IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 22, 2020, accepted June 18, 2020, date of publication June 25, 2020, date of current version July 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005022

Exposure-Structure Blending Network for High Dynamic Range Imaging of Dynamic Scenes

SANG-HOON LEE, HAESOO CHUNG, AND NAM IK CHO, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, INMC, Seoul National University, Seoul 08826, South Korea

Corresponding author: Nam Ik Cho (nicho@snu.ac.kr)

This work was supported by the SK hynix Inc.

ABSTRACT This paper presents a deep end-to-end network for high dynamic range (HDR) imaging of dynamic scenes with background and foreground motions. Generating an HDR image from a sequence of multi-exposure images is a challenging process when the images have misalignments by being taken in a dynamic situation. Hence, recent methods first align the multi-exposure images to the reference by using patch matching, optical flow, homography transformation, or attention module before the merging. In this paper, we propose a deep network that synthesizes the aligned images as a result of blending the information from multi-exposure images, because explicitly aligning photos with different exposures is inherently a difficult problem. Specifically, the proposed network generates under/over-exposure images that are structurally aligned to the reference, by blending all the information from the dynamic multi-exposure images. Our primary idea is that blending two images in the deep-feature-domain is effective for synthesizing multi-exposure images that are structurally aligned to the reference, resulting in better-aligned images than the pixel-domain blending or geometric transformation methods. Specifically, our alignment network consists of a two-way encoder for extracting features from two images separately, several convolution layers for blending deep features, and a decoder for constructing the aligned images. The proposed network is shown to generate the aligned images with a wide range of exposure differences very well and thus can be effectively used for the HDR imaging of dynamic scenes. Moreover, by adding a simple merging network after the alignment network and training the overall system end-to-end, we obtain a performance gain compared to the recent state-of-the-art methods.

INDEX TERMS Computational photography, convolutional neural network, high dynamic range imaging.

I. INTRODUCTION

Dynamic ranges of standard cameras are too narrow when compared with those of most scenes around us. Also, they cannot capture too bright or dark regions that have illumination values out of the ranges of normal camera-settings. Thus, high dynamic range (HDR) imaging, which is a technique to capture and express HDR scenes in a single image, has been studied to overcome the limitation. The most common approach is to take a sequence of low dynamic range (LDR) images with different exposures and fuse them to an HDR image [7], [13]. Then, the HDR image is appropriately tonemapped to the display dynamic range [8], [35], [43]. Another approach is directly synthesizing a tonemapped-like image as a weighted sum of LDR images, which is called exposure fusion [31], [34], [39].

However, the classical HDR imaging methods produce plausible results only when the camera is fixed on a tripod, and when all the objects in the scene do not move, which are too impractical or limited conditions. In a practical situation, there are misalignments in the multi-exposure photos, due to background and foreground motions. If the misaligned photos are merged into an HDR image, most regions in the result are blurred due to the background motion, and there can be ghost artifacts around the moving objects. To alleviate these problems, advanced HDR imaging methods include the algorithm to align the input images to the reference that is usually the image with medium exposure. For some examples, the alignment is implicitly or explicitly conducted using patch matching [18], optical flow [23], homography transformation [52], or attention module [53]. However, the alignment methods

The associate editor coordinating the review of this manuscript and approving it for publication was Yong Yang.

do not produce satisfying results when there are complicated background differences, nonrigid motion, or illumination difference, which results in low-quality results, sometimes with fatal artifacts. Thus, we still need efforts to improve the alignment step before merging the multi-exposure images because the alignment quality is critical for reducing the blurring and ghost artifacts.

In this respect, we propose an end-to-end convolutional neural network (CNN) that first generates well-aligned multi-exposure images and then merges the aligned images. It has been noted that explicitly aligning differently exposed images is not an easy task, as evidenced in the previous works. Thus we focus on the alignment part while we use a simple network for the merging part. For synthesizing well-aligned differently exposed images, our main idea is to blend the images at the deep-feature-domain, rather than to blend or transform the images at the pixel- or shallow-feature-domain. Then the blended features are used to construct an aligned image. Precisely, we design a building block called Exposure-Structure Blending Network (ESBN), which generates an under- or over-exposure image that is structurally aligned to the reference image. The ESBN consists of two encoders that extract deep features from two differently exposed images separately, several residual blocks that blend the deep features, and a decoder that constructs the aligned image. If we have N LDR exposure images, then we use $N-1$ ESBNs to generate all the aligned exposure images. Then, the aligned images are fed to a simple merging network, which consists of a densely connected residual block.

In the experiment, we first train the ESBN and merging network separately to show the effectiveness of ESBN over the conventional alignment methods. We show that the ESBN works well for quite large exposure differences, and thus any existing (even non-dynamic) HDR imaging methods can also use the outputs of our ESBN for generating a plausible HDR image. Also, we train the overall network end-to-end to validate a performance gain of joint alignment and merging networks. From the extensive qualitative and quantitative comparisons, it is shown that separate training of ESBN and merging network produces better results than existing state-of-the-art methods, and the end-to-end training yields even better performance.

The main contribution of our work is the ESBN, which blends the exposure and structure information from differently exposed and dynamically changing images, and as a result, generates multi-exposure images that are structurally aligned to the reference. The aligned images can also be used as the inputs to existing (non-)dynamic HDR imaging methods. The other contribution is an end-to-end HDR imaging network that combines the alignment and a simple merging network, which provides state-of-the-art performance.

II. RELATED WORK

The conventional HDR imaging methods generate a final HDR image H as a weighted sum of input LDR images in the HDR domain (generally in the 32-bit RGBE format), which can be expressed as

$$H = \sum_{n=1}^{N} W_n H_n, \tag{1}$$

where N is the number of LDR images, H_n is the n-th input image in the HDR domain, and W_n denotes the weight corresponding to H_n. In dynamic scenes, however, the methods face limitations such as blurring or ghost artifacts due to background and/or foreground motions. To overcome the limitations, some methods detect regions with motions in H_n and then reject them by reducing W_n. Also, other methods first align H_n and then compute W_n. In recent years, patch-based optimization and CNN-based learning methods have also been proposed for the better HDR generation.

A. REJECTING REGIONS WITH MOTIONS

The first approach is to detect regions with foreground motions in input images and reject them in the merging process. The methods with this approach assume that multi-exposure images have no background motions or they are already aligned globally, where each of the methods has a different way of measuring the motion. For some examples, Jacobs *et al.* [20] computed the local entropy of input images to detect regions with motions. Heo *et al.* [16] and An *et al.* [1], [3] computed the correlation between the images to reject the moving regions. Zhang and Cham [55] analyzed the magnitude and orientation of gradients to classify regions with and without motions. An *et al.* [2] also attempted to find the moving regions by measuring the zero-mean normalized cross-correlation. Lee *et al.* [29] and Oh *et al.* [40] used rank minimization to find outlier regions, and Yan *et al.* [54] synthesized ghost-free HDR images based on the sparse representation. However, these methods are not appropriate for the scenes with large background motions because they ignored some of the misaligned regions that might contain valuable information for improving the results. We believe that it is essential to have aligned images to use the information from different exposures as much as possible, in contrast to the methods that discard the moving regions.

B. ALIGNMENT BEFORE MERGING

The second approach is to align input images to the reference that is usually the image with medium exposure or the best-exposed image chosen by quality measures. The images having the background and/or foreground motions compared to the reference are aligned using geometric transformation and/or optical flow. For some specific examples, Ward *et al.* [51] aligned backgrounds using translation transformation, and Tomaszewska and Mantiuk [49] used homography transformation. However, since these methods cannot deal with foreground motions, Bogoni [5] additionally used optical flow to align moving objects. Also, Kang *et al.* [24] computed the optical flow using a method in [33] to estimate local motions. Zimmer *et al.* [57] estimated the optical flow by minimizing an energy function that includes gradient and smoothness terms. Hu *et al.* [17] found regions with dense correspondences between the images to align them. These methods work well in general but not in challenging cases such as nonrigid motion. Hence the final HDR images sometimes show the artifacts. We believe that having well-aligned multi-exposure images is the most important for the successful HDR imaging of dynamic scenes, and thus focus on the generation of aligned multi-exposure images in the deep learning framework.

C. PATCH-BASED RECONSTRUCTION

The third approach is to reconstruct an HDR image patch by patch, unlike previous pixel-wise methods. It first finds patches with dense correspondences between the input and reference, and then reconstructs patches of final HDR image by solving an optimization problem. Specifically, Sen *et al.* [46] reconstructed an HDR image by restoring missing information in saturated regions of the reference image using other input images. Hu *et al.* [18] proposed a patch-based reconstruction of multi-exposure images, which are then used as the input to the existing merging methods. These methods generally yield high-quality HDR images with less artifact, but they require high computational complexity due to the repetition of the optimization for all patches.

D. DEEP-LEARNING-BASED METHODS

Based on many successful CNN-based image restoration methods, there have also been several HDR imaging methods using deep networks. They take a stack of LDR images as input and produce an HDR image using a CNN, sometimes with a pre-processing step such as the optical flow or homography transform. Precisely, Kalantari and Ramamoorthi [23] aligned input images using an optical flow method in [32] and merged the aligned images into an HDR image using a CNN. Wu *et al.* [52] first aligned background motions using homography transformation and merged the aligned images using an encoder-decoder architecture. Yan *et al.* [53] used attention modules to detect useful regions and misaligned regions, and adopted dilated residual dense blocks to merge the images. These CNN-based methods not only merge input LDR images into an HDR image but also restore the information of saturated regions. Unlike the previous methods, we design a network that generates an aligned image along with a merging network. Specifically, the alignment network generates an image with a certain exposure to have the same structure as the one with different exposure. Compared to the previous methods that perform image alignment by optical flow or homography transform, our network can be trained end-to-end along with the merging network, and hence performs faster at the inference. Also, compared to the attention module that implicitly aligns the image by focusing on similar areas across the images, our method generates accurately aligned images by blending the information from multi-exposure images. Hence, our alignment network can also be used as an efficient pre-processing step for other HDR imaging methods.

E. SINGLE-IMAGE HDRI

Unlike previously mentioned methods that use bracketed multi-exposure images as the input, there is another approach that generates an HDR image from a single LDR input. Specifically, some of the methods belonging to this approach generate virtual multi-exposure images from the input and then blend them with the appropriate weight maps [11], [14], [21], [26], [41], [42], [48], [50]. Also, there are reverse tone mapping operators (rTMOs) that map the LDR to the HDR as presented in [4], [19], [27], [37], [38]. Recently, CNN-based single-image HDR and rTMO have also been presented in [6], [9], [10], [23], [30]. Specifically, a CNN-based exposure fusion was introduced in [23]. Also, a CNN-based rTMO was also proposed in [10], where they generate virtual multi-exposure images by using a CNN, and then fuse them for generating the HDR image. In [9], they focused on the saturated regions and predicted the contents of those regions to generate an HDR image. Also, a GAN-based inverse tone mapping was proposed in [30]. There is also a method that jointly performs HDRI and SR from a single input [47]. In summary, the purpose of single-image HDR or rTMO is to generate an HDR image from an LDR input, whereas ours does from multiple LDR images with different exposures. These methods have advantages in that there are no ghost artifacts in the output because they use a single input. However, when we have well-taken or well-aligned multi-exposure images, multi-input HDRI methods would perform better than the single-input HDRI because more information is available from the multi-exposure inputs, especially in the dark and washed-out regions [53].

III. PROPOSED METHOD

A. OVERALL PIPELINE

The overall HDR imaging architecture is a cascade of an alignment network and a merging network, as shown in Fig. 1. The alignment network consists of two ESBNs, one for generating the under-exposure image with the structure of the reference and the other for over-exposure. Then, the aligned and reference images are stacked and forwarded to the merging network, which consists of a one-step residual network. The alignment and merging networks can be trained separately, or the overall network can be trained end-to-end, where the latter yields better performance as expected.

Formally describing the process in Fig. 1, our network takes as input a set of three LDR images $\mathcal{I} = \{I_1, I_2, I_3\}$, and outputs an HDR image H. The input images I_1, I_2, and I_3 have low, medium, and high exposures, respectively, and they are presumed to have both background and foreground motions. We set the image with medium exposure I_2 as the reference, and the ESBNs generate the aligned multi-exposure images $\{I_1^a \text{ or } I_3^a\}$ from $\{I_1 \text{ or } I_3\}$ and I_2. As a result, the alignment

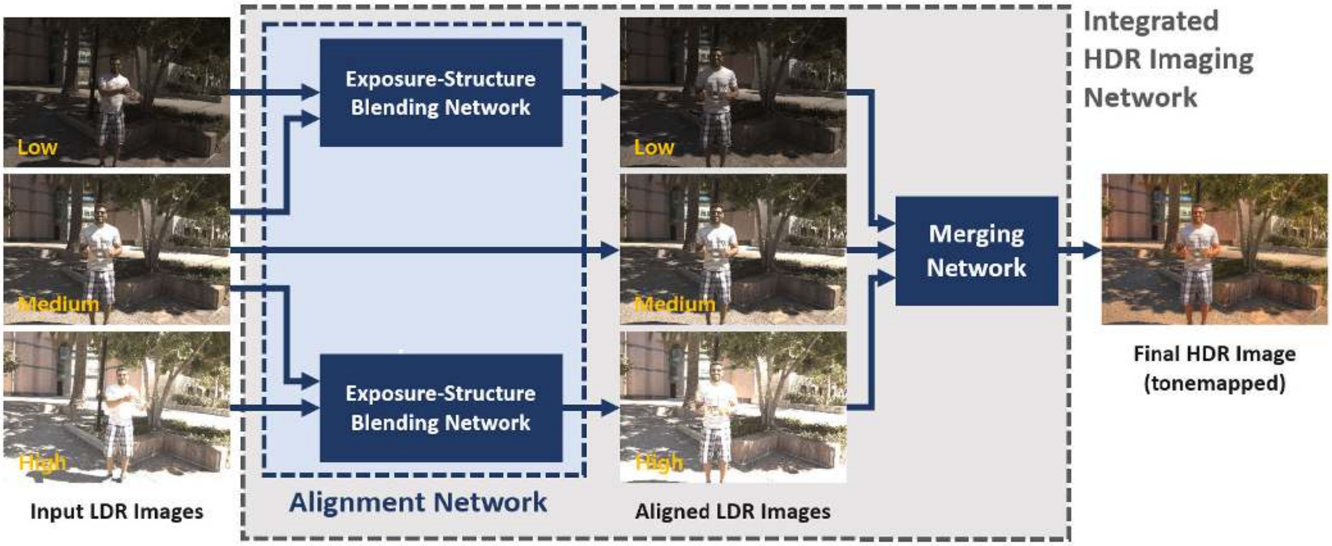

Low
Medium
High
Input LDR Images
Exposure-Structure Blending Network
Exposure-Structure Blending Network
Alignment Network
Low
Medium
High
Aligned LDR Images
Merging Network
Integrated HDR Imaging Network
Final HDR Image (tonemapped)

FIGURE 1. The overall pipeline of our HDR imaging method. First, structurally aligned multi-exposure images are generated in the Alignment Network, which consists of exposure-structure blending networks (ESBNs). Then, the aligned images are forwarded to the merging network, which generates the final HDR image.

network gives a set of aligned LDR images $\mathcal{I}^a = \{I_1^a, I_2^a, I_3^a\}$, where $I_2^a = I_2$, and the merging network generates H from $\mathcal{I}^a$.

B. ALIGNMENT NETWORK

Some recent HDR imaging methods explicitly align input LDR images using optical flow [23] or homography transformation [52]. Both methods align the background motions well, but the former produces some artifacts in the nonrigid foreground objects. The latter has the limitation that it can align only one homography transformation, *i.e.,* the homography either for background or a large planar foreground object. In addition to these shortcomings, they are difficult to be parallelized and take long computation times. For example, the alignment process using the optical flow in [23] takes a half minute approximately.

To alleviate this problem, we propose a CNN that generates the aligned image from the ones with foreground/background motions and also with a large difference in illumination. When we wish to synthesize an image with the illumination of I_s (namely the *source* image), $s = 1$ or 3, that have the same structure as I_r (*reference*), $r = 2$, we may consider the architectures shown in Fig. 2: (1) Use only I_s as the input, and the CNN geometrically transforms the structure of I_s to the ground truth (that has the illumination of I_s and the structure of I_r). That is, this CNN is the geometric transformer. (2) Conversely, only I_r is input to the CNN, which changes its illumination to that of I_s so that it becomes close to the ground truth. The role of this CNN is just to change the illumination of I_r. (3) For increasing the information to be used in the CNN, we use both I_r and I_s as the input to the CNN. This network blends the images starting from the spatial domain and finally construct the image similar to the ground truth. (4) I_r and I_s go through different encoders so that their features are separately extracted, and they are blended in the feature domain and reconstructed to be close to the ground truth.

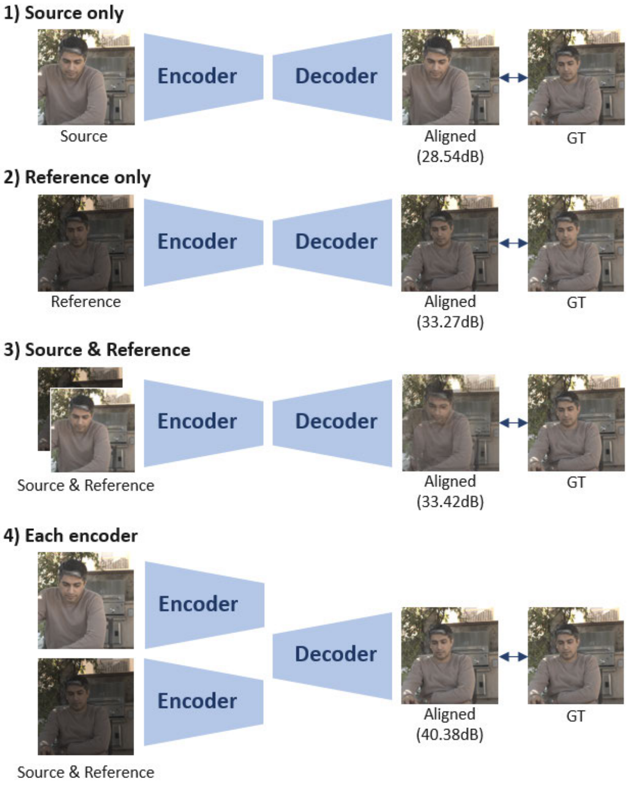
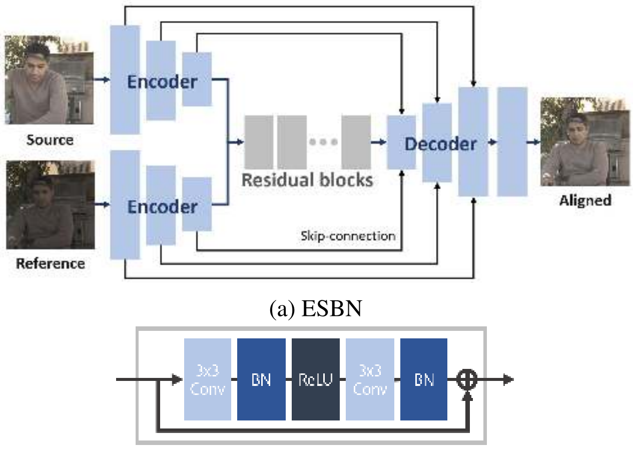


FIGURE 2. Four different architectures of CNN to generate the different exposure image aligned to the reference.

Fig. 2 also shows the result of each method, showing the closeness of the output to the ground truth in terms of PSNR.



(a) ESBN

(b) Residual block

FIGURE 3. (a) The details of ESBN. It consists of two encoders which extract the features from their input images, and residual blocks that blend the features, and a decoder that reconstructs the aligned image from the features. It also has skip-connections between the layers. (b) The details of the residual block in ESBN. It consists of two 3 × 3 convolution layers followed by batch normalizations and a ReLU.

As expected, using both I_r and I_s, and blending them in the feature domain synthesizes the best aligned image, which is our ESBN. The details of ESBN is shown in Fig. 3, which is inspired from the performances of *Unet* [44] and its advanced version, *Image Transformation Network* [22]. While they take a single image as the input of their encoder-decoder architecture and extract features using an encoder, the ESBN takes two images, I_s and I_r, and has two corresponding encoders to extract features from the images. Then, they are blended through several convolution layers and reconstructed by a decoder. Formally, the ESBN is described as a function

$$\hat{I}_s^a = f_{E_s}(I_s, I_r), \tag{2}$$

which means that the source I_s and the reference I_r are blended through the network f_{E_s} to generate the output $\hat{I}_s^a$.

To be more precise with the structure, each encoder has three encoding layers that are 5 × 5 convolution layers with the stride of 2, followed by batch normalizations and leaky ReLUs. The first layer of the encoder produces the feature map with 64 channels, and the second and third 128 and 256, respectively. The two encoders have their own parameters instead of sharing the same parameters. The feature maps from the upper and lower encoders are concatenated and then fed to the residual blocks [15]. There are nine residual blocks in total, each of which consists of 3 × 3 convolution layers, batch normalization, and a ReLU. Finally, the features are fed to the decoder to generate the aligned image. The decoder consists of three deconvolution layers with the stride of 1/2, which are followed by batch normalizations and ReLUs. Finally, there is an output layer, which is a 5 × 5 convolution layer with the stride of 1, followed by tanh(·).

The ESBN $f_{E_s}(I_s, I_r)$ is trained by minimizing ℓ_2 distance between the aligned source and its ground truth, *i.e.,* by minimizing the loss function defined as

$$\mathcal{L}_{E_s} = \|\hat{I}_s^a - I_s^a\|_2, \tag{3}$$

where I_s^a is the ground truth for $\hat{I}_s^a$. We also tried ℓ_1 loss, but the ℓ_2 loss produces quantitatively better aligned images (See Section. V-B1).

As shown in Fig. 1, two ESBNs, $f_{E_1}(I_1, I_2)$ for under-exposure and $f_{E_3}(I_3, I_2)$ for over-exposure, constitute the alignment network. The role of the alignment network is to produce under- and over-exposure images whose structures are aligned to that of the reference. Formally, the output from the alignment networks are summarized as

$$\hat{I}_1^a = f_{E_1}(I_1, I_2), \tag{4}$$

$$\hat{I}_3^a = f_{E_3}(I_3, I_2), \tag{5}$$

$$and \quad \hat{I}_2^a = I_2. \tag{6}$$

C. MERGING NETWORK

Recent HDR imaging methods [23], [52] reconstruct the final HDR image after aligning the input LDR images. But, since the aligned images still have artifacts or misalignments, they used somewhat complex networks to compensate for the misalignments while merging the images. On the other hand, since our ESBN works better than the optical flow or homography transformation, as will be demonstrated in the experiments, we use a simple network for the merging.

The proposed merging network extracts features from three aligned LDR images and reconstructs the final HDR images, as shown in Fig. 4. We adopt the residual learning with a residual dense block as in [28], [56], while the original residual dense network includes three residual dense blocks. Each aligned image passes through 3 × 3 convolution layers and leaky ReLU, resulting in 64 feature maps. All the features are concatenated into 192 (64 × 3) channels, which are again reduced to 64 by using the 1 × 1 convolution layer. Then, it is fed to the residual dense block, which consists of six convolution layers followed by leaky ReLU. Finally, the HDR image is produced from the last output layer, which is a 3 × 3 convolution layer followed by tanh(·).

For the residual learning, the output layer takes as input the sum of feature maps from the residual dense block and the reference. Formally, the merging network works as

$$\hat{H} = f_M(\mathcal{I}^a), \tag{7}$$

where $\hat{H}$ is the estimated HDR image and $\mathcal{I}^a$ is the set of aligned LDR images. The network is trained by minimizing ℓ_1 distance between the tonemapped estimated HDR image and the tonemapped ground truth HDR image, and the loss function $\mathcal{L}_M$ is defined as

$$\mathcal{L}_M = \|\mathcal{T}(\hat{H}) - \mathcal{T}(H)\|_1, \tag{8}$$

where H is the ground truth HDR image and $\mathcal{T}$ denotes the tonemapping operation. Since HDR images need to be tonemapped to the range of the targeting display, we compute the loss function in LDR domain instead of HDR domain

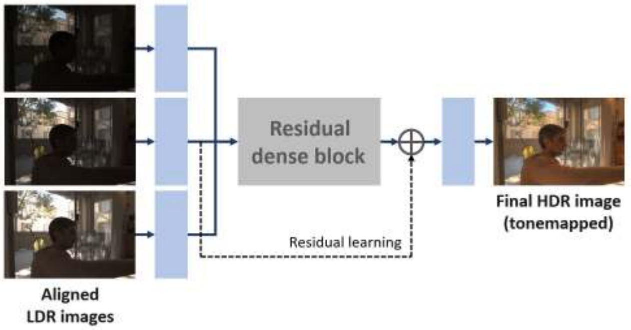


FIGURE 4. The architecture of the proposed merging network. The network extracts the features from the aligned LDR images and reconstructs the final HDR image. We use a residual dense block and apply a residual learning to produce high-quality HDR images.

as stated in [23]. We use a differentiable tonemapper, μ-law, which is defined as

$$\mathcal{T}(H) = \frac{\log(1+\mu H)}{\log(1+\mu)}, \tag{9}$$

where μ is set to 5,000.

D. INTEGRATED HDR IMAGING NETWORK

The alignment network and merging network can be trained separately. However, it will be shown that end-to-end learning works better than the separate training, by minimizing the overall integrated loss

$$\mathcal{L}_I = \mathcal{L}_{E_1} + \mathcal{L}_{E_3} + \mathcal{L}_M. \tag{10}$$

IV. DATASETS

A. KALANTARI DATASET AND GROUND TRUTH ALIGNED IMAGES

For training the HDR imaging networks, we need multi-exposure images and ground truth HDR images. There is a dataset with ground truth image for the static scene as in [12], but it is not easy to construct the dataset for the dynamic scenes that the camera and objects are moving. Recently, Kalantari *et al.* [23] made such a well-prepared dataset, where each set contains three dynamic exposure photos and a ground truth HDR image. Camera response functions of all multi-exposure LDR images in the dataset are calibrated using a gamma function. Hence, we train and test our network with the Kalantari dataset. However, in our case, we further need ground truth aligned images I_1^a and I_3^a for training the alignment network. Thus, we create them from the ground truth HDR image. Specifically, we darken or lighten the illumination of the ground truth HDR image for each exposure by the gamma function used in [23].

B. PREPROCESSING

The ESBN takes six-channel images as the input, while general networks for image processing take three-channel images. In [23], they showed that the performance of their network could be improved by using both LDR images and linearized HDR ones as the input to the network. The reason for this result is that LDR images are suitable to extract features and to detect saturated regions, while HDR images contribute to improve the robustness of the network for the wide range of exposures. We follow this method for preparing the input to the network. Specifically, though Fig. 3 illustrates that the ESBN takes two images (over-exposed and medium-exposed), each image is actually the concatenation of LDR and its corresponding HDR images like [23].

C. PATCH GENERATION

Since a large number of training examples are required to train our networks, cropped patches instead of original images are used for the training. Specifically, 256 × 256 overlapping patches are cropped from 1500 × 1000 original images with a stride of 64, and around 18K patches are generated. Then, they are increased to 140K by data augmentation such as rotation and flipping. Furthermore, patches without background or foreground motions are removed, and the remaining 133K patches are finally used to train the networks efficiently.

V. EXPERIMENTAL RESULTS

A. EVALUATION METRICS

For the quantitative evaluation, we use three metrics, PSNR, SSIM, and HDR-VDP-2 [36]. Specifically, we use PSNR and SSIM to evaluate the aligned results, and we use HDR-VDP-2, PSNR, and SSIM to show the quality of the final HDR image. When computing PSNR and SSIM of HDR results, we compute them both in HDR domain (PSNR-L and SSIM-L) and LDR domain (PSNR-T and SSIM-T) after tonemapping.

B. ABLATION STUDIES

1) COMPARISON OF LOSS FUNCTIONS

We try both ℓ_1 and ℓ_2 loss for training the networks because they have pros and cons, which is not totally predictable. Hence, we need to choose one of them experimentally. In general, the ℓ_1 is more robust to outliers than the ℓ_2, because the outliners make the squared values of differences (ℓ_2) too large. On the other hand, the ℓ_2 is known to work more stable when pairs of training images do not have significant differences [45].

We first compare the performances of the proposed alignment network and merging network trained with different loss functions. Our alignment network is trained using ℓ_1 and ℓ_2 loss, and the alignment results on Kalantari test sets [23] are listed in Table. 1. The results indicate that the alignment network trained using ℓ_2 yields better performance for the alignment. Similarly, the merging network is trained using ℓ_1 and ℓ_2, and the reconstructed HDR results are compared in Table. 2. It can be seen that ℓ_1 shows better performance in the case of the merging network.

2) COMPARISON OF SEPARATE AND JOINT TRAINING

In Section. III-D, we explained that our alignment network and merging network can be trained separately, or jointly

TABLE 1. Comparisons of the alignment network trained with different loss functions. The best results are shown in boldface.

	PSNR	SSIM
ℓ_1 loss	39.5832	0.9815
ℓ_2 loss	**40.3849**	**0.9880**

TABLE 2. Comparison of the merging network trained with different loss functions. The best results are shown in boldface.

	PSNR-T	SSIM-T	PSNR-L	SSIM-L
ℓ_1 loss	**44.0313**	**0.9914**	**41.1796**	**0.9871**
ℓ_2 loss	43.6762	0.9860	40.7847	0.9845

TABLE 3. Comparison of the separate and joint training of the proposed network. The best results are shown in boldface.

	PSNR-T	SSIM-T	PSNR-L	SSIM-L
Seperate	44.0313	0.9914	41.1796	0.9871
End-to-end	**44.5478**	**0.9930**	**41.3625**	**0.9877**

trained as a single network. In this subsection, we compare the results of these training methods in Table. 3, which shows that the joint training brings performance gains.

C. COMPARISONS WITH STATE-OF-THE-ART METHODS

1) IMAGE ALIGNMENT

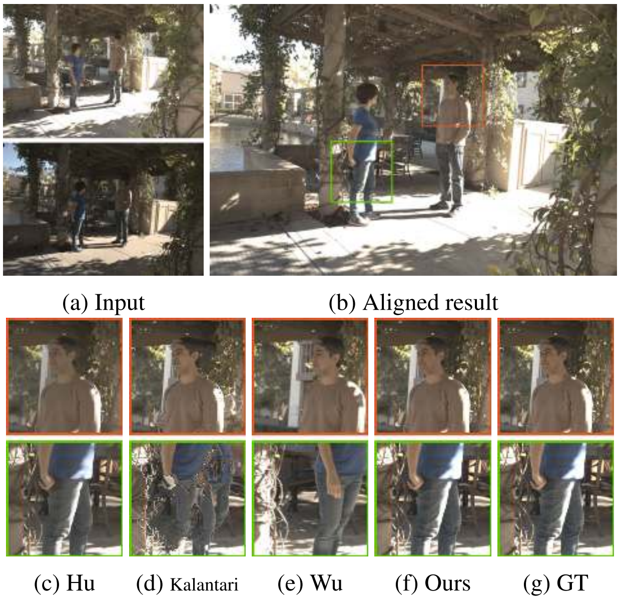
We first compare the alignment accuracy by ESBN and other alignment methods employed in previous HDR imaging methods on the test sets in Kalantari dataset. In Fig. 5, the source image with high exposure is aligned to the reference using the proposed and other alignment methods. It can be seen that the method of Hu *et al.* [18] using patch matching reconstructs the aligned image without artifacts, but it has low contrast and loses details. The method of Kalantari *et al.* [23] using optical flow failed to estimate motions and produces some artifacts in the foreground regions. The method of Wu *et al.* [52] using homography transformation does not align the foreground motions, while it aligns the background. Our alignment network generates the aligned image very close to the ground truth image. We also compute PSNR and SSIM of the aligned results in Table. 4 to compare the performance quantitatively. It can be seen that our alignment method shows better performance than the others.

TABLE 4. Comparisons of alignment results by ESBN and other alignment methods. The best results are shown in boldface.

	Hu [18]	Kalantari [23]	Wu [52]	Ours
PSNR	34.4117	32.5710	28.3794	**40.3849**
SSIM	0.9683	0.9351	0.8614	**0.9880**

2) HDR RECONSTRUCTION

We compare the HDR imaging results of our method with several state-of-the-art methods on test images in Kalantari dataset. In Fig. 6, we visualize the reconstructed HDR images of various methods, which are tonemapped with Photomatix to be visualized on the LDR displays. The patch-based reconstruction methods by Sen *et al.* [46] and Hu *et al.* [18] do not well reconstruct the saturated regions on the branches and the wall. They fail to find correct corresponding patches in the saturated regions and reconstruct the regions with the patches in the sky. The CNN-based methods of Kalantari *et al.* [23] and Wu *et al.* [52] reconstruct the saturated regions in some degree, but they fail to reconstruct the details of the shadow of the tree on the wall. On the contrary, our HDR imaging network can reconstruct both the saturated regions and the details. The results of Yan *et al.* [53] are not shown here because the source code is not yet provided.



(a) Input (b) Aligned result

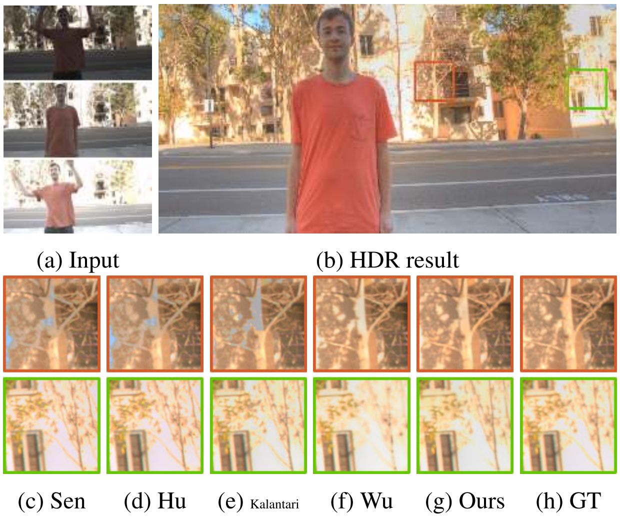
(c) Hu (d) Kalantari (e) Wu (f) Ours (g) GT

FIGURE 5. Comparsions of our ESBN and other alignment methods. (a) A source image and a reference image. (b) The aligned image using our ESBN. Magnification of red and green boxes by (c) Hu *et al.* [18], (d) Kalantari *et al.* [23], (e) Wu *et al.* [52], (f) our method, and (g) the ground truth.

In Table. 5, we also compute PSNR, SSIM and HDR-VDP-2 between the HDR results and the corresponding ground truth images to evaluate the methods quantitatively. It can be seen that our network with separate training of alignment and merging network provides better performance than the others. Moreover, the joint training of our network brings a performance gain. We quote the quantitative evaluation results of Yan *et al.* [53] from their paper, since their source code is not available.

We also compare the running times in Table. 6, by executing the source codes[1] provided by the authors. The codes were run on a PC with i7-4790 and Titian XP, for the LDR images of 1500 × 1000. The execution time for image alignment is measured separately, except for Sen and Yan that directly reconstruct the HDR image. It can be seen that our alignment process has the shortest running time because it is performed by the GPU, while other alignment methods are performed by the CPU. Furthermore, our merging net-

[1] Since the source code of the alignment using homography transformation in [52] is not provided, we implement the alignment code which can generate similar results.



(a) Input (b) HDR result

(c) Sen (d) Hu (e) Kalantari (f) Wu (g) Ours (h) GT

FIGURE 6. Comparsions of proposed and recent HDR imaging methods. (a) A set of input LDR images. (b) HDR result of our method. Magnification of red and green boxes by (c) Sen *et al.* [46], (d) Hu *et al.* [18], (e) Kalantari *et al.* [23], (f) Wu *et al.* [52], (g) our method, and (h) the ground truth.

TABLE 5. Comparisons of HDR results of the proposed method with state-of-the-art methods. The best results are shown in boldface.

	PSNR-T	SSIM-T	PSNR-L	SSIM-L	HDR-VDP-2
Sen [46]	41.11	0.9815	38.82	0.9749	59.38
Hu [18]	34.87	0.9701	31.72	0.9520	57.05
Kalantari [23]	42.70	0.9878	41.21	0.9845	62.98
Wu [52]	41.65	0.9857	40.88	0.9849	62.70
Yan [53]	43.62	-	41.03	-	62.30
Ours (seperate)	44.03	0.9914	41.18	0.9871	63.02
Ours (joint)	**44.55**	**0.9930**	**41.36**	**0.9877**	**63.24**

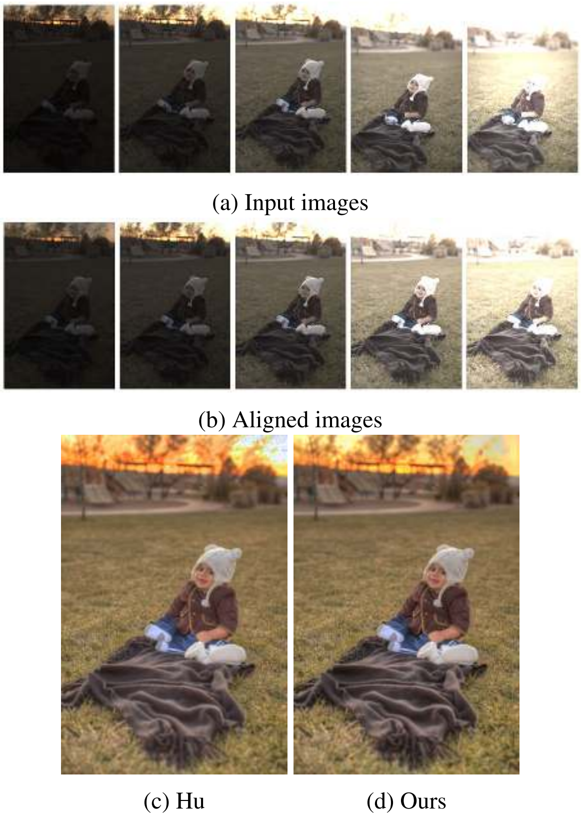
TABLE 6. Comparisons of average execution times of the proposed method with state-of-the-arts.

	Alignment(s)	Merging(s)	Total(s)
Sen [46]	-	-	312.853
Hu [18]	185.722	1.712	187.434
Kalantari [23]	29.893	1.558	31.451
Wu [52]	0.501	0.239	0.740
Yan [53]	-	-	0.320
Ours	0.231	0.172	0.403

work takes less time than those of Kalantari *et al.* [23] and Wu *et al.* [52] because our network has a simple structure.

D. APPLICATION TO THE CASE OF MORE NUMBERS OF EXPOSURES

Note that Fig. 1 shows the case of three-exposures ($N = 3$), but we believe that we can straightforwardly extend this structure for $N > 3$. However, end-to-end learning for $N > 3$ is currently not possible, like all the recent deep-learning-based dynamic scene HDR imaging methods that use Kalantari dataset for the training [23], [52], [53]. For the extension, we need to make a well-prepared dataset with more numbers of exposures, which would be much more complicated than the three-exposure case. Hence, very high-cost for the dataset construction is the limitation of our and existing methods.



(a) Input images

(b) Aligned images

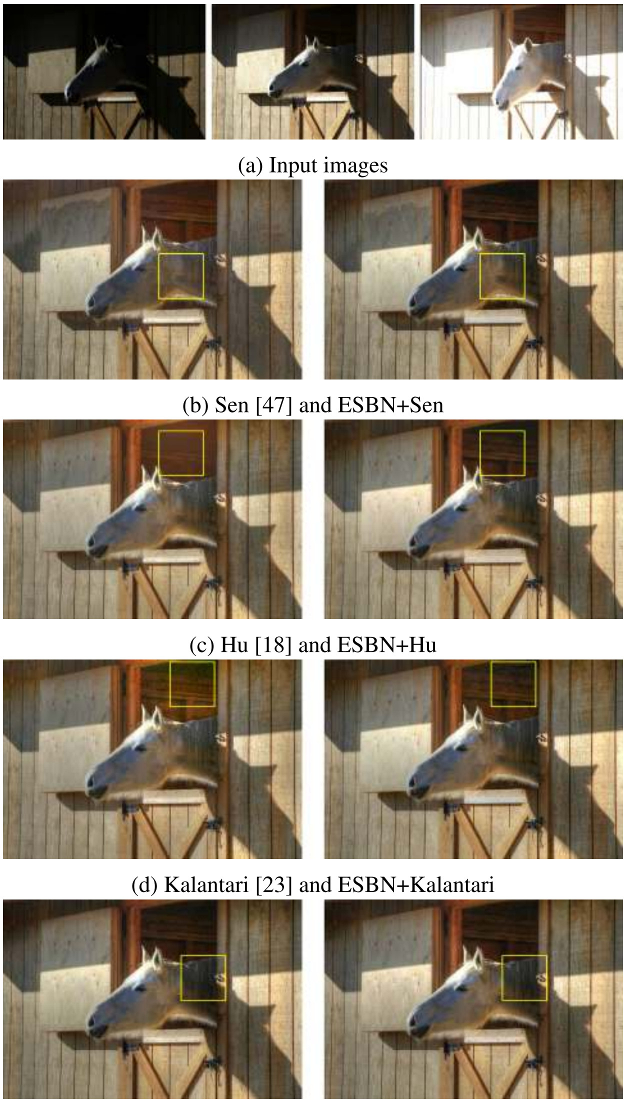
(c) Hu (d) Ours

FIGURE 7. (a) An image set with $N = 5$ which contains images with significant exposure differences in Sen dataset. (b) Alignment results using our ESBNs. The comparison of HDR results by (c) Hu *et al.* [18], where they used the merging method of [39] and (d) our alignment method method, also with merging by [39].

But, we show that our ESBN, which is separately trained, can be applied for the merge of a larger number of exposures. Precisely, our network generates aligned images with significant exposure differences very well, and hence can be used as a pre-processing step for any conventional HDR imaging methods for $N > 3$. For example, we apply our ESBNs to an image set with $N = 5$ in Sen dataset [46], and the alignment results are shown in Fig. 7. It can be seen that our ESBNs generate the aligned images very well, even though the exposure differece is very large. Note that the ESBN is not retrained for the Sen dataset (the ESBN trained with the Kalantari dataset is applied in this case), which shows its generalization performance. In Fig. 7, we also compare the HDR results using the original version of the HDR imaging method in [18] and the modified version where the alignment method is substituted with ours. The latter reconstructs more natural sky, while some artifacts are shown in the former. In summary, we believe that our ESBN is also effective for the case of $N > 3$, and can be used as a preprocessing step for other merging methods as well.

E. PRE-PROCESSING FOR OTHER HDR IMAGING METHODS

To show the effectiveness of our ESBN as a preprocessor for other HDR imaging methods, we compare the HDR results


(a) Input images

(b) Sen [47] and ESBN+Sen

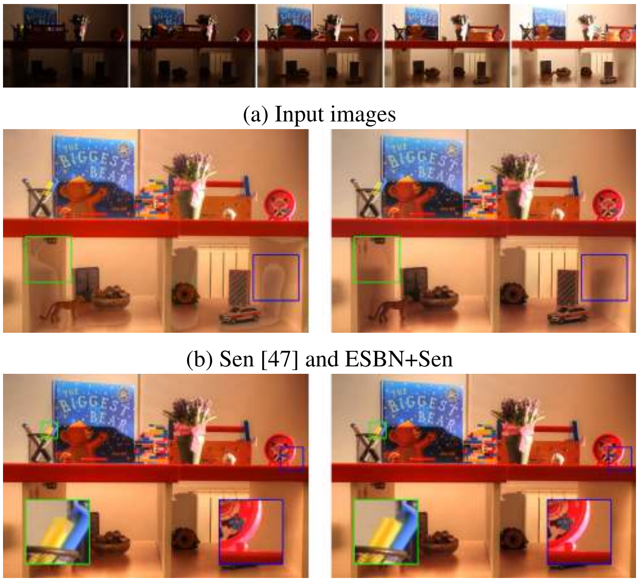
(c) Hu [18] and ESBN+Hu

(d) Kalantari [23] and ESBN+Kalantari

(e) Wu [53] and ESBN+Wu

FIGURE 8. (a) An input image set with $N = 3$ by courtesy of Sing Bing Kang [24]. (b) (c) (d) (e) Comparisons of the HDR results by the original and the modified version of the state-of-the-art methods.

by the original version of the state-of-the-art methods and the modified versions which include our ESBN. Precisely, the aligned multi-exposure images are first generated by our ESBN and then the final HDR image is constrcuted by the original HDR imaging method. In Fig. 8, we compare the HDR results of an image set with $N = 3$. The methods of Sen *et al.* [46] and Kalantari *et al.* [23] produce some artifacts, but those disappear in the modified version of each method. The method of Hu *et al.* [18] fails to reconstruct the dark region, but the modified method succeeds. The modified method of Wu *et al.* [52] reconstructs the mane of the horse with more details. In addition, we compare the HDR results of an image set with $N = 5$ in Fig. 9. The modified method of Sen *et al.* [46] reconstructs the yellow cap of the pen and the highlighted side of the clock more naturally than before, and the modified method of Hu *et al.* [18] causes less artifact.


(a) Input images

(b) Sen [47] and ESBN+Sen

(c) Hu [18] and ESBN+Hu

FIGURE 9. (a) An input image set with $N = 5$ by courtesy of Karaduzovic-Hadziabdic [25]. (b) (c) Comparisons of the HDR results by the original and the modified version of the state-of-the-art methods.

TABLE 7. Comparisons of the original version of the state-of-the-art HDR imaging methods and the modified version where our ESBN is used as a preprocessing. The better results are shown in boldface.

	PSNR-T	SSIM-T	PSNR-L	SSIM-L
Sen [46]	41.11	0.9815	38.82	0.9749
ESBN+Sen [46]	**41.35**	**0.9863**	**39.70**	**0.9800**
Hu [18]	34.87	0.9699	31.72	0.9520
ESBN+Hu [18]	**38.99**	**0.9870**	**36.31**	**0.9808**
Kalantari [23]	42.70	0.9877	**41.21**	0.9845
ESBN+Kalantari [23]	**43.03**	**0.9908**	41.11	**0.9860**
Wu [52]	**41.65**	0.9860	40.88	**0.9858**
ESBN+Wu [52]	41.60	**0.9890**	**41.44**	0.9856

We also compute PSNR and SSIM of the HDR results of test images in Kalantari dataset [23] for the modified methods. Table. 7 shows that the state-of-the-art HDR imaging methods are quantitatively improved, and especially the method of Hu *et al.* [18] is largely improved by using the ESBN as a preprocessing step.

VI. CONCLUSIONS

We have proposed an end-to-end network for the HDR imaging from a set of multi-exposure LDR images with camera and object motions. The conventional deep-learning approaches for the HDR imaging adopted optical flow, homography transform, or attention mechanism to alleviate the problems caused by motions. Unlike the previous methods, we use a deep network to generate aligned images and then merge the aligned LDR images by another network. The proposed alignment network attempts to synthesize an image that has the exposure of under- or over-exposed image and the structure of standard exposure image, and thus we can have a set of well-aligned LDR images as a result. Then a simple merging network can synthesize a quality HDR image without ghosting or blurring artifacts. The overall network, *i.e.,* the cascade of the alignment and merging networks, is trained end-to-end so that joint learning of alignment-merging is attempted. Extensive experiments show that the proposed network shows better quantitative and qualitative results than state-of-the-art methods. In addition, we showed that the alignment network effectively works for generating the aligned images with significant illumination differences, and it can be used as a pre-processing step for other HDR imaging methods. Our code and dataset will be released at https://github.com/tkd1088/ESBN.

REFERENCES

[1] J. An, S. J. Ha, and N. I. Cho, "Reduction of ghost effect in exposure fusion by detecting the ghost pixels in saturated and non-saturated regions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1101–1104.

[2] J. An, S. J. Ha, and N. I. Cho, "Probabilistic motion pixel detection for the reduction of ghost artifacts in high dynamic range images from multiple exposures," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 42, Dec. 2014.

[3] J. An, S. H. Lee, J. G. Kuk, and N. I. Cho, "A multi-exposure image fusion algorithm without ghost effect," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1565–1568.

[4] F. Banterle, K. Debattista, A. Artusi, S. Pattanaik, K. Myszkowski, P. Ledda, and A. Chalmers, "High dynamic range imaging and low dynamic range expansion for generating HDR content," *Comput. Graph. Forum*, vol. 28, no. 8, pp. 2343–2367, 2009.

[5] L. Bogoni, "Extending dynamic range of monochrome and color images through fusion," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Sep. 2000, pp. 7–12.

[6] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.

[7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annu. Conf. Comput. Graph. Interact. Techn.* New York, NY, USA: ACM/Addison-Wesley, 1997, pp. 369–378.

[8] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002.

[9] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, Nov. 2017.

[10] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–10, Nov. 2017.

[11] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[12] B. Funt and L. Shi, "The rehabilitation of MaxRGB," in *Proc. Color Imag. Conf.* Springfield, VA, USA: Society Imaging Science Technology, 2010, pp. 256–259.

[13] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 215–222.

[14] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[16] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha, "Ghost-free high dynamic range imaging," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 486–500.

[17] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 499–512.

[18] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1163–1170.

[19] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *Vis. Comput.*, vol. 30, no. 5, pp. 507–517, May 2014.

[20] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Comput. Graph. Appl.*, vol. 28, no. 2, pp. 84–93, Mar. 2008.

[21] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[22] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 694–711.

[23] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–144, 2017.

[24] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.

[25] K. Karaduzovic-Hadziabdic, T. J. Hasic, and R. Mantiuk. (2017). *Multi-Exposure Image Stacks for Testing HDR Deghosting Methods, Dataset*. [Online]. Available: https://www.repository.cam.ac.uk/handle/1810/261766, doi: 10.17863/CAM.6881.

[26] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, 2003.

[27] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *Proc. 27th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2014, pp. 49–56.

[28] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[29] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1045–1049, Sep. 2014.

[30] S. Lee, G. H. An, and S.-J. Kang, "Deep recursive HDRI: Inverse tone mapping using generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 596–611.

[31] S.-H. Lee, J. S. Park, and N. I. Cho, "A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1737–1741.

[32] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.

[33] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell. (IJCAI)*, Vancouver, BC, Canada, Aug. 1981, pp. 674–679. [Online]. Available: https://ri.cmu.edu/pub_files/pub3/lucas_bruce_d_1981_2/lucas_bruce_d_1981_2.pdf

[34] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1717–1721.

[35] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, p. 68, Aug. 2008.

[36] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, p. 40, 2011.

[37] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–8, Dec. 2009.

[38] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 631–648, Jan. 2017.

[39] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, Mar. 2009.

[40] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[41] J. S. Park, J. W. Soh, and N. I. Cho, "High dynamic range and super-resolution imaging from a single image," *IEEE Access*, vol. 6, pp. 10966–10978, 2018.

[42] J. S. Park, J. W. Soh, and N. I. Cho, "Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 20263–20283, Jul. 2019.

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

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[45] G. J. Rosa, "The elements of statistical learning: Data mining, inference, and prediction by Hastie, T., Tibshirani, R., and Friedman, J.," *Biometrics*, vol. 66, no. 4, p. 1315, 2010.

[46] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–203, 2012.

[47] J. W. Soh, J. S. Park, and N. I. Cho, "Joint high dynamic range imaging and super-resolution from a single image," *IEEE Access*, vol. 7, pp. 177427–177437, 2019.

[48] J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Process. Lett.*, vol. 10, no. 10, pp. 289–292, Oct. 2003.

[49] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proc. WSCG, 15th Int. Conf. Central Eur. Comput. Graph., Vis. Comput. Vis. Co-Oper. EUROGRAPHICS*. Pilsen, Czech Republic: Univ. West Bohemia Pilsen, Jan./Feb. 2007, pp. 49–56. [Online]. Available: https://dspace5.zcu.cz/handle/11025/11002?locale=en

[50] T.-H. Wang, C.-W. Chiu, W.-C. Wu, J.-W. Wang, C.-Y. Lin, C.-T. Chiu, and J.-J. Liou, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 470–484, Apr. 2015.

[51] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *J. Graph. Tools*, vol. 8, no. 2, pp. 17–30, Jan. 2003.

[52] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 117–132.

[53] Q. Yan, D. Gong, Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1751–1760.

[54] Q. Yan, J. Sun, H. Li, Y. Zhu, and Y. Zhang, "High dynamic range imaging by sparse representation," *Neurocomputing*, vol. 269, pp. 160–169, Dec. 2017.

[55] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.

[56] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[57] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 405–414, Apr. 2011.