RENDERING VIRTUAL OBJECTS INTO REAL SCENE BY USING A DIFFUSE SPHERE

Yuji Kanda[1], Fumihiko Sakaue[1], Jun Sato[1], Kazuhisa Ishimaru[2] and Masayuki Imanashi[3]

[1]*Nagoya Institute of Technology, 466-8555 Nagoya, Japan*

[2]*NIPPON SOKEN Inc., 445-0012 Nishio, Japan*

[3]*DENSO Corporation, 448-8661 Kariya, Japan*

Keywords: Augmented Reality, Diffuse Sphere, Light Source Distribution, Shading, Shadows.

Abstract: In this paper, we propose an efficient method for rendering virtual objects into real scene images. We put a diffuse sphere into the scene as a reference object, and use the direct relationship between the intensity of the diffuse sphere and the intensity of virtual objects for rendering the virtual objects in images. We also generate shadows of the virtual objects by using shadows of the reference sphere. As a result, arbitrary virtual objects and their shadows can be rendered quite efficiently from a single diffuse sphere put in the scene. The proposed method is tested in real image experiments, and evaluated quantitatively comparing with the existing method.

1 INTRODUCTION

Rendering virtual objects into real scenes is very useful for making movies and for designing augmented reality systems. For generating realistic augmented reality images, it is very important to generate accurate shade and shadow of virtual objects in images.

The existing methods generate shade and shadow of objects by estimating light source distributions of the scene (Marschner and Greenberg, 1997). The spherical mirror is often used for measuring light source distributions (Debevec, 1998; Kanbara and Yokoya, 2004). In these methods, a camera observes the spherical mirror put in the scene, and the light source distributions are measured from the camera image directly. Although these methods are simple and efficient, cameras in general do not have enough dynamic range to measure the light source distributions directly. Thus, these methods require to take multiple images changing the shutter speed of the camera. For solving the problem, known diffuse objects such as diffuse spheres were also used for estimating the light sources (Zhang and Yang, 2001; Wei, 2003; Takai et al., 2004; Weber and Cipolla, 2001). These methods use the boundary lines on diffuse objects, which separate illuminated and un-illuminated areas for estimating light source orientations. However, extracting these boundary lines on diffuse objects is not easy in practice. Furthermore, it is difficult for these methods to estimate light source distributions which consist of many light sources. As shown in these research, accurate estimation of light source distributions is not easy in practice.

Figure 1: The reference sphere and a plane put in the scene. We use a diffuse sphere and a plane for generating shading information and shadows of virtual objects.

Thus, we in this paper propose an efficient method for rendering virtual objects into real scenes without estimating light source distributions as far as possible. In our method, we use a diffuse sphere put in the scene as a reference object as shown in Fig. 1. Our method separates the task of shading and the task of shadowing. We separate these two tasks, since we generate shading information without estimating light source distributions of the scene, while we generate shadows by estimating the light source distributions.

For generating shading information of virtual ob-

Kanda Y., Sakaue F., Sato J., Ishimaru K. and Imanishi M..
RENDERING VIRTUAL OBJECTS INTO REAL SCENE BY USING A DIFFUSE SPHERE.
DOI: 10.5220/0003850202790285
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 279-285
ISBN: 978-989-8565-03-7

jects, we use the direct relationship on intensity between the virtual object and the diffuse sphere. The proposed method is very efficient and reasonably accurate, since this method does not need to estimate light source distributions, and does not suffer from the estimation error of light source distributions.

For generating the shadow of virtual objects, we use the shadow of reference diffuse sphere. If we have a known object on a ground plane, the shadow of the object on the plane is a very useful cue to estimate light source distributions. Sato et al. (Sato et al., 2003) proposed a method for estimating light source distributions from shadows on a plane, which are generated by known objects. However, their method assumes that there is no inter reflection between the object and the plane. Hence, their method suffers from the inter reflections, if the reflectance of the object is large. Thus, we in this paper estimate light source distributions from shadows of a diffuse sphere taking account of the effect of inter reflections between the sphere and the plane. As a result, virtual objects and their shadows can be rendered quite efficiently and accurately from a single diffuse sphere put in the scene.

In the following sections, we first propose an efficient method for generating shading information of virtual objects from a reference diffuse sphere without using light source information. We next propose a method for generating shadows by estimating light source distributions from the shadow of the reference sphere. We in this paper assume all the light sources are at infinity, and virtual objects put into the scene do not emit any light and are Lambertian.

2 SYNTHESIZING VIRTUAL OBJECTS INTO THE SCENE

The standard method for synthesizing virtual objects into real scene is based on the estimation of light sources of the scene. Suppose we have M light sources at infinity, whose directions are represented by unit vectors $\mathbf{s}_k$ $(k = 1, \cdots, M)$, and magnitudes are $\mathbf{E}_k = [E_k^R, E_k^G, E_k^B]^\top$ $(k = 1, \cdots, M)$ in red, green and blue color channels. If the object surface is Lambertian, the image intensity $\mathbf{I}_i = [I_i^R, I_i^G, I_i^B]^\top$ of ith point on the surface illuminated by $\mathbf{E}_k$ $(k = 1, \cdots, M)$ can be described as follows:

$$\mathbf{I}_i = \sum_{k=1}^{M} \mathbf{C}\mathbf{R}_i\mathbf{E}_k v_{ik} \mathbf{n}_i^\top \mathbf{s}_k \tag{1}$$

where $\mathbf{n}_i$ denotes the surface normal at the ith point, and v_{ik} denotes a visibility function, which takes 1 if the ith point is illuminated from the kth light source, and takes 0 in the other case. $\mathbf{R}_i$ is a 3×3 diagonal matrix whose diagonal components are the reflectance of the surface point ρ_i^R, ρ_i^G, ρ_i^B in R, G and B components as follows:

$$\mathbf{R}_i = \begin{bmatrix} \rho_i^R & 0 & 0 \\ 0 & \rho_i^G & 0 \\ 0 & 0 & \rho_i^B \end{bmatrix} \tag{2}$$

$\mathbf{C}$ denotes the characteristic function of the camera, which represents the response of the camera in R, G and B channels, i.e. the gain of output signals with respect to input signals. In general the crosstalk among R, G and B channels is very small and negligible, and thus we assume the matrix $\mathbf{C}$ is also diagonal as follows:

$$\mathbf{C} = \begin{bmatrix} C^R & 0 & 0 \\ 0 & C^G & 0 \\ 0 & 0 & C^B \end{bmatrix} \tag{3}$$

where, C^R, C^G and C^B denote gains in each channel. We in this paper assume that the gamma correction does not exist in the camera, and its characteristic function is linear.

From (1), we find that we need light source information $\mathbf{E}_k$ in each orientation $\mathbf{s}_k$ and camera characteristics $\mathbf{C}$ for synthesizing the image intensity of virtual objects. Thus, the existing methods estimate light source information of the scene by using various cues, such as shadows and shading, and measure the characteristic function of the camera by using color chart etc. However, the results suffer from the measurement errors of these components.

Thus, we in this paper propose a method which enables us to compute the image intensity of virtual objects without estimating light sources and camera characteristic functions. In our method, we put a diffuse sphere in the scene as a reference object as shown in Fig. 1, and use its image information for synthesizing virtual objects into the scene. Since the sphere has surface normals in all the directions, the surface normal of any point on a virtual object in the scene has its reference surface normal on the sphere. Although not all the surface normals are visible in a single camera image, all the visible surface normals of virtual objects have their references in the visible part of the sphere, if we assume affine projection.

Now, let us consider an image intensity $\mathbf{I}'_j$ of a jth point on the reference sphere, whose surface normal $\mathbf{n}'_j$ is identical with the surface normal $\mathbf{n}_i$ of ith point on the virtual object, i.e. $\mathbf{n}'_j = \mathbf{n}_i$. Then, $\mathbf{I}'_j$ can be described as follows:

$$\mathbf{I}'_j = \sum_{k=1}^{M} \mathbf{C}\mathbf{R}'_j\mathbf{E}_k v'_{jk} \mathbf{n}_i^\top \mathbf{s}_k \tag{4}$$

where, $\mathbf{R}'_j$ is the reflectance matrix of the jth point on the reference sphere, and v'_{jk} is a visibility function of the point with respect to the kth light source. If the object is convex, v_{ik} is identical with v'_{jk}. If the object is not convex, v_{ik} is not completely identical with v'_{jk}. However, in most of the case, v_{ik} is close to v'_{jk}. Thus, we in this paper assume v_{ik} and v'_{ik} are identical.

Since both $\mathbf{C}$ and $\mathbf{R}$ are diagonal, (1) and (4) can be described as follows:

$$\mathbf{I}_i = \mathbf{R}_i\mathbf{L} \tag{5}$$

$$\mathbf{I}'_j = \mathbf{R}'_j\mathbf{L} \tag{6}$$

where,

$$\mathbf{L} = \sum_{k=1}^{M} \mathbf{C}\mathbf{E}_k v_{ik} \mathbf{n}_i^\top \mathbf{s}_k \tag{7}$$

From (5) and (6), we find that the image intensity $\mathbf{I}_i$ of the virtual object can be computed from the image intensity $\mathbf{I}'_j$ of the reference sphere as follows:

$$\mathbf{I}_i = \mathbf{R}_i \mathbf{R}_j'^{-1} \mathbf{I}'_j \tag{8}$$

As a result, we do not need to estimate light source distributions nor characteristic function of cameras for synthesizing virtual objects into the real scene images. Thus, the proposed method is very simple and efficient.

Since the image of the reference sphere is digital and the surface normals of the reference sphere are discrete in practice, we compute the image intensity $\mathbf{I}'_j$ of a point on the reference sphere, whose surface normal is identical with that of the virtual object, by using the linear interpolation. To ensure a reasonable accuracy of synthetic image generation, the reference sphere in the image should not be too small.

3 SYNTHESIZING SHADOW OF VIRTUAL OBJECTS

We next consider a method for generating shadow of objects. For realistic virtual images, it is very important to generate realistic shadows on the floor. We generate shadow of virtual objects by using the same reference sphere used for shading virtual objects.

Suppose the reference sphere is put on a planar Lambertian surface. Then, the shadows of the sphere appears on the planar surface, and they are visible from the scene camera. These image shadows provide us useful information for generating the shadows of virtual objects put in the scene. Sato et al. (Sato et al., 2003) proposed an efficient method for estimating light source distributions from shadows of scene

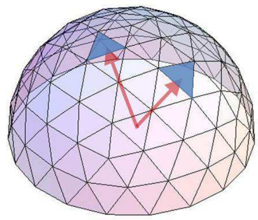


Figure 2: Sampling of light source distributions on a geodesic dome. Each patch on the dome represents the direction of each light source.

objects. We extend their method for estimating light source distributions and generating shadows from the reference sphere. Sato et al. assumed that there is no inter reflection between the scene objects and the planar surface, and estimated the light source distribution from the shadows of the scene objects on the planar surface. This assumption is reasonable if the reflectance of objects is small, and the form factors between the object and the plane are also small. The form factor is the visibility between two patches on surfaces, and it becomes large if the relative orientation between normals of these two patches is close to π radian and these patches are close to each other. In our method, we use a reference sphere put on a plane. The reference sphere and the plane are very close to each other at around the contact point of these two objects. Furthermore, the relative orientation of two surfaces is close to π radian at around the contact point. Thus, inter reflections are not negligible in the estimation of light source distributions. Therefore, we propose a method for estimating light source distributions and generating shadows taking account of inter reflections.

3.1 Sampling of Light Source Distribution

In order to estimate light source distributions under multiple lights, we represent the light source distributions by using a geodesic dome as shown in Fig.2. In this model, a light source distribution is represented by the magnitude of light source $\mathbf{E}_k = [E_k^R, E_k^G, E_k^B]^\top$ $(k = 1, \cdots, M)$ in M light source directions $\mathbf{s}_k$ $(k = 1, \cdots, M)$, which are represented by M patches on the geodesic dome. Since $\mathbf{s}_k$ $(k = 1, \cdots, M)$ is predefined on the geodesic dome, the estimation of a light source distribution is same as the estimation of $\mathbf{E}_k$ $(k = 1, \cdots, M)$.

3.2 Light Source Distributions from Shadows of Reference Objects

Let us consider a reference object, which consists of a sphere and a plane with known reflectance, as shown in Fig. 1. Although many other objects exist in the 3D space, we consider these other objects as light sources at infinity, and these light sources illuminate the reference object. Hence, the inter reflection occurs between the reference sphere and the reference plane, but it does not occur between the reference object and other objects in the 3D space. This assumption is valid in most of the case, since the distance between the reference object and other objects are much larger than the distance between the reference sphere and the reference plane in general. Thus, we model the 3D world by the reference object and light sources at infinity, which are represented by a geodesic dome.

Suppose we have N patches on the reference object, and M patches on the geodesic dome. Then, M patches on the geodesic dome emit light, and the inter reflection occurs among N patches on the reference object. Then, the following radiosity equation holds:

$$\mathbf{I}_i = \sum_{j=1}^{N} F_{ij}\mathbf{R}_i\mathbf{I}_j + \sum_{k=1}^{M} G_{ik}\mathbf{C}\mathbf{R}_i\mathbf{E}_k \tag{9}$$

where, the first term in the right side of (9) represents the inter reflection and the second term represents the direct illumination by light sources on the geodesic dome.

F_{ij} denotes the form factor between ith and jth patches on the reference object, and G_{ik} denotes the form factor between ith patch on the reference object and kth patch on the geodesic dome. These are described as follows:

$$F_{ij} = \frac{A_j v_{ij} \cos\theta_i \cos\theta_j}{\pi r^2} \tag{10}$$

$$G_{ik} = \frac{A_k v_{ij} \cos\theta_i}{\pi} \tag{11}$$

where, A_i denotes the area of ith patch, and θ_i denotes the angle between the surface normal of ith patch and a ray between two patches. r is the distance between two patches. Note, r^2 term does not exist in G_{ik}, since each patch on the geodesic dome represents a light source at infinity, and its energy does not decrease with respect to the distance r. Also, each patch on the geodesic dome is front-parallel, and thus $\cos\theta_k = 1$. This is why $\cos\theta_k$ does not exist in (11).

Since we have N patches on the reference object and M patches on the geodesic dome, we can derive the following system of linear equations from (9).

$$\begin{bmatrix} G_{11}\mathbf{R}_1 & \cdots & G_{1M}\mathbf{R}_1 \\ \vdots & & \vdots \\ G_{N1}\mathbf{R}_N & \cdots & G_{NM}\mathbf{R}_N \end{bmatrix} \begin{bmatrix} \mathbf{E}'_1 \\ \vdots \\ \mathbf{E}'_M \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 - \sum F_{1j}\mathbf{R}_1\mathbf{I}_j \\ \vdots \\ \mathbf{I}_N - \sum F_{Nj}\mathbf{R}_N\mathbf{I}_j \end{bmatrix} \tag{12}$$

where, $\mathbf{E}'_k$ is a light source distribution multiplied by the camera characteristic function $\mathbf{C}$ as follows:

$$\mathbf{E}'_k = \mathbf{C}\mathbf{E}_k \qquad (k = 1, \cdots, M) \tag{13}$$

We in this paper estimate $\mathbf{E}'_k$ instead of $\mathbf{E}_k$, which enables us to generate virtual shadows without knowing camera parameters $\mathbf{C}$. If $N \geq M$, we can estimate the light source distribution $\mathbf{E}'_k$ $(k = 1, \cdots, M)$ by solving (12) taking account of the inter reflection between the reference sphere and the reference plane.

3.3 Synthesizing Shadows

Once the light source distribution is estimated, the shadow of virtual objects can be generated from the distribution.

Let us consider an image point P in the ith patch on the ground plane, whose image intensity is $\mathbf{I}_P$ and reflectance is $\mathbf{R}_P$. Let G_{ik} be the form factor between the ith patch on the ground plane and kth patch on the geodesic dome before putting virtual objects, and let G'_{ik} be that after putting the virtual objects. Both G_{ik} and G'_{ik} can be computed easily from the geometric relationship between the virtual objects and the light source directions. Then, if we neglect inter reflections, the image intensities of point P before and after putting the virtual objects are described as follows:

$$\mathbf{I}_P = \sum_{k=1}^{M} G_{ik}\mathbf{R}_P\mathbf{E}'_k \tag{14}$$

$$= \mathbf{D}\,\mathbf{r}_P \tag{15}$$

$$\mathbf{I}'_P = \sum_{k=1}^{M} G'_{ik}\mathbf{R}_P\mathbf{E}'_k \tag{16}$$

$$= \mathbf{D}'\,\mathbf{r}_P \tag{17}$$

where, $\mathbf{D}$ and $\mathbf{D}'$ are diagonal matrices, whose diagonal components are $\sum_{k=1}^{M} G_{ik}\mathbf{E}'_k$ and $\sum_{k=1}^{M} G'_{ik}\mathbf{E}'_k$ respectively, and $\mathbf{r}_P$ is a three vector which consists of the three diagonal components of $\mathbf{R}_P$.

Then from (15) and (17), we find that the image intensity $\mathbf{I}'_P$ of point P after putting the virtual objects can be computed from the image intensity $\mathbf{I}_P$ of the point before putting the virtual objects as follows:

$$\mathbf{I}'_P = \mathbf{D}'\mathbf{D}^{-1}\mathbf{I}_P \tag{18}$$

As shown in (18), the image intensity of shadows can be computed without knowing reflectance $\mathbf{R}_P$ of the ground plane and camera parameters $\mathbf{C}$.

Figure 3: The image of the reference sphere and synthesized virtual objects (person, pod).

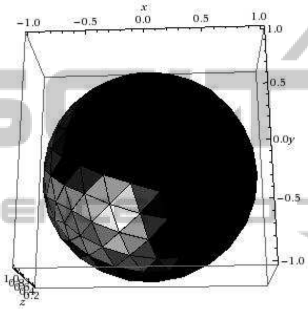


Figure 4: The light source distribution estimated from Fig. 3 by using the proposed method.

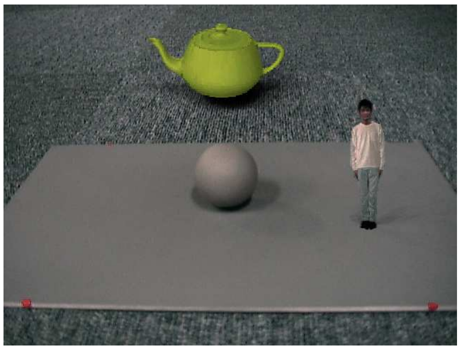


Figure 5: The image of the reference sphere and synthesized virtual objects (person, pod).

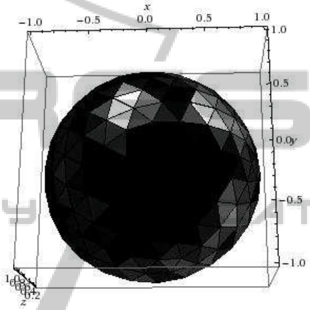


Figure 6: The light source distribution estimated from Fig. 5 by using the proposed method.

4 EXPERIMENTS

4.1 Real Image Experiments

We next show the results of some real image experiments. In these experiments, we synthesize virtual objects into real scene images. The virtual objects are rendered by using the method shown in section 2, and their shadows are rendered by using the method shown in section 3.

Fig. 3 shows the image of a reference sphere and a reference plane. The intensity of the sphere was used for rendering virtual objects into the image, and the shadow of the sphere was used for estimating the light source distributions and rendering shadows of the virtual objects into the image. Fig. 4 shows the light source distribution estimated from Fig. 3 by using the proposed method. The light source distribution was measured at 225 points on the geodesic dome. Fig. 3 shows virtual objects (i.e. person, pod) and their shadows rendered by using the proposed method. Fig. 5 and Fig. 6 show another example of image synthesis under a different light source distribution. As shown in these figures the shade and shadow of the virtual objects were generated properly, although the light source distributions of the scene are very complex.

4.2 Accuracy Evaluation

We next evaluate the accuracy of the proposed method comparing with other methods. In this evaluation, we use synthetic images for quantitative evaluation.

Fig. 7 shows the synthetic 3D scene of the reference object used in our experiments. The reference sphere and the plane are Lambertian and their reflectance is 0.6 in R, G and B. The light source distribution used for generating the scene is shown in Fig. 8. There are 75 light sources on the geodesic dome. Fig. 9 shows an image of the scene observed by a virtual camera. The number of the observed patches of the reference object was 83, and that of the ground plane was 374. The Gaussian image noises with the standard deviation of 1.0 were added to the image intensity of each point, while the range of image intensity is from 0 to 255. We estimated the light source distribution of this scene from the image shown in Fig. 9 by using the proposed method. We also estimated the light source distribution by using the existing method proposed by Sato et al. (Sato et al., 2003) for comparison.

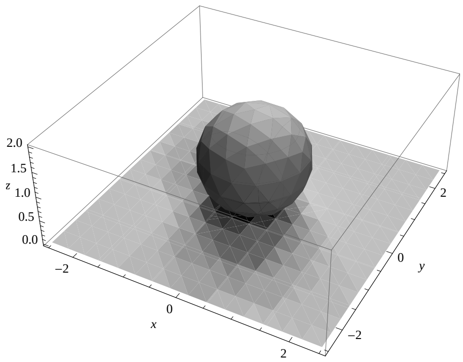


Figure 7: 3D scene used in our simulation experiments.

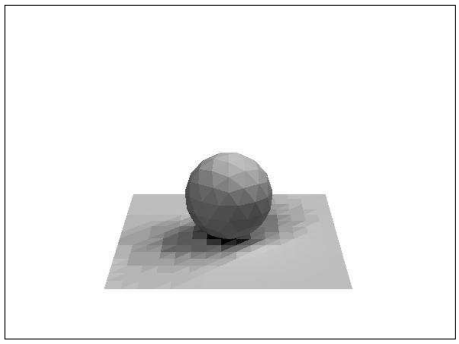


Figure 9: The synthetic image generated by projecting the 3D scene shown in Fig. 7.

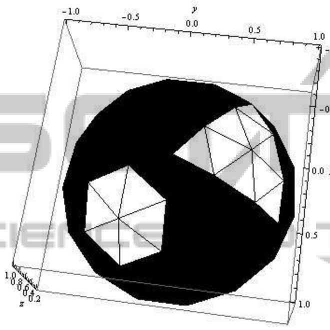
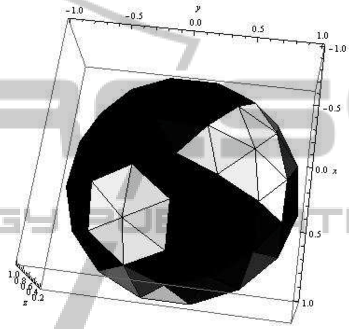
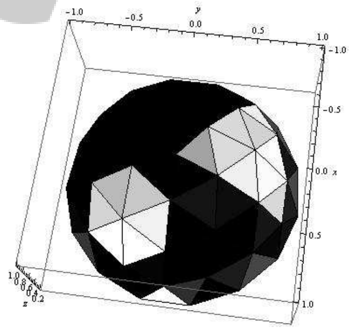


Figure 8: The light source distribution used in the simulation experiments.



(a) Proposed method.



(b) Existing method.

Figure 10: The light source distribution estimated by using the proposed method (a) and the existing method (b). The ground truth is shown in Fig. 8.

Fig. 10 (a) shows the light source distribution estimated by using the proposed method, and Fig. 10 (b) shows that of the existing method. From Fig. 10 and Fig. 8, we find that the proposed method provides us better estimation of light source distribution comparing with the existing method. The RMS error of the estimated light source distribution divided by the maximum ground truth light source magnitude was 0.14 in the proposed method and was 0.40 in the existing method.

We next evaluated the accuracy of estimated light source distribution changing the reflectance of the reference object. Fig. 11 shows the RMS error of light source distributions estimated from the proposed method and the existing method. The horizontal axis shows the reflectance of the reference object, and the vertical axis shows RMS error divided by the maximum light source magnitude. As shown in this figure, if the reflectance of the reference object is small, the error of estimation is small in both methods. However, if the reflectance of the reference object is large, the accuracy of the existing method (Sato et al., 2003) degrades drastically, while the proposed method still provides us good accuracy. This is because the existing method suffers from inter reflections, when the reflectance of the reference object is large.

5 CONCLUSIONS

In this paper, we proposed an efficient method for synthesizing virtual objects into real scene images by using a diffuse reference sphere. The proposed method uses the direct relationship between the intensity of the reference sphere and the intensity of virtual objects for generating the shading information of objects. Also, the proposed method generate shadows of

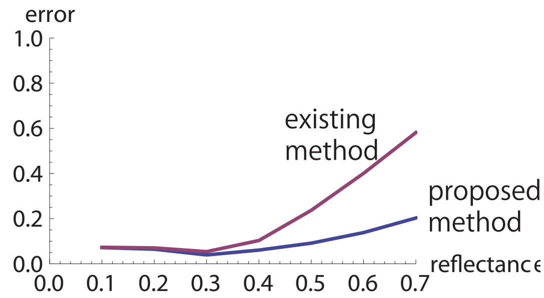


Figure 11: The accuracy of estimated light source distributions. If the reflectance of the reference object is large, the accuracy of the existing method (Sato et al., 2003) degrades drastically, while that of the proposed method does not.

the virtual objects by using the shadows of the reference sphere. As a result, the virtual objects and their shadows can be rendered quite efficiently from a single diffuse sphere put in the scene.

We tested the proposed method by using real scene images. We also evaluated the accuracy of the proposed method by using a synthetic 3D scene comparing with the existing method. The proposed method can synthesize virtual objects quite efficiently with reasonable accuracy, and thus it is very useful.

The future work includes the extension of the proposed method to the video streams, so that moving virtual objects and their shadows are rendered properly into real scene videos.

REFERENCES

Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. SIGGRAPH 98*, pages 189–198.

Kanbara, M. and Yokoya, N. (2004). Shadow representation for augmented reality by estimating light environment and shape of real scene. In *Proc. MIRU*, volume 2, pages 247–252.

Marschner, S. and Greenberg, D. (1997). Inverse lighting for photography. In *Proc. Color Imaging Conference*, pages 262–265.

Sato, I., Sato, Y., and Ikeuchi, K. (2003). Illumination from shadows. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(3):290–300.

Takai, T., Nishimura, K., Maki, A., and Matsuyama, T. (2004). Difference sphere: an approach to near light source estimation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 98–105.

Weber, M. and Cipolla, R. (2001). A practical method for estimation of point light sources. In *Proc. British Machine Vision Conference*, pages 471–480.

Wei, J. (2003). Robust recovery of multiple light source based on local light source constant constraint. *Pattern Recognition Letters*, 24(1–3):159–172.

Zhang, Y. and Yang, Y. (2001). Multiple illuminant direction detection with application to image synthesis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(8):915–920.