

APPEARANCE-BASED REPRESENTATION AND RENDERING OF CAST SHADOWS

Dong-O Kim¹, Sang Wook Lee², and Rae-Hong Park¹

¹Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea

hyssop@sogang.ac.kr, rhpark@sogang.ac.kr

²Department of Media Technology, Graduate School of Media, Sogang University, Seoul, Korea

slee@sogang.ac.kr

ABSTRACT

Appearance-based approaches have the advantage of representing the variability of an object's appearance due to illumination change without explicit shape modeling. While a substantial amount of research has been performed on the representation of object's diffuse and specular shading, little attention has been paid to cast shadows which are critical for attaining realism when multiple objects are present in a scene. We present an appearance-based method for representing shadows and rendering them in a 3D environment without explicit geometric modeling of shadow-casting object. A cubemap-like illumination array is constructed, and an object's shadow under each cubemap light pixel is sampled on a plane. Only the geometry of the shadow-sampling plane is known relative to the cubemap lighting. The sampled images of both the object and its cast shadows are represented using the Haar wavelet, and rendered on an arbitrary 3D background of known geometry. Experiments are carried out to demonstrate the effectiveness of the presented method.

KEYWORDS

Appearance, Diffuse, Specular, Shading, Rendering, Cubemap, Haar wavelet, Illumination

1. INTRODUCTION

The main advantage of the appearance-based representation is that without requiring any information about object shape and BRDF, it accounts for the variability of object/image appearance due to illumination or viewing change using a low-dimensional linear subspace. Much research has been carried out for representing object's diffuse and specular appearance for image synthesis and recognition. However, there have been few appearance-based approaches to shadow synthesis and rendering and this motivated us to develop a method to represent a shadow using basis images for re-rendering and re-lighting. Shadows are critical for achieving realism when multiple objects/backgrounds are rendered. Without shadows, it is impossible to realistically render the image/object appearance into synthetic scenes.

Initial appearance-based approaches have represented a convex Lambertian object without attached and cast shadows using a 3-D linear subspace determined with three images taken under linearly independent lighting conditions [1,2]. Higher-dimensional linear subspaces have been used for object recognition to deal with non-Lambertian surfaces, shadows and extreme illumination conditions [3-6]. A set of basis images that span a linear subspace are constructed by applying PCA (principal component analysis) to a large number of images captured under different light directions.

Instead of the PCA-driven basis images, analytically specified harmonic images have recently been investigated for spanning a linear subspace. A harmonic image is an image under lighting distribution specified by spherical harmonics. Basri and Jacobs used a 9-D linear subspace defined with harmonic images for face recognition under varying illumination [8]. Lee *et al.* has shown that input images taken under 9 lights can approximate a 9-D subspace spanned by harmonic images for an object or a class of objects, and presented a method for determining 9 light source directions [10]. Harmonic images have been recently used for efficient object rendering under complex illumination by Ramamoorthi and Hanrahan and by Sloan *et al.* [9, 12]. All of the above harmonic image-based approaches require object geometry and BRDF to compute harmonic images. Despite all the progress in object modeling, obtaining an accurate model for shape and BRDF from a complex real object is still a highly challenging problem. Sato *et al.* developed a method for analytically deriving basis images of a convex object from its images taken under a point light source without requiring object model for shape and BRDF [7]. They also developed a method to determine a set of lighting directions based on the object's BRDF and the angular sampling theorem on spherical harmonics.

When an object's images are captured under various illumination conditions, the images of the object's cast shadows can also be captured simultaneously with a plane placed on the opposite side of the light. A set of those shadow images can be represented using appropriate basis functions and their linear combination can be re-projected for rendering onto a new synthetic or other 3-D scenes. We develop a method for representing shadow appearance for cubemap illumination geometry and physically construct a cubemap-like lighting apparatus for shadow image sampling/capture. The light-sampling distance is determined by the desired cubemap resolution. In this specific illumination array, the size of the light pixel of the cubemap increases as the sampling distance increases (or as the resolution decreases). Therefore, the size of the light pixel limits the bandwidth of the cast shadows and prevents aliasing due to insufficient sampling. We use the Haar wavelet for basis functions to represent appearance of both object and shadow. There have been approaches to efficient object and shadow rendering using linear combination of basis images. Nimeroff *et al.* [11] used pre-computed basis images for rendering a scene including shadows and developed basis functions for natural skylight. Sloan *et al.* [12] used spherical harmonics for pre-computed radiance transport (PRT), Ng *et al.* developed PRT techniques based on wavelet and cubemap [13, 14], and Zhou *et al.* also used wavelet for precomputed object and shadow fields (POF and PSF) under moving objects and illuminations [15]. All of these approaches require object geometry and BRDF for pre-computing PRT or POF/PSF. On the other hand, the goal of the work presented in this paper is to estimate shadow projection in a cubemap using only sampled shadow images expanded on a set of wavelet basis images. No model for object shape and BRDF is necessary and all that is required is the calibration of the shadow sampling plane with respect to the light array.

The rest of this paper is organized as follows. Section 2 presents the proposed approach to the sampling, representation and rendering of shadows into a 3-D scene. Sections 3 shows experimental results and Section 4 concludes this paper.

2. OBJECT AND SHADOW APPEARANCE

2.1. Light Cube and Image Capture

In computer graphics, cubemaps are used frequently for representing area lighting such as environment map. A wavelet basis is suitable for approximating light distribution on a cubemap, containing area lights that vary from the size of cubemap pixel to the size of entire sky. A light-array cube is constructed as shown in Figure 1 and an image of object and shadow is captured under each cube pixel light. For our experimental study, the cube has a 32×32 light array only on the top side and a shadow sampling plane was place near the bottom side. Calibration is

performed to establish the relationship between the light array, shadow sampling plane and camera. Each light pixel element is constructed in a rectangular compartment with a white power LED on the top and diffuse on the bottom as illustrated as shown in Figure 1. Seen from the bottom, the light element looks uniformly bright square due to the diffuser and the gap between the light squares is less than 2 mm.

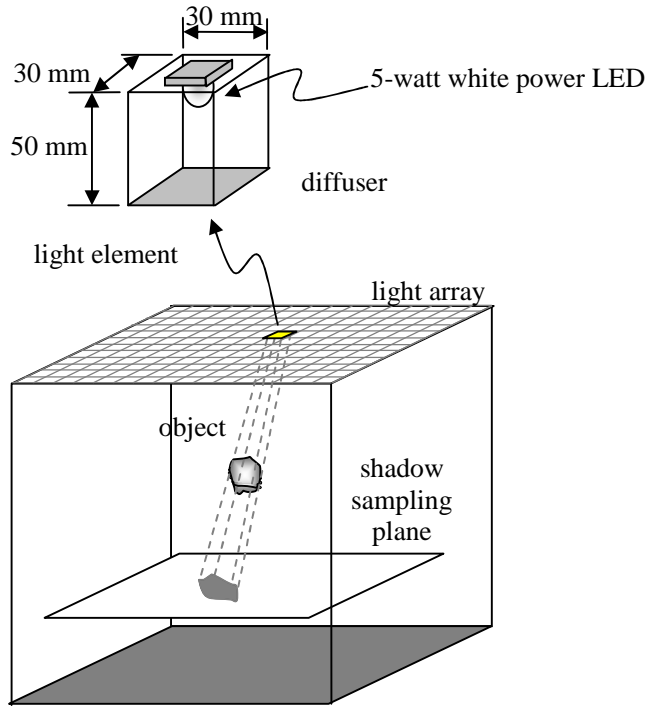


Figure 1. Cubemap light array, light element and shadow sampling plane

2.2. Shadow Sampling in Light Cube

Let us consider the shadow sampling-bandwidth relationship in the light cube. Figure 2 illustrates shadow casting geometry and shadow intensity for a point-like small object. The relationship between the distance d_s between shadow samples and the distance d_l between light sources is given as:

$$d_s = d_l \frac{h_s}{h_l}, \quad (1)$$

where h_s and h_l denote the distance between the object and the shadow sampling plane and that between the light array plane and the object, respectively.

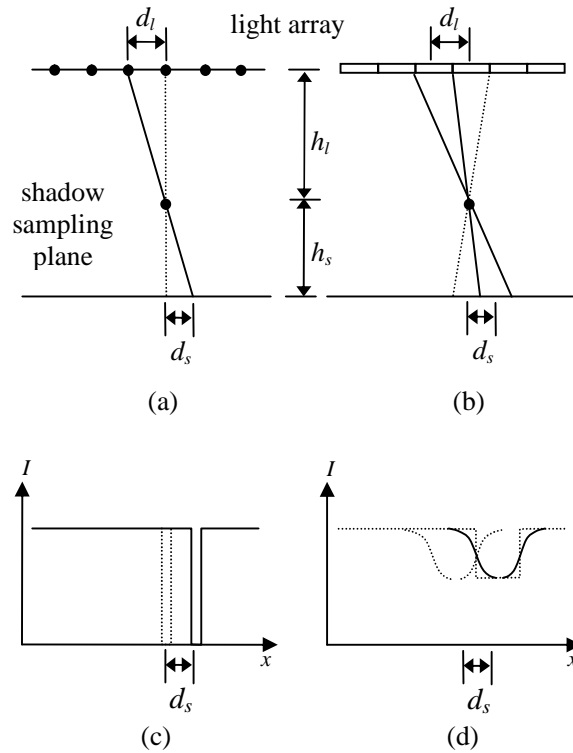


Figure 2. Shadows for point object: (a) geometry for point light source, (b) geometry for area source, (c) intensity for point light source, and (d) intensity for area light source.

Point light sources cast hard shadow with wide bandwidth as shown in Figure 2 (a) and (c). This type of shadow cannot be rendered effectively using a linear combination of sampled images without aliasing when they are sampled under a finite number of point light sources. Only in the restricted case that the shadow sampling plane is right beneath the object, it may be possible to have sufficient sampling rate due to small d_s .

An area light source spreads shadow shape and reduces the shadow bandwidth as depicted in Figure 2 (b) and (d). Although the spread is rectangular in an ideal case, it normally appears as a smoother function probably due to light diffraction at the source and scattering at the object. The area sources closely packed in the light array simply blurs the shadows and thus serves as anti-aliasing filter. Figure 3 shows the shadow casting geometry where the object and the shadow sampling plane are both tilted with respect to the light array. In this case, the shadow spread is not

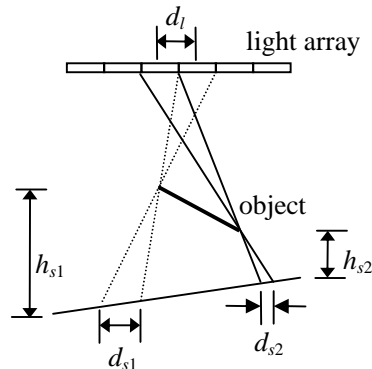


Figure 3. Shadows for tilted object and shadow sampling plane.

constant but depends on the ratio of h_s and h_l , and thus provides varying degree of anti-aliasing blurring. This analysis should be valid for attached shadows within an object and as well as sampled cast shadows.

2.3. Image Representation and Shadow Mapping

The intensity I observed at point \mathbf{x} on an object surface can be written as

$$I(\mathbf{x}) = \int_0^{2\pi} \int_0^{\pi/2} L(\mathbf{x}, \theta_i, \phi_i) V(\mathbf{x}, \theta_i, \phi_i) \rho(\mathbf{x}, \theta_i, \phi_i; \theta_o, \phi_o) \cos \theta_i \sin \theta_i d\theta_i d\phi_i \quad (2)$$

where L denotes the lighting and V is the visibility function which indicates whether the light reach at a point on the surface, ρ is the BRDF function, and (θ_i, ϕ_i) and (θ_o, ϕ_o) denote incoming and outgoing directions of light, respectively. For a fixed camera, the term $V(\mathbf{x}, \theta_i, \phi_i) \rho(\mathbf{x}, \theta_i, \phi_i; \theta_o, \phi_o) \cos \theta_i$ can be simplified to the reflectance kernel $R_V(\mathbf{x}, \theta_i, \phi_i)$ including the visibility information. Therefore, the intensity I of a surface point can be written as:

$$I(\mathbf{x}) = \int_0^{2\pi} \int_0^{\pi/2} L(\mathbf{x}, \theta_i, \phi_i) R_V(\mathbf{x}, \theta_i, \phi_i) \sin \theta_i d\theta_i d\phi_i. \quad (3)$$

Equation (3) can represent the intensity on both object surface and shadow sampling plane. Instead of pre-computing $R_V(\mathbf{x}, \theta_i, \phi_i)$ for each surface point with synthetic objects as in [14], we use the shadow image on the shadow capturing plane as a light mask for the scene underneath the object. Shadow image for novel illumination is synthesized by linearly combining basis images of the sampled shadows. We employed the Haar wavelet for the representation and compression of a large number of sampled shadow images. Although there have been few directly relevant studies on the choice of basis for this type of shadow images, the wavelet showed better results than spherical harmonics for compression of a cubemap image [14] and for the estimation of illumination from shadows [16].

Shadow mapping from the shadow sampling plane to a synthetic scene is illustrated in Figure 4. For a given vertex \mathbf{v} in the synthetic scene and light source in the direction \mathbf{l} , the relationship between \mathbf{v} and its projected point \mathbf{p} in the shadow sampling plane $P: \mathbf{n} \cdot \mathbf{x} + d = 0$ is given as $\mathbf{M}\mathbf{v} = \mathbf{p}$ where the projection matrix \mathbf{M} is defined as [17]:

$$\mathbf{M} = \begin{pmatrix} \mathbf{n} \cdot \mathbf{l} + d - l_x n_x & -l_x n_y & -l_x n_z & -l_x d \\ -l_y n_x & \mathbf{n} \cdot \mathbf{l} + d - l_y n_y & -l_y n_z & -l_y d \\ -l_z n_x & -l_z n_y & \mathbf{n} \cdot \mathbf{l} + d - l_z n_z & -l_z d \\ -n_x & -n_y & -n_z & \mathbf{n} \cdot \mathbf{l} \end{pmatrix} \quad (4)$$

where \mathbf{l} denotes the direction vector from light source to vertex and \mathbf{n} is the surface normal. This mapping should be performed for the constant area light using Equation (2). If the BRDF ρ is approximated as constant ($\bar{\rho}$) over the solid angle to which the area light source is subtended, Equation (2) can be rewritten as:

$$I(\mathbf{x}) = L\bar{\rho} \int_0^{2\pi} \int_0^{\pi/2} V(\mathbf{x}, \theta_i, \phi_i) \cos \theta_i \sin \theta_i d\theta_i d\phi_i. \quad (5)$$

This means that the mean of the shadow image intensities in the area subtending to the solid angle at \mathbf{v} can be used for irradiance weighting in rendering the vertex \mathbf{v} . In other words, a linearly combined shadow image for an illumination distribution is a visibility map for lighting elements in the cube.

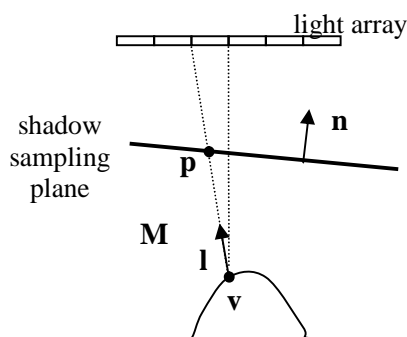


Figure 4. Shadow mapping from the shadow sampling plane.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

Using the cube illumination system, several objects and their cast shadows are captured. Total of $32 \times 32 = 1024$ sample images are acquired and Figures 5(c) and 6(c) show one of the 1024 sample images for two objects. They show both an object and its cast shadow.

Figure 5 compares a real scene and a rendered scene. Figures 5(a) and 5(b) show environment maps having one-pixel and nine-pixel lights, respectively. Figures 5(c) and 5(d) are a sampled view and a 9-light rendered view of a toy airplane illuminated from environment maps shown in Figures 5(a) and 5(b), respectively. Figures 5(e) and 5(f) show the zoomed views of the object and Figures 5(g) and 5(h) show the zoomed views of the shadows. Figure 5 shows that specularities, cast shadows and attached shadows appear more softly in the rendered image with 9 adjacent lights than in the one-light sampled image. As shown in the wing region of the airplane in figure 5(e) and figure 5(f), the self shadows in the sampled image look sharper than those in the rendered image with 9 light sources. Figure 6 shows rendered results of a toy tree under novel illumination. It can be seen in Figure 6 that the specularity in the base region of the toy tree is substantially more dispersed in the 9-light rendered image than in the one-light sampled image. The shadow is much softer in the 9-light rendered image.

Figure 7 show the rendered images with a synthetic scene under a novel illumination. Figure 7(a) (7(c)) show the cast shadows on the synthetic scene under the illumination given by the environment map shown in Figure 7(b) (7(d)). Figure 7(c) shows that the soft shadow is realistically rendered.

Only one light array is used on the top side of the light cube in our implementation and experiments. The construction of a more complete light cube with up to six light arrays and multiple shadow sampling planes is a subject of our future work. In this paper, we show the results of shadow rendering only under simple sets of adjacent lighting elements. However, any complex environment map in the form of cubemap can be used to generate novel illumination. This paper is focused only on shadow sampling and rendering. In the future, we intend to compare several bases such as wavelet and spherical harmonics for their effectiveness in representing specular and diffuse appearance as well as cast/attached shadows.

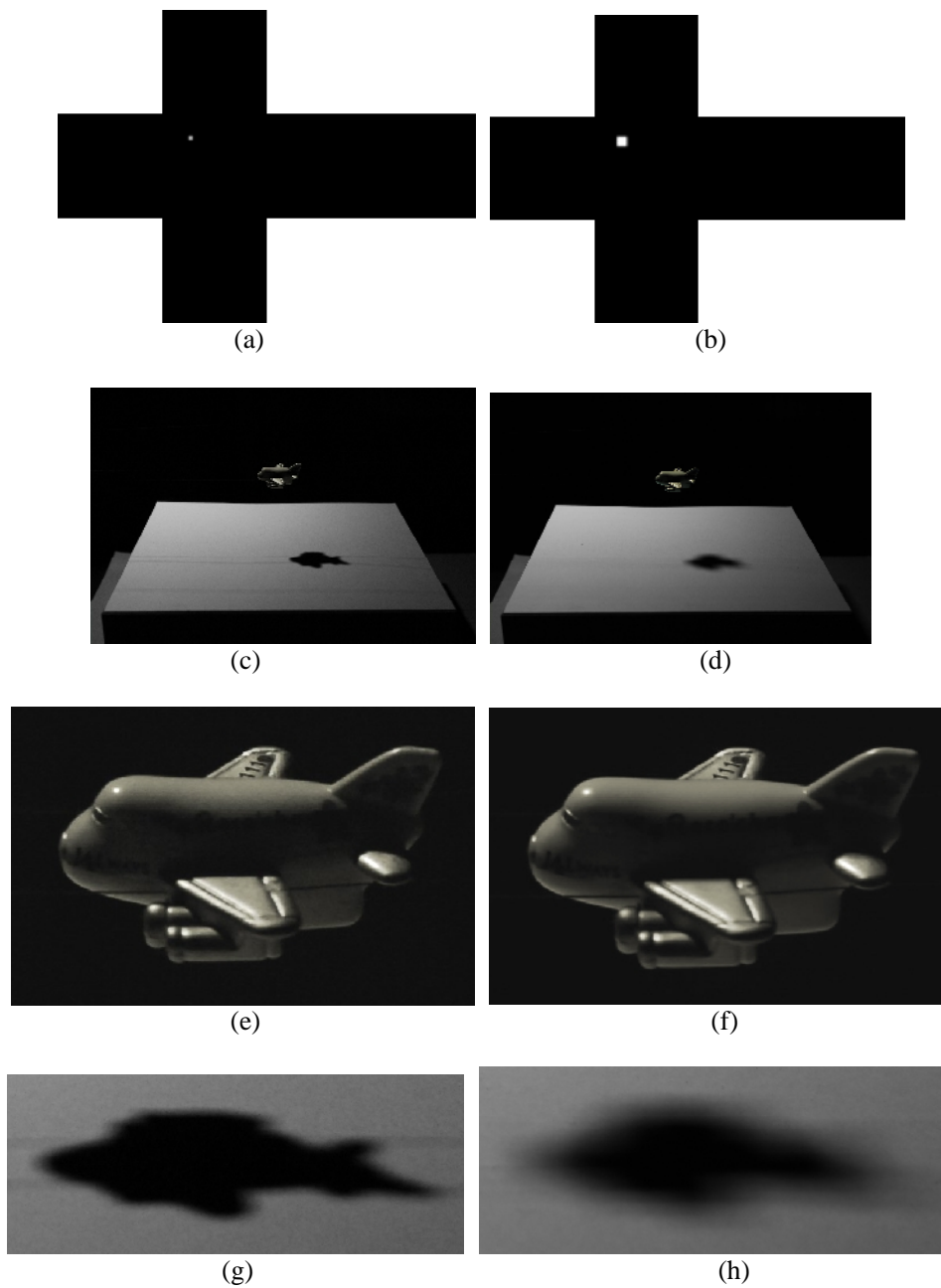


Figure 5. Rendering in the shadow sampling plane under a novel illumination; (a) cubemap (single light pixels), (b) cubemap (9 light pixels), (c) rendering onto the shadow sampling plane, (d) rendering onto the shadow sampling plane, (e) object appearance, (f) object appearance, (g) real shadow appearance, (h) shadow appearance.

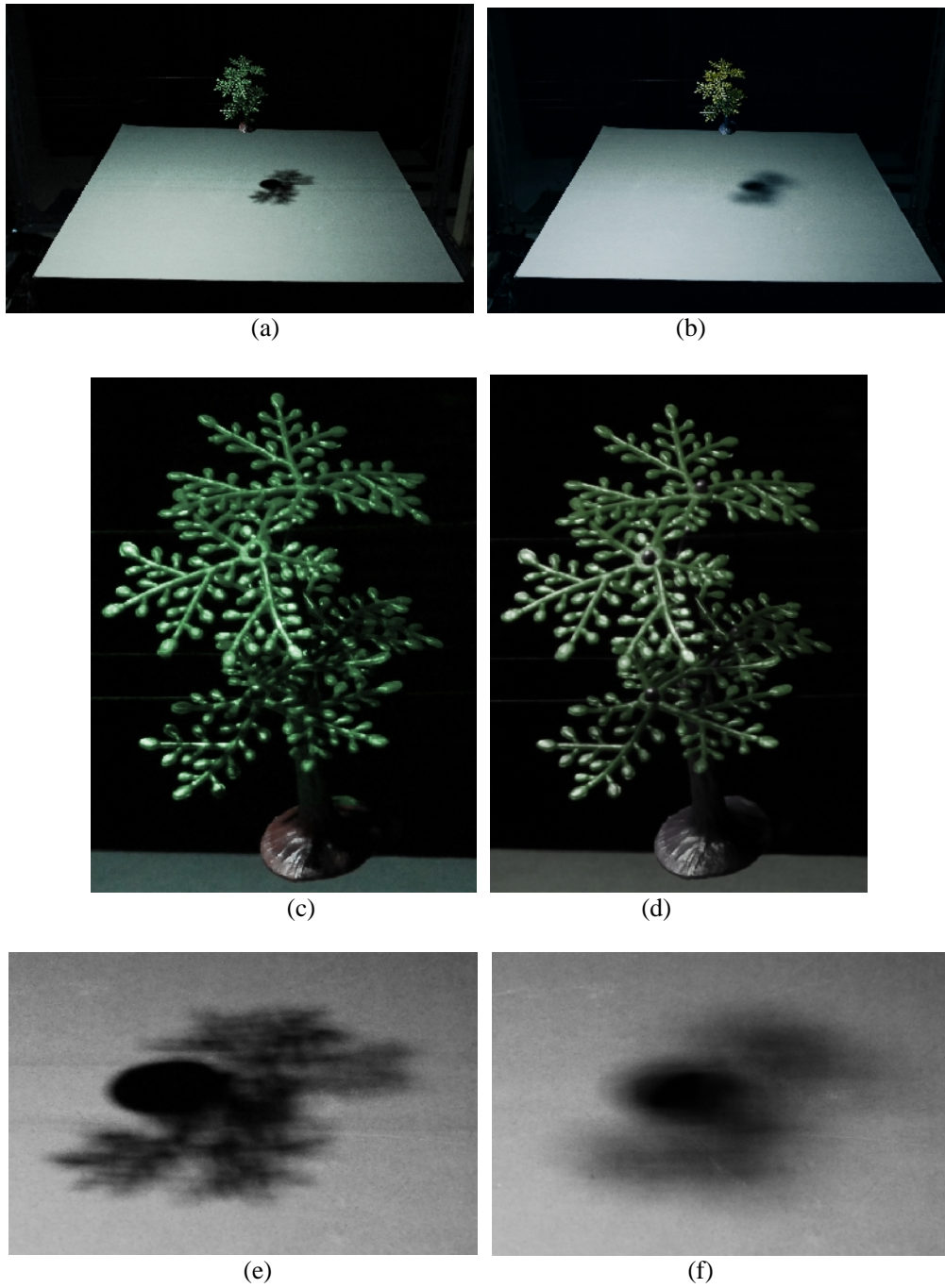


Figure 6. Rendering in the shadow sampling plane under a novel illumination with 1 pixel light and 9 pixels light, respectively; (a) rendering onto the shadow sampling plane, (b) rendering onto the shadow sampling plane, (c) object appearance, (d) object appearance, (e) real shadow appearance, (f) shadow appearance.

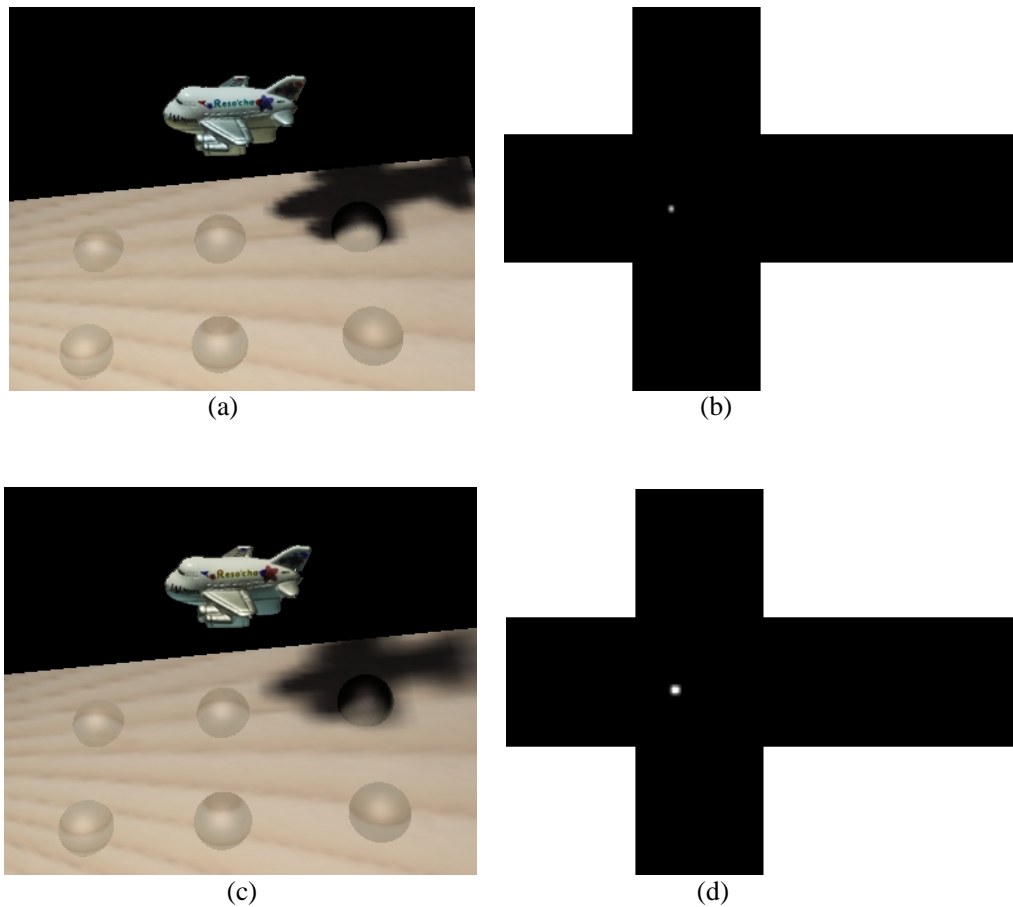


Figure 7. Shadow rendering; (a) shadow rendering result, (b) environment map with 1 pixel light, (c) shadow rendering result, (d). environment map with 4 pixels light.

4. CONCLUSIONS

We present an appearance-based method for representing shadows and rendering them into a synthetic scene without explicit geometric modeling of shadow-casting object. The method utilizes the special geometry of cubemap-like illumination with closely packed area light elements and the resolution of synthesized shadows degrades gracefully without aliasing as the resolution of the cubemap lighting decreases. We constructed a cube light with a light array on its top and conducted experiments to show the efficacy of the presented approach.

ACKNOWLEDGEMENTS

The second author's work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MOE) No. NRF-2012R1A1A2009461.

REFERENCES

- [1] A. Shashua, (1997) "On photometric issues in 3D visual recognition from a single image," *Int. J. Computer Vision*, Vol. 21, pp. 99–122.
- [2] H. Murase and S. Nayar, (1995) "Visual learning and recognition of 3-D objects from appearance," *Int. J. Computer Vision*, Vol. 14, No. 1, pp. 5–24.
- [3] P. Hallinan, (1994) "A low-dimensional representation of human faces for arbitrary lighting conditions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 995–999.
- [4] A. Yuille, D. Snow, R. Epstein, and P. Belhumeur, (1999) "Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability," *Int. J. Computer Vision*, Vol. 35, No. 3, pp. 203–222.
- [5] A. Georghiades, D. Kriegman, and P. Belhumeur, (1998) "Illumination cones for recognition under variable lighting: Faces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 52–59, 1998.
- [6] A. Georghiades, D. Kriegman, and P. Belhumeur, (2001) "From few to many: Generative models for recognition under variable pose and illumination," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 643–660.
- [7] I. Sato, T. Okabe, Y. Sato, and K. Ikeuchi, (2003) "Appearance sampling for obtaining a set of basis images for variable illumination," *Proc. IEEE Int. Conf. Computer Vision 2003*, pp. 800–807, Nice, France.
- [8] R. Basri and D. Jacobs, (2001) "Lambertian Reflectance and Linear Subspaces," in *Proc. IEEE Intl. Conf. Computer Vision*, pp. 383–389.
- [9] R. Ramamoorthi and P. Hanrahan, (2001) "A signal-processing framework for inverse rendering," in *Proc. SIGGRAPH'01*, pp. 117–128.
- [10] K. C. Lee, J. Ho, and D. Kriegman, (2001) "Nine points of light: Acquiring subspaces for face recognition under variable lighting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition 01*, pp. 519–526.
- [11] J. S. Nimeroff, E. Simoncelli, and J. Dorsey, (1994) "Efficient rerendering of naturally illuminated environments," in *5th Eurographics Workshop on Rendering*, pp. 359–373, Darmstadt, Germany.
- [12] P. Sloan, J. Kautz, and J. Snyder, (2002) "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments," *ACM Trans. Graphics*, Vol. 21, No. 3, pp. 527–536.
- [13] R. Ng, R. Ramamoorthi, and P. Hanrahan, (2003) "All-frequency shadows using non-linear wavelet lighting approximation," *ACM Trans. Graphics*, Vol. 22, No. 3, pp. 376–381.
- [14] R. Ng, R. Ramamoorthi, and P. Hanrahan, (2004) "Triple product integrals for all-frequency relighting," *ACM Trans. Graphics*, Vol. 23, No. 3, pp. 477–487.
- [15] K. Zhou, Y. Hu, S. Lin, B. Guo, and H. Shum, (2005) "Precomputed shadow fields for dynamic scenes," *ACM Trans. Graphics*, Vol. 24, No. 3, pp. 1196–1201.
- [16] T. Okabe, I. Sato and Y. Sato, (2004) "Spherical harmonics vs. Haar wavelets: basis for recovering illumination from cast shadows," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition 2004*, pp. I-50-57.
- [17] E. Haines and T. Moeller, (2002) *Real-Time Rendering*, Natick, MA, USA, A K Peters.

Authors

Dong-O Kim received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea, in 1999 and 2001, respectively. Currently, he is working toward the Ph.D. degree in electronic engineering at Sogang University. His current research interests are image quality assessment and physics-based computer vision for computer graphics.

Rae-Hong Park was born in Seoul, Korea, in 1954. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center (DTV image/video enhancement), Samsung Electronics Co., Ltd., Suwon, Korea. In 2012, he spent a sabbatical year in Digital Imaging Business (R&D Team) and Visual Display Business (R&D Office), Samsung Electronics Co., Ltd., Suwon, Korea. His current research interests are video communication, computer vision, and pattern recognition. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996. Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University. He is a co-recipient of the Best Student Paper Award of the IEEE Int. Symp. Multimedia (ISM 2006) and IEEE Int. Symp. Consumer Electronics (ISCE 2011).

Sang Wook Lee received the BS degree in electronic engineering from Seoul National University, Seoul, in 1981, the MS degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Seoul, in 1983, and the PhD degree in electrical engineering from the University of Pennsylvania in 1991. He is currently a professor of media technology at Sogang University, Seoul. He was an assistant professor in computer science and engineering at the University of Michigan (1994-2000), a postdoctoral research fellow and a research associate in computer and information science at the University of Pennsylvania (1991-1994), a researcher at the Korea Advanced Institute of Science and Technology (1985-1986), a research associate at Columbia University (1984-1985), and a researcher at LG Telecommunication Research Institute (1983-1984). His major field of interest is computer vision, with an emphasis on BRDF estimation, optimization for computer vision, physics-based vision, color vision, range sensing, range data registration, and media art.