ORIGINAL ARTICLE

Shading-aware shadow detection and removal from a single image

Xinyun Fan[1] · Wenjun Wu[1] · Ling Zhang[2] · Qingan Yan[3] · Gang Fu[1] · Zipei Chen[1] · Chengjiang Long[4] · Chunxia Xiao[1]

Abstract

Shadow removal is a challenging problem due to its sensitivity to lighting and material conditions. In this paper, we propose a shading-aware shadow processing algorithm, which can automatically detect and remove complex shadows from a single color image. Our framework consists of two key steps. We firstly conduct a shadow-preserving filter upon the image which will effectively remove the image texture while preserving the shadow and shading information. Shadow regions are estimated by establishing a confidence map from the filtered image incorporating depth cue. We then develop a shading-aware optimization framework to remove shadows and recover shading in these regions. The extensive experimental results show that the proposed algorithm produces visually compelling results in a series of challenging images and it can handle complex shadows in both indoor and outdoor scenes. Quantitative and qualitative comparisons with current state-of-the-art methods strongly demonstrate the efficacy of our proposed approach.

Keywords Shadow removal · Complex shadow · Image processing

✉ Chunxia Xiao
  cxxiao@whu.edu.cn

  Xinyun Fan
  xinyunfan@whu.edu.cn

  Wenjun Wu
  815012933@qq.com

  Ling Zhang
  zhling@wust.edu.cn

  Qingan Yan
  qingan.yan@jd.com

  Gang Fu
  xyzgfu@gmail.com

  Zipei Chen
  2019202110105@whu.edu.cn

  Chengjiang Long
  cjfykx@gmail.com

[1] School of Computer, Wuhan University, Wuhan 430072, China

[2] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China

[3] JD.com American Technologies Corporation, Mountain View, CA 94043, USA

[4] JD Digits, Mountain View, CA 94043, USA

1 Introduction

Shadow is a ubiquitous natural phenomenon in our daily life. Although shadows can provide useful clues for illumination estimation [46], scene depiction [20] and object shapes [32], shadows also degrade the performance of some applications, such as object recognition [7], object tracking [29] and intrinsic image decomposition [25]. Therefore, it is a fundamental problem to detect and remove shadows from single images and will definitely be beneficial for computer vision and graphics communities.

Shadow removal involves three main challenges. First, for the image with complex shadows like a surface with both soft and hard shadow, accurate shadow detection is challenging. Second, there are usually texture details losing on hard shadow boundaries, which will induce visual artifacts on these boundaries during shadow removing [40,48]. Finally, to obtain visually consistent shadow removal results, the shading information should be preserved in the shadow-free image [42].

To overcome the above challenges, we propose an automatic shadow detection and removal method by jointly exploring color cues as well as depth information. First, based on the observation that shadows essentially appear as smooth and continuous regions, we develop a shadow-preserving
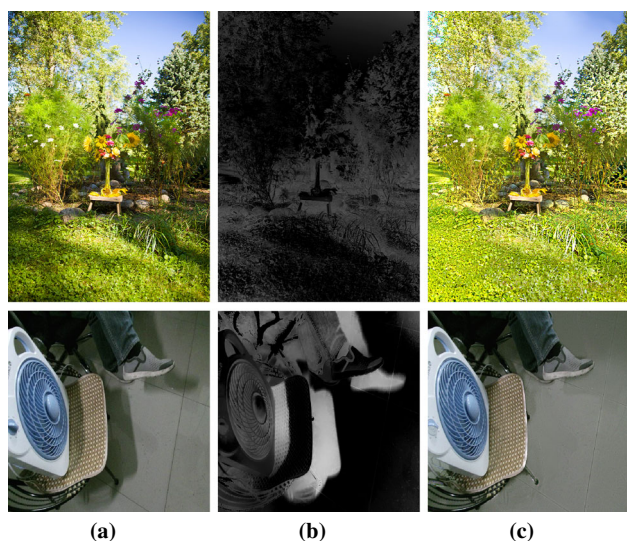
Fig. 1 Our visualization results of shadow detection and removal. From left to right are **a** input images, **b** shadow confidence maps and **c** shadow removal results.

filter which effectively removes the image texture while preserving the shadow and shading information. This filter enables to estimate more accurate shadow confidence map and benefits better shadow detection results. Second, to effectively detect the complex shadow including both hard/soft shadow and local/global shadow, we develop a shadow confidence propagation scheme for shadow detection, which works well on detecting shadows in various environments. Finally, based on the shadow confidence, we present a novel shading-aware shadow removal optimization model. It can effectively remove the shadows while preserving the shading information of the shadow regions well. Moreover, the chromaticity and texture details under the shadow regions are well recovered, as illustrated in Fig. 1.

In summary, our major contributions are threefold as follows:

- We propose a novel shadow-preserving texture filter. Such a filter is immune to texture, boundary and noise, and enhances shadow and shading information which enables to better shadow confidence estimation.
- We design a shadow confidence propagation scheme, which propagates the local shadow boundary confidence to the global scene adaptively, enabling to detect both local and global complex shadows.
- We propose a shading-preserving shadow removal framework which can effectively remove the complex shadow as well as recovering the shading, chromaticity and texture details under the shadow regions.

Our proposed method have been demonstrated on a series of challenging images, including indoor and outdoor scenes

with hard and soft shadows. We also compare it with current state-of-the-art methods and show its superior performance in both shadow detection and removal.

2 Related work

Shadow detection methods Several user assistance-based shadow detection methods [2,14,48] work well on simple shadow scenes, whereas they often require extra tedious user interaction for complex scenarios. Many automatic shadow detection methods have been proposed. For example, Lalonde et al. [22] built a shadow detector for photographs of outdoor scenes, and this method focused on detecting ground shadow. Guo et al. [15] compared pairs of regions and determined whether they were in the same illumination conditions. Nevertheless, this detector may group soft shadows with non-shadow regions and may fail in the case of multiple light sources. Shadow detection for complex shadows is still a challenging problem.

Recently, deep learning based on convolutional neural network has achieved great success in computer vision community and also has been exploited in shadow detection [17,24,31,49]. For example, Nguyen et al. [31] detected shadow using conditional generative adversarial networks. Hu et al. [17] analyzed image context and detected shadow in direction-aware manner. To obtain satisfying results, deep learning methods depend on large and good training data. For some complex scenes, the training data are difficult to collect, and the results are not satisfactory.

Shadow removal methods Traditional shadow removal methods were proposed mainly based on gradient domain manipulation [11,26,30]. To receive satisfactory results, these methods depend on accurate shadow edges detection for shadow-free image reconstruction. Shadow matting was also exploited in shadow detection and removal [6,14,23,39], while these methods do not preserve the shading well in the shadow-free image. Inspired by the color transfer theory [34], several shadow removal methods have been proposed based on illumination or color transferring [35,41,47,48]. To generate good results, for each patch in the shadow regions, these methods need to find corresponding non-shadow regions for illumination transferring.

Several deep learning-based shadow removal methods have also been proposed [9,18,33,37,45]. Ding et al.[9] removed the shadow in an attentive recurrent way, and Wei et al. [38] preprocessed the shadow image with inpainting method. To address the lack of training data, Hu et al. [18] adopted unpaired data to diverse the samples, and Cun et al. [8] designed a shadow matting generative adversarial network to synthesize realistic shadow images. Obviously, the performance of those learning-based methods heavily depend on the training dataset. However, for images with
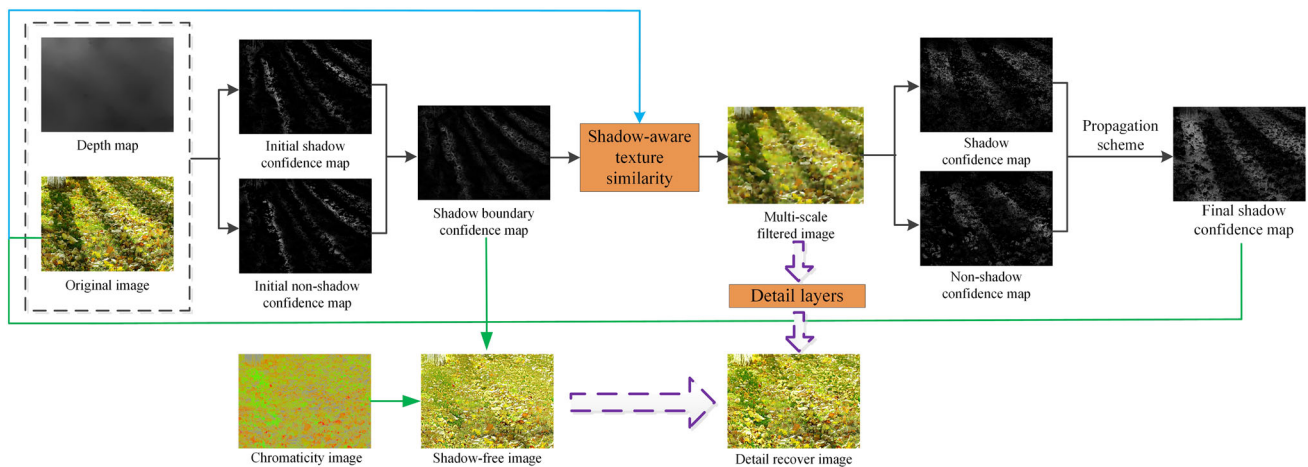
Fig. 2 Overview of the proposed shadow detection and removal

complex shadows, it is difficult to construct the ground-truth shadow-free images, which makes these methods not sufficiently adept at processing complex shadows. In contrast, the optimization-based methods do not depend on those training dataset and have better generalization ability to handle complex scenes.

As the depth acquisition devices become more and more popular, some researchers have resorted to depth cues for shadow removal [42] and intrinsic image decomposition [3,4,16,19]. Xiao et al. [42] applied depth information to remove shadows in RGB-D image and produced impressive shadow detection and removal results. However, this method is not immune to image texture when performing shadow detection and does not work well for preserving shading in shadow removal results. In this work, we also fall into this category with depth cues. To address existing defects, we make a shadow-preserving texture filter before shadow detection and design a shading-preserving shadow removal framework to get the final result.

3 Shadow detection and removal

Our goal in this paper is to accurately detect the shadows in a single image as well as to effectively remove them while keeping the texture and shading in it. To this end, we first propose an effective shadow detection algorithm that utilize a shadow-preserving filter to effectively remove the textures while preserving the shadow and shading information, and shadow regions are estimated by establishing a confidence map from the filtered image incorporating the depth map (Sect. 3.1). Then, we develop a shading-aware optimization algorithm to remove the shadows and recover the shading in these regions. The details of the image will be recovered by

adding the detail layers in a weighted average way (Sect. 3.2). The framework of the overall algorithm is shown in Fig. 2.

3.1 Automatic shadow detection

Natural photographs usually contain complex textures which will affect the accuracy of shadow detection. Inspired by [5,43], we propose a shadow-preserving bilateral filter for shadow detection. The pipeline of our proposed automatic shadow detection is shown in Fig. 3.

For a complex image, the depth map of the scene would be helpful for shadow confidence estimation. To obtain more accurate shadow confidence map, we apply the depth information of the image into our method. We can acquire the depth map using low-cost depth sensors, such as MS Kinect, or via learning-based methods. Recently, many image depth estimation methods have been proposed [12,13,21,27,43]. In this paper, for input image without depth map, we apply the method [12] to estimate the depth map.

3.1.1 Shadow-preserving texture filter

The proposed shadow-preserving texture filter is defined as

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega_p} f(\|q - p\|) g(\|S_q - S_p\|) I_q, \tag{1}$$

where p represents the current pixel, Ω_p is the local neighborhood of p, q represents a pixel in Ω_p, S is the shadow-aware texture measure of input image I and k_p is a normalizing parameter. The spatial kernel f and the range kernel g are Gaussian functions. This is a modification of the bilateral texture filter [5] with the shadow-aware texture similarity instead of structure-aware texture similarity. With the guid-
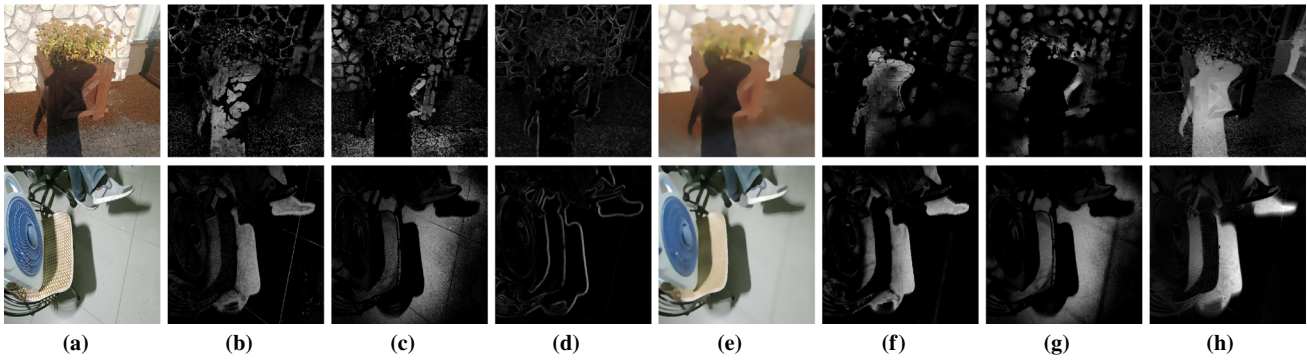
Fig. 3 Overview of our shadow detection. Given an input image (**a**), we first compute the initial shadow confidence (**b**) and the initial non-shadow confidence (**c**) with the method in [42], and further estimate the initial shadow boundary (**d**). Then we estimate the shadow confidence map (**f**) and the non-shadow confidence map (**g**) of the filtered image

(**e**) obtained by our shadow-preserving texture filter. Finally, we design a structure-aware confidence propagation scheme to interpolate (**f**) and (**g**) around the boundary to other pixels, and obtain the final shadow confidence map (**h**)

ance of S, our filter can preserve the shadow boundaries, as shown in Fig. 4.

We assume that texture signal usually has smaller amplitude than shadow boundary. So we first find the patches which contain the shadow boundary, and compute the likelihood of these patches Ω_p via estimating its shadow range $\Psi(\Omega_p) = C_B^{\max}(\Omega_p) - C_B^{\min}(\Omega_p)$. C_B is the shadow boundary confidence and it will be introduced in the next section. $C_B^{\max}(\Omega_p)$ and $C_B^{\min}(\Omega_p)$ denote the maximum and the minimum shadow boundary confidence in Ω_p. Intuitively, a patch with the maximum shadow range means a maximal probability of containing the shadow boundary. The shadow-aware texture measure S_p at p is:

$$S_p = C_B^{\mathrm{avg}}(\Omega_q), \tag{2}$$

where the pixel q has the largest $\Psi(\Omega_q)$ among the neighbor pixels of p. $C_B^{\mathrm{avg}}(\Omega_q)$ is the average shadow boundary confidence of the region Ω_q, and Ω_q is the local neighborhood of q.

3.1.2 Shadow confidence estimation

For each pixel p, the shadow confidence is related to the feature similarity between the pixel and its neighbor pixels q, which is defined as:

$$\alpha_{pq} = \alpha_{pq}^c \cdot \alpha_{pq}^n \cdot \alpha_{pq}^d, \tag{3}$$

where α_{pq}^c, α_{pq}^n and α_{pq}^d represent the similarity of chromaticity, normal and spatial location between p and q. $\alpha_{pq}^c = \exp(-\frac{\|\mathrm{ch}(I_p) - \mathrm{ch}(I_q)\|_2}{2\sigma_{\mathrm{ch}}^2})$, $\alpha_{pq}^d = 1 - \frac{\|\bar{p} - \bar{q}\|}{\max\limits_{q \in \Omega_p} \|\bar{p} - \bar{q}\|}$, and α_{pq}^n is estimated by solving the following optimization function:

$$\underset{\{\alpha_{pq}^n\}}{\mathrm{argmin}} \sum_{p \in I} \left\| n(p) - \sum_{q \in \Omega_p} \alpha_{pq}^n n(q) \right\|^2.$$

Here, $\mathrm{ch}(I_p)$, $n(p)$ and $\bar{p}$ are the chromaticity, normal and 3D spatial location of the pixel p, σ_{ch} is a positive parameter controlling the sensitivity of the similarity (typically is set to 0.1), and Ω_p denotes the local neighborhood of p.

With the feature similarity between p and its neighbors, we calculate the corresponding weighted average intensity $m_p = \frac{1}{\sum_{q \in \Omega_p} \alpha_{pq}} \sum_{q \in \Omega_p} \alpha_{pq} I_q$ and then estimate the initial shadow confidence C_S and the non-shadow confidence C_U using the method in [42]. The functions are as follows:

$$C_S = \frac{1 - \exp\left(-\frac{\max(m_p - I_p, 0)^2}{2\sigma^2}\right)}{|\Omega_p|} \sum_{q \in \Omega_p} \alpha_{pq}, \tag{4}$$

$$C_U = \frac{1 - \exp\left(-\frac{\max(I_p - m_p, 0)^2}{2\sigma^2}\right)}{|\Omega_p|} \sum_{q \in \Omega_p} \alpha_{pq}. \tag{5}$$

The visual maps of these two variables are shown in Fig. 3b, c, respectively. The shadow boundary confidence C_B (Fig. 3d) can be obtained by computing the windowed total variation and windowed inherent variation with C_S and C_U.

The shadow-preserving filtered image (Fig. 3e) effectively removes the texture and noise. We can estimate more accurate shadow confidence C_S and non-shadow confidence C_U with the shadow-preserving filtered image, and the results are illustrated in Fig. 3f, g.

3.1.3 Shadow confidence optimization

Due to the properties of feature similarity, the shadow confidence C_S is better estimated around the shadow boundaries.
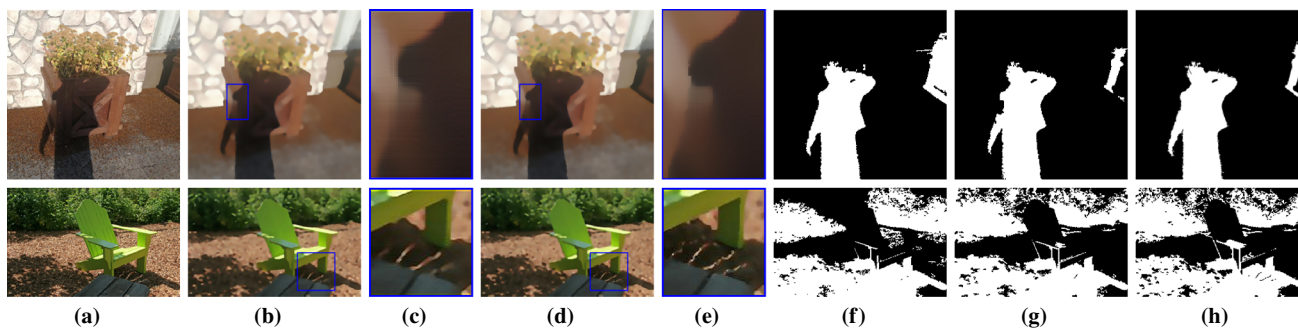
Fig. 4 Image filtering comparisons. From left to right are **a** input images, **b** result using bilateral texture filter [5], **c** region of interest marked in blue in **b**, **d** result using multi-scale shadow-preserving texture filter, **e** region of interest marked in blue in **d**, **f**–**h** shadow detection results on **a**, **b** and **d**, respectively

So as to enhance the estimation of the rest shadow regions far from the shadow boundaries, we apply a structure-aware confidence propagation to interpolate the confidence C_S and C_U around the boundary to other regions, and get a more comprehensive shadow confidence map.

Let n be the number of pixels in the image. The shadow confidence s_i of pixel p_i is then obtained by minimizing the following function:

$$\sum_{i=1}^{n} C_S(p_i)(s_i - 1)^2 + \sum_{i=1}^{n} C_U(p_i)s_i^2 + \sum_{i,j} w_{ij}(s_i - s_j)^2. \tag{6}$$

The first term encourages the pixel p_i with large shadow confidence $C_S(p_i)$ to get a large value (close to 1). The second term enables the pixel p_i with large non-shadow confidence $C_U(p_i)$ to take a small value (close to 0). The last term is a smooth term. For every adjacent pixel pair (i, j), the weight w_{ij} is the element of matting Laplacian matrix [1]. As the filtered image J is piecewise smoothed with no oscillating texture variations, we can effectively propagate the shadow confidence and non-shadow confidence using the structure of J, and obtain higher-quality shadow confidence.

As shown in Fig. 3h, the optimized shadow confidence map C_S is more accurate than the initial one. The shadow regions are more highlighted, and the gradient information around shadow boundary is preserved as well, which will benefit our shadow removal in the next step.

To further remove the effects of noise and texture structures while detecting the shadows, we propose a multi-scale shadow confidence estimation method. In each filtering, by modifying the filter parameter (window size Ω and standard deviation σ_s), the texture is progressively smoothed, and shadow boundary is progressively refined. The proposed method is summarized in Algorithm 1.

Algorithms 1: Multi-scale image shadow detection

Input: Image I, Iterative n_{iter}, Parameter Ω, σ_s
Output: Shadow-preserved texture filter image J, Shadow boundary map C_B, Shadow confidence map C_S.
$J^0 = I$
for $i = 1$ to n_{iter} **do**
    $C_B^i \leftarrow$ Shadow boundary detection on J^{i-1}
    $J^i \leftarrow$ Shadow preserving texture filter on J^{i-1}
        guided by C_B^i
    $\Omega, \sigma_s \leftarrow 0.5 \times \Omega, \sigma_s$
end
$C_S \leftarrow$ Shadow confidence map computing on J^{iter}

The advantages of the proposed shadow detection scheme are as follows: (1) Our method is more immune to texture, noise, and receives better shadow and shading information which enables to better shadow detection results; (2) with the local shadow boundary confidence and the global shadow propagation strategy, our method can detect not only local shadow areas, but more complex shadows in the scene. Figures 5 and 6 show the shadow confidence map comparisons with the method in [42].

3.2 Shading-aware shadow removal

3.2.1 Shadow removal

Like [39], the shadow factor can be modeled in the form of:

$$\beta = I/F, \tag{7}$$

where F is the shadow-free image and β is a three-channel fractional shadow factor each in $[0, 1]$ for scaling the respective color channel. In this paper, we use the normal information from depth for shadow detection and removal. Our aim is to achieve the shadow-free image F-preserving shading and the shadow factor β excluding shading, as shown in Fig. 6.
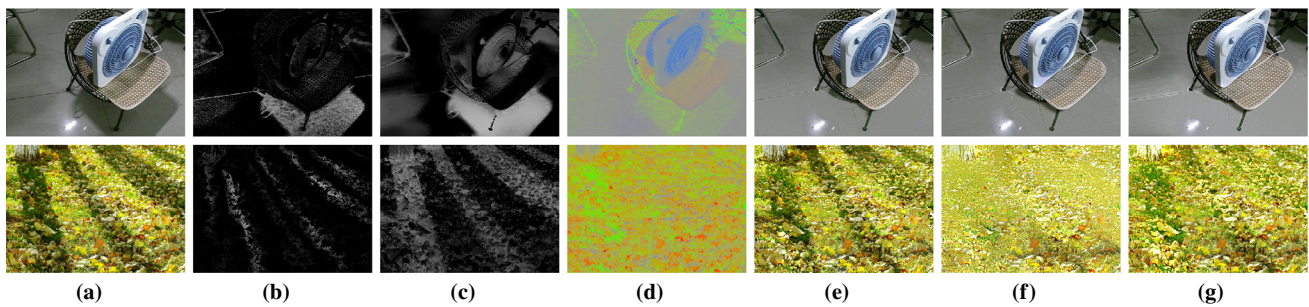
Fig. 5 Effect of the chromaticity term. From left to right are **a** input images, **b**, **c** shadow confidence maps of method [42] and our proposed method, **d** chromaticity images of **a**, **e** shadow removal results of method [42] with our shadow confidence maps **c** as input, and **f**, **g** our shadow removal results without and with chromaticity term, respectively
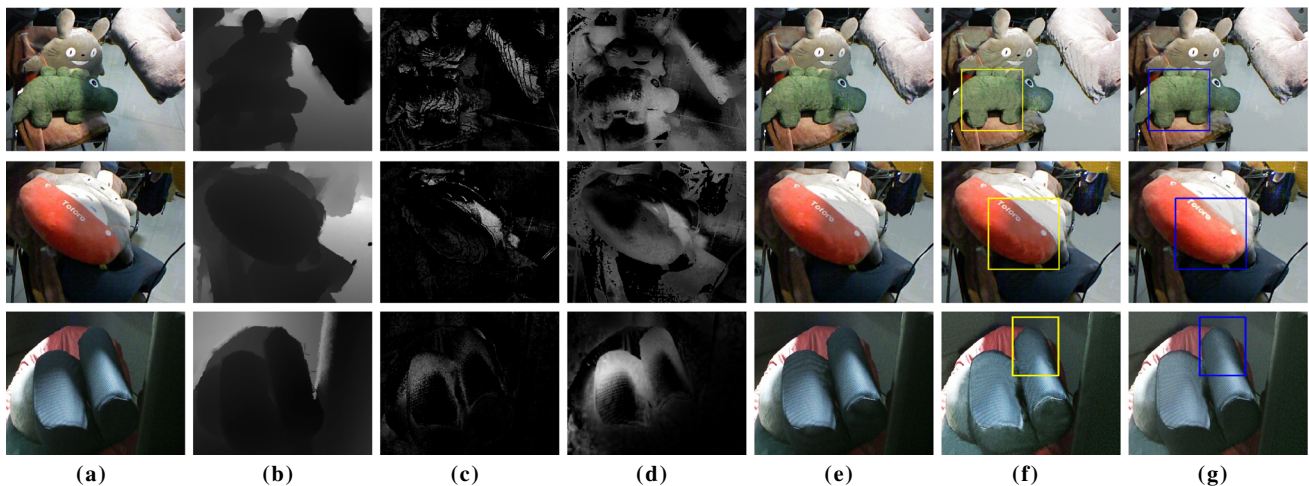

Fig. 6 Effect of the shading-preserving term. From left to right are **a** input images, **b** depth maps, **c**, **d** shadow confidence maps of method [42] and our proposed method, **e** shadow removal results of method [42] with our shadow confidence maps **d** as input, **f**, **g** our shadow removal results without and with the shading-preserving term, respectively. Note that the yellow boxes are the part with poor shading, while the same place in the blue boxes preserve the shading well

To estimate the shadow-free image F and the corresponding shadow factor β, we propose the following energy equation for shadow removal:

$$E(F, \beta) = E_{\text{data}}(F, \beta) + \lambda_1 E_{\text{smooth}}(F, \beta)$$
$$+ \lambda_2 E_{\text{chro}}(F) + \lambda_3 E_{\text{const}}(\beta). \quad (8)$$

This energy model contains four terms: data term E_{data}, shading-preserving smoothing term E_{smooth}, chromaticity term E_{chro} and constant term E_{const}. The balanced weights λ_1, λ_2 and λ_3 are set to 1, 0.5 and 1, respectively, in our experiments.

Data term As we aim to decompose the input image into a product of shadow-free and shadow factor components, we enforce this as a soft constraint via the data fitting term E_{data}. We assume monochromatic, white illumination and apply the fitting constraint to per color channel, i.e., $I_c \approx F_c \cdot \beta_c$, $c \in \{R, G, B\}$. To make the decomposition more robust to

white illumination deviations, we use per-channel weights w_c in the constraint:

$$E_{data}(x) = \omega_{iw}(x) \sum_{c \in \{R, G, B\}} \omega_c \cdot \|I_c - F_c \cdot \beta_c\|^2, \quad (9)$$

where $\{\omega_R, \omega_G, \omega_B\} = \{0.299, 0.587, 0.114\}$. In addition, based on the observation [28] that low-intensity pixels are more sensitive to the image noise, and pixels with higher intensity provide more decomposition reliability, we incorporate the image intensity weight $\omega_{iw}(x) = 1 - \omega_{\text{intensity}} \cdot (1 - |I(x)|)$ in our data term, where $|I(x)|$ is the image intensity and $\omega_{\text{intensity}}$ is the adjustable balance weight.

Shading-preserving smoothing term To obtain visually realistic result of shadow removal, the shading component should be preserved in the shadow-free image F. Our basic assumption is that pixels with similar features, including

chromaticity, normal and spatial locations, are likely to have the same color or illumination. Let R_s be the shadow regions containing neighboring pixels, and two pixels p and q with a large similarity α_{pq} tend to have the same shadow-free image F. We define the shading-preserving term as:

$$E_{sp} = \sum_{p \in R_s} C_S(p) \cdot \| F_p - \sum_{q \in \Omega_p} \alpha_{pq} \cdot F_q \|^2, \tag{10}$$

where $C_S(p)$ is the shadow confidence for pixel p and Ω_p denotes the local spatial neighbors of p.

When $C_S(p)$ has a large value, which indicates the hard shadow, the smooth constraint on shadow-free image should be more enforced; thus, the recovered illumination could vary with the scene shape and produce more realistic results.

We also define shadow boundary-aware smoothness constraint on β. Inspired by the Retinex theory, which have demonstrated that total variation has good performance in promoting illumination smoothness, we adopt the relative total variation (RTV) [44] for producing smooth β. We define the shadow map smoothness regularizer as follows:

$$E_{sm} = \sum_{p \in R_s} (1 - |C_B(p)|)(H(\beta_p) + V(\beta_p)), \tag{11}$$

where $H(\beta_p)$ and $V(\beta_p)$ denote the horizontal and vertical relative total variation (RTV) measure. In this equation, when $|C_B(p)|$ has a large value, which indicates the shadow boundary, the shading smoothness should be less enforced.

With the above smoothing constraints on both shadow-free image F and the shadow matte β, the smoothing constraint term is defined as:

$$E_{\text{smooth}} = E_{sp} + E_{sm}. \tag{12}$$

Chromaticity term We assume that the chromaticity of the input image is not altered by illumination effects such as shading and shadows [10]. In this case, the chromaticity of the unknown shadow-free image F should be the same as that of the input image. With this assumption, we define the following soft constraint as:

$$E_{\text{chro}}(F) = \| c(x) - c_F(x) \|^2, \tag{13}$$

where $c(x) = I(x)/|I(x)|$ is the chromaticity of the input image and c_F is the chromaticity of the shadow-free image F, $c_F(x) = F(x)/|F(x)|$. To avoid division by zero, we further rewrite this term as:

$$E_{\text{chro}}(F) = \left\| \frac{I(x)}{|I(x)| + \xi} - \frac{F(x)}{|F(x)| + \xi} \right\|^2, \tag{14}$$

where ξ is a regularization parameter and is typically set 0.0001 in our experiments.

Constant term We pick out the reliable lit pixels that should maintain their β colors and enforce their values to be 1:

$$E_{\text{const}}(\beta) = \sum_{p \in N_b} \| \beta_p - 1 \|^2, \tag{15}$$

where the N_b is the reliable lit region that is neither high shadow confidence pixels nor their neighbors.

As illustrated in Fig. 5, using chromaticity prior, the texture and chromaticity under the shadow regions are better recovered. In Fig. 6, we can observe that using the shading-preserving term in the smooth term, the shading of the shadow regions is better reconstructed, and the recovered illumination varies with the scene shape. Figures 5 and 6 also show the shadow removal comparisons with method [42]. For fair comparisons on shadow removal step, both our method and method [42] use the same shadow confidence maps as input.

3.2.2 Image detail recovering

Although our shadow removal method can recover the texture detail well in most cases, for some extremely complicated cases, where the shadow regions are too dark and have heavy noise, or the edges information and texture details in the shadow regions have been weakened seriously due to the illumination occluding, our previous method may not work well, as illustrated in Fig. 7b. To make the method more robust and better recover the texture details, we add a multi-scale texture recovering in our method. In the previous steps, using the proposed shadow-preserving texture filter, we can extract a multi-scale detail levels D^i from the original image I and $D^i = J^i - J^{i-1}$. We combine the details into the final results in a spatially varying manner using the weighted average.

Let $I_{\text{ini}}^{\text{free}}$ be the initial shadow removal result and $I_{\text{enhance}}^{\text{free}}$ be the enhanced image, then:

$$I_{\text{ehance}}^{\text{free}} = I_{\text{ini}}^{\text{free}} + C_S \cdot \sum_{i=1}^{m} U^i D^i, \tag{16}$$

where m is the scale of the processing which is usually set to be 3, $U^i = G_\sigma * e^{(|D_i - C_i|)}$ and $C_p^i = \frac{\sum_{q \in \Omega_p} \nabla |I_q^i|}{n}$. Ω_p is a local neighborhood of pixel p, and n is the number of pixels in Ω_p. G_σ is the Gaussian convolution, which is used to locally smooth the weight. C_S is the shadow confidence map, which reflects the density of shadow in each pixel. By multiplying
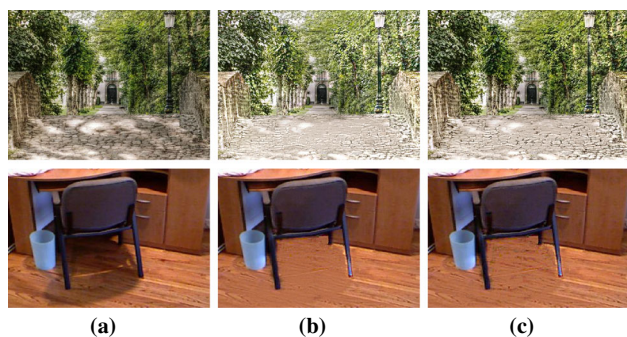
(a) **(b)** **(c)**

Fig. 7 Visualization of the recovered texture detail. From left to right are **a** input images, **b** shadow removal results and **c** the detail-recovered results

C_S, the details in shadow regions can be enhanced efficiently. In Fig. 7, we present the image detail recovering results. We can observe that the texture details are effectively recovered in the shadow region.

4 Experiments

To illustrate the effectiveness of our method, we perform our shadow detection and removal on different datasets, and compare our method with other state-of-the-art methods quantitatively and qualitatively. All our results are implemented with MATLAB R2016a, and all our experiments are executed on the machine that equipped with Intel(R) Core(TM) i5-7400 CPU @ 3.00GHz with 8GB RAM. For an image of size 640×480, our method generally takes 5–7 min for shadow removal and detection, where it spends 30–40 s for depth estimation and shadow detection, and takes about 4–6 min for performing shadow removal.

4.1 Datasets and evaluation metrics

Datasets for shadow detection Two benchmark datasets are employed in shadow detection. The first one is the SBU Shadow Dataset [36], which contains 4089 training images and 638 testing images. It includes a wide variety of scenes and covers various types of pictures. The second benchmark dataset we employed is the ISTD Shadow Dataset [37]. It includes 1340 training images and 530 testing images, and covers 135 different types of ground materials.

Evaluation metrics for shadow detection We employ the accuracy (ACC) and the balance error rate (BER) metrics to quantitatively evaluate the shadow detection performance, as defined in [17].

Datasets for shadow removal The comparison is conducted on SRD [33] and ISTD [37] datasets, and both of them have the shadow-free images. The first benchmark dataset [33]

contains 3088 images, and the second benchmark dataset [37] contains 1870 images.

Evaluation metrics for shadow removal We conduct quantitative comparisons on shadow removal using the root mean square error (RMSE) between the produced shadow removal results and the corresponding ground-truth image in the LAB color space, and compute RMSE values.

4.2 Comparison with shadow detection methods

In Fig. 8, we compare our results with some state-of-the-art shadow detection methods [15,17,37,42,48] on the benchmark datasets and some other images. Among these methods, the method [15] is based on handcrafted features, the method [42] applies the RGB-D images, and the method [48] incorporates user interactivity; the last two [17,37] are deep learning-based methods. In order to achieve the best results of those paper, the existing results in the paper are directly used as the comparison results, and the rest results are generated using implementations provided by the authors or reproduced according to the paper with recommended parameter setting. Also, note that, in these examples, the depth map used for both [42] and our method is estimated using [12]. It can be observed that incorporating shadow-preserving filtering techniques as well as depth maps, our methods work better for these images. Due to deep learning methods heavily depend on the variety of the training data, for some scenes that are hard to obtain the ground-truth training data, these methods do not work well.

Table 1 presents the quantitative comparisons with the state-of-the-art methods on the shadow detection accuracy for the two benchmark datasets. Note that we normalize the shadow confidence map and generate binary masks for [42] and our method for comparisons. We compare the binary mask against the ground truth on both the ISTD dataset [37] and SBU dataset [36]. The two datasets include lots of large-scale scenes, which can benefit to evaluate the performance of our algorithm comparatively and objectively. Our method has achieved one of the best performances on the both datasets.

4.3 Comparison with shadow removal methods

In Fig. 9, we compare our results with various kinds of state-of-the-art shadow removal methods [15,33,35,37,42,48] on the benchmark datasets and some other images we collect. The last three [33,35,37] are deep learning-based shadow removal methods. For fair comparison, the shadow removal results of other methods are generated using implementations provided by the authors or reproduced according to the paper with recommended parameter setting.

In Table 2, we can see the quantitative comparisons on shadow removal using the root mean square error (RMSE) in the LAB color space. The comparison is conducted on SRD

Fig. 8 Visual comparison on shadow detection results. From left to right are **a** input images, **b** results of Guo [15], **c** results of Xiao [42], **d** results of Zhang [48] incorporating user interactivity, **e** results of Wang [37], **f** results of DSC [17], **g** our refined shadow confidence maps and **h** the binary results based on **g**

Table 1 Quantitative comparisons on shadow detection using BER and ACC (BER: smaller is better, ACC: larger is better)

DA	MD	Guo [15]	Xiao [17]	Zhang [42]	Wang [48]	DSD [37]	Ours
SBU	ACC	0.86	0.765	0.96	0.93	*0.97*	**0.98**
	BER	25.03	31.11	7.13	11.34	*5.59*	**3.62**
ISTD	ACC	0.84	0.734	0.95	**0.98**	0.93	*0.97*
	BER	27.16	30.37	8.56	**3.85**	12.21	*5.33*

The best and second best results are marked in textbold and textitalic, respectively. MD is shorthand for method

dataset [33] and ISTD dataset[37]. We evaluate the performance of different methods on shadow regions, non-shadow regions and the whole image, as shown in Table 2. The results demonstrate that our removal results perform better for illumination recovery in shadow regions and have the smallest difference from ground-truth shadow-free images.

User study As some images have no ground truths, we evaluated the quality of 40 shadow removal images by user tests. We performed a user study with 50 random volunteers to validate the effectiveness of our proposed method. For each volunteer, we randomly show them the shadow removal results of our approach and other six methods [15,33,35,37,42,48]. All the results are labeled to avoid potential unfair comparison. Once a volunteer has finished browsing all the shadow removal results for each image, a survey is conducted to collect the feedbacks on the following questions:

Fig. 9 Visual comparison on shadow removal results. From left to right are **a** input images, **b** results of Guo [15], **c** results of Xiao [42], **d** results of Zhang [48], **e** results of Qu [33], **f** results of Wang [37], **g** results of AGAN [35] and **h** results of our proposed approach

- Q1: Which one exhibits the best overall shadow removal result?
- Q2: Which one best recovers the illumination of the shadow regions?
- Q3: Which one introduces the least visual artifacts?
- Q4: Which one has the least destruction on the non-shadow regions?
- Q5: Which one preserves the clearest textures?

For each image, and for each question, the volunteer should select the best methods. Table 3 illustrates the survey results.

Discussions Deep learning-based shadow detection and removal methods have achieved convincing results for some input images. However, the performance of these methods heavily depends on the training dataset. The current training data usually contain images with simple shadow regions, as it is relatively easy to obtain the training data. For some complex scenes, the training data are difficult to collect. For example, as shown in Fig. 9, the shadow-free images (ground-truth data) are difficult to collect. In these cases, those deep learning methods do not work well. In contrast, our optimization-based method can produce satisfactory results just by tuning a small number of parameters.

Parameter influence We have explored the effect of changing the parameter setting, as shown in Fig. 10. To illustrate the effect of each parameter, we give each parameter a different value at a time while keeping other parameters unchanged and see how the shadow removal results vary with this parameter. Our method is not sensitive to parameter variations, and

Table 2 Quantitative comparisons on shadow removal using RMSE (smaller is better) on the SRD and ISTD datasets

	R	I_{in}	Guo [15]	Xiao [42]	Zhang [48]	Qu [33]	Wang [37]	AGAN [35]	Ours
SRD	S	42.38	29.89	10.16	*9.56*	11.78	17.82	17.63	**8.63**
	N	**4.56**	6.47	7.51	6.90	*4.84*	5.48	7.83	5.32
	A	14.41	12.60	8.13	7.24	*6.64*	7.95	15.97	**6.13**
ISTD	S	32.67	18.95	11.56	*9.77*	12.54	10.33	9.78	**9.16**
	N	**6.83**	7.46	7.75	7.12	7.44	*6.93*	7.67	7.23
	A	10.97	9.30	8.14	8.16	8.24	*7.47*	8.16	**7.41**

We compute the RMSE for input images (I_{in}) and the corresponding shadow removal results using the methods in different regions (S: shadow regions, N: non-shadow regions, A: all the regions). The best and second best results are marked in textbold and textitalic, respectively

Table 3 Survey results of user study by collecting the users' feedback to five questions, i.e., Q1, Q2, Q3, Q4 and Q5

Method	Question ID					Mean (%)
	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)	Q5 (%)	
Guo [15]	8.70	4.30	12.55	14.75	11.25	10.31
Xiao [42]	8.35	5.05	9.18	13.10	14.24	9.98
Zhang [48]	9.27	11.23	15.15	19.55	15.31	14.10
Qu [33]	1.75	1.37	3.65	6.40	7.00	4.03
Wang [37]	2.05	1.60	3.05	5.40	6.15	3.65
AGAN [35]	1.40	2.10	4.02	6.10	8.10	4.34
Ours	68.48	74.35	52.40	34.70	37.95	53.58

(a) original **(b)** σ_{ch}=0.1 **(c)** σ_{ch}=0.6 **(d)** λ_1=0.5 **(e)** λ_1=2 **(f)** λ_2=0.05 **(g)** λ_2=5 **(h)** λ_3=0.5 **(i)** λ_3=2
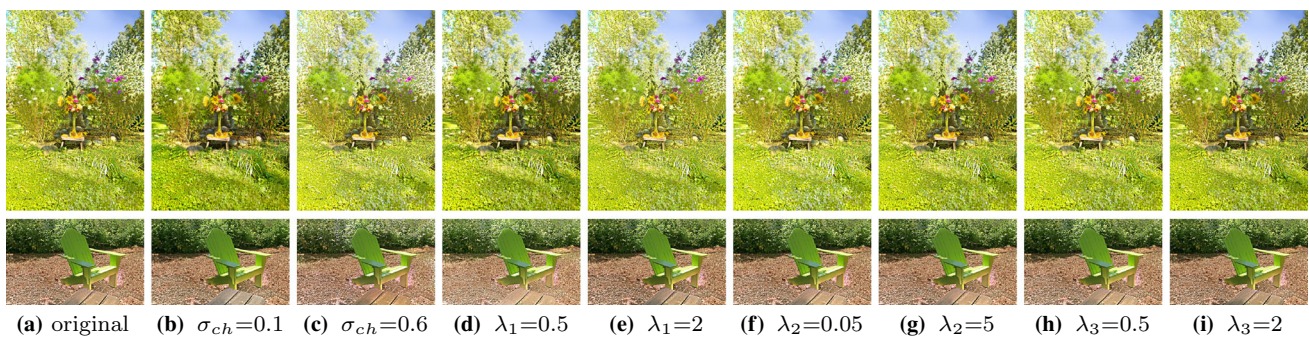
Fig. 10 Effect of parameters. **a** Result with default parameter setting described in our main paper ($\sigma_{ch} = 0.3$, $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\lambda_3 = 1$). **b–j** Results with different parameter settings. Note that we only change one parameter value at a time while keeping other parameter values fixed

the default parameter setting we set in the main paper can be used to tackle images from other benchmarks well.

Limitations Our methods also have some limitations. One limitation is that as we do not incorporate the semantic object recognition in our shadow detection and removal system. Hence, some dark regions, such as the legs of the chair and dark textures of the floor, will be mistakenly detected as shadow regions. In this case, when we perform shadow removal on these regions, it will achieve unsatisfied results, as illustrated in the bottom of Figs. 7 and 9. In addition, computational cost is currently a bottleneck to our algorithm.

5 Conclusion and future work

In this paper, we have proposed a shading-aware shadow detection and removal algorithm. We first introduce a shadow-preserving texture filter and apply shadow confidence method for shadow confidence estimation. With the benefits of the shadow confidence map, we then develop a shading-aware shadow removal method. Our method can effectively remove the complex shadows, and in particular, our method works much better on recovering the shading of the shadow regions. In the future, we would like to extend our current method to handle video shadow detection and removal.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Anat, L., Dani, L., Yair, W.: A closed-form solution to natural image matting. IEEE TPAMI **30**(2), 228–242 (2007)
2. Arbel, E., Helor, H.: Shadow removal using intensity surfaces and texture anchor points. IEEE TPAMI **33**(6), 1202–1216 (2011)
3. Barron, J.T., Malik, J.: Intrinsic scene properties from a single rgb-d image. In: CVPR, pp. 17–24 (2013)
4. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: ICCV, pp. 241–248 (2013)
5. Cho, H., Lee, H., Kang, H., Lee, S.: Bilateral texture filtering. ACM TOG **33**(4), 1–8 (2014)
6. Chuang, Y.Y., Goldman, D.B., Curless, B., Salesin, D.H., Szeliski, R.: Shadow matting and compositing. In: ACM SIGGRAPH, pp. 494–500 (2003)
7. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with HSV color information. In: Intelligent Transportation Systems, pp. 334–339 (2002)
8. Cun, X., Pun, C.M., Shi, C.: Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. AAAI (2020)
9. Ding, B., Long, C., Zhang, L., Xiao, C.: Argan: attentive recurrent generative adversarial network for shadow detection and removal. In: ICCV, pp. 10213–10222 (2019)
10. Eh, L., Jj, M.: Lightness and retinex theory. J. Opt. Soc. Am. **61**(1), 1–11 (1971)
11. Finlayson, G.D., Hordley, S.D., Lu, C., Drew, M.S.: On the removal of shadows from images. IEEE TPAMI **28**(1), 59–68 (2006)
12. Godard, C., Aodha, O.M., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: CVPR, pp. 6602–6611 (2017)
13. Godard, C., Mac Aodha, O., Firman, M., Brostow, G.J.: Digging into self-supervised monocular depth estimation. In: ICCV, pp. 3828–3838 (2019)
14. Gryka, M., Terry, M., Brostow, G.J.: Learning to remove soft shadows. ACM TOG **34**(5), 1–15 (2015)
15. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: CVPR, pp. 2033–2040 (2011)
16. Hachama, M., Ghanem, B., Wonka, P.: Intrinsic scene decomposition from rgb-d images. In: ICCV, pp. 810–818 (2015)
17. Hu, X., Fu, C.W., Zhu, L., Qin, J., Heng, P.A.: Direction-aware spatial context features for shadow detection and removal. In: IEEE TPAMI, pp. 7454–7462 (2019)
18. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-shadowgan: learning to remove shadows from unpaired data. In: ICCV, pp. 2472–2481 (2019)
19. Jeon, J., Cho, S., Tong, X., Lee, S.: Intrinsic image decomposition using structure-texture separation and surface normals. In: ECCV, pp. 218–233 (2014)
20. Karsch, K., Sunkavalli, K., Hadap, S., Carr, N., Jin, H., Fonte, R., Sittig, M., Forsyth, D.: Automatic scene inference for 3d object compositing. ACM TOG **33**(3), 1–15 (2014)
21. Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: Fourth International Conference on 3D Vision, pp. 239–248 (2016)
22. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Detecting ground shadows in outdoor consumer photographs. Lect. Notes Comput. Sci. **6312**, 322–335 (2010)
23. Le, H., Samaras, D.: Shadow removal via shadow image decomposition. In: ICCV, pp. 8578–8587 (2019)
24. Le, H., Vicente, Y., Tomas, F., Nguyen, V., Hoai, M., Samaras, D.: A+ D net: training a shadow detector with adversarial shadow attenuation. In: ECCV, pp. 662–678 (2018)
25. Li, Z., Snavely, N.: Learning intrinsic image decomposition from watching the world. In: CVPR, pp. 9039–9048 (2018)
26. Liu, F., Gleicher, M.: Texture-consistent shadow removal. In: ECCV, pp. 437–450 (2008)
27. Liu, F., Shen, C., Lin, G.: Deep convolutional neural fields for depth estimation from a single image. In: CVPR, pp. 5162–5170 (2015)
28. Meka, A., Richardt, C., Theobalt, C.: Live intrinsic video. ACM TOG **35**(4), 109 (2016)
29. Mikic, I., Cosman, P.C., Kogut, G.T., Trivedi, M.M.: Moving shadow and object detection in traffic scenes. In: IEEE TPAMI, vol. 1, pp. 321–324 (2000)
30. Mohan, A., Tumblin, J., Choudhury, P.: Editing soft shadows in a digital photograph. IEEE Comput. Graph. Appl. **27**(2), 23–31 (2007)
31. Nguyen, V., Vicente, T.F.Y., Zhao, M., Hoai, M., Samaras, D.: Shadow detection with conditional generative adversarial networks. In: ICCV, pp. 4520–4528 (2017)
32. Okabe, T., Sato, I., Sato, Y.: Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions. In: ICCV, pp. 1693–1700 (2009)
33. Qu, L., Tian, J., He, S., Tang, Y., Lau, R.W.H.: Deshadownet: a multi-context embedding deep network for shadow removal. In: CVPR, pp. 2308–2316 (2017)
34. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2002)
35. Sidorov, O.: Conditional GANS for multi-illuminant color constancy: revolution or yet another approach? In: CVPR (2019)
36. Vicente, T.F.Y., Hou, L., Yu, C.P., Hoai, M., Samaras, D.: Large-scale training of shadow detectors with noisily-annotated shadow examples. In: ECCV, pp. 816–832 (2016)
37. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: CVPR, pp. 1788–1797 (2018)
38. Wei, J., Long, C., Zou, H., Xiao, C.: Shadow inpainting and removal using generative adversarial networks with slice convolutions. Comput. Graph. Forum **38**, 381–392 (2019)
39. Wu, T.P., Tang, C.K., Brown, M.S., Shum, H.Y.: Natural shadow matting. ACM TOG **26**(2), 8 (2007)
40. Xiao, C., Gan, J.: Fast image dehazing using guided joint bilateral filter. Vis. Comput. **28**(6–8), 713–721 (2012)
41. Xiao, C., She, R., Xiao, D., Ma, K.L.: Fast shadow removal using adaptive multi-scale illumination transfer. Comput. Graph. Forum **32**(8), 207–218 (2013)
42. Xiao, Y., Tsougenis, E., Tang, C.: Shadow removal from single rgb-d images. In: CVPR, pp. 3011–3018 (2014)
43. Xu, D., Ricci, E., Ouyang, W., Wang, X., Sebe, N.: Multi-scale continuous crfs as sequential deep networks for monocular depth estimation, pp. 161–169 (2017)
44. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. ACM TOG **31**(6), 1–10 (2012)

45. Zhang, L., Long, C., Zhang, X., Xiao, C.: Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In: AAAI (2020)
46. Zhang, L., Yan, Q., Liu, Z., Zou, H., Xiao, C.: Illumination decomposition for photograph with multiple light sources. IEEE Trans. Image Process. **26**(9), 4114–4127 (2017)
47. Zhang, L., Yan, Q., Zhu, Y., Zhang, X., Xiao, C.: Effective shadow removal via multi-scale image decomposition. Vis. Comput. **35**(6–8), 1091–1104 (2019)
48. Zhang, L., Zhang, Q., Xiao, C.: Shadow remover: image shadow removal based on illumination recovering optimization. IEEE Trans. Image Process. **24**(11), 4623–4636 (2015)
49. Zheng, Q., Qiao, X., Cao, Y., Lau, R.W.: Distraction-aware shadow detection. In: CVPR, pp. 5167–5176 (2019)

Xinyun Fan received her B.Sc. degree from Northwest A&F University in 2018. Currently, she is a postgraduate at the School of Computer, Wuhan University, China. Her research interests include image and video editing, and image enhancement.

Wenjun Wu received her B.Sc. degree from Wuhan University in 2016 and received her M.Sc. degree at the School of Computer of Wuhan University in 2019. Her research interests include image editing and computer vision.

Ling Zhang received her B.Sc. degree in computer science and technology from Wuhan Donghu University in 2009 and received M.Sc. degrees in instructional technology from Central China Normal University in 2012 and her Ph.D. degree at Wuhan University in 2017. Currently, she is a postdoctoral researcher in School of Computer Science and Technology, Wuhan University of Science and Technology. Her research interests include image and video editing, and computational photography.

Qingan Yan received his Ph.D. degree at Computer Science Department of Wuhan University in 2017. Before that, he got his M.S. and B.E. degree in Computer Science from Southwest University of Science and Technology, and Hubei Minzu University, in 2012 and 2008, respectively. He is currently working at JD.com Silicon Valley Research Center in California as a research scientist. His research interests include 3D reconstruction, geometric scene perception, visual correspondence and image synthesis.

Gang Fu is working toward the Ph.D. degree at the School of Computer, Wuhan University. His research interests include intrinsic image decomposition, specularity detection and removal, and image enhancement.

Zipei Chen received his B.Sc. degree in software engineering from Xi'an University of Finance and Economics in 2019. Currently, he is a postgraduate at Computer Science Department of Wuhan University. His research interests include image editing and active learning.

Chengjiang Long is an Adjunct Professor at University at Albany, SUNY since Aug 2018. He worked at Kitware as a Computer Vision Researcher/Senior RD Engineer between 2016 and 2020. He also worked as an Adjunct Professor at Rensselaer Polytechnic Institute (RPI) from Jan 2018 to May 2018. He received the M.S. degree in Computer Science from Wuhan University in 2011 and a B.S degree in Computer Science and Technology from Wuhan University in 2009. He got his Ph.D. degree in Computer Science from Stevens Institute of Technology in 2015. His research interests involve computer vision, machine learning, artificial intelligence and computer graphics. He is a member of IEEE and AAAI.

Chunxia Xiao received his B.Sc. and M.Sc. degrees from the Mathematics Department of Hunan Normal University in 1999 and 2002, respectively, and Ph.D. degree from the State Key Lab of CAD&CG of Zhejiang University in 2006. Currently, he is a professor at the School of Computer, Wuhan University, China. From October 2006 to April 2007, he worked as a postdoc at the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, and during February 2012 to February 2013, he visited University of California Davis for 1 year. His main interests include computer graphics, computer vision and machine learning. He is a member of IEEE.