論文 / 著書情報
Article / Book Information

Title	Interactive Relighting in Single Low-Dynamic Range Images
Authors	Jung-Hsuan Wu, Suguru Saito
Citation	ACM Transactions on Graphics, Vol. 36, No. 2, Article 18
Pub. date	2017, 4
Note	(C) ACM 2017. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in ACM Transactions on Graphics, http://dx.doi.org/10.1145/3034185.

Interactive Relighting in Single Low-Dynamic Range Images[1]

JUNG-HSUAN WU and SUGURU SAITO
Tokyo Institute of Technology

(a) (b) (c) (d)

Fig. 1: Users can freely edit the illumination in low-dynamic range (LDR) photographs and obtain realistic relit results with our method. (a) Input LDR image. (b) Result of relighting by moving existing light sources to back of the altar. (c)(d) Results in which the illumination seems to be occluded by chairs or windows are produced by inserting multiple light sources and placing them at different positions.

This paper[1] addresses the relighting of outdoor and large indoor scenes illuminated by non-distant lights, which has seldom been discussed in previous works. We propose a method for users to interactively edit the illumination of a scene by moving existing lights and inserting synthetic lights into the scene that requires only a small amount of user annotation and a single low-dynamic range (LDR) image. We achieve this by adopting a top-down approach that estimates the scene reflectance by fitting a diffuse illumination model to a photograph. This approach gains stability and robustness by estimating the camera, scene geometry, and light sources in sequence and by using a confidence map, which is a per-pixel weight map. The results of our evaluation demonstrate that the proposed method can estimate a scene accurately enough for realistic relighting of images. Moreover, the experimental results of our user studies show that the synthesized images are so realistic as to be almost indistinguishable from real photographs.

Categories and Subject Descriptors: I.4.1 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual Reality*; I.4.8 [**Computer Graphics**]: Scene Analysis—*Depth Cues and Shading*

Additional Key Words and Phrases: image-based relighting, image-based modeling

1. INTRODUCTION

Relighting of a single image is required by many applications pertaining to movie and photo editing, but it is difficult because the scene must be accurately estimated in advance. Many approaches have been developed to estimate outdoor scenes illuminated by a distant light source (i.e., the sun) that emits parallel light rays. In contrast to these methods, in this work we focus on outdoor and large indoor scenes illuminated by non-distant light sources, which is challenging because the light rays are nonparallel.

Most image-based relighting techniques require manually created scene geometries, multiple photographs, or special equipment. Our goal is to develop an interactive illumination manipulation method that produces relit results that might not be physically correct but are visually consistent and plausible, requiring only a small amount of user annotations and a single LDR image. To this end, we have developed a method that estimates and reproduces the luminance distribution from a single LDR image by adopting simplified models of camera, light source, scene reflectance, and 3D geometry.

In our method, first, the user is asked to specify a region that contains the illuminated bright light spot and to trace the boundaries between the walls and the ground. Th method then estimates the luminance distribution of a 3D scene by finding the parameters of camera, light source, and scene reflectance that reproduce the luminance distribution in the image. A pseudo 3D scene is then created with the found parameters and the user can change the lighting in the image, after which the results are synthesized in real-time (Fig. 1).

The primary technical contribution of this paper is our proposal of a top-down approach that stably and robustly estimates the luminance distribution and recovers the scene reflectance from one LDR image. This approach utilizes a physical-based model to describe the illumination in a photograph with several parameters and it estimates the parameters in sequence so that the estimation gains stability. It also adopts a confidence map, which is a per-pixel weight map that alleviates the effect of occlusions, shadows, overlapping, and highlight-clipping in LDR images by assigning low weight to the pixels influenced by these phenomena so that the estimation gains robustness. Our method is applicable to the removal of bright light spots, which are produced by non-distant light sources, from LDR images and works best for images of scenes in which the walls are perpendicular to the ground and the walls and ground are Lambertian and plane-like surfaces.

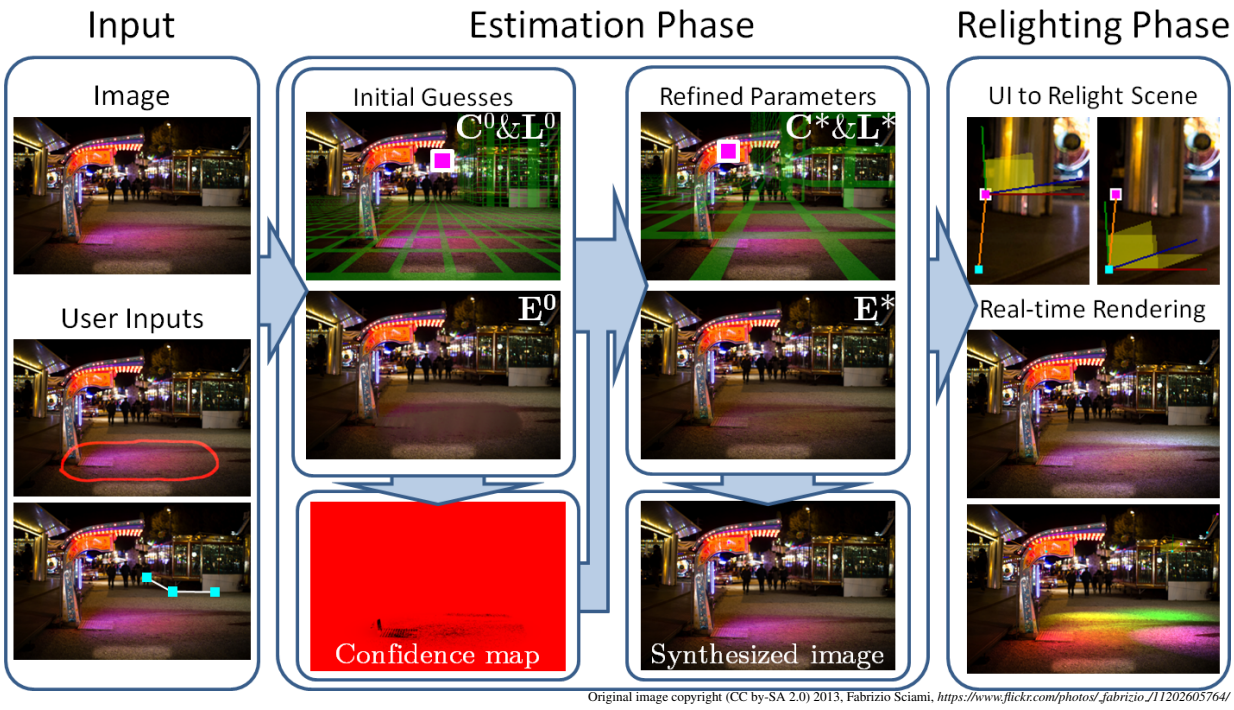


Fig. 2: Our method for relighting a scene from a single image. Given an input image and user annotations of the light spot regions as well as the edges between the walls and ground, our method estimates and reconstructs the scene. It only takes a few minutes to finish annotating and less than two minutes for computation. Users can then interactively adjust the lights to relight the scene.

2. RELATED WORK

Light Source and Illumination Estimation

Many methods have been proposed to estimate light source, as it is an important issue in the field of computer graphics and computer vision. Cao and Shah [2005] proposed a method to recover the direction of a light source by using the relationship between the occluder and shadow in an image. More robust methods [Hara et al. 2008; Liu et al. 2009; Lopez-Moreno et al. 2010; Chandraker and Ramamoorthi 2011; Lalonde et al. 2012] improved the estimation by taking the sky, ground, or coarse 3D geometries into consideration. However, the assumption utilized in these methods, namely, that the illumination was produced by one distant light source (i.e., the sun), prevents them from being applied to most indoor and nighttime outdoor scenes.

For non-distant light sources, specular reflections are explicit clues to the estimation of the light source [Hara et al. 2005; Lagger and Fua 2006; 2008]. However, specular reflections are highly sensitive to the normal of object surfaces and thus it is difficult to estimate scenes with complex geometries, e.g., grass, from specular reflections. Diffuse reflections, while also providing clues to help estimate light sources [Lopez-Moreno et al. 2013], change much more smoothly than specular reflections. Therefore, we chose to estimate the light source by using diffuse reflections rather than specular ones.

Other methods [Zhou and Kambhamettu 2002; Matusik et al. 2003; Zickler et al. 2006; Xu and Wallace 2008; Tunwattanapong et al. 2011; Lombardi and Nishino 2012; Ren et al. 2015] require multiple input images or special equipment while capturing the image to collect the data required to estimate illumination. Our method differs from these in that it only needs a single LDR image and a small amount of user annotations and no access to geometry or any devices other than a digital camera.

For readers interested in state-of-the-art techniques of illumination estimation, please read the paper presented by Kronander et al. [2015]. This survey paper goes over the pros and cons of many methods in detail.

Recovery of Surface Reflectance

There have been many works [Shen et al. 2008; Hsu et al. 2008; Bousseau et al. 2009; Dong et al. 2011; Carroll et al. 2011; Garces et al. 2012] proposed to recover the reflectance of surfaces from an image. However, these methods do not take the 3D geometries into consideration, so the illumination is not changeable. Yu et al. [1999], Boivin and Gagalowicz [**?**], Hachama et al. [2015], and Shi et al. [2015] proposed methods to recover the reflectance from real photographs with pre-knowledge of real scene geometry, and the method proposed by Karsch et al. [2011] utilizes a semi-automatically created pseudo 3D geometry to improve the recovery of albedo. These methods assume that the illumination changes smoothly, which is usually not true for outdoor scenes illuminated by non-distant area light. In contrast, our illumination model allows sudden changes in illumination and is applicable to high-contrast lighting.

Image-based Modeling

Horn [1989] proposed a technique of shape-from-shading that estimates the surface normal of objects from photographs in which a Lambertian object is illuminated by light coming from different directions. Many approaches [Barron and Malik 2012; Li et al. 2013; Zoran et al. 2014] were subsequently presented to improve the shape-from-shading techniques. In contrast to these bottom-up

approaches, which recover the 3D shape of objects by the relationships between nearby pixels, our top-down approach first defines a physically based illumination model and then identifies the parameters that reproduce the luminance distribution in the input photograph.

User assistance is also applicable to recover the 3D scene geometry from a single image [Criminisi et al. 2000; Oh et al. 2001; El-hakim et al. 2005; Colburn et al. 2013]. A simple but efficient method to construct the geometry of a scene was developed by Iizuka et al. [2011]. However, their method assumes that the focal length of the camera is known. This assumption makes it difficult for users who are not familiar with cameras to use their method. Zheng et al. [2012] proposed a method that generates a scene geometry by minimizing the difference between user annotations and the synthesized 3D geometry. In contrast, our method estimates scene geometries and the focal length of the camera by minimizing the error in diffuse illumination with limited user input.

3. OVERVIEW

In this paper, we propose a method that estimates the luminance distribution in an image by finding the parameters of camera, albedo, and light source that minimize the pixel-wise difference between the luminance distribution in the synthesized image and that in an input photograph with user annotations that indicate the region of the illuminated bright light spots as well as the boundaries between the ground and walls, as shown in Fig. 2. After the parameters are estimated, the user is allowed to manipulate the light source to relight the scene.

We assume that the scene in a photograph is Lambertian, that the 3D geometry consists of one horizontal flat ground and zero or more planar walls perpendicular to the ground, that the camera is an ideal pinhole camera located at the origin with pitch rotation only, and that the scene is illuminated by one or more non-distant area light sources. With these assumptions, the illumination in the photograph is determined with the parameters of camera, albedo, and light source.

With the assumptions of 3D geometry and camera, it is sufficient to reconstruct the scene's 3D geometry with the camera parameters (focal length and pitch angle) and the user annotations [Hartley and Zisserman 2004; Jiang et al. 2009]. Let $\mathbf{S}(\mathrm{C})$, L, $\boldsymbol{\rho}$, and L_E denote the 3D geometry created with camera parameters C, the light source, the albedo, and the environment light, respectively. The synthetic color image, $\mathbf{I}'$, is composed of the diffuse illumination color, $\mathbf{D}$, and the environment illumination color, $\mathbf{E}$, by

$$\mathbf{I}'\big(\mathbf{S}(\mathrm{C}),\mathrm{L},\boldsymbol{\rho},\mathrm{L}_E\big)=\mathbf{D}\big(\mathbf{S}(\mathrm{C}),\mathrm{L},\boldsymbol{\rho}\big)+\mathbf{E}\big(\boldsymbol{\rho},\mathrm{L}_E\big), \tag{1}$$

where

$$\mathbf{D}\big(\mathbf{S}(\mathrm{C}),\mathrm{L},\boldsymbol{\rho}\big)=\boldsymbol{\rho}*\mathrm{diff}\big(\mathbf{S}(\mathrm{C}),\mathrm{L}\big),\quad \mathbf{E}\big(\boldsymbol{\rho},\mathrm{L}_E\big)=\mathrm{L}_E\,\boldsymbol{\rho},$$

$\mathrm{diff}(\mathbf{S}(\mathrm{C}),\mathrm{L})$ returns the incident diffuse light emitted from L to the scene whose 3D geometry is $\mathbf{S}(\mathrm{C})$, and $*$ denotes the per-pixel multiplication.

The basic concept underlying the proposed method is that the unknowns (C, L, $\boldsymbol{\rho}$, and L_E) are solved by minimizing the overall difference within the user-specified region for light spot, which is denoted by $\mathbb{R}$, between $\mathbf{I}'$ and the input image $\mathbf{I}$, that is,

$$[\mathrm{C}^*,\mathrm{L}^*,\boldsymbol{\rho}^*,\mathrm{L}_E^*]=\underset{\mathrm{C},\mathrm{L},\boldsymbol{\rho},\mathrm{L}_E}{\arg\min}\sum_{p\in\mathbb{R}}\Big(\boldsymbol{\alpha}_p\big(\mathbf{I}'_p(\mathbf{S}(\mathrm{C}),\mathrm{L},\boldsymbol{\rho},\mathrm{L}_E)-\mathbf{I}_p\big)\Big)^2, \tag{2}$$

where $\boldsymbol{\alpha}$ denotes the confidence map, a per-pixel weight that has low weight for pixels located in the shadow, the region overlapped with other light spots, or having a highlight-clipping effect due to the limited displayable range of the LDR image, and high weight otherwise. The confidence map contributes to the robustness of the estimation. The bold symbols $\boldsymbol{\alpha}$, $\mathbf{I}'$, $\mathbf{S}(\mathrm{C})$, $\boldsymbol{\rho}$, $\mathbf{I}$, $\mathbf{D}$, and $\mathbf{E}$ denote per-pixel parameters, the non-bold uppercase symbols C, L, and L_E are non-pixel-wise vector parameters, and the symbols with a superscripted star denote the parameters obtained through the estimation to make the equations easier to understand.

The unknown parameters are estimated by minimizing the objective function in Eq. 2. However, there are three problems that make this estimation difficult.

The first problem is the ambiguity between L, L_E, and $\boldsymbol{\rho}$. The same image is synthesized if the intensity of L and L_E is multiplied by a certain number n and $\boldsymbol{\rho}$ is divided by n, that is, $\mathbf{I}'(\mathbf{S}(\mathrm{C}),\mathrm{L},\boldsymbol{\rho},\mathrm{L}_E)=\mathbf{I}'(\mathbf{S}(\mathrm{C}),n\mathrm{L},\boldsymbol{\rho}/n,n\mathrm{L}_E)$. Instead of decomposing $\mathbf{E}$ into L_E and $\boldsymbol{\rho}$, we rewrite Eqs. 1 and 2 as

$$\mathbf{I}'(\mathbf{S}(\mathrm{C}),\mathrm{L}',\mathbf{E})=\mathbf{D}(\mathbf{S}(\mathrm{C}),\mathrm{L}',\mathbf{E})+\mathbf{E},\ \mathbf{D}=\mathbf{E}*\mathrm{diff}(\mathbf{S}(\mathrm{C}),\mathrm{L}') \tag{3}$$

and

$$[\mathrm{C}^*,\mathrm{L}'^*,\mathbf{E}^*]=\underset{\mathrm{C},\mathrm{L}',\mathbf{E}}{\arg\min}\sum_{p\in\mathbb{R}}\Big(\boldsymbol{\alpha}_p\big(\mathbf{I}'_p(\mathbf{S}(\mathrm{C}),\mathrm{L}',\mathbf{E})-\mathbf{I}_p\big)\Big)^2, \tag{4}$$

respectively, where $\mathrm{L}'=\mathrm{L}/\mathrm{L}_E$. Since $\boldsymbol{\rho}$ and L_E do not appear in Eq. 4, the ambiguity is resolved.

The second problem is the large number of unknown variables that $\mathbf{E}$ contains, since it consists of the environment illumination colors of each pixel within $\mathbb{R}$. Estimating the environment illumination $\mathbf{E}_p$ for each pixel p in $\mathbb{R}$ is very time-consuming and difficult because there are so many pixels (usually up to tens of thousands) and there is no pre-knowledge of the lighting coming from the surrounding environment.

Instead of estimating $\mathbf{E}_p$ for each pixel p by minimizing the error in Eq. 4, we choose to iteratively estimate C and L' with the initial guess, $\mathbf{E}^0$, and then directly recover $\mathbf{E}_p^*$ with the estimated $\mathrm{C}^*$and L'^*. We took this approach because our method of generating the initial guess, which is detailed in the supplementary document, can create $\mathbf{E}^0$ well enough and the confidence map $\boldsymbol{\alpha}$ helps to reduce the influence of regions that are not recovered in $\mathbf{E}^0$. Furthermore, there are far fewer errors produced by $\mathbf{E}_p$ than by C and L' in Eq. 2, since C and L' influence the diffuse illumination over the whole user-specified light spot region and $\mathbf{E}_p$ only affects one pixel. Therefore, it is possible to estimate C and L' with $\mathbf{E}^0$.

The third problem is the instability of the optimization. $\mathbf{I}'$ is the result of two projections (from the light to the 3D scene and from the 3D scene to the screen), so there exist local minima and the optimization is unstable.

To achieve a stable optimization, we introduce a constraint that the optical axis of the light source follows the camera and light position to reduce local minima. This constraint is expressed by the direction adjustment function $\mathrm{T}(\cdot)$, which adjusts the optical axis of L' in the 3D scene space to make it always direct to the corresponding position in the 3D scene of a fixed location in the 2D image (as discussed in Section 6.2), which guarantees that Eq. 3 synthesizes the light spot at the same location in the 2D image. With the direction adjustment function $\mathrm{T}(\cdot)$, the estimation of C and L' by Eq. 4 becomes stable.

With the solutions above, the unknown parameters can be estimated by Eq. 4. Our optimization starts from initial guesses C^0, L'^0, and $\mathbf{E}^0$, and then iteratively refines C and L'. The refining of C and L' is separated into two steps because C is a global parameter that affects the appearance in a 2D image of all synthetic light spots and L' only controls the appearance of one light spot. The environment

Fig. 3: Two common patterns of light spots: oval-shaped light spot (left) and comet-shaped (right).

illumination $\mathbf{E}^*$ is recovered by removing the diffuse illumination produced by the estimated $\mathrm{C}^*$and L'^*.

The user interfaces to assign user inputs and relight the scene are presented in Section 4 of this paper. The method that generates the confidence map $\boldsymbol{\alpha}$ is introduced in Section 5, and we explain the estimation and relighting processes in Section 6. Sections 7 and 8 evaluate the proposed method by experiments and user studies, respectively. Section 9 demonstrates and discusses the results of this paper, and we conclude in Section 10 with a brief summary and mention of future work.

4. USER INTERFACE

Annotation for estimates

Because it is extremely difficult and expensive to automatically detect the light spot areas in an image, the user is asked to specify them. The user specifies light spots by circling them one-by-one (Fig. 2) and then annotates them by type in accordance with their shape (oval-shaped or comet-shaped, Fig. 3). The different types of light spots are estimated by the same process but different initial guesses are used. The user also inputs the lines along the boundaries between the walls and the ground, as shown in Fig. 2.

Manipulation for relighting

The user is allowed to move and rotate the light sources intuitively, as shown in the relighting phase of Fig. 2(top-right). The red, green, and blue lines correspond to the x-, y-, and z-axes, and the yellow quadrilaterals indicate the xy-, xz-, and yz-planes. Users move and rotate the light sources by dragging on these lines and planes. The other properties of the light source (e.g., color and beam angle) are controlled by slide bars in another control panel.

5. CONFIDENCE MAP

The light spots in photographs may be partially occluded or overlapped with other light spots. There is also a highlight-clipping effect at pixels whose intensity is higher than the storable range of LDR images. These may significantly reduce the estimation accuracy, so the confidence map $\boldsymbol{\alpha}$ is proposed to alleviate their influence.

The idea is very straightforward. As shown in Fig. 4, an approximation of the luminance distribution of a light spot in an image, $\mathbf{M}$, is defined and high weight is assigned to pixels whose intensity is close to $\mathbf{M}$; low weight otherwise.

As presented by Wang et al. [2007], we define $\mathbf{M}$ with a 2D Gaussian function G_l with a peak set to 1, a scaling term ω, and a shifting term m.

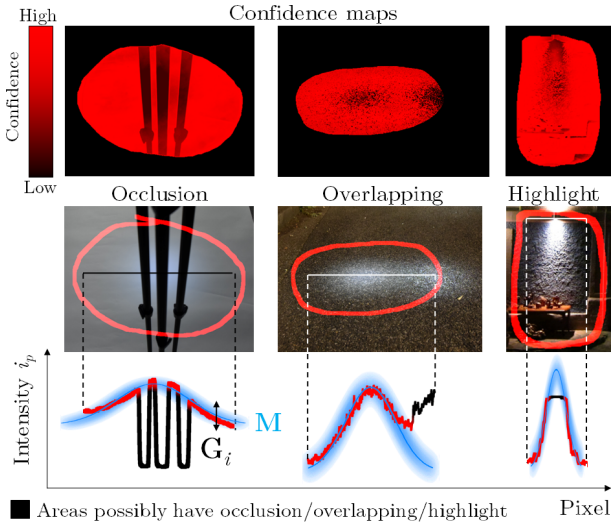


Fig. 4: The confidence maps (top) and input images with user annotation (middle). The bottom row shows how the confidence map works: $\mathbf{M}$ (blue curves) is the approximation of the luminance distribution (red and black curves) of the light spot in an input image. Pixels whose intensity is largely different from $\mathbf{M}$ (black regions) have low confidence; otherwise, high confidence.

$$\mathbf{M}_p = \omega\, \mathrm{G}_l(p') + m, \tag{5}$$

The physical interpretation of G_l, ω, and m are the approximation of the Lambert's cosine function, the intensity of the light source, and the environment illumination, respectively. p' is the transformed position of pixel p after rotation centered at μ, which is the centroid of the pixels in $\mathbb{R}$. This rotation is adopted to align the orientation of G_l to that of the light spot in an image. The orientation of the light spot is analyzed by principal component analysis (PCA), and the position of pixel p is transformed by

$$p' = \mathrm{R}_l \cdot (p - \mu) + \mu, \tag{6}$$

where

$$\mu = \frac{\sum_{p\in\mathbb{R}} i_p\, p}{\sum_{p\in\mathbb{R}} i_p}, \quad i_p = \max\Big(\mathrm{int}\big(\mathbf{I}_p\big) - \mathrm{int}\big(\mathbf{E}_p^0\big), 0\Big),$$

R_l denotes the rotation matrix that rotates the x-axis to the first principal component, and $\mathrm{int}(\cdot)$ returns the intensity of the given rgb color.

The mean of the 2D Gaussian function G_l is set to μ. The two standard deviations σ_{lx} and σ_{ly} of G_l, ω, and m are obtained by minimizing the difference between $\mathbf{M}$ and the input photograph, that is,

$$[\sigma_{lx}, \sigma_{ly}, \omega, m] = \underset{\sigma_{lx},\sigma_{ly},\omega,m}{\arg\min} \sum_{p\in\mathbb{R}} \boldsymbol{\lambda}_p \Big(\mathrm{int}\big(\mathbf{I}_p\big) - \mathbf{M}_p\Big)^2, \tag{7}$$

where $\boldsymbol{\lambda}_p$ is the weight for preventing the influence of shadows and highlight-clipping effect, and is defined as

$$\boldsymbol{\lambda}_p = \begin{cases} 0, & \text{if } \mathrm{int}\big(\mathbf{I}_p\big) \geq \mathrm{threshold}_i \\ i_p, & \text{otherwise.} \end{cases} \tag{8}$$

Here, $\mathrm{threshold}_i$ is set to 254/255 while the intensity of the image is from 0 to 1. In our implementation, Eq. 7 is solved by

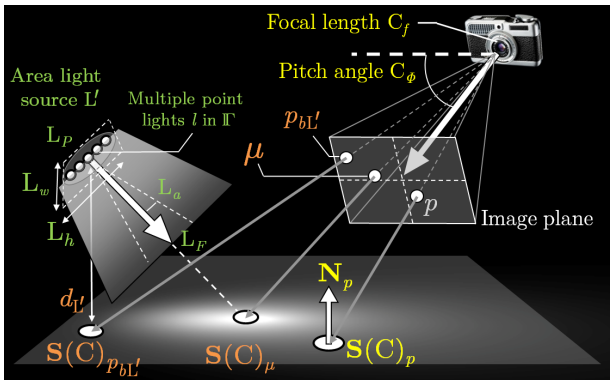


Fig. 5: Parameters necessary for computing the color of a pixel p. Parameters of light source are colored green and parameters of camera and 3D geometry are yellow. Variables of orange are used to generate the initial guess of the light source.

Table I. : Parameters of camera and light source. Symbols with uppercase subscript denote vectors of scalar and those with lowercase subscript are scalars.

	Symbol	Description
C	C_f	The focal length of the camera scaled by a world-to-pixel factor
	C_ϕ	The pitch angle of the camera, which is expressed in radian
L′	L_P	The center of the area light source
	L_F	The optical axis of the light source
	L_a	The beam angle of the light beam
	L_w, L_h	The width and height of the area light source
	L_k	The decline rate of luminance from optical axis to slant
	L'_C	The color ratio between the light emitted from light source and environment light

JCOBYLA [2012], which is the Java implementation of Powell's method of non-linear derivative-free constrained optimization [Powell 1964].

Another 1D Gaussian function G_i is also adopted to model the effects of the texture and bumpiness of the surface as well as to evaluate the influence of the occlusion, overlapping with the other light spots, and highlight-clipping effect. The mean μ_i and variance σ_i of G_i are computed by

$$\mu_i = \frac{1}{\|\mathbb{R}\|}\sum_{p\in\mathbb{R}}\left(\text{int}(\mathbf{I}_p) - \mathbf{M}_p\right), \text{ and}$$
$$\sigma_i = \sqrt{\frac{1}{\|\mathbb{R}\|}\sum_{p\in\mathbb{R}}\left(\left(\text{int}(\mathbf{I}_p) - \mathbf{M}_p\right) - \mu_i\right)^2}. \quad (9)$$

Ideally, μ_i should be zero if the initial guess of environment illumination, $\mathbf{E}^0$, which is used in Eq. 6, is perfectly recovered as an unlit image. However, in our experiments, $\mathbf{E}^0$ is usually different from the ground truth, so non-zero μ_i helps fit the luminance distribution better.

With $\mathbf{M}$ and G_i, the confidence of pixel p is defined as

$$\boldsymbol{\alpha}_p = G_i\Big(\text{int}(\mathbf{I}_p) - \mathbf{M}_p\Big). \quad (10)$$

Our method gains robustness and high estimation accuracy with the confidence map $\boldsymbol{\alpha}$ as shown in Section 7.

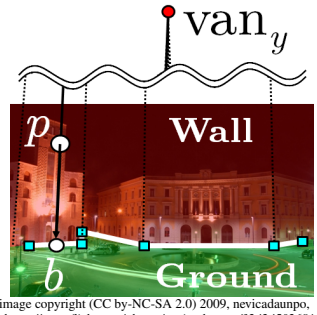


Fig. 6: The x-coordinate and z-coordinate of the position in the 3D scene of a pixel p that is on the wall are the same as that of a pixel b, whose corresponding position in the 3D scene, $\mathbf{S}(C)_b$, is the point on the ground and is right below the corresponding position in the 3D scene of pixel p, $\mathbf{S}(C)_p$. Pixel b is the intersection of the user-annotated walls' boundaries (white line segments and cyan squares) and the straight line that passes through the y-axis vanishing point and p (black arrow).

6. ESTIMATION AND RELIGHTING

Our method estimates the camera and light source parameters by generating initial guesses and then refines them in sequence, and the environment illumination is then recovered by removing the diffuse illumination produced with the estimated camera and light source. Table I lists the parameters of the camera C and the light source L′, which are also illustrated in Fig. 5. The symbols with uppercase subscript denote vectors and those with lowercase subscript are scalars. In this section, the model of the 3D geometry and diffuse illumination are introduced in Section 6.1 and the direction adjustment function $T(\cdot)$ is defined in Section 6.2. The methods that refine the parameters of camera and light source are explained in Section 6.3. Finally, the method to recover the environment illumination and generate a relit image is presented in Section 6.4.

6.1 Model of 3D Geometry and Diffuse Illumination

3D Geometry

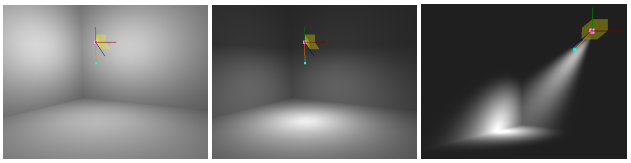
By assuming that the ground is the plane of $y = -1$, along with assumptions for the camera and 3D geometry mentioned in Section 3, the corresponding position of p in a 3D scene is given by

$$\mathbf{S}(C)_p = \begin{cases} \left(\frac{-p_x}{p_y\mathfrak{c} + C_f\mathfrak{s}}, -1, \frac{p_y\mathfrak{s} - C_f\mathfrak{c}}{p_y\mathfrak{c} + C_f\mathfrak{s}}\right)^T & \text{, if } p \in \text{ ground} \\ \left(X_b, -Z_b\frac{p_y\mathfrak{c} + C_f\mathfrak{s}}{p_y\mathfrak{s} - C_f\mathfrak{c}}, Z_b\right)^T & \text{, if } p \in \text{ wall} \end{cases} \quad (11)$$

where (p_x, p_y) denote the coordinates of a pixel p in a 2D image, $\mathfrak{s}$ and $\mathfrak{c}$ are $\sin(C_\phi)$ and $\cos(C_\phi)$, respectively, and X_b and Z_b are the x- and z-coordinates of the position in a 3D scene corresponding to the pixel b, which is the intersection of the user-specified boundaries between the ground and wall and the straight line that passes through the y-axis vanishing point and the pixel p, as illustrated in Fig. 6. The supplementary document explains how to derive Eq. 11.

The normal of a pixel, $\mathbf{N}_p$, in the 3D scene is obtained by normalizing the cross product of the tangent vectors at the surface, which are computed by the gradients of $\mathbf{S}(C)_p$ along the x- and y-axis of the image.

Diffuse Illumination



(a) (b) (c)

Fig. 7: The illuminations synthesized with Eq. 12. (a) Room illuminated by an omni-directional light source such as a standing lamp. (b) Room illuminated by a fluorescent lamp-like light source. (c) Corner of the room illuminated by a spotlight.

The diffuse illumination is assumed to be produced by an area light source, which is approximated by a set of point light sources radially arrayed on an elliptic disk centered at L_P, which is perpendicular to the optical axis L_F and whose size is determined by L_w and L_h. The diffuse illumination of a pixel p is given by

$$\mathrm{diff}(\mathbf{S}(\mathrm{C})_p, \mathrm{L}') = \mathrm{L}'_C \sum_{\mathrm{l}\in\mathbb{\Gamma}} \frac{\mathrm{V}_{\mathrm{l}_P} \cdot \mathbf{N}_p}{\|\mathrm{V}_{\mathrm{l}_P}\|} \frac{1}{\|\mathrm{V}_{\mathrm{l}_P}\|^2} \mathrm{Beam}(\mathrm{V}_{\mathrm{l}_P}, \mathrm{L}_a, \mathrm{L}_k), \tag{12}$$

where $\mathrm{V}_{\mathrm{l}_P}$ denotes the vector from $\mathbf{S}(\mathrm{C})_p$ to a point light source l and is defined as

$$\mathrm{V}_{\mathrm{l}_P} = \mathrm{L}_P + \Delta\mathrm{l}_P - \mathbf{S}(\mathrm{C})_p,$$

$\mathbb{\Gamma}$ indicates the set of point lights that approximate the area light L', and $\Delta\mathrm{l}_P$ is the relative position of each point light l to L_P.

The first term in Eq. 12 follows Lambert's cosine function, and the second term represents the inverse-square law that the intensity decreases over distance due to the spread of light rays. The third term $\mathrm{Beam}(\cdot)$, which controls the decline of luminance from optical axis to slant in light cone, is defined as

$$\mathrm{Beam}(\mathrm{V}_{\mathrm{l}_P}, \mathrm{L}_a, \mathrm{L}_k) = \begin{cases} (1 - \frac{\theta}{\mathrm{L}_a})^{\mathrm{L}_k} & , \text{if } \theta < \mathrm{L}_a \\ 0 & , \text{otherwise} \end{cases}, \tag{13}$$

where θ denotes the angle between vectors $-\mathrm{V}_{\mathrm{l}_P}$ and L_F and is computed as

$$\theta = \cos^{-1}\left(\frac{-\mathrm{V}_{\mathrm{l}_P} \cdot \mathrm{L}_F}{\|\mathrm{V}_{\mathrm{l}_P}\|}\right).$$

Note that when $\mathrm{L}_k = 0$ and $\mathrm{L}_a \geq \pi$, the value of $\mathrm{Beam}(\cdot)$ is always one and Eq. 12 is equivalent to an omni-directional area light source.

This diffuse illumination model allows the synthesizing of diffuse illuminations produced by most kinds of light sources. Figure7(a) shows a room illuminated by an omni-directional light source such as a standing lamp, (b) shows the room illuminated by a fluorescent lamp-like light source, and in (c), a corner of the room is illuminated by a spotlight. In addition, it is also efficient for the estimation because there is only one solution that minimizes the objective function in Eq. 4. The proof of the uniqueness of the solution to our diffuse illumination model is explained in the supplementary document.

6.2 Constraint of Optical Axis

In order to estimate the parameters stably while avoiding local minima, we propose the direction adjustment function $\mathrm{T}(\cdot)$, which takes as input the camera parameters C, the centroid μ of the light spot in 2D image space, and the central position of the area light, L_P, and returns the optical axis L_F as

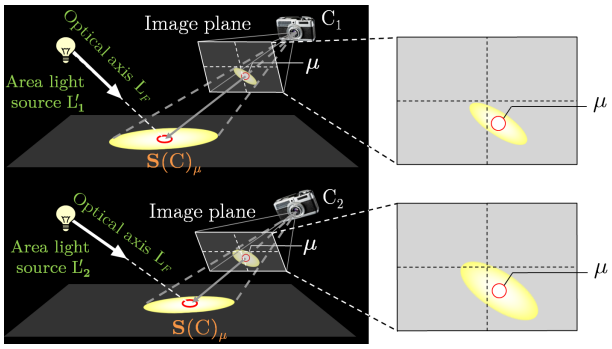


Fig. 8: The direction adjustment function $\mathrm{T}(\cdot)$ adjusts the optical axis in the 3D scene space of the light source L' to make it always emit light toward $\mathbf{S}(\mathrm{C})_\mu$ (left) so that the centroid of the light spot is synthesized at the same location in 2D image space (right).

$$\mathrm{L}_F = \mathrm{T}(\mathrm{C}, \mu, \mathrm{L}_P) = \frac{\mathbf{S}(\mathrm{C})_\mu - \mathrm{L}_P}{\|\mathbf{S}(\mathrm{C})_\mu - \mathrm{L}_P\|}, \tag{14}$$

Setting L_F to the vector from L_P to $\mathbf{S}(\mathrm{C})_\mu$ guarantees that L' emits light toward μ, so the centroid of the light spot is synthesized at the same location in 2D screen space, as shown in Fig. 8.

6.3 Estimation of Camera and Light Source

Let the symbols with the superscript k indicate the parameters refined after k iterations. The camera parameters $\mathrm{C} = \{\mathrm{C}_f, \mathrm{C}_\phi\}$ are refined by

$$[\mathrm{C}^{k+1}, p^{k+1}_{\mathrm{bL}'}, d^{k+1}_{\mathrm{L}'}] = \underset{\mathrm{C}, p_{\mathrm{bL}'}, d_{\mathrm{L}'}}{\arg\min} \sum_{p\in\mathbb{R}} \left(\boldsymbol{\alpha}_p\left(\mathbf{I}'_p(\mathbf{S}(\mathrm{C}), \mathrm{L}'^k_{\mathrm{T}(\mathrm{C},\mu,p_{\mathrm{bL}'},d_{\mathrm{L}'})}, \mathbf{E}^0) - \mathbf{I}_p\right)\right)^2,$$
$$\text{subject to } 0 < \mathrm{C}_f, -\tfrac{\pi}{2} \leq \mathrm{C}_\phi \leq \tfrac{\pi}{2}, 0 \leq d_{\mathrm{L}'} \tag{15}$$

where $\mathrm{L}'^k_{\mathrm{T}(\mathrm{C},\mu,p_{\mathrm{bL}'},d_{\mathrm{L}'})}$ is the light source whose L_P is computed by

$$\mathrm{L}_P = \mathbf{S}(\mathrm{C}^k)_{p^k_{\mathrm{bL}'}} + d^k_{\mathrm{L}'}\mathbf{N}_{p_{\mathrm{bL}'}}, \tag{16}$$

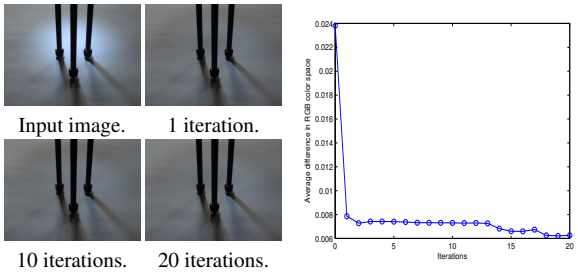
L_F is computed by Eq. 14, and the other parameters L'_C, L_a, L_w, L_h, and L_k are the same as those of L'^k. The supplementary document explains Eq. 16 in more detail.

After C^{k+1} is obtained and the 3D geometry of the scene is constructed, the parameters of light source $\mathrm{L}' = \{\mathrm{L}_P, \mathrm{L}_F, \mathrm{L}_a, \mathrm{L}_w, \mathrm{L}_h, \mathrm{L}_k, \mathrm{L}'_C\}$ are refined by

$$\mathbf{L}'^{k+1} = \underset{\mathrm{L}'}{\arg\min} \sum_{p\in\mathbb{R}} \left(\boldsymbol{\alpha}_p\left(\mathbf{I}'_p(\mathbf{S}(\mathrm{C}^{k+1}), \mathrm{L}', \mathbf{E}^0) - \mathbf{I}_p\right)\right)^2.$$
$$\text{subject to } 0 < \mathrm{L}_a, 0 < \mathrm{L}'_C, 0 \leq \mathrm{L}_w, 0 \leq \mathrm{L}_h, 0 \leq \mathrm{L}_k \tag{17}$$

JCOBYLA[Auders Gustafsson 2012], which is used to solve Eq. 7, is also adopted to solve Eqs. 15 and 17 in our implementation.

For the images in which more than one light spot region is specified by the user, C^{k+1}_j, $p^{k+1}_{\mathrm{bL}'j}$, and $d^{k+1}_{\mathrm{L}'j}$ are first estimated by Eq. 15 with the pixels within the region $\mathbb{R}_j$, where $\mathbb{R}_j$ is the j-th user-specified light spot region. Then, the C^{k+1}_j that minimizes the overall error of all light spots is chosen as the final camera parameter,



(a) (b)

Fig. 9: (a) The input image and the environment illumination $\mathbf{E}^*$ recovered by removing the light spot with C^k and L'^k, where $k = 1$, 10, and 20. (b) The average difference between $\mathbf{I}$ and $\mathbf{I}'$. The x-axis indicates the number of iterations k and the y-axis illustrates the average difference.

that is,

$$\mathrm{C}^{k+1} = \underset{\mathrm{C}_j^{k+1}}{\arg\min} \sum_{i=1}^{n} \sum_{p \in \mathbb{R}_i} \left(\boldsymbol{\alpha}_p \left(\mathbf{I}'_p \left(\mathbf{S}(\mathrm{C}_j^{k+1}), \mathrm{L}'^k_{j\mathrm{T}(\mathrm{C}_j^{k+1}, \mu_i, p^{k+1}_{b\mathrm{L}'i}, d^{k+1}_{\mathrm{L}'i})}, \mathbf{E}^0 \right) - \mathbf{I}_p \right) \right)^2, \tag{18}$$

where μ_i is the centroid of the light spot in $\mathbb{R}_i$.

We found through experimentation that refining C and L' once is sufficient to synthesize visually plausible results. Refining with more iterations improves the estimation accuracy a little bit but takes much more time. Figure 9(a) shows the input image and the environment illumination $\mathbf{E}^*$ recovered with C^k and L'^k, where $k = 1, 10$, and 20, and 9(b) illustrates the average error computed by our objective function, which is the average difference between the input photograph $\mathbf{I}$ and the image $\mathbf{I}'$ that is synthesized with C^k, L'^k, and $\mathbf{E}^0$ in $\mathbb{R}$. Since we aim to build an interactive method, we decided to refine C and L' only once in all experiments in this paper. Additional experimental results demonstrating that one iteration is sufficient are explained in Section 7.

6.4 Recovery of Environment Illumination and Relighting

Let $\mathrm{C}^*$and $\mathbb{L}'^*$ denote the estimated camera and the set of estimated light sources, respectively. The environment illumination is recovered by removing the diffuse illumination synthesized by Eq. 4 as

$$\mathbf{E}^* = \frac{\mathbf{I}}{1 + \sum_{\mathbf{L}'^*_i \in \mathbb{L}'^*} \mathrm{diff}\left(\mathbf{S}(\mathrm{C}^*), \mathbf{L}'^*_i\right)}. \tag{19}$$

After the environment illumination $\mathbf{E}^*$ is recovered, users are allowed to relight the scene by adding and removing light sources and by changing their parameters. The new diffuse illumination, $\mathbf{D}''$, is synthesized as the sum of the diffuse illumination produced by each light source in $\mathbb{L}''$, which is the new set of light sources, and the final relit image $\mathbf{I}''$ is obtained by adding $\mathbf{D}''$ to $\mathbf{E}^*$, that is,

$$\mathbf{I}'' = \mathbf{D}'' + \mathbf{E}^*, \tag{20}$$

where

$$\mathbf{D}'' = \sum_{\mathrm{L}''_i \in \mathbb{L}''} \mathbf{E}^* * \mathrm{diff}\left(\mathbf{S}(\mathrm{C}^*), \mathrm{L}''_i\right).$$

7. EVALUATIONS

We performed experiments to evaluate the accuracy and robustness of the estimation by our method. The estimation accuracy of camera and light parameters is evaluated in Section 7.1. The robustness against the condition of light spots captured incompletely is evaluated in Sections 7.2 and 7.3. The robustness against rough and sketchy user-specified light spot regions is evaluated in Section 7.4.

7.1 Estimation Accuracy of Parameters

We prepared 30 synthetic test images rendered by Blender [Blender 2015] for this experiment. Each test image was synthesized with one of two settings of 3D geometries (planes and bumpy surfaces), one of two kinds of material (with and without specularity), one of four textures (grass, carpet, brick, and wood plank), and one random illumination condition. Both direct and indirect illumination were synthesized to render realistic images.

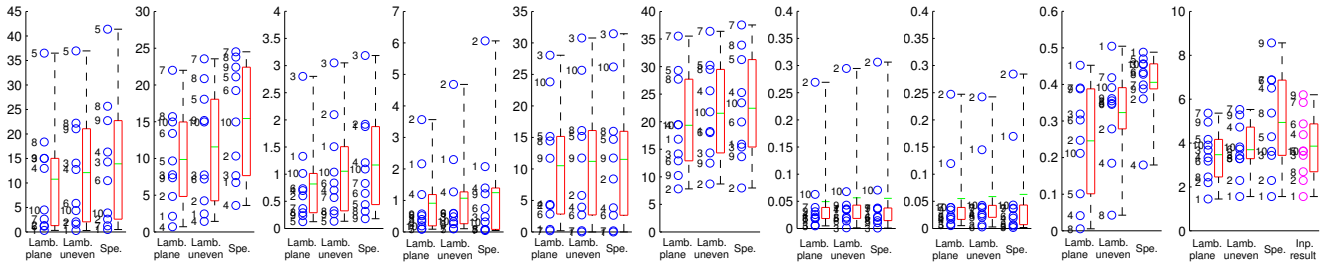
All 3D geometries consisted of a horizontal ground, a W-shaped vertical wall, and multiple vertical pillars embedded in the wall and all illumination conditions were composed of the environment illumination and the diffuse illumination produced by one area light source. The parameters of the area light source were randomly generated to produce light spot(s) on the ground, the wall, and/or the pillars. The intensity of the brightest point in the light spot(s) produced by the area light source was from 4 to 8 times brighter than the average intensity of the unlit area, which was about the same ratio as that between the brightest point in the light spot region and the environment illumination in the photographs used in this paper. The environment light was set to $(1, 1, 1)$.

The camera was an ideal pinhole camera set up to look at the center of the image plane. The focal length, angle of pitch rotation, and position of the camera were randomly generated. The resolution of test images was set to 1024×800.

Table II. : Comparison of estimation errors among different scenes

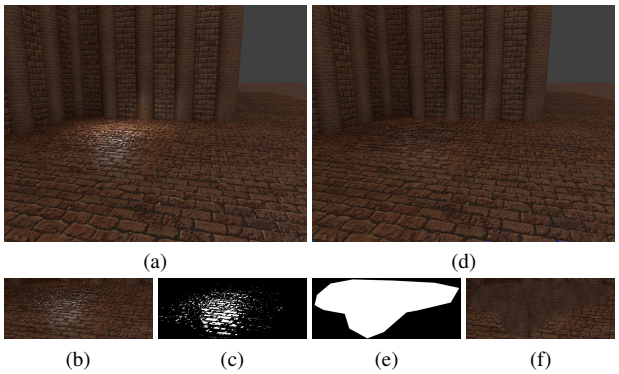
Param.	Plane		Uneven		Specular		Unit
	Mean	Std.	Mean	Std.	Mean	Std.	
C_f	10.76	11.34	12.11	11.80	13.90	13.15	Degree
C_ϕ	9.85	6.66	11.57	7.95	15.43	7.82	Degree
L_P	0.82	0.79	1.05	0.92	1.17	0.93	Camera height
L'_C	0.91	1.13	1.07	1.43	1.25	1.82	Int. of $\mathbf{E}$
L_F	10.51	9.82	11.21	10.72	11.51	10.81	Degree
L_a	19.33	8.96	21.53	8.81	22.43	9.61	Degree
L_w	0.05	0.08	0.06	0.09	0.06	0.09	Camera height
L_h	0.06	0.07	0.06	0.08	0.06	0.09	Camera height
L_k	0.25	0.15	0.32	0.13	0.41	0.09	None
$\mathbf{E}^*$	3.47	1.22	3.69	1.24	4.94 (3.85)	2.26 (1.50)	CIELAB Δ

Figure 10 and Table II summarize the estimation errors of our method for different settings of 3D geometries and materials. The estimation error of C_f is the difference between the field of view decided by the estimated C_f and that decided by the ground truth. Errors of C_ϕ, L_a, and L_k are the difference between the estimated results and the ground truth. The optical axis L_F is evaluated by the angle between the estimated result and the ground truth. Errors of L_P, L_w, and L_h are the differences in the position and size of the light source divided by camera height and error of L'_C is the difference in intensity relative to that of environment illumination, as our method only estimates scaled position, size, and intensity for light



(a) C_f. (b) C_ϕ. (c) L_P. (d) L'_C. (e) L_F. (f) L_a. (g) L_w. (h) L_h. (i) L_k. (j) $\mathbf{E}^*$.

Fig. 10: Errors of estimated parameters against different settings of the scene. The vertical axes show the error values and the units of each parameter are listed in Table II. Results for different 3D geometry and material settings are plotted separately in three columns and the numbers on the left of circles indicate the corresponding test images, which are shown in the supplementary document. The green lines and red rectangles on the right of the columns of circles visualize the mean and the range from the 25th percentile to 75th percentile, respectively. Enlarged plots are included in the supplementary document.



(a) (d) (b) (c) (e) (f)

Fig. 11: Removal of strong specular reflections. (a) Input image. (b) Our method removed the diffuse illumination but there exist noticeable artifacts caused by strong specular reflections. (c) A mask indicating the hole to be filled-in is generated by setting thresholds to (b). (d) Visually plausible result is produced by PatchMatch because our method removes the diffuse illumination in advance. (e) A user-specified mask of entire illuminated area. (f) The result of applying PatchMatch to fill-in the user-specified mask (e).

sources. The estimation error of $\mathbf{E}^*$ is the pixel-wise root-mean-square error (RMSE) in CIELAB color space. The computation of RMSE only takes pixels whose intensity is affected by the area spot light source into account.

From Fig. 10, we can see that the RMSE of all test images without specular reflections is less than or very close to 5, which means that the recovered $\mathbf{E}^*$ is very similar to the ground truth. However, RMSE obviously increases for test images with specular reflections because the areas with strong specular reflections (see Fig. 11(a)) are usually too bright to capture the physical color and texture in LDR images and thus result in noticeable artifacts, as shown in Fig. 11(b).

In cases where specular reflections remain in diffuse removal results, post-processing with the PatchMatch algorithm [2009] is effective. Since our method is a top-down approach, it is robust enough to remove diffuse illumination. After removing the diffuse illumination, the specular reflection regions are easily specified by a mask image created by thresholds (Fig. 11(c)) and are successfully filled-in by the PatchMatch algorithm because the patch distance it adopts is no longer affected by the diffuse illumination. Consequently, a visually plausible environment illumination (Fig. 11(d)) is produced. In contrast, if a large mask that covers the entire illuminated area (Fig. 11(e)) is applied to the PatchMatch algorithm, it is very difficult to produce a high-quality inpainted result (Fig. 11(f)). The numbers in parentheses in Table II indicate the RMSE of the inpainted results and Fig. 10(j) illustrates the RMSE of the inpainted results by magenta circles. As shown, the post-processing with the PatchMatch algorithm obviously alleviates the RMSE of $\mathbf{E}^*$ recovered from test images with specular reflections.

Table III. : Comparison between our method and the method proposed by Karsch et al.

Param.	Ours		Karsch		Unit
	Mean	Std.	Mean	Std.	
C_f	12.26	11.77	50.53	40.98	Degree
C_ϕ	12.28	7.62	33.72	21.35	Degree
L_P	1.01	0.86	21.97	17.56	Camera height
L'_C	1.08	1.44	3.95	4.11	Int. of $\mathbf{E}$
$\mathbf{E}^*$	4.03	1.72	9.77	6.25	CIELAB Δ

From this experiment, we found that our method may underestimate the beam angle and result in large error ($\geq$ 25 degrees) in L_a when the light source has a beam angle larger than 140 degrees because the intensity of diffuse illumination on the outer side of the light spot is usually too low to be separated from the environment illumination. However, our method still produces visually plausible $\mathbf{E}^*$ because people merely notice the difference on the outer side of the light spot.

For the other parameters of camera and light source, the experimental results indicate that our method estimates them accurately enough for visually plausible relighting of images showing various 3D geometries and illumination conditions. From Fig. 10, we can see that the errors only slightly increase when the settings (i.e., bumpiness of surfaces and specularity) deviate from our assumptions. The errors of C_ϕ and L_F are within the tolerable ranges indicated by [Jan J. Koenderink and Pont 2004; Lopez-Moreno et al. 2010] for all test images, but the parameters estimated from some of the test images are quite different from the ground truth. The results labeled 1 and 2 in Fig. 10 show significantly large errors in L'_C, L_w, and L_h, which are caused by an incomplete light spot, and the C_f estimated from the light spot on a plane that is almost parallel to the image plane is usually largely different from the ground

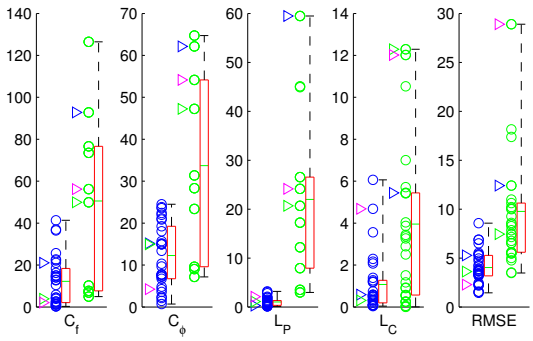


Fig. 12: Comparison between our method (left) and the method proposed by Karsch et al. [2011] (right). The green lines and red rectangles on the right of the columns of circles visualize the mean and the range from the 25th percentile to the 75th percentile. The triangles indicate the errors of the images in Fig. 13. Enlarged plots are included in the supplementary document.

truth because in such cases, changing C_f results in zooming in or zooming out from the light spot. However, our top-down approach optimizes the other parameters to produce light spots that are very similar to that in the input image, so we can still recover visually plausible $\mathbf{E}^*$ even though some of the parameters of camera and light source are largely different from the ground truth.

We performed an extra evaluation of estimation accuracy in which we compared our method with the one proposed by Karsch et al. [2011]. Their method also estimates non-distant area light sources as well as the albedo, and both their method and ours rely on user inputs. For their method, we manually marked the shape and position of the light source in a 2D image, specified two pairs of line segments indicating the vanishing points, and aligned the corners of the pseudo 3D geometry to the corresponding corners in the image as shown in the second row of Fig. 13. Because their method only estimates the position, and because color for light source and these two parameters are the most important and fundamental parameters to produce illuminations, we only evaluate these two when estimating the light source. The albedo recovered by their method is multiplied by the intensity of environment illumination before computing RMSE.

Figure 12 visualizes the distribution of estimation errors of Karsch et al.'s method (green circles) and ours (blue circles) and Table III summarizes the mean and standard deviation of the estimation errors, from which we can see that the errors of camera parameters estimated by their method are larger than ours for some test images, because in those test images (Fig. 13(a) shows one of them), there are no clear edges for their method to find vanishing points from which to estimate camera parameters. In contrast, the luminance distribution of the light spot provides enough clues for our method to estimate camera parameters more accurately. Figure 12 also shows that the errors in parameters of light source estimated by their method are larger than those estimated by ours. These differences may be caused by the adopted assumptions of light source. Their assumption is that the light source is a flat polygon that is on plane(s) in the scene (such as the ceiling or walls), which is different from many kinds of light source in outdoor scenes (such as streetlamps). In contrast, our method does not have any such assumption for light source, so it estimates light source parameters more accurately. As for the albedo, Table III shows that the RMSE of the albedo produced by their method is larger than that of $\mathbf{E}^*$ recovered by our method. This is because Color Retinex[Land et al. 1971; Grosse et al. 2009], which their method uses to create the initial guess, computes albedo by integrating pixel-wise derivatives, and the integration process accumulates small errors that occur on the boundaries of divided regions, as shown in the top row of Fig. 13(b), in which the ground is divided by the striped lines of the wooden texture, and results in the extremely bright region shown in the third row of Fig. 13(b). For images that do not contain divided regions as shown in Fig. 13(c), the RMSE of their results are less than or close to 5, but their pixel-wise intrinsic images decomposition algorithm produces blurred results, while ours are clear.

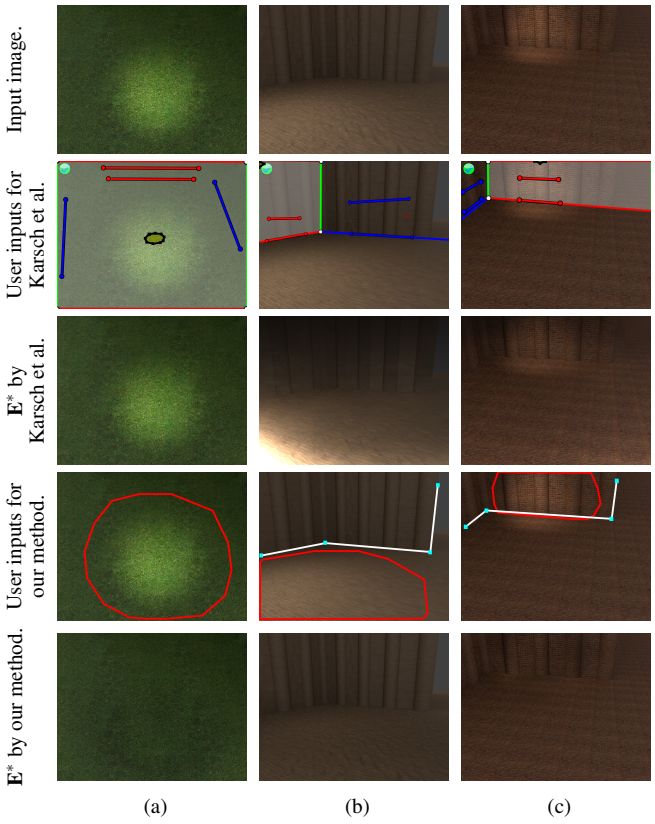
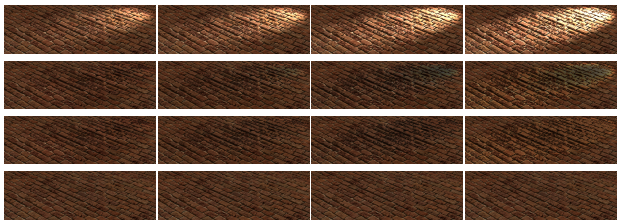


Fig. 13: Picked up test images used in comparison with the method proposed by Karsch et al., the user annotations, and the estimated environment illumination. (a) The test image that their method has large estimation errors in camera parameters. (b) The test image that their method has large RMSE in the recovered environment illumination. (c) The test image that their method produces blurred result. The estimation errors of (a), (b), and (c) are indicated by the blue, magenta, and green triangles in Fig. 12, respectively.

The experimental results show that our top-down approach works best for images that follow our assumption of planar and Lambertian surfaces, and although the estimation errors increase, it is applicable to uneven surfaces and specular reflections, which are beyond the assumption. In cases where specular reflections still remain in the diffuse removal result, the user can remove them by manually marking them with a threshold and then inpainting by the PatchMatch algorithm. The comparison of our method and Karsch et al.'s shows that our method produces a more visually plausible environment illumination than theirs does for images with bright light spots produced by a non-distant light source.



(a) 2.4% (b) 4.7% (c) 9.4% (d) 15.5%

Fig. 14: (a) to (d) : The top row is a zoomed-in part of original images, the second row is $\mathbf{E}^*$ recovered by our method, the third row is the results inpainted by PatchMatch, and the fourth row is the ground truth images. The percentages indicate the ratio of overbright areas in the light spot in original images.

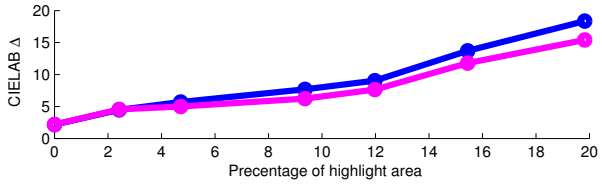


Fig. 15: The line graph shows the relationship between the percentage of overbright area and the RMSE between the ground truth and the recovered $\mathbf{E}^*$ (blue) and the result after inpainting by PatchMatch (magenta).

7.2 Robustness against Highlight-Clipping Effect

The comet-shaped light spot with highlight-clipping effect is a difficult case for our method. The LDR image may not capture physical color and texture for the regions affected by the highlight-clipping effect, and the influence of the highlight-clipping effect may not be alleviated by the confidence map explained in Section 5 because the luminance distribution of a comet-shaped light spot may be different from the 2D Gaussian function G_l in Eq. 5. We performed the following experiment to evaluate the robustness of our method against such a case.

We prepared seven test images for this experiment. The test images were synthesized with the scene explained in Section 7.1 and a light source illuminating the ground at grazing angle. All parameters of the light source were fixed, except for the intensity, which was progressively increased to gradually saturate the brightness of the light spot.

As in Section 7.1, the pixel-wise RMSE in the CIELAB color space between $\mathbf{E}^*$ and the ground truth image is computed. Note that only the pixels that are inside the user-specified light spot region $\mathbb{R}$ are used to compute the RMSE.

The line graph in Fig. 15 shows the relationship between the percentage of overbright area and the RMSE of $\mathbf{E}^*$ recovered by our method (blue) and that after inpainting by PatchMatch (magenta) for the test images. From this graph, we can see that the RMSE is lower than or very close to 5 for test images with a percentage of overbright area less than 4.7%, which means that the recovered $\mathbf{E}^*$s are very similar to the ground truth for these test images. For test images with a large percentage of overbright area, the RMSE increases, which indicates that our method is affected by the highlight-clipping effect.

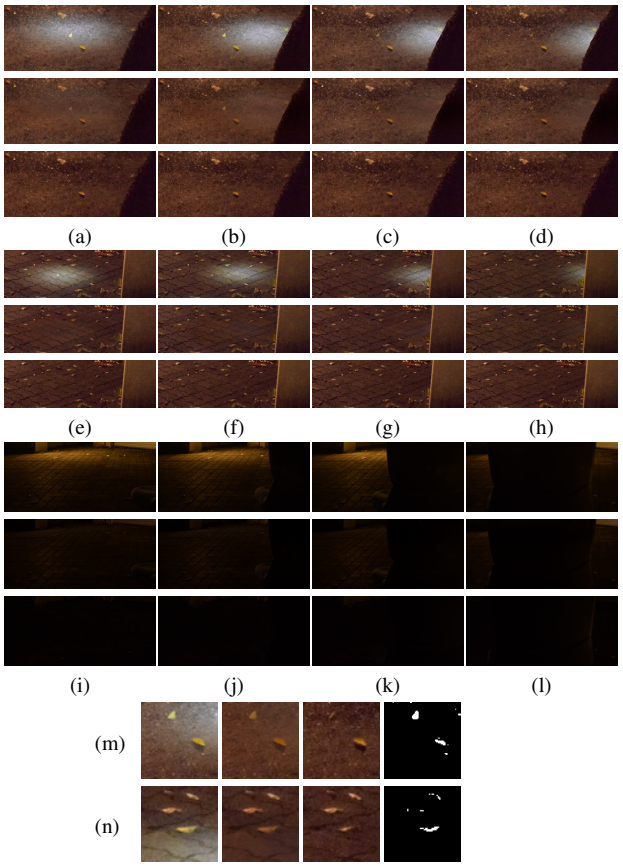


Fig. 16: (a) to (l) : Top rows are zoomed-in parts of original images, second rows are $\mathbf{E}^*$, and third rows are the ground truth images. (m) and (n): Zoomed-in parts of (b) and (f), where the columns from left to right are the original image, $\mathbf{E}^*$, and ground truth image. The rightmost columns illustrate pixels whose color difference in CIELAB color space between $\mathbf{E}^*$ and the ground truth image is greater than 5 (marked with a white color).

7.3 Robustness against Occlusion and Overlapping with Another Light Spot in Real Images

In this experiment, we examined how well our method can estimate a light spot whose appearance is incomplete or is overlapping with another light spot. We prepared five sets of test images and also their ground truth images. The images were taken with a Nikon D5300 digital camera and stored in Adobe RGB color space with 8bit JPEG format.

The first two sets consist of test images taken in outdoor scenes in which there is a light spot on the ground that is produced by an LED flashlight. The light spot is partially occluded by two different objects in these two sets. One of the objects is a tree that is darker than the ground and the other is a half-wall that is brighter than the ground. The test images in the third set show a scene lit by a standing pathway lamp instead of an LED flashlight and the light spot is occluded by an office chair placed at different locations. These three sets of test images and their ground truth images, in which there is no light spot, were prepared to evaluate the robustness against occlusion.

The remaining two sets were prepared for the evaluation of robustness against overlapping light spots. Both sets consist of test

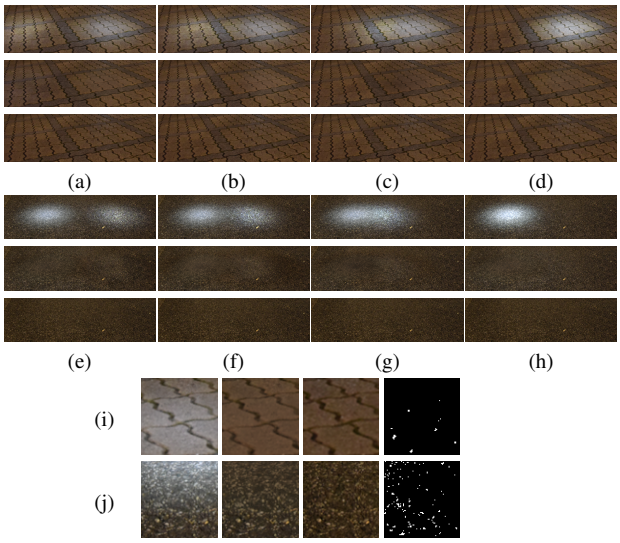


Fig. 17: (a) to (h) : Top rows are zoomed-in part of original images, second rows are the $\mathbf{E}^*$, and third rows are ground truth images. (i) and (j): Zoomed-in parts of (c) and (g), where the columns from left to right show the original image, $\mathbf{E}^*$, ground truth image, and the pixels whose color difference in CIELAB color space between $\mathbf{E}^*$ and the ground truth image is greater than 5 (marked with a white color).

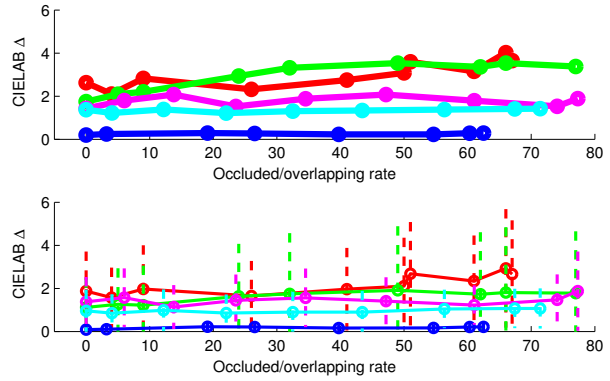


Fig. 18: (Top) The line graph illustrates the relationship between the percentage of occluded/overlapping area and the RMSE for test images in the five sets. (Bottom) The lines and dotted lines show the mean and standard deviation, respectively, of color difference (in CIELAB color space) of pixels in the user-specified light spot region in each test image.

images featuring one fixed light spot and one varying light spot. The fixed light spot is always at the same location in all test images, and the varying light spot appears at different positions and overlaps with the fixed light spot. The light spots in the test images in these two sets appear on the ground made of two different materials: brick and asphalt. The ground truth images of these two sets show the same scene without any light spot on the ground.

The influence of incompleteness of light spots upon our main purpose, which is the relighting of one image, is first evaluated by visual plausibility, as shown in Figs. 16 and 17. The second rows in Fig. 16(a) to (l) and 17(a) to (h) show the recovered $\mathbf{E}^*$ and the third rows are the ground truth images. The differences between $\mathbf{E}^*$ and the ground truth images are barely visible, demonstrating that $\mathbf{E}^*$ recovered by our method is quite similar to the ground truth.

Table IV. : Times users chose relit result as original photograph

Operation	Times chosen as original image	Times shown in experiment	Percentage
Reducing	167	264	63.3%
Enhancing	41	110	37.3%
Moving/ inserting	50	110	45.5%
Total	258	484	53.3%

Second, as in the experiment in Section 7.1, the pixel-wise RMSE in CIELAB color space between $\mathbf{E}^*$ and the ground truth image is computed as an objective evaluation for this experiment. Note that only the pixels that are inside the user-specified light spot region $\mathbb{R}$ (or the union of $\mathbb{R}$s for the test images from the sets of overlapped light spots) are used to compute the RMSE.

In Fig. 18, the line graph illustrates the relationships between RMSE and the percentage of occluded or overlapping area for test images from the five sets. The dotted lines show the means and standard deviations of color difference (in CIELAB color space) of the pixels in the user-specified light spot region in each test image. From this figure, we can see that the RMSE is low (≤ 5) in every test image. This means that the estimated $\mathbf{E}^*$ looks very similar to the ground truth image at most pixels. Large color differences occurred at the small bumps in Fig. 16(m)(n), the gaps between bricks in Fig. 17(i), and the small concaves of the asphalt road in Fig. 17(j). Since our method reconstructs the scene with flat planes, these are inevitable errors. However, these errors are difficult to actually perceive due of the inability of humans to notice inconsistencies between the illumination and the surface normals, as pointed out by [James T. Todd 1983; Yuri Ostrovsky and Sinha 2005].

7.4 Influence of User Input

The estimation with our method depends heavily upon the annotations by users, especially those of the light spot region, which does not have clear boundaries in general. We explored how the user annotations of light spot regions affect the estimation results by means of an experiment with a variety of regions annotated by several users.

The same as the experiment in Section 7.3, this experiment evaluates visual plausibility by observing the recovered environment illumination $\mathbf{E}^*$. Figure 19 displays the $\mathbf{E}^*$ obtained with different user annotations. As shown, our method can accurately estimate and remove the light spot even with rough and sketchy user annotations that circle a wide area. However, if the user strokes across the central part of the light spot (the two images at the bottom of Fig. 19(a), for instance), or if the user circles the light spot too tightly (the image at the bottom of Fig. 19(d)), the result becomes inaccurate. Fortunately, it only takes about 30 seconds to a few minutes for the user to annotate the image and see the estimated result, so the strokes can be redrawn to circle a wider region to obtain a better result.

8. USER STUDY

In addition to the comparisons with ground truth, we evaluated the quality of relit images and the usability of our method by user tests.

(a) (b) (c) (d)

Fig. 20: Some of the original photographs (top rows) and the relit result (bottom rows) by reducing intensity of the light spot (a), enhancing intensity of the light spot (b), and moving/inserting synthetic lights (c) in our user test. During this test, a red circle is shown to indicate the region that is relit (d).

(a) (b) (c) (d)

Fig. 19: (a)(d) User annotations. (b)(c) Top rows show the ground truth images and second to fourth rows are the environment illumination $\mathbf{E}^*$ recovered with different user annotations.

Fig. 21: Enhancing intensity of light spots in original image (top) often causes over-brightened light spots (bottom).

8.1 Side-by-Side User Test

With this test, we evaluated the quality of the relighting results generated by our method. Here, 22 photographs that include indoor scenes and outdoor landscapes taken from various viewpoints are relit by reducing or enhancing the intensity of light spots, or by moving/inserting synthetic lights. Figure 20(a)(b)(c) shows instances of the relit results with different operations (reducing, enhancing, and moving/inserting) and their original photographs used in this test. Both the relit image and the original photograph were displayed side-by-side to participants with the red circle in both images indicating the relit region, as shown in Fig. 20(d). Participants were asked to choose which of the two images they felt was the original photograph. Twenty-two individuals participated in this experiment. The results are summarized in Table IV. The rows indicate the count and percentage chosen as original photographs for the synthetic images that were relit by reducing, enhancing, or moving/inserting lights.

The first row shows that the participants tended to choose the synthetic images that were relit by lights whose intensity was reduced. In contrast, enhancing the intensity of the light spots had the worst score, as shown in the second row. One reason that diminished the visual plausibility of the image relit by enhanced lights is the use of over-brightened lights, as suggested in Fig. 21. Another possible reason may be that the term "lighting" in the instruction "The lighting in one of these two images was modified. Please choose one that was not modified" in this test biased the answer and worked against the brighter images. Since the overall score of the three rows is close to the chance of random guess (50%), our relighting results are fairly realistic. Comments from some of the participants indicate that the noticeable inconsistency in an image

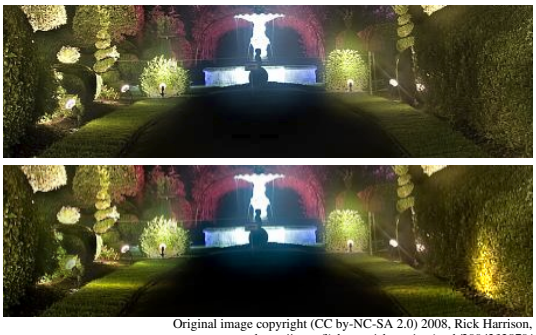


Fig. 22: The original image (top) and the result relit by inserting synthetic lights on the fence on the right (bottom).

is increased if there is a light spot but no corresponding light source, as with the light spot on the right in Fig. 22. These comments suggest that inserting a light source object into the image increases the visual plausibility of the results. Therefore, combining our method with an existing object-inserting method or tool would enable the synthesis of more realistic and consistent results.

8.2 Usability Test

We evaluated the effectiveness of our method as part of the image editing process by comparing the two fundamental processes for realistic relighting with the conventional Photoshop CS6 photo-editing software. The two processes are 1) inserting a new light spot and 2) removing an existing light spot. We asked four participants to insert and remove the light spot in photographs by using our method and by using the conventional photo-editing software and then we compared the results and time spent with both methods. All participants were familiar with the basic operations of the photo-editing software, but it was the first time for most of them to use our method, so we briefly explained how to use it before the test began.

Figure 23(a) shows the results of inserting new light spots into a photograph by our method (top), by manual painting with the photo-editing software (middle), and by the lighting effect function in the photo-editing software (bottom). As shown, our result and the manually painted result are more visually acceptable than the rendering result of the lighting effect function because the latter does not consider the 3D geometry of the scene. From this experiment, we found that users can paint visually plausible light spots only if they know the exact appearance of the light spots in the scene; otherwise they may create a visually unacceptable result. In contrast, our method parameterizes the light spots by several intuitive parameters (position, optical axis, color, etc.), enabling users to insert realistic light spots even without precise knowledge of their appearance.

Figure 23(b) shows the results of removing existing light spots from the photograph by our method (middle) and by manual painting with the stamp tool of the photo-editing software (bottom). Manual removal by the photo-editing software did not preserve the texture of the wall, while in contrast, our method produced a more visually plausible result.

With our method, it took participants 22 to 90 seconds to draw circles around the light spot regions and to specify the boundaries of the wall and 1 to 3.5 minutes for them to add and move one light, while with the photo-editing software, it took 10 to 20 minutes for them to paint a new light and about 20 minutes to erase one light spot.

(a) (b)

Fig. 23: (a) Results of inserting new light spot by our method (top), by manual painting with photo-editing software (middle), and by lighting effects rendered by the photo-editing software (bottom). (b) Input image (top) and results of removing existing light spots by our method (middle) and by manual painting with photo-editing software (bottom).

These results show that our method is more effective for inserting new light spots and removing existing light spots and that it produces more visually pleasant results than conventional photo-editing software.

9. RESULTS AND DISCUSSION

We implemented our method on an Intel i7-3770 3.4GHz, 16GB RAM machine with a GeForce GTX 660 graphics card. The resolution of all input images in this paper ranged from 400k to 1M pixels. It takes 5 to 120 seconds to estimate the camera parameters and reconstruct the 3D geometry, and about 10 to 70 seconds to estimate one light source.

Our method is applicable to the relighting of photographs of outdoor and large indoor scenes illuminated by non-distant light sources, as shown in Figs. 24 and 25. In addition to these scenes, our method can be applied to light spots in the shadows in daytime scenes. Figure 26 shows the relit result of a daytime scene in which there are light spots in the shadow produced by sunlight leaking from the trees. However, our method cannot remove or reproduce the shadows because it adopts a simplified 3D geometry model. This may cause inconsistencies in the relit results, as shown around the pedestrians in the top-right of Fig. 24.

The experimental results indicate that for the images of scenes consisting of bumpy and/or non-Lambertian surfaces, although the error in $\mathbf{E}^*$ increases, our method can still produce visually plausible $\mathbf{E}^*$. In the most difficult case for our method, which is the image containing comet-shaped light spots with overbright areas, artifacts may appear when the percentage of overbright area in the light spots is more than 4.7%.

We also found through experiments that our method failed in three cases. The first case is when our method cannot handle the light spots appearing on complicated 3D geometries such as the

Fig. 24: Input images (left) and relit results (right).

Fig. 25: Input images (left) and relit results (right).

Fig. 26: A daytime scene (left) and the result relit by our method (right).

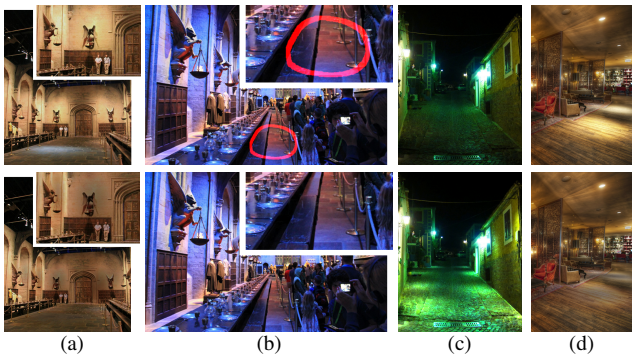
statues on the wall in Fig. 27(a), although the experimental results in Section 7.1 had shown that our method is applicable for light spots on an uneven surface. The diffuse illumination on the statues at the center of the light spots is too different from that on a planar surface. Therefore, it is difficult for our method to completely remove the diffuse illumination in such cases.

The second case is when the whole user-specified light spot region is illuminated by light sources other than white color. In such case, our method may generate incorrect environment illumination (Fig. 27(b)) because our optimization process estimates the color of the light source as the ratio of colors to the environment illumination instead of the absolute value. This same reason causes our method to sometimes generate incorrect diffuse illumination for images like the ones shown in Fig. 27((c)top), in which the whole ground is illuminated by green light. Our method recovers the environment illumination as green in this case, and the synthetic light spots on the ground only reflect the green color (Fig. 27((c)bottom)).

The third case in which our method may fail is when the tone of the input photograph has been modified by nonlinear functions such as tone-mapping. Our illumination model cannot fit the luminance distribution in such photographs because the illumination in those photographs is different from the physical luminance distribution. An example of such a case is shown in Fig. 27(d).

10. CONCLUSION AND FUTURE WORK

We introduced an interactive method that estimates the luminance distribution and surface reflectance and produces visually plausible relit images. It requires only a few user annotations that are easy to specify and requires no information about the scene or any aid from



(a) (b) (c) (d)

(a) original image copyright (CC by 2.0) 2012. Karen Roe, *https://www.flickr.com/photos/karen_roe/7172834407/*
(b) original image copyright (CC by 2.0) 2012, Karen Roe, *https://www.flickr.com/photos/karen_roe/7172825377/*
(c) original image copyright (CC by-NC-SA 2.0) 2010, nick p, *https://www.flickr.com/photos/nickperez/5760682356/*
(d) original image copyright (CC by-NC-SA 2.0) 2012, Trey Ratcliff, *https://www.flickr.com/photos/stuckincustoms/4079229245/*

Fig. 27: Images for which it is difficult for our method to produce visually plausible results. Top: input images and (b) also shows the user annotation of light spot region. Bottom: (a)(b)(d) are the recovered environment illuminations and (c) is the relit result.

special tools. We achieved it by the top-down approach, which defines simplified models of the area light, camera, and 3D geometry and finds the best parameters to reproduce the luminance distribution to fit that in an input photograph. The experimental results show that although the estimation error of environment illumination increases, it may still produce visually plausible environment illumination for images of scenes consisting of bumpy and/or non-Lambertian surfaces, which means that the proposed method is robust enough for relighting images. The user studies show that the synthetic results are indistinguishable from real photographs and that the relighting process of our method is easy to use.

There are five topics that we would like to address in future work. First, the proposed method focuses on estimating diffuse illumination rather than removing the shadows and inserting the light source objects. The shadows and light source objects may cause inconsistencies in the relit results. Therefore, developing a method that combines the proposed method with shadow-removal methods such as [Wu et al. 2007; Panagopoulos et al. 2009] to remove the shadows and developing a method that can realistically insert light source objects into the image by using the estimation results of the proposed method may improve the quality of the resultant images.

Second, as stated in Section 7.1, the pixels affected by the highlight-clipping effect in LDR images (the specular reflections in Fig. 11(a) and the red square in the top row of Fig. 23(b)) may result in untextured blank regions in the recovered environment illumination. The PatchMatch algorithm [Barnes et al. 2009] is adopted to alleviate this artifact in our implementation. We think it is possible to improve the quality of the resultant images by taking the estimated light source and 3D geometry into account.

Third, adopting other methods such as [Carroll et al. 2011] and [Bell et al. 2014] may produce the initial guess of albedo more accurately. The estimated parameters of camera and light source may also be more accurate.

Then, the manipulation of the relighting process of the proposed method can be improved by combining it with lighting design methods such as [Boyadzhiev et al. 2013] to create a more easy-to-use relighting system.

Finally, as a possible application, our method provides a useful tool to calibrate the parameters of a camera with a single image for cases where there are no clues for 3D geometries. Image forgery detection (e.g., [Johnson and Farid 2005] and [Kee et al. 2013]) is another interesting potential application of our method. The process in our method that calculates the difference between the luminance distribution in an image and that produced by our physically well-approximated models would be useful for such purpose.

REFERENCES

AUDERS GUSTAFSSON, C. 2012. Constrained optimization by linear approximation for java. http://www.codeproject.com/Articles/508513/Derivative-free-nonlinear-optimization-for-NET-and.

BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph. 28,* 3 (July), 24:1–24:11.

BARRON, J. T. AND MALIK, J. 2012. Color constancy, intrinsic images, and shape estimation. In *ECCV '12*. 57–70.

BELL, S., BALA, K., AND SNAVELY, N. 2014. Intrinsic images in the wild. *ACM Trans. Graph. 33,* 4 (July), 159:1–159:12.

BLENDER. 2015. 2.72 release. https://www.blender.org/.

BOUSSEAU, A., PARIS, S., AND DURAND, F. 2009. User-assisted intrinsic images. *ACM Trans. Graph. 28,* 5 (Dec.), 130:1–130:10.

BOYADZHIEV, I., PARIS, S., AND BALA, K. 2013. User-assisted image compositing for photographic lighting. *ACM Trans. Graph. 32,* 4 (July), 36:1–36:12.

CAO, X. AND SHAH, M. 2005. Camera calibration and light source estimation from images with shadows. In *IEEE CVPR '05*. 918–923.

CARROLL, R., RAMAMOORTHI, R., AND AGRAWALA, M. 2011. Illumination decomposition for material recoloring with consistent interreflections. *ACM Trans. Graph. 30,* 4 (July), 43:1–43:10.

CHANDRAKER, M. AND RAMAMOORTHI, R. 2011. What an image reveals about material reflectance. In *IEEE ICCV '11*. 1076–1083.

COLBURN, A., AGARWALA, A., HERTZMANN, A., CURLESS, B., AND COHEN, M. F. 2013. Image-based remodeling. *IEEE TVCG 19,* 1, 56–66.

CRIMINISI, A., REID, I., AND ZISSERMAN, A. 2000. Single view metrology. *Int. J. Comput. Vision 40,* 2 (Nov.), 123–148.

DONG, Y., TONG, X., PELLACINI, F., AND GUO, B. 2011. Appgen: Interactive material modeling from a single image. *ACM Trans. Graph. 30,* 6 (Dec.), 146:1–146:10.

EL-HAKIM, S., WHITING, E., AND GONZO, L. 2005. 3d modelling with reusable and integrated building blocks. In *7 th Conference on Optical 3-D Measurement Techniques*. 3–5.

GARCES, E., MUNOZ, A., LOPEZ-MORENO, J., AND GUTIERREZ, D. 2012. Intrinsic images by clustering. *Comput. Graph. Forum 31,* 4 (June), 1415–1424.

GROSSE, R., JOHNSON, M. K., ADELSON, E. H., AND FREEMAN, W. T. 2009. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *IEEE ICCV '09*. 2335–2342.

HACHAMA, M., GHANEM, B., AND WONKA, P. 2015. Intrinsic scene decomposition from rgb-d images. In *IEEE ICCV '15*. 810–818.

HARA, K., NISHINO, K., AND IKEUCHI, K. 2005. Light source position and reflectance estimation from a single view without the distant illumination assumption. *IEEE TPAMI 27,* 4 (Apr.), 493–505.

HARA, K., NISHINO, K., AND IKEUCHI, K. 2008. Mixture of spherical distributions for single-view relighting. *IEEE TPAMI 30,* 1 (Jan.), 25–35.

HARTLEY, R. AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, Second ed. Cambridge University Press.

HORN, B. K. P. 1989. Shape from shading. MIT Press, Cambridge, MA, USA, Chapter Obtaining Shape from Shading Information, 123–171.

HSU, E., MERTENS, T., PARIS, S., AVIDAN, S., AND DURAND, F. 2008. Light mixture estimation for spatially varying white balance. *ACM Trans. Graph. 27,* 3 (Aug.), 70:1–70:7.

IIZUKA, S., KANAMORI, Y., MITANI, J., AND FUKUI, Y. 2011. Efficiently modeling 3D scenes from a single image. *IEEE CG&A 32,* 6, 18–25.

JAMES T. TODD, E. M. 1983. Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance 9,* 4, 583–595.

JAN J. KOENDERINK, A. J. V. D. AND PONT, S. C. 2004. Light direction from shad(ow)ed random gaussian surfaces. *Perception 33,* 12, 1405–1420.

JIANG, N., TAN, P., AND CHEONG, L.-F. 2009. Symmetric architecture modeling with a single image. *ACM Trans. Graph. 28,* 5 (Dec.), 113:1–113:8.

JOHNSON, M. K. AND FARID, H. 2005. Exposing digital forgeries by detecting inconsistencies in lighting. In *Proceedings of the 7th Workshop on Multimedia and Security*. 1–10.

KARSCH, K., HEDAU, V., FORSYTH, D., AND HOIEM, D. 2011. Rendering synthetic objects into legacy photographs. *ACM Trans. Graph. 30,* 6 (Dec.), 157:1–157:12.

KEE, E., O'BRIEN, J. F., AND FARID, H. 2013. Exposing photo manipulation with inconsistent shadows. *ACM Trans. Graph. 32,* 3 (July), 28:1–28:12.

KRONANDER, J., BANTERLE, F., GARDNER, A., MIANDJI, E., AND UNGER, J. 2015. Photorealistic rendering of mixed reality scenes. *Comput. Graph. Forum 34,* 2 (May), 643–665.

LAGGER, P. AND FUA, P. 2006. Using specularities to recover multiple light sources in the presence of texture. In *ICPR*. 587–590.

LAGGER, P. AND FUA, P. 2008. Retrieving multiple light sources in the presence of specular reflections and texture. *Comput. Vis. Image Underst. 111,* 2 (Aug.), 207–218.

LALONDE, J.-F., EFROS, A. A., AND NARASIMHAN, S. G. 2012. Estimating the natural illumination conditions from a single outdoor image. *Int. J. Comput. Vision 98,* 2 (June), 123–145.

LAND, E. H., JOHN, AND MCCANN, J. 1971. Lightness and retinex theory. *Journal of the Optical Society of America 61,* 1, 1–11.

LI, H., VOUGA, E., GUDYM, A., LUO, L., BARRON, J. T., AND GUSEV, G. 2013. 3d self-portraits. *ACM Trans. Graph. 32,* 6 (Nov.), 187:1–187:9.

LIU, Y., QIN, X., XU, S., NAKAMAE, E., AND PENG, Q. 2009. Light source estimation of outdoor scenes for mixed reality. *Vis. Comput. 25,* 5 (Apr.), 637–646.

LOMBARDI, S. AND NISHINO, K. 2012. Reflectance and natural illumination from a single image. In *ECCV '12*. 582–595.

LOPEZ-MORENO, J., GARCES, E., HADAP, S., REINHARD, E., AND GUTIERREZ, D. 2013. Multiple light source estimation in a single image. *Comput. Graph. Forum 32,* 8, 170–182.

LOPEZ-MORENO, J., HADAP, S., REINHARD, E., AND GUTIERREZ, D. 2010. Compositing images through light source detection. *Computers & Graphics 34,* 6, 698–707.

LOPEZ-MORENO, J., SUNDSTEDT, V., SANGORRIN, F., AND GUTIERREZ, D. 2010. Measuring the perception of light inconsistencies. In *ACM APGV '10*. 25–32.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Trans. Graph. 22,* 3 (July), 759–769.

OH, B. M., CHEN, M., DORSEY, J., AND DURAND, F. 2001. Image-based modeling and photo editing. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. 433–442.

PANAGOPOULOS, A., SAMARAS, D., AND PARAGIOS, N. 2009. Robust shadow and illumination estimation using a mixture model. In *IEEE CVPR '09*. 651–658.

POWELL, M. J. D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal 7,* 2, 155–162.

REN, P., DONG, Y., LIN, S., TONG, X., AND GUO, B. 2015. Image based relighting using neural networks. *ACM Trans. Graph. 34,* 4 (July), 111:1–111:12.

SHEN, L., TAN, P., AND LIN, S. 2008. Intrinsic image decomposition with non-local texture cues. In *IEEE CVPR '08*. 1–7.

SHI, J., DONG, Y., TONG, X., AND CHEN, Y. 2015. Efficient intrinsic image decomposition for rgbd images. In *ACM VRST '15*. 17–25.

TUNWATTANAPONG, B., GHOSH, A., AND DEBEVEC, P. 2011. Practical image-based relighting and editing with spherical-harmonics and local lights. In *CVMP '11*. 138–147.

WANG, L., WEI, L.-Y., ZHOU, K., GUO, B., AND SHUM, H.-Y. 2007. High dynamic range image hallucination. In *Eurographics Symposium on Rendering '07*. 321–326.

WU, T.-P., TANG, C.-K., BROWN, M. S., AND SHUM, H.-Y. 2007. Natural shadow matting. *ACM Trans. Graph. 26,* 2 (June).

XU, S. AND WALLACE, A. M. 2008. Recovering surface reflectance and multiple light locations and intensities from image data. *Pattern Recogn. Lett. 29,* 11 (Aug.), 1639–1647.

YU, Y., DEBEVEC, P., MALIK, J., AND HAWKINS, T. 1999. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *ACM SIGGRAPH '99*. 215–224.

YURI OSTROVSKY, P. C. AND SINHA, P. 2005. Perceiving illumination inconsistencies in scenes. *Perception 34,* 11 (Nov.), 1301–1314.

ZHENG, Y., CHEN, X., CHENG, M.-M., ZHOU, K., HU, S.-M., AND MITRA, N. J. 2012. Interactive images: cuboid proxies for smart image manipulation. *ACM Trans. Graph. 31,* 4 (July), 99:1–99:11.

ZHOU, W. AND KAMBHAMETTU, C. 2002. Estimation of illuminant direction and intensity of multiple light sources. In *ECCV '02*. 206–220.

ZICKLER, T., RAMAMOORTHI, R., ENRIQUE, S., AND BELHUMEUR, P. N. 2006. Reflectance sharing: Predicting appearance from a sparse set of images of a known shape. *IEEE TPAMI 28,* 8 (Aug.), 1287–1302.

ZORAN, D., KRISHNAN, D., BENTO, J., AND FREEMAN, B. 2014. Shape and illumination from shading using the generic viewpoint assumption. In *NIPS '14*. 226–234.