An Improved Augmented-Reality Framework for Differential Rendering Beyond the Lambertian-World Assumption

Aijia Zhang, Yan Zhao, *Member, IEEE*, and Shigang Wang, *Member, IEEE*

Abstract—In augmented reality, it is important to achieve visual consistency between inserted virtual objects and the real scene. As specular and transparent objects can produce caustics, which affect the appearance of inserted virtual objects, we herein propose a framework for differential rendering beyond the Lambertian-world assumption. Our key idea is to jointly optimize illumination and parameters of specular and transparent objects. To estimate the parameters of transparent objects efficiently, the psychophysical scaling method is introduced while considering visual characteristics of the human eye to obtain the step size for estimating the refractive index. We verify our technique on multiple real scenes, and the experimental results show that the fusion effects are visually consistent.

Index Terms—Augmented reality, specular and transparent objects, global illumination, light estimation, material estimation, joint optimization

1 Introduction

In augmented reality, visual consistency between virtual objects and real scenes is a challenging problem that has been studied in recent years. The key to solving visual consistency is to estimate three-dimensional (3D) geometries, parameters of surface materials, and illumination of real scenes. Various methods have been proposed to insert virtual objects into a real scene using differential rendering to achieve a plausible result. Most of the studies only solved the effects of light sources on virtual objects and did not consider the effects of caustics (Fig. 1) produced by specular and transparent objects on virtual objects. Our goal in this paper is to obtain plausible results when virtual objects are inserted around specular and transparent objects.

To insert virtual objects into a scene with specular and transparent objects, a few important aspects must be addressed: recognition of specular and transparent objects; scene reconstruction; and estimation of illumination and material of objects. Specular and transparent objects in a real scene must be recognized prior to material estimation and differential rendering. Most existing methods for recognizing transparent objects [1], [2], [3] cannot distinguish specular objects from transparent objects. Hence, we distinguish them based on their physical properties. Scene reconstruction yields a complete geometry of the scene and camera parameters of different views.

The focus of our study is to estimate the illumination and material of objects. Most previous methods regarding illumination estimation assumed that the material of real objects was Lambertian [4], [5], [6], [32]; therefore, the applicable scene was simple. Recently, a few methods [7], [8], [9], [10] have been proposed to recover the surface material and illumination simultaneously. However, these methods do not recognize specular objects or consider transparent objects, which is not conducive to inserting virtual objects.

To summarize, our contributions are as follows:

(1) We developed a new framework based on a global illumination model to estimate the joint global optimization of illumination and the material of specular and transparent objects, as well as solved the problem of inserting virtual objects around specular or transparent objects. To the authors knowledge, no algorithm has been published for solving the effects of caustics produced by real scene objects on virtual objects.

(2) We used the psychophysical scaling method and the maximum-likelihood difference scaling (MLDS) to derive the minimum change in refractive index of the transparent object when the human eyes can perceive the change in caustics of the transparent object. The minimum change in refractive index was used as the step size when estimating the refractive index.

(3) To distinguish between transparent and specular objects, we utilized the different interactions between light and the surface of objects.

2 Related Work

A classification of illumination estimation methods has been proposed by Ren H. *et al.* [11]. Various methods have been proposed for synthesizing virtual objects into real scenes. In this paper, we address the effect of specular and transparent objects on virtual objects in a scene. Hence, we focus on studies related to recognizing specular and transparent

- *The authors are with the College of Communication Engineering, Jilin University, Changchun 130000, China. E-mail: 296529503@qq.com, zhao_y@jlu.edu.cn, wangshigang@vip.sina.com.*

Manuscript received 22 June 2019; revised 28 May 2020; accepted 17 June 2020. Date of publication 23 June 2020; date of current version 27 Oct. 2021.
(Corresponding author: Yan Zhao.)
Recommended for acceptance by A. Ghosh.
Digital Object Identifier no. 10.1109/TVCG.2020.3004195

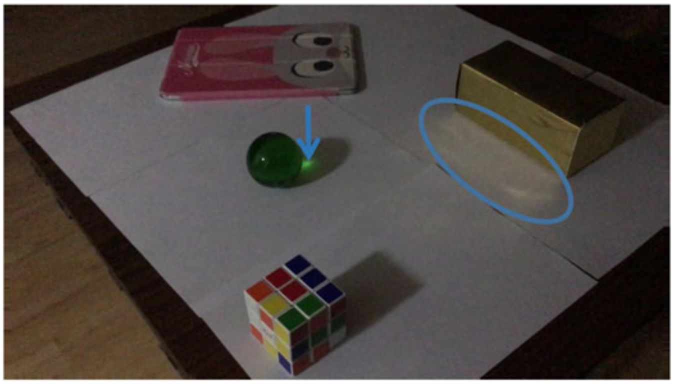

Fig. 1. Caustics produced by specular and transparent objects. Caustics produced by specular and transparent objects are denoted by the ellipse and pointed by the arrow, respectively.

objects, estimating illumination and materials, and differential rendering for virtual object insertion in augmented reality.

Recognition of Specular and Transparent Objects. Studies pertaining to transparent object recognition are few. Lysenkov*et al.* [12] used the weakness of a Kinect sensor in the perception of transparent objects to extract a region of interest. The extracted region of interest was then entered into Grabcut [13] for segmentation. Ji *et al.* [3] improved [12] to obtain more accurate results by excluding noise regions that could not be identified in [12]. In addition, the work of [2] used a learning and classification method based only on color images, where the images were decomposed into smaller areas to search for clues on transparent objects. However, most existing methods recognize specular objects as transparent objects. Light is refracted and reflected on the surface of transparent and specular objects, respectively, which causes the depth to be inconsistent in multiview depth images captured using a RGB-D camera. Furthermore, both types of objects will produce invalid areas on depth images. Most of the existing methods rely on depth inconsistency and invalid regions to extract transparent objects. Therefore, specular objects are recognized as transparent objects. In our method, the difference in radiance produced by the interaction of light with different material surfaces was used to further distinguish specular and transparent objects.

Estimation of Illumination and Material. In early studies, auxiliary information such as probe spheres [14], [15] or shadows were used to estimate the illumination in a scene. Panagopoulos *et al.* [16] regarded the light environment as a von Mises–Fisher hybrid distribution and used the expectation maximization algorithm to solve the various parameters in a mixed distribution. Shadows were detected using a 3D model, which did not depend on the shadow projection surface. The same research group [17] proposed a method based on a high-order Markov random field illumination model, using a voting algorithm to initialize and analyze the shadows, which yielded a more robust result. Moreover, a highlight point is a type of auxiliary information. Jiddi*et al.* [18] estimated the position of a light source from a sequence of images with specular reflection. They captured the position of the highlight point pixels in the image sequences and calculated the direction of the incident light. The position determination of the light source was similar to computing the intersection point of multiple 3D lines.

Owing to technological development, various methods without auxiliary information have been proposed. Boom *et al.* [5] used color and depth images to estimate point light source in cases where a scene was assumed as a Lambertian surface. Karaoglu*et al.* [19] improved the method of [5] by classifying image surface segments based on surface attributes. These attributes were used in a supervised learning program to derive important levels of different surfaces for correct lighting results. The optimization method afforded more accurate results compared with those in [5]. Liu *et al.* [4] utilized an environment map to estimate illumination and recover natural light by estimating the coefficients of the spherical harmonic basis function. A global illumination model was used to estimate illumination in [32]. More accurate results of the direction of light sources were obtained after considering internal reflections. Owing to the development of deep learning, some methods for illumination estimation using neural networks have been proposed [20], [21], [22], [36]. Gardner *et al.* [20] introduced an end-to-end network that converted limited scene illumination information to an HDR image. It is a convenient method because the scene model and material properties of the objects do not require estimation.

The joint estimation of reflectance and lighting has been proposed in recent years. Wu *et al.* [7] developed a joint optimization objective that can constrain illumination, spatially varying bidirectional reflectance distribution function (BRDF), normal information, and other information. The objective optimization is to estimate some variables and retain the others. Ignoring occlusions and inter-reflections in this method may result in inaccuracies in material estimation. AzinoviÂ´c *et al.* [8] used an inverse path tracing algorithm to jointly estimate reflectance and illumination using 3D scenes and images from different views as input. This method takes into account internal inter-reflections and occlusions and achieves better results. Using deep learning to jointly estimate illumination and material is another approach. Georgoulis *et al.* [33] presented a two-step approach that estimated material parameters and natural illumination. First, a single image depicting a single-material specular object was input to a convolution neural network (CNN) to obtain a reflection map. Subsequently, the reflection map was fed to the CNN and decomposed into reflection parameters and an illumination map. However, the input image was a single object with a single material, which limited the applicability of this method. In [34], a network was trained to regress diffuse albedo and normal maps, from which the illumination was obtained. The proposed network leveraged offline multiview stereos as self-supervision, and the work established a statistical natural illumination to ensure the accuracy of the results. A neural inverse rendering network was proposed in [35], which can predict albedo, normal, and illumination maps. The key of this method is that the residual appearance renderer can train datasets to learn complex appearances, such as internal reflections and shadows.

Differential Rendering for Virtual Objects Insertion in Augmented Reality. The rendering of virtual objects into realistic real scenes have been long studied. Debevec [23] first

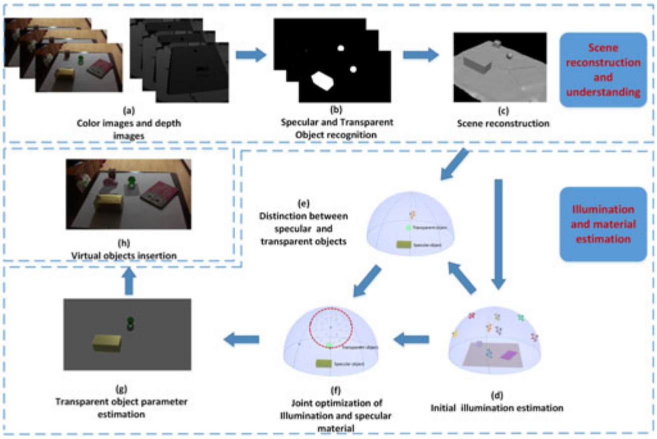
(a)
Color images and depth images
(b)
Specular and Transparent Object recognition
(c)
Scene reconstruction
Scene reconstruction and understanding
(h)
Virtual objects insertion
(e)
Distinction between specular and transparent objects
Transparent object
Specular object
Illumination and material estimation
(g)
Transparent object parameter estimation
(f)
Joint optimization of Illumination and specular material
(d)
Initial illumination estimation

Fig. 2. Workflow of our system. Our system is categorized into three stages: scene reconstruction and understanding; illumination and material estimation; and virtual object insertion.

proposed differential rendering to demonstrate the effect between real-world and virtual models. Karsch *et al.* [37] estimated the depth information and light source from the input single-image by data-driven method. Subsequently, they performed differential rendering to compose virtual objects into the input image. The methods described in [4] and [38] also utilized differential rendering to insert virtual objects into an image, which can not only depict the effect of the shadow of the virtual object on the real world, but also vice versa. Our proposed method considers a more complex case: the effects of the caustics of real specular and transparent objects on virtual objects.

3 Approach Overview

The overall framework of our system is illustrated in Fig. 2. First, at the scene reconstruction and understanding stage, we input the depth and color images of multiple views captured by a Kinect sensor (Fig. 2a). Specular and transparent objects in the scene were recognized by processing depth and color images (Fig. 2b). The complete model was reconstructed using the combination of visual hull and Kinectfusion algorithms (Fig. 2c).

Next, at the illumination and the material estimation stage, as shown in Fig. 2d, we established a photon emission hemispherical model to estimate the initial light source position. During the estimation of the initial light source, inter-reflections between the surfaces were considered. We utilized the initial estimated illumination and the different interactions between light and the object surface to distinguish transparent and specular objects (Fig. 2e). As shown in Fig. 2f, a uniform sampling was performed near the initial light source position on the hemispherical surface, and illumination as well as the parameters of specular and transparent objects were jointly optimized. Finally, using differential rendering [23], virtual objects can be realistically inserted around the specular and transparent objects.

4 Scene Reconstruction and Understanding

Recovering a scene geometry is the first step in our method. In [4], [6], objects in a scene were assumed to be Lambertian surfaces, which can easily recover the scene geometry. In our approach, our goal was to manage the effects of caustics produced by specular and transparent objects on virtual objects, which requires the processing of transparent and specular objects. Existing methods for recognizing transparent objects [1], [2], [3] cannot distinguish between specular and transparent objects. We describe how to distinguish them to improve the existing method for recognizing transparent objects in Section 5.2. Subsequently, we categorized our study into two parts: first, both transparent and specular objects were recognized in the scene; next, the geometry of the entire scene including transparent and specular objects were reconstructed.

Although we did not improve the scene reconstruction method, these two steps are key to obtain the final plausible result. The two above mentioned parts were implemented as follows. First, it was assumed that the initial poses of each input color image and depth image can be obtained using the Kinectfusion algorithm [24], and that the truncated signed distance function (TSDF) [25] was available to represent the original model. We utilized the noise region search algorithm in [3] to obtain the approximate region where specular and transparent objects were located. Subsequently, the regions of specular and transparent objects were reprojected onto color images, which can be used as initial information for the color image segmentation algorithm. We extracted the silhouettes of the specular and transparent objects using the improved Grabcut algorithm in [3]. The steps above enabled both transparent and specular objects to be recognized. The camera poses obtained by the Kinectfusion algorithm were not accurate, which will affect subsequent reconstruction of specular and transparent objects and material estimation. We used the camera pose optimization described in [31] to optimize the camera poses. When optimizing the camera pose, a geometric model excluding the area of transparent and specular objects was used. As for reconstructing the entire scene, we adopted a method similar to that in [3,26]. Combining the silhouettes of specular and transparent objects, we recovered the models of specular and transparent objects using the visual hull method. The TSDF model was combined with the specular and transparent object models to obtain the complete model.

5 Joint Estimation of Illumination and Material Parameters

Owing to the special material of transparent and specular objects, the caustics produced by them will affect the appearance of virtual objects in the combined virtual and real objects in augmented reality. Therefore, our goal is to jointly estimate illumination E, the parameters of the specular object BRDF model f_r, and the parameters of the transparent object material T in the scene. We established a joint optimization with regard to the photometric error, defined as follows:

$$\underset{E,\{f_r,T\}_x}{\operatorname{argmin}} \sum_i \sum_{x \in X_i} \|S_i(x) - L(\omega_o;x,E)\|^2, \tag{1}$$

where x represents the point of the object of interest, X_i is the set of corresponding points in the image S_i, and $L(\omega_o;x,E)$ is the reflected radiance in the output direction ω_o. Eq. (1) enables the estimated appearance of the object to match with the object in the captured image S_i as much as possible.

As the objective function in Eq. (1) is nonconvex, to obtain the global optimal solution of illumination and material parameters, we designed the following steps. First, in Section 5.1, we estimate the initial light source based on inverse photon mapping, which considers inter-reflections between the surfaces. Subsequently, the illumination optimization and the estimation of the specular object BRDF model are converted into a convex optimization problem to jointly obtain the solution (Section 5.3). In Section 5.4, illumination optimization and the estimation of transparent object parameter are jointly computed using the linear least absolute error. Finally, two optimization problems are considered simultaneously. When the minimum sum of their objective functions is obtained, the corresponding illumination and material parameters are the final estimated results.

5.1 Initial Illumination Estimation

When estimating the initial illumination, transparent and specular objects were not considered, and other objects were assumed to be Lambertian surfaces. We utilized the inverse photon mapping method [32] to estimate the position and intensity of light sources. In the algorithm, a photon emission hemispherical model was established, as shown in Fig. 2d, in which the entire scene was surrounded by a hemisphere. Point light sources were distributed evenly on the hemisphere surface. A large number of photons from different directions were emitted into the scene from a specified point light source on the hemisphere.

Next, photons were emitted from the point source in the photon emission hemisphere model into the scene, and photon tracing was then performed. Assuming that the power of the j-th point light source is I_j and the number of photons emitted from each point light source is q, the energy of each photon $\Delta\Phi(\omega_p)$ from the j-th point light source is expressed as

$$\Delta\Phi(\omega_p) = \frac{I_j}{q}, \tag{2}$$

During photon tracing, a photon map was created to store the coordinates at the collision point, the incident direction of the photon, and the incoming photon energy simultaneously. As multiple light sources existed in the photon emission hemisphere model, multiple photon maps were created, and the information of photon emitted by each light source was stored in the corresponding photon map.

After photon tracing was completed under multiple light sources, ray tracing was performed from different views. The reflected radiance of the first object in the scene hit by the light ray is defined as

$$L(x,\omega_{\mathrm{o}}) = \int_{\Omega} f(x,\omega_i,\omega_o)E(x,\omega_i)(n_x \cdot \omega_i)d\omega_i, \tag{3}$$

where L is the reflected radiance at point x in direction ω_o, Ω is the hemisphere of the incoming direction, f is the BRDF at point x, ω_i is the direction of the incoming light ray, E is the incoming radiance, and n_x is the surface normal at point x. According to the photon mapping algorithm [27], we can rewrite Eq. (3) as follows:

$$L(x,\omega_o) \approx \frac{1}{\pi d^2(x)} \sum_{p=1}^{n} f(x,\omega_p,\omega_o)\Delta\Phi(\omega_p), \tag{4}$$

where n is the number of photons collected near point x using the photon map, and $d(x)$ is the distance of the point x from the farthest photon of the n photons collected. The material of the object is assumed to be Lambertian; therefore, L was only determined by the location at point x. Setting f to 1 in Eq. (4) enables the reflected radiance at point x to be expressed as

$$L(x) \approx \frac{1}{\pi d^2(x)} \sum_{p=1}^{n} \Delta\Phi(\omega_p). \tag{5}$$

Eq. (5) represents the reflected radiance corresponding to point x under a certain light source. The reflected radiance of point x generated by the j-th light source in photon emission hemisphere model is denoted by $L_j(x)$. Each photon energy $\Delta\Phi(\omega_p)$ from the j-th light source can be expressed using Eq. (1). Substituting Eq. (2) into (5) yields

$$L_j(x) \approx \frac{1}{\pi d_j^2(x)} \times \frac{n \times I_j}{q}, \tag{6}$$

where $d_j(x)$ is the radius at point x in the j-th light source. In this case, Eq. (1) becomes

$$\underset{E}{\operatorname{argmin}} \sum_{i=1}^{m} \sum_{x \in X_{\mathrm{i}}} \left\| S_i(x) - \sum_{j=1}^{k} \frac{1}{\pi d_{ji}^2(x)} \times \frac{n \times I_j}{q} \right\|^2, \tag{7}$$

where m is the number of images, $S_i(x)$ is the gray value of the pixel at point x of the i-th image, and $d_{ji}(x)$ is the radius at point x of the i-th image in the j-th light source. k is the number of light sources set on the hemisphere. We used the initial illumination optimization method in [32] to solve Eq. (7). The final results of position and intensity of the light sources were recorded in the collection of the illumination estimation result.

5.2 Distinction Between Transparent and Specular Objects

When light reaches the surface of a transparent object, it will be refracted into the object, and the light inside the object may be scattered, refracted, or absorbed. Because the interaction between a transparent object and light is extremely complex, we assumed that the light was only refracted when it entered the transparent object and disregarded the Fresnel reflection on the object surface. Whereas light rays reaching the surface of a specular object is reflected. Using this difference in physical properties, we distinguished whether the object identified in Section 4 is a transparent object or a specular object. A model that is not certain whether it is a transparent object or a specular object is selected out. First, the model was assumed to be a specular object; therefore, most lights were reflected. In Eq. (3), if f is set to 1 and E is the illumination information in the collection of the illumination estimation result, then Eq. (3) can be written as

TABLE 1
Pseudo-Code of Joint Optimization for Illumination and Material of the Specular Object

```
for each of the illumination combinations do
    estimate diffuse albedo and k clusters;
    for each of k clusters do
        estimate parameters of Ward BRDF for the cluster
    end for
end for
```

$$L(x)=\sum_{i=1}^{num} I_i(n_x\cdot\omega_i), \tag{8}$$

where *num* is the number of light sources in the collection of the illumination estimation result. Using Eq. (8), the sum of the pixel intensity value of the points on the surface of the specular object at different views is estimated as follows: $b_1, b_2, \ldots b_h$, where *h* is the number of different views participating in the calculation. The selected model is then assumed to be a transparent object. The refractive index of the object model is assumed to be 1.2. The illumination information in the collection of illumination estimation result is used to render the transparent object under the same *h* views, and the sum of the pixel intensity value of the object at different views is obtained as follows: $t_1, t_2, \ldots t_h$. If $|\sum_{i=1}^{h} b_i-\sum_{i=1}^{h} c_i| < |\sum_{i=1}^{h} t_i-\sum_{i=1}^{h} c_i|$, then the object is a specular object, where c_i is the sum of the gray value of the object pixels on the captured image from the *i*-th view. If $|\sum_{i=1}^{h} b_i-\sum_{i=1}^{h} c_i| > |\sum_{i=1}^{h} t_i-\sum_{i=1}^{h} c_i|$, then the object is a transparent object.

5.3 Joint Optimization for Parameters of Illumination and Material of Specular Object

First, we uniformly sampled the area near the position of each light source in the collection of the illumination estimation result, and the number of sampling points near each light source was *g* (Fig. 2f). The intensity of each sample point light source was the intensity of the corresponding light source in the collection of the illumination estimation result. Therefore, the illumination was set to $(g+1)^{num}$ possible illumination combinations. The illumination was set as one of the $(g+1)^{num}$ illumination combinations. To segment the different color clusters of the specular object, the diffuse albedo of the object was estimated first. For *k* clusters, triangles with similar diffuse albedos were classified into the same cluster by the K-means algorithm.

Next, the material parameters were estimated more accurately in different clusters. Similar to a previous study regarding reflectance estimation [28], we represent the material of the specular objects using an isotropic Ward BRDF model:

$$f(\rho_d,\rho_s,\sigma)=\frac{\rho_d}{\pi}+\frac{\rho_s\exp(-\tan^2\theta_{\mathbf{h}}/\sigma^2)}{4\pi\sigma^2\sqrt{\cos\theta_{\mathbf{i}}\sin\theta_{\mathbf{o}}}}, \tag{9}$$

where ρ_d is the diffuse albedo; ρ_s is the specular albedo; σ is the roughness parameter; $\mathbf{i}$ and $\mathbf{o}$ are the incident and viewing directions, respectively; $\mathbf{h}$ represents the half vector between $\mathbf{i}$ and $\mathbf{o}$ ($\mathbf{h}=(\mathbf{i}+\mathbf{o})/|\mathbf{i}+\mathbf{o}|$); $\theta_{\mathbf{i}}$, $\theta_{\mathbf{o}}$, and $\theta_{\mathbf{h}}$ are the incident, viewing, and half vectors with respect to the surface normal, respectively. The parameters to be estimated for the specular object are ρ_d, ρ_s, and σ. The transparent object is not considered. In this case, Eq. (1) can be written as

$$\underset{\rho_d\rho_s\sigma}{\arg\min}\sum_i\sum_{x\in X_i}\|S_i(x)-L(\omega_o;x)\|^2, \tag{10}$$

where X_i is the set of the specular object points. Substituting Eq. (9) into Eq. (3), $L(\omega_o;x)$ is expressed as

$$L(\omega_o;x)=\sum_{i=1}^{num}\left(\frac{\rho_d}{\pi}+\frac{\rho_s\exp(-\tan^2\theta_{\mathbf{h}_{i_x}}/\sigma^2)}{4\pi\sigma^2\sqrt{\cos\theta_{\mathbf{i}_{i_x}}\sin\theta_{\mathbf{o}_{i_x}}}}\right)I_i(n_x\cdot\omega_i), \tag{11}$$

where I_i is $E(x,\omega_i)$ in Eq. (3). To estimate the parameters of the BRDF model, many methods utilized the GaussNewton algorithm and other nonlinear least-squares minimization techniques. The final result may be trapped into a local minimum owing to the nonconvexity of the BRDF model. Therefore, we referred to [29] and converted the problem into a convex problem to obtain the optimal solution. All gray values *S* of the pixels in X_i in Eq. (10) constituted a column vector *S*, and all reflected radiance *L* corresponding to the points in X_i in Eq. (10) constituted a column vector *L*. We employed the branch-and-bound search combined with the second-order cone programming proposed in [29] to obtain the optimal solution of the following problem:

$$\min_{\rho_d\rho_s\sigma} e \text{ s.t.}\|S-L\|_2\le e, \rho_d\ge 0, \rho_s\ge 0, 0\le\sigma\le 1 \tag{12}$$

Eq. (12) was applied to estimate the parameters of the Ward BRDF model of the specular object for different clusters. *k* error values *e* corresponding to *k* clusters were added to e_{all}. For different combinations of light sources, we achieved $(g+1)^{num}$ e_{all}. These values were then stored for joint optimization with parameters of transparent objects. In Table 1, the pseudo code expresses the process of joint optimization for the illumination and material of the specular object.

5.4 Joint Optimization for Parameters of Illumination and Material of Transparent Object

One of the purposes of this study is to simulate the effects of the caustics produced by transparent objects on virtual objects. The parameters that primarily affect the caustics include the refractive index and color attenuation coefficient, which are estimated in two stages.

Refractive index estimation is the first stage. First, illumination is set as one of the $(g+1)^{num}$ illumination combinations. We estimated the refractive index by caustics and the transparent object at different refractive indices. As the photon mapping rendering method can express caustics, the scene was rendered using photon mapping. Therefore, we minimized the following objective function derived from Eq. (1):

$$\begin{gathered} r_o=\underset{refr}{\operatorname{argmin}}\sum_i\sum_{x\in X_i}\left\|S_i(x)-L_{refr}(\omega_o;x)\right\|^2 \\ e'=\|S_i(x)-L_{r_o}(\omega_o;x)\|^2, \end{gathered} \tag{13}$$

where L_{refr} is the reflected radiance at point *x* rendered by photon mapping; *refr* is the refractive index of the transparent object; X_i is the set of corresponding points in the image S_i; r_o and e' are the corresponding refractive index and objective function error value, respectively, when the objective function is optimal. The relationship between the refractive index *refr* and L_{refr} is complex. To simplify the optimization of Eq. (13), we uniformly sampled the refractive index *refr*. For different sampled refractive indices, the scene was rendered multiple times based on photon mapping to obtain the objective function value at different refractive indices. The refractive index corresponding to the minimum objective function value was the optimal solution.

Furthermore, to not disturb the fusion effect, the refractive index sampling step size was set as the minimum change in the refractive index that enabled the human eye can recognize the change in caustics produced by the transparent object. We propose using the psychophysical scaling method and MLDS to derive the refractive index sampling step size. Based on [30], we obtained the minimum change in the refractive index that enabled the human eye to recognize the change in caustics produced by the transparent object as λ.

The refractive index of common transparent objects varies from 1.2 to 2. Therefore, the interval of the estimated refractive index *refr* was [1.2, 2], and *refr* was sampled in the step size of λ. We optimized Eq. (13) at different sample points of refractive index to obtain the optimal refractive index of the transparent object and the corresponding objective function value e' for a certain illumination in the $(g+1)^{num}$ illumination combinations.

Subsequently, the illumination was set as other illumination combinations, resulting in $(g+1)^{num}$ optimal refractive indices and objective function values e'.

Finally, under the same illumination, e_{all} in Section 5.3 and e' were added to e_{Σ}. The illumination combination, the parameters of the specular object BRDF model, and the transparent object refractive index corresponding to the minimum e_{Σ} were the final optimization results.

In the second stage, the Beer–Lambert law was introduced to estimate the color attenuation coefficient. For the estimation of color attenuation coefficients σ_r, σ_g, and σ_b, we used the following objective functions:

$$\begin{aligned} &\underset{\sigma_r}{\operatorname{argmin}} \sum_{i=1}^{H} \left| r_{0_i} \exp(-\sigma_r d_i) - r_i \right| \\ &\underset{\sigma_g}{\operatorname{argmin}} \sum_{i=1}^{H} \left| g_{0_i} \exp(-\sigma_g d_i) - g_i \right| \qquad (14) \\ &\underset{\sigma_b}{\operatorname{argmin}} \sum_{i=1}^{H} \left| b_{0_i} \exp(-\sigma_b d_i) - b_i, \right| \end{aligned}$$

where σ_r, σ_g, and σ_b are the attenuation coefficients of red, green, and blue channels, respectively. H is the total number of pixels involved in the calculation; d_i is the transmission distance of light; and r_{0_i}, g_{0_i}, and b_{0_i} are red, green, and blue channel gray values obtained using the optimized refractive index and illumination to render the transparent object at d_i=0, respectively. r_i, g_i, and b_i are the red, green, and blue channel gray value of the captured color image, respectively. In the optimization of Eq. (14), the objective function was transformed into a linear problem using the logarithm to obtain the optimal solution.

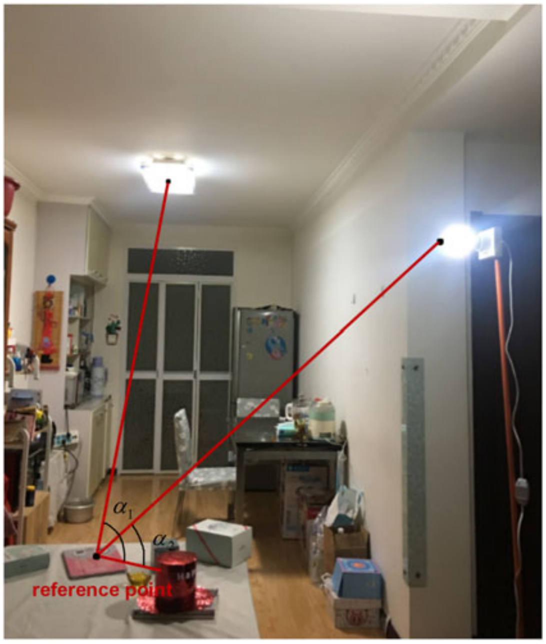

Fig. 3. Example of scene showing light sources and reference point.

6 Implementation Details

Parameter of Hemispherical Model. In the initial illumination estimation, we established a photon emission hemispherical model. The diameter of the hemisphere was set to twice the minimum diameter of the circle that can surround the entire scene. The number of light sources on the hemispherical model was set to 36 in our experiments. The direction of the photons emitted from the light source was limited to the hemisphere.

Handling of Transparent Objects. The refractive index of a transparent object was estimated using the transparent object and the caustics produced by it. In this step, we specified the 3D bounding box of the transparent object. The 3D bounding box was expanded by four times to obtain the space around the transparent object to depict the caustics of the transparent object. Using the reduced space instead of the entire scene to estimate the refractive index reduced the computation required.

7 Results and Discussion

To evaluate our approach, we have created eleven real scene databases. The input color images of real scene databases were captured using a second-generation Kinect sensor with a resolution of 1920×1080. The scene was lighted by an area light source or by an area light source plus a daylight simulator bulb. We manually recorded the angle α, which is the angle between the line connecting the real light source to the reference point in the scene and the plane of the reference point, as shown in Fig. 3. The angle β was between the line connecting the estimated light source to the reference point and the plane where the reference point was located. The error angle $\theta = |\alpha - \beta|$ was used to measure the accuracy of the illumination estimation. All experiments were performed on MATLAB and Visual Studio with an Intel Core i7 4.2 GHz CPU and 16 GB of memory.

In the initial illumination estimation phase, because the caustics produced by the specular and transparent objects

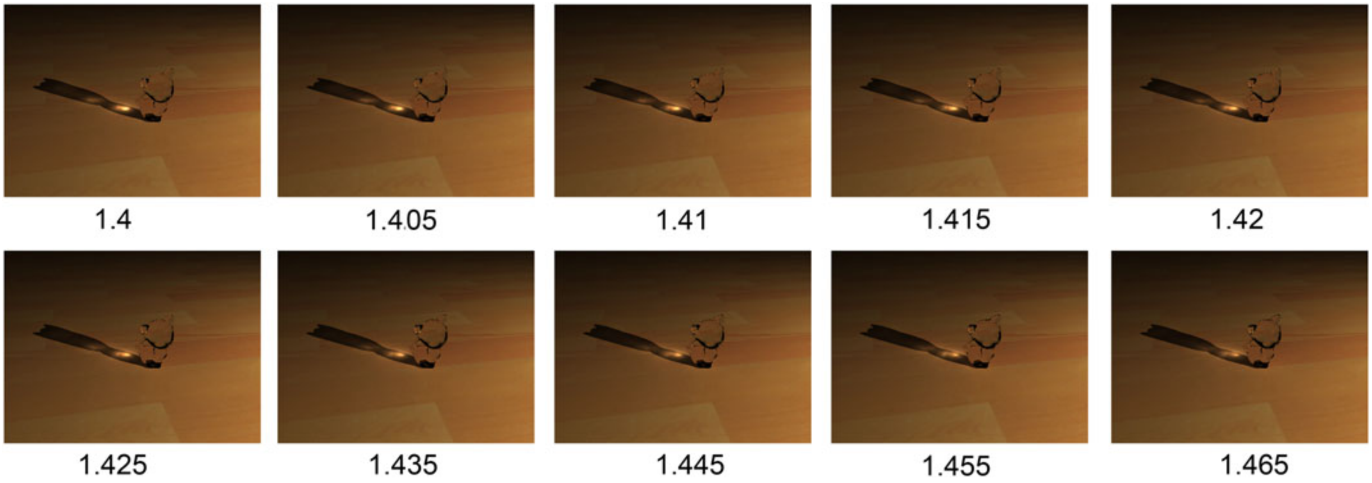

Fig. 4. Different refractive index images. As the refractive index of the transparent object increased, the caustics changed.

were not considered. We set the number of emitted photons to 1000000. The effects of caustics must be considered when estimating the refractive index and inserting virtual objects; therefore, the number of emitted photons was set to 8000000, which yielded the correct generation of caustics in most cases.

In the calculation of the minimum change in the refractive index, the steps performed were as follows:

Stimuli. Stimulus comprise multiple computer-generated transparent object images rendered using photon mapping or bidirectional path tracing. The scene was illuminated using a point light source, and the refractive index of the transparent object varied from 1.4 to 1.465 in 10 steps, as shown in Fig. 4. Another transparent object (Fig. 5a) was rendered with the refractive index varying from 1.5 to 1.565. In addition to the point light source, we set different types of light sources, such as an area light source and a spot light source to render the transparent objects. The selected transparent objects had simple or fine shapes (Figs. 5b and 5c). The transparent object in Fig. 5b was illuminated by an area light source, and its refractive index varied from 1.3 to 1.38. The scene in Fig. 5c was rendered by a spot light source, where the refractive index of the transparent cup was from 1.3 to 1.375. Four experiments with different transparent objects and refractive index intervals were performed.

Subjects. Eight subjects participated in the experiment. One was the author, and the others were unaware of the purpose of the experiment. All of them have normal color vision and normal acuity.

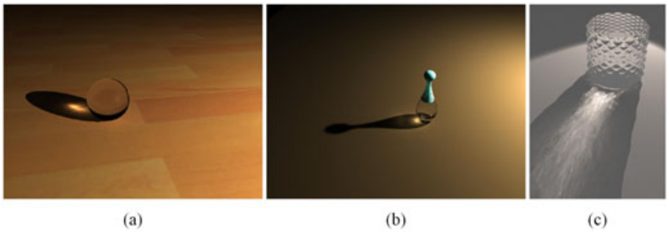

Fig. 5. Selected transparent objects for rendering. Every transparent object was rendered at different refractive indices, resulting in multiple images as stimulus.

Procedure. The subject used both eyes to view the monitor at a distance of 55 cm. In each trial, the subjects were presented with a 2 x 2 array of images (Fig. 6) and asked to indicate which pair (above or below) showed a greater within-pair difference in caustics. The subjects observed all combinations of refractive index in random order across 210 trials. Sufficient time was allotted to the subjects to respond to each trial.

Fig. 7 shows the estimated perceptual scales for eight subjects in different refractive index intervals along with the mean. For each transparent object, different refractive index steps were selected. We plotted the polylines of the experiments with the same steps in one graph to obtain Figs. 7a, 7b, and 7c. As shown in Fig. 7, when the change in refractive index was greater than 0.01, the human eye perceived a change in the caustics.

We evaluated the efficiency of our method in eight real scenes, where the objects in the scene included diffuse, specular, and transparent objects. Under the same experimental

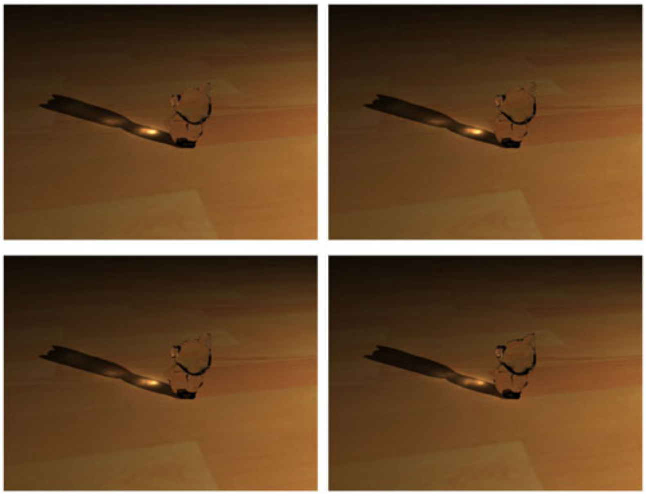

Fig. 6. 2 x 2 array of images. For each trial, the observer observed two pairs of images and decided the pair with the greater difference in caustics (above or below). The refractive indices of the transparent object in the above and below pair of images were 1.4 , 1.405, and 1.415, 1.455, respectively.

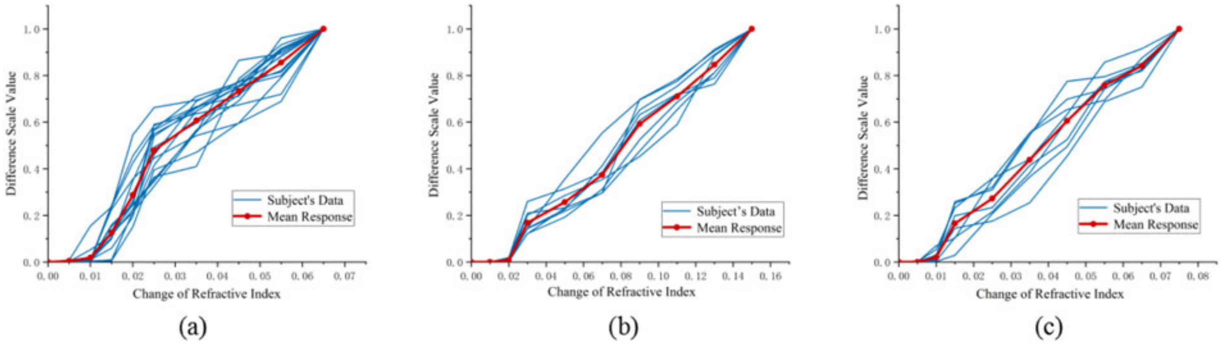

Fig. 7. Estimated difference scale for eight subjects in the maximum-likelihood difference-scaling task.

condition, the error angle obtained by our method was compared with the error angles produced using the methods in [6] and [32]. Table 2 shows the average value of the error angle comparison results. The method in [6] uses a local illumination model for illumination estimation, without considering the reflection between objects. A global illumination model was used inversely for light source estimation in our method. From the comparison between our method and that in [6], it was discovered that the accuracy of illumination estimation can be improved by considering the internal reflection of the objects. Comparing our method with that in [32], the highlight phenomenon of specular objects and the caustics of transparent objects enable not only the material of specular and transparent objects to be estimated, but also render the illumination estimation result more accurate.

In the distinction between transparent and specular objects, we ignored the Fresnel reflection of the model and assumed it to be transparent. The Fresnel reflection caused a transparent object to exhibit specular reflections, as shown in Fig. 8. The proposed method was utilized to distinguish whether the water bottle in Fig. 8 was a transparent or specular object. As multiple images from different perspectives were used for calculation, the object was still identified as a transparent object. Therefore, the presence of the Fresnel reflection did not affect the accuracy of the result.

To depict the result of the joint optimization of illumination and object material, we extracted the transparent and specular objects and rendered objects using the estimated parameters. As shown in Fig. 9, the transparent and specular objects were distinguishable, and their appearance was recovered. Figs. 9c and 9e show the original images, whereas Figs. 9d and 9f show the rendered images. In Figs. 9d and 9f, the black areas were caused by the incomplete reconstruction of the scene. However, this problem did not affect the insertion of virtual objects in the real scene. We will provide the fusion results later.

TABLE 2
Pixel Error Comparison for Multiple Scenes

Method	Chen *et al.* [6]	Zhang *et al.* [32]	Our Method
Error angle	19.1° ± 6.4°	13.8° ± 2.5°	8.4° ± 2°

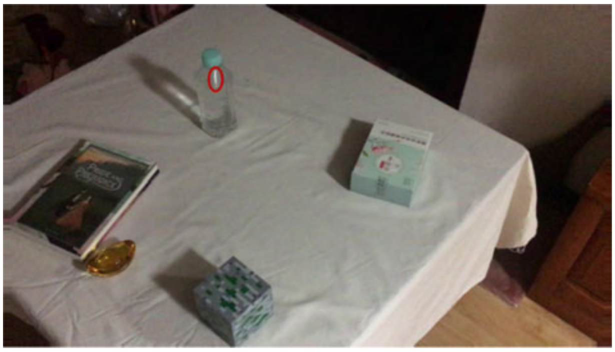

Fig. 8. Example scene showing a transparent water bottle with Fresnel reflection. Fresnel reflection produced by the transparent water bottle is denoted by the red ellipse.

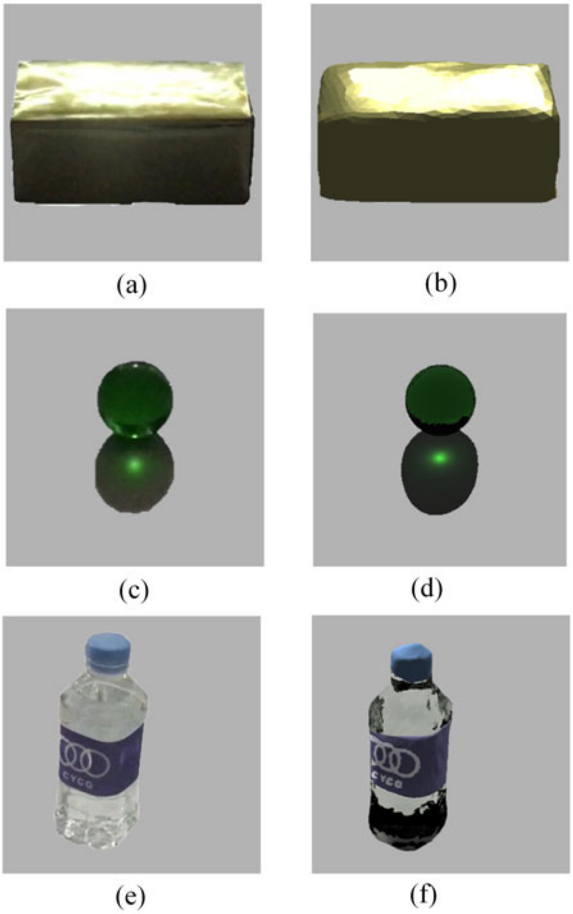

Fig. 9. Appearance recovery of transparent and specular objects. Left column is the original photograph of the objects, and right column is the image rendered using estimated illumination and object parameters.

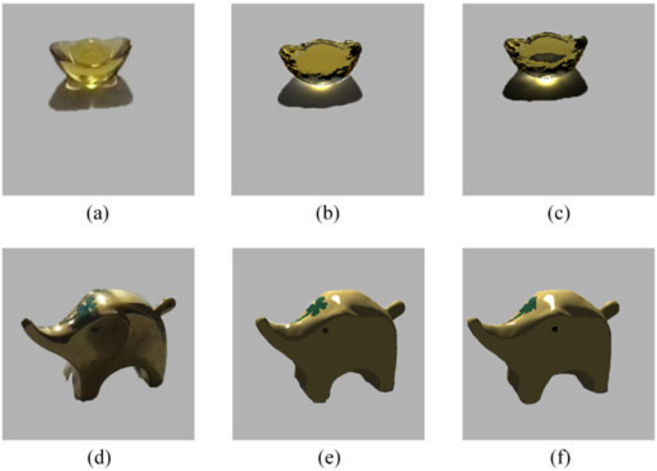

Fig. 10. Comparisons of recovery appearance with and without camera pose optimization. (a) and (d) show the original photographs of the objects. (b) and (e) show the photographs of recovered objects after adding camera pose optimization; (c) and (f) show the photographs of recovered objects without camera optimization.

Fig. 10 shows the effect of camera pose optimization on the recovery results. The optimized camera poses yielded a more accurate joint estimation of the materials and illumination.

The main purpose of our study is to insert virtual objects around transparent and specular objects in a real scene while maintaining a coherent appearance. Differential rendering [23] was required to insert virtual objects into the real scene. The scene was rendered twice using the estimated lighting and material of the transparent and specular objects, i.e., the one without virtual objects and the one with virtual objects. Provided that the material of the other diffuse objects are consistent in the two renderings, our fusion images can be completed by differential rendering. We compare our fusion result with [19] in Fig. 11. The estimation result of [19] can only handle the shadow direction of virtual objects; however, our method can process the effects of transparent and specular objects on virtual objects. In addition, transparent objects in a real scene were not considered in [19]. As shown in Figs. 11l and Figs. 11o, the effects of the shadows from the transparent objects on the virtual objects cannot be expressed. In scene 7, the caustic effect of the water bottle is shown on the virtual hat using our method. A method that does not consider transparent objects is used as in [19]; therefore, the caustics of the water bottle cannot be shown on the virtual hat. In scene 8, as the caustics effect of the transparent object is not obvious, Figs. 11z and 11A are not different significantly. In scenes 1, 2, 3, 5, and 6, comparing with the study in [19], the caustics effect of specular objects is shown on virtual objects using our method. In our experiments, to demonstrate the caustics effect more clearly, the materials of the caustics receivers in real scenes and the inserted virtual objects are all diffuse. In addition, we show comparison of fusion images to ground truth in Fig. 12, in which the image obtained using our method is compared with the image generated by a method that is without considering inter-reflection and real objects with complex materials. Comparing the appearance of white virtual object in Figs. 12b and 12c, our method can recover the color bleeding of the real blue box. Also, the effect of caustics produced by the transparent ball and specular red box on virtual objects can be depicted utilizing our method as shown in Figs. 12e, 12f, 12h and 12i. Moreover, our fusion images are plausible by comparing the inserted virtual objects with the ground truth.

Our method is widely applicable. When no transparent and specular objects exist in the scene, the initial light estimation method can be used to estimate the illumination. Joint optimization can be performed provided that transparent or specular objects exist in the scene. In Fig. 13, we plot the curves with respect to the error angle on the x-axis and the photometric error in Eq. (1) (log scale) on the y-axis. As shown, the optimal result can be obtained by separately optimizing the illumination and the material of the transparent or specular object. Therefore, to reduce the calculation time, when both specular and transparent objects exist in the scene, the specular object can be selected for the joint optimization of illumination and material as generating the caustics is time consuming. The optimized illumination can then be directly used to estimate the parameters of the transparent object.

The joint optimization required approximately 30 min, in which 30, 20, and 50 percent of the time were spent on the initial illumination estimation, optimization for illumination and material of specular object, and estimation of transparent object parameters, respectively. Owing to the many generations of caustics of the transparent objects, the most time consuming step was the estimation of the transparent object parameters. Using an improved real-time caustics generation algorithm or exploiting the computing power of the GPU, the time for estimating the transparent object parameters can be reduced significantly.

Our method may be less successful in some cases. First, the visual hull was based on contours to recover the object models but failed to recover concavities in objects. As shown in Figs. 10a and 10b, the concavity of the ingot was not completely displayed, which resulted in some differences between the effects of the final rendered caustics and the real caustics. Subsequently, the specular objects sometimes produced highlights at the edges, which caused the extracted contours to shrink slightly. In Fig. 11q, because the contour extraction of the elephant was reduced, part of the shadow of the inserted virtual object occluded the real object. Next, when identifying areas of transparent and specular objects, nonspecular opaque objects were identified as specular objects occasionally. The mouse in Fig. 14 was identified as a possible area where specular or transparent objects existed; however, this did not affect the final differential rendering. In addition, as for outdoor environments, estimating directional light sources like the sun should be possible in our method. If there are multiple light sources in the indoor scene, the light sources may interfere with each other. For example, light sources with high intensity value will cover the surrounding light sources with low intensity value, which makes it difficult to discern the number of light sources by images. Finally, as shown in Figs. 12a and 12b, rendering the virtual object with the estimated point light source may cause some differences between the rendered virtual white box and the real white box in shadow and inter-reflection. However, people may not be able to perceive the difference when there is only the inserted virtual object is displayed without the real object for comparison.

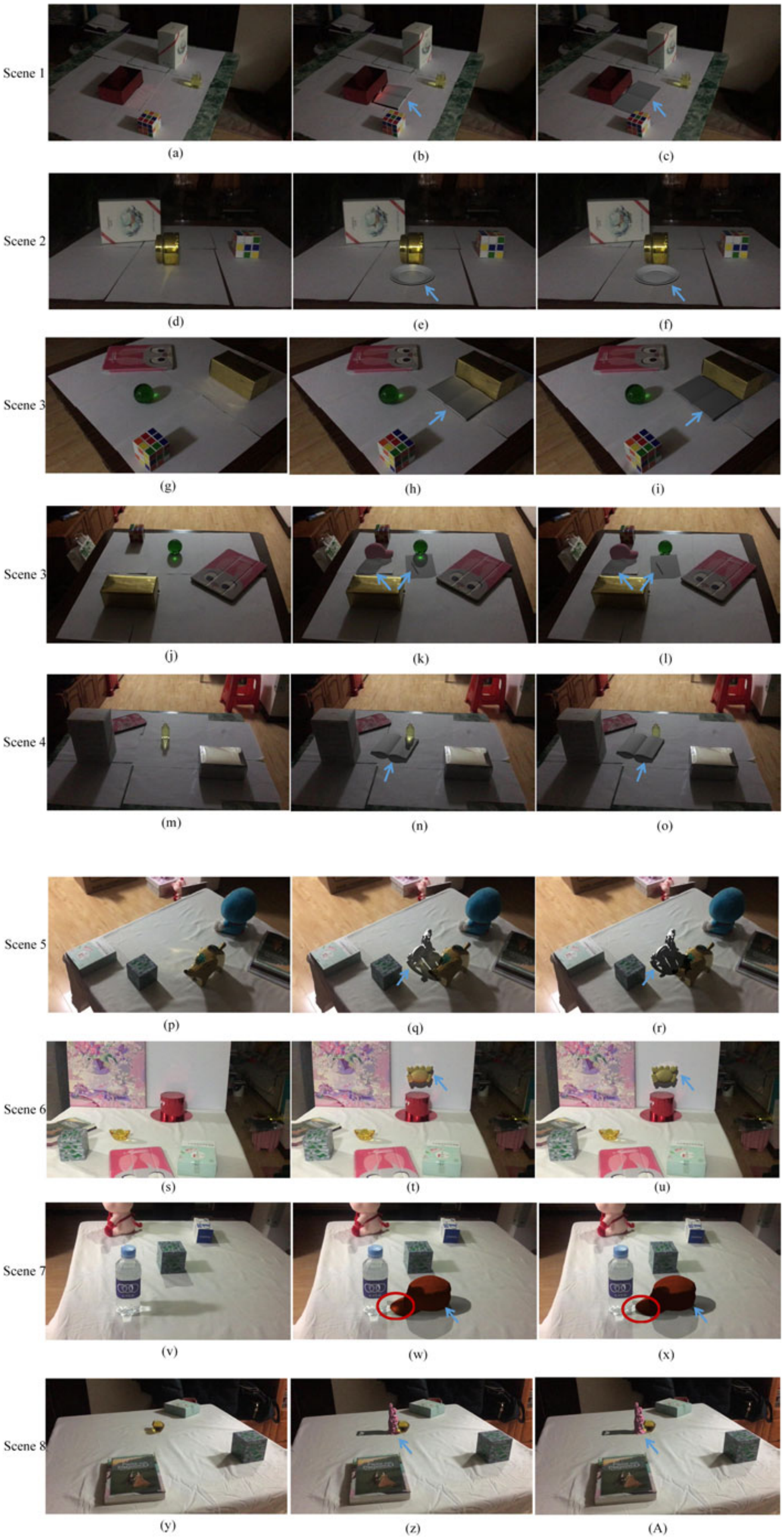

Fig. 11. Comparison of several fusion results. From the left column to the right: original captured scene images, fusion results using our proposed method, and fusion results using [19]. The object pointed by the arrow is the inserted virtual object.

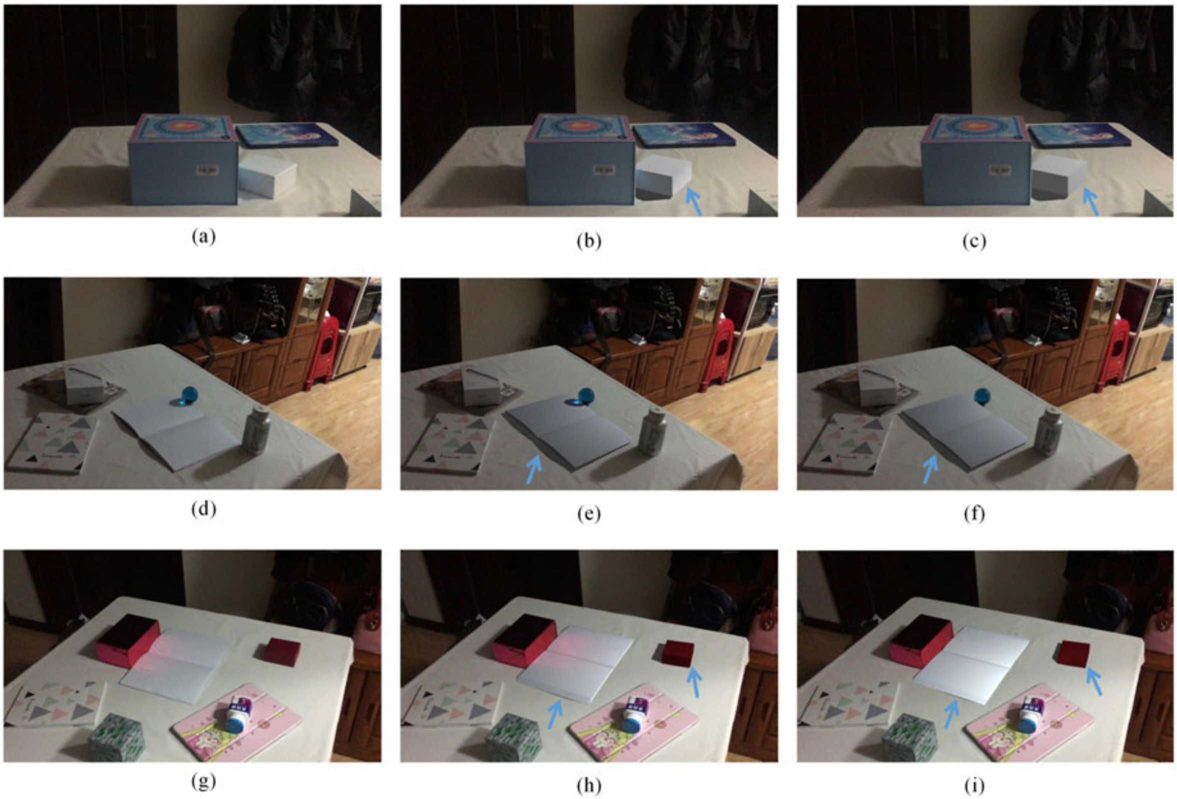

Fig. 12. Comparison of fusion images to ground truth. From the left column to the right: Ground truth images with real objects; fusion images with inserted virtual objects using our method and [6]. The object pointed by the arrow is the inserted virtual object.

8 Conclusions And Future Work

In this paper, we solved the problem of differential rendering for augmented reality beyond the Lambertian-world assumption. The proposed framework can jointly estimate the materials of transparent objects and specular objects in real world, and simulate the effect of caustics on virtual inserted objects. We combined the psychological scale method to estimate the refractive index of transparent objects, so that virtual objects can be inserted around the transparent objects realistically under the condition of

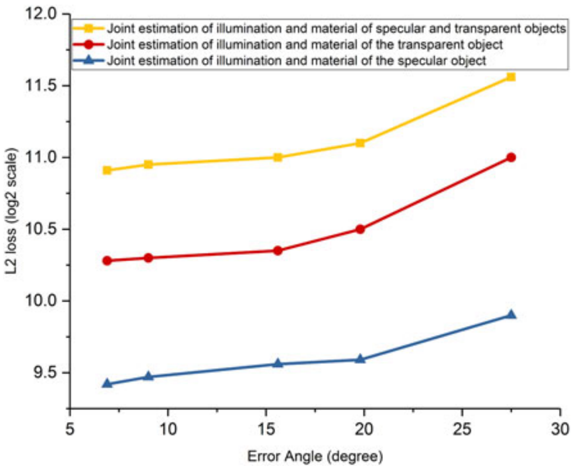

Fig. 13. Relationship between photometric error (Eq. (1)) and angle error of light source under different conditions. In different cases, photometric error is reduced as the error angle decreases.

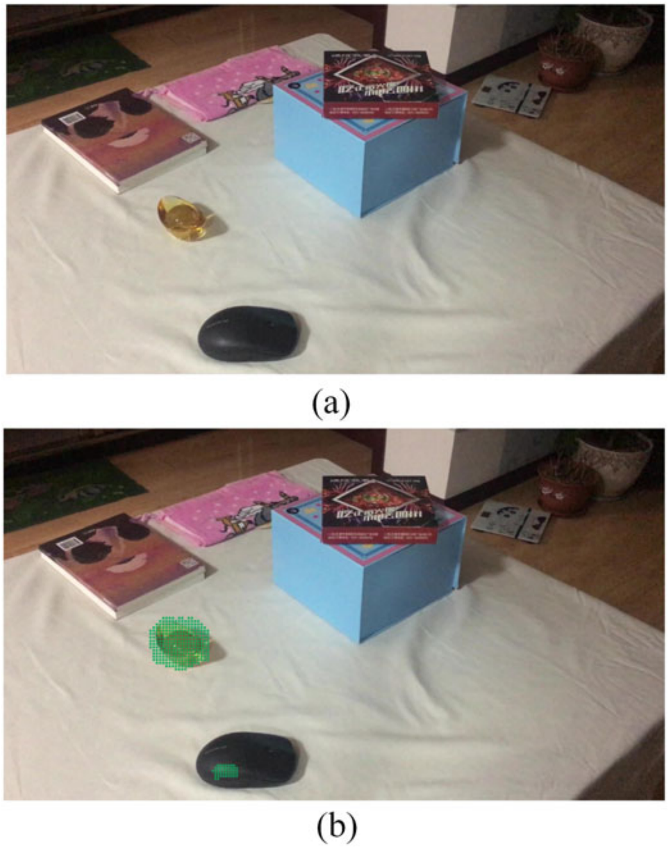

Fig. 14. Possible areas of transparent and specular objects. (a) shows the original image; (b) shows possible areas of specular and transparent objects, which are indicated by the green grid in (b).

considering the visual characteristics of human eyes. Our method can be used to obtain a more compelling fusion effect in an augmented reality system with real objects of complex materials.

In the future, we would like to extend our approach to analyze more complex specular objects, such as anisotropic objects. To obtain a more plausible fusion result, we will consider Fresnel reflections when estimating the parameters of transparent objects.

Acknowledgments

The authors would like to thank the authors of [7] and Zhenxing Li for the help and suggestions. This work was supported by the National Key R&D Plan (2017YFB1002900) and the National Natural Science Foundation of China (No.61771220 and No. 61631009).

References

[1] N. Alt, P. Rives, and E. Steinbach, "Reconstruction of transparent objects in unstructured scenes with a depth camera," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 4131–4135.

[2] A. Torres-Gomez and W. Mayol-Cuevas, "Recognition and reconstruction of transparent objects for augmented reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2014, pp. 129–134.

[3] Y. Ji, Q. Xia, and Z. Zhang, "Fusing depth and silhouette for scanning transparent object with RGB-D sensor," *Int. J. Opt.*, vol. 2017, pp. 1–11, 2017.

[4] B. Liu, K. Xu, and R. R. Martin, "Static scene illumination estimation from videos with applications," *J. Comput. Sci. Technol.*, vol. 32, no. 3, pp. 430–442, 2017.

[5] B. J. Boom, S. Orts-Escolano , X. X. Ning, S. McDonagh, P. Sandilands, and R. B. Fisher, "Interactive light source position estimation for augmented reality with an RGB-D camera," *Comput. Animation Virt. Worlds*, vol. 28, no. 1, 2015, Art. no. e1686

[6] X. W. Chen, X. Jin, and K. Wang, "Lighting virtual objects in a single image via coarse scene understanding," *Sci. China Inf. Sci.*, vol. 57, no. 9, pp. 1–14, 2013.

[7] H. Wu, Z. Wang, and K. Zhou, "Simultaneous localization and appearance estimation with a consumer RGB-D camera," *IEEE Trans. Vis. Comput. Graphics.*, vol. 22, no. 8, pp. 2012–2023, Aug. 2016.

[8] D. Azinovic, T.-M. Li, A. Kaplanyan, and M. Niener, "Inverse path tracing for joint material and lighting estimation," in *Proc. CVPR.*, 2019, pp. 1–14.

[9] G. Oxholm and K. Nishino, "Shape and reflectance estimation in the wild," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 376–389, Feb. 2016.

[10] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1670–1687, Aug. 2015.

[11] H. Ren, H. Qiu, F. He, and K. Leng, "A survey on image-based approaches of synthesizing objects," in *Proc. Int. Conf. Virt. Reality Vis.*, 2017, pp. 264–269.

[12] I. Lysenkov, V. Eruhimov, and G. Bradski, "Recognition and pose estimation of rigid transparent objects with a kinect sensor," in *Proc. Robot., Sci. Syst.*, 2012, pp. 273–280.

[13] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[14] S. Lee, and S. K. Jung, "Estimation of illuminants for plausible lighting in augmented reality," in *Proc. Int. Symp. Ubiquitous Virt. Reality.*, 2011, pp. 17–20.

[15] M. Kanbara and N. Yokoya, "Real-time estimation of light source environment for photorealistic augmented reality," in *Proc. Int. Conf. Pattern Recognit.*, 2004, pp. 911–914.

[16] A Panagopoulos, D Samaras, and N Paragios, "Robust shadow and illumination estimation using a mixture model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 651–658.

[17] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, "Illumination estimation and cast shadow detection through a higher-order graphical model," in *Proc. CVPR.*, 2011, pp. 673–680.

[18] S. Jiddi, P. Robert, and E. Marchand, "Reflectance and Illumination Estimation for Realistic Augmentations of Real Scenes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2017, pp. 244–249.

[19] S . Karaoglu, Y. Liu, T. Gevers, and A. W. M. Smeulders, "Point light source position estimation from RGB-D images by learning surface attributes," *IEEE Trans. Image Process*, vol. 26, no. 11, pp. 5149–5159, Nov. 2017.

[20] M. Gardner *et al.*, "Learning to predict indoor illumination from a single image," *ACM Trans. Graph.*, vol. 36, no. 6, 2017, pp. 1–14.

[21] Y. Hold-Geoffroy , K. Sunkavalli, S. Hadap, E. Gambaretto, and J. Lalonde, "Deep Outdoor Illumination Estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, Art. no. 2373–2382.

[22] W. Henrique, P. Donald , and J.-F., Lalonde, "Learning to estimate indoor lighting from 3D objects," in *Proc. Int. Conf. 3D Vis.*, 2018, pp. 199–207.

[23] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. SIGGRAPH*, 1998, pp. 189–198.

[24] R. A. Newcombe, S. Izadi, and O. Hilligesetal, "KinectFusion: Real-time dense surface mapping and tracking," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, pp. 127–136 2011.

[25] B. Curless and M. Levoy, "Volumetric method for building complex models from range images," in *Proc. 23rd Annu. Conf. Comput. Graph. Interactive Techn.*, pp. 303–312, 1996.

[26] K. Narayan, J. Sha, A. Singh, and P. Abbeel, "Range sensor and silhouette fusion for high-quality 3D scanning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 3617–3624.

[27] H. W. Jensen, "Global illumination using photon maps," in *Proc. Eurogr. Workshop Rendering Techn.*, 1996, pp. 21–30.

[28] K. Kim, J. Gu, S. Tyree, P. Molchanov, M. Niener, and J. Kautz, "A lightweight approach for on-the-fly reflectance estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 20–28.

[29] C. Yu, Y. Seo, and S. Lee, "Global optimization for estimating a BRDF with multiple specular lobes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 319–326.

[30] C. Charrier, L. Maloney, H. Cherifi, and K. Knoblauch, "Maximum likelihood difference scaling of image quality in compression-degraded images," *J. Opt. Soc. America A.*, vol. 24, no. 11, pp. 3418–3426, 2007.

[31] Q.-Y. Zhou and V. Koltun, "Color map optimization for 3D reconstruction with consumer depth cameras," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, 2014.

[32] A. J. Zhang, Y. Zhao, and S. G. Wang, "Illumination estimation for augmented reality based on a global illumination model," *Multimedia Tools Appl.*, vol. 78, no. 23, pp. 33487–33503, 2019.

[33] S. Georgoulis, K. Rematas, and T. Ritschel, "Reflectance and natural illumination from single-material specular objects using deep learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1932–1947, Aug. 2018.

[34] Y. Yu and W. A. P. Smith, "InverseRenderNet: Learning single image inverse rendering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3150–3159.

[35] S. Sengupta, J. Gu, K. Kim, G. Liu, D. Jacobs, and J. Kautz, "Neural inverse rendering of an indoor scene from a single image," in *Proc. ICCV.*, 2019, pp. 8598–9607.

[36] S. Song and T. Funkhouser, "Neural illumination: Lighting prediction for indoor environments," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6911–6919.

[37] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, and H. Jin, "Automatic Scene Inference for 3D object compositing," *ACM Trans. Graph.*, vol. 33, no. 3, pp. 1–15, 2014.

[38] L. Gruber, T. Trummer, and D. Schmalstieg, "Real-time photometric registration from arbitrary geometry," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2012, pp. 119–128.

Aijia Zhang received the BSc degree in communication engineering, from Jilin University, China, in 2017. She is currently working toward the PhD degree with Jilin University. Her main research interest is illumination estimation in 3D scene.

Yan Zhao (Member, IEEE) received the BS degree in communication engineering, from the Changchun Institute of Posts and Telecommunications, in 1993, the MS degree in communication and electronic, from the Jilin University of Technology, in 1999, and the PhD degree in communication and information system, from the Jilin University, in 2003. She currently is a professor of communication engineering. Her research interests include image and video processing, multimedia signal processing.

Shigang Wang (Member, IEEE) received the BS degree, from Northeastern University, in 1983, the MS degree in communication and electronic, from the Jilin University of Technology, in 1998, and the PhD degree in communication and information system, from Jilin University, in 2001. He is currently a professor of communication engineering. His research interests include image and video coding, multidimensional signal processing.