Illumination Decomposition for Photograph With Multiple Light Sources

Ling Zhang, Qingan Yan, Zheng Liu, Hua Zou, and Chunxia Xiao, *Member, IEEE*

***Abstract*— Illumination decomposition for a single photograph is an important and challenging problem in image editing operation. In this paper, we present a novel coarse-to-fine strategy to perform illumination decomposition for photograph with multiple light sources. We first reconstruct the lighting environment of the image using the estimated geometry structure of the scene. With the position of lights, we detect the shadow regions as well as the highlights in the projected image for each light. Then, using the illumination cues from shadows, we estimate the coarse illumination decomposed image emitted by each light source. Finally, we present a light-aware illumination optimization model, which efficiently produces the finer illumination decomposition results, as well as recover the texture detail under the shadow. We validate our approach on a number of examples, and our method effectively decomposes the input image into multiple components corresponding to different light sources.**

***Index Terms*— Illumination cues, illumination decomposition, multiple light sources, shadows.**

I. Introduction

DECOMPOSING and editing the illumination of a photograph is a fundamental photo editing task, which is widely used in image color editing [1]–[3], object composition [4], [5] and image relighting [6]. The illumination we observe at each pixel is the result of complex interactions between the lighting and the reflectance of materials in the scene. A feasible approach for illumination decomposition is first to estimate the information of the light sources in the scene, and then separate the shading into components corresponding to different light sources. With the decomposed illumination, many applications, such as shadow editing, object recoloring and scene relighting, can be developed.

However, illumination decomposition for a single image with multiple light sources is a challenging work. First, we have to know the light information and the illumination distribution in the scene. For outdoor scene in the day, the sun is usually the only light source. While for the indoor photograph, there are usually several lights in the scene, and some lights might be out of the view volume of the camera, which makes the lighting analysis to be a difficult problem. Second, we need to separate the illumination emitted by each light source, for both the direct illumination and indirect (ambient) illumination. As it is an under-constrained problem, even given the light source information, finding a linear or nonlinear combination to meet the decomposition requirement for each light is quite a difficult task.

Several methods have been proposed for image illumination decomposition. Carroll *et al.* [1] decomposed an input image into multiple basis sources, which correspond to direct lighting and indirect illumination from each material. This method focuses on one direct light. It is not practical for illumination decomposition with multiple light sources. Nayar *et al.* [6] exploited high frequency illumination for separating the direct and global illumination components of a scene measured by a camera and illuminated by a light source. This method also handles the scene illuminated by one direct light. Furthermore, the method requires several input images. Different from previous methods, in this paper, we present an illumination decomposition method for a single photograph with multiple lights sources.

We exploit a coarse-to-fine strategy for separating the input image into multiple components corresponding to different light sources (Fig. 12). Specifically, our method consists of the following main steps. With the help of the geometry structure of the scene, we first reconstruct the light information in the scene and the illumination distribution of the image, and detect the shadow regions in the projected image corresponding to each light. Then, using the extracted illumination cues from shadows, we estimate the initial decomposed image emitted by each light via illumination editing propagation. Finally, we present a light-aware illumination optimization model to improve the initial decomposition results, which efficiently produces the finer illumination decomposition results. We present our algorithm overview in Fig. 5.

Our method can perform the illumination decomposition task with only a single image as input, which makes our method more practical than previous methods. In addition, in contrast to Carroll *et al.* [1], our decomposition algorithm does not depend heavily on the intrinsic image decomposition results. As intrinsic image decomposition itself is a difficult underconstrained problem, especially for scene with complex illumination conditions and materials, our method only uses

Manuscript received December 11, 2016; revised May 3, 2017 and May 26, 2017; accepted May 31, 2017. Date of publication June 6, 2017; date of current version June 23, 2017. This work was supported in part by NSFC under Grant 61472288 and Grant 61672390, in part by NCET under Grant NCET-13-0441, in part by the State Key Laboratory of Software Engineering under Grant SKLSE-2015-A-05, in part by the Foundation of Key Research Institute of Humanities and Social Science at Universities under Grant 16JJD870002, and in part by the Chinese Ministry of Education. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaochun Cao. *(Corresponding author: Chunxia Xiao.)*

L. Zhang, Q. Yan, H. Zou, and C. Xiao are with the State Key Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: lingzhang@whu.edu.cn; yanqingan@whu.edu.cn; zouhua08@gmail.com; cxxiao@whu.edu.cn

Z. Liu is with the National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430072, China (e-mail: liu.zheng.jojo@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2712283

1057-7149 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the coarse intrinsic decomposition results for light information estimation. Our method can deal with light sources with different colors (Fig. 10), and we need not to pre-process white balance for the input image by eliminating color casts due to differing lights [7], [8]. We demonstrate the applicability of our method on a number of examples, including some illumination editing applications.

In summary, the main contributions of the paper are as follows:

- Reconstrcut the light environment and illumination distribution (including shadow and highlight) of the input image using the geometry structure of the scene.
- Propose an illumination propagation method to calculate the initial illumination decomposition for each light using the extracted illumination cues from shadows.
- Present an illumination decomposition optimization model to improve the coarse decomposition results, which effectively recovers the texture details and produces more visually pleasing results.

II. Related Work

In this section, we review the two groups of the most related works to our illumination decomposition system: image decomposition and illumination recovery.

A. Image Decomposition

Image decomposition is a fundamental problem in digital image processing, which decomposes the image into different components by decomposition method to extract the required image information. Intrinsic image decomposition is one of the basic image decomposition. Since the notion of intrinsic images was first introduced in [9], many intrinsic image decomposition methods have been proposed [10], mainly due to its potential wide applications in computer graphics and vision. The common methods are to decompose the input image into reflectance image and shading image. According to the Retinex theory [11], in general, large intensity gradients correspond to reflectance changes and shading is smoother. Based on this theory, several methods [12]–[14] applied the pixel gradient and texture configuration for intrinsic image decomposition. The global sparsity prior is also a useful cue for image decomposition, and these methods consider that the natural image is dominated by a relatively small set of material colors [15]–[18]. As automatic method is difficult for tackling the complex images, user interaction has been incorporated into intrinsic image decomposition [19]–[22] for producing more accurate results. Using surface normals, depth or point clouds also have been incorporated in intrinsic decomposition [23]–[25].

Another common image decomposition is illumination decomposition. One natural way for image illumination decomposition is using the intrinsic image decomposition results, that is, with the decomposed shading or illumination component, the shading image is further decomposed into components corresponding to different light sources. Carroll *et al.* [1] proposed a user-assisted illumination decomposition method for a single image for material editing. They first decomposed the input image into illumination and reflectance components employing interactive intrinsic image decomposition [19]. Then they further decomposed illumination component with multiple basis sources, corresponding to direct lighting and indirect illumination from each material. Similar to [1], our approach also requires some user interactions, but our approach addresses the illumination decomposition for image with multiple direct light sources, furthermore, as the intrinsic image decomposition itself is also a difficult problem, our method only uses the coarse intrinsic image decomposition results for light estimation..

There are also some methods for image decomposition without using intrinsic images. Seitz *et al.* [26] proposed an inverse light transport method for illumination decomposition, and Bai *et al.* [27] separated an image into a sum of components for each light bounce. Using high frequency illumination patterns, Nayar *et al.* [6] separated the direct and global components of a scene lit by a single light source with a wide variety of global illumination effects. O'Toole *et al.* [28] decomposed light transport into direct and indirect lighting, and they separated the direct and indirect components for uniform lighting. Recently, Dong *et al.* [2] decomposed the illumination for recoloring diffuse surfaces with consistent interreflections. Note that, these methods usually used a large number of images captured using active lighting under controlled conditions.

B. Illumination Recovery

Illumination recovery is also known as inverse lighting, which is to recover the illumination distribution for a scene from the appearance of objects located in the scene. To perform illumination recovery, the light sources of the image needs to be identified. Hara *et al.* [29] presented two methods for recovering the light source position from a single image, and the light source position helps to improve the illumination recovery accuracy. Zhang and Yang [30] extracted multiple illumination information from the shading of a sphere. Mercier *et al.* [31] proposed a framework to automatically recover an object shape, reflectance properties, and light sources from a set of images. Kenji *et al.* [32] used EM algorithm to simultaneously estimate both the number of point light sources and the reflectance property of an object. Relying on a combination of weak cues extracted from different portions of the image, Lalonde *et al.* [33] estimated the illumination for a single outdoor image. By using a single image to estimate the light number and positions is a difficult problem. Inspired by [34] and [35], we explore shadow as well user assistance to resolve this problem.

Several methods have been proposed for illumination recovery from a single image with cast shadows, and these methods are based on the idea that images with cast shadows can be sparsely represented. Sato *et al.* [36] recovered the illumination distribution of a scene with shadows cast by an object of known shape. They introduced an adaptive sampling framework for illumination distribution estimation. Okabe *et al.* [37] employed a Haar wavelet basis to recover lighting in images with cast shadows. Ravi *et al.* [38] analyzed that the cast

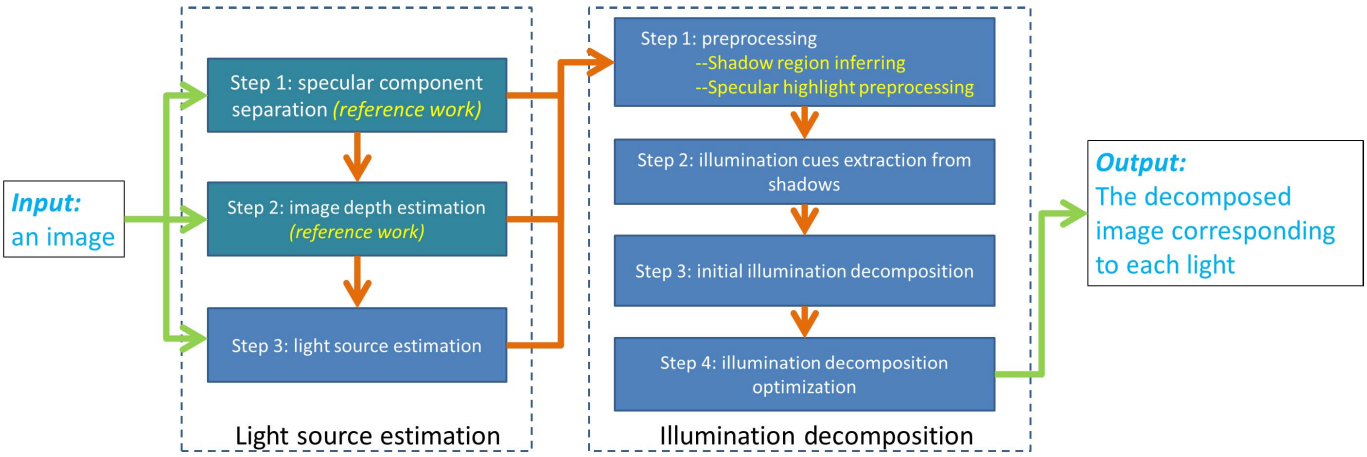


Fig. 1. The pipeline of the proposed illumination decomposition system.

shadows can be represented using convolutions and Fourier basis functions. Based on the observation that the image set produced by a Lambertian scene with cast shadows can be represented by a sparse set of images generated by directional light sources, Xue *et al.* [39] exploited sparse representation for recovering the illumination of a scene from a single image with cast shadows, given the geometry of the scene. Sparse representation for images with cast shadow is efficient due to dimensionality reduction. These methods handle the scene illuminated by one direct light. Furthermore, the method [39] requires several images. In contrast, our input is a single photograph with multiple direct lights, and the output is multiple decomposed images corresponding to different direct light sources.

III. Overview

Different from previous works in image illumination decomposition that deal with image with one direct light in the scene, we process image with multiple direct lights and decompose the input image into multiple components corresponding to different lights. Let I denote the input image. N is the number of light sources in the input image. I^k represents the illumination decomposition image for the k*th* light source, and we try to extract I^k respectively from the input image I.

Our illumination decomposition approach consists of two main stages: light source estimation and illumination decomposition. Fig. 1 gives a brief overview of the proposed system. Light source estimation stage provides light environment of the image for the following illumination decomposition. Illumination decomposition stage focuses on illumination calculation for each light by using the information deduced from the image. We demonstrate our method using an example in Fig. 5.

A. Light Source Estimation

We first separate specular highlights (if exist) from the input image and estimate the depth map for the image, which are used as the assistance information for our algorithm. Then we determine the number and initial position of the lights in input image with some user interaction, and refine the estimated initial light positions by optimizing an energy function.

B. Illumination Decomposition

This section is the core content of the proposed method. We infer shadow regions for each light and analyze the lighting condition in the image. Then, we extract the illumination cues from shadows and apply propagation strategy to estimate initial illumination decomposition results. Finally, we utilize a light-aware model to optimize the initial decomposition results to produce more realistic illumination decomposition results.

IV. Light Source Estimation

Lighting is a significant element on image appearance. By reconstructing the lighting condition of a scene, we thus can get more accurate illumination decomposition results. In order to effectively estimate the number of the light sources and their corresponding physical positions, we first detect specular highlights (if exist) within the input image and estimate its associate depth map. As shadows are posed by the occlusion of light transmission, we thus utilize the orientation of shadows to predict the initial position of light sources. With the information of depth map, specular highlights and shadows, we build an energy function to estimate the final position of light sources.

A. Specular Component Separation

The specular highlights reveal useful cues for light position estimation. As shown in Fig. 2, the input image has two lights, which accordingly cause highlights appearing on the vase. We use chromaticity-based specular removal technique [40], [41] to separate the specular component from I and estimate the chromaticity of different lights. Similar to [41] we assume that multiple point lights exist in the scene and all light sources have the same color (white light).

For I, we use the algorithm of [40] to separate its specular component. The separated specular component A^0 (Fig. 2(b)) will be used in the light position estimation process and the estimated light color can be edited in illumination editing. We use the information around the specular region to repair the specular region. The repaired image with specular removed is denoted as I^0.

(a) (b) (c)

Fig. 2. Specular component separation. (a) is the input image. (b) is the separated specular image. (c) is the image with specular removed.

(a) (b)

Fig. 3. Image depth estimation. (a) is the input images. (b) is the predicted depth images.

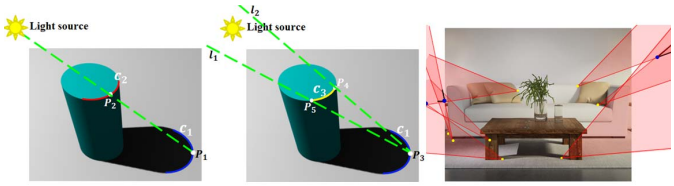
B. Image Depth Estimation

The illumination information in an image with multiple lights is complex, which makes the shadow detection and shading analysis difficult for the illumination decomposition task. The depth map (the geometric structure) of the scene facilitates the analysis of illumination distribution and shadow labeling.

We use the deep convolutional neural field (DCNF) model [42] to extract the depth map from the input image, which formulates the depth estimation as a deep CRF learning problem. Fig. 3 presents two image depth estimation results using DCNF technique. Based on the computed depth map, we estimate the geometry structure of the scene and get the 3D global coordinate of each point in the image.

C. Initial Light Information Estimation

Given a photograph, the lights in the scene are often outside of the view frustum. To accurately infer the number of the light sources is a difficult problem, especially for scenes with large number of light sources. In this paper, instead of automatically determining the number of the light sources, we apply the geometric constraints technique [34], [35], [43] between shadows, shadow-casting objects and the light sources to identify the number of lights in the scene. Meanwhile, we can estimate an initial position for each light.



(a) (b) (c)

Fig. 4. Physical consistency between shading, shadows and lighting. (a) The blue curve c_1 is the cast shadow contour in the planar surface and the red curve c_2 is the corresponding curve on the shadow-casting object with respect to the light. Specifying a point P_1 in c_1, according to the imaging principle, there is a point (P_2) in c_2, and the ray connecting point P_1 and P_2 intersects the light source. (b) P_3 is a point in curve c_1, and curve c_3 on the cylinder model is the region that may cast shadow point P_3. P_4 and P_5 are the ends of curve c_3. l_1 and l_2 are rays connected point P_3 and the ends of curve c_3. Point P_3, ray l_1 and l_2 combine into a wedge region corresponding to point P_3. (c) The number and the initial position estimation for the lights in an image using the physically consistency. The yellow points and red wedge regions are labeled by user interactions.

In an image, the shading and shadows should be physically consistent for a light source, as shown in Fig. 4(a). According to [34] and [35] for a shadow region, we select a point in shadow and set a wedge region that encompasses the corresponding object casting the shadow (Fig. 4(b)). If some shadows are cast by the same light source, there will be an overlap between the wedges for these shadows, and the projection of the light source may be located in this overlap region. Hence, we can determine the number and the initial position of light sources depending on the overlaps between all wedge regions in the projected image. As shown in Fig. 4(c), by applying the wedge-shaped constraints in the image, we conclude that there are three light sources in the image, and the light sources are located in the region between the black lines. In our method, we consider the vertex of the black region (the blue point) as the initial position of light source.

D. Light Source Estimation

With the estimated depth map and the separated specular component, we proceed to optimize the initial light position. As shading, shadow and specularity information in a scene can provide important cues for estimating multiple illuminant directions [43]. Inspired by [4] and [35], we present the following energy function to optimize the initial position of each light:

$$\arg\min \{\sum_{p\in P}\{\alpha_1(A^0-A)^2+\alpha_2(B^0-B)^2+\alpha_3(C^0-C)^2\} +\alpha_4\sum_{k=1}^{N}(L_k^{'}-L_k)^2\}, \quad (1)$$

where A^0 is the separated specular image, and B^0 is the shading image which can obtain by performing intrinsic image decomposition [20] on I^0. C^0 is the region size of the most prominent shadow in the input image, which is detected using the close-form matting method [44], [45], as shown in Fig. 5(h). $L_k^{'}$ is the initial position of the k*th* light. L_k is the optimized light position of the k*th* light. N is the number of

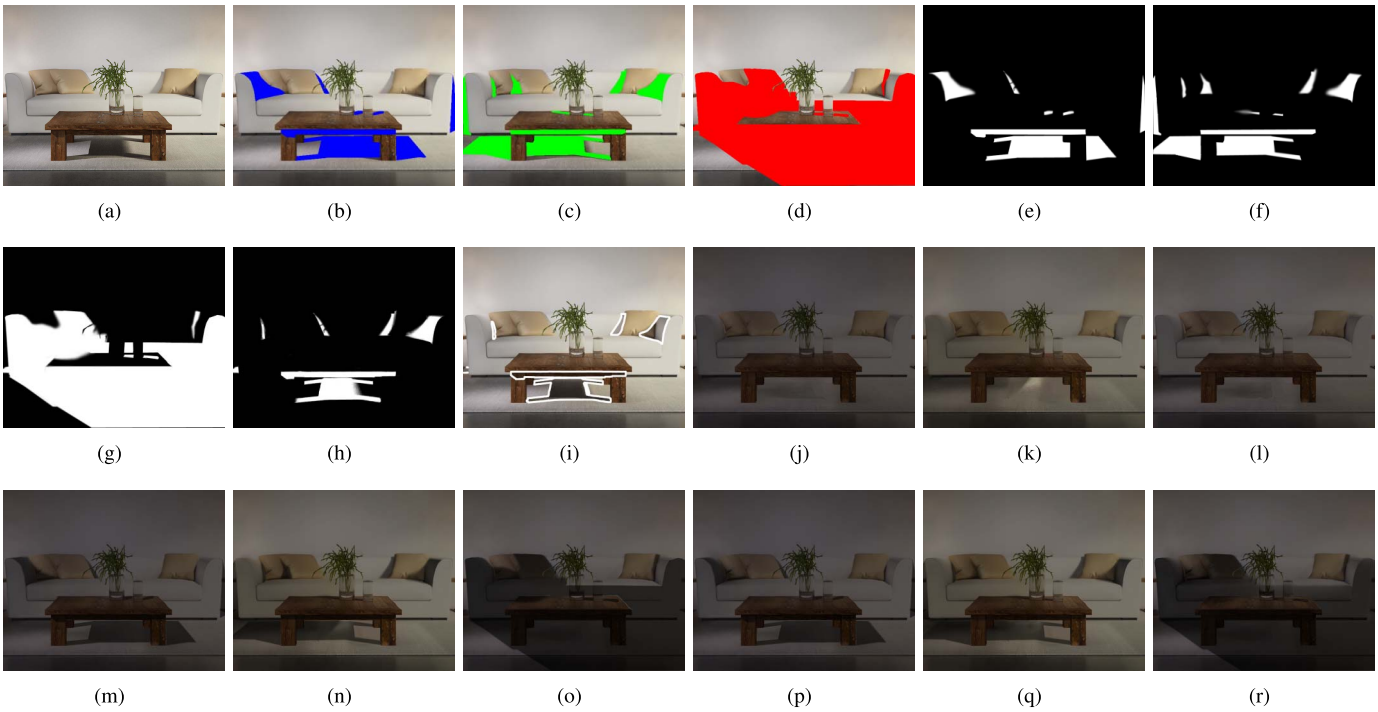


Fig. 5. Illumination decomposition system overview. (a) is the input image. (b), (c) and (d) are the inferred shadow regions for each light in the projected image (the blue region, the green region and the red region). (e), (f) and (g) are the estimated shadow mask map based on the inferred shadow regions for each light. (h) is the detected prominent shadow mask for the input image. (i) is the prominent shadow boundary of the input image. (j), (k) and (l) are the coarse decomposed results without shadows. (m), (n) and (o) are the coarse decomposed results with shadows. (p), (q) and (r) are the final decomposed results for each light.

TABLE I
THE POSITION OF THE LIGHTS IN FIG. 4

Light source	The estimated position		
	x	y	z
Light 1	-0.5766	1.2031	0.3515
Light 2	-1.3422	0.2164	-0.4589
Light 3	-0.0672	0.6686	-0.3886

lights in the image. α_1, α_2, α_3 and α_4 are balance parameters. Note that, in our paper we ignore the size of light sources.

We minimize the energy function (Eq. (1)) by using an iterative method to obtain the optimal light positions, which progressively approximate the true light positions in each iteration. Our optimization begins with the initial light positions and the estimated depth map. At each iteration, according to the Phong model, we redraw the scene and produce a new re-rendered image. Consequently, we obtain a new specular image A, a new shading image B and a new prominent shadow size C. Using these information, we solve Eq. (1) using Gradient Descent Method to get new light positions. This is an iteration of our optimization process. With the estimated new light positions, we continue the next iteration. We stop the iterative process when the following conditions are satisfied: the maximum number of iterations is reached, or the energy difference between two iterations is less than a threshold value. Table I is the estimated light position for each light in Fig. 3(a). Fig. 6 is the energy curve of the optimized process corresponding to Table I, which shows how the energy decreases over iterations.

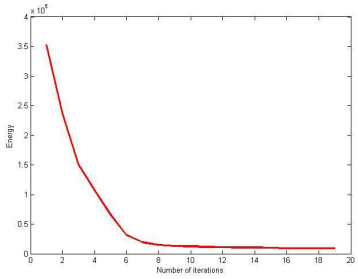


Fig. 6. The energy curve for the optimized process corresponding to Table I.

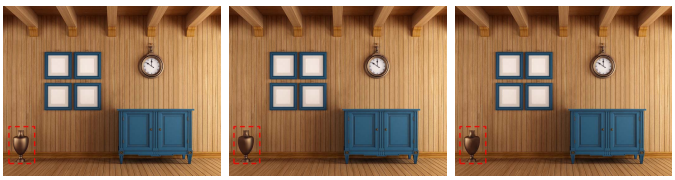
V. ILLUMINATION DECOMPOSITION

With the obtained illumination information and the geometry structure of the scene, we accomplish our illumination decomposition task for each light using the illumination cues from shadow regions.

A. Preprocessing

Before the illumination decomposition, we have to analyze the relationship between lights and shadows. Meanwhile, for the highlight image, the inappropriate highlights in the image may influence the accuracy of the decomposition results. Then, we also have to deal with the highlights corresponding to each light (if the image has specular highlights).

1) Shadow Region Inferring: With complex lighting environment, some shadows corresponding to one light may be not obviously observed in the input image. To overcome this issue we infer shadow regions for each light by exploring



(a) (b) (c)

Fig. 7. Specular highlights preprocessing. (a) is the input image with two highlights. (b) is the image only with highlights from 1*st* light. (c) is the image only with highlights from 2*ed* light.

the estimated geometric structure of the scene and the direction of the light. As shown in Fig. 5, the colored regions in (b), (c) and (d) are the shadows corresponding to three different light sources, respectively. We use a trimap to display the shadow regions for each light, which white region represents shadow and black region represent nonshadow. Correspondingly, we obtain a shadow mask map with smooth boundaries for each light, as shown in Fig. 5(e, f, g). Note that, the shadow regions here contain both cast shadows and attached shadows, where attached shadow is an important component to generating realistic images. With the inferred shadow regions, we can identify the lighting condition in each region.

Moreover, we apply the closed-form matting method [44] to interactively detect the prominent shadows in the input image. Fig. 5(h) is the mask map of detected shadows in Fig. 5(a). With this information, we then can identify the prominent shadow boundaries in the input image, as shown in Fig. 5(i).

2) Specular Highlight Preprocessing: For the highlight image, with the geometry structure of the scene and the position of this light, we can identify the relationship between the highlights and lights. In order to get a decomposed result with correct highlight information we have to remove the highlights that are not generated by the current light. As shown in Fig. 7, there are two lights on the vase (the red box in (a)), and the highlights are from the two different lights. With the highlights removing, (b) is the image preserving highlights from 1*st* light while (c) is the image preserving highlights from 2*nd* light. The image only with highlights from k*th* light is denoted as I_h^k.

B. Illumination Cues From Shadows

With the observation that there is no illumination contribution from the light source on the shadow regions where this light source is blocked, we can extract illumination cues from shadow regions. With these illumination cues, we can estimate the initial decomposition illumination in some special regions for each light source. We assume that there is a fixed ambient illumination for the input image, and the ambient illumination is also appeared in the decomposition images.

S_k is the shadow regions cast by the k*th* light source in image. If $S_k \cap S_t \neq \varnothing$ and $t \neq k$, we consider there are overlaps between S_k and S_t. We employ a user-assisted way to extract the illumination cues for each light, and draw strokes to specify sample region F_k in each S_k. Note that, the shadow samples from different S_k may not have the same lighting environment. These shadow samples from different S_k do not need to share similar material. For each F_k, we specify a sample region G_k in nonshadow region that G_k and F_k share similar material. Different shadow sample regions with similar material can share the same nonshadow sample region. As shown in Fig. 8, the shadow samples in (a), (c) and (d) share similar material, and there is only one corresponding nonshadow sample in each image. In Fig. 8(b), the two shadow samples have different materials, so each shadow sample has a corresponding nonshadow sample.

With these samples, we calculate the illumination cues for each light in shadow samples. Take a scene with three light sources as an example ($N = 3$), the specified three shadow samples share different materials. Accordingly, we specified three nonshadow samples corresponding to the shadow samples. Let μ_{F_1}, μ_{F_2}, μ_{F_3} be the average intensity of the three shadow samples F_1, F_2, F_3; μ_{G_1}, μ_{G_2}, μ_{G_3} be the average intensity of the corresponding nonshadow samples G_1, G_2, G_3, We assume F_k and G_k have the similar average reflectance R_{F_k}. Let the illumination intensities for three direct lights are L_1, L_2, L_3, respectively, and the intensity of ambient illumination is L_a. We assume that in sample F_k only the kth light is occluded. Then, we have:

$$\begin{cases} (L_2 + L_3 + L_a)R_{F_1} = \mu_{F_1} \\ (L_1 + L_3 + L_a)R_{F_2} = \mu_{F_2} \\ (L_1 + L_2 + L_a)R_{F_3} = \mu_{F_3} \\ (L_1 + L_2 + L_3 + L_a)R_{F_1} = \mu_{G_1} \\ (L_1 + L_2 + L_3 + L_a)R_{F_2} = \mu_{G_2} \\ (L_1 + L_2 + L_3 + L_a)R_{F_3} = \mu_{G_3}. \end{cases} \tag{2}$$

With Eq. (2), we get the relationship between R_{F_1}, R_{F_2} and R_{F_3}, that is: $R_{F_2} = \frac{\mu_{G_2} R_{F_1}}{\mu_{G_1}}$ and $R_{F_3} = \frac{\mu_{G_3} R_{F_1}}{\mu_{G_1}}$. Then, we compute the mean intensity for each light in sample F_1 (containing the ambient lighting):

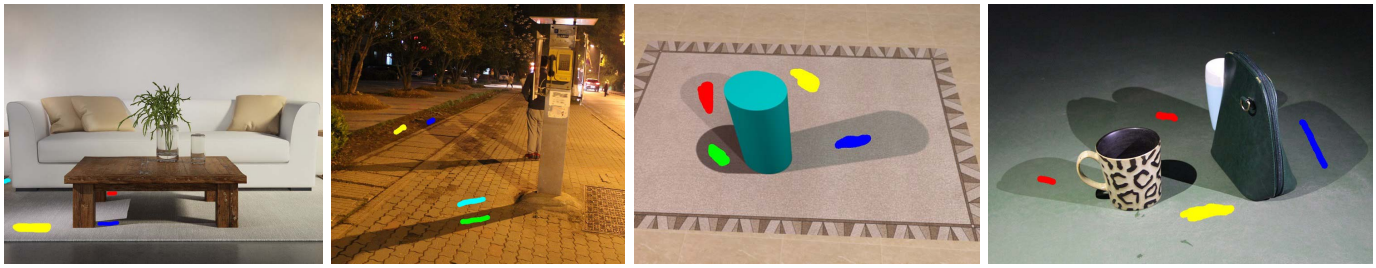
$$\begin{cases} (L_1 + L_a)R_{F_1} = \dfrac{\mu_{F_2}\mu_{G_1}}{\mu_{G_2}} + \dfrac{\mu_{F_3}\mu_{G_1}}{\mu_{G_3}} - \mu_{G_1} \\ (L_2 + L_a)R_{F_1} = \mu_{F_1} + \dfrac{\mu_{F_3}\mu_{G_1}}{\mu_{G_3}} - \mu_{G_1} \\ (L_3 + L_a)R_{F_1} = \mu_{F_1} + \dfrac{\mu_{F_2}\mu_{G_1}}{\mu_{G_2}} - \mu_{G_1}. \end{cases} \tag{3}$$

Meanwhile, we calculate the ratio between the direct illumination and the ambient illumination via Eq. (2): $t_k = \frac{L_k}{L_a}$, for each light. Similarly, the mean intensity for each light in sample F_2 and F_3 can be obtained. Note that, for scene with N lights, the mean intensity for each light in each sample can be calculated in the similar way, and the general formula for computing the mean intensity for the kth light in sample F_1 is that:

$$(L_k + L_a)R_{F_1} = \sum_{t=1, t\neq k}^{N} \left(\frac{\mu_{F_t}\mu_{G_1}}{\mu_{G_t}}\right) - \mu_{G_1}. \tag{4}$$

This general formula is validated for scenes with four lights and five lights.

Let k be the current light source. I_i is the intensity value at pixel i in the input image (if the input image has highlights, we use the image only with highlights from kth light, that is, we use I_h^k instead of I). E_i^k denotes the estimated intensity



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

Fig. 8. Shadow samples and corresponding nonshadow samples. The strokes with same color in shadow regions are samples from the same S_k, which have the same lighting environment. Different color regions have different lighting environment.

value at pixel $i \in F_t$ in the decomposition image I^k, then based on Eq. (3), we get:

$$E_i^k = \frac{(L_k + L_a)R_{F_t} \times I_i}{\mu_{F_t}}, \quad i \in F_t,\ t \in \{1, 2, \cdots, N\},\ t \neq k. \tag{5}$$

Using above method, we can obtain the decomposed intensity values in each F_t for each decomposition image I^k.

C. Initial Illumination Decomposition

After obtaining the decomposition results for each light in the specified shadow regions, we use these decomposition results as the illumination cues for the initial illumination estimation. The basic idea is that, with the illumination cues, we propagate the intensities in known region to unknown regions.

For pixel i and its adjacent pixel j in the input image I, with the same illumination condition, we have

$$\begin{cases} I_i = (L_d + L_a)R_i \\ I_j = (L_d + L_a)R_j, \end{cases} \tag{6}$$

where R_i and R_j are the reflectance at pixel i and pixel j. L_d and L_a are the direct and ambient illumination at the two pixels. Then, we get:

$$\frac{R_i}{R_j} = \frac{I_i}{I_j}. \tag{7}$$

Similarly, for the pixel i and its adjacent pixel j in the decomposition image I^k, if the illumination condition is the same at the two pixels, we have

$$\frac{R_i}{R_j} = \frac{I_i^k}{I_j^k}, \tag{8}$$

where I_i^k and I_j^k are the intensity at pixel i and pixel j in the decomposition image I^k. With Eq. (7) and Eq. (8), we get the relationship between the input image and the decomposition image at pixel i and pixel j:

$$\frac{I_i}{I_j} = \frac{I_i^k}{I_j^k}. \tag{9}$$

So using the known intensity I_j^k, we estimate the value of I_i^k: $I_i^k = \frac{I_i \times I_j^k}{I_j}$. With this result, we propagate the illumination from known regions to unknown regions.

Eq. (9) is valid only if pixel i and pixel j have the same illumination condition in the input image. While on the shadow boundaries, the lighting condition between adjacent pixels may be different. In this situation, Eq. (9) will not be valid. Instead, we use weighted mean of adjacent pixels to calculate the pixel on shadow boundaries. We denote shadow boundaries as P, as shown in Fig. 5(i).

Similar to the editing propagation [46], we propagate the intensity from $F_t \subset S_t$ to other region. We denote the known region as $Z_k = \{F_t, t \in \{1, \cdots, N\}, t \neq k\}$ and the rest unknown regions as U. We estimate the intensity value in region U through propagating outwards from the boundary of Z_k. We denote region U by layers, starting from the boundaries of Z_k. The width of each layer is one pixel. Let total number of layers be N_k. For facilitating the illumination propagation, we introduce an array of tags $flag[*]$. $index(i)$ is the index number of pixel i. If $i \in Z_k$, $flag[index(i)] = 1$; else, $flag[index(i)] = 0$.

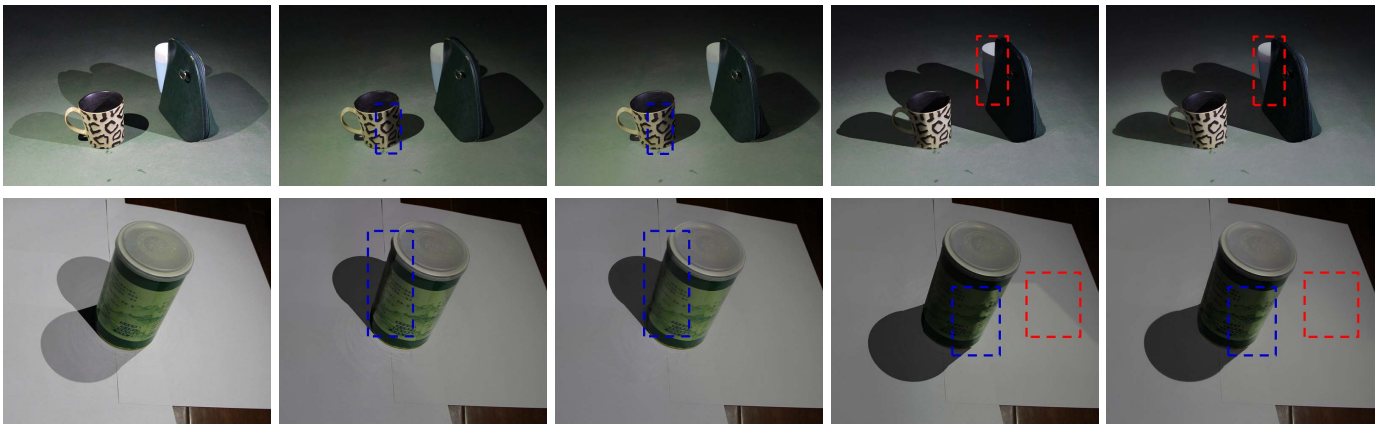
Let n be the current layer, we initialize n to be 1. L_n represents the region located in the current layer. We repeat the following operations until $n > N_k$:

If $i \in L_n$ and $flag[index(i)] = 0$,

$$E_i^k = \begin{cases} \dfrac{\sum\limits_{j \in N(i)} w_{ij} E_j^k}{\sum\limits_{j \in N(i)} w_{ij}} & i \in P,\ flag[index(j)] = 1 \\ \dfrac{\sum\limits_{j \in N(i)} w_{ij} (\frac{I_i}{I_j} \times E_j^k)}{\sum\limits_{j \in N(i)} w_{ij}} & i \notin P,\ flag[index(j)] = 1, \end{cases} \tag{10}$$

where $N(i)$ is a local neighborhood of pixel i, w_{ij} measures the similarity between pixel i and pixel j, and $w_{ij} = \exp(-\frac{(I_i - I_j)^2}{2\sigma^2})$. If $flag[index(i)] = 0$ and the value of this pixel has been computed, we let $flag[index(i)] = 1$, and n adds 1. By estimating the unknown pixel using the adjacent pixels with estimated intensity value, we can estimate the decomposed intensity values in U for the k*th* light.

Using above propagation strategy, we calculate the coarse decomposed image E^k for the k*th* light, and the illumination in the shadow regions is also recovered, as shown in Fig. 5(j, k, l).



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

Fig. 9. The coarse decomposed results and the optimized results for each light. (a) is input images. (b) and (d) are initial illumination decomposition for each light. (c) and (e) are the final optimized results for each light.

1) Shadow Fusion: Shadow is an important component for images. We denote the estimated shadow mask for the k*th* light as s^k, which is a trimap, as shown in Fig. 5(e, f, g). We composite the cast shadows and the attached shadows onto the decomposed result E^k applying the composting equation [47], and our composting function is:

$$C_i^k = \alpha_i D_i^k + (1 - \alpha_i) E_i^k, \tag{11}$$

where C_i^k is the intensity at pixel i of the decomposition image I^k containing shadows, which is a linear combination of the foreground intensity D_i^k and the background intensity E_i^k, weighted by the fusion operator α_i^k. D_i^k describes the shadow intensity value at pixel i for the k*th* light in shadow regions. As there is no light striking the scene surface of shadow regions, we consider $D_i^k = 0$. E_i^k is the estimated decomposed result which we have obtained before. We set $\alpha_i = \begin{cases} 0 & s_i^k = 0 \\ \frac{s_i^k - A_i^k}{255} & s_i^k \neq 0 \end{cases}$, where s_i^k is the intensity value of pixel i in shadow map s^k, and A_i^k is the intensity value produced by the ambient illumination at pixel i with the lighting condition of light k.

As the illumination in shadow regions on the decomposed image E^k is recovered for each light, we have $(L_k + L_a)R_i = E_i^k$. Based on equations in Eq. (2), we know that $L_k = t_k L_a$. Then, the ambient intensity with the lighting condition of light k at pixel i is: $A_i^k = L_a R_i = \frac{E_i^k}{t_k + 1}$. In Fig. 5, (m, n, o) are the decomposition results with both cast shadows and attached shadows.

D. Illumination Decomposition Optimization

We have got the initial decomposed results for each light. But the results may be coarse, especially for scene with complex illumination and materials. Because of the propagation errors, there may be some artifacts in the coarse results, as shown in the initial results in Fig. 9(d) (the red box in the second row). In addition, the shadows may be composed unnaturally. For example, in Fig. 9, the boundary of the attached shadow is too hard (the blue box) or the texture of the object is not clear (the red box in the first row). To improve the initial results and achieve finer decomposition results for each light, we formulate the following energy function to optimize the initial results:

$$E = E_{data} + \lambda_1 E_{smooth} + \lambda_2 E_{\det ail}. \tag{12}$$

This energy model contains three terms: data item E_{data}, smoothing item E_{smooth} and detail-preserving item E_{detail}. Parameters λ_1 and λ_2 are the balance weights. We use I_i^k to denote the desired decomposed value for the k*th* light at pixel i.

1) Data Item: This item is used to obtain a reasonable decomposition for each light source:

$$E_{data} = \sum_k \sum_i (I_i^k - C_i^k)^2, \tag{13}$$

where $k \in \{1, \cdots, N\}$, C_i^k is the intensity at pixel i of the initial estimated decomposition image for the k*th* light source.

2) Smoothing Item: Generally, the final decomposed result tends to be similar for local window with similar appearance. We use a smoothing item to constrain the local similarity of similar appearance. To enforce this policy for each decomposed image, we define the smoothing term as:

$$E_{smooth} = \sum_k \sum_i \sum_{j \in N(i)} z_{ij} (I_i^k - I_j^k)^2, \tag{14}$$

where $N(i)$ is a local window with center at pixel i. The affinity coefficient z_{ij} measures the appearance similarity between pixel i and pixel j and is defined as $z_{ij} = \exp(-\frac{(I_i - I_j)^2}{2\sigma^2})$.

3) Detail-Preserving Item: Smoothing process may lead to texture detail blurring, as illustrated in Fig. 14(b). To preserve the details of decomposed image, we minimize the difference of the gradient information between the input image and the desired decomposition image. The decomposed image contains only the shadows cast by the current light source, and the gradient from shadow boundaries created by other light source in the input image is not desirable to present. Thus, we need to perform special treatment on the shadow boundaries. To address this problem, we add a weighting factor

in the term to distinguish the useful gradient and undesirable gradient. The detail-preserving item is defined as follows

$$E_{detail} = \sum_k \sum_i w_{ij} (\nabla I_i^k - \nabla I_i)^2, \tag{15}$$

where ∇ is the gradient operator. The weighting factor w_{ij} is designed as:

$$w_{ij} = \begin{cases} \exp(-\frac{(a_i - a_j)}{2\sigma^2}), i \in \overline{P_k} \\ \exp(-\frac{(I_i - I_j)}{2\sigma^2}), \ else, \end{cases} \tag{16}$$

where $\overline{P_k}$ represents the shadow boundaries which are not created by the k*th* light source, a_i and a_j are the values of channel a in the Lab space for the input image at pixel i and pixel j.

We solve the optimization problem using Gauss-Seidel method. In the iterative optimization process, we use the estimated coarse decomposed results C_i^k as the initial value. As shown in Fig. 5(p, q, r) and Fig. 9(d, g), the decomposition optimization model significantly improves the initial results. The artifacts are eliminated and the details are effectively recovered. The transition on the boundaries of attached shadow is smoother. The decomposed results, taking both the cast shadow and the attached shadow into account, appear visually natural and realistic.

VI. Experiments and Discussions

In this section, we present a variety of illumination decomposition results to validate the performance of the proposed method. We also apply our method on image illumination editing. We perform our method on both synthetic images and natural images, and run our algorithm using C++ on a PC machine equipped with Pentium (R) Dual-Core CPU E5200@2.50GHz with 2GB RAM. In our experiments, the size of $N(i)$ is set to 13×13 and σ is set to 10. In addition, we set $\alpha_1 = 2$, $\alpha_2 = 1$, $\alpha_3 = 1$ and $\alpha_4 = 1.3$ in Eq. (1)

A. Decomposed Results

In our algorithm, there are three parts requiring manual interventions. First, we need users to select the shadow keypoint and set a wedge region to determine the number and the initial position of the lights, as shown in Fig. 4(c). Second, when we use the close-form matting method [44] to detect the prominent shadow, we needs users to input some scribbles for distinguishing shadow and non-shadow regions. Third, we require users to specify some samples in shadow and nonshadow regions, respectively, as shown in Fig. 8, which are used to calculate the illumination cues for each light.

Fig. 10 shows the decomposed results for a synthetic image with three light sources, which compares the decomposed result with ground truth image for each light. The synthetic images in our experiment are rendered using 3D modeling software Autodesk Maya2013. The three lights in Fig. 8 share different light color. The first and the second light are white light, and the color of the third light is pink. Using our illumination decomposition method, we get the decomposed image for each light, as shown the second row in the Fig. 8.

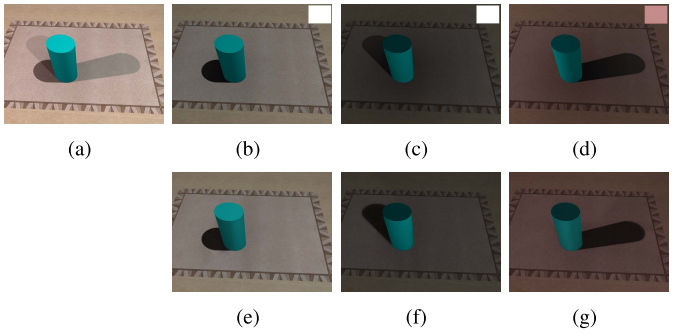


Fig. 10. Illumination decomposition results for a synthetic image with three light sources. (a) is the input image. (b), (c) and (d) are the ground truth decomposition images for the 1*st*, 2*ed*, 3*rd* light, respectively, and the color box in each ground truth image is the color of the corresponding light source. (e), (f) and (g) are our decomposed results for each light.

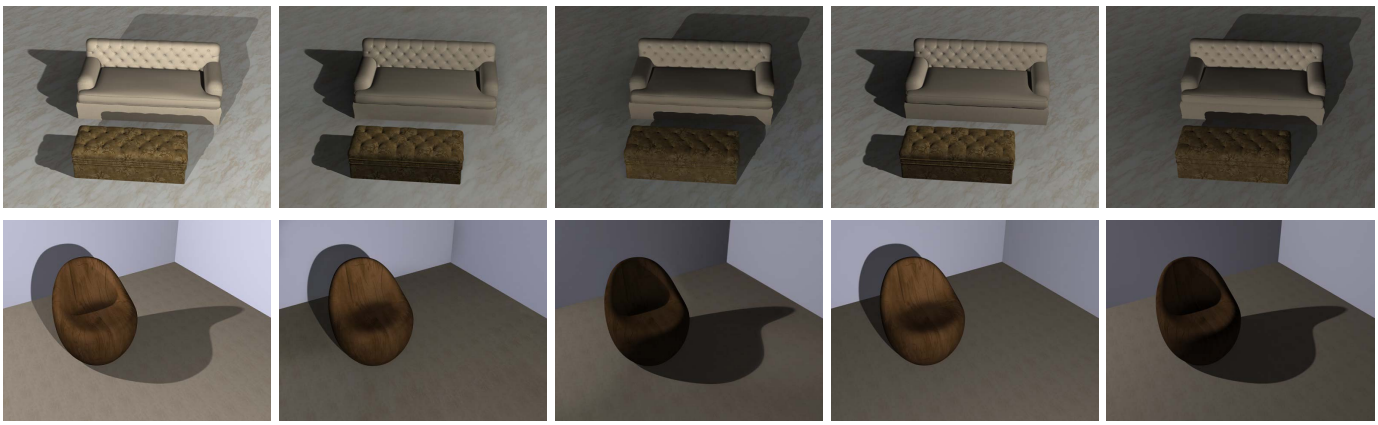
TABLE II
Comparing the Estimated Light Position With the Ground Truth for Images in Fig. 10 and Fig. 11

Image	Light source	Ground truth			The estimated position		
		x	y	z	x	y	z
Image 1	Light 1	1.065	-0.005	3.727	1.077	-0.006	3.659
	Light 2	1.527	0.010	0.114	1.485	0.009	0.123
	Light 3	-0.680	-0.016	0.005	-0.702	-0.015	0.046
Image 2	Light 1	1.650	-0.414	2.170	1.720	-0.432	2.300
	Light 2	-0.305	-0.433	2.275	-0.330	-0.478	2.327
Image 3	Light 1	0.750	-0.031	1.272	0.810	-0.036	1.325
	Light 2	-0.041	0.039	1.280	-0.052	0.400	1.331

Both the illumination and color of our results are very close to the ground truth images (the first row). In Table II, the data for Image 1 compare the estimated position for each light in Fig. 10 with ground truth position. The estimated positions using our method are close to the actual positions.

In Fig. 11, we perform our method on another two synthetic images. There are two light sources in each scene. For comparison purpose, we also present the ground truth image for each light source, as shown in Fig. 11(d, e). In the first image, the wooden chair cast shadows on both the floor and the wall, which have different materials. The second scene contains two objects: a sofa and a stool. One shadow from the stool is projecting on the sofa and the floor, and this shadow region contains two materials. Our coarse-to-fine strategy successfully estimates the decomposed image for each light in the scene, as shown in Fig. 11(b, c). Our decomposition results are visually similar to the ground truth images, and the hue and brightness are consistent with the ground truth. In our decomposed images, the texture details are effectively recovered, and the composed attached shadows are also visually natural. Table II compares the estimated position for images in Fig. 11 with ground truth position. Image 2 denotes image in the first row and Image 3 is image in the second row. The table shows that the estimated positions using our method approximates the actual positions well.

In Fig. 12 and Fig. 13, we present experiments for outdoor scenes with multiple lights at night. As shown in the first and the second column in Fig. 12, each image contains two directional lights. The first image has two separate objects, and includes multiple shadow overlapping regions. The second is



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

Fig. 11. Illumination decomposition results for synthetic images. (a) is input images. (b) and (c) are our image decomposition results for each light source. (d) and (e) are the ground truth for each light source.

Fig. 12. Illumination decomposed results for nature images with two lights. Images in first row are the input images. Images in the second and the third row are the corresponding decomposed results for each light.

a scene under the streetlight at night. Such kinds of images appear regularly among the outdoor scene at night. Fig. 11 is another outdoor image with four direct lights. There are three identical cylinders in the input image, but these cylinders have different light conditions. Besides a global light for these three cylinders, the first cylinder (from near to far) blocks three local lights and the second cylinder blocks a local light. The decomposed results in Fig. 12 and in Fig. 13 are both natural. In these real images, our illumination decomposed strategy works well and the details on the ground and the objects are effectively recovered.

We also present experiments for indoor images, as shown in Fig. 14 and the third and the fourth column in Fig. 12. The image in the fourth column in Fig. 12 has highlights. We first preprocess the highlights and remove highlights which are not generated by the current light, as shown in Fig. 7. Then, we perform the decomposed algorithm, and the decomposed results contain only highlights from the current light. The scene of the input image in Fig. 14 has three lights. Applying our illumination decomposed strategy, we get the decomposed results for each light, as shown in Fig. 14(b, c, d). With different light environment, the decomposed images have different intensity and the shadow regions are corresponding to different light position.

Fig. 15 shows the decomposed results for another two indoor images, which are close-range scenes. Due to the

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

Fig. 13. Illumination decomposed results for natural image with four lights. (a) is the input image. (b), (c), (d) and (e) are the decomposed results for each light.

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

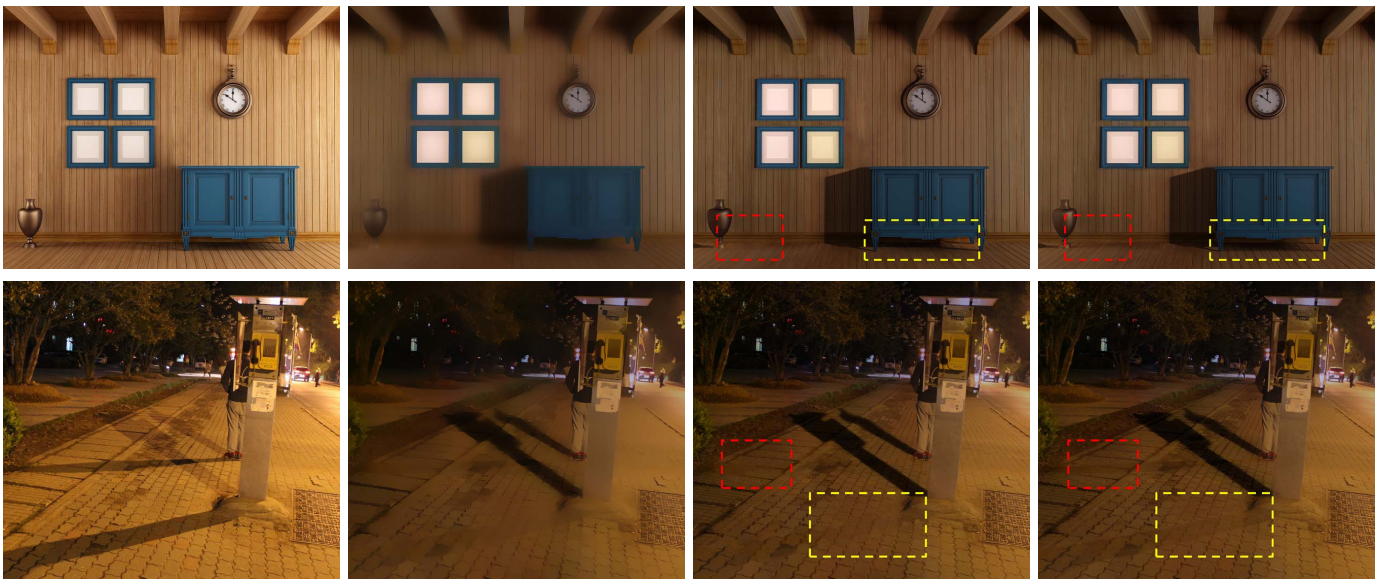
Fig. 14. Illumination decomposed results for natural image with three lights. (a) is the input image. (b), (c) and (d) are the decomposed results for each light.

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

Fig. 15. Illumination decomposed results for indoor images with highlights. (a) are input images and each scene has two direct lights. (b) and (c) are the decomposed results for each light without specular highlight preprocessing. The red boxes are the highlights. (d) and (e) are the decomposed results with highlights preprocessing.

occlusion between objects, the shadows are complex. The two input images in Fig. 15 have two lights and both contain clear highlights. Without specular highlights preprocessing, the highlights preserve in each of the decomposed results (Fig. 15(b, c)), where these results are actually not correct and the specular information does not match the light direction corresponding to the shadows. The incorrect highlights (red box) reduce the visual reality. In our system, with highlights preprocessing, the decomposed results for each light (Fig. 15(d, e)) are more natural, and the highlights for each corresponding light are correctly re-rendered.

In Fig. 16, we present two experiments for the different balance weights in our illumination decomposition optimization model (Eq. 12). In Eq.11, $\lambda_1 = 0$ measures the degree of the appearance smoothing and $\lambda_2 = 0$ measures the sharpness of the texture details. With $\lambda_2 = 0$, the detail-preserving item will not work in our system, which results in a smoothing result with detail blurring. As shown in Fig. 16(b), the surface textures in two results are both blurred compared with our results with $\lambda_2 = 1.2$. In our illumination decomposition model, we take the texture detail into account and use a weighting factor w_{ij} in Eq. 15 to discriminate the desirable and undesirable shadow boundaries. This makes our method produce visually convincing results. When using consistent $w_{ij} = \exp(-\frac{(I_i - I_j)^2}{2\sigma^2})$ in all regions, the weighting factor does not process the shadow boundaries separately. In this situation,



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

Fig. 16. Numerical analysis for parameters in illumination optimization model. (a) are two input images. (b) are the optimized results for one light with $\lambda_1 = 1.0$, $\lambda_2 = 0.0$. (c) are our optimization results with $\lambda_0 = 1.0$, $\lambda_1 = 1.2$. (d) are optimized results with consistent weighting factor w_{ij}.

the detail-preserving item works by keeping the gradient of the input image to the decomposition image in consistent way, the shadow boundaries not corresponding to the current light may cause undesirable artifacts in the final decomposed result, as shown the color box in Fig. 16(d). To address this problem, our definition for w_{ij} which distinguish the useful and undesirable gradient produces much better results with details recovered in the surface (as shown in Fig. 16(c)).

The time consumption of our method depends on the size of the input image and the size of the local neighborhood used in our system. Typically, for an image with size of 1074×691 (the first row in Fig. 12) and the local neighborhood with size of 13×13, it takes about 6 seconds for initial illumination estimation and takes 12 seconds for solving the optimization equation (Eq. 12).

B. Illumination Editing

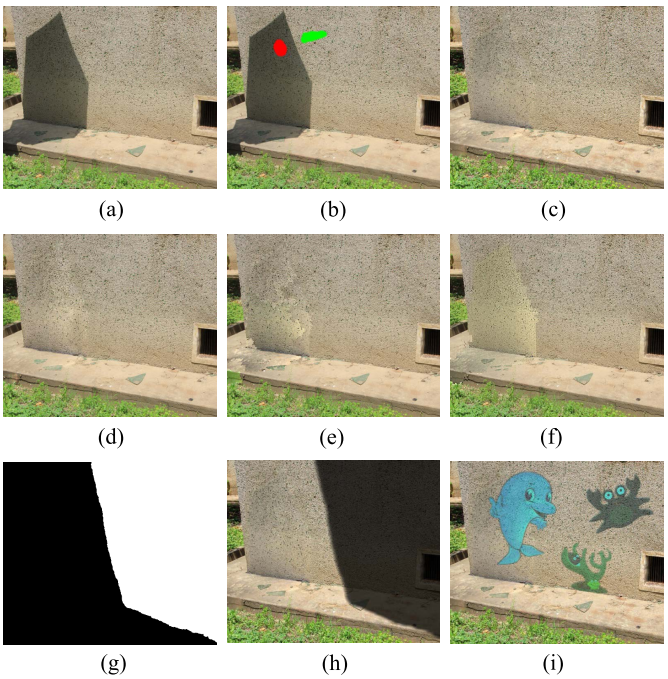
Our illumination decomposition method can be extended to image shadow editing. Like the illumination decomposition, we specify samples in shadow region and non-shadow region, as shown in Fig. 17(b). F_1 is the shadow sample, and G_1 is the nonshadow sample. Let μ_1, μ_2 be the average intensity of the two samples, respectively. The shadow-free intensity at pixel i in F_1 can be estimated as: $E_i = \frac{\mu_2 \times I_i}{\mu_1}$, $i \in F_1$. Similar to the section 4.2, after we have estimated the illumination for F_1, we can recover the illumination in other shadow regions by propagating the illumination from F_1 to other shadow regions. Let S be the shadow regions, and then the recovered intensity for the shadow regions is represented as:

$$E_i = \frac{\sum_{j \in N(i)} w_{ij} (\frac{I_i}{I_j} \times E_j)}{\sum_{j \in N(i)} w_{ij}}, \quad (i \in S) \cap (i \notin F_1)$$

$$flag[index(j)] = 1. \quad (17)$$

Fig. 17(c) is the shadow-free result using our proposed illumination processing system, which effectively recover the illumination and texture details in shadow region. We also present some shadow removal results produced by three existing methods [48], [49], [50], as shown Fig. 17(d, e, f). Compared with these three methods, our shadow removing result is more consistent with the surrounding environment.

We can also edit the shadow-free image using the composting equation (Eq. 11). As shown Fig. 17(h), we composite a new shadow mask (Fig. 17(g)) onto the shadow free image,



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 17. Shadow image editing. (a) is input image. (b) displays samples in image, red region is shadow sample, green region is non-shadow sample. (c) is our shadow removal result. (d) is shadow removal result of [Ling *et al.* 2015]. (e) is shadow removal result of [Xiao *et al.* 2013]. (f) is shadow removal result of [Shor and Lischinski 2008]. (g) is new shadow mask. (h) is shadow editing result. (i) is object composting result.

which produces a shadow on the wall and ground. Similarly, given a color image, we can blend the color image into the wall, producing a wall painting, as shown in Fig. 17(i).

Limitations: Our method processes the image with multiple lights using the information of shadows. But, if the pixel intensities in shadow regions are closed to zero, our illumination cues from shadows will fail to work, which will affect the subsequent decomposed process. Furthermore, for image with heavy noises in shadow regions, the decomposed results may also contain some noises in these regions. Another limitation is that, the illumination cues are obtained from some user interactions for specifying samples. In practice, an automatic illumination decomposition method is desirable.

VII. Conclusion and Future Work

In this paper, we have presented a coarse-to-fine illumination decomposition system for image with multiple light sources. We first reconstruct the light information using the estimated geometry structure of the scene. Then, we estimate the initial illumination decomposed result for each light with the extracted illumination cues from shadows. Finally, we develop a light-aware optimization model to optimize the coarse decomposition and produce finer decomposed results for each light. Although the decomposed results may not be physically accurate, the results are visually pleasing with texture details effectively recovered. Finally we also have extended our method to some applications, such as shadow removal and shadow editing. We believe our method can facilitate other image editing operations, such as material recoloring and illumination-aware object composting.

In our system, we need to use some interactions to obtain the illumination cues from the specified sample regions. In the future, based on the sophisticated illumination environment analysis, we would like to develop an automatic method to perform the illumination decomposition for image with multiple lights. In addition, our method can get the information of image illumination and shadow, which may be helpful for shading image generation. So extending these ingredients into intrinsic image decomposition or shape from shading would be an interesting future work. Finally, it is also an interesting research direction to extend our method to decompose the video illumination. For video data, the illumination may be dynamic, and we need to estimate the depth map for the video data. For the input video streaming, the depth map for each frame should be estimated in real time, and the depth maps should be spatial-temporally coherent. Similar to image case, this research direction is also a challenging task.

Acknowledgment

The authors would like to thank all the anonymous reviewers for their insightful comments and constructive suggestions.

References

[1] R. Carroll, R. Ramamoorthi, and M. Agrawala, "Illumination decomposition for material recoloring with consistent interreflections," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 76–79, 2011.

[2] B. Dong, Y. Dong, X. Tong, and P. Peers, "Measurement-based editing of diffuse albedo with consistent interreflections," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–11, 2015.

[3] Q. Zhang, C. Xiao, H. Q. Sun, and T. Feng, "Palette-based image recoloring using color decomposition optimization," *IEEE Trans. Image Process*, vol. 26, no. 4, pp. 1952–1964, Apr. 2017.

[4] K. Karsch, V. Hedau, D. Forsyth, and D. Hoiem, "Rendering synthetic objects into legacy photographs," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 61–64, 2011.

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

[6] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 935–944, Jul. 2006.

[7] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 15–19, 2008.

[8] I. Boyadzhiev, K. Bala, S. Paris, and F. Durand, "User-guided white balance for mixed lighting conditions," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 439–445, 2012.

[9] H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision System*, A. Hanson and E. Riseman, Eds. New York, NY, USA: Academic, 1978, pp. 3–26.

[10] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *Proc. 12th ECCV*, 2012, pp. 57–70.

[11] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 57, no. 1, pp. 1–11, 1967.

[12] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1459–1472, Sep. 2004.

[13] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *Proc. IEEE CVPR*, Jan. 2008, pp. 1–7.

[14] Q. Zhao, P. Tan, Q. Dai, and L. Shen, "A closed-form solution to retinex with nonlocal texture constraints," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1437–1444, Jul. 2012.

[15] L. Shen, C. Yeo, and B.-S. Hua, "Intrinsic image decomposition using a sparse representation of reflectance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2904–2915, Dec. 2013.

[16] G. Elena, M. Adolfo, L. Lopez-Moreno, and G. Diego, "Intrinsic images by clustering," *Comput. Graph. Forum*, vol. 13, no. 4, pp. 1415–1424, 2012.

[17] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–12, 2014.

[18] S. Bi, X. Han, and Y. Yu, "An $L1$ image transform for edge-preserving smoothing and scene-level intrinsic decomposition," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–78, 2015.

[19] A. Bousseau, S. Paris, and F. Durand, "User assisted intrinsic images," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 89–97, 2009.

[20] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Proc. IEEE CVPR*, Jun. 2011, pp. 3481–3487.

[21] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 425–436, Apr. 2013.

[22] N. Bonneel, K. Sunkavalli, J. Tompkin, D. Sun, S. Paris, and H. Pfister, "Interactive intrinsic video editing," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–10, 2014.

[23] T. Luo, J. Shen, and X. Li, "Accurate normal and reflectance recovery using energy optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 2, pp. 212–224, Feb. 2015.

[24] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Proc. IEEE ICCV*, Dec. 2013, pp. 241–248.

[25] P. Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, "Coherent intrinsic images from photo collections," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 439–445, 2012.

[26] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos, "A theory of inverse light transport," in *Proc. IEEE ICCV*, Oct. 2005, pp. 1440–1447.

[27] J. Bai, M. Chandraker, T. T. Ng, and R. Ramamoorthi, *A Dual Theory of Inverse and Forward Light Transport*. Berlin, Germany: Springer, 2010.

[28] M. O'Toole, R. Raskar, and K. N. Kutulakos, "Primal-dual coding to probe light transport," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 13–15, 2012.

[29] K. Hara, K. Nishino, and K. Ikeuchi, "Determining reflectance and light position from a single image without distant illumination assumption," in *Proc. IEEE ICCV*, vol. 1. Oct. 2003, pp. 560–567.

[30] Y. Zhang and Y. H. Yang, "Illuminant direction determination for multiple light sources," in *Proc. CVPR*, vol. 1. 2000, pp. 269–276.

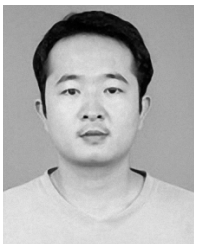
[31] B. Mercier, D. Meneveaux, and A. Fournier, "A framework for automatically recovering object shape, reflectance and light sources from calibrated images," *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 77–93, 2007.

[32] K. Hara, K. Nishino, and K. Ikeuchi, "Mixture of spherical distributions for single-view relighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 25–35, Jan. 2008.
[33] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Estimating natural illumination from a single outdoor image," in *Proc. IEEE ICCV*, Sep. 2009, pp. 183–190.
[34] E. Kee, J. F. O'brien, and H. Farid, "Exposing photo manipulation with inconsistent shadows," *ACM Trans. Graph.*, vol. 32, no. 3, pp. 167–186, 2013.
[35] E. Kee, J. F. O'brien, and H. Farid, "Exposing photo manipulation from shading and shadows," *A CM Trans. Graph.*, vol. 33, no. 5, pp. 1935–1946, 2014.
[36] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination from shadows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 290–300, Mar. 2003.
[37] T. Okabe, I. Sato, and Y. Sato, "Spherical harmonics vs. haar wavelets: Basis for recovering illumination from cast shadows," in *Proc. IEEE CVPR*. vol. 1. Jun. 2004, pp. I-50–I-57.
[38] R. Ramamoorthi, M. Koudelka, and P. Belhumeur, "A Fourier theory for cast shadows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 288–295, Feb. 2005.
[39] M. Xue, L. Haibin, and D. W. Jacobs, "Illumination recovery from image with cast shadows via sparse representation," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2366–2377, Aug. 2011.
[40] H. Shen and Q. Cai, "Simple and efficient method for specularity removal in an image," *Appl. Opt.*, vol. 48, no. 14, pp. 2711–2719, 2009.
[41] N. Neverova, D. Muselet, and A. Trémeau, "Lighting estimation in indoor environments from low-quality images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 380–389.
[42] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5162–5170.
[43] Y. Li, H. Lu, and H.-H. Shum, "Multiple-cue illumination estimation in textured scenes," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1366–1373.
[44] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
[45] C. Xiao, M. Liu, D. Xiao, Z. Dong, and K.-L. Ma, "Fast closed-form matting using a hierarchical data structure," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 1, pp. 49–62, Jan. 2014.
[46] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 646–653, 2006.
[47] T.-P. Wu, C.-K. Tang, M. S. Brown, and H.-Y. Shum, "Natural shadow matting," *ACM Trans. Graph.*, vol. 26, no. 2, 2007, Art. no. 8.
[48] Z. Ling, Q. Zhang, and C. Xiao, "Shadow remover: Image shadow removal based on illumination recovering optimization," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4623–4636, Nov. 2015.
[49] C. Xiao, D. Xiao, L. Zhang, and L. Chen, "Efficient shadow removal using subregion matching illumination transfer," *Comput. Graph. Forum*, vol. 32, no. 7, pp. 421–430, 2013.
[50] Y. Shor and D. Lischinski, "The shadow meets the mask: Pyramid-based shadow removal," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 577–586, 2008.

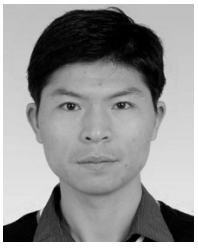
Ling Zhang received the B.Sc. degree in computer science and technology from Wuhan Donghu University in 2009 and the M.Sc. degree in instructional technology from Central China Normal University in 2012. She is currently pursuing the Ph.D. degree with the School of Computer, Wuhan University, China. Her current research interests include image and video editing, and computational photography.

Qingan Yan received the B.Sc. degree in computer science from Hubei University for Nationalities, China, in 2008, and the M.Sc. degree in computer science from the Southwest University of Science and Technology, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Computer, Wuhan University. His current research interests include 3D modeling, matching, and scene and shape analysis.

Zheng Liu received the B.Sc. and M.Sc. degrees in computer science and technology from the China University of Geosciences in 2006 and 2009, respectively, and the Ph.D. degree in instructional technology from Central China Normal University in 2012. From 2013 to 2014, he held a post-doctoral position with the School of Mathematical Sciences, University of Science and Technology of China. He is currently a Lecturer with the National Engineering Research Center of Geographic Information System, China University of Geosciences. His main interests include digital geometry processing and image and video processing.



Hua Zou received the B.Sc. and M.Sc. degrees from the Xi'an University of Architecture and Technology in 2001 and 2005, respectively, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2009. From 2015 to 2016, he visited the University of Pittsburgh. He is currently an Associate Professor with the School of Computer, Wuhan University, China. His main interests include image and video processing and pattern recognition.



Chunxia Xiao (M'11) received the B.Sc. and M.Sc. degrees from the Mathematics Department, Hunan Normal University, in 1999 and 2002, respectively, and the Ph.D. degree from the State Key Laboratory of CAD&CG, Zhejiang University, in 2006. From 2006 to 2007, he held a post-doctoral position with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, and from 2012 to 2013, he visited the University of California at Davis. He is currently a Professor with the School of Computer, Wuhan University, China. His main interests include computer graphics, computer vision, and machine learning.