



Method for reconstructing a high dynamic range image based on a single-shot filtered low dynamic range image

BIN LIANG,^{1,2} DONGDONG WENG,^{1,2,*} YIHUA BAO,² ZIQI TU,¹ AND LE LUO¹

¹Beijing Engineering Research Center of Mixed Reality and Advanced Display, Beijing Institute of Technology, Beijing 100081, China

²AICFVE of Beijing Film Academy, Beijing 100081, China

*crgj@bit.edu.cn

Abstract: Traditional cameras are limited by sensors and cannot directly capture single-shot high dynamic range (HDR) images. We propose an improved single-shot HDR image reconstruction method that uses a single-exposure filtered low dynamic range (FLDR) image. First, by adding an optical filter in front of the camera lens, a FLDR image with different RGB channel exposure states and luminance ranges can be captured in a single-shot, unlike the traditional LDR image. Second, a deep inverse tone mapping network (DITMnet) with multibranch features extraction and multioutput images synthesis is designed to reconstruct an HDR image from a single FLDR image. Experimentally, under different exposure states and color spaces, our method outperforms similar algorithms.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

High dynamic range (HDR) images have greater dynamic range and irradiance contrast and contain richer real-world lighting information than low dynamic range (LDR) images, so they have been applied to many fields [1–11], such as photorealistic editing, physics-based rendering, games, movies, and medicine. Common imaging sensors suffer from limitations in the process of capturing light and cannot directly acquire HDR images. Therefore, special hardware devices [5–8] or fusion algorithms [9–11] have also been generated.

Traditional cameras obtain RGB color images with a single and fixed exposure state through Bayer color filter arrays [12]. Optical filters can selectively change the exposure time and imaging effect of images [13–15]. Loading an optical filter in front of the camera lens can change the light transmittance of each band to form a simple and efficient imaging device and is equivalent to changing the spectral response rate of the camera CCD so that the exposure state and luminance range of each channel in the filtered image are different. Although this filtered image acquisition method reduces the luminance of each channel, it increases the luminance range. Especially in areas such as virtual rendering, the distribution of the luminance information is relatively important.

The most classic methods [9–11] merge an HDR image from multiexposure LDR images of a static scene. These LDR images must undergo strict pixel alignment operations; otherwise, ghost artifacts or tear marks may easily occur. In practice, it is difficult for us to obtain suitable multiexposure images due to object movement or operation errors. Therefore, inverse tone mapping (IMT, or reverse tone mapping, RTM)-based methods [16–20] for reconstructing an HDR image from a single LDR image are currently a popular research topic. Most of the traditional ITM methods [16–20] are based on the idea of a heuristic model, extending the pixel value of the LDR image to a higher luminance range but sometimes destroying the structural features of the original image (see the feature similarity (FSIM) evaluation index scores in Table 1).

In recent years, the application of deep convolutional neural networks (CNNs) to ill-posed or inverse problem image processing has been very suitable [21–28] due to the powerful learning ability and good structural expansion characteristics of these networks.

Table 1. Average score of all methods on the different exposure states and HDR datasets based on LDR data. Bold values indicate the best values. 0.025in

Dataset	Fairchild HDR												Funt HDR											
	Color				HSL				RGB				HSL											
Metrics	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM
State																								
LEO	65.25	2107.4	20.39	0.834	65.56	1695.8	17.85	0.838	65.48	1887.4	22.94	0.868	66.12	1479.5	18.71	0.882								
AEO	68.44	1184.4	21.72	0.902	68.76	733.3	22.35	0.904	70.05	815.6	24.32	0.947	70.72	456.6	24.38	0.954								
KOEO	68.36	1531.4	21.06	0.874	68.75	1060.9	19.21	0.876	70.46	886.8	24.01	0.944	70.95	554.7	22.26	0.951								
HPEO	66.39	1702.5	20.28	0.843	66.83	1323.1	17.95	0.848	68.48	1411.1	22.19	0.905	69.02	1079.8	18.75	0.912								
HDRCNN	70.19	1141.7	21.18	0.893	70.40	804.7	20.26	0.894	72.34	712.9	24.15	0.951	72.99	448.1	23.37	0.958								
DRTMO	69.60	1304.2	21.76	0.869	69.72	1061.8	19.87	0.869	72.07	599.3	23.57	0.930	72.66	345.7	24.28	0.942								
Ours	71.35	593.7	22.08	0.916	71.52	457.1	24.17	0.915	73.43	211.1	23.42	0.962	73.99	103.6	29.72	0.967								
State																								
Under-exposure																								
LEO	61.64	3972.3	20.06	0.723	61.97	3592.9	14.49	0.734	62.01	4300.3	21.69	0.685	62.42	3555.8	13.83	0.718								
AEO	65.49	1593.2	23.48	0.900	65.88	1157.3	20.47	0.910	66.04	1761.4	26.08	0.919	66.70	1141.9	20.84	0.938								
KOEO	65.08	1213.5	22.62	0.902	65.44	886.7	20.64	0.903	65.41	841.8	25.68	0.935	65.75	553.6	22.79	0.946								
HPEO	64.86	1302.1	21.66	0.888	65.31	1021.6	19.35	0.897	66.35	1195.5	23.99	0.930	66.86	859.4	20.40	0.942								
HDRCNN	68.65	945.9	23.27	0.923	68.86	679.6	22.74	0.929	70.23	847.9	26.18	0.950	70.94	535.8	24.00	0.962								
DRTMO	68.35	665.7	22.75	0.922	68.39	475.0	23.10	0.922	69.78	457.7	24.19	0.957	70.30	285.4	25.80	0.963								
Ours	70.21	401.9	22.80	0.945	70.30	293.4	25.90	0.944	71.55	197.0	24.28	0.972	72.13	112.6	29.11	0.975								
State																								
Normal-exposure																								
LEO	60.10	3234.2	20.39	0.72	60.40	2740.6	14.40	0.741	60.00	4901.6	21.41	0.698	60.32	3965.0	12.79	0.737								
AEO	62.24	4992.3	25.04	0.800	62.77	4090.4	13.32	0.832	62.75	8328.1	26.78	0.783	63.53	6427.4	11.22	0.840								
KOEO	61.62	1895.1	24.65	0.862	61.99	1454.7	17.67	0.875	59.87	3447.9	27.02	0.804	60.17	2697.5	14.71	0.833								
HPEO	62.39	3464.2	23.67	0.845	62.91	2825.5	15.11	0.875	62.96	6881.0	25.83	0.814	63.68	5375.2	12.13	0.867								
HDRCNN	66.22	1941.2	25.06	0.876	66.41	1577.8	18.69	0.890	67.45	1654.5	27.13	0.906	68.08	1350.2	20.26	0.922								
DRTMO	65.42	1070.5	23.63	0.891	65.54	782.0	20.70	0.897	66.12	1233.8	24.66	0.919	66.62	976.7	20.94	0.927								
Ours	67.25	664.9	23.71	0.929	67.29	508.1	24.55	0.932	67.72	748.5	25.03	0.942	68.27	646.1	24.12	0.947								
State																								
Total results																								
LEO	62.33	3104.7	20.28	0.758	62.64	2676.4	15.58	0.771	62.50	3696.4	22.02	0.751	62.95	3000.1	15.11	0.779								
AEO	65.39	2590.0	23.41	0.87	65.80	1993.7	18.71	0.882	66.28	3635.0	25.73	0.883	66.98	2675.3	18.81	0.911								
KOEO	65.02	1546.7	22.77	0.879	65.39	1134.1	19.17	0.885	65.25	1725.5	25.57	0.894	65.62	1268.6	19.92	0.910								
HPEO	64.54	2156.3	21.87	0.859	65.02	1723.4	17.47	0.873	65.93	3162.6	24.00	0.883	66.52	2438.1	17.09	0.907								
HDRCNN	68.35	1342.9	23.17	0.898	68.56	1020.7	20.56	0.904	70.01	1071.8	25.82	0.935	70.67	778.0	22.54	0.947								
DRTMO	67.79	1013.5	22.71	0.894	67.88	772.9	21.23	0.896	69.32	763.6	24.14	0.936	69.86	536.0	23.67	0.944								
Ours	69.60	553.5	22.86	0.930	69.70	419.5	24.88	0.930	70.90	385.5	24.24	0.959	71.46	287.4	27.65	0.963								

This paper is guided and designed based on Computational Photography [29]: First, a combined optics with an optical filter and a traditional camera is used to focus light on an image plane; Then a digital sensor captures the focus light through the electrical process; Finally, a powerful

deep learning algorithm is used to compute and reconstruct HDR images from the information obtained by the sensor.

In summary, our contributions are as follows:

(1) An improved deep ITM network (DITMnet) [30] with multibranch features extraction and the multioutput image synthesis is proposed. The experimental results show that the DITMnet method achieves better reconstruction results than other similar algorithms not only on conventional data sets but also on the special data sets we constructed.

(2) A single-shot filtered LDR (FLDR) image acquisition method [30] is proposed. Compared to the use of a traditional LDR image, the use of an FLDR image obtained with different exposure states and luminance ranges for each channel in a single shot facilitates the reconstruction of a better HDR image.

(3) A more comprehensive, large-scale FLDR-LDR-HDR database covering indoor, outdoor, natural scenery is built. Each scenario includes 2 HDR images, 9 multiexposure LDR images, and corresponding FLDR images. The code is published on GitHub.

Figure 1 shows and compares the reconstruction of all HDR reconstruction methods based on a single LDR image. All the LDR images are generated from HDR predicted images by the tone mapping (TM) of Kim [31] without adjustment of the parameters. We control the variables as much as possible to more fairly compare the performance of all the algorithms.

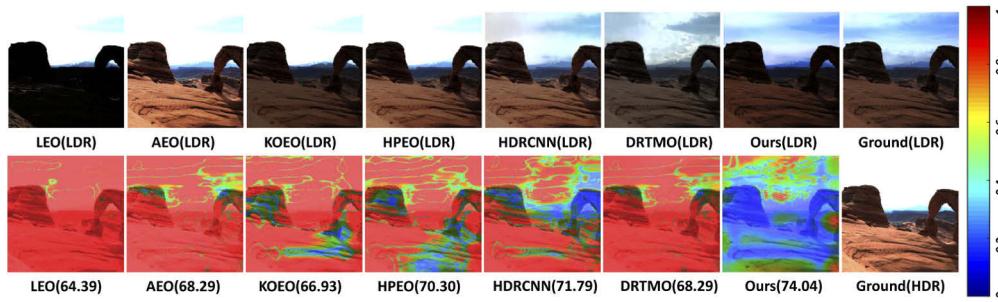


Fig. 1. Performance comparison of different HDR reconstruction methods. **Top:** LDR images are generated from predicted HDR images by the tone mapping of Kim without adjustment of the parameters. **Bottom:** Probability map and quality scores obtained from the HDR-VDP-2 metric for comparison of HDR reconstruction performance. If the error is small, it is shown in blue; otherwise, it is shown in red. Higher scores indicate better performance.

The remainder of this paper is organized as follows. First, the related work is reviewed in Section 2, and then the algorithm flow of our single-shot HDR reconstruction method is described in Section 3. Next, the effectiveness of our DITMnet network and single-shot FLDR acquisition method is gradually verified in various experiments in Section 4. Section 5 summarizes our current work, limitations, and future improvements.

2. Related work

Optical filter: If the irradiance contrast in the scene is relatively large, it is likely that complete and effective scene information cannot be obtained with a single shot. The most classic method merges the multiexposure images to preserve the effective information in the scene [9–11]. Multiple-exposure images can be obtained by changing the camera aperture size or integration time. However, for a single shot, we want to keep the aperture constant to prevent artifacts, and it is not possible to change the integration time of the camera. Therefore, obtaining images in different exposure states through optical filters is an excellent option.

Optical filters (shown in Fig. 2) are used to strobe the desired band and reflect (or absorb) other bands. Nezam et al. [13] proposed a degree-of-polarization (DOP)-based differential-group-delay (DGD) monitor using an optical filter such that the DGD monitoring range and DOP dynamic range were dramatically increased. Johannes et al. [14] used optical filters and filter wheels to obtain high fidelity color images. Shrestha et al. [15] used a pair of filters and a binocular stereo camera to construct a six-channel color image acquisition method for estimating the spectral reflectance and value of each pixel. Although these methods are not single-shot acquisition systems, they still give us clear inspiration: with the help of a suitable optical filter, an image with different exposures can be obtained in a single shot.

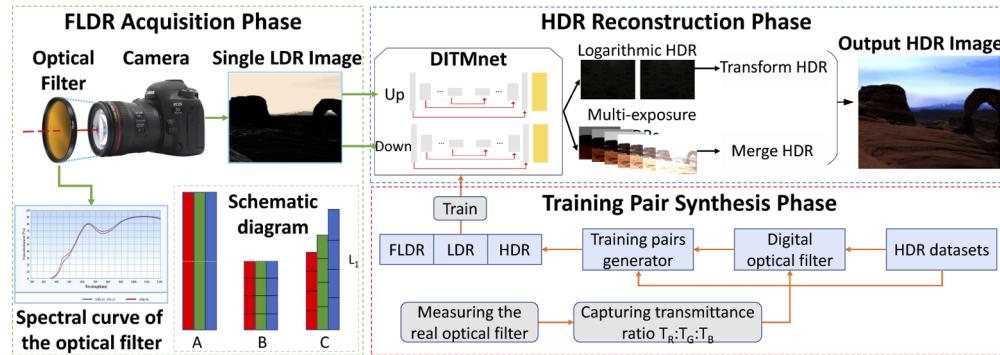


Fig. 2. Overview of the proposed method, consisting of the **training pair synthesis phase** (red virtual box), single-shot **FLDR acquisition phase** (green virtual box), and **HDR reconstruction phase** (blue virtual box). In the synthesis phase, the ternary pairs are generated by virtual cameras and digital optical filters from HDR datasets. In the acquisition phase, FLDR images are obtained in a single shot using physical optical filters and real Canon cameras. In the reconstruction phase, two different types of initial HDR images are generated from a single-exposure FLDR image using the DITMnet network and are merged into a final HDR image. In the **schematic diagram**, **A** represents the original luminance; **B** represents an ordinary camera, which can acquire an LDR image with the same exposure and luminance range in one shot; **C** represents a filtered camera, which can acquire an FLDR image with different exposures and luminance range in one shot.

HDR reconstruction: According to the number of input images, there are two main types of methods: one method merges a single HDR image from multiexposure LDR images [9, 10], and the other method generates an HDR image from a single-exposure LDR image based on ITM [16–20].

Landis [16] proposed a power function model to extend the luminance of LDR images because the power function has functional range expansion capabilities. Akyuz [17] proposed a linear transformation method combined with gamma calibration. Banterle [18] proposed a smooth filtering algorithm to reconstruct HDR content. Kovaleski and Oliveira [19] proposed an enhancement algorithm based on cross-bilateral filtering. Subsequently, Huo et al. [20] further expanded the work and removed the thresholds used by Kovaleski and Oliveira [19].

In recent years, CNNs have also made great achievements in HDR image reconstruction [21–28]. Zhang and Lalonde [21] proposed an autoencoder network to predict outdoor environmental parameters to reconstruct HDR images, but this parameter is applicable to only an outdoor sunlight environment model. Eilertsen et al. [22] used the U-net architecture to predict log-domain HDR images by estimating the luminance values of the saturated regions, which is equivalent to information completion without effective constraints in a blank area. Endo et al. [23] used an improved U-net architecture to indirectly predict an HDR image by merging multiexposure LDR predicted images [9]. This method requires considerable memory and the training speed is slow.

Marnerides et al. [24] used a multibranch CNN network without the "upsampling" operation [32] to directly reconstruct HDR images, but reasonable formula derivation and explanation of the processing of multiscale image features are lacking. Lee et al. [27] used a generative adversarial network (GAN) with two generators G^{plus} and G^{minus} to generate seven LDR images in different exposure states through three inference phases. Although this method reduces the cost, it does not consider the end-to-end training of the entire model, which easily results in error accumulation. Liu et al. [28] used three specialized CNNs to model and reverse the HDR-to-LDR image formation pipeline to reconstruct an HDR image from a single LDR image. This method fine-tunes the entire model end-to-end to reduce error accumulation, but it requires a considerable amount of memory.

Based on our previous thesis [30], a single-shot FLDR image acquisition method and DITMnet network are constructed, and a single shot can be used to obtain an FLDR image with different exposure states and luminance ranges for three channels. Then, the DITMnet network with multibranch feature extraction and multioutput images synthesis is used to reconstruct an HDR image.

3. Our method

Figure 2 shows an overview of the process of our work. The process includes the training pair synthesis phase, the single-shot FLDR acquisition phase, and the HDR reconstruction phase. In the synthesis phase, a set of ternary training pairs is generated from HDR datasets by virtual cameras and digital optical filters, and this set includes a ground-truth HDR image, multiexposure traditional LDR images and corresponding filtered FLDR images; in the acquisition phase, FLDR images are obtained in a single shot using a physical optical filter and a real Canon camera; in the reconstruction phase, two different types of initial HDR images are generated from a single-exposure FLDR image using the DITMnet network and then merged into a final HDR image.

3.1. Single-shot FLDR image acquisition method

The single-shot FLDR image acquisition method in this article employs a common camera and optical filters. Through a single shot, an FLDR image with different exposure and luminance ranges for each channel can be obtained. The schematic diagram is as follows:

As shown in Fig. 2, regarding the principle of single-shot FLDR image acquisition, assuming that A represents the luminance range of the scene, ordinary cameras are limited by sensors and can measure only a part of the luminance range of the scene (as shown in subfigure B). It is impossible to record any information outside this range (as shown above the dotted line L_1 in the subfigure). Our improved filtered camera can independently change the transmission of each channel and obtain an FLDR image with different exposure states and luminance ranges through a single shot (as shown in subfigure C).

In particular, the FLDR images collected by our method are very different from the traditional LDR images: first, the exposure state and luminance range of each channel are different. In other words, a wider range of scene luminance information can be measured, and the adaptability to scenes with large irradiance contrast is improved; second, the method is equivalent to changing the spectral response curve of the camera for each channel and reducing the luminance resolution of the image, which leads to the phenomenon of color cast.

In this article, we used a real physical SJB130 optical filter, a KONICA MINOLTA cs-200 chromameter and an X-RITE 24 color checker card to build an optical filter transmittance measurement device, which is used to measure the transmittance ratio of the optical filter.

In general, although our device reduces the luminance resolution of the image, it increases the luminance range. The DITMnet network proposed in this paper focuses on restoring the color and adjusting the luminance of the image while retaining a more reliable luminance distribution.

3.2. Network architecture

We assume that there are two mapping relationships between LDR images and HDR images: Corresponding mapping relationship between LDR image and logarithmic domain HDR image (see Appendix A), association mapping between multiple exposure LDR images (see Appendix B). In this paper, our key purpose is simultaneously to obtain a logarithmic domain HDR image and several multiexposure LDR images through a single exposure LDR image. Therefore, this article designs a DITMnet network with multibranch features extraction and multioutput images synthesis, including a 3D U-net branch, local-detail branch, and fusion-output branch, as shown in Fig. 3. Intuitively, the 3D U-net branch can output multiexposure gradient images better than the 2D U-net branch. Although a U-net with skip layers can retain some of the image local features, some artifacts [32] often appear in the generated results because of the upsampling layers. Therefore, the local-detail branch without upsampling can better extract local features of the images and eliminate artifacts. The fusion-output branch is used to fuse the image features of the first two branches and generate two different types of predicted images. Sufficient network depth can invert the camera response curve (CRF) well, and multiexposure output images can be used to study the effect of exposure on imaging, which is also a motivation for the design of the DITMnet structure.

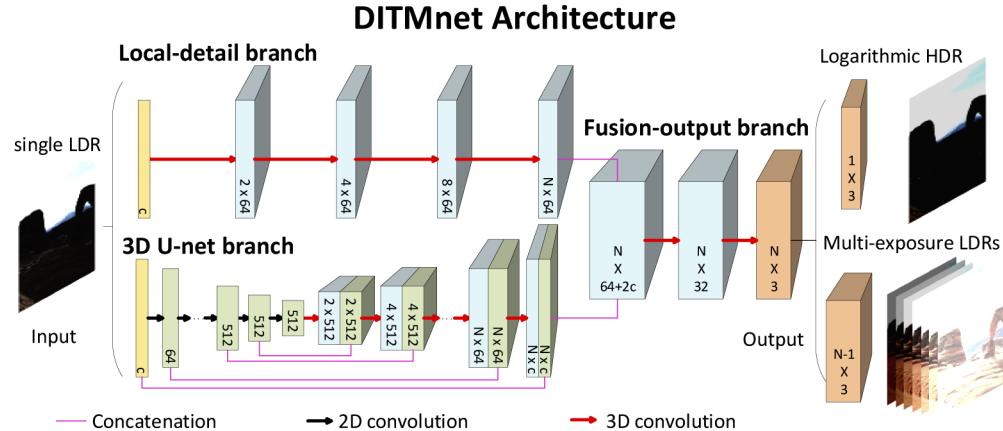


Fig. 3. DITMnet network architecture, including 3D U-net branch, local-detail branch, and fusion-output branch. A single LDR image is input, and DITMnet can output N images under the constraints of two loss functions, L_1 and L_{cos} (see Section 3.3).

Note that the complete model of this article is divided into the upmodel and downmodel and that the structure of the two sub-models is completely consistent (as shown in Fig. 3). When an LDR image is input, the upmodel infers the multiexposure LDR images whose exposure time gradually increases based on the exposure state of the input image. Similarly, the downmodel infers images whose exposure time gradually decreases. In the training phase, each sub-model needs to be separately trained end-to-end. In the testing phase, we gather the output results of the two submodels to merge them into a final HDR image.

3D U-net branch: The 3D U-net branch is based on a convolutional encoder-decoder network and extracts the global abstract features of the input image. The encoder contains eight 2D convolutional layers, and the decoder contains eight 3D deconvolutional layers. We reserve the skip concatenation operation between the encoder and decoder. Note that before the concatenation operation, the tensor of the encoder needs to be expanded to match the tensor size of the decoder. Given a $W \times H \times c$ three-dimensional tensor (W, H, c represent the width, height and channel, respectively), an $N \times W \times H \times 2c$ four-dimensional tensor can be generated. Here, we set $N = 8$.

Local-detail branch: The local-detail branch is based on a 3D deconvolutional network without upsampling and extracts the local-detail features of the input images. Given a $W \times H \times c$ three-dimensional tensor, an $N \times W \times H \times f$ four-dimensional tensor can be generated. f represents the convolution kernel size of the 3D deconvolutional layers, and $f = 64$.

Fusion-output branch: The fusion-output branch is a network of two 3D convolutional layers without downsampling and is used to fuse the image features of the first two branches and generate two different types of predicted images, including a logarithmic domain HDR image and $N - 1$ multiexposure LDR images. Given an $N \times W \times H \times (2c + f)$ four-dimensional tensor, an $N \times W \times H \times c$ four-dimensional tensor can be generated.

Batch normalization and the leaky rectified linear unit (ReLU) function are also applied to all layers, except the output layer, whose activation function is the sigmoid function.

3.3. Loss function

In our DITMnet network, the upexposure model and the downexposure model share the same network structure. When the output multiexposure LDR images are sorted from dark to bright, the parameters of the upexposure model can be obtained; otherwise, the parameters of the down-exposure model are obtained. In this section, the upexposure model is taken as an example, and the meanings of the symbols used are defined as follows:

D_j represents the j .th training pair, which consists of a logarithmic HDR image H_j , $T + 1$ filtered FLDR images $I_{j,i}^F$ and $T + 1$ conventional LDR images $I_{j,i}$. $I_{j,i}^F$ and $I_{j,i}$ represent the i .th FLDR and LDR image on D_j , respectively. T represents the number of camera exposures, N represents the number of output images of a single submodel, and K represents the total number of pixels of a single output image.

Following the paper [24], two loss functions L_1 and L_{cos} are used, and defined in Eq. (1):

$$\begin{aligned}\mathcal{L}_1 &= \frac{1}{NK} \|X_{gt} - X_{ot}\|_1, \\ \mathcal{L}_{cos} &= 1 - \frac{1}{K} \sum_{k=1}^K \frac{X_{gt}^k \cdot X_{ot}^k}{\|X_{gt}^k\|_2 \cdot \|X_{ot}^k\|_2}, \\ \mathcal{L} &= \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{cos}.\end{aligned}\quad (1)$$

where, $X_{gt} = \{H_j, I_{j,i+1 \rightarrow j,i+N}\} \oplus O_i$ and $X_{ot} = M_i \circ G(I_{j,i}^F, \theta)$. H_j represents a $W \times H \times c$ logarithmic HDR image, and $I_{j,i+1 \rightarrow j,i+N}$ represents a $\min(N - 1, T - i) \times W \times H \times c$ tensor from the $(i + 1)$.th to the $\min(N + i, T + 1)$.th LDR image. \oplus and O_i are a concatenation operator and a $\min(i + N - T - 1, 0) \times W \times H \times c$ zero tensor, respectively, and M_i and \circ are an $N \times W \times H \times c$ mask tensor and an elementwise product operator, respectively. If $\min(N, T + 1 - i)$ is less than $(T + 2 - i)$, the value of M_i is 1; otherwise, it is 0. $G(I_{j,i}^F, \theta)$ represents an $N \times W \times H \times c$ output tensor when the i .th image $I_{j,i}^F$ is input, and θ represents the network model parameters. Note that when the input image is $I_{j,i}$, the model is LDR2HDR, and when the input image is $I_{j,i}^F$, the model is FLDR2HDR. Detailed experimental instructions are provided in Section 4. Here, $\lambda_1 = 24$ and $\lambda_2 = 1$.

3.4. Merging an HDR image

When a single-shot FLDR image is input in DITMnet, the upexposure and down-exposure submodel outputs 2 log-domain HDR images and $2(N - 1)$ multiexposure LDR images.

First, the first initial HDR image can be obtained according to Eq. (2):

$$H_j^{log} = 10^{(H_j^{up} + H_j^{down})/2}, \quad (2)$$

where H_j^{up} and H_j^{down} represent logarithmic domain HDR prediction images of the upexposure model and the downexposure model, respectively.

Next, we sort the $2(N - 1)$ multiexposure LDR prediction images from the brightest to the darkest and then average the $(N - 1).th$ and $N.th$ images as the basic image v_i , so we have $(2N - 1)$ images. To avoid the selection of an image that is too bright or too dark, the absolute difference between each image v_j and the base image v_i cannot exceed a certain threshold η , i.e., $|v_j - v_i| < \eta$. Here, $\eta = 24$. Thus far, the second initial HDR image can be obtained using the method ofDebevec et al. [9] according to Eq. (3):

$$H_j^{merge} = Debevec(LDRs, times), \quad (3)$$

In this paper, the lowest exposure is assumed to be 1/128 seconds. Since the precise exposure time cannot be obtained, the result H_j^{merge} is a "pseudo" HDR image. To obtain more accurate results, we use a weighted parameter α to synthesize the two initial images H_j^{log} and H_j^{merge} into a final HDR image according to Eq. (4):

$$H_j^{final} = (1 - \alpha) \frac{H_j^{log}}{\max(H_j^{log})} + \alpha \frac{H_j^{merge}}{\max(H_j^{merge})}. \quad (4)$$

Here, $\alpha = 0.6$.

3.5. Creating training pairs

To train and validate our network, we collected 3 different HDR datasets from Fairchild-HDR, Funt-HDR, and DML-HDR to create training pairs. These datasets cover a variety of different types of scenarios, such as indoor, outdoor, and natural landscapes. To adapt to our DITMnet network, this paper designs virtual cameras with digital optical filters, randomly crops the HDR datasets and generates more than 80K sets of special ternary data pairs. Each set of data contains a logarithmic domain HDR image, 9 multiple-exposure traditional LDR images and corresponding FLDR images. All images are 256×256 . The Fairchild-HDR has approximately 36K data pairs, and 85% of the data is randomly selected as the training set. In particular, the Funt-HDR data, which contain 4K ternary data pairs, are directly used for performance testing without any training to eliminate the dependence of the other HDR reconstruction methods on the datasets. The Funt-HDR dataset is used as the benchmark dataset to test the performance of all HDR reconstruction methods. In addition, the DML-HDR has approximately 43K data pairs, and 85% of the data is randomly selected to retrain our model in an end-to-end manner on the FLDR images. None of the test data are included in the training data.

The process of creating training pairs is as follows:

Step 1: HDR dataset normalization Eq. (5) is used to convert the HDR image to the log-domain and then normalize the image, where μ represents the maximum value in the HDR dataset and H_i^{ori} represents the $i.th$ HDR original image.

$$H_i = \frac{\lg(1 + H_i^{ori})}{\lg(1 + \mu)}, \quad (5)$$

Step 2: Digital optical filter According to the characteristics of physical optical filters, we simulate real filters by adjusting the transmission ratio of the RGB channels. In particular, $T_R : T_G : T_B = 1 : 1 : 1$ is equivalent to a virtual camera without optical filters, and a traditional LDR image can be obtained. We used the optical filter transmittance measurement device designed in Section 3.1 to estimate the transmittance of the real optical filter as $T_R : T_G : T_B = 0.7316 : 0.6839 : 0.4945$. In addition, in the experiments in Section 4.3, we tested various filtering schemes to explore the performance of different optical filters for HDR reconstruction.

Step 3: Virtual camera generator Eq. (6) is used to simulate the cameras, and five representative camera response functions (CRF) were selected from the database of response function (DoRF) [33] by the K-means algorithm:

$$\begin{aligned} L_{ij} &= f(H_i \Delta t_j), \\ \Delta t_j &= \frac{1}{\tau^{T/2}}, \dots, \frac{1}{\tau^2}, \frac{1}{\tau}, 1, \tau, \tau^2, \dots, \tau^{T/2}. \end{aligned} \quad (6)$$

where H_i represents the scene radiance value of the $i.th$ HDR image, Δt_j represents the $j.th$ exposure, f represents the CRF, L_{ij} represents the luminance value of the LDR image, and τ represents exposure interval. Here, $\tau = \sqrt{2}$, $T = 8$.

4. Experiments and analysis

The DITMnet network was implemented with the TensorFlow library on a PC with i7 CPU, 32 GB RAM and NVIDIA GTX 1080 Ti GPU. The network adopted Adam optimizer with a momentum term of 0.5 and a batch size of 1. The initial learning rate was $5e - 3$, then decreased to $1e - 4$; the total training (including all ablation experiments) took approximately 39 days.

In this section, four experiments are designed to gradually explore and verify our contributions. The specific arrangements are as follows. First, the evaluation methods and metrics are introduced in detail in Section 4.1. Then the performance comparison experiments based on LDR2HDR are designed in Section 4.2. These experiments compare the performance of our DITMnet network with that of other similar HDR methods on traditional LDR images and verify the effectiveness of the multibranch features extraction and multioutput images synthesis strategies in the ablation experiments. Next, comparison experiments based on FLDR2HDR are designed in Section 4.3. These experiments verify the effectiveness of our single-shot FLDR image acquisition method and explore the reconstruction performance of different filtering schemes. In Section 4.4 and 4.5, two HDR reconstruction experiments based on retraining FLDR images and real FLDR images are designed to verify the effectiveness of our DITMnet network on synthetic and real FLDR images, respectively.

4.1. Evaluation methods and metrics

In this paper, we compared seven different HDR reconstruction methods, including Landis (LEO) [16], Akyuz et al. (AEO) [17], Kovaleski and Oliveira (KOEO) [19], Huo et al. (HPEO) [20], HDRCNN [22], DRTMO [23] and our DITMnet.

For qualitative and quantitative evaluation, we used four evaluation metrics, HDR-VDP-2 (visual difference predictor, VDP) [34], mean square error (MSE), peak signal noise ratio (PSNR), and FSIM, to compare the prediction performance. Among them, VDP [34] is the most important evaluation indicator. The probability map and quality score between the HDR predicted image and the HDR ground-truth image are calculated to evaluate the performance of HDR reconstruction. The probability map indicates the average difference between two HDR images that would be noticed by an observer. The Q score indicates quality through a mean opinion score. The Q score is 100 for the best quality and vice versa. The last three metrics, MSE, PSNR, and FSIM, were secondary evaluation indicators used to evaluate the performance regarding LDR images that were generated from HDR predicted images by Kim's TM [31] without adjustment of the parameters.

In addition, we compared the performance in two color spaces: RGB space, which was used to compare the color information distribution of the images, and HSL space, which used to compare the luminance information distribution.

4.2. Experiment based on LDR2HDR

To verify the validity of our DITMnet network, we compare the performance of our DITMnet network with that of similar HDR methods based on traditional LDR images in this section. The preset models and parameters published by the original authors are used for all the models. The input images are the same traditional LDR images, and the generated HDR prediction images are compared with the ground-truth HDR images. Figure 4 (left) shows a visual qualitative comparison of these methods on the LDR test datasets, and Table 1 shows a statistical quantitative comparison.

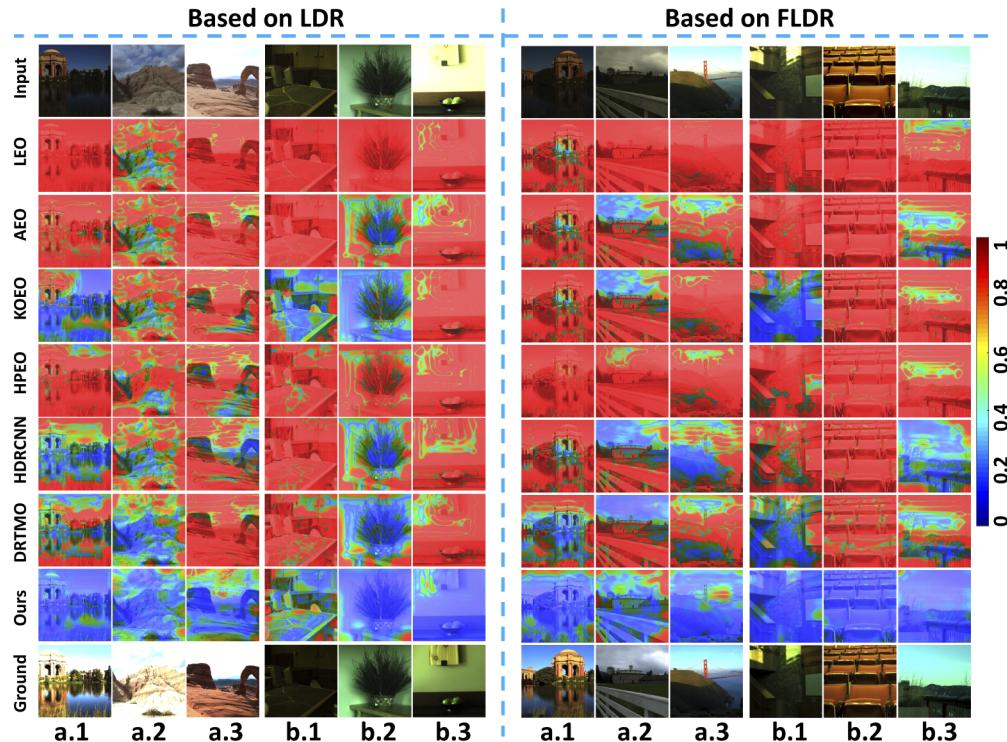


Fig. 4. Visualization probability maps of all HDR reconstruction methods based on traditional LDR (left) and our FLDR (right). **a** and **b** represent the Fairchild-HDR and Funt-HDR test datasets, respectively. 1, 2, and 3 indicate under-/ normal-/ overexposed states, respectively. Blue indicates low errors and red indicates high errors.

As seen in Fig. 4 (left), the first four traditional methods (LEO [16], AEO [17], KOEO [19], and HPEO [20]) effectively preserve local edges and the general structure of the images. However, these methods cannot recover the details and features of the images well (see the MSE and FSIM metric scores in Table 1). Because these methods use fixed models and parameters to reconstruct HDR images, they cannot adapt to changing scenes and exposure states. As seen in Table 1, in the RGB color space, the PSNR metric score of AEO [17] is very high, but the VDP metric score is very poor. This shows that AEO [17] performs well on LDR images after the tone mapping of the HDR predicted images but performs poorly on HDR images. The other three methods perform worse than AEO [17]. Compared with that of traditional methods, the performance of the last three deep learning methods (HDRCNN [22], DRTMO [23], and our DITMnet network) is greatly improved. Table 1 shows that in the HSL color space, the performance scores of our DITMnet network almost exceed those of other methods; in the RGB color space, except for the

PSNR metric, our DITMnet network's evaluation scores are far above those of the other models, especially for the most important VDP metric.

In other words, our DITMnet network reconstructs an HDR image from a single traditional LDR image better than the other models, especially in HSL color space.

In addition, we designed two ablation experiments to verify the effectiveness of the multibranch features extraction and multioutput images synthesis strategies in the DITMnet network.

Ablation experiment 1: Cut one of the branches, and separately train the network parameters of the other branch. After the loss function stabilizes, output the prediction results from all the test images. Compare the average quality scores of each individual branch. The results are shown in Fig. 5 and Table 2. (EX 1). As seen in Fig. 5, the image produced by the 3D U-net branch is somewhat sharp and has some artifacts [32]; the image generated by the local-detail branch lacks some low-frequency information, the details are not clear enough, and the image is slightly fuzzy. The images produced by both branches are very close to the ground-truth, which not only eliminates artifacts but also enhances the low-frequency information, and the details are clearer than those of the other two branches.

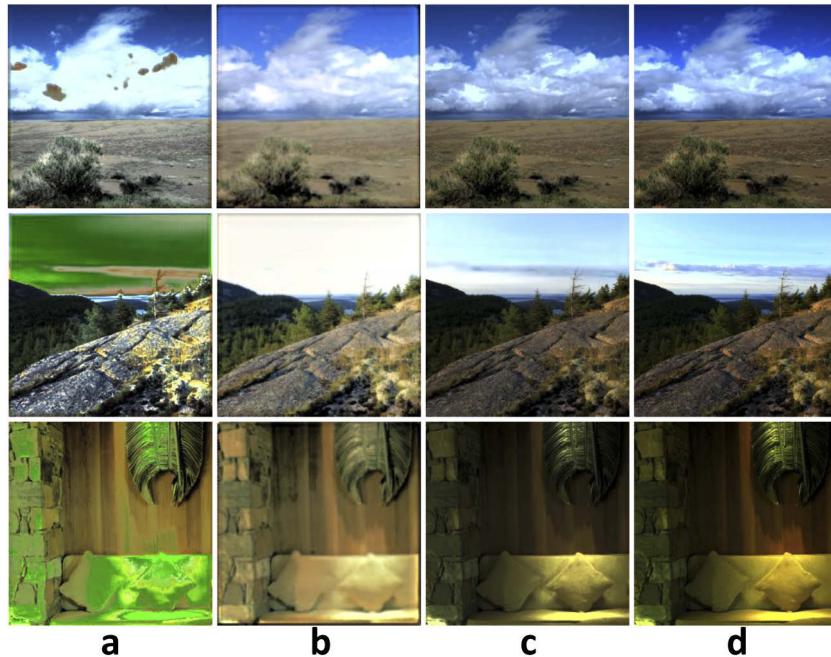


Fig. 5. Comparison of each branch. (a) shows the predicted results of only the 3D U-net branch; (b) shows the predicted results of only the Local-detail branch; (c) shows the predicted result of both branches cooperating; (d) shows the corresponding ground-truth.

Ablation experiment 2: Similar to the operation of ablation experiment 1, cut one of the output modes, and separately train the network parameters of the other output mode. Compare the average quality scores of the three output methods. LCM represents the log-domain conversion method, which is the first initial HDR prediction image in this paper; MEM represents the multiexposure merging method [9], which is the second one; our DITMnet method (Ours) combines the two initial HDR images into a final HDR image by weighting the value α . The results are shown in Table 2 (EX 2). When no special strategy [22] is used for processing, the HDR reconstruction performance score of the LCM method is slightly worse than that of the MEM method; our DITMnet network uses a synthesis strategy, and its performance score is better than those of the other two methods.

Table 2. Average quality score of the two ablation experiments on the two HDR datasets. Bold values indicate the best values.

Metrics	VDP											
	Dataset		Fairchild-HDR		Funt-HDR		Dataset		Fairchild-HDR		Funt-HDR	
Color	RGB	HSL	RGB	HSL		Color	RGB	HSL	RGB	HSL		
EX 1	Comparisons to Different Branches				EX 2	Comparisons of Different Outputs						
3D U-net	62.25	62.82	65.52	66.22		LCM	68.28	68.47	67.93	68.54		
Local-detail	63.15	63.26	63.64	64.32		MEM	68.79	68.83	70.65	71.18		
Ours	69.60	69.70	70.85	71.39		Ours	69.60	69.70	70.85	71.39		

The experimental results show that our DITMnet network has an outstanding ability to extract details, effectively reduces artifacts and further improves the HDR reconstruction performance.

4.3. Experiment based on FLDR2HDR

To verify the effectiveness of our single-shot FLDR image acquisition method, we compared the performance of all the methods on the FLDR images. The models and parameters of all methods are the same as those in Section 4.2, but the input is FLDR images instead of LDR images. Figure 4 (right) shows a visual qualitative comparison of the different HDR reconstruction methods, and Table 3 shows a statistical quantitative comparison.

First, we performed a comparative analysis of the group in Table 3. A comparison of the first four traditional methods shows that their VDP metric scores are low, indicating that these methods exhibit poor performance on HDR reconstruction, while AEO [17] has high PSNR metric scores, indicating that this method performs well on tone-mapped images. In general, the traditional methods perform poorly. The last three deep learning methods exhibit much higher scores for the various indicators. In summary, in both the HSL color space and RGB color space, the performance scores of our method typically surpass those of other methods, especially on the most important VDP metric.

Second, we performed a comparative analysis between Tables 1 and 3, as shown in Fig. 6. The models and parameters for all methods remain the same, but the input image is an FLDR image instead of an LDR image. All the existing methods not only can successfully predict a reasonable HDR image but also can further improve the performance of HDR reconstruction, verifying the effectiveness of our single-shot FLDR image acquisition method. Compared with that of other similar HDR reconstruction methods, our method still maintains the best performance.

Finally, we arranged and combined the estimated optical filter transmittances $T_1 = 0.7316$, $T_2 = 0.6839$, and $T_3 = 0.4945$ to generate six different optical filters. In addition, we generated a virtual neutral density optical filter with a transmittance of 0.6839 to explore the effects of different filtering schemes on HDR image reconstruction. As seen in Fig. 7, the neutral density optical filtering scheme equally attenuate the light in all bands, and the effect is equivalent to reducing the exposure time of the camera. This filtering scheme cannot specifically attenuate certain channels that are too bright or too dark, which is not conducive to the HDR reconstruction images. Our acquisition method can select suitable filters for different scenes and is more adaptable to scenes with large contrast between light and dark. The performance comparison results between the second line and the third line show that our acquisition method is more conducive to the reconstruction of HDR images, but the results between the first line and the fourth line show that our HDR reconstruction method exhibits slight chromatic aberration in some scenes.

In summary, these experiments show that our single-shot FLDR image acquisition method is more conducive to HDR image reconstruction, especially in terms of luminance information distribution, than traditional LDR image or neutral density filtered image acquisition methods.

Table 3. Average score of all methods on the different exposure states and HDR datasets based on FLDR data. Bold values indicate the best values. 0.025in

Dataset	Fairchild HDR										Funt HDR							
Color	RGB					HSL					RGB				HSL			
Metrics	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE	PSNR	FSIM	VDP	MSE
State	Under-exposure																	
LEO	65.99	2996.6	20.23	0.817	66.23	2319.6	15.68	0.815	67.41	1776.3	22.85	0.896	67.94	1161.3	19.03	0.904		
AEO	69.64	1958.1	21.34	0.874	69.64	974.3	20.56	0.874	71.86	1291.3	23.63	0.933	71.98	593.7	23.06	0.936		
KOEO	68.88	2666.9	20.75	0.835	69.19	1739.9	16.56	0.833	71.46	1531.9	23.37	0.923	71.77	855.3	20.15	0.927		
HPEO	66.66	2804.5	19.94	0.807	67.08	2072.6	16.18	0.806	68.66	2008.0	21.75	0.883	69.26	1363.0	17.72	0.886		
HDRCNN	70.84	2368.7	20.86	0.850	70.87	1700.9	16.94	0.848	73.17	1332.9	23.42	0.929	73.42	768.0	20.56	0.934		
DRTMO	69.86	2286.9	21.55	0.827	69.83	1841.6	16.79	0.830	71.85	963.6	23.32	0.906	72.43	576.1	22.14	0.920		
Ours	71.02	1094.8	21.85	0.882	71.19	846.5	20.47	0.882	73.13	340.9	23.17	0.943	73.74	170.4	27.86	0.951		
State	Normal-exposure																	
LEO	62.53	3890.8	20.26	0.765	62.85	3328.2	14.98	0.771	62.98	3897.5	22.31	0.735	63.45	3214.7	14.88	0.759		
AEO	67.36	1564.0	22.93	0.893	67.38	729.3	21.50	0.909	68.29	1288.2	25.46	0.936	68.47	591.5	22.86	0.948		
KOEO	66.93	1738.1	22.08	0.883	67.07	1065.8	19.19	0.885	68.02	942.9	25.10	0.947	68.14	460.8	23.42	0.951		
HPEO	65.65	1804.0	21.07	0.860	66.09	1241.0	18.44	0.865	67.81	1234.6	23.27	0.927	68.16	806.4	20.52	0.931		
HDRCNN	70.25	1223.4	22.61	0.907	70.04	616.7	21.96	0.915	71.82	840.8	25.48	0.955	71.92	355.3	25.13	0.963		
DRTMO	69.57	1082.3	22.42	0.896	69.45	759.4	20.93	0.899	71.38	396.7	24.02	0.951	71.77	193.3	26.75	0.960		
Ours	70.89	639.3	22.51	0.929	70.97	435.9	23.56	0.927	72.80	251.7	23.95	0.967	73.21	106.3	29.02	0.972		
State	Over-exposure																	
LEO	61.20	2968.2	20.21	0.721	61.50	2430.1	15.16	0.745	61.02	3293.9	21.18	0.694	61.31	2478.2	14.79	0.742		
AEO	64.03	2986.8	24.42	0.840	64.40	1760.1	18.05	0.889	64.27	4774.1	26.71	0.838	64.75	2980.9	15.84	0.893		
KOEO	62.12	1671.9	23.75	0.875	62.49	1182.5	18.92	0.886	61.04	2008.0	26.55	0.852	61.33	1476.5	17.65	0.875		
HPEO	63.92	2018.3	22.87	0.871	64.38	1261.3	18.80	0.904	64.63	3462.4	25.35	0.872	65.10	2187.5	17.19	0.918		
HDRCNN	68.24	1509.5	24.35	0.889	68.21	747.0	21.75	0.919	69.18	1277.9	26.95	0.922	69.60	741.7	22.91	0.944		
DRTMO	67.25	901.4	23.30	0.907	67.13	561.7	22.26	0.917	67.89	850.1	24.53	0.934	68.25	584.3	23.55	0.947		
Ours	68.88	537.3	23.37	0.937	68.76	367.9	24.74	0.938	70.00	292.1	24.81	0.959	70.39	163.4	28.07	0.966		
State	Total results																	
LEO	63.24	3285.2	20.23	0.768	63.53	2692.6	15.28	0.777	63.80	2989.2	22.12	0.775	64.23	2284.8	16.24	0.802		
AEO	67.01	2169.6	22.90	0.869	67.14	1154.6	20.04	0.891	68.14	2451.2	25.27	0.902	68.40	1388.7	20.59	0.925		
KOEO	65.98	2025.6	22.19	0.864	66.25	1329.4	18.22	0.868	66.84	1494.3	25.00	0.907	67.08	930.9	20.41	0.918		
HPEO	65.41	2209.0	21.29	0.846	65.85	1525.0	17.81	0.858	67.03	2235.0	23.46	0.894	67.51	1452.3	18.47	0.912		
HDRCNN	69.78	1700.6	22.61	0.882	69.71	1021.5	20.22	0.894	71.39	1150.5	25.29	0.935	71.64	621.7	22.87	0.947		
DRTMO	68.89	1423.5	22.42	0.877	68.80	1054.3	19.99	0.882	70.38	736.8	23.96	0.931	70.82	451.2	24.15	0.943		
Ours	70.26	757.1	22.58	0.916	70.31	550.1	22.93	0.916	71.98	294.9	23.98	0.956	72.45	146.7	28.32	0.963		

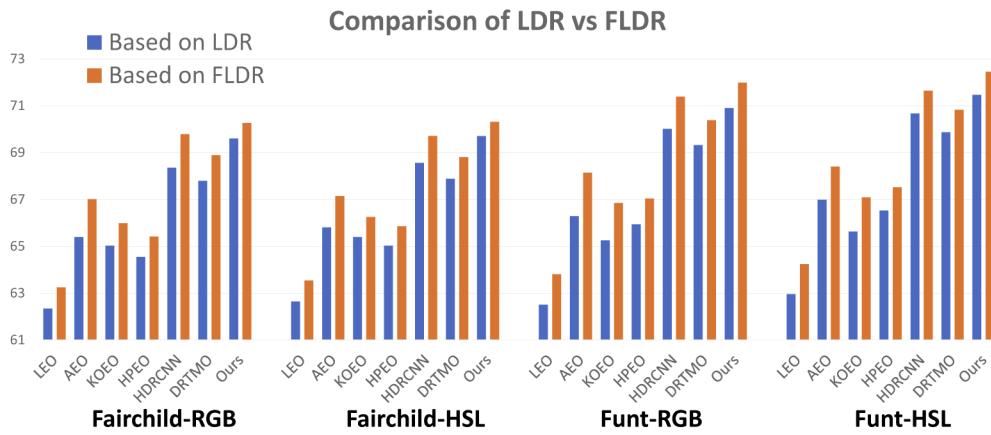


Fig. 6. Comparison of the performance scores of all HDR reconstruction methods under different input images. Blue means based on LDR images, and brown means based on FLDR images. **Fairchild** and **Funt** test datasets and **RGB** and **HSL** color spaces are included.

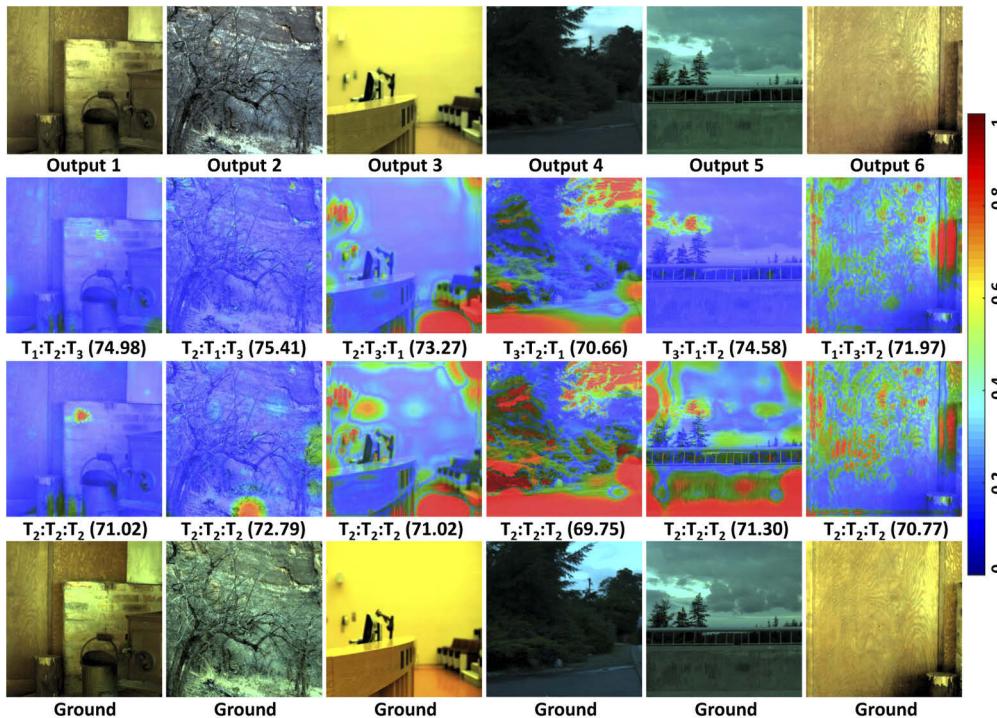


Fig. 7. Performance comparison of HDR reconstruction from FLDR images with different filtering schemes. **The first line** represents the HDR prediction images from six different filtering schemes; **the second line** represents the probability maps and quality scores of these six HDR prediction images; **the third line** represents the probability maps and quality scores of the HDR prediction images from the neutral density filtering scheme; **the fourth line** represents the ground truth HDR images.

4.4. Experiment based on retrained FLDR2HDR

Furthermore, we retrained the DITMnet network in an end-to-end manner on the FLDR images of the DML-HDR database. To verify the effectiveness of our complete system, we generated HDR images by the new parameters of our DITMnet and FLDR images as input and compared the performance of all methods in HSL color space. The models and parameters of the other methods are the same as those in Section 4.2, but the input is LDR images. Figure 8 (left) shows a visual qualitative comparison of the different HDR reconstruction methods on the retrained FLDR images, and Table 4 shows a statistical quantitative comparison.

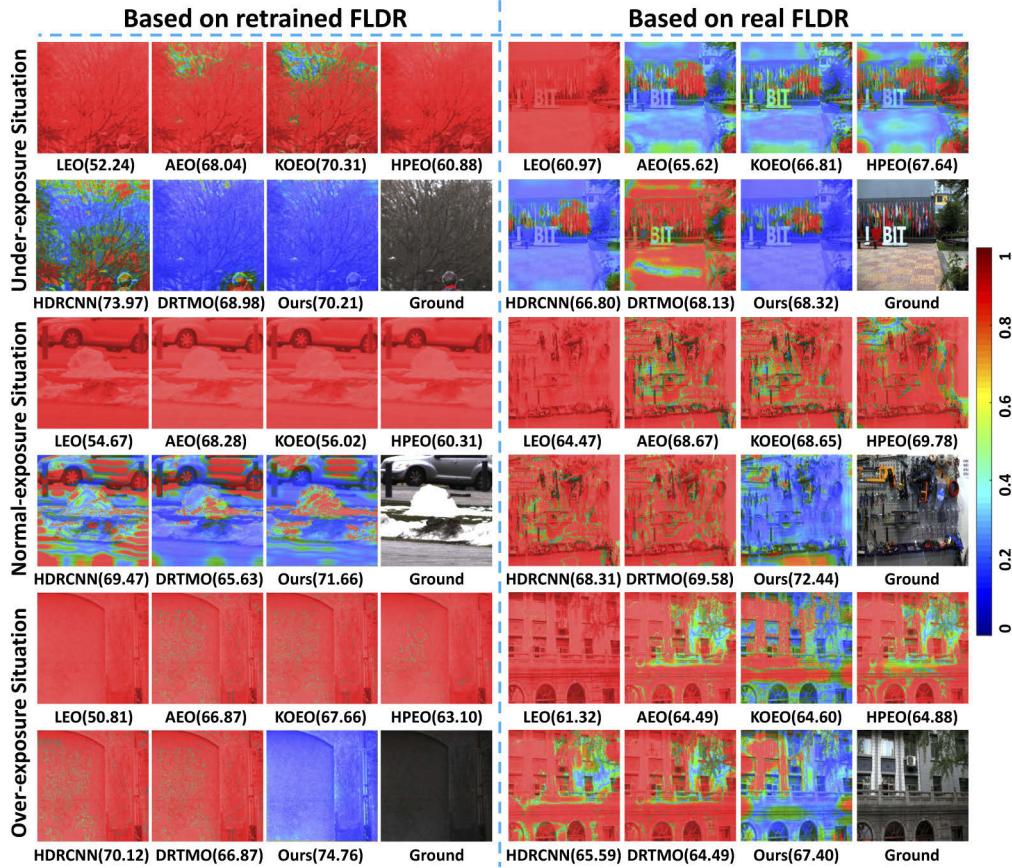


Fig. 8. Visual comparison probability maps and quality scores on different exposure states (under-/ normal-/ overexposure) on the retrained FLDR images (left) and real FLDR images (right), respectively.

As seen in Fig. 8 (left), the performance of the last three deep learning methods is significantly better than that of the first four traditional methods. When the irradiance changes drastically in some scenes, such as outdoor woods and snow, the performance of HDRCNN [22] and DRTMO [23] are better than those of the other traditional algorithms, but still below ours. When the irradiance changes smoothly, such as indoor corridors, the performance of our method is considerably more prominent than that of the other methods. Table 4 shows that the performance scores of our DITMnet network almost exceed those of other methods.

In summary, our complete system is more conducive to reconstruct HDR images under various conditions.

Table 4. Average score of all methods on the different exposure states on the DML-HDR database. Bold values indicate the best values. 0.025in

Method	LEO	AEO	KOE	HPEO	HDRCNN	DRTMO	Ours
State							
Under-exposure							
VDP	56.07	62.73	62.51	57.83	65.77	67.20	67.91
MSE	2032.40	623.00	567.30	1086.80	428.10	528.20	213.30
PSNR	18.67	23.46	22.65	18.91	24.06	23.13	27.81
FSIM	0.8583	0.9399	0.9455	0.9104	0.9554	0.9535	0.9688
State							
Normal-exposure							
VDP	54.73	58.84	57.19	58.27	63.02	62.83	64.23
MSE	3464.60	2564.10	1172.00	1701.80	859.60	809.10	514.10
PSNR	14.48	17.43	20.20	18.46	23.14	21.90	26.19
FSIM	0.7268	0.8758	0.9021	0.8948	0.9302	0.9307	0.9558
State							
Over-exposure							
VDP	53.14	54.03	53.56	54.76	59.75	54.03	61.14
MSE	4982.60	7035.50	12092.50	5549.20	1225.60	1640.60	1187.70
PSNR	12.20	11.72	8.36	13.03	20.46	18.18	21.53
FSIM	0.6660	0.7630	0.5073	0.7994	0.8942	0.8657	0.8982
State							
Total results							
VDP	54.65	58.53	57.75	56.95	62.85	61.35	64.43
MSE	3493.20	3407.53	4610.60	2779.27	837.77	992.63	638.37
PSNR	15.12	17.54	17.07	16.80	22.55	21.07	25.18
FSIM	0.7504	0.8596	0.7850	0.8682	0.9266	0.9166	0.9409

4.5. Experiment based on real photographs

To verify the effectiveness of our single-shot HDR reconstruction system for real images, we used a Canon EOS-1D X Mark II camera with a real SJB130 optical filter to obtain several sets of multiple-exposure images and synthesized them to HDR images as ground truth using standard methods [9]. In each group, three different exposure FLDR images were selected as input to verify the effectiveness of our method on real FLDR images. Figure 8 (right) shows the

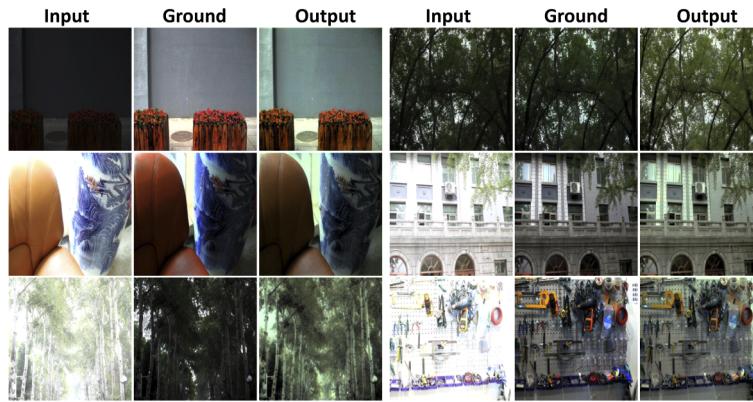


Fig. 9. Reconstruction examples of real Canon camera images in different scenarios.

probability maps and quality scores of all the methods on the real FLDR images. Figure 9 shows examples of reconstructions of HDR images from real camera images in different scenarios from our single-shot HDR reconstruction method.

In short, our method exhibits not only excellent performance on synthetic images but also good reconstruction effects on real images.

5. Conclusion and future work

This paper presents an improved single-shot HDR reconstruction system. This system can obtain an FLDR image with different exposure states and luminance ranges for three channels in a single shot through an optical filter and a camera, retaining more reliable luminance information, and can reconstruct an HDR image from a single FLDR image through the DITMnet network proposed in this paper. The experimental results show that our method exhibits better HDR reconstruction performance than previous methods and provides research guidance for the design of single-shot HDR reconstruction equipment.

Limitations. At present, our method requires a large number of computer resources. If the method relies on GPU computing, then at least 9 G of memory is required for smooth performance. Increasing the number of output images would yield better performance and a higher dynamic range for HDR images, but more computing resources and better hardware would be needed.

Future work. Next, we will further optimize the network and add a preprocessing sub-network to combine FLDR and LDR data to further enhance the image color reconstruction. Second, we will improve the network and achieve more automation. In the multioutput image synthesis step of this paper, the weight is obtained from a series of small sample data, which is relatively reasonable but not necessarily optimal. In the future, we intend to design a subnetwork to automatically learn more reasonable weights.

Appendix

Appendix.A: According to Eq. (7), the mapping relation of the log-domain HDR image is obtained:

$$\begin{aligned} H_i &= h^{-1}(L_i), h^{-1} : [0, 255] \rightarrow \mathbb{R}^+, \\ \ln H_i &= \ln(h^{-1}(L_i)) = k(L_i). \end{aligned} \quad (7)$$

where $k = \ln h^{-1}$.

Therefore, there is a corresponding mapping relationship between the LDR image and the HDR image, as given by Eq. (7).

Appendix.B

The camera imaging process according to Debevec et al. [9] is shown in Eq. (8):

$$\begin{aligned} L_{i,j} &= f(H_i \Delta t_j), \\ f^{-1}(L_{i,j}) &= H_i \Delta t_j, \end{aligned} \quad (8)$$

Without loss of generality, it is assumed that Δt_k is linearly scaled to $\alpha \Delta t_j$ with a coefficient α . Let $g = f^{-1}$, $L_{i,k}$ represents the LDR image generated under the exposure Δt_k , then

$$\begin{aligned} H_i &= \frac{g(L_{i,j})}{\Delta t_j} = \frac{g(L_{i,k})}{\Delta t_k} = \frac{g(L_{i,k})}{\alpha \Delta t_j}, \\ g(L_{i,j}) &= \frac{g(L_{i,k})}{\alpha}, \\ L_{i,j} &= f\left(\frac{g(L_{i,k})}{\alpha}\right). \end{aligned} \quad (9)$$

Therefore, there is an association mapping relationship between multiple-exposure LDR images, as given in Eq. (9).

Funding

the Key-Area Research and Development Program of Guangdong Province (No.2019B010149001); Higher Education Discipline Innovation Project (B18005); National Natural Science Foundation of China (No.61731003).

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).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *ACM SIGGRAPH 2008 classes*, Association for Computing Machinery, (2008), 1–10.
2. P. Fei, Z. Yu, X. Wang, P. J. Lu, Y. Fu, Z. He, J. Xiong, and Y. Huang, "High dynamic range optical projection tomography (hdr-opt)," *Opt. Express* **20**(8), 8824–8836 (2012).
3. C. Marchessoux, P. L. De, O. Vanovermeire, and L. Albani, "Clinical evaluation of a medical high dynamic range display," *Med. Phys.* **43**(7), 4023–4031 (2016).
4. D. Giassi, B. Liu, and M. B. Long, "Use of high dynamic range imaging for quantitative combustion diagnostics," *Appl. Opt.* **54**(14), 4580–4588 (2015).
5. A. A. Adeyemi, N. Barakat, and T. E. Darcie, "Applications of digital micro-mirror devices to digital optical microscope dynamic range enhancement," *Opt. Express* **17**(3), 1831–1843 (2009).
6. F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli, "Flexisp: A flexible camera image processing framework," *ACM Trans. Graph.* **33**(6), 1–13 (2014).
7. A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum* **35**(2), 153–163 (2016).
8. G. Chen, L. Li, W. Jin, J. Zhu, and F. Shi, "Weighted sparse representation multi-scale transform fusion algorithm for high dynamic range imaging with a low-light dual-channel camera," *Opt. Express* **27**(8), 10564–10579 (2019).
9. P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, Association for Computing Machinery, (2008), 1–10.
10. M. Á. Martínez-Domingo, E. M. Valero, J. Hernández-Andrés, S. Tominaga, T. Horiuchi, and K. Hirai, "Image processing pipeline for segmentation and material classification based on multispectral high dynamic range polarimetric images," *Opt. Express* **25**(24), 30073–30090 (2017).
11. 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. Multimed.* **17**(4), 470–484 (2015).
12. I. Pekkucuksen and Y. Altunbasak, "Multiscale gradients-based color filter array interpolation," *IEEE Trans. on Image Process.* **22**(1), 157–165 (2013).
13. S. R. M. Nezam, L. Yan, J. E. McGeehan, Y. Shi, A. E. Willner, and S. Yao, "Enhancing the dynamic range and dgd monitoring windows in dop-based dgd monitors using symmetric and asymmetric partial optical filtering," *J. Lightwave Technol.* **22**(4), 1094–1102 (2004).
14. J. Brauers and T. Aach, "Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras," *IEEE Trans. on Image Process.* **20**(2), 496–505 (2011).
15. R. Shrestha, A. Mansouri, and J. Y. Hardeberg, "Multispectral imaging using a stereo camera: Concept, design and assessment," *EURASIP J. on Adv. Signal Process.* **2011**(1), 57 (2011).
16. H. Landis, "Production-ready global illumination," Siggraph course notes **16**, 11 (2002).
17. A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do hdr displays support ldr content? a psychophysical evaluation," *ACM Trans. Graph.* **26**(3), 38 (2007).
18. F. Banterle, P. Ledda, K. Debattista, A. Chalmers, and M. Bloj, "A framework for inverse tone mapping," *The Vis. Comput.* **23**(7), 467–478 (2007).
19. R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *2014 27th SIBGRAPI Conf. Graph. Patterns Images*, (IEEE, 2014), pp. 49–56.

20. Y. Huo, Y. Fan, D. Le, and V. Brost, "Physiological inverse tone mapping based on retina response," *Vis. Comput.* **30**(5), 507–517 (2014).
21. J. Zhang and J.-F. Lalonde, "Learning high dynamic range from outdoor panoramas," in *2017 IEEE International Conference on Computer Vision*, (IEEE, 2017), pp. 4529–4538.
22. 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.* **36**(6), 1–15 (2017).
23. Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.* **36**(6), 1–10 (2017).
24. D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," *Comput. Graph. Forum* **37**(2), 37–49 (2018).
25. K. R. Prabhakar, V. S. Srikanth, and R. V. Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), pp. 4714–4722.
26. B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimed. Tools Appl.* **76**(1), 631–648 (2017).
27. S. Lee, G. Hwan An, and S.-J. Kang, "Deep recursive hdri: Inverse tone mapping using generative adversarial networks," in *Proceedings of the European Conference on Computer Vision*, (2018), pp. 596–611.
28. Y.-L. Liu, W.-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang, "Single-image hdr reconstruction by learning to reverse the camera pipeline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), pp. 1651–1660.
29. R. Raskar and J. Tumblin, *Computational photography: mastering new techniques for lenses, lighting, and sensors* (AK Peters, Ltd., 2009).
30. B. Liang, D. Weng, Y. Bao, Z. Tu, and L. Luo, "Reconstructing hdr image from a single filtered ldr image base on a deep hdr merger network," in *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, (IEEE, 2019), pp. 257–258.
31. M. H. Kim and J. Kautz, "Consistent tone reproduction," in *Proceedings of the Tenth IASTED International Conference on Computer Graphics and Imaging*, (ACTA Press Anaheim, 2008), 152–159.
32. A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill* **1**(10), e3 (2016).
33. M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?" in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2 (IEEE, 2003), pp. II–602.
34. M. Narwaria, R. Mantiuk, M. P. Da Silva, and P. Le Callet, "Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imaging* **24**(1), 010501 (2015).