MergeNet: Single High Dynamic Range Image Reconstruction Method

Bin Liang
MRAD of Beijing Institute of Technology

Dongdong Weng*
MRAD of Beijing Institute of Technology and AICFVE of Beijing Film Academy

Ruikang Ju
MRAD of Beijing Institute of Technology

Lulu Feng
MRAD of Beijing Institute of Technology

ABSTRACT

The high dynamic range (HDR) environment mapping could provide great dynamic range and irradiance contrast for improving the fidelity of Computer Graphics Rendering, however, the acquisition of HDR images is much harder for most scenarios. We propose the improved deep merger network (MergeNet) to reconstruct an HDR image from a single filtered low dynamic range (FLDR) image with the feature extraction ability of deep learning methods and the band transmission characteristics of optical filters. Qualitative and quantitative comparisons have been executed for our method and other similar ones with multiple evaluation indicators. Experimental results show that our MergeNet performs favorably against state-of-the-art HDR image reconstruction methods.

CCS CONCEPTS

• **Computing methodologies**; • **Computer graphics**; • **Image processing**;

KEYWORDS

HDR reconstruction, inverse tone mapping, Artificial augmented/virtual realities

ACM Reference Format:
Bin Liang, Dongdong Weng*, Ruikang Ju, and Lulu Feng. 2021. MergeNet: Single High Dynamic Range Image Reconstruction Method. In *2021 6th International Conference on Multimedia Systems and Signal Processing (ICMSSP 2021), May 22–24, 2021, Shenzhen, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3471261.3471266

1 INTRODUCTION

With the remarkable growth of computer calculation ability, there are more and more application scenarios of AR / VR / MR (referred to as XR). When the geometric model of virtual object remains unchanged, the fidelity of the complete scene mainly depends on the illumination consistency between the virtual objects and the real ones. The high dynamic range (HDR) environment mapping provides a good solution with greater dynamic range and irradiance contrast than low dynamic range (LDR) images. Unfortunately, the devices [1, 2] to directly generate high-quality HDR images are still too expensive to obtain.

Traditional algorithms [3, 4] recover an HDR image from multiple LDR images. However, they often produce ghost artifacts or tear marks due to pixel alignment issues. Another way is to employ inverse (reverse) tone mapping operators [5-10] (iMTO, or rTMO) to directly infer an HDR image from a single LDR image, which is an ill-posed problem. Most of them build empirical models to obtain higher dynamic range and more details from LDR images. Recently, deep learning has achieved great success in various computer vision tasks [11-15]. Owing to its strong learning ability and great structure expansion characteristics, convolutional neural networks (CNNs) are extremely suitable for those tasks.

However, there remains two main difficulties on reconstructing an HDR image from a single LDR image. Firstly, multiple-exposures LDR images are only available for static scenes to avoid artifacts or tears. Secondly, HDR images have higher dynamic range and more irradiance contrast than LDR ones, which requires a large amount of data to add credible texture, detail, contrast, and other information.

To solve the problems above, we propose the MergeNet to obtain a 32-bit HDR image from an 8-bit filtered LDR image in an end-to-end manner. Firstly, we add an optical filter in front of the camera to obtain a filtered LDR image with different exposure and luminance ranges for each channel [14, 15]. Then input the filtered image to MergeNet to obtain two types of predicted images, a logarithmic HDR image and multiple LDR images in different exposure. Finally, we merge two HDR reference images to a final HDR image by weighted blending.

In summary, our contributions as follows:

(1) An improved deep Merger Network is proposed. It can reconstruct a HDR image from a single filtered LDR image, Experiments has shown that our method has a beneficial effect on HDR images reconstruction.
(2) A complete generation pipeline by the filtered image simulation system is improved. It could create a relatively complete dataset including filtered LDR images, traditional LDR images, and logarithmic HDR images.

This paper is organized as follows. Firstly, we briefly review the relevant work in section2. and present our design ideas and details of MergeNet in section3. Next, we design several experiments to compare the performance of our method with other state-of-the-art methods. Finally, we summarize our work in section5.

ICMSSP 2021, May 22–24, 2021, Shenzhen, China
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-9037-8/21/05…$15.00
https://doi.org/10.1145/3471261.3471266

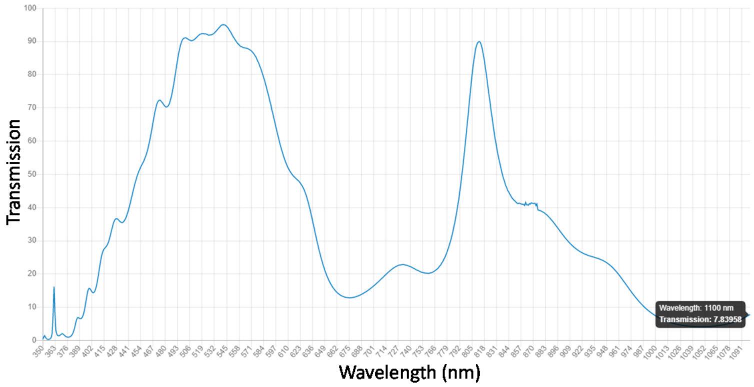

Figure 1: An optical filter from Omega. (https://www.omegafilters.com/product/3653).

2 RELATED WORK

Many HDR reconstruction methods have been proposed over the years. The traditional methods generally expand LDR to HDR. Furthermore, deep learning has also been used for similar problems. The following subsections will discuss these topics.

2.1 Optical filter

Optical filters are manufactured by vapor-depositing optical films on the glass surface to attenuate (absorb) certain wavelength bands, while reflecting (or absorb) other bands, as shown in figure 1. Brauers et al. [16] obtained High-fidelity color image with optical filters to separate the visible electromagnetic spectrum into several passbands. In the previous work [14, 15], we tried to build a single-shot FLDR acquisition system with optical filter and common camera. The validity of this system in improving image dynamic range and irradiance contrast is verified by simulation and experiments.

In this paper, we use an optical filter to perform simulation image acquisition. The RGB transmission of different optical filters is different. Here we set the ratio as $T_{rgb} = T_r : T_g : T_b = 0.6 : 0.9 : 0.3$ to verify the validity of the workflow.

2.2 Inverse tone mapping operator

Over the past decades, many research works have focused on HDR image reconstruction. The most well-known methods [3, 4] are to merge an HDR image from several multi-exposure LDR images. However, this method is primarily applicable to static scenes, which regularly produces ghosting and tearing artifacts.

The other way proposed to reconstruct an HDR image from a single LDR image is called inverse tone mapping operator (iTMO) as follows:

$$H_i = h^{-1}(L_i), \ where \ h^{-1} : [0, \ 255] \rightarrow \mathbb{R}^+ \tag{1}$$

Where, L_i denotes the $i.th$ LDR image irradiance values, h^{-1} denotes the inverse function of the tone mapping, and H_i denotes the $i.th$ HDR image irradiance values.

One of the earliest algorithms is the power function model proposed by Landis [5], who proposed the power function model to extend the luminance of LDR images. Akyüz [6] proposed the linear transformation method with gamma calibration, and found it more consistent with the visual perception of users. Meylan [7] proposed the iTMO algorithm for processing images with specular highlights. Banterle [8] proposed the expanded iTMO method with the smooth filtering algorithm to reconstruct missing information. Kovaleski et al. [9] proposed the cross-bilateral filtering enhancement algorithm. Huo et al. [10] further expanded the method without the thresholds.

2.3 Deep learning

In recent years, deep learning has been extensively used in iTMO image processing. Zhang et al. [11] propose an autoencoder to reconstruct an HDR image for outdoor environments. Eilertsen et al. [12] utilized the U-Net architecture to directly predict a log-domain HDR image. Endo et al. [13] exploited an improved U-Net architecture to indirectly predict an HDR image from multi-exposure LDR images.

Based on the previous thesis [14, 15], we propose a hybrid method with a single-shot FLDR acquisition system and MergeNet architecture to predict an HDR image from a single filtered LDR image. Specifically, after feeding a filtered LDR image, our method will obtain two types of HDR reference images, and then synthesize a final HDR image. Figure 2 displays the overview flow of our work.

3 HDR MERGER NETWORK

We believe that there is an inherent gradual relationship among the multiple exposure images, and a corresponding relationship between the LDR image and the HDR image.in order to learn those two mapping relationships, we constructed the special single-input and multi-output network structure.

3.1 FLDR images generation

In this paper, we create a special set of paired images in the form of {FLDR images, LDR images, HDR images}. The FLDR images are used as input, and other two pairs of images are regarded as ground-truth. Especially, LDR images are special FLDR ones when the RGB transmission as $T_{rgb} = T_r : T_g : T_b = 1.0 : 1.0 : 1.0$.

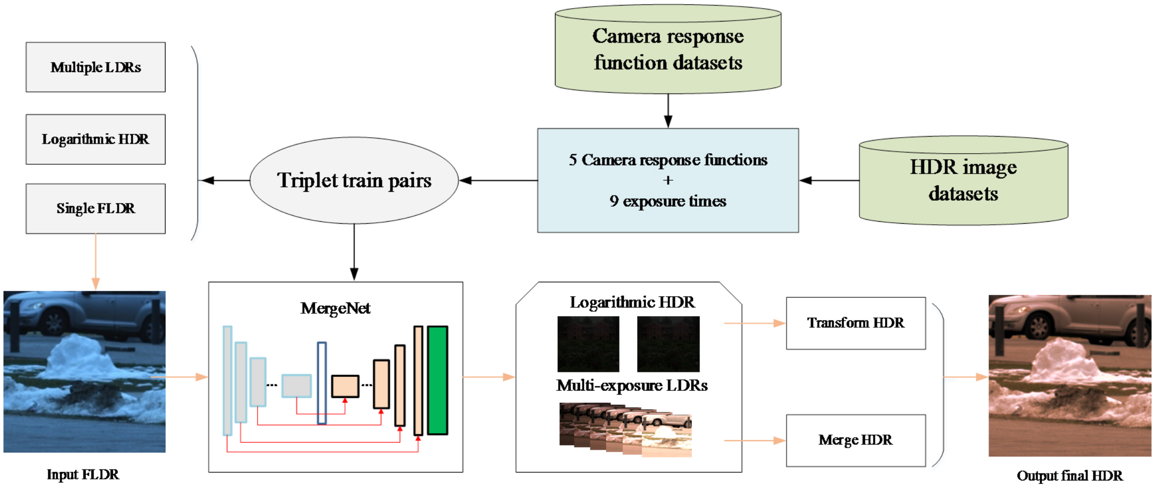

Figure 2: An overview of the proposed method. Our flow consists of generating training dataset and reconstructing HDR phases. In generating phase, the triplet pairs are produced from HDR datasets by simulating cameras. Next, we use CNNs to forecast a logarithmic HDR image and seven multi-exposure LDR images from a single filtered LDR image. The final HDR image is then produced from these images.

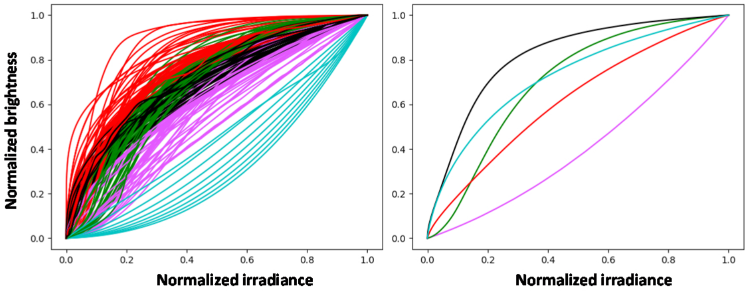

Figure 3: We choose five representative curves (right) from DoRF Curves database (left) to create training data.

Firstly, we convert original HDR images H_i^{ori} into log-domain, where μ represents the max value of the original dataset, $\odot$ represents a channel-wise product operator.

$$H_i^{log} = \frac{log_{10}\left(1 + T_{rgb} \odot H_i^{ori}\right)}{log_{10}\left(1 + \mu\right)} \tag{2}$$

Then, we simulate the camera imaging process as follows:

$$L_{i,j} = f\left(H_i^{log} \Delta t_j\right) \tag{3}$$

Where f denotes the camera response function (CRF), $\Delta t_j = \frac{1}{\tau^{T/2}}, \ldots, \frac{1}{\tau^2}, \frac{1}{\tau}, 1, \tau, \tau^2, \ldots, \tau^{T/2}$ denotes exposure duration, $L_{i,j}$ denotes pixel values of the LDR images. Here $\tau = \sqrt{2},\ T = 8$. Moreover, we use the K-means algorithm to select five representative CRFs from Database of Response Functions (DoRF) [17] (see Figure 3). All CRFs are monotonic, and normalized in the range of [0, 1].

3.2 Network architecture

The MergeNet is a fully convolutional encoder-decoder network. Given a $W \times H \times c$ LDR image (W, H, c respectively denote width, height, color channels), it generates an $N \times W \times H \times c$ four dimensions tensor which consists of a log-domain HDR image and $N-1$ multi-exposure LDR images. The encoder consists of nine layers whose filter kernels are (64, 128, 256, 512, 512, . . ., 512), the decoder also consists of nine layers whose filter kernels are (512, . . ., 512, 256, 128, 64, 3).

Meanwhile, we add skip layers follow the extension of the U-net and residual units between encoder layers and decoder layers. Batch normalization and leaky ReLU function are applied to all the convolution layers, except the final output layer, which is the sigmoid function.

3.3 Loss function

The symbolic meanings are defined as follows: D_j denotes the $j.th$ set of training pairs, H_j denotes the log-domain HDR image , $I^F_{j,i}$ denotes the $i.th$ FLDR image, $I_{j,i}$ denotes the $i.th$ LDR image.

Our MergeNet has an up-model and down-model with the same architecture but different weight parameters. Take up-model as an example, the down-model just needs to inverse the order, namely,

$$\mathcal{L}_{j,\ i} = \left\|I_{gt} - I_{ot}\right\|_1 = \left\|\left\{H_j, I_{j,i+1\to j,i+N}\right\} \oplus O_j - M_i \circ G\left(I^F_{j,i}, \theta\right)\right\|_1 \tag{4}$$

$$\mathcal{L}_{sum} = \sum_{D_j,\ j=1}^{all} \sum_{i=1}^{T} \mathcal{L}_{j,i} \tag{5}$$

Where $\|I_{gt} - I_{ot}\|_1$ is the L_1 distance between ground-truth I_{gt} and prediction I_{ot}. H_j denotes a $W \times H \times c$ log-domain HDR image, $I_{j,i+1\to j,i+N}$ denotes $(N-1)\times W\times H\times c$ LDR images. $\oplus$ and O_j are a concatenation operator and a $min\{j+N-T-1, 0\} \times W \times H \times c$ zero tensor, M_j and $\circ$ are a $N \times W \times H \times c$ mask tensor and an element-wise product operator. $G(I^F_j, \theta)$ is an $N\times W\times H\times c$ tensor when inputting a filter image I^F_j and θ is the sub-model parameters.

3.4 Merging HDR images

Input a FLDR image, the MergeNet would obtanin 2 log-domain HDR images and $2(N-1)$ multi-exposure LDR images. Firstly, we calculate the first initial HDR image from two logarithmic HDR predicted images:

$$H^{log}_j = 10^{\left(H^{up}_j + H^{down}_j\right)/2} \tag{6}$$

Then, we sort $2(N-1)$ multi-exposure LDR images from the darkest to the brightest, and average the $(N-1).th$ and $N.th$ image as a base image. Thus, we have $(2N-1)$ images. To avoid choosing the too bright or too dark one, the absolute difference between the selected image v_j and the base image v_i cannot exceed certain threshold η, i.e., $|v_j - v_i| < \eta$. Here we set $\eta = 24$. Thus, we get the second initial HDR image by Debevec et al. [3]:

$$H^{merge}_j = Debevec\ (LDRs,\ times) \tag{7}$$

Finally, we get the last HDR image from two above initial HDR images as following:

$$H^{final}_i = (1-\alpha)\, H^{log}_i + \alpha \frac{H^{merge}_i}{\max\left(H^{merge}_i\right)} \times \max\left(H^{log}_i\right) \tag{8}$$

Here we set $\alpha = 0.6$

4 EXPERIMENTS

Our MergeNet is implemented under the Chainer library using Python language on a PC with an i7 CPU, 32GB RAM, and NVIDIA GTX 1080Ti GPU. The network parameters are initialized using the pre-model of Endo et al. [13]. We adopt the Adam optimizer with momentum term of 0.5 and an initial learning rate of $5e-3$.

4.1 Dataset and Evaluation metrics

We collected 2 different HDR datasets from DML-HDR and Funt-HDR to create training pairs. The data sets not only contain a variety of videos and images, but also come from different scenarios, including indoor, garden, snowfield, and natural scenery. There are 48k pairs of images, all of which are 512×512. In this work, we use 90% of data to train network, the remaining data are used to as the test set.

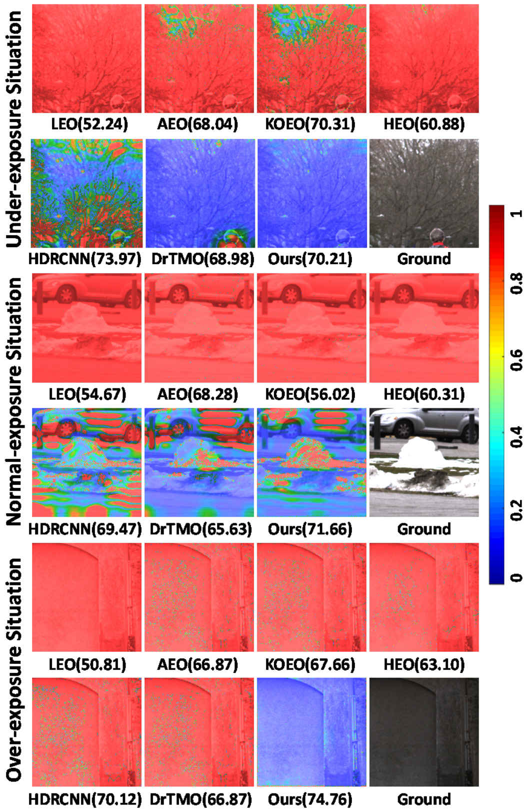

Figure 4: Visual comparison probability maps and quality scores on different exposure states.

We used five evaluation metrics—mean square error (MSE), peak signal noise ratio (PSNR), structural similarity (SSIM), feature similarity and visual difference predictor (VDP) [18]—to compare seven different HDR reconstruction methods, including Landis (LEO) [5], Akyüz et al. (AEO) [6], Kovaleski et al. (KOEO) [9], Huo et al. (HEO) [10], HDRCNN [12], DrTMO [13] and our MergeNet.

4.2 Comparisons with ground-truth

To verify the validity of our MergeNet method, we demonstrate its effectiveness between the predicted result and ground-truth. Figure 4 shows a visual qualitative comparison, and Table 1 shows a statistical quantitative comparison. The different intensities denote probabilistic errors, blue denotes low errors and red denotes high errors. Quality score is 100 for the best quality. The test set has a total number of 600 scenes, each of them consists of three different exposure situations: under-, normal-, and over-exposure. The average quality score and SSIM are 64.43 and 0.9008, respectively. The results show that the results of our MergeNet is closer to the real HDR data.

Table 1: Average values of the different exposure states for all methods. Bold values indicate the best value

Method	MSE	PSNR	SSIM	FSIM	VDP
Under-exposure					
LEO	2032.4	18.67	0.6035	0.8583	56.07
AEO	623.0	23.46	0.8669	0.9399	62.73
KOEO	567.3	22.65	0.8414	0.9455	62.51
HEO	1086.8	18.91	0.7141	0.9104	57.83
HDRCNN	428.1	24.06	0.8756	0.9554	65.77
DrTMO	528.2	23.13	0.8716	0.9535	67.20
Ours	**213.3**	**27.81**	**0.9476**	**0.9688**	**67.91**
Normal-exposure					
LEO	3464.6	14.48	0.3959	0.7268	54.73
AEO	2564.1	17.43	0.7945	0.8758	58.84
KOEO	1172.0	20.20	0.8235	0.9021	57.19
HEO	1701.8	18.46	0.7915	0.8948	58.27
HDRCNN	859.6	23.14	0.8794	0.9302	63.02
DrTMO	809.1	21.90	0.8780	0.9307	62.83
Ours	**514.1**	**26.19**	**0.9215**	**0.9558**	**64.23**
Over-exposure					
LEO	4982.6	12.20	0.3926	0.6660	53.14
AEO	7035.5	11.72	0.6366	0.7630	54.03
KOEO	12092.5	8.360	0.0280	0.5073	53.56
HEO	5549.2	13.03	0.6899	0.7994	54.76
HDRCNN	1225.6	20.46	**0.8353**	0.8942	59.75
DrTMO	1640.6	18.18	0.7867	0.8657	54.03
Ours	**1187.7**	**21.53**	0.8334	**0.8982**	**61.14**
Total-exposure					
LEO	3493.2	15.12	0.4640	0.7622	54.65
AEO	3407.5	17.54	0.7660	0.8596	58.53
KOEO	4610.6	17.07	0.5643	0.7850	57.75
HEO	2779.3	16.80	0.7318	0.8682	56.95
HDRCNN	837.8	22.55	0.8634	0.9266	62.85
DrTMO	992.6	21.07	0.8454	0.9166	61.35
Ours	**638.4**	**25.18**	**0.9008**	**0.9409**	**64.43**

4.3 Comparisons with state-of-the-art

Furthermore, we also compare our MergeNet method with other 6 iTMO methods. As shown in Figure 4, the first four methods (LEO [5], AEO [6], KOEO [9], HEO [10]) can preserve the general structure and local edges. However, they cannot recover the details of complex situations. Compared with the methods above, the last three deep learning approaches (HDRCNN [12], DrTMO [13], and MergeNet) have greatly improved the result. Among them, our approach maintains a higher quality score and a lower visual perception probability error than other existing methods.

In addition, Table 1 shows a quantitative evaluation among those methods. Compared with other methods, our MergeNet performs obviously well in almost all evaluation metrics. Although the performance of our method does not fully exceed HDRCNN [12] under the over-exposure situation, the gap is very small. In summary, the results show that our MergeNet method yields a more favorable performance than other HDR reconstruction methods.

5 CONCLUSION

This paper proposes an improved deep Merger Network to reconstruct HDR image from a single filtered LDR image with the feature extraction ability of deep learning methods and the band transmission characteristics of optical filters. The results demonstrate that our MergeNet method performs favorably against state-of-the-art HDR image reconstruction methods. Simultaneously, we have verified the validty of HDR image reconstruction with the FLDR image.

Limitation. At present, our HDR reconstruction method exhibits slight chromatic aberration [15], which can be solved by increasing the network depth and training epochs. In addition, by increasing the number of output images, the MergeNet can reconstruct a better-performing HDR image, but requires more computing power and better hardware.

Future work. Firstly, we will choose more types of HDR datasets to improve the reconstruction performance in complex exposure scenes. Next, we will reset the filter ratio more reasonable to match the physical optical filter for constructing a mixed dataset with the synthetic images and real images.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.61731003) and the Key-Area Research and Development Program of Guangdong Province (No.2019B010149001) and the 111 Project (B18005).

REFERENCES

[1] Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pajk, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., Kautz, J., Pulli, K. 2014.Flexisp: A flexible camera image processing framework. ACM Transactions on Graphics, 33(6), 231:1-231:13.

[2] Serrano, A., Heide, F., Gutierrez, D., Wetzstein, G., Masia, B. 2018. Convolutional Sparse Coding for High Dynamic Range Imaging. Computer Graphics Forum. 2016, 35(2): 153-163.

[3] Debevec P E, Malik J. 2008. Recovering high dynamic range radiance maps from photographs. ACM SIGGRAPH 2008 classes. 2008: 1-10.

[4] Tom Mertens, Jan Kautz, and Frank Van Reeth. 2007. Exposure Fusion. In Proc. of Pacific Graphics 2007: 382–390.

[5] LANDIS H. 2002. Production-ready global illumination. In SIGGRAPH Course Notes 16: 87–101.

[6] Akyüz A O, Fleming R, Riecke B E, *et al.* 2007. Do HDR displays support LDR content? A psychophysical evaluation. ACM Transactions on Graphics (TOG), 26(3), 38-es.

[7] Meylan L, Daly S, Süsstrunk S. 2007.Tone mapping for high dynamic range displays.In Human Vision and Electronic Imaging XII. International Society for Optics and Photonics, 2007, 6492: 649210.

[8] Banterle F, Ledda P, Debattista K, *et al.* A framework for inverse tone mapping. The Visual Computer, 2007, 23(7): 467-478.

[9] Kovaleski, Rafael P., and Manuel M. Oliveira. 2014. High-quality reverse tone mapping for a wide range of exposures. 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, 2014: 49-56.

[10] Huo Y, Yang F, Dong L, *et al.* 2014. Physiological inverse tone mapping based on retina response. The Visual Computer, 2014, 30(5): 507-517.

[11] Zhang, Jinsong, and Jean-François Lalonde. 2017. Learning High Dynamic Range from Outdoor Panoramas. Proceedings of the IEEE International Conference on Computer Vision. 2017: 4519-4528.

[12] Eilertsen G, Kronander J, Denes G, *et al.* 2017. HDR image reconstruction from a single exposure using deep CNNs. ACM transactions on graphics (TOG), 2017, 36(6): 1-15.

[13] Endo, Yuki, Yoshihiro Kanamori, and Jun Mitani. 2017. Deep Reverse Tone Mapping. ACM Trans. Graph., 2017, 36(6): 177:1-177:10.

[14] Liang, B., Weng, D., Bao, Y., Tu, Z., & Luo, L.. 2019. Reconstructing HDR Image from a Single Filtered LDR Image Base on a Deep HDR Merger Network. 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE, 2019: 257-258.

[15] Bin, L., Weng, D., Bao, Y., Tu, Z., & Luo, L.. 2020. Method for reconstructing a high dynamic range image based on a single-shot filtered low dynamic range image. Optics Express, 2020, 28(21): 31057-31075.

[16] Brauers, Johannes, and Til Aach. 2011. Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras. IEEE Transactions on Image Processing, 2010, 20(2): 496-505.

[17] Grossberg, Michael D., and Shree K. Nayar. 2003. What is the Space of Camera Response Functions?. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. IEEE, 2003, 2: II-602.

[18] Narwaria M, Mantiuk R, Da Silva M P, *et al.* 2015. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. Journal of Electronic Imaging, 2015, 24(1): 010501.