Deep Inverse Tone Mapping Optimized for High Dynamic Range Display

Katsuhiko Hirao
Waseda University
Tokyo, Japan
k_hirao@katto.comm.waseda.ac.jp

Zhengxue Cheng
Waseda University
Tokyo, Japan
zxcheng@katto.comm.waseda.ac.jp

Masaru Takeuchi
Waseda University
Tokyo, Japan
takeuchi@katto.comm.waseda.ac.jp

Jiro Katto
Waseda University
Tokyo, Japan
katto@katto.comm.waseda.ac.jp

***Abstract*—The popularity of high dynamic range (HDR) makes the inverse tone mapping become an important technique for HDR display. In this paper, we propose a convolutional neural network (CNN) based inverse tone mapping method to generate a high-quality HDR image from one single standard dynamic range (SDR) image. First, we present a CNN design with a three-channel input, which considers both luminance and chrominance. Second, we propose to use overlapped inputs to remove the boundary artifacts, caused by zero padding in CNN. Experimental results demonstrate the high quality of our generated HDR images compared to the ground truth and conventional inverse tone mapping methods.**

Keywords—convolutional neural networks, high dynamic range imaging, inverse tone mapping

I. Introduction

High dynamic range (HDR) image can avoid over and under exposure to improve viewing experience. One method to get an HDR image is to take bracketed standard dynamic range (SDR) images and merge them. However, this method can't be applied to existing SDR images. On the other hand, inverse tone mapping operators (iTMOs) can get an HDR image from a single SDR image. In most iTMOs, only the luminance channel is processed, and the chrominance channels are intact. However, it is expected that SDR color images can be converted to higher quality images optimized for HDR environment by SMPTE ST 2084 (ST 2084) including wide color gamut by ITU-R BT.2020 (BT.2020) and high resolution such as 3840x2160 (4K). Therefore, we propose a deep learning based method to estimate color graded images for HDR environment from one SDR image.

II. Related Work

Kuo et al. [2012] proposed a histogram based iTMO, which had different responses with different scene characteristics by including scene classification [1]. Huo et al. [2013] proposed an iTMO inspired by the property of the Human Visual System (HVS) with low computational complexity [2]. Kovaleski et al. [2014] proposed an iTMO, which uses cross bilateral filtering to compute smooth brightness enhancement functions (BEFs) that preserve sharp edges [3]. Masia et al. [2009;2015] improved inverse tone mapping based on a gamma expansion by providing a new way for automatic parameter calculation from the image statics [4,5].

The CNN based HDR imaging has also been investigated. Kalantari et al. [2017] used CNNs to merge SDR images aligned with optical flow into an HDR image [6]. Endo et al. [2017] proposed a CNN based iTMO [7]. They estimated an HDR image from a single SDR image indirectly by inferring bracketed SDR images using CNN and merging them. Eilertsen et al. [2017] also proposed a CNN based iTMO [8]. Their proposed CNN that operate in the log domain reconstruct an HDR image by estimating information of saturated area [8].

III. Proposed Algorithm

A. CNN design for Inverse Tone Mapping

Fig. 1 shows an overall architecture of our network. Inspired by [6], we use a network with four fully convolutional layers. We use rectified linear unit (ReLU) as an activation function in the layers except for the last layer. We find that it is slightly better not to use sigmoid in the last layer for activation function in our model. We apply zero padding in all convolutional layers to guarantee that the sizes of input and output in our network are the same. Learning the SDR2HDR function *F* requires the iTMO parameters of *W*. This is achieved through minimizing the loss function between an estimated HDR image *F(X;W)* and a corresponding ground truth image *Y*, where *X* is an SDR image by using stochastic gradient descent (SGD). Given a set of estimated HDR images $F(X_i;W)$ and ground truth images Y_i, we minimize the Mean Squared Error (MSE) loss:

$$\mathrm{L(W)} = \frac{1}{N}\sum_{i=1}^{N}\|F(X_i;W) - Y_i\|^2, \tag{1}$$

where *N* is the number of training samples.

B. Removal of Boundary Artifacts

In the testing phase, most CNN approaches are evaluated by down sampled images. Considering the practical use, our results are evaluated by high resolution images (4K). Due to the limitation of VRAM, we divide an input image into nine

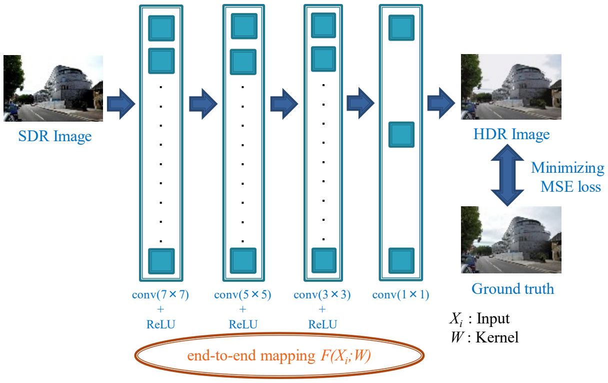

Fig. 1. The overall architecture of our network.

978-1-5386-5041-7/18/$31.00 ©2018 IEEE

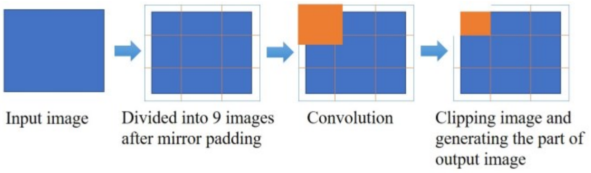
Input image
Divided into 9 images after mirror padding
Convolution
Clipping image and generating the part of output image

Fig. 2. Image splitting and merging process by overlapping and clipping.

sub-images as the inputs to our CNN. Then, we combine divided outputs into one image. However, in this way, we notice boundaries in the output images, because zero paddings in convolutional layers generate dark areas at edges of divided output images and appears as noticeable boundaries when they are combined. In order to remove the boundaries, we propose to use overlapped inputs and then clip images to proper size, which is shown in Fig. 2. We use mirror padding for input image as a pre-processing step. Thus, we remove edge areas of divided output images.

IV. Dataset

When we design CNN-based iTMOs, dataset of popular HDR formats such as Radiance HDR and OpenEXR formats was difficult to train because of the different distribution of HDR images. Instead, we use LUCORE [9] which was recently published for HDR display references. We find LUCORE is better than the dataset of HDR formats above to train CNN.

LUCORE is the UHD/HDR evaluation image set built by IMAGICA Corp. HDR and SDR images (TIFF 16bit) in LUCORE conform to BT.2020, ST 2084 and BT.2020, Gamma 2.2, respectively. HDR and SDR datasets of LUCORE have been done by optimal color grading with the display whose max luminance is 1000nits and 100nits, respectively. By using LUCORE, we can train our CNN to transform color graded images for SDR display into the ones for HDR display. Therefore, it can be expected that the estimated HDR images are optimized for HDR environment. The number of input and output channels for CNN is not one (luminance) but three (RGB) to take advantage of optimal color grading of LUCORE. HDR images of LUCORE is TIFF format, therefore, it's easier to train than images of HDR formats. Furthermore, estimated HDR images can be transformed not only to HDR formats by applying ST 2084 EOTF but also to video by using ffmpeg to make sequential output images into HDR video with HDR metadata (BT.2020 and ST 2084). We use one or two scenes of LUCORE for training and extract 40x40x3 image patches. Finally, the total number of SDR images for training is 254,200. Note that the resulting images are not included in the training data.

V. Experimental Results

TABLE I. Results of with and without our splitting process

	With splitting process	Without splitting process
PSNR	30.87	30.84
HDR-VDP-2 for SDR	47.81	47.94

TABLE II. Comparoison results with existing methods

	Kuo [2012]	Huo [2013]	Kovaleski [2014]	Maisa [2015]	Ours
HDR-VDP-2 for HDR	38.49	37.94	25.99	22.03	39.82

We perform two experiments. The first experiment is that we compare our results of with and without splitting process with ground truth HDR images for showing the effectiveness of splitting process. The second experiment is that we compare our results with several existing inverse tone mapping methods. To perform experiments using existing methods, we use the HDR Toolbox provided by Banterle at el. [2011] for HDR imaging processing [10]. As objective evaluations, we use Peak Signal to Noise Ratio (PSNR) and HDR-VDP-2 [11]. HDR-VDP-2 can be applied to not only SDR images but also HDR images to predict quality score, i.e., mean-opinion-score (MOS). We use PSNR and HDR-VDP-2 for SDR images when we compare our results with ground truth because our CNN output is first given by TIFF 16bit images and we use HDR-VDP-2 for HDR images when we compare our results with existing inverse tone mapping methods. Note that HDR-VDP-2 does not support BT.2020 and it may not work properly and is calculated with 4K resolution, 23-inch display and 0.86 meters of viewing distance. We performed the training with 30,000 iterations of which iteration counts are carefully adjusted. We find too many iterations cause the estimated HDR images to be noisy and banding artifacts to be appeared in the areas of smooth gradation. We prepare 38 images for testing and evaluate the results by calculating mean PSNR and HDR-VDP-2 of these testing images. The first experimental results are shown in Table 1. Table 1 shows that the Q score of HDR-VDP-2 for SDR images is about 48 that corresponds roughly to four of MOS score. Table 1 also shows that our patch-based splitting method does not affect the values of PSNR and HDR-VDP-2. However, when focusing the boundary artifacts area, PSNR and viewing experiences are improved significantly, as shown in Fig. 3. The results of mean PSNR and HDR-VDP-2 by comparing our results with several existing methods are shown in Table 2. Table 2 shows that the Q score of our results is better than those of other existing methods. The examples of estimated and ground truth HDR images are shown in Fig. 4. Fig. 4 shows that our HDR estimated images provide higher quality. Pixels of united output image near boundaries are not processed with continuous values, however, we can't see boundaries thanks to our splitting process method. Therefore, our approach can be applied to high resolution images regardless of VRAM limitation.

Our results show high quality, however, we find that our results of very high luminance areas such as sunlight are not so good (as shown in bottom rows of Fig. 4). Fig. 5 shows histograms of our HDR estimated image and ground truth HDR image of bottom rows of Fig. 4 that are normalized to 0-1. The data distribution of ground truth HDR image and our estimated HDR images are very similar. However, there are several pixels of which values are one in the ground truth HDR images but are not in our estimated HDR images. This may be because LUCORE does not have enough

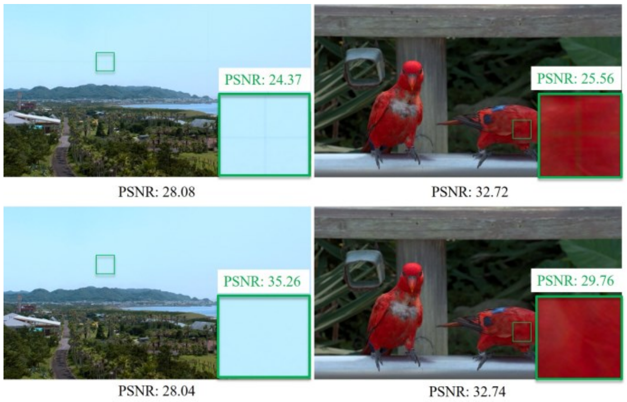

Fig. 3. Comparison results with and without our splitting process method. PSNR of overall image and the part of boundary artifacts area (3x3) are written in black letters and green letters respectively.

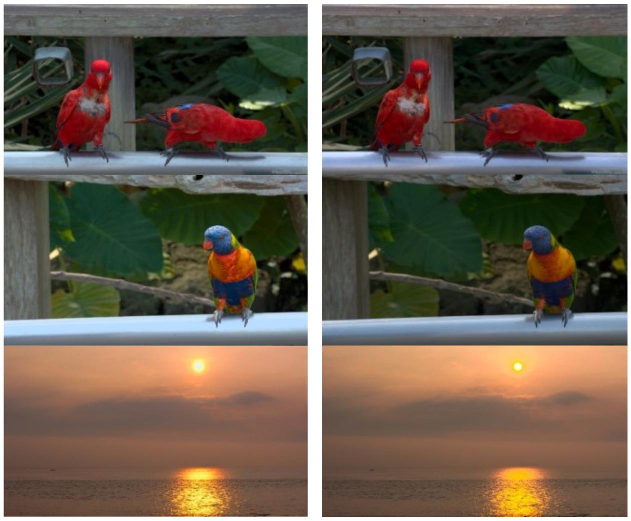

Fig. 4 Results of the estimated HDR images. Images on the left show our estimated HDR images and images on the right show the ground truth images. These images are shown to capture the frame of videos with HDR metadata. Note that the real HDR images can only be shown by HDR display

images of very high luminance areas and pixel values of these areas are not well-trained.

VI. Conclusions

In this paper, we proposed a deep learning-based method to estimate color graded images for HDR display from ones for SDR display. Our method is friendly to high resolution images owing to our proposed splitting processing and boundary removal methods. Besides, the estimated HDR images can be easily transformed not only to HDR formats by applying ST 2084 EOTF but also to HDR video by using ffmpeg to make sequential output images into HDR video with HDR metadata. We evaluate our proposed method by comparing our estimated HDR images with those of several conventional inverse tone mapping methods and ground truth HDR images using PSNR and HDR-VDP-2 as objective image quality assessment. Our estimated HDR images are better than four conventional methods in HDR-VDP-2. In our detailed investigation of estimated HDR images, we found that our method does not perform well at very high luminance areas such as sunlight because of the dataset used. However, our results are natural in most areas with 30,000 iterations, which are well-trained.

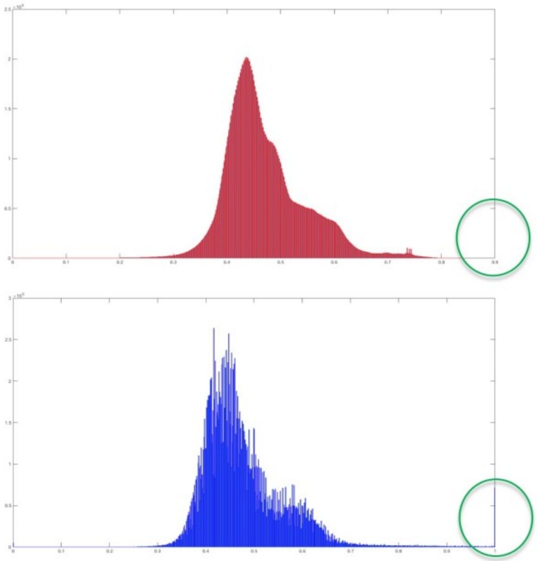

Fig. 5 The histograms of estimated HDR image and ground truth HDR images using the scene of bottom rows of Fig. 4 respectively. The highest values of ground truth HDR images expressed as sunlight are not found in our estimated HDR results.

References

[1] Kuo, Pin-Hung, Chi-Sun Tang, and Shao-Yi Chien. "Content-adaptive inverse tone mapping." Visual Communications and Image Processing (VCIP), 2012 IEEE. IEEE, 2012.

[2] Huo, Yongqing, Yang Fan, Dong Le and Vincent Brost, "Physiological inverse tone mapping based on retina response." The Visual Computer 30.5: pp. 507-517, 2014.

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

[4] Masia, Belen, Agustin Sandra, Fleming W. Roland, Sorkine Oiga and Gutierrez Diego, "Evaluation of reverse tone mapping through varying exposure conditions." ACM transactions on graphics (TOG). Vol. 28. No. 5. ACM, 2009.

[5] Masia, Belen, Ana Serrano and Diego Gutierrez, "Dynamic range expansion based on image statistics." Multimedia Tools and Applications 76.1: pp.631-648, 2015.

[6] Kalantari, Nima Khademi, and Ravi Ramamoorthi. "Deep high dynamic range imaging of dynamic scenes." ACM Trans. Graph 36.4: 144, 2017.

[7] Endo, Yuki, Yoshihiro Kanamori, and Jun Mitani. "Deep reverse tone mapping." ACM Trans. Graph 36.6, 2017.

[8] Eilertsen, Gabriel, Kronander Joel, Denes Gyorgy, Mantiuk Rafat K and Unger Jonas, "HDR image reconstruction from a single exposure using deep CNNs." ACM Transactions on Graphics (TOG) 36.6: 178, 2017.

[9] https://www.imagica.com/news/lucore/, in Japanese.

[10] Banterle, Francesco, Artusi Alessandro and Chalmers Alan, "Advanced high dynamic range imaging: theory and practice". AK Peters/CRC Press, 2011.

[11] Mantiuk, Rafat, Kim Kil Joong, Rempel G. Allan and Heidich Wolfgang, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions." ACM Transactions on graphics (TOG). Vol. 30. No. 4. ACM, 2011.