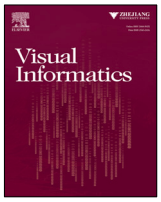
Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf



Highlights

Interactive lighting editing system for single indoor low-light scene images with corresponding depth maps

Visual Informatics xxx (xxxx) xxx

Zhongyun Bao, Gang Fu, Lian Duan, Chunxia Xiao*

- We propose a single indoor image relighting method under the interactive target lighting.
- we achieve the coarse-to-fine depth image and decompose the input RGB image.
- We apply our proposed method to the single indoor low-light scene image relighting.

Graphical abstract and Research highlights will be displayed in online search result lists, the online contents list and the online article, but will not appear in the article PDF file or print unless it is mentioned in the journal specific style requirement. They are displayed in the proof pdf for review purpose only.

Please cite this article as: Z. Bao, G. Fu, L. Duan et al., Interactive lighting editing system for single indoor low-light scene images with corresponding depth maps. Visual Informatics (2022), https://doi.org/10.1016/j.visinf.2022.08.001.

VISINF: 150

ARTICLE IN PRESS

Visual Informatics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Interactive lighting editing system for single indoor low-light scene images with corresponding depth maps

Zhongyun Bao, Gang Fu, Lian Duan, Chunxia Xiao *

School of Computer Science, Wuhan University, Wuhan, 430072, Hubei, China

ARTICLE INFO

Article history:
Received 19 December 2021
Received in revised form 17 August 2022
Accepted 27 August 2022
Available online xxxx

Keywords:
Image processing
Interactive relighting
Spherical harmonic lighting
Depth map
Intrinsic image decomposition

ABSTRACT

We propose a novel interactive lighting editing system for lighting a single indoor RGB image based on spherical harmonic lighting. It allows users to intuitively edit illumination and relight the complicated low-light indoor scene. Our method not only achieves plausible global relighting but also enhances the local details of the complicated scene according to the spatially-varying spherical harmonic lighting, which only requires a single RGB image along with corresponding depth map. To this end, we firstly present a joint optimization algorithm, which is based on geometric optimization of depth map and intrinsic image decomposition avoiding texture-copy, for refining the depth map and obtaining the shading map. Then we propose a lighting estimation method based on spherical harmonic lighting, which not only achieves the global illumination estimation of the scene, but also further enhances local details of the complicated scene. Finally, we use a simple and intuitive interactive method to edit the environment lighting map to adjust lighting and relight the scene. Through extensive experimental results, we demonstrate that our proposed approach is simple and intuitive for relighting the low-light indoor scene, and achieve state-of-the-art results.

1. Introduction

Image relighting (Bao et al., 2022) aims to relight image and enhance the image scene details, which has a wide range of applications in the fields of image editing (Zhang et al., 2015; Haouchine et al., 2017; Gui and Zeng, 2019; Fu et al., 2021), scene understanding (Sun et al., 2020), etc. Particularly, with the development of smart devices and image editing software, users expect more convenient and high-quality lighting editing of images under various lighting environments. However, it is difficult because the lighting and shadows in the scene of the low-light environment are very complicated and cannot be represented by a simple lighting model, which makes the edited illumination exhibit serious defects in the bright part, lack of realism, and disorder scene lighting. Although most of the existing image-based relighting techniques achieve image relighting, they show poor relighting results in complicated low-light indoor scenes, and cannot provide users with simple and intuitive interactive relighting operations.

Previous method (Wu and Saito, 2017) (IRSI) achieves better lighting editing and enhancement effects for simple scenes by using manual interaction to mark the light spot for estimating the position and intensity of the light source. However, this relighting method requires accurate scene geometry information, user annotation, multiple or even a large number of photographs with specific requirements or special precision equipment. This method shows poor relighting in complex scenes, especially for low-light indoor scenes. Garon et al. (2019) (FSV) proposed a real-time method to estimate spatially-varying indoor lighting from a single RGB image. Given an image and a 2D location in the image, this method estimates a 5-th order spherical harmonic representation of the lighting at the given location. Although it achieves spatially-varying lighting, it needs large datasets to train the model and the user cannot perform intuitive lighting editing and interaction. Also, for the low-light indoor complex scenes, it presents poor relighting results.

Different from these methods, in this paper, we are inspired by such a fact that due to the complexity of lighting in the scene with a low-light environment, lighting adjustment cannot be performed by only estimating light sources. We fully combine the shading maps and depth maps to achieve the lighting editing and enhance results for low-light scene. We propose an interactive lighting editing and relighting algorithm, which is concentrated on complicated indoor low-light scenes. We not only provide users with a simple and intuitive interactive way to edit lighting and reasonably relight scenes in low-light conditions, but also effectively estimate the local lighting information of the scene to adjust the illumination and enhance the local details of complicated scenes.

* Corresponding author.
E-mail address: cxxiao@whu.edu.cn (C. Xiao).

https://doi.org/10.1016/j.visinf.2022.08.001

Fig. 1. Our algorithm for relighting a complex indoor scene from a single image under low-light environment. It only requires an RGB-D image as input to achieve global relighting (the first row) as well as spatially-varying lighting (the second row) for the scene using an intuitive and simple interactive method.

We firstly take an RGB-D image as input, which are used to perform depth map refinement and intrinsic image decomposition to avoid texture-copy by a joint optimization framework. Then we combine the fined depth map to estimate the lighting from the shading map, and propose a lighting estimation algorithm, which uses spherical harmonic lighting model and lighting detail maps together to represent complex environment light. Finally, we use a simple interactive relighting method by editing the environment light map to intuitively adjust lighting and relight the scene for user. The experimental results show that the method can produce more plausible lighting results than previous methods, enhance the local details of scenes, and provide a better foundation for applications such as image inpainting and image content retrieval. Fig. 1 comprehensively shows the global relighting effects for the complicated scene (the first row) and the local detail enhancement effects of the scene (the last row).

To sum up, the contributions of our work are as follows:

1. We design a single indoor scene image relighting algorithm that takes a source RGB image and the corresponding depth map as input to produce a relighting indoor scene image. It not only achieves plausible global illumination for the entire scene, but also enhances the local details of the complex scene according to global–local SH lighting.
2. We propose a joint optimization framework that achieves a coarse-to-fine depth map and decomposes the input RGB image into reflectance map, shading map and lighting detail map, it can also estimate the lighting information of a scene.
3. Our joint optimization framework is compatible with our proposed lighting model and complex indoor scene. Note that our algorithm provides users with a simple and intuitive interactive light editing system.

2. Related works

2.1. Depth map refinement

The general idea of the RGB image-based depth map optimization method is to first estimate the lighting condition and the material of the RGB image, and then use the lighting to optimize the details of the coarse depth map. Böhme et al. (2010) first proposed a TOF (time of flight) camera depth image optimization method. This method considers that the intensity image of a scene and their depth image meet a certain lighting relationship, and establishes a lighting constraint relationship with the depth image to optimize the depth map. Yu et al. (2013) proposed a method for Kinect depth map optimization. It uses the block-based depth map repairing method to maintain the structural geometry of the depth map, and then combines the estimated spherical harmonic lighting, the reflectance map and the repaired depth map to get the final optimization result. Or-El et al. (2015) established a real-time depth optimization framework for scene with complex reflectivity, which can deal with natural light or multiple light sources. Zollhöfer et al. (2015) used SDF (signed distance function) to represent the surface of the object, and combined SRGB to directly optimize SDF. Kadambi et al. (2015) combined the normal information of the surface from the image by combining the polarization information of the image and the rough depth map to obtain a more accurate depth map. The advantage of these methods is that RGB images are easy to obtain, and the image quality and resolution of the visible light sensors have been better developed. When the textures of image surface is complicated in the realistic scene, and the depth is optimized, these textures will be left on the depth map, which causes obvious texture-copy problem. To this end, we propose a joint optimization algorithm that is based on geometric optimization of depth maps as well as intrinsic image decomposition to avoid texture-copy problems, and provide a key step for later implementation of more plausible lighting.

2.2. Intrinsic image decomposition

Intrinsic image decomposition is mainly divided into two categories. One mainly achieves better discrimination between the reflectance and the lighting of image by adding more constraints to them. Rother et al. (2011) added a new prior information to the reflectivity, assuming that the color value of the reflectance is in a sparse set of tones, and that the decomposition of the reflectivity component is not required to know the boundary information. More methods (Shen and Yeo, 2011; Shen et al., 2013) observed that adjacent pixels with similar hues tend to have the same reflectivity value in the reflectivity map. Li and Brown (2014) divided the image into two layers according to the gradient histogram and established a probabilistic model framework for optimization. These constraints usually aimed at the daytime or daily indoor environment.

The others use statistical learning to distinguish between the reflectance and the light components. For example, in the method (Tappen et al., 2003), a classifier is used to identify

whether the local gray blocks of an image belong to reflectance or illumination. The method (Tappen et al., 2006) used nonlinear regression to decompose the image into shading map, reflectance map, and noise map. In recent years, with the development of deep learning, many intrinsic image decomposition methods have achieved good results. The method (Fan et al., 2017) combined deep neural networks and is sensitive to edge information, and the method (Lettry et al., 2018) Jointly estimated reflectance and illumination using fully convolutional networks.

2.3. Lighting estimation and editing

Many lighting editing methods firstly estimate the position, direction, and intensity of the light source, then edit the properties of the light source, then combine them to obtain a light-edited image. Cao and Shah (2005) proposed a method for estimating the linear light source of a scene. The methods (Chandraker and Ramamoorthi, 2011; Hara et al., 2007) further utilized a variety of information in the scene to improve the accuracy of the linear light source estimation, such as the sky and ground area information in large outdoor scenes. However, the focus of our paper is to estimate and edit the lighting in low-light environments, so we cannot simply assume that the light source in the scene is a linear light source. Xu et al. (2018) proposed an optimal sparse sampling method based on deep learning, which trains a given light source position and the formed image to adjust the linear light source.

The methods (Cao and Shah, 2005; Lagger and Fua, 2006) focused on the estimation of nonlinear light sources and proposed a specular-based light source estimation method. Although the 3D information of the objects in the scene can be obtained through some methods, the scenes in the real environment often exhibit complex geometric structures, and the highlights have a strong sensitivity to the normal of the object surface, making this type of method only suitable for simple geometric model.

Some methods achieve more accurate lighting estimation by giving multiple images and special equipment, for example the methods (Boyadzhiev et al., 2013; Ren et al., 2015; Tunwattanapong et al., 2011) used multiple images in different lighting environments to achieve the effect of light editing. However, these methods have high requirements for shooting, which require us to fix the camera position as well as changing the area illuminated by the scene lighting. Wu and Saito (2017) proposed an interactive LDR image enhancement method. This method can achieve better real lighting editing and enhancement effects. However, the scenario they assumed is relatively simple, and the method is only suitable for large outdoor scenes at night, and needs to construct a geometric scene through manual interaction, which has certain requirements for user input.

3. Overview

In this paper, we propose an approach that estimates illumination of large and complicated indoor scenes under the low-light conditions and relights the scenes by an intuitive and simple interactive method. We not only achieve plausible global relighting, but also enhance the local details of complex scenes by spatially-varying lighting. The detailed algorithm structure is shown in Fig. 2.

3.1. Geometry optimization of depth maps

Many existing works use shape-from-shading techniques to optimize the details of the depth map, and get better results. However, since the texture information of the object surface is unknown, most existing methods assume it to be a constant or sparse image block, which is brought into the depth map. The phenomenon that texture components remaining in the depth map is called the texture-copy, which will further affect the results of light editing and make the relighted image appear additional texture defects. To avoid the problem of texture-copy and achieve better lighting effects, we propose a geometry optimization algorithm for depth map that avoids texture-copy problem. Inspired by methods (Or-El et al., 2015; Xu et al., 2012), we first decompose the image into a texture layer and a structure layer, and then use the structure layer to optimize the depth map for solving the texture-copy problem. We also find that the relative total variation model (RTV) (Xu et al., 2012) is suitable for removing textures of low-light images while retaining their structural information. The RTV model is defined as follows:

$$\mathrm{RTV} = \sum_{d\in\{x,y\}} \frac{\nu_d(p)}{\varsigma_d(p)+\varepsilon}, \tag{1}$$

where p refers to the position of each pixel in the image, d refers to the horizontal x and vertical y directions, ν is the total variation of the window, ς is the inherent variation of the window, and ε is a small value to avoid division by 0. The windowed total variation ν is defined as:

$$\nu_d(p) = \sum_{q\in\Omega(p)} g(p,q)\left|(\partial_d S)_q\right|, \tag{2}$$

where $\Omega(p)$ represents a rectangular area centered on pixel p, the size of which is 3×3, S is the image to be solved that contains only structural information, and $g(p,q)$ is a weight function. The windowed intrinsic variation ς_d is defined as:

$$\varsigma_d(p) = \left|\sum_{q\in\Omega(p)} g(p,q)(\partial_d S)_q\right|. \tag{3}$$

The RTV model can effectively represent the structure in the original gray image. To separate the texture and structure in the original grayscale image, a minimization equation needs to be solved, which is defined as follows:

$$\arg\min_{S_{ir}} \sum_p (S(p)-I(p))^2 + \lambda_s RTV(p), \tag{4}$$

where I and S indicate the original gray-scale image and the structure image, respectively. Note that the first term is designed to make the structure map and the original gray image as close as possible, and the second term is to make the structural component in the image as prominent as possible.

After the texture of the image is removed, we can approximate this component as a constant. Then we use spherical harmonic illumination based on the spherical harmonic function proposed by the method (Landis, 2002) to solve this problem.

In the process of optimizing depth maps of low-light scenes, although the textures of original RGB images are removed, they still contain the shadow and highlight information, so the assumption of Lambert model is not suitable for representing them. To this end, based on the RGBD-Fusion method, we first build a non-Lambertian lighting model, then use spherical harmonic function to estimate the lighting of images and constrain the reflectivity, and finally solve the energy equation to obtain an optimized depth map. The lighting model for the structural image in this paper is expressed by

$$S(p) = \rho(p)\mathcal{H}(p) + \beta(p), \tag{5}$$

where ρ represents the component with a larger gradient in the image, mainly includes the larger spatial scale reflectance and shadow, and shadow with larger spatial scale, $\mathcal{H}$ is the spherical harmonic illumination model, and β is the highlight information. After establishing a non-Lambertian lighting model, we first

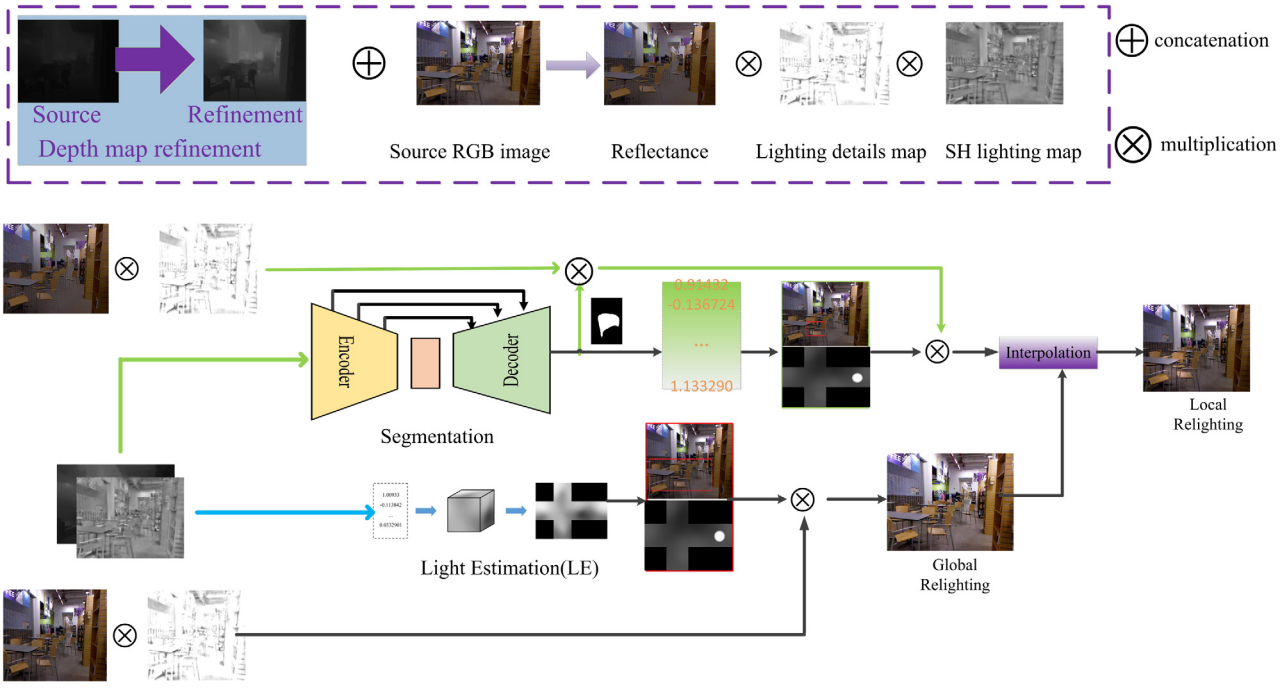


Fig. 2. The overall structure of proposed method. It consists of lighting editing and relighting mechanisms. The core idea of the method is to estimate the illumination of large and complicated indoor scenes under the low-light conditions, and achieve plausible global and local scenes relighting by an intuitive and simple interactive way. In the purple box, the left side of the arrow indicates the input of the system, and the right side of the arrow indicates the sub-result of the system output.

assume that the reflectance map is a constant to solve the illumination $\mathcal{H}(p)$ by the method (Wu et al., 2014). In addition, although the original depth map is coarse, the illumination estimation is not sensitive to its geometric details. Therefore, we assume that the values of ρ and β are always 1 and 0, respectively, and establish a spherical harmonic illumination model and solve its coefficients. Since image only contains structural information, most of the content is distributed in the low-frequency part, and the light can be used by a lower order spherical harmonic coefficient. We thus use a first-order spherical harmonic model $\mathcal{H}(p)$, which is defined as follows:

$$\mathcal{H}(p) = \overrightarrow{h_1}\widetilde{N}(p) + h_0, \tag{6}$$

where $\overrightarrow{h_1}$ and h_0 are the 1st and 0th order spherical harmonic coefficients respectively; $\widetilde{N}(p)$ is the unit normal of the object surface at point p. In the lighting model (5), when the structure map S and lighting remain constant, the details of the depth map have greater correlation with ρ and β. To avoid the effect of reflectivity, shadows, and highlights in the structure map on the depth map, we establish the following constraint relationship $\Re(p)$ for ρ.

$$\Re(p) = \sum_{q\in\Omega(p)} W_s(p,q)W_d(p,q)(\rho(p) - \rho(q)), \tag{7}$$

where W_s is the constraint function combined with the structure diagram, which is defined as:

$$W_s(p,q) = \begin{cases} 0, \|S(p) - S(q)\|^2 \geq \tau \\ \exp(\frac{\|S(p)-S(q)\|^2}{2\sigma_s^2}), otherwise \end{cases} \tag{8}$$

where τ is the threshold of the reflectivity change of the structure map. This function indicates that in a field of point p of the structure map, the larger difference between the pixel value of another point and the value of point p is, the smaller $W_s(p,q)$ will be. W_d is the constraint function combined with the depth map and is defined as:

$$W_d(p,q) = \exp(-\frac{\|D(p) - D(q)\|^2}{2\sigma_d^2}). \tag{9}$$

Similarly, W_d measures the relationship between the depth map and the reflectance map. When the depth varies drastically, the difference in reflectance will be greater and a relatively small W_d will be obtained. Because ρ represents the reflectivity with a large gradient and the highlights have little effect on it, we firstly let β be constant at 0, and then solve a minimization equation problem to find ρ:

$$\arg\min_{\rho} \sum_{p} \|\rho(p)\mathcal{H}(p) - \mathcal{S}(p)\|^2 + \lambda_\rho \|\Re(p)\|^2. \tag{10}$$

3.2. Intrinsic image decomposition

In this paper, we combine the refined depth image obtained in 3.1 and the shadow boundary confidence image obtained by shadow detection to decompose a given original RGB image into a reflectance map, a shading map, and a shadow map. In addition, to reduce the ambiguity between the reflectance map and the shading map of the image, we utilize an energy minimization equation including data term, smooth term, and constant term to solve unknown quantities, and get the decomposed images. The energy optimization model is as follows:

$$\begin{aligned} E(R,S,H) =& E_{data}(R,S,H\$) + \lambda_s E_{smooth}(R,S,H) \\ &+ \lambda_c E_{const}(R,S,H), \end{aligned} \tag{11}$$

where variables R, S, and H are the reflectance map, shading map, and shadow map to be solved, respectively; E_{data} is the data term that enables the solved reflectance map, shading map, and shadow map to reconstruct the original RGB map as much as possible, E_{smooth} is the smooth term and E_{const} is the constant term. λ_s and λ_c control the magnitude of the smooth and constant terms, respectively.

Data constraint. We assume that the light in the environment is white, that is, the shading map is a gray image. Therefore, for each color channel of an RGB image, a data constraint is established, that is, $I_c \approx S_c R_c H_c$, $c \in \{R, G, B\}$. In addition, to make the combination of the three channels of the image closer to white, we add a weight coefficient ω_c on each channel. The data term is defined as follows:

$$E_{data} = \omega_{iw} \sum_{c\in\{R,G,B\}} \omega_c |I_c - S_c - R_c - H_c|^2, \tag{12}$$

where $\omega_R, \omega_G, \omega_B = 0.299, 0.587, 0.114$, ω_{iw} is the image intensity weight coefficient.

Smooth constraint. The smooth term contains three energy constraint terms: smooth reflectance E_{rs}, smooth light E_{ss}, and smooth shadow E_{hs}. The smooth term is defined as follows:

$$E_{smooth}(R, S, H) = E_{rs}(R) + \gamma_s E_{ss}(S) + \gamma_h E_{hs}(H), \tag{13}$$

where γ_s and γ_h are the light and shadow smoothing weight coefficients, respectively. We assume that the reflectance map is composed of some constant image blocks, which is the assumption of reflectance sparsity. This assumption can be achieved by the following energy minimization function:

$$E_{rs}(R) = \sum_{p\in R} |C_B(p)| \sum_{q\in N} \omega^R_{p,q} \left\| R_p - R_q \right\|^p, \tag{14}$$

here $\omega^R_{p,q}$ is the weight of the reflectance block, N represents the pixels, and $C_B(p)$ is the shadow boundary confidence image, which can be calculated as follows:

$$C_B(p) = \sqrt{V^B(p) V^D(p)}, \tag{15}$$

where D is the refined depth information, and $V^{B,D}(p)$ is represented as follows:

$$V^{B,D}(p) = \sqrt{\frac{L_x^{\{B,D\}}(p)^2 + L_y^{\{B,D\}}(p)^2}{\mathcal{D}_x^{\{B,D\}}(p)^2 + \mathcal{D}_y^{\{B,D\}}(p)^2 + \varepsilon}}, \tag{16}$$

where L and $\mathcal{D}$ mean intrinsic variational function and full variational function.

For the shading map, the corresponding constraint is as follows:

$$E_{ss}(S) = \sum_p (1 - C_B(p)) \left(S_p - \sum_{q\in N_p} \omega^N_{pq} S_q \right)^2, \tag{17}$$

where ω^N_{pq} can be obtained by the following optimization function:

$$\arg\min_{\omega^N_{pq}} \sum_{p\in P_N} \left\| N(p) - \sum_{q\in N_I(p)} \omega^N_{pq} N(q) \right\|^2. \tag{18}$$

For the shadow map, we need a smooth constraint that is sensitive to shadow boundaries:

$$E_{hs}(H) = \sum_p (1 - |C_B(p)|) \sum_{q\in N_p} \left\| H_p - H_q \right\|^2, \tag{19}$$

where N_p represents pixels adjacent to space p.

Constant constraint. The constant constraint contains two terms, the hue term and the shadow constant term, which is defined as follows:

$$E_{const}(R, H) = E_{ch}(R) + \alpha E_{sc}(H), \tag{20}$$

where $E_{ch}(R)$ is the reflectance hue constraint and $E_{sc}(H)$ is constant constraint of the shadow. The definition of the reflectance hue constraint is as follows:

$$E_{ch}(R) = \left\| \frac{\mathbf{I}}{I + \varepsilon} - \frac{\mathbf{R}}{R + \varepsilon} \right\|^2, \tag{21}$$

where $\mathbf{I}$ and $\mathbf{R}$ indicate the source RGB image and reflectance image, respectively. For constant constraint of the shadow, we select pixels in the reliable area, as well as constraining their shadow value to 1 and keeping the color value constant. The constraint is defined as follows:

$$E_{const}(H) = \sum_{p\in N_b} \left\| H_p - 1 \right\|^2, \tag{22}$$

where N_b is a reliable region, which is defined as a region with strong confidence in the shadow boundary and its adjacent pixels.

3.3. Interactive editing illumination and relighting

One of the core ideas of our algorithm is to propose a simple and intuitive interactive illumination editing method to relight complex indoor scenes under low-light conditions. To this end, an illumination editing and relighting method based on spherical harmonic lighting is proposed. For the complex low-light scene, the method can represent it by combining spherical harmonic lighting model and lighting detail maps, and can also intuitively adjust or enhance it through interactive spherical harmonic parameter adjustment way. The algorithm not only achieves global illumination of complex scenes, but also enhances local details of scenes by spatially-varying lighting.

Spherical harmonic lighting model. Since the lighting editing is only related to shading maps and shadow maps, the goal of this paper is to find a way to edit them more efficiently and conveniently. In addition, we are dealing with objects with complex scenes and unknown lighting sources under low-light conditions indoor. We thus use the spherical harmonic lighting model to represent the illumination of the scene by a higher degree spherical harmonic function. We define the light decomposition model as follows:

$$S(p) = \mathcal{L}(p) \mathcal{SH}(p), \tag{23}$$

where $\mathcal{L}$ is the lighting detail map and $\mathcal{SH}$ is the spherical harmonic lighting map and its calculation method is as follows:

$$\mathcal{SH}(p) = \sum_{i=1}^{N} c_i Y_i(p), \tag{24}$$

where Y is the spherical harmonic basis function, N is the number of spherical harmonic basis functions, and c is the spherical harmonic coefficient.

Lighting estimation and decomposition. We firstly assume that the lighting detail map is constant one and estimate the spherical harmonic parameters, then uses these parameters to render the spherical harmonic lighting map, which is used to calculate the lighting detailed map by combining the original shading map. To facilitate the calculation and representation, we construct all spherical harmonic basis functions of the depth image into a matrix

$$A = \begin{pmatrix} Y_0(P_0) & \dots & Y_N(P_0) \\ \dots & \dots & \dots \\ Y_0(P_n) & \dots & Y_N(P_n) \end{pmatrix}, \tag{25}$$

where $p_0 \dots p_n$ are all pixels in the depth map. To make spherical harmonic lighting as close as possible to the original light and dark map, the following relationship should be satisfied:

$$\arg\min_{\mathrm{h}} \left\| S(p) - A \cdot \mathbf{h} \right\|, \tag{26}$$

where $\mathbf{h}$ is a column vector composed of spherical harmonic coefficients, that is,

$$\mathbf{h} = (A^T A)^{-1} A^T S(p). \tag{27}$$

Fig. 3. Spherical harmonic light decomposition results. From left to right, they are the shading map, lighting detail map and spherical harmonic lighting map, respectively.

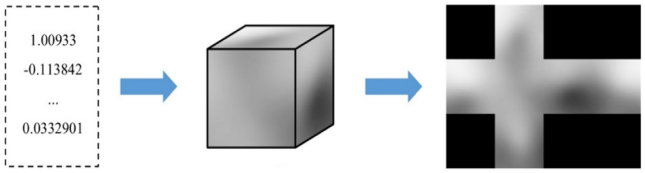


Fig. 4. The rendering process of environment lighting map. From left to right, they are the spherical harmonic coefficient, cube map and light distribution map, respectively.

After using a spherical harmonic lighting model to approximate the shading maps, the lighting details can be directly solved by the following formula:

$$\mathcal{L}(p) = \frac{S(p)}{A \cdot \mathbf{h} + \varepsilon}. \tag{28}$$

The decomposition results of the lighting detail map and spherical harmonic lighting are shown in Fig. 3. The lighting detailed map is mainly caused by high-frequency local lighting in the scene, while the spherical harmonic lighting mainly represents the impact of external environment light on the scene.

In this paper, we use spherical harmonic lighting to represent environment light, and there is no need to assume a single light source. We achieve the lighting editing on the overall spherical harmonic lighting. As we know, the spherical harmonic coefficients cannot directly represent the lighting situation, and an interactive method is needed to realize the intuitive adjustment of the lighting. We thus design a lighting editing method that adjusts the environment light map to achieve the editing of lighting in the scene.

Environment lighting representation. Cube maps or spherical maps are commonly used in computer graphics to represent environment lighting. Cube mapping was first proposed by the method (Greene, 1986), and it can better maintain the original lighting shape of the scene, and more intuitively shows the area corresponding to each direction, this paper therefore uses cube maps to represent the distribution of environment light. For a fixed spherical harmonic coefficient, the corresponding environment light is also fixed. Through the estimation of spherical harmonic illumination, a set of spherical harmonic coefficients are calculated in this paper, and our goal is to render spherical harmonic coefficients into cube maps. For the convenience of calculation, the length of the sides of the cube map is set to 1. The rendering process of environment lighting map is shown in Fig. 4.

According to Fig. 4, the spherical harmonic parameters can be used to obtain the lighting in each space direction. Therefore, we first map each pixel to the space position of the cube map, and then all points on the cube map are rendered to generate a lighting map for each face.

User interaction and editing. We change the environment light by editing the environment light map, which adds lines or white spots on the map through user interaction to increase the area light, fills the bright area with black to remove the original lighting, or increases the overall brightness of the map to enhance the overall image lighting. However, the user's input has great uncertainty. If these maps containing random patterns are directly added to the original scene as environment light, the rendered image will show obvious defects. Therefore, after the environment lighting map is edited by the user, we resample the new environment lighting map to estimate new spherical harmonic parameters. Since spherical harmonic lighting represents low-frequency lighting information, spherical harmonic parameters are robust to user input. Even if the environment light edited by the user is chaotic, after resampling the environment light, some high-frequency details in the interaction process will not be brought into the light, which can better reconstruct the environment light. As shown in Fig. 5, the white line in (b) is the light added by the user. After re-evaluating the spherical harmonic lighting, this area forms a softer area light.

Fig. 5. Environment lighting of user interaction. From left to right, they are original environment lighting, user interaction and re-estimated environment lighting, respectively.

Global relighting. Note that the spherical harmonic function is defined on the spherical surface. When estimating spherical harmonic parameters, it is necessary to ensure uniform sampling on the spherical surface. An RGB image with the depth map is projected into a spatial point cloud, and then converted into a spherical coordinate system, the center area is often sparse and the corners of the image are denser. In order to eliminate the impact of this sampling error, we first use a normal distribution to generate a three-coordinate value sequence. That is, $(x, y, z) \sim N(0, 1)$, these points are uniformly distributed in all directions of the origin, and then this group of points is normalized to obtain a group of uniformly distributed points, which are located on a spherical surface with a radius of 1, denoted as s. We apply the method (Green, 2003) to sample the environment light map and calculate the spherical harmonic parameters:

$$c_i = \frac{4\pi}{N} \sum_{j=1}^{N} M(s) Y_j(s), \tag{29}$$

where $M(s)$ represents the brightness value of the vector from the origin to the point s on the environment light map. After sampling and editing the spherical harmonic lighting, we combine the spherical harmonic lighting map, lighting detail map and reflectance map to re-render the scene, the rendering method is as follows:

$$I'(p) = R'(p)\mathcal{L}(p)\mathcal{SH}(p). \tag{30}$$

Spatially-varying relighting. Although the global illumination achieves the relighting in a certain direction of the scene and the overall brightness enhancement in the whole scene by a single spherical harmonic basis, global illumination cannot faithfully represent scene lighting for all surface points simultaneously, in other words, it cannot accurately enhance the local details of the complex scene because of the characteristics of environment light. To this end, we also propose a spatially-varying relighting method to enhance the local details by relighting the local scene.

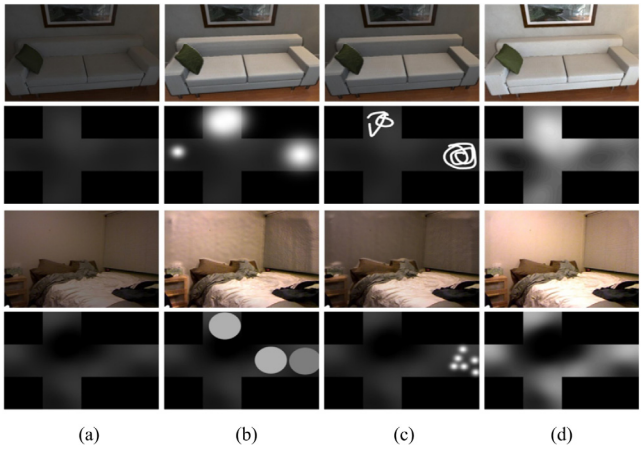


Fig. 6. Low light image light editing and enhancement effects, where the odd rows are RGB images and the even rows are the corresponding environment light editing process; For image column (a) are the original images, (b) and (c) are the results after adding a light source; (e) is the result of overall light enhancement.

Inspired by the method (Maier et al., 2017), we first apply Mask-RCNN (He et al., 2017) to segment the target scene, then we assume that the scene is partitioned into N local scene blocks with N sets of different lighting coefficients, for each segmented local scene block, we use unknown lighting coefficients L with a set of spherical harmonic coefficients up to the second order to represent the lighting information by the same method as global relighting.

4. Experiments

4.1. Implementation and parameter setting

In this paper, we use MATLAB to implement the depth map optimization and intrinsic image decomposition, the testing machine is configured with a processor Intel i3-4130 and 12 GB of memory. The RTV adjustment parameter is set to $\lambda_s = 0.015$, the texture filter window size is set to 3, and the structure map constraint adjustment parameter $\sigma_s = 0.22$, The reflectivity map constraint adjustment parameters are set to $\sigma_d = 7.071$ and the minimum weight coefficients $\lambda_\rho, \lambda_{\beta 2}, \lambda_{\beta 1}, \lambda_{d1}, \lambda_{d2}$ are set to 0.1, 1, 1, 0.004, and 0.0075, respectively. For the optimization of a depth map, our method takes about one minute. In the experiment, when the depth map optimization method and intrinsic image decomposition method are used for preprocessing, the total time is about three minutes.

We use ICL-NUIM and NYU datasets to implement lighting editing. Illumination editing and relighting are implemented using C ++, Pytorch and OpenCV. The size of each image used in the experiment is 640 × 480. The light decomposition process took an average of 0.78 s on the test machine, and the light synthesis process took an average of 1.01 s. User interaction is performed on the expanded ambient light map through image editing software. The editing process mainly includes the addition of light sources, the enhancement of overall lighting, and the removal of existing lighting. After editing multiple scenes, the results are shown in Fig. 6. It can be seen that editing the ambient light map can effectively adjust the scene lighting, and the adjusted scene remains realistic. Note that (b) and (c) of Fig. 6 show the relighting results from different lighting directions, and (d) shows the relighting results by lighting up the whole scenes.

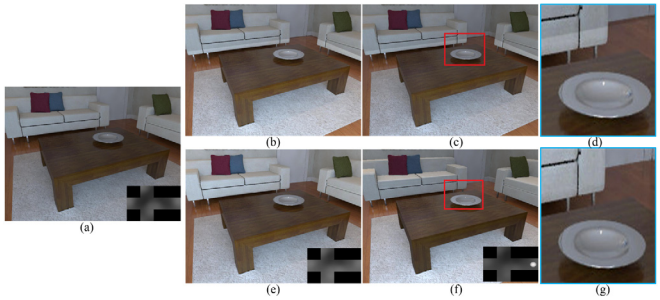


Fig. 7. Comparison of the results of this method and the IRSI, (a) is the original image; (b) and (c) are the light editing results by the IRSI; (e) and (f) are our relighting results; (d) and (g) are local details of (c) and (f).

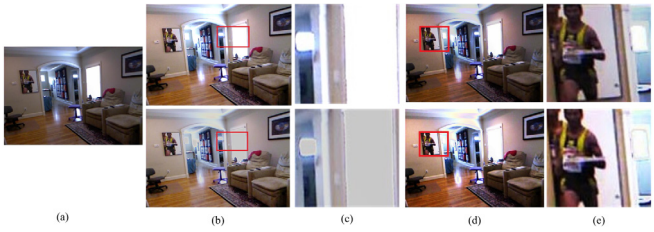


Fig. 8. Visual relighting results for our method and the FSV. (a) is the source image; (b) and (d) are global relighting images and local relighting images respectively. The first row results are produced by our method, and the second row results are obtained by the FSV. (c) and (e) are their local detail images.

4.2. Comparison with state-of-the-art methods

We compare the lighting editing effects of our method to IRSI (Wu and Saito, 2017) and the method FSV (Garon et al., 2019). The IRSI is an interactive lighting editing method. Firstly, the user marks the three-dimensional plane and lighting information in the scene, and then the lighting in the scene is edited by moving an existing light source or inserting a new light source. IRSI mainly aims to adjust models, including point light source, linear light source and spotlight, while our method is to adjust the environment light. To make the comparison results effective, when using two methods to edit the illumination, we only add straight-line lighting in the direction perpendicular to the ground, or add the lighting to the entire scene. Fig. 7 is the comparison results between our method and the IRSI. Among them, (b) and (e) enhance the overall lighting, and (c) and (f) add lighting directly above the floor. It can be seen from the blocks (d) and (g) that the results of the IRSI method have unnatural brightness mutations at the interface of planes such as sofas. This is because IRSI assumes that the scene consists of several simple planes. This assumption is reasonable in some outdoor simple building scenes, but it often causes the relighted object to be distorted in complex low-light indoor scenes. Due to making full use of the scene depth map information, our method can better edit lighting for complex scenes, and achieves better results in detail than the IRSI method. For example, in our algorithm, the brightness of the plate in the scene changed with the addition of vertical lighting and obtain plausible relighting effect.

The method FSV proposes a real-time method to estimate spatially-varying indoor lighting from a single RGB image. For a given image with the corresponding 2D local block, their method estimates a 5th order spherical harmonic representation of the lighting at the given location. Although this method achieves local illumination estimation and relighting results, it relies on a large number of training data sets. Comparing with our method, as shown in Fig. 8, this method is more sensitive to the complex indoor scenes and overexposed places, and obtains the poor

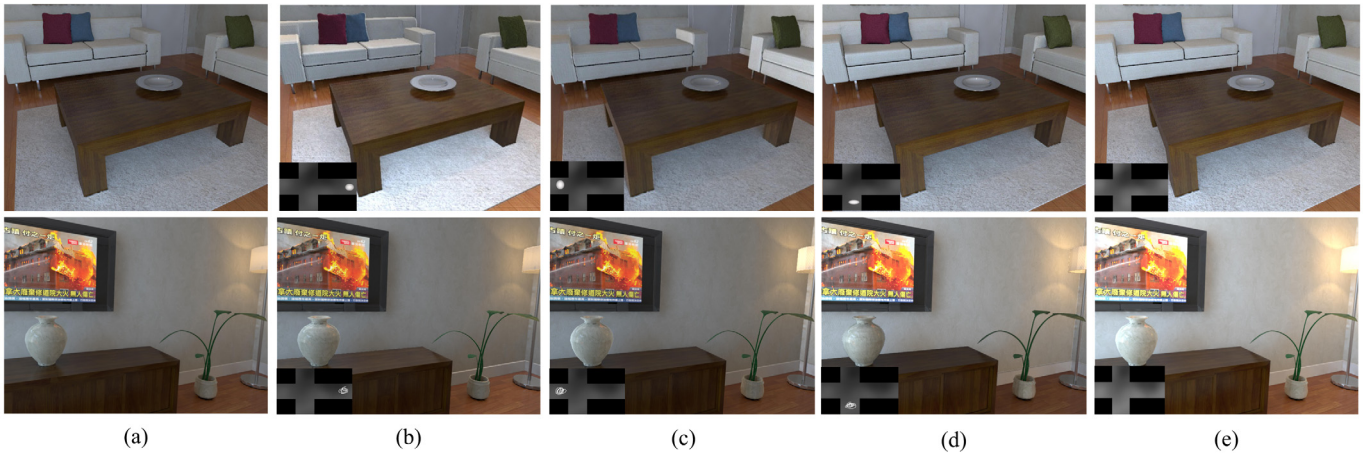


Fig. 9. The results of global relighting in different complex indoor scenes using our method; (a) are original images, (b), (c) and (d) are the results of global relighting, they are relighted from a certain direction of the scene by adding a light source from the top, left and front of the scene, respectively. (e) are the results of overall relighting.

Fig. 10. Comparison of low-light image enhancement results between our method and LIME (under the same conditions: y=0.7). From left to right, they represent the original image, the enhancement image for LIME, the relighting image of ours and Local details comparison of experimental effects from LIME and our algorithm.

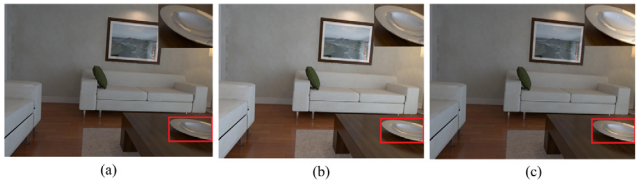


Fig. 11. The results of global relighting and local relighting of scene. (a) is the original image, (b) is the global-relighting image and (c) is the local-relighting image.

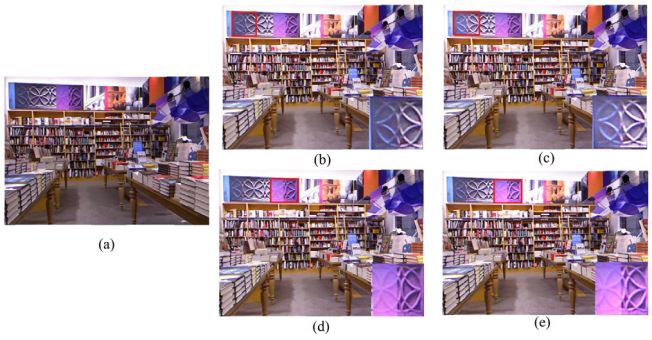


Fig. 12. The results of our global relighting and global–local relighting in complicated indoor scene. (a) is original image, (b) and (d) are the global relighting, (c) and (e) achieve the global-local relighting by combining the global and local spherical harmonic coefficients.

relighting effects. On the contrary, our algorithm has a strong tolerance to these factors, it can obtain better relighting effects.

Global relighting. To verify the relighting effects of our algorithm in different low-light indoor complex scenes, we select two sets of indoor scenes to visualize their relighting effects by using our method. In Fig. 9, we choose two different indoor scenes, which are 640 × 480. We use interactive editing way to add light sources in different directions of the scenes to relight the scenes. The first column images are original images. From the second column to the fourth column, they are the results of global relighting by adding a light source from the top, left and front of the scene, respectively. The last column images are the results of overall relighting. It can be seen from the appealing Fig. 9, our algorithm can achieve relighting in different indoor complex scenes, the relighting effects are relatively natural and plausible, and the scene illumination under low light conditions is effectively enhanced. In addition, our method not only can realize the light editing, but also can enhance the low-light image by performing gamma correction on the ambient light or editing the environment light into a constant image with higher brightness. In addition, we compare with the LIME method (Guo et al., 2016). LIME is a low-light enhancement method based on intrinsic image decomposition, combined with gamma correction for lighting adjustment. To make other results faithful to their methods, we use the same gamma correction parameters as LIME to adjust the environment light map. As shown in Fig. 10, the parameter y = 0.7 is used for this result. Although the overall brightness level of the two is relatively close, there is a big difference in details: LIME does not consider the three-dimensional information of the image, so the entire image is enhanced, which will make the original dark areas enhanced with more defects. In our method, because the original environment light of the scene is enhanced, the result is closer to the real lighting enhancement, and the darker area of the enhanced original image has fewer defects.

Local relighting. Although we have achieved global relighting of indoor complex scenes under low-light conditions, and the lighting effect achieved is more plausible, it cannot accurately enhance the local details of the complex scenes. To this end, we propose a spatially-varying relighting algorithm to achieve local detail enhancement of scenes. The visualization enhancement effects are shown in Figs. 11 and 12. In Fig. 11, we achieve global relighting (b) and local relighting (c) on the original image (a), respectively. We can see that the lighting details of the plate in the scene with three different lighting conditions are obviously different from magnifying the local block. Especially, for local relighting, the lighting detail of the plate in image (c) is clearer than

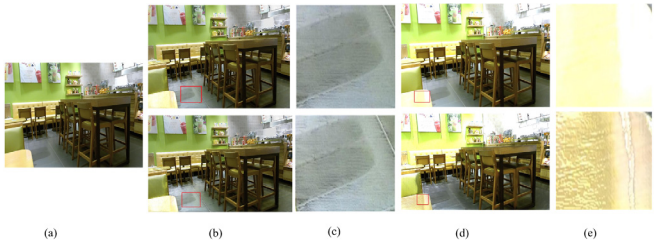


Fig. 13. Visual relighting results for the single indoor images using our method with and without depth map optimization. 1st row: relighted images with depth map optimization; 2st row: relighted images without depth map optimization. (a) Input. (b) and (c) are the relighted images from a single direction and the corresponding detail maps respectively; (d) and (e) are the overall relighted images and the corresponding local detail maps respectively.

others. Fig. 12 shows that the local relighting in the scene is better approximated with global–local spherical harmonic coefficients compared to only with global spherical harmonic coefficients. According to comparison results between image (b) and image (c), or between image (d) and image (e), we can see that the global–local relighting can highlight the details of the scene more clearly from the magnified local blocks.

4.3. Ablation study

To verify the effectiveness of our joint optimization algorithm, we show the quantitative results of our method with and without depth map optimization. In Fig. 13, we can see that the method with depth map optimization (i.e. full framework) achieves more reasonable relighting compared to approach without depth map optimization. From the lighting detail images (c) or (e), the former obtains better relighting effects including reasonable lighting and clear texture details, while the latter produces unnatural phenomena such as texture disorder and artifacts. Through experimental analysis, we can see that the latter phenomenon is attributed to texture copying from the images and coarse depth map defects.

4.4. User study

To further evaluate the quality of relighting images produced by our method, we follow the method (Li et al., 2018) to conduct a perceptual study to further evaluate our method. We select 60 indoor scene lighting images for studying. Thirty images are from the real-world and the others are generated by our method. The resolution of all images is set to 640×480.

Then we recruited 60 participants from a school campus, including professional image processing researchers, Photoshop users, and senior art scholars. We divide each image into three visual levels: (1) Real: realistic illumination image, (2) Fake: unrealistic relighting image with artifacts, (3) Uncertain: uncertain image which they cannot make a decision, and ask them to give their judgment results. The results of user studies from different methods are shown in Table 1. As we can see, for our work, 63.2% of real images are judged to be real images and the other real images are judged to be fake image or uncertain image. The methods FSV, IRSI have 52.7%, 55.4% of real images are judged to be real images, respectively. Besides, 54.1% of the synthesis images from our method are judged to be real image, which has almost the same judgment result as the real images. This further proves that our method has good performance in processing relighting task.

Table 1
Perceptual study results from different methods.

Method	Data type	Real	Fake	Uncertain
FSV	Real/Composite	0.527/0.496	0.312/0.375	0.161/0.129
IRSI	Real/Composite	0.554/0.501	0.273/0.324	0.173/0.175
Ours	Real/Composite	0.632/0.531	0.251/0.326	0.117/0.133

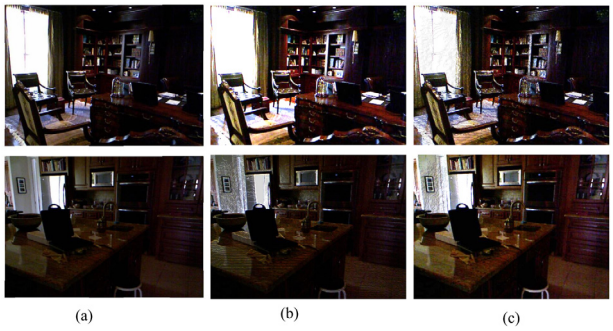


Fig. 14. Visual failure cases: input source images (a), output the relighting images (b) and (c).

4.5. Limitations

Our method takes RGB-D images as input, however, many existing low-light indoor images with more complex scenes are difficult to obtain depth information. Even current deep learning methods cannot accurately estimate the corresponding depth information, which will bring challenges for relighting. Our algorithm produces failed results as shown in Fig. 14. It can be seen that the relighted images (b) and (c) obtain failed effects with messy textures and border artifacts. We believe that we can directly use end-to-end deep learning methods to achieve better relighting for complicated indoor scenes in the future.

5. Conclusion

In this paper, we have proposed an interactive lighting editing algorithm that relights the indoor complicated scene under low-light conditions. It not only achieved global relighting for the indoor scenes, but also enhances the local details of complicated scenes requiring only a single RGB image. To this end, we first proposed a joint optimization framework to optimize the input depth image to avoid texture-copy phenomenon, and produce a fine depth map to achieve the intrinsic decomposition of the RGB image. Then we proposed a lighting estimation algorithm based on spherical harmonic function and combined with the fine depth map to estimate the lighting information of the shading map from decomposition.

Acknowledgment

This work is partially supported by NSFC (No.61972298) and Bingtuan Science and Technology Program (No.2019BC008).

CRediT authorship contribution statement

Zhongyun Bao: Conceptualization, Methodology, Software, Writing-Original draft preparation, Visualization, Investigation. **Gang Fu:** Data curation, Writing- Reviewing and Editing. **Lian Duan:** Software, Visualization, Investigation. **Chunxia Xiao:** Supervision, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bao, Z., Long, C., Fu, G., Liu, D., Li, Y., Wu, J., Xiao, C., 2022. Deep image-based illumination harmonization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 18542–18551.

Böhme, M., Haker, M., Martinetz, T., Barth, E., 2010. Shading constraint improves accuracy of time-of-flight measurements. Comput. Vis. Image Underst. 114 (12), 1329–1335.

Boyadzhiev, I., Paris, S., Bala, K., 2013. User-assisted image compositing for photographic lighting. Acm Transactions on Graphics 32 (4CD), 1–12.

Cao, X., Shah, M., 2005. Camera calibration and light source estimation from images with shadows. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'05, Vol. 2, IEEE, pp. 918–923.

Chandraker, M., Ramamoorthi, R., 2011. What an image reveals about material reflectance. In: 2011 International Conference on Computer Vision. IEEE, pp. 1076–1083.

Fan, Q., Yang, J., Hua, G., Chen, B., Wipf, D., 2017. A generic deep architecture for single image reflection removal and image smoothing. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3238–3247.

Fu, G., Zhang, Q., Zhu, L., Li, P., Xiao, C., 2021. A multi-task network for joint specular highlight detection and removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 7752–7761.

Garon, M., Sunkavalli, K., Hadap, S., Carr, N., Lalonde, J.F., 2019. Fast spatially-varying indoor lighting estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6908–6917.

Green, R., 2003. Spherical harmonic lighting: The gritty details. In: Archives of the Game Developers Conference. Vol. 56, p. 4.

Greene, N., 1986. Environment mapping and other applications of world projections. IEEE Comput. Graph. Appl. 6 (11), 21–29.

Gui, Y., Zeng, G., 2019. Joint learning of visual and spatial features for edit propagation from a single image. Vis. Comput.

Guo, X., Li, Y., Ling, H., 2016. LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26 (2), 982–993.

Haouchine, N., Roy, F., Courtecuisse, H., Niener, M., Cotin, S., 2017. Calipso: Physics-based image and video editing through CAD model proxies. Vis. Comput. (3).

Hara, K., Nishino, K., Ikeuchi, K., 2007. Mixture of spherical distributions for single-view relighting. IEEE Trans. Pattern Anal. Mach. Intell. 30 (1), 25–35.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

Kadambi, A., Taamazyan, V., Shi, B., Raskar, R., 2015. Polarized 3d: High-quality depth sensing with polarization cues. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3370–3378.

Lagger, P., Fua, P., 2006. Using specularities to recover multiple light sources in the presence of texture. In: 18th International Conference on Pattern Recognition. ICPR'06, Vol. 1, IEEE, pp. 587–590.

Landis, H., 2002. Global illumination in production. ACM SIGGRAPH 2002 Course #16 Notes.

Lettry, L., Vanhoey, K., Van Gool, L., 2018. DARN: a deep adversarial residual network for intrinsic image decomposition. In: 2018 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 1359–1367.

Li, Y., Brown, M.S., 2014. Single image layer separation using relative smoothness. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2752–2759.

Li, C., Zhou, K., Wu, H.T., Lin, S., 2018. Physically-based simulation of cosmetics via intrinsic image decomposition with facial priors. IEEE Trans. Pattern Anal. Mach. Intell. 1.

Maier, R., Kim, K., Cremers, D., Kautz, J., Nießner, M., 2017. Intrinsic3d: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3114–3122.

Or-El, R., Rosman, G., Wetzler, A., Kimmel, R., Bruckstein, A.M., 2015. Rgbd-fusion: Real-time high precision depth recovery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5407–5416.

Ren, P., Dong, Y., Lin, S., Tong, X., Guo, B., 2015. Image based relighting using neural networks. ACM Trans. Graph. 34 (4), 111.

Rother, C., Kiefel, M., Zhang, L., Schölkopf, B., Gehler, P.V., 2011. Recovering intrinsic images with a global sparsity prior on reflectance. In: Advances in Neural Information Processing Systems. pp. 765–773.

Shen, L., Yeo, C., 2011. Intrinsic images decomposition using a local and global sparse representation of reflectance. In: CVPR 2011. IEEE, pp. 697–704.

Shen, L., Yeo, C., Hua, B.S., 2013. Intrinsic image decomposition using a sparse representation of reflectance. IEEE Trans. Pattern Anal. Mach. Intell. 35 (12), 2904–2915.

Sun, Y., Miao, Y., Chen, J., Pajarola, R., 2020. PGCNet: Patch graph convolutional network for point cloud segmentation of indoor scenes. Vis. Comput. 36 (10).

Tappen, M.F., Adelson, E.H., Freeman, W.T., 2006. Estimating intrinsic component images using non-linear regression. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'06, Vol. 2, IEEE, pp. 1992–1999.

Tappen, M.F., Freeman, W.T., Adelson, E.H., 2003. Recovering intrinsic images from a single image. In: Advances in Neural Information Processing Systems. pp. 1367–1374.

Tunwattanapong, B., Ghosh, A., Debevec, P., 2011. Practical image-based relighting and editing with spherical-harmonics and local lights. In: 2011 Conference for Visual Media Production. IEEE, pp. 138–147.

Wu, J.H., Saito, S., 2017. Interactive relighting in single low-dynamic range images. ACM Trans. Graph. 36 (2), 18.

Wu, C., Zollhöfer, M., Nießner, M., Stamminger, M., Izadi, S., Theobalt, C., 2014. Real-time shading-based refinement for consumer depth cameras. ACM Trans. Graph. (ToG) 33 (6), 200.

Xu, Z., Sunkavalli, K., Hadap, S., Ramamoorthi, R., 2018. Deep image-based relighting from optimal sparse samples. ACM Trans. Graph. 37 (4), 126.

Xu, L., Yan, Q., Xia, Y., Jia, J., 2012. Structure extraction from texture via relative total variation. ACM Trans. Graph. 31 (6), 139.

Yu, L.F., Yeung, S.K., Tai, Y.W., Lin, S., 2013. Shading-based shape refinement of rgb-d images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1415–1422.

Zhang, L., Zhang, Q., Xiao, C., 2015. Shadow remover: Image shadow removal based on illumination recovering optimization. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 24 (11), 4623–4636.

Zollhöfer, M., Dai, A., Innmann, M., Wu, C., Stamminger, M., Theobalt, C., Nießner, M., 2015. Shading-based refinement on volumetric signed distance functions. ACM Trans. Graph. 34 (4), 96.