Fast and Accurate Illumination Estimation Using LDR Panoramic Images for Realistic Rendering

Haojie Cheng, Chunxiao Xu, Jiajun Wang, Zhenxin Chen, and Lingxiao Zhao

Abstract—A high dynamic range (HDR) image is commonly used to reveal stereo illumination, which is crucial for generating high-quality realistic rendering effects. Compared to the high-cost HDR imaging technique, low dynamic range (LDR) imaging provides a low-cost alternative and is preferable for interactive graphics applications. However, the limited LDR pixel bit depth significantly bothers accurate illumination estimation using LDR images. The conflict between the realism and promptness of illumination estimation for realistic rendering is yet to be resolved. In this paper, an efficient method that accurately infers illuminations of real-world scenes using LDR panoramic images is proposed. It estimates multiple lighting parameters, including locations, types and intensities of light sources. In our approach, a new algorithm that extracts illuminant characteristics during the exposure attenuation process is developed to locate light sources and outline their boundaries. To better predict realistic illuminations, a new deep learning model is designed to efficiently parse complex LDR panoramas and classify detected light sources. Finally, realistic illumination intensities are calculated by recovering the inverse camera response function and extending the dynamic range of pixel values based on previously estimated parameters of light sources. The reconstructed radiance map can be used to compute high-quality image-based lighting of virtual models. Experimental results demonstrate that the proposed method is capable of efficiently and accurately computing comprehensive illuminations using LDR images. Our method can be used to produce better realistic rendering results than existing approaches.

Index Terms—Illumination estimation, LDR panoramic image, image-based lighting, realistic rendering

1 Introduction

Realistic rendering helps improve immersion of virtual models in real-world scenes [1]. The illumination model plays an important role in realistic rendering and is used to simulate the light transmission [2]. High dynamic range (HDR) images can be used to provide sufficient and precise illumination information for realistic rendering [3]. The HDR imaging technique is able to accurately measure the illumination of a real-world scene and generate an omnidirectional HDR radiance map using specific photographing techniques and gadgets, e.g., a high-resolution HDR camera [4], mirrored sphere [2] and fisheye lens [5]. However, these ways of capturing HDR images have two limitations. First, multiplexing exposures in the time domain are usually employed and may take a long period of time to photograph an HDR image. Therefore, instantly acquiring HDR images of a dynamic environment is currently an extremely challenging task for real-time applications such as mixed reality (MR). Second, HDR photograph devices are usually much more expensive than ordinary photograph devices and less popular in the commercial market.

Inferring accurate and realistic illumination from a low dynamic range (LDR) image as inferred from an HDR image offers a beneficial trade-off between the hardware cost, capturing speed and image quality [6]. Each pixel intensity of the LDR image is determined by the camera imaging system and subsequent post-processing steps, such as tone mapping [7]. Parameters used in these photographing processes are crucial for accurate illumination estimation. Since the LDR imaging technique produces a limited pixel bit depth, important lighting information is usually discarded. Several original photographing parameters cannot be directly obtained or easily derived from a single LDR image. Therefore, predicting real-world illuminations using LDR images is actually a tough challenge.

Traditional image-based illumination estimation methods are based on certain characteristics of statistical priors, e.g., color mixture of local edge regions [8] and image noise distribution [9]. Nonetheless, obtaining such information can be cumbersome and time-consuming. In recent years, deep learning methods have successfully shown their potential to estimate illuminations with weak constraints. Some research studies [10], [11], [12] focused on inferring the panoramic HDR radiance map from an image of a limited field of view (FOV). Some other methods [13], [14] observed the surface reflection or geometrical information of objects in an image to predict properties of light sources. However, unknown geometries and arbitrarily complex scenes still cannot be easily handled only using trained models.

- *Haojie Cheng, Chunxiao Xu, Jiajun Wang, and Lingxiao Zhao are with the University of Science and Technology of China, Hefei 230052, China, and also with the Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215123, China. E-mail: {cmm0418, feimos}@mail.ustc.edu.cn, {wangjj, hitic}@sibet.ac.cn.*
- *Zhenxin Chen is with the Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215123, China. E-mail: chenzhenxin@sibet.ac.cn.*

Manuscript received 1 July 2022; revised 7 September 2022; accepted 7 September 2022. Date of publication 12 September 2022; date of current version 10 November 2023.
This work was supported in part by the China National Key R&D Plan under Grant 2020YFC2003900.
(Corresponding author: Lingxiao Zhao.)
Recommended for acceptance by D. Iwai.
This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2022.3205614, provided by the authors.
Digital Object Identifier no. 10.1109/TVCG.2022.3205614

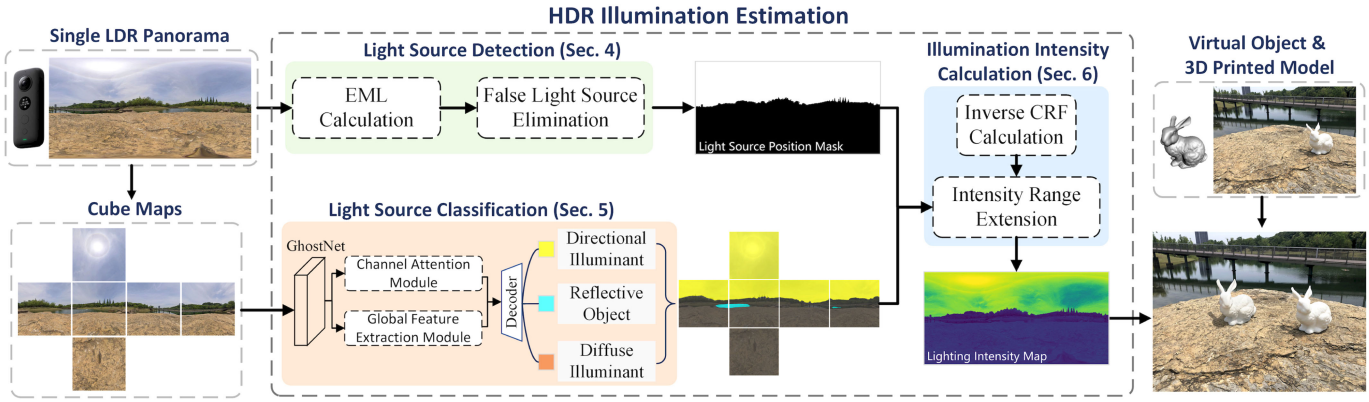


Fig. 1. Overview of our illumination estimation approach using a single LDR panorama for realistic rendering.

In this paper, a novel approach is proposed for illumination estimation in realistic rendering using panoramic LDR images. Our approach tries to solve the conflict between realism and promptness of image-based illumination estimation. Basically in our idea, the realistic illumination model is formulated into three important aspects, namely location, type and intensity of light sources. Light sources are first differentiated from non-illuminants according to their patterns of pixel value distribution in LDR images. Based on the knowledge that shading and shadowing effects of realistic rendering are related to light source type, true light source detections are then classified into different types with various illumination behaviors. Lighting intensities can be more precisely calculated based on previously estimated locations and types of light sources.

As elaborated in Fig. 1, our method consists of three steps. A panoramic image is first converted to an equivalent melanopic lux (EML) image [15] for the initial positioning of light sources. We develop a new algorithm to extract distinctive characteristics of true light sources during the exposure attenuation procedure and remove false light source detections. Second, a new convolutional neural network (CNN) model is employed to further classify detected light sources. Finally, inspired by the camera imaging mechanism, the inverse camera response function (CRF) is estimated to generate the sensor irradiance. Then, the dynamic range of pixel intensities in the LDR panorama is extended by taking previously estimated lighting parameters into account. In our prototype software program, these intensive computations are implemented on the GPU using CUDA with hardware acceleration. Experimental results demonstrate that the proposed method can efficiently handle arbitrarily complex scenes and provide realistic illuminations for creating better rendering effects.

The main contributions of our paper are outlined as follows:

- We develop a hybrid method to efficiently estimate precise illuminations using LDR panoramic images. It computes multiple parameters of light sources and can robustly handle various illumination conditions in both indoor and outdoor scenes.
- A new algorithm that efficiently and accurately identifies locations of light sources in LDR images is proposed. It precisely characterizes behavior patterns of illuminants during the image exposure attenuation procedure and removes false light sources detections.
- We develop an integrative method to calculate lighting intensities. Our method takes the correspondence between light source types and lighting effects into account and classifies detected light sources using a new CNN model. It can precisely recover the CRF and extend the dynamic range of lighting intensities based on light source locations and types.

2 Related Research

To realistically render a virtual model, key properties including its material, geometry and illumination should be taken into account [6]. In this section, previous research on simulating the real-world illumination is outlined.

Using Reference Objects. A reference object can be used as the light probe to capture the HDR radiance map in real-world scenes. Debevec [2] first demonstrated that the mirrored ball-like light probe can be inserted into the scene space to collect information of the omnidirectional HDR illumination. Similarly, LeGendre et al. [16] leveraged three different reflective spheres to reveal different lighting cues in a single exposure. However, inserting additional gadgets into the scene space may sacrifice the spatial scope of a user activity. To address this problem, various methods using reference objects in the image as light probes have been proposed. Park et al. [17] inferred the HDR illumination by referring to a special object with homogeneous materials and mostly convex shapes. Some other methods estimate illuminations from faces [14], [18] or known 3D objects [13], [19], [20]. However, it is usually unrealistic for arbitrary real-world scenes to contain these reference objects.

Light Source Detection. The accurate positioning of light sources is crucial for high-quality illumination estimation. For real-time applications, some methods pursue the efficiency over the precision. Rhee et al. [21] developed an interactive MR system, in which light sources were identified for real-time image-based shadowing. However, the boundary calculation of light sources is fairly rough in their method. Morgand et al. [22] computed the brightest point on the convex surface of an object and predicted its specularity according to the new viewing position. To differentiate true light sources from false detections, Fu et al. [23] made use of a large-scale specularity dataset for detecting

highlights. Karsch et al. [24] segmented an image into superpixels using SLIC [25] and trained a binary classifier to detect visible light sources in the image. Similarly, Gardner et al. [10] manually annotated light sources from the SUN360 dataset [26] and trained two logistic regression classifiers for identifying different types of light sources. The light source boundary calculated using their method can be improved.

Prediction Based on Images of Limited FOV. The research on predicting wide-range HDR illuminations directly from the image of a limited FOV is becoming increasingly popular. For outdoor scenes, Lalonde et al. [27] applied a data-driven method to combine weak cues, e.g., shadows and sky colors, for estimating illuminations. Zhang et al. [28] used the Lalonde-Matthews sky model [29] to robustly handle all weather conditions. For indoor scenes, Gardner et al. [10] built the Laval Indoor HDR Dataset. Based on this dataset, Zhan et al. [30], [31] regressed the illumination distribution in real geometry spaces. Li et al. [32] presented a complete indoor scene reconstruction framework to estimate geometry, reflectance and illumination. To robustly handle any circumstances including indoor and outdoor scenes, Karsch et al. [24] developed a user-guided interface to create 3D structures of a scene. Subsequently, they [33] leveraged the SUN360 database [26] to further predict out-of-view illuminations. Due to the limited FOV, predictions of light source locations given by these methods are usually inaccurate. This may significantly reduce the rendering quality, especially when rendering virtual models with specular material surfaces. In our work, panoramic images are used for accurate illumination calculation.

Tonemapping Inversion. Inverse tone-mapping operators (iTMOs) can be used to enrich details of a single LDR image and better simulate a true HDR image. Independent of camera parameters, Lin et al. [8] analyzed the effect of the camera sensor on edge color distribution and predicted CRF. Rempel et al. [34] proposed a method using edge stopping function to reduce the influence of brightness enhancement in saturated regions, while the distortion was minimized using a human visual system model [35]. Some other methods [36], [37] synthesized multiple LDR images with different exposures and used them to reconstruct corresponding HDR images. With the fast development of deep learning techniques, CNN has been widely employed to recover missing details in over-exposed image regions [38], [39]. Santos et al. [40] presented a feature masking mechanism to avoid relying on invalid information in saturated regions. In general, iTMOs pay more attention to recovering details in saturated regions. But the accuracy of predicted intensities still needs to be improved.

3 Realistic Illumination Estimation Using LDR Images

Our research work tries to accurately estimate illuminations in real-world scenes using normal digital imaging techniques with low costs of acquisition and computation. As a costly imaging technique, HDR imaging can capture illuminations of a far higher dynamic range of exposures than low-cost LDR imaging. Light sources can be detected using a simple thresholding algorithm in an HDR image. Since the LDR pixel bit depth is rather limited, different types of objects in an LDR image may have almost identical pixel values. This brings significant difficulties for correctly differentiating their lighting properties. To resolve this issue, our solution starts with a more in-depth analysis of the rendering function [41] for figuring out key aspects of image-based illumination estimation:

$$L_o(p,\omega_o) = \int_{\Omega} f(p,\omega_o,\omega_i) L_i(p,\omega_i) |cos\theta_i| \, d\omega_i, \tag{1}$$

θ_i is the angle between an incident light direction ω_i and the surface normal at a point p on the object surface. When a panoramic image is used as a radiance map, the pixel intensity corresponding to ω_i is equal to the radiance $L_i(\cdot)$. More details are given in [41]. As described by the rendering function, *locations* (i.e., positions and areas) and *intensities* of light sources primarily determine the shading and shadowing effects on rendered objects. Therefore, accurate positioning and intensity estimation of light sources is critical for high-quality realistic rendering.

3.1 Patterns of Image Exposure Attenuation for Light Source Detection

The main subject of light source positioning is to correctly extract illuminant areas by precisely identifying pixels belonging to light sources. In particular, light sources that contribute significantly to illuminations around the camera should be accurately detected. As in our above discussion, it is not advisable to simply threshold pixel values of the limited bit depth for detecting light sources in LDR images. More discriminative LDR image features shall be extracted and analyzed to better differentiate illuminants and non-illuminants.

It can be acknowledged that the illuminant area of a visible light source normally exhibits a uniform illumination distribution of high pixel values in the image captured by the camera. A non-illuminant area may also have high pixel values but normally has a non-uniform illumination distribution. Such a rule applies to both HDR images and LDR images and can be taken as a strategic basis for distinguishing between light sources and other objects. To analyze the illumination distribution within an image area, a traditional way is to make use of histogram analysis. However, this is computationally expensive and does not suit for interactive graphics applications in our case. We tried to explore LDR image features that characterize the uniform distribution of high-value pixels for detecting light sources along with a new perspective other than stationary pixel information. We found that changes in light source areas are much more isotropic than in non-illuminant areas during the process of image exposure attenuation (Fig. 2). Such dynamic characteristics of light sources inspire a promising way for the fast and accurate light source detection in LDR images. This idea is somehow similar to the object detection based on temporally dynamic features represented in the image sequence of a video [42]. However, it is not based on motion patterns.

Our idea for positioning light sources is to identify illuminant areas in LDR images with uniform high-value pixel distribution. It extracts dynamic patterns of image exposure

Fig. 2. Light sources and non-luminous objects exhibit different characteristics during the image exposure attenuation procedure. When the image exposure gradually decreases, true illuminants behave isotropic changes, while non-luminous objects represent brightness convergences on their surfaces toward light sources.

attenuation instead of relying on LDR pixel values of limited bit depth. Light source locations are initially detected based on EML [15]. Since areas unbelonging to light emitters in an LDR image may also have high pixel values, false detections are usually inevitable. To further remove false light source detections, the LDR image exposure is gradually attenuated, during which patterns of how initially detected illuminant areas change are analyzed. Initial light source detections that behave isotropy in this procedure remain as true illuminants.

3.2 Realistic Illumination Intensity Recovery from LDR Images

As inspired by Eq. (1), illumination intensities shall be precisely calculated after positioning light sources to generate realistic shading and shadowing effects. Our work tries to develop an approach that allows reconstructing HDR-like illumination intensities from LDR images. Real-world illumination intensities E usually have a high dynamic range. Illumination intensities estimated directly from LDR pixel values can hardly support realistic rendering due to a significant reduction of lighting information.

Normal digital camera sensors can only capture 8-bit illumination intensities and result in a limited pixel bit depth. The LDR image I captured using a camera imaging system can be represented as:

$$I = \mathcal{F}(\mathcal{Q}(E)), \tag{2}$$

An ordinary LDR camera first compresses and clips HDR intensities into a limited range. The compressed image (i.e., image irradiance) can be formulated as $I_{\mathcal{Q}} = \mathcal{Q}(E) = min(E, 1)$. To match the human perception of the scene, a non-linear CRF (denoted by $\mathcal{F}$) is then used to adjust the contrast of $I_{\mathcal{Q}}$. Inspired by Eq. (2), an inverse camera imaging mechanism can be simulated to precisely estimate illumination intensities. An existing inverse CRF algorithm [43] can be adopted to optimally adjust the color of an LDR image. The contribution of light sources to the environment lighting can also be leveraged to extend the dynamic range of illumination intensities calculated using LDR images.

We also noticed that different illuminants may present various intensities in HDR images. However, due to the fact that pixel dynamic range is severely squeezed, this phenomenon does not occur in LDR images. This defect of an LDR image troubles the accurate estimation of illumination intensities. According to [44], there can be various types of light sources in a real-world scene. The luminous level of a light source corresponds to its specific type to a certain extent. The type information can further help to determine the realistic illumination intensity of a detected light source and shall be taken into account in the accurate illumination estimation. Therefore, in our approach, previously detected light sources are classified before their illumination intensities are calculated.

Deep learning techniques have been widely used for the image classification. Such data-driven approaches can better satisfy the demands of both low time cost and high accuracy than traditional classifiers (e.g., machine learning methods). For this purpose, a new CNN model for the semantic segmentation is adopted to classify detected light sources in our approach. It contains the lightweight global feature extraction and channel attention modules. The training dataset contains indoor and outdoor scenes and can offer preferable robustness and facilitation for our illumination estimation.

4 Light Source Detection in LDR Images

Light source detection is a prerequisite step for estimating illuminations. Our light source detection method consists of two steps: initial light source positioning and the removal of false detections.

4.1 Initial Light Source Positioning

In computer graphics, the RGB color model has been widely used to support human perception and digital image display in electronic systems. To recover the lighting information in an LDR image as much as possible, the dynamic range of its RGB pixel values is first expanded by mapping it to an EML image [15]. The resulting EML image has a wide dynamic range and can be used to quantify circadian lighting.

An LDR image pixel is classified as part of a light source if its EML value is larger than a proper threshold value, and vice versa. To produce a minimal set of discrete light sources, a breadth-first search is performed to determine connectivities between light source masks. In such a way, the effect caused by the limited pixel bit depth for detecting light sources in an LDR image can be alleviated. Nonetheless, the corresponding EML value is still limited by the original RGB pixel value. False detections are usually generated and need to be further removed.

4.2 Removal of False Light Source Detections

According to photon absorption and scattering mechanisms [45], the image exposure attenuation represents characteristic illumination features that can be used to distinguish between light sources and non-luminous objects. Based on the theory in Section 3.1, all the pixels of each light source detection can be used as input to analyze the light distribution. For real-time applications, we can detect light sources in each frame of the image series. Our time-independent algorithm for removing false light source detections in an LDR image is briefly described in Alg. 1.

In digital imaging processing [45], compared with the low-intensity irradiance, high-intensity irradiance is not easily affected by changes in different camera exposure settings. Accordingly, if the exposure of an LDR image decreases gradually, pixels of low-intensity values should vary more sensitively than pixels of high-intensity values.

The attenuated intensity of a pixel I' can be defined as the following:

$$I'(\lambda) = max\left(I \cdot \left(\hat{I}^{\sigma} + \lambda\right)^{\frac{1}{\sigma}}, 0\right), \tag{3}$$

where I and $\hat{I}$ are the initial and normalized intensities of a pixel respectively. The factor σ is used to approximately simulate inverse CRF. $\lambda \in (-1, 0]$ denotes the attenuation offset for the dynamic exposure attenuation simulation.

Algorithm 1. Removal of False Light Source Detections

Input: The region I of the initial light source $\mathcal{O}$ cropped from an LDR image.
Output: $\mathcal{O}$ is a real light source or not.

```
Split I into M regular sub-regions;
l = 0;
for m ← 1 to M do
    mth region is pooled to a patch of 2 × 2;
    foreach I'_k(λ_n) in Eq. (4) calculated by Eq. (3) do
        Construct global attenuation matrix C_m;
    end
    Calculate projected matrix S_m using Eq. (5);
    Calculate standard deviation R_m using Eq. (6);
    if R_m ≠ 0 then
        l = l + 1;
    end
end
if l > 0 then
    O is a false light source;
else
    O is a real light source;
end
```

To quantify the exposure attenuation, a simple way is to average attenuations of all pixels. However, camera sensor noises can cause subtle variations of measured pixel values and are usually inevitable [42]. To solve this issue, the area of each initial light source detection is pooled to the 2×2 resolution to reduce camera quantization errors before λ decreases. This operation also helps improve the computational efficiency of our algorithm. But for large light source detections that have irregular boundaries, their geometric structures are ignored by the pooling operation. In such a case, more false detections can be generated and the detection accuracy may decrease. To solve this problem, our method splits each irregular initial light source detection into M relatively regular patches using the method based on image moment functions [46]. More details are given in the supplementary material of our paper, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TVCG.2022.3205614.

During the image exposure attenuation process, all attenuated intensities of the kth pixel $\hat{I}'_k = \{I'_k(\lambda_1), I'_k(\lambda_2), \ldots, I'_k(\lambda_N)\}$ are integrated into a local attenuation function. To globally analyze the illumination of the mth region ($m \in [1, M]$), local attenuation functions of all pixels can be further integrated into the global attenuation function $\mathbf{C}_m$:

$$\mathbf{C}_m = \sum_{n=1}^{N} \sum_{k=1}^{4} I'_k(\lambda_n). \tag{4}$$

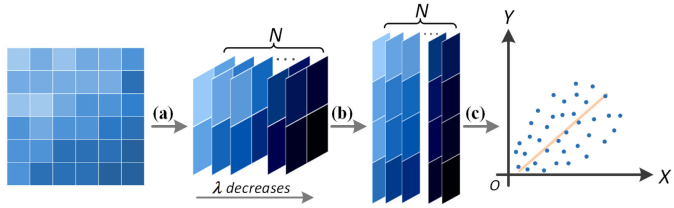


Fig. 3. Our method to support the quantification of the exposure attenuation performed N times: (a) Pooling pixels of the detection to N patches of 2×2; (b) Patches are sorted according to their intensities from the largest to the smallest and reformed to 4×1; (c) N patches are plotted onto a 2D orthogonal system and a fitted eigenvector is computed to verify whether the corresponding detection is a true light source.

When λ decreases for N times, an initial light source detection can be first converted to a set of N patches (Fig. 3a). Each patch has only four pixels. These patch pixel values usually have subtle differences from each other. $\mathbf{C}_m$ is a linear mixture of source signals. These source signals can be compressed using linear projection and dimensionality reduction algorithms. Characteristics of the illumination convergence can then be practically analyzed. For this reason, pixels of an initial light source are adjusted from $\mathbf{C}_m \in \mathbb{R}^{2\times2\times N}$ to $\mathbf{D}_m \in \mathbb{R}^{4\times N}$ (Fig. 3b). Regarding that higher intensities make more important contributions to the illumination effect, all pixel intensities of the mth region are sorted from the largest to the smallest. $\mathbf{D}_m$ can then be projected onto a plane. In our case, the plane orthogonal to a vector $\mathbf{1} = [1\,1\,1\,1]^T$ is used for such a projection (Fig. 3c).

$$\mathbf{S}_m = P \cdot \mathbf{D}_m, \tag{5}$$

where $P \in \mathbb{R}^{2\times4}$ denotes the orthogonal projection matrix. The global attenuation function $\mathbf{D}_m$ can be converted to a two-dimensional function $\mathbf{S}_m$. This helps analyze the illumination variation of an object in the image more intuitively.

Since a false detection is not a true light source, the illumination in its area should gradually converge in a certain direction. The main direction of the mth region in an image can be measured by:

$$R_m = \mathcal{M}(\mathcal{T}(\mathbf{S}_m)), \tag{6}$$

where $\mathcal{M}(\cdot)$ is used for calculating the standard deviation and $\mathcal{T}(\cdot)$ denotes the eigenvalue calculation of a matrix using a singular value decomposition solution. If all R_m are close to zero, the main direction of the local lighting cannot be distinguished. In such a case, the object is a real light source, and vice versa.

5 Light Source Classification

Different types of light sources normally have different lighting characteristics. As described in Section 3.2, an accurate classification of light sources helps determine realistic illumination intensities. In our work, a CNN-based semantic segmentation model is designed to classify light sources previously detected in an LDR image.

5.1 Dataset for Training Our CNN Model

For training a robust semantic segmentation model, a large panoramic image dataset with light source annotations is

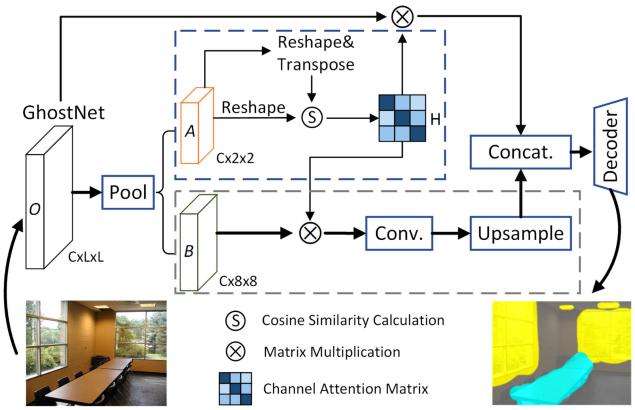


Fig. 4. An overview of our light source classification using a CNN-based semantic segmentation model. The channel attention and global feature extraction modules are included in blue and gray dash boxes respectively.

required. However, there are no such image datasets publicly available. It would also cost a pretty much effort to conduct annotating. Fortunately, the ADE20K dataset [47] can be used to train our CNN model and offers a sufficient amount of indoor and outdoor stuff/object labels and image-level scene descriptors. Based on different object properties and human experiences, labels in the ADE20K dataset represent four classes, including directional illuminants (or infinite illuminants), diffuse illuminants (or non-directional illuminants), reflective objects and non-luminous objects. All images in this dataset should be resized to the same resolution of $L \times L$ for training our CNN model.

5.2 CNN Architecture Design

Parsing complex scenes can be very challenging due to mismatched relationships and inconspicuous classes. These problems are related to global information of different receptive fields and contextual relationships. A CNN model configured with suitable attention modules can help obtain more reliable illumination prediction results.

The CNN architecture of our light source classification model is illustrated in Fig. 4. The input requires RGB images and the output image contains classified light source segmentations. To guarantee a suitable computational efficiency, the GhostNet [48] model is selected as the backbone of our CNN model structure to perform fast image feature extraction. Feature maps are then fed into two layers of different scales using the average pooling. The lower resolution layer $A \in \mathbb{R}^{C\times 2\times 2}$ is used as the input to the *channel attention module*. The higher resolution layer $B \in \mathbb{R}^{C\times 8\times 8}$ is used as the *global feature extraction module*, which is finally concatenated with the local feature layer $O \in \mathbb{R}^{C\times L\times L}$ extracted using the GhostNet model.

Each channel map of high-level feature maps can be treated as a class-specific response. The channel attention module is built to maintain explicit interdependencies between different channel maps. As in the blue dash box shown in Fig. 4, the feature layer $A \in \mathbb{R}^{C\times 2\times 2}$ is reshaped to $A' \in \mathbb{R}^{C\times 4}$. Then a cosine similarity calculation is performed between A' and its transpose to obtain the channel attention map $H \in \mathbb{R}^{C\times C}$:

$$h_{rc} = \beta \cdot \frac{exp(A'_r \cdot A'_c)}{exp(\|A'_r\| \cdot \|A'_c\|)}, \tag{7}$$

where h_{rc} measures the impact of the cth channel on the rth channel. When network parameters gradually converge during the model training, channel weights become increasingly reliable. We set a scale factor β that gradually increases along with training iterations.

For most real-world scenes, the total image area covered by light sources is usually less than that covered by non-luminous objects. To avoid inaccurate classification results caused by the unbalanced intra-class distribution, the global loss function $\mathcal{L}$ used for training our CNN model is defined as a linear combination of the weighted dice loss function $\mathcal{L}_{Dice}$ and the focal loss function [49] $\mathcal{L}_{FL}$:

$$\mathcal{L} = \omega_1 \cdot \mathcal{L}_{Dice}(\zeta) + \omega_2 \cdot \mathcal{L}_{FL}, \tag{8}$$

where weighted factors are set as $\omega_1 = 0.1$ and $\omega_2 = 0.9$ to denote the importance of different loss functions. Values in the array $\zeta = \{0.3, 0.3, 0.3, 0.1\}$ correspond to weights of reflective objects, diffuse illuminants, directional illuminants and non-luminous objects respectively.

5.3 Optimization of Classification Results

The proposed CNN model classifies detected light sources in an LDR image according to their luminous properties instead of luminous states. For example, a lamp may be turned off and the night sky is almost non-luminous. Nonetheless, they are still identified as light sources using our CNN model. Therefore, we need to recognize light sources that are truly luminous. Our classification results should be further optimized according to the actual luminousness in the panoramic image.

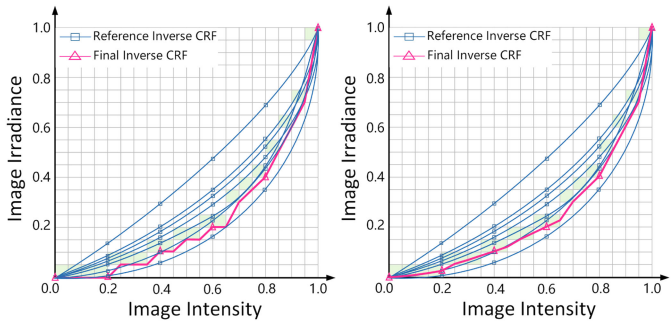
To verify real illuminants, semantic segmentation results $\mathbb{P}$ need to be compared with the light source mask $\mathbb{Q}$ obtained using our detection algorithm in Section 4. If a classified light source in $\mathbb{P}$ has a large intersection with $\mathbb{Q}$, the type of this light source can be determined as the semantic segmentation result. The ADE20K dataset contains images of most objects in the world. However, the light source that has a small area or complex geometry may be incorrectly classified as a non-luminous object due to the scene complexity in practice. To avoid missing any true light sources detected in Section 4, our method defines unclassified objects highlighted in $\mathbb{Q}$ as diffuse illuminants.

6 Calculation of the Illumination Intensity

After light source detection and classification, illumination intensities are accurately calculated based on the locations and types of light sources. As discussed in Section 3.2, our illumination intensity estimation method consists of two steps, i.e., the inverse CRF estimation followed by extending the illumination intensity range.

6.1 Inverse CRF Estimation

The real-world radiance normally has a linear distribution [43]. According to the non-linear $\mathcal{F}$ in Section 3.2, pixel distributions in the image I are also non-linear. The inverse



(a) *Initial Result* (b) *Optimized Result*

Fig. 5. Optimization of $\mathcal{G}$ in LDR images. Adding the gradient constraint factor to the initial voting method [43] ensures that the final inverse CRF curve monotonically increases.

CRF (denoted by $\mathcal{G} = \mathcal{F}^{-1}$) aims to convert the non-linear I to its linear version $I_{\mathcal{Q}}$. Each I has the unique $\mathcal{G}$, but all inverse CRFs have the same characteristics [39]. First, each $\mathcal{G}$ should monotonically increase. Second, minimum and maximum input values in the image I should be mapped to minimum and maximum output values respectively in the image $I_{\mathcal{Q}}$: $\mathcal{G}(0) = 0$ and $\mathcal{G}(1) = 1$.

S patches are first chosen from I. In the yth patch ($y \in [1, S]$), pixels in a specific row or column are scanned to provide V reference inverse CRFs: $\mathbb{G}^y = \{\mathbf{g}_1, \ldots, \mathbf{g}_V\}$. Then, a voting method $\mathcal{H}$ developed by Sharma et al. [43] is used to fit the optimized inverse CRF of the yth patch: $G^y = \mathcal{H}(\mathbb{G}^y)$. Furthermore, $\mathcal{H}$ is applied to all patches to compute the final inverse CRF of I: $\mathcal{G} = \mathcal{H}(\{G^1, \ldots, G^S\})$.

As shown in Fig. 5, some reference inverse CRF curves can be discretized by creating $B \times B$ mesh grids. Subsequently, the number of curves intersecting with each grid cell is counted. For each column in this mesh, the cell that has the highest count is used to determine the final inverse CRF. If adjacent cells in the horizontal direction contain the same number of curves, note that $\mathcal{H}$ violates the requirement that the inverse CRF curve should monotonically increase [45]. In the yth patch, the derivative of the inverse CRF denoted by $K^y(\varrho) = \partial G^y / \partial \varrho$ should be positive. ϱ is a variable evenly sampled within [0,1]. If $K^y(\varrho) \leq 0$, the new derivative of the yth inverse CRF at ϱ can be defined as:

$$\hat{K}^y(\varrho) = \|K^y(\varrho)\| + d \cdot \varrho, \tag{9}$$

where $d = 1/B$ indicates the side length of each grid cell. The gradient constraint $\hat{K}^y$ ensures the monotonicity of $\mathcal{G}$. After the estimation of $\mathcal{G}$, $I_{\mathcal{Q}}$ still has a low dynamic range. If $I_{\mathcal{Q}}$ is used as a radiance map in the image-based lighting, rendering results cannot be convincingly realistic.

6.2 Extending the Dynamic Range of Illumination Intensities

To accurately extend and predict E as the HDR image, results of the previous light source positioning (Section 4) and classification (Section 5) steps are used to further optimize the calculation of lighting intensities. Each light source contributes to illuminating non-luminous objects in the scene. The contribution of the ith light source can be estimated as:

$$\xi_i = \frac{E_{env} \cdot E_i}{\|E_{env}\| \cdot \|E_i\|}, \tag{10}$$

where E_{env} indicates the mean of the ambient lighting and E_i indicates the mean of the ith light source pixel values in the grayscale space of $I_{\mathcal{Q}}$. Specifically, E_{env} represents the illumination component of the LDR image excluding all light source areas $\mathbb{O}$.

Given a point at $p(u, v)$ in $I_{\mathcal{Q}}$, its pixel value in the grayscale space X_p is an essential benchmark. To calculate E_{env}, the LDR image can be decomposed into an illumination (or shading) component $U_{\mathcal{Q}}$ and a reflectance (or albedo) component $R_{\mathcal{Q}}$ [50]. $R_{\mathcal{Q}}$ greatly disturbs the estimation accuracy of E_{env}. To solve this problem, the intrinsic image decomposition method based on the retinex algorithm [50] is employed to extract $U_{\mathcal{Q}}$ of $I_{\mathcal{Q}}$. E_{env} in Eq. (10) can be optimized as:

$$E_{env} = \begin{cases} \left(\sum_{u=1}^{W}\sum_{v=1}^{H} U_{\mathcal{Q}}(p)\right)/Z, & if\ p \in \mathbb{O} \\ 0, & elif\ p \notin \mathbb{O} \end{cases}, \tag{11}$$

where W and H represent the width and height of a panoramic image respectively. Z indicates the total number of pixels in $I_{\mathcal{Q}}$ excluding light source areas.

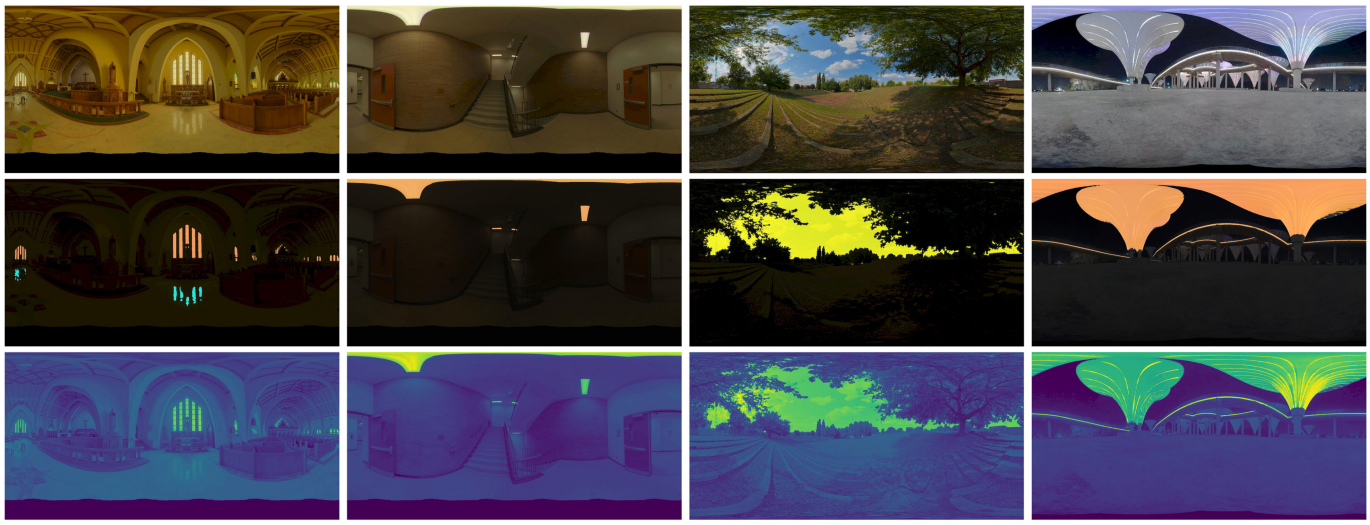
Directional and diffuse illuminants $\mathbb{X}$ usually have greater contributions to the scene illumination than reflective objects $\mathbb{Y}$. Additionally, the illumination contribution of the ith light source to its surrounding environment ξ_i can be used as an important reference factor to estimate correct illumination intensities. The extended illumination intensity $E(X_p)$ can be calculated as:

$$\begin{aligned} E(X_p) &= \begin{cases} \xi_i \cdot \mathbf{F}(X_p), & if\ p \in \mathbb{X} \\ \mu \cdot \xi_i \cdot \mathbf{F}(X_p), & elif\ p \in \mathbb{Y} \\ \hat{X}_p, & else \end{cases}, \\ \mathbf{F}(X_p) &= \hat{X}_p \cdot max\left(1,\ Q/(1 + e^{-\gamma(X_p - E_{env})})\right) \end{aligned} \tag{12}$$

where μ is a factor to suppress intensities of reflective objects. Q indicates the preset maximum intensity of the predicted HDR image. The design of $\mathbf{F}(\cdot)$ is inspired by tone-mapping algorithms [35], [38] and the nature of the sigmoid function. The sigmoid curve mimics the photoreceptor response of cameras and can preserve contrast in mid-tones. $\hat{X}_p$ denotes the normalized pixel value. The factor γ is used to control the gradient of adjacent predicted HDR values mapped from adjacent gray values.

For a panoramic video clip, subtle temporal differences between adjacent frames may cause illumination intensities to vary with discontinuities. This inevitably leads to flickering effects when the panoramic video clip is used as a dynamic background for lighting in realistic rendering. To obtain smooth illumination intensities estimated for all frames of a panoramic video clip, the temporal smoothing function is applied on the tth frame. It is defined as the following for the calculation of ξ_i and γ:

$$\{\xi_i^t, \gamma^t\} = \arg\min_{\xi_i, \gamma} \left(\sum_{X_p=0}^{255} (E^t(X_p) - E^{t-1}(X_p)) \right), \tag{13}$$



(a) Bright Indoor (b) Dark Indoor (c) Bright Outdoor (d) Dark Outdoor

Fig. 6. Results of our illumination estimation method. Images of the middle row show diffuse illuminants, directional illuminants and reflective objects marked in orange, yellow and blue respectively. The illumination intensity map in the third row is color-coded from yellow (high intensity) to blue (low intensity).

Eq. (13) can be solved based on the Levenberg-Marquardt (LM) algorithm [51]. After illumination intensities have been calculated on the panoramic image, their values are then mapped to an HDR radiance map and used for realistically rendering virtual models.

7 Evaluations

This section documents our comprehensive evaluations for assessing the effectiveness and applicability of our method. More results can be found in the supplementary material, available online.

7.1 Experiments

To facilitate interactive realistic rendering applications, our method for estimating illuminations using LDR images concerns not only the complete and accurate analysis of light sources but also the high computational efficiency.

Implementation Details. Our experiments were performed using a single NVIDIA GeForce GTX1060 GPU with 3GB DDR3. Algorithms were programmed in C++ and using CUDA ver. 10.0, with a UI developed using Qt. Various panoramic LDR images were tone-mapped from the Laval HDR dataset [10] or captured using the Insta360 ONE X camera for verifications. To balance the computational efficiency with the lighting estimation quality, the resolution of the panoramic image was adjusted to 2000×1000. In our experiment of applying the proposed CNN model, L was set to 473. The Adam optimizer [52] was used with the momentum parameter of 0.9 and the batch size of 32. The initial learning rate was set as 0.001 and decayed exponentially by a factor of 0.9 in every 30 steps.

Existing Methods for Comparison. In our qualitative and quantitative evaluations, results of our method are compared with several existing state-of-the-art methods, including algorithms of Eilertsen et al. [38], Liu et al. [39] and Santos et al. [40]. All approaches were executed at the same viewpoint and along the same viewing direction in our experiment. In addition, we took the illumination computed using the HDR image captured by the camera as the ground truth for our evaluation.

Physically-Based Rendering. For the large-scale and automated testing, we developed a physically-based path tracer. Virtual models were placed in the center of virtual scenes. The estimated HDR panorama was used as a radiance environment map to provide omnidirectional illuminations. To improve the computational efficiency, the multiple importance sampling (MIS) method was used to combine the illumination importance sampling and the BSDF importance sampling [53]. Specifically, the piecewise-constant probability distribution function [3] was leveraged to calculate the importance distribution of illuminations.

7.2 Qualitative Evaluation

To visually evaluate the proposed method, its lighting estimation quality is analyzed and compared with existing methods, in terms of the correctness of the light source parameter calculation and the realistic rendering effect.

7.2.1 *Visual Assessment of Illumination Estimation Results*

Geometric structures and illumination conditions of real-world scenes are generally diverse and complex. To verify the adaptability of our method, several indoor and outdoor circumstances under bright or darker illumination conditions were employed in our testing (Fig. 6). More results are given in the supplementary material of our paper, available online.

Our method adopts a non-learning method to calculate light source positions in LDR images. It ensures that boundaries of light sources are precisely calculated (e.g., the bright sky area clipped by trees shown in Fig. 6c). To improve the computational efficiency, only light sources that provide major contributions to illuminate the surrounding area of the viewpoint are identified. Illuminants or reflective objects that emit weak lights were classified as non-luminous objects. Since the estimated illumination intensity has a

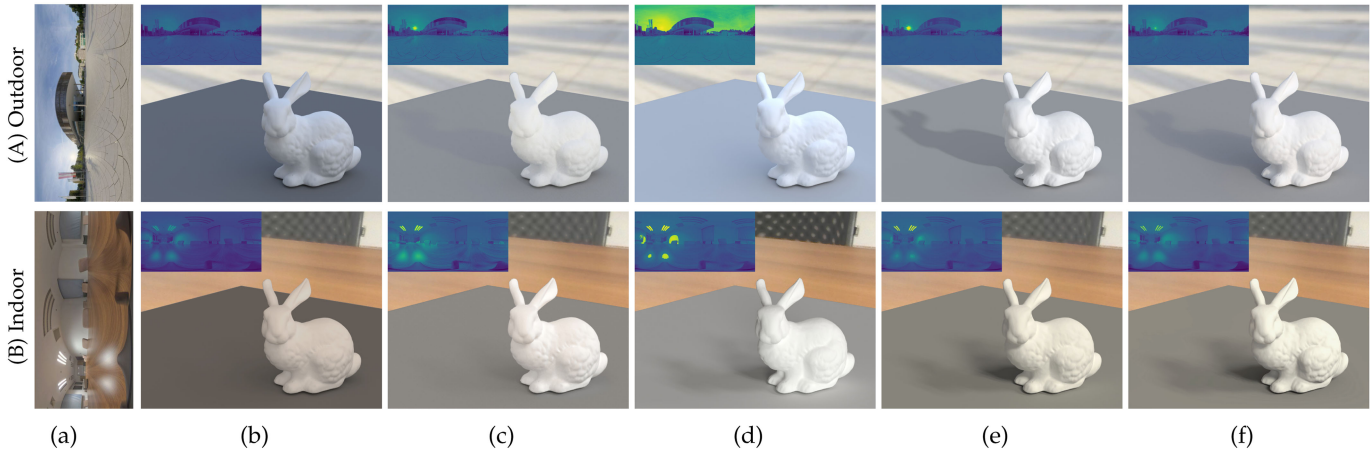


Fig. 7. Ablation study of our illumination estimation method. (a) Input LDR panoramas. (b) Lighting effects directly using the LDR panorama as the radiance map. (c) Lighting effects using intensities estimated without detecting light sources. (d) Lighting effects using intensities estimated without classifying light sources. (e) Lighting effects using intensities estimated by the complete proposed method. (f) Ground truth lighting using the HDR panorama as the radiance map.

high dynamic range, reconstructed illumination intensity maps are converted to color-coded maps to clearly display all details. Benefit from light source classification, our method is able to handle various complex real-world scenes and calculate satisfactory HDR illuminations.

7.2.2 Ablation Study of Illumination Estimation

If the LDR panoramic image is directly used as the environment radiance map, it just has a low dynamic range of pixel values. Rendered surfaces of virtual models cannot be brightly illuminated in this case (Fig. 7b). To verify the benefit and necessity of each step of our method, an ablation study was conducted using LDR panoramic images of indoor and outdoor scenes in our experiment. Since the illumination intensity calculation relies on the outputs of previous steps in the pipeline of our method (Fig. 1), we mainly focus on evaluating the contributions by the light source detection and classification steps.

When only the light source classification is applied without performing the light source detection before the calculation of illumination intensities, some objects (e.g., the floor and table) may be incorrectly identified as reflective illuminants without the location information of true light sources. Such objects are normally large and can significantly deviate estimated illumination intensities and then affect the rendering of virtual models. In such a case, the surface of the virtual bunny is uniformly illuminated and cannot represent realistic shading and shadowing effects, as shown in Fig. 7c.

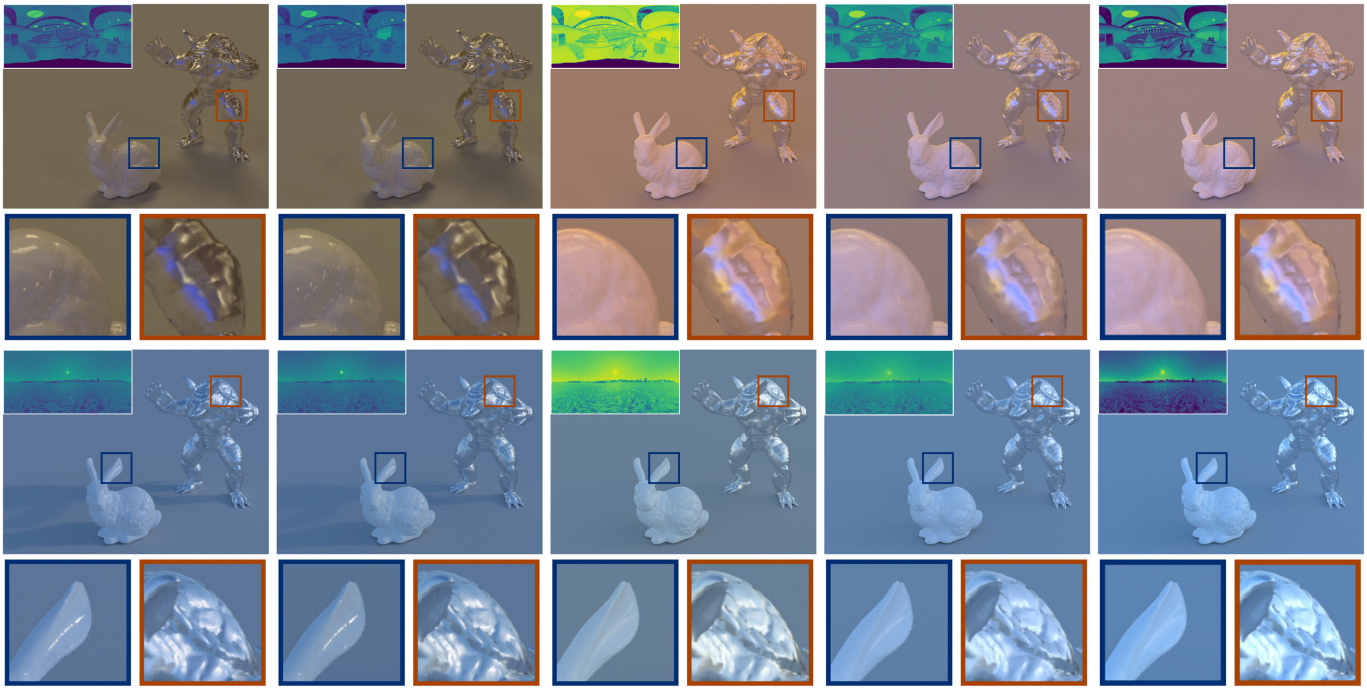
When light sources are detected but not classified before the calculation of illumination intensities, our method cannot compatibly handle various scenes without the type information of illuminants. For example, in the sunny outdoor, identifying the sun as directional illuminant generates sharp shadows around virtual models. Otherwise, virtual models seem to be displayed in the overcast outdoor. In contrast, illumination intensities of reflective objects are normally low. Without correctly recognizing reflective objects, illumination intensities of reflected lights on the table and wall will be almost the same as ceiling lamps. This results in over illuminated and unrealistic rendering effects on the surface of a virtual model under indoor circumstances, as shown in Fig. 7d.

A complete integration of all operations in our illumination estimation pipeline yields the desired realistic rendering result in our ablation study, as shown in Fig. 7e. Our method composes accurate and comprehensive illumination parameters estimated from the LDR panoramic image. The shading and shadowing effects rendered based on our illumination estimation look almost identical to the ground truth rendered using the HDR panoramic image.

7.2.3 Visual Comparison of Rendering Quality

To intuitively compare realistic rendering results of our method with those of other approaches, two classical models of specular material were rendered after sampling 512 times per pixel. The captured HDR radiance map is used as the ground truth illumination (Fig. 8a). By comparing all panoramic radiance maps, it can be found that the ambient lighting estimated using other methods is brighter than the ground-truth lighting. Their rendering results are greatly affected by the ambient lighting (i.e., E_{env} in Eq. (10)) of the LDR image. Besides, existing methods [38], [39], [40] paid more attention to recovering details in saturated local regions in the LDR image rather than calculating accurate light source intensities. Therefore, virtual models cannot cast realistic shadows on the ground. When lighting parameters are comprehensively calculated using our method, these drawbacks can be better solved as shown in Fig. 8b.

The above comparison is implemented in a static scene. For real-time applications with dynamically varying scene environments, the rendering result with smooth lighting is desired. To compare temporal consistencies of illuminations estimated by different methods, consecutive frames in a video clip were used as environment radiance maps to render a virtual bunny model of the metallic material (Fig. 9). With benefit from the proposed temporal smoothness function in Eq. (13), lighting parameter differences between adjacent video frames are reduced to a minimum. It helps avoid flickering effects of illuminations in dynamic scenes.



(a) Ground truth (b) Proposed method (c) Eilertsen et al. [38] (d) Liu et al. [39] (e) Santos et al. [40]

Fig. 8. Visual comparison of realistic rendering results between our method and existing state-of-the-art methods for indoor and outdoor scenes. HDR radiance maps predicted using different methods are shown in attached panoramic images.

7.2.4 *Rendering Effects of Virtual versus Real Objects*

Intensities of the HDR radiance map normally are not exactly consistent under real-world illuminations. In our experiment, the rendering effect of the virtual model is also compared with its physical version. In Fig. 10, three classical 3D models were placed and photographed in various real-world circumstances. The panoramic camera was placed where the virtual model was rendered.

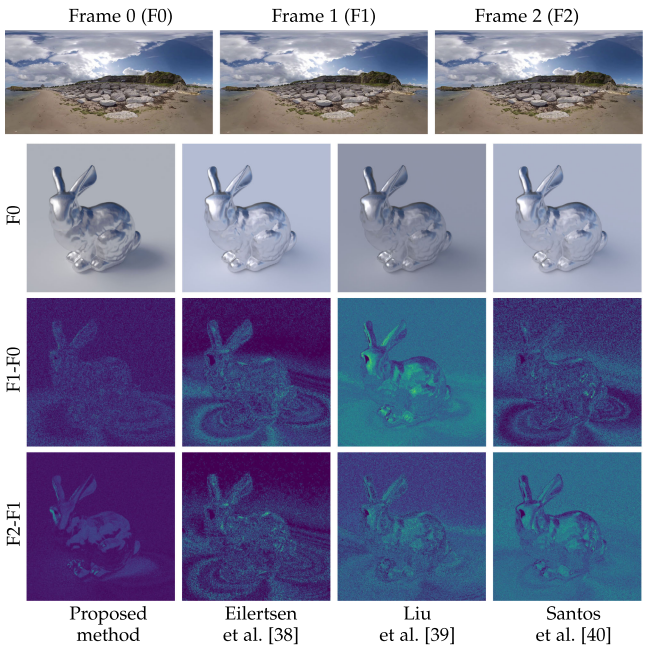


Fig. 9. The comparison between rendering results of three consecutive video frames. Two lower rows represent difference maps between adjacent frames.

We manually adjusted the pose and position of virtual models to closely match with photographs of real 3D printed models in the same scene. Besides, we referred to the 3D printed model to set the optimal material and color of virtual models as

(a) Real objects (b) Proposed method

Fig. 10. Comparing displays of virtual models rendered using different methods and photographs of 3D printed objects captured under different circumstances.

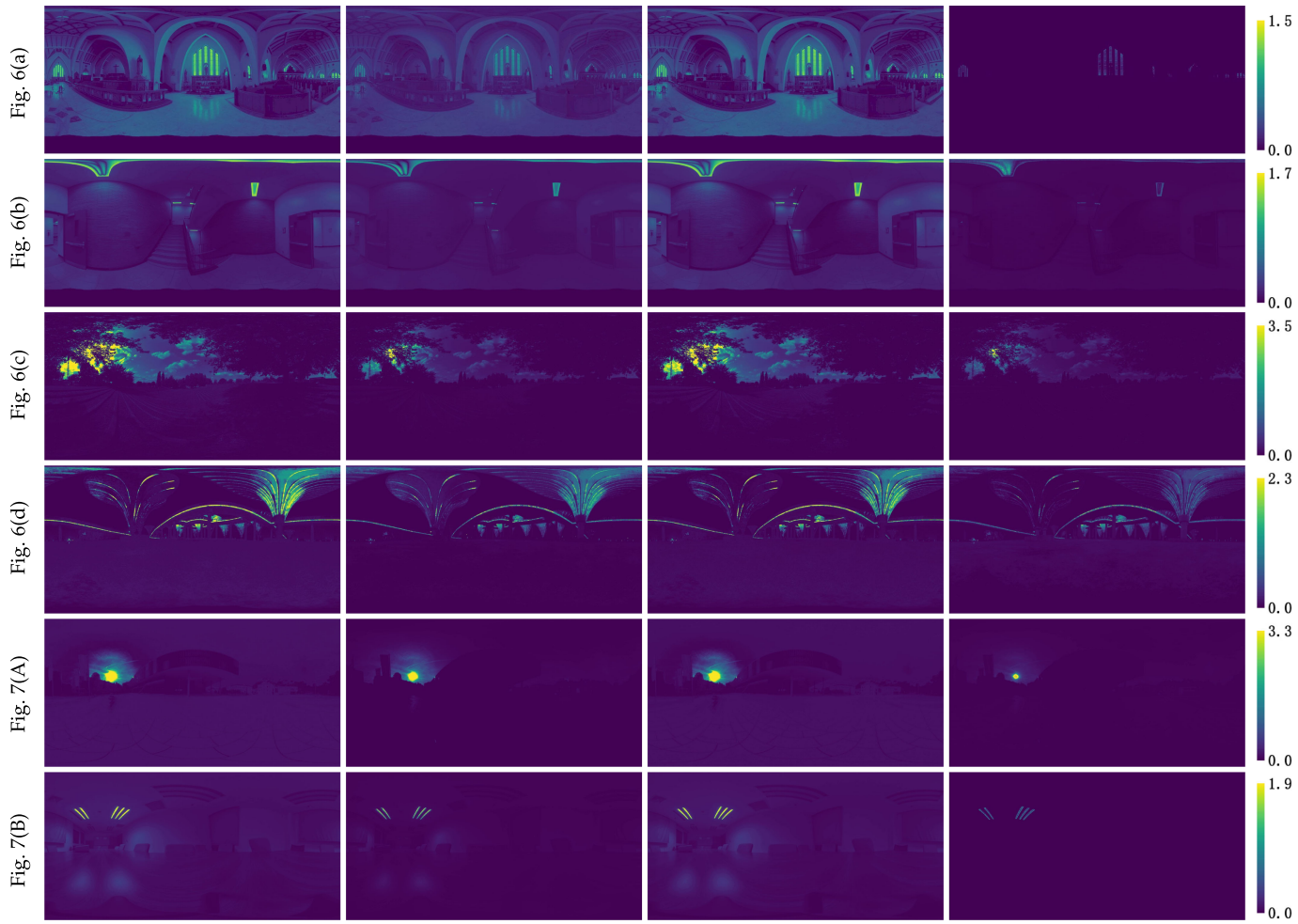


Fig. 11. Quantitative comparison of the lighting estimation accuracy in several scenes presented in Section 7.2. The color-coded map indicates the difference map between the radiance map predicted using different methods and the ground-truth HDR radiance map.

much as possible for rendering. In our comparison, although the physically real and virtual models look slightly different, rendering results demonstrate that synthetic rendering effects generated using our method can almost mix the spurious with the genuine. Our rendering results help people perceive the synthetic entities more immersively using the realistic rendering technique. More comparison results between our method and existing state-of-the-art methods are given in the supplementary material of our paper, available online.

7.3 Quantitative Evaluation

To quantitatively evaluate our method, four experiments were conducted to assess rendering quality and efficiency.

7.3.1 Assessment of the Lighting Estimation Accuracy

To intuitively evaluate illumination estimation errors, the difference map $\mathcal{D}$ between the ground truth E_{gt} and the illumination E_{pred} predicted using different methods are shown in Fig. 11. To clearly display error maps, the error range is compressed and the error at p in an image is computed as:

$$\mathcal{D}(p) = log_{10}\|E_{pred}(p) - E_{gt}(p)\|. \quad (14)$$

Fig. 11 demonstrates that higher estimated errors are concentrated near the light source. Since light sources generally have high-intensity illuminations, the light source illumination is hard to estimate accurately, especially in outdoor scenes. Benefit from the function $\mathbf{F}(\cdot)$ in Eq. (12), the higher gray value in an LDR image corresponds to the higher HDR value. Therefore, the lighting intensity around the light source can be optimally extended. As shown in Fig. 11d, our method can produce much fewer errors than other methods.

Three different metrics, i.e., RMSE, PSNR and SSIM, were further employed to quantify illumination estimation errors against the ground truth in our evaluation as shown in Table 1. These metrics were calculated on the difference maps between the ground truth HDR illumination and predicted illuminations using different methods (Fig. 11). As proven by quantitative evaluation results, the proposed method outperforms other methods in all testing scenes. The scores of Santos et al. [40] are generally close to or slightly higher than Eilertsen et al. [38]. Liu et al. [39] can generate realistic contrast and saturation in estimated HDR images and obtain relatively good scores, but the illumination intensity calculation of their method still needs to be improved.

TABLE 1
The Quantitative Comparison of our Approach and Other Existing Methods

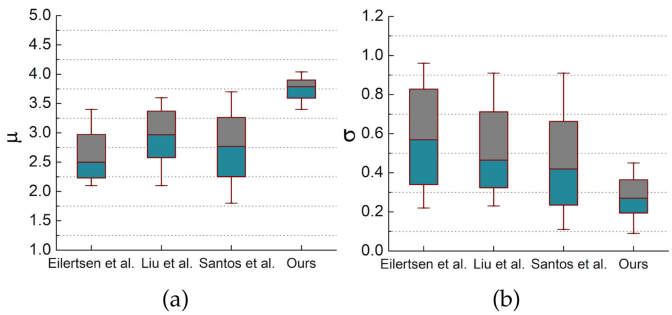
Method		Fig. 6a	Fig. 6b	Fig. 6c	Fig. 6d	Fig. 7a	Fig. 7b
Eilertsen et al. [38]	RMSE↓	0.18	0.30	0.21	0.25	0.32	0.24
	PSNR↑	16.64	18.57	10.94	11.95	11.92	11.73
	SSIM↑	0.45	0.47	0.67	0.65	0.57	0.42
Liu et al. [39]	RMSE↓	0.16	0.25	0.20	0.19	0.22	0.21
	PSNR↑	26.55	25.76	27.08	19.47	19.47	26.01
	SSIM↑	0.64	0.66	0.75	0.75	0.69	0.76
Santos et al. [40]	RMSE↓	0.17	0.30	0.20	0.21	0.29	0.23
	PSNR↑	16.18	22.04	10.02	11.52	13.17	10.31
	SSIM↑	0.46	0.48	0.69	0.66	0.61	0.46
Proposed Method	RMSE↓	**0.13**	**0.19**	**0.16**	**0.18**	**0.20**	**0.12**
	PSNR↑	**34.76**	**32.46**	**38.29**	**29.88**	**37.31**	**33.28**
	SSIM↑	**0.86**	**0.84**	**0.81**	**0.82**	**0.80**	**0.81**

(↑ indicates higher values are better and ↓ indicates lower values are better.)

7.3.2 User Study

How humans visually perceive illumination estimation results and realistic rendering effects somehow may not be truly measured by a quantitative assessment using computers. Therefore, a user study of subjective assessment was conducted to evaluate the quality of illumination estimation results. 30 participants aging from 20 to 50 years old were invited to this user study. Besides, participants were engaged in technical or non-technical occupations, with or without lighting estimation experiences.

In our comparison, we selected six real-world scenes and four algorithms as outlined in Table 1. First, all resulting images of realistic rendering were randomly ordered for each participant to avoid cueing him or her. After watching each rendering result for 30 seconds, the participant was then asked to give a score from 1 to 5 (from the worst to the best). Every participant repeated this experiment three times. In addition, we designed several evaluation criteria, e.g., the position accuracy of shadings and shadows, to control the individual bias of each participant. More details are given in the supplementary material of our paper, available online. All tested algorithms were able to process selected real-world scenes and all participants evaluated all rendering results. In such a way, the conclusion of this user study can be mainly affected by the subjective opinions of these participants. Fig. 12 shows statistics of the mean value μ and standard deviation σ of scores given by the 30 participants. The method with a higher μ and a lower σ is capable of handling various real-world scenes robustly and producing stable lighting estimation results. The proposed method obtained the average score of $\mu = 3.75$ and $\sigma = 0.28$ and outperformed other methods. This user study demonstrates that our method is favored by most involved participants.



(a) (b)

Fig. 12. User study. (a) Mean value of participants' scores. (b) Standard deviation of participants' scores.

7.3.3 Ablation Study of CNN-based Light Source Classification

As described in Section 5.2, the proposed CNN architecture for light source classification mainly involves the GhostNet backbone, global spatial extraction module and channel attention module. To evaluate the effectiveness of each component, we conducted repeated experiments with several settings as listed in Table 2. For such an ablation study, the adjusted ADE20K dataset was divided into 18K/2K images for training and validation respectively. Two different metrics, i.e., *pixel-wise accuracy* (Pixel Acc.) and *mean of class-wise intersection over union* (Mean IoU), were employed for the quantitative analysis in this ablation study.

Table 2 shows that the global feature extraction and channel attention modules can helpfully support the GhostNet backbone to improve the light source classification accuracy. In our experiment, adding these two modules to the network yields a higher performance up to $88.34\%/71.14\%$ in terms of Pixel Acc. and Mean IoU. Since the panorama dataset with light source annotations currently does not publicly exist, we manually annotated 300 panoramas captured by ourselves and collected from public panorama datasets. To minimize the discrepancy between the perspective training data and

TABLE 2
Ablation Study of our CNN-Based Light Source Classification

Method	Mean IoU (%)	Pixel Acc. (%)
GhostNet (Backbone)	63.91 ± 0.22	85.76 ± 0.19
GhostNet + GF	65.30 ± 0.18	86.31 ± 0.11
GhostNet + CA	67.87 ± 0.19	87.40 ± 0.12
GhostNet + GF + CA	$\mathbf{71.01 \pm 0.13}$	$\mathbf{88.19 \pm 0.15}$

('GF' and 'CA' denote global feature extraction module and channel attention module respectively.)

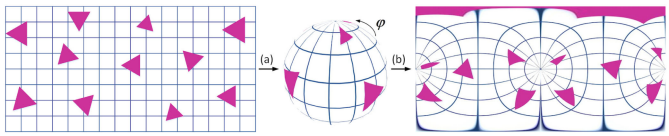


Fig. 13. Panorama warping: (a) The original panorama is first mapped onto a unit sphere and then rotated to a certain pitch angle (e.g., $\varphi = 90°$). (b) The spherical image is projected to a new panorama using the spherical projection.

the panoramic testing data, all panoramic images were converted to skybox images (i.e., a total of 1.8K images) for external testing. This testing scheme yields $87.21\%/68.91\%$ in terms of Pixel Acc. and Mean IoU. Due to the difference between training and testing data, testing scores slightly decrease. Nevertheless, final results are still acceptable to satisfy our requirement.

7.3.4 Analysis of the Computational Efficiency

Fast lighting estimation not only facilitates the user interaction, but also makes it possible to implement a dynamic illumination effect with light sources varying over time.

The panoramic image in existing panorama datasets was generally photographed with a camera vertically above the ground. However, for most head-mounted or hand-held MR devices, their cameras can capture images in arbitrary poses due to hand or head rotation. According to the principle of equirectangular image, different viewing directions of camera lenses lead to different geometric structures of panoramas. In such a case, the object near the upper or lower border of a panorama may be divided into multiple parts (Fig. 13). To verify that our method can efficiently handle varying numbers of light sources with different geometric structure complexities, the panoramic image can be gradually distorted into different structures. We encourage readers to view this process in our supplementary video. With GPU-accelerated implementation, the proposed illumination estimation method can achieve an average of 39 FPS. The measured elapsed time included operations of light source detection (average 126 FPS), light source classification (average 87 FPS) and illumination intensity calculation (average 102 FPS).

8 Limitations and Future Work

Spatially-Varying Illumination. The proposed method focuses on estimating illuminations near the panoramic camera. In MR applications, virtual objects are often arbitrarily placed, not just near the camera. The lighting is generally spatially-varying especially in indoor scenes: light sources at different positions in the scene create different lighting conditions due to occlusions and non-uniform light distributions. This is a potential material for further improvement of our method. We will research geometry reconstruction combined with the proposed work in this paper to estimate spatially-varying illuminations using a single panoramic image.

Saturated Region Reconstruction. Due to camera exposure settings and low dynamic display, texture details in some bright regions are inevitably lost (Fig. 14a). The proposed method focuses on predicting HDR illuminations based on LDR pixel values rather than recovering missing details in over-exposed regions. Illumination intensities in over-exposed regions estimated by our method become undesirably consistent (Fig. 14b). But the missing details may have relatively lower intensities in the ground-truth lighting (Fig. 14c). To resolve this issue, applying the algorithm of saturated region reconstruction can further improve the accuracy of our illumination estimation.

(a) Input image (b) Ground truth (c) Proposed method

Fig. 14. Limitations of our method: missing texture details in over-exposed regions cause inaccurate lighting estimation.

9 Conclusion

To provide efficient and high-quality illumination estimation for interactive realistic rendering, a novel lighting estimation method using LDR panoramic images is proposed in this paper. Our approach consists of three steps to calculate the key properties of light sources. First, an efficient light source detection method is applied to facilitate the accurate positioning of light sources. It extracts illuminant characteristics during the image exposure attenuation procedure and can precisely differentiate true light sources from non-luminous objects in LDR images. Second, a semantic segmentation model is designed to accurately identify types of detected light sources. The type information of light sources can help precisely recover HDR illumination intensities from LDR images. Third, a new method based on inverse camera imaging mechanism is developed to estimate illumination intensities in LDR images. It comprehensively takes the location and type information of light sources into account. Compared with existing methods, our method offers better lighting estimation quality and computational efficiency against environmental complexities.

References

[1] J. S. Nimeroff, E. Simoncelli, and J. Dorsey, "Efficient re-rendering of naturally illuminated environments," *Photorealistic Rendering Techniques*. Berlin, Germany: Springer, 1995, pp. 373–388.
[2] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. 25th Annu. Conf. Comput. Graph. Interactive Techn.*, 1998, pp. 189–198.
[3] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[4] T. Bertel, M. Yuan, R. Lindroos, and C. Richardt, "OmniPhotos: Casual 360° VR photography with motion parallax," in *Proc. SIGGRAPH Asia Emerg. Technol.*, 2020, pp. 1–2.

[5] M. Knecht, C. Traxler, O. Mattausch, W. Purgathofer, and M. Wimmer, "Differential instant radiosity for mixed reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2010, pp. 99–107.

[6] H. Barrow, J. Tenenbaum, A. Hanson, and E. Riseman, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst.*, vol. 2, pp. 3–26, 1978.

[7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annu. Conf. Comput. Graph. Interactive Techn.*, 2008, pp. 1–10.

[8] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. 1–8.

[9] Y. Matsushita and S. Lin, "Radiometric calibration from noise distributions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[10] M.-A. Gardner et al., "Learning to predict indoor illumination from a single image," *ACM Trans. Graph.*, vol. 36, no. 6, Nov. 2017, Art. no. 176.

[11] Y. Hold-Geoffroy, A. Athawale, and J. Lalonde, "Deep sky modeling for single image outdoor lighting estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6920–6928.

[12] P. P. Srinivasan, B. Mildenhall, M. Tancik, J. T. Barron, R. Tucker, and N. Snavely, "Lighthouse: Predicting lighting volumes for spatially-coherent illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8077–8086.

[13] S. Gibson, T. Howard, and R. Hubbold, "Flexible image-based photometric reconstruction using virtual light sources," *Comput. Graph. Forum*, vol. 20, no. 3, pp. 203–214, Jul. 2001.

[14] D. A. Calian, J.-F. Lalonde, P. Gotardo, T. Simon, I. Matthews, and K. Mitchell, "From faces to outdoor light probes," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 51–61, May 2018.

[15] B. Jung and M. Inanici, "Measuring circadian lighting through high dynamic range photography," *Lighting Res. Technol.*, vol. 51, no. 5, pp. 742–763, 2019.

[16] C. Legendre et al., "DeepLight: Learning illumination for unconstrained mobile mixed reality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5911–5921.

[17] J. Park, H. Park, S. E. Yoon, and W. Woo, "Physically-inspired deep light estimation from a homogeneous-material object for mixed reality lighting," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 2002–2011, May 2020.

[18] R. Yi, C. Zhu, P. Tan, and S. Lin, "Faces as lighting probes via unsupervised deep highlight extraction," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 317–333.

[19] H. Weber, D. Prévost, and J. Lalonde, "Learning to estimate indoor lighting from 3D objects," in *Proc. IEEE Int. Conf. 3D Vis.*, 2018, pp. 199–207.

[20] J. Lopez-Moreno, E. Garces, S. Hadap, E. Reinhard, and D. Gutierrez, "Multiple light source estimation in a single image," *Comput. Graph. Forum*, vol. 32, no. 8, pp. 170–182, Aug. 2013.

[21] T. Rhee, L. Petikam, B. Allen, and A. Chalmers, "MR360: Mixed reality rendering for 360° panoramic videos," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 4, pp. 1379–1388, Apr. 2017.

[22] A. Morgand, M. Tamaazousti, and A. Bartoli, "A multiple-view geometric model of specularities on non-planar shapes with application to dynamic retexturing," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 11, pp. 2485–2493, Nov. 2017.

[23] G. Fu, Q. Zhang, Q. Lin, L. Zhu, and C. Xiao, "Learning to detect specular highlights from real-world images," in *Proc. ACM Int. Conf. Multimedia*, 2020, pp. 1873–1881.

[24] K. Karsch, V. Hedau, D. Forsyth, and D. Hoiem, "Rendering synthetic objects into legacy photographs," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–12, Dec. 2011.

[25] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "A multiple-view geometric model of specularities on non-planar shapes with application to dynamic retexturing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, May 2012.

[26] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba, "Recognizing scene viewpoint using panoramic place representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2695–2702.

[27] H. Weber, D. Prévost, and J.-F. Lalonde, "Learning to estimate indoor lighting from 3D objects," in *Proc. Int. Conf. 3D Vis.*, 2018, pp. 199–207.

[28] J. Zhang, K. Sunkavalli, Y. Hold-Geoffroy, S. Hadap, J. Eisenman, and J. Lalonde, "All-weather deep outdoor lighting estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 150–10 158.

[29] J. Lalonde and I. Matthews, "Lighting estimation in outdoor image collections," in *Proc. Int. Conf. 3D Vis.*, 2014, pp. 131–138.

[30] F. Zhan et al., "EMLight: Lighting estimation via spherical distribution approximation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 3287–3295.

[31] F. Zhan et al., "GMLight: Lighting estimation via geometric distribution approximation," *IEEE Trans. Image Process.*, vol. 31, no. 3, pp. 2268–2278, Mar. 2022.

[32] Z. Li, M. Shafiei, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker, "Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2475–2484.

[33] K. Karsch et al., "Automatic scene inference for 3D object compositing," *ACM Trans. Graph.*, vol. 33, no. 3, pp. 1–15, 2014.

[34] A. G. Rempel et al., "Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 39–es, 2007.

[35] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," in *Proc. ACM SIGGRAPH*, 2008, pp. 1–10.

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

[37] 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.*, 2018, pp. 596–611.

[38] 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, 2017.

[39] Y.-L. Liu et al., "Single-image HDR reconstruction by learning to reverse the camera pipeline," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1651–1660.

[40] M. S. Santos, T. I. Ren, and N. K. Kalantari, "Single image HDR reconstruction using a CNN with masked features and perceptual loss," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 80:1–80:10, 2020.

[41] J. T. Kajiya, "The rendering equation," *ACM SIGGRAPH Comput. Graph.*, vol. 20, pp. 143–150, 1986.

[42] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Sep. 2017.

[43] A. Sharma, R. T. Tan, and L.-F. Cheong, "Single-image camera response function using prediction consistency and gradual refinement," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 19–35.

[44] D. Salomon, *The Computer Graphics Manual*. Berlin, Germany: Springer, 2011.

[45] M. D. Grossberg and S. K. Nayar, "Modeling the space of camera response functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1272–1282, Oct. 2004.

[46] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and applications*. Singapore: World Scientific, 1998.

[47] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20K dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 633–641.

[48] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1577–1586.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Conf. Comput. Vis.*, 2017, pp. 2980–2988.

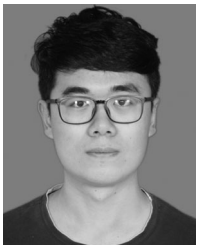
[50] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[51] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, no. 2, pp. 164–168, 1944.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[53] L. J. Guibas and E. Veach, "Robust Monte Carlo methods for light transport simulation," Ph.D. dissertation, 1997.

Haojie Cheng received the BSc degree from Anhui University of Science and Technology, Huainan, China, in 2018. He is currently working toward the PhD degree with the University of Science and Technology of China, Hefei, China. He is currently performing his research work in the Healthcare Information Technology Innovation Center, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China. His research interests include realistic scene rendering, interactive 3D medical visualization based on GPU, and the preoperative surgical simulation and navigation using VR and AR.



Chunxiao Xu received the BSc degree from Shandong University, in 2019. He is currently working toward PhD degree with the University of Science and Technology of China, Heifei, China. Since July 2019, he has been conducting his scientific research work on GPU based 3D volume rendering at Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China. He is currently working on 3D volume rendering with enhanced illumination model for real-time VR and AR applications.

Jiajun Wang received the MSc degree from Southeast University, Nanjing, China, in 2019. During this period, he conducted his research about acoustic emission in the Research Center of Condition Monitoring and Fault Diagnosis. Then he began his career in Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China. He is currently focusing on the research of 3D medical visualization and artificial intelligence for clinical applications. He is also performing his work as a software engineer.

Zhenxin Chen received the BEng degree from the Harbin Institute of Technology, Harbin, China, in 2016, and the MEng degree from Shandong University, Jinan, China, in 2019. He is currently performing his research work in the Healthcare Information Technology Innovation Center, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China. His research interests include interactive 3D medical visualization, biological signal, and medical image processing.

Lingxiao Zhao received the MSc and PhD degrees, in 2004 and 2011, respectively. He worked as a MSc student and then a PhD researcher in the computer graphics and visualization group of the Delft University of Technology (TU Delft), the Netherlands, from 2002 to 2008. During this period, he conducted his research work as a part of the advanced development of CT imaging and visualization for computer-aided diagnosis in virtual colonography inside of Philips Healthcare, Best, the Netherlands. In 2009, he began his career in commercial healthcare software development. In 2011, he joined the Imaging Dept. of Philips Healthcare and continued to work for the R&D of medical visualization algorithms and software products. In 2014, he moved to Philips Research, HTC, Eindhoven, the Netherlands, and contributed for two incubating projects of Philips Healthcare. Since September 2015, he started his new position in Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China. He is currently a professor and the director of the Healthcare Information Technology Innovation Center in SIBET. He is also a guest professor in the Biomedical Engineering Dept. of University of Science and Technology of China, Hefei, China. His current research mainly focuses on image processing, 3D graphics and visualization, cloud computing, machine learning, and deep learning.