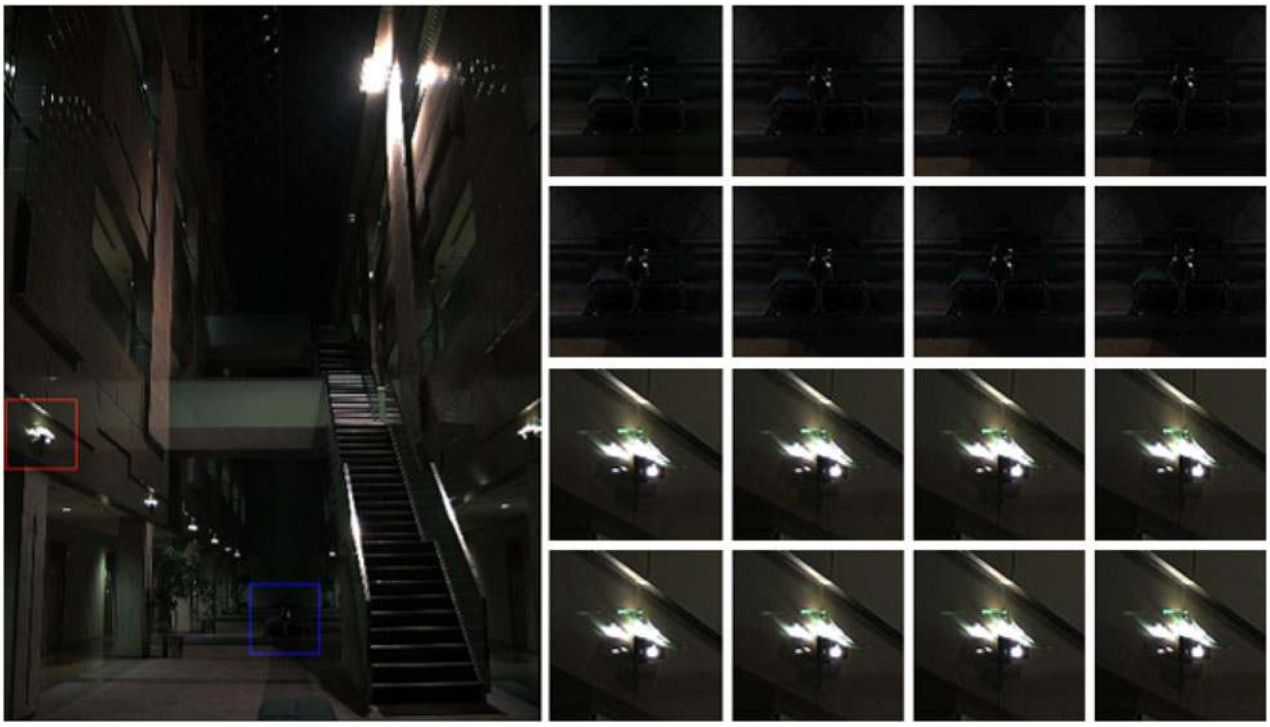

Fig. 12. Visual quality comparison of AtriumNight with different methods at 2.0 bpp. (On the left is the original image, top two rows on the right are non-salient regions of reconstructed image, and two bottom rows on the right are salient ones. From left to right, the first row is original image, official implementation of JPEG XT profile A,B,C, the second row is Choi's method10 [10], Wei's method 21 [21], Feyiz's method 22 [22], and our proposed method, respectively.)

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

Fig. 13. Visual quality comparison of BigFogMap with different methods at 2.7 bpp. (On the left is the original image, top two rows on the right are non-salient regions of reconstructed image, and two bottom rows on the right are salient ones. From left to right, the first row is original image, official implementation of JPEG XT profile A,B,C, the second row is Choi's method [10], Wei's method [26], Feyiz's method [27], and our proposed method, respectively.)

of quality range. Experimental result shows that a moderate parameter can control the quality range well. Then, we use them to guide extension layer compression, and compare it with JPEG XT profiles A, B, C. Using HDR-VDP-2 as evaluation metric, our method outperforms JPEG XT standard in both objective and subjective quality. The main reason is that our method balances the bit rate between salient regions and non-salient ones. This is a desirable trade-off to ensure higher quality of information that is significant to human eyes, and lower quality of redundant imperceptible information.

As to the running time, our proposed method is slower than JPEG XT, because saliency map and relative quality need to be calculated in both encoder and decoder side. The computational complexity of calculating saliency map is scheme dependent, while relative quality is easily computed. So in practice, a simple saliency extraction algorithm can keep the running time almost the same as JPEG XT.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61906008, 61632006, 61672066, 61976011, 61671070), Scientific Research Project of Beijing Educational Committee (KM202010005013), and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research.

REFERENCES

[1] Reinhard E.; Stark M.; Shirley P.; Ferwerda J.: Photographic tone reproduction for digital images, in *Conf. on Computer Graphics and Interactive Techniques*, Vol. 21, ACM, 2002, San Antonio, Texas, 267–276.

[2] Koz A.; Dufaux F.: A comparative survey on high dynamic range video compression. *Proceedings Volume 8499, Applications of Digital Image Processing XXXV*, San Diego, California, United States. 2012.

[3] Artusi A.; Mantiuk R.K.; Richter T.; Korshunov P.: JPEG XT: A compression standard for hdr and wcg images. *IEEE Signal. Process. Mag.*, **33** (2) (2016), 118–124.

[4] Richter T.; Artusi A.; Ebrahimi T.: JPEG XT: A new family of jpeg backward-compatible standards. *IEEE Multimedia*, **23** (3) (2016), 80–88.

[5] Khan I.R.: A nonlinear quantization scheme for two-layer hdr image encoding. *Signal Image Video Process.*, **10** (5) (2015), 1–6.

[6] Iwahashi M.; Hamzah F.A.B.; Yoshida T.; Kiya H.: Two layer coding of HDR images with noise bias compensation, in *Asia-Pacific Signal and Information Processing Association Summit and Conf.*, IEEE, Hong Kong, 2015, 436–439.

[7] Pendu M.L.; Guillemot C.; Thoreau D.: Template based inter-layer prediction for high dynamic range scalable compression, in *IEEE Int. Conf. Image Processing*, *IEEE,* Quebec City, QC, Canada, 2015.

[8] Korshunov P.; Ebrahimi T.: Context-dependent jpeg backward-compatible high-dynamic range image compression. *Optical Eng.*, **52** (10) (2013), 102006-1–102006-11.

[9] Fujiki T.; Adami N.; Jinno T.; Okuda M.: High dynamic range image compression using base map coding, in *Asia-Pacific Signal and Information Processing Association Summit and Conf.*, IEEE, Hollywood, CA, USA, 2013, 1–4.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press

[10] Choi S.; Kwon O.J.; Lee J.; Kim Y.: A jpeg backward-compatible image coding scheme for high dynamic range images. *Digit. Signal Process.*, **67** (2017), 1–16.

[11] Feng T.; Abhayaratne C.: Visual saliency guided high dynamic range image compression, in *2018 26th European Signal Processing Conf. (EUSIPCO)*, IEEE, Rome, Italy, 2018, 166–170.

[12] Laurent I.; Christof K.; Ernst N.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20** (11) (1998), 1254–1259.

[13] Cheng M.M.; Zhang G.X.; Mitra N.J.; Huang X.; Hu S.M.: Global contrast based salient region detection, in *Computer Vision and Pattern Recognition*, Vol. 37, IEEE, Colorado Springs, CO, USA, 2011, 409–416.

[14] Hou X.; Zhang L.: Saliency detection: A spectral residual approach, in *Computer Vision and Pattern Recognition*, 2007. CVPR '07. IEEE Conf. on Vol. 2007, IEEE, Minneapolis, Minnesota, USA, 1–8.

[15] Achanta R.; Estrada F.; Wils P.; Sï£¡ï£¡sstrunk S.: Salient region detection and segmentation, in *Computer Vision Systems*, Vol. 5008, Springer, Berlin Heidelberg, 2008, 66–75.

[16] Goferman S.; Zelnik-Manor L.; Tal A.: Context-aware saliency detection, in *IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, San Francisco, CA, USA, 2010.

[17] Dong Y.; Pourasad M.; Nasiopoulos P.: Human visual system based saliency detection for high dynamic range content. *IEEE Trans. Multimedia*, **18** (4) (2016), 549–562.

[18] Xie D.; Jiang G.; Shao H.: Saliency detection for high dynamic range images via global and local cues, in *Int. Conf. Geo-Informatics in Resource Management and Sustainable Ecosystem*, Springer, Singapore, 2016, 43–51.

[19] Brémond R.; Petit J.; Tarel J.P.: Saliency maps of high dynamic range images, in *European Conf. on Computer Vision*, Springer, Berlin, Heidelberg, 2010, 118–130.

[20] Rana A.; Valenzise G.; Dufaux F.: Learning-based tone mapping operator for efficient image matching. *IEEE Trans. Multimedia*, **21** (1) (2018), 256–268.

[21] Rana A.; Valenzise G.; Dufaux F.: Learning-based adaptive tone mapping for keypoint detection, in *2017 IEEE Int. Conf. Multimedia and Expo (ICME)*, IEEE, Hong Kong, 2017, 337–342.

[22] Rana A.; Singh P.; Valenzise G. *et al.*: Deep tone mapping operator for high dynamic range images. *IEEE Trans. Image Process.*, **29**, (2019), 1285–1298.

[23] The Greg Ward's online HDR image set [Online]. Available at: http://www.anyhere.com/gward/hdrenc/pages/originals.html. Last accessed on: April 26, 2020.

[24] Mantiuk R.; Kim K.J.; Rempel A.G.; Heidrich W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, **30** (4) (2011), 1–14.

[25] JPEG official homepage [Online]. Available at: http://www.jpeg.org. Last accessed on: April 26, 2020.

[26] Wei Z.; Wen C.; Li Z.: Local inverse tone mapping for scalable high dynamic range image coding. *IEEE Trans. Circuits Syst. Video Technol.*, **28** (2) (2018), 550–555.

[27] Feyiz K.; Kamisli F.; Zerman E. *et al.*: Statistical analysis and directional coding of layer-based HDR image coding residue, in *IEEE Int. Workshop on Multimedia Signal Processing*, IEEE, London-Luton, UK, 2017, 1–6.

[28] Wang Z.; Bovik A.C.; Sheikh H.R. *et al.*: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, **13** (4) (2004), 600–612.

Jin Wang received the B.S. degree in computer science from Huazhong University of Science and Technology(HUST), Wuhan, China in 2006 and the M.S. and the Ph.D. degrees in computer science from Beijing University of Technology(BJUT), Beijing, China in 2010 and 2015, respectively. He is currently a lecturer in the Faculty of Information Technology at BJUT. His areas of interests include digital image processing, image/video compression and computer vision.

Shenda Li received the M.S. degree in software engineering from Beijing University of Technology, Beijing, China, in 2019. His research interests is image compression and HDR image processing.

Qing Zhu received the M.S. and the Ph.D. degrees in electronic information and communication from Waseda University in 1994 and 2000 respectively. After graduation, she worked in the Institute of Information and Communication of Waseda University. She is currently a Professor with the Faculty of Information Technology, Beijing University of Technology. Her research interests include multimedia information processing technology, virtual reality technology and information integration technology.

https://doi.org/10.1017/ATSIP.2020.15 Published online by Cambridge University Press