

International Journal of Image and Graphics
© World Scientific Publishing Company

LIGHTING AND SHADOW INTERPOLATION USING INTRINSIC LUMIGRAPHS

YASUYUKI MATSUSHITA

*Microsoft Research Asia, 3/F, Beijing Sigma Center No.49, Zhichun Road, Hai Dian District,
Beijing China 100080, yasumat@microsoft.com*

SING BING KANG

*Microsoft Research, One Microsoft Way, Redmond, Washington 98052-6399, U.S.A.,
sbkang@microsoft.com*

STEPHEN LIN

HEUNG-YEUNG SHUM

XIN TONG

*Microsoft Research Asia, 3/F, Beijing Sigma Center No.49, Zhichun Road, Hai Dian District,
Beijing China 100080, {stevelin ; hshum ; xtong}@microsoft.com*

Received (05 July 2003)

Revised (09 July 2003)

Densely-sampled image representations such as the light field or Lumigraph have been effective in enabling photorealistic image synthesis. Unfortunately, lighting interpolation with such representations has not been shown to be possible without the use of accurate 3D geometry and surface reflectance properties. In this paper, we propose an approach to image-based lighting interpolation that is based on estimates of geometry and shading from relatively few images. We decompose light fields captured at different lighting conditions into *intrinsic images* (reflectance and illumination images), and estimate view-dependent scene geometries using multi-view stereo. We call the resulting representation an *Intrinsic Lumigraph*. In the same way that the Lumigraph uses geometry to permit more accurate view interpolation, the Intrinsic Lumigraph uses both geometry and intrinsic images to allow high-quality interpolation at different views and lighting conditions. The joint use of geometry and intrinsic images is effective in computing *shadow masks* for shadow prediction at new lighting conditions. We illustrate our approach with images of real scenes.

Keywords: Lumigraph; Shadow Warping; Multi-view Stereo; Intrinsic Images.

1. Introduction

In recent years, much progress has been made in image-based rendering. One class of such methods relies on densely sampled images, such as the light field¹⁰ and the Lumigraph⁶. Another class requires an accurate physically-based rendering algorithm and sufficiently detailed geometric and material properties of the scene and

light sources^{3,18,23}. Others require all of the above information²².

Methods that rely on densely sampled images have the advantage that they do not require accurate geometry, which in practice requires a high-quality and range finder. However, this advantage is achieved at the expense of a large database. In addition, it is not possible to relight the scene using these current image-based representations, with the exception of Wong *et al.*²¹, who use dense sampling of camera locations and illumination conditions (and hence may not be practical for real scenes). Methods that permit scene relighting typically need a detailed and accurate 3D geometric model in order to extract surface properties in the form of a Bidirectional Reflectance Distribution Function (BRDF). Usually, such models can only be acquired using expensive range finders, and even then, the shapes used as examples tend to be simple. Nimeroff *et al.* proposed another approach¹⁴ which uses steerable linear basis functions to accomplish re-rendering of a scene under a directional illuminant at an arbitrary orientation. One drawback of the method is that it requires a huge basis set to handle narrow illuminants.

We are motivated by the need for a more *practical* approach to interpolate lighting appearance of a scene that has sparsely sampled lighting conditions. Equivalent results can be obtained using densely sampled lighting conditions with high compression^{26,27}, however, such methods rely on precise geometry, which is often difficult to obtain from light fields.

Our proposed method requires only images (light fields) as input, and assume that the camera positions associated with these images are known. The light fields are captured under a relatively small set of different lighting conditions. From these light fields, we can extract two separate datasets: view-dependent geometries using stereo, and *intrinsic images* using the method proposed by Weiss²⁰. These datasets are used to predict shadow movement with changing light conditions.

2. Prior Work

Much of the work on realistic rendering relies on reflectance modeling and known 3D geometry. A representative approach in this area is presented by Sato *et al.*¹⁸, which merges multiple range datasets to yield a single 3D model. This shape is subsequently used for diffuse-specular separation and reflectance estimation. They showed results for single objects with no shadows. Wood *et al.*²² also use color images and laser range scans. Their range datasets are merged manually to produce a global 3D model. Subsequently, a function that associates a color to every ray originating from a surface is constructed and compressed. Magda *et al.*²⁴ developed a method to reconstruct the geometry of objects with arbitrary BRDFs from a scene's incident light field. Using this method, Koudelka *et al.*²⁵ developed an image-based rendering technique in which novel images of an object can be synthesized under arbitrarily specified illumination conditions. The approach correctly handles self-shadowing and interreflections using the recovered geometry. In contrast, our method does not require accurate scene geometry but uses rough

geometry to compute shadow shapes under intermediate lighting conditions.

Yu *et al.*²³ compute surface BRDFs based on Ward's anisotropic BRDF model¹⁹ from multiple images and a 3D model. They assume that at least one specularity is observed per surface. On the other hand, Boivin and Gagalowicz² propose a technique for recovery of a BRDF approximation from a single image based on iterative analysis by synthesis (or *inverse rendering*¹³). The emittances of the light sources are assumed known. This is an extension of Fournier *et al.*'s work⁴, which assumes perfectly diffuse surfaces, and Loscos *et al.*¹¹, who additionally considered textured surfaces. Marschner and Greenberg¹³ directly estimate the BRDF model of Lafortune *et al.*⁸ from an image and a surface model. Malzbender *et al.*¹² proposed a space and time efficient method for encoding an object's diffuse lighting response as the light position varies with respect to the surface, using a set of coefficients.

Debevec³ uses global illumination for augmented reality applications. He uses local geometry and manually computes reflectance parameters, with which objects can be inserted with realistic-looking interreflections. In a series of works geared for augmented reality, Sato *et al.* estimate the illumination distribution from shadows¹⁷, and subsequently from the brightness distributions in shadows¹⁶.

In our work, we rely on *intrinsic images* as a means for predicting shadows. Intrinsic images are a mid-level description of scenes first proposed by Barrow and Tenenbaum¹. A given image of a scene can be decomposed into a *reflectance image* and an *illumination image*. Various methods have been proposed to compute this decomposition, with piecewise constant reflectances using the Retinex algorithm⁹, with all-reflectance/all-illumination classification using wavelets⁵, and with maximum-likelihood (ML) estimation assuming time-constant reflectance and time-varying illumination²⁰.

3. Overview

An overview of our system is illustrated in Figure 1. The inputs to our method are a number of light fields, each captured under a different illumination condition. Once the light fields are acquired, view-dependent depth maps are then computed at the sampled camera positions using a multi-view stereo algorithm.

In addition, we decompose the light fields into intrinsic images in a manner similar to Weiss's method²⁰ (which handles a single image stream). For each camera and lighting position, the pair of intrinsic images consists of an illumination image that exhibits shading and shadowing effects, and a reflectance image that displays the unchanging reflectance property of the scene. The illumination images are used to identify pixels that contain cast shadows or attached shadows, which result when a surface area is occluded from the light source. These shadow masks are used in conjunction with shadows predicted by the scene geometry to estimate shadow appearance for novel lighting directions.

We call this new representation the *Intrinsic Lumigraph*, because it uses both geometry and intrinsic images for view reconstruction. When interpolating lighting

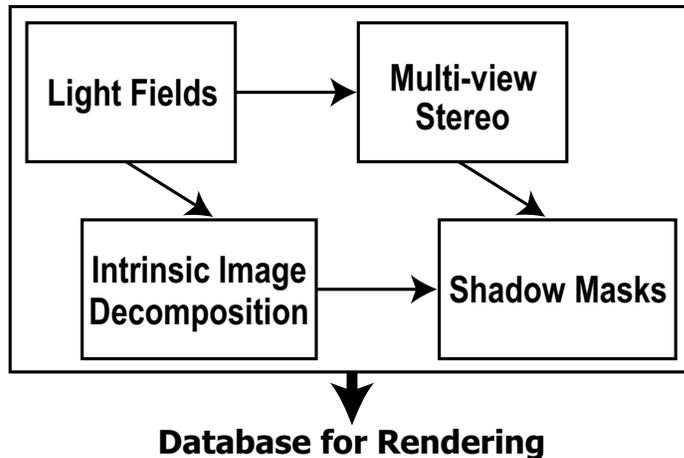


Fig. 1. System overview.

conditions of the scene, the diffuse reflection and shading can be well-approximated by interpolation of illumination images; however, shadows generally do not appear realistic when linearly combined. Our method for predicting shadow appearance enables us to synthesize images with much more accurate lighting interpolation.

4. Constructing the Intrinsic Lumigraph

In this section, we detail the process of constructing the Intrinsic Lumigraph. We first describe the capture of light fields under various illumination conditions, and then outline our algorithm for multi-view geometry. We next present our method for computing the intrinsic images, followed by the determination of shadow masks.

4.1. Capturing Light Fields under Different Lighting Conditions

We capture our light fields using the imaging setup shown in Figure 2. The camera is digitally controlled to capture images at predefined positions on a 2D grid. Each light field consists of an image sequence along a linear path that is captured under a fixed illumination condition, where the light source used is approximately a point light source. Since it is often the case that the camera’s optical axis is not perpendicular to the camera plane, or the camera is not in the precise position, we adopted the technique developed by Deng et al.²⁸ to accomplish the positional error correction and rectification of the light field.

4.2. Generating View-dependent Geometries

Using the captured light fields, we compute depth maps at each camera position using a multi-view stereo algorithm. The stereo algorithm is based on the work of

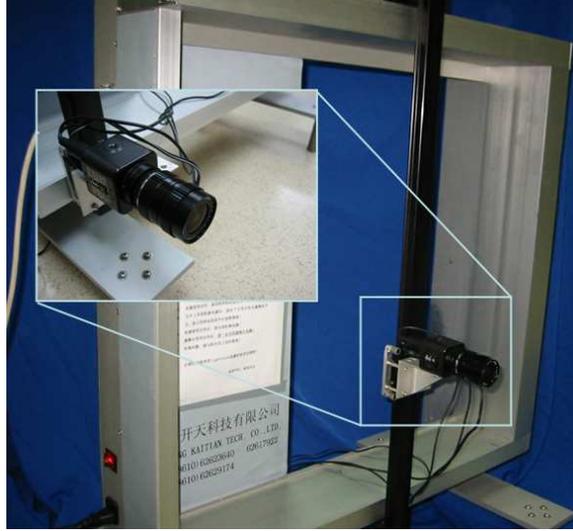


Fig. 2. Light field capture device.

Kang *et al.*⁷; it was chosen because it is very simple to implement and is very effective in handling occlusions. To improve the depth estimates, we linearly combine depth estimates from separate light fields taken under different lighting conditions. Depth estimates from areas that are more highly textured are favored.

From a sequence of N light fields, for each reference view we first obtain N estimated depth maps $D(n)$ and N confidence maps $C(n)$ using the multi-view stereo algorithm. The confidence map $C(n)$ is computed using the local matching error variance, which provides an indication of the reliability of the estimated depths. We use these confidence maps to refine the depth values through weighted averaging, i.e.,

$$D(x, y) = \frac{\sum_n D(n, x, y) \cdot C(n, x, y)}{\sum_n C(n, x, y)}. \quad (1)$$

An alternative method for refining the estimated depth values is to use the local Hessian of the local brightness distribution. The eigenvalues of the local Hessian are correlated with the degree of local texturedness; the higher the amount of texture, the more reliable the depth estimates tend to be in general. To be conservative, we use the minimum eigenvalues as a measure of depth reliability and as a means for weighting the depth estimates. The Hessian is obtained from the differential method of SSD(sum of squared differences):

$$E(u, v) = \Sigma_{k,l} (I_1(x + u + k, y + v + l) - I_o(x + k, y + l))^2. \quad (2)$$

In Equation (2), k and l correspond to the width and height of the subimage window

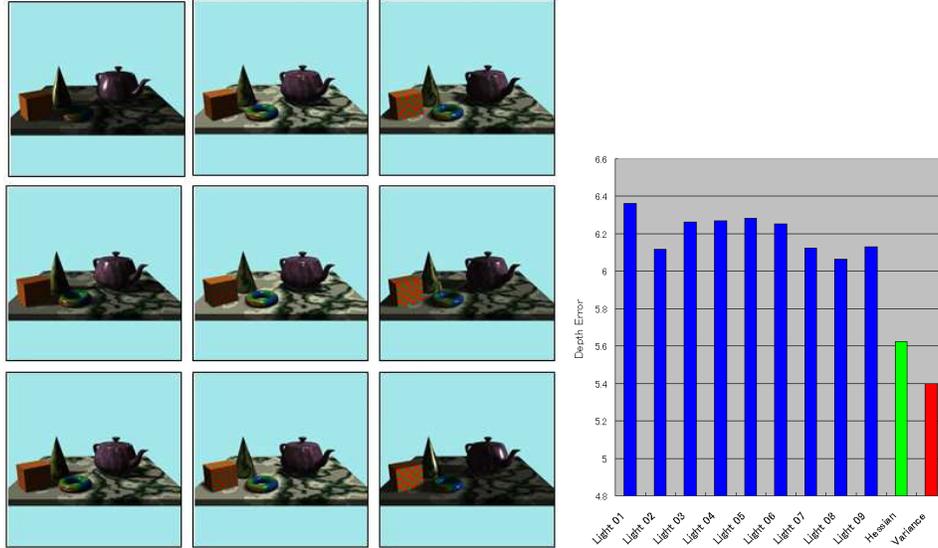


Fig. 3. Illumination sampling (left) and comparison of mean depth errors (right). The nine blue bars correspond to mean depth errors for each of the light fields, the green bar is the error when the Hessian (5×5 window) is used, and the red bar is the error obtained when the matching error variance (5×5 window) is used.

respectively, while u and v represent the movement of the window along x and y -axis. The differential method uses a local Taylor series expansion of the intensity function

$$\begin{aligned}
 & E(u + \Delta u, v + \Delta v) \\
 &= \sum_{k,l} (I_1(x + u + \Delta u + k, y + v + \Delta v + l) - I_0(x + k, y + l))^2 \\
 &\simeq \sum_{k,l} (I_1(x + u + k, y + v + l) + \nabla I_1 \cdot (\Delta u, \Delta v)^T - I_0(x + k, y + l))^2 \\
 &= \sum_{k,l} (\nabla I_1 \cdot (\Delta u, \Delta v)^T)^2 + \sum_{k,l} (\nabla I_1 \cdot (\Delta u, \Delta v)^T) e_{k,l} + E(u, v) \quad (3)
 \end{aligned}$$

where $\nabla I_1 = (I_x, I_y) = \nabla I_1(x + u + k, y + v + l)$ is the intensity gradient and $e_{k,l}$ is the term inside the brackets in Equation (2). Minimizing w.r.t. $(\Delta u, \Delta v)$, we obtain a 2×2 system of equations

$$\begin{bmatrix} \sum_{k,l} I_x^2 & \sum_{k,l} I_x I_y \\ \sum_{k,l} I_x I_y & \sum_{k,l} I_y^2 \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \begin{bmatrix} \sum_{k,l} I_x e_{k,l} \\ \sum_{k,l} I_y e_{k,l} \end{bmatrix} \quad (4)$$

The matrix on the left hand side is referred to as the *Hessian* of the system.

Both methods produce comparable results, which are significantly better than the depth maps generated from any one light field alone. We tested this on a synthetic light field with known 3D geometry, and compared our results that merge the depth estimates from all the light fields to one that uses only a single light field. The results can be seen in Figure 3. In this experiment, we used nine light

fields of a synthetic scene under different illumination directions (left). Each light field has 9×9 images, and only the central image (used as the reference) is shown in Figure 3. In the rest of this work, the local matching error variance is used to improve accuracy of the depth values. The advantage of employing stereo using the Hessian images is that the process is done by only one execution of depth estimation, while the method using the local matching error variance requires n executions of depth estimation when n illumination conditions are obtained.

We chose to compute the local view-dependent geometries because the stereo algorithm, while good, does not produce perfectly accurate geometry. In addition, some degree of photometric variation along the image sequence usually exists, making the direct production of a single accurate global 3D geometry from images very difficult. The local geometries encode such photometric variation, since they are *highly locally photoconsistent*. The stereo algorithm has the tendency to maximize this behavior.

4.3. Extracting Intrinsic Images

We applied Weiss's ML estimation method²⁰ to derive intrinsic light fields. Given a sequence of N light fields with varying illumination, it is decomposed into a single reflectance light field and N illumination light fields. With images of $u \times v$ in size from $s \times t$ view points under n different illumination conditions, we can denote this decomposition as follows:

$$I(s, t, u, v, n) = R(s, t, u, v) \cdot L(s, t, u, v, n) \quad (5)$$

where $I(s, t, u, v, n)$, $R(s, t, u, v)$, and $L(s, t, u, v, n)$ are an input light field sequence, a reflectance light field, and an illumination light field sequence, respectively. In the log domain, (5) is written as (6):

$$i(s, t, u, v, n) = r(s, t, u, v) + l(s, t, u, v, n) \quad (6)$$

For each of M derivative filters $\{f_m\}$, a filtered reflectance light field \hat{r}_m is estimated by taking the median of filtered input light fields:

$$\hat{r}_m(s, t, u, v) = \text{median}_n \{i(s, t, u, v, n) \star f_m\} \quad (7)$$

Finally, $R(s, t, u, v)$ is recovered by deconvolution of the estimated filtered reflectance light fields \hat{r}_m .

4.4. Estimating Shadow Masks

A major difficulty in lighting interpolation is the realistic generation of shadows. To compute shadow masks for real scenes, our approach first infers shadow pixels from the illumination intrinsic image. The intensities in an illumination intrinsic image represent magnitudes of incident irradiance, so image areas with low values indicate shadowed regions. To determine a suitable threshold for distinguishing shadowed from non-shadowed pixels, we adopt the clustering technique of Otsu¹⁵.

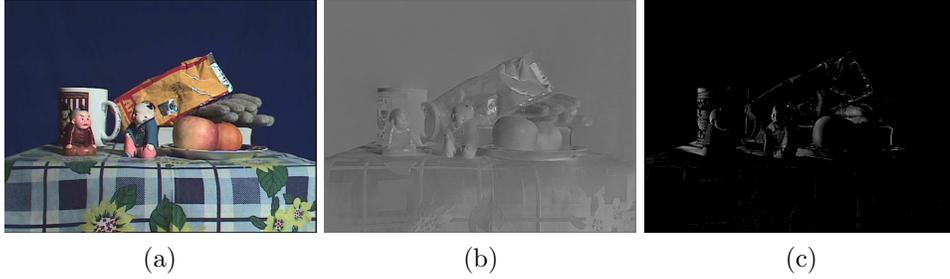


Fig. 4. Example of an illumination image and its shadow mask counterpart. (a) Original image, (b) Illumination image, (c) Shadow masks.

The threshold is computed from maximizing the between-class scatter by minimizing within-class variances. In our case, we classify pixels into two classes, shadowed and lit, assuming shaded pixels can be categorized into either of them. We first create an intensity histogram for each illumination image to obtain the probability density function $p(i)$ where i indicates an intensity value. Assuming that shadowed pixels have relatively lower intensity than lit pixels, we define the cumulative probability functions P for shadowed (P_s) and lit (P_l) area using a threshold value T .

$$P_s(T) = \sum_{i=i_{min}}^T p(i), \quad P_l(T) = \sum_{i=T}^{i_{max}} p(i) \quad (8)$$

In the same manner, we define the mean of the shadowed (μ_s) and lit (μ_l) area as functions of the threshold T as

$$\mu_s(T) = \sum_{i=i_{min}}^T ip(i), \quad \mu_l(T) = \sum_{i=T}^{i_{max}} ip(i). \quad (9)$$

Finally, the optimum threshold value T_{opt} is computed from the following equation.

$$T_{opt} = \arg \max_T \{P_s(T) \cdot P_l(T) \cdot (\mu_s(T) - \mu_l(T))^2\} \quad (10)$$

To estimate shadow masks, the optimum threshold value T_{opt} is determined for each illumination image. These thresholds are then used to extract the shadow masks. Even though these shadow masks include both attached shadows and dimly shaded pixels, the quality of shadow warping is not significantly affected for two reasons. First, the shadow masks are later intersected with geometrically-derived shadows. This operation removes most non-shadow pixels. Second, the shadow masks are treated as alpha mattes that represent the degree of shadow darkness. As a result, brighter pixels contribute less to the final rendered shadow. A shadow mask computed in this manner is shown in Figure 4(c).

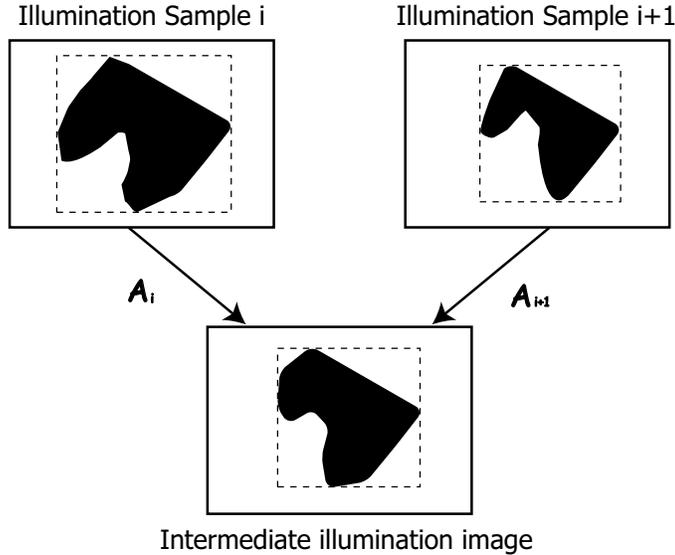


Fig. 5. Illustration of subimage registration over the geometric-based shadow blobs. Changes of intermediate shadows' shape are represented by transformation matrices from neighboring bases, i.e. the geometric-based shadows under sampled illumination conditions.

4.5. Shadow Warping

While the technique described in the previous section allows us to estimate shadow regions for images at sampled illumination conditions, it cannot be applied to intermediate lighting directions. This is because the associated images are not available. Since we are not able to predict the shape of intermediate shadow masks from intrinsic images, we instead predict the general *shadow distortion* between the sampled lighting conditions using the shadows cast from the view-dependent geometries. Although these geometries are not highly accurate, their shadows can be computed for arbitrary light directions, and the distortions in shadow shape as a light source moves from one sampled position to another can nevertheless be helpful in morphing the shadows computed from intrinsic images.

In this process, we first estimate light source type (point / directional) and lighting directions of captured images with some user interaction. By clicking on several pairs of corresponding shadow and object points in an image, the light source position can be determined by least-squares triangulation. With the light position and the estimated geometry, the resulting shadows can be computed. We can also compute the geometric-based shadows for light positions between the sampled illumination directions.

After computing the geometric-based shadows, the changes in the geometric-based shadows are computed by the region-based transformation matrices. Assuming each shadow blob to be a subimage region, we employed subimage registration

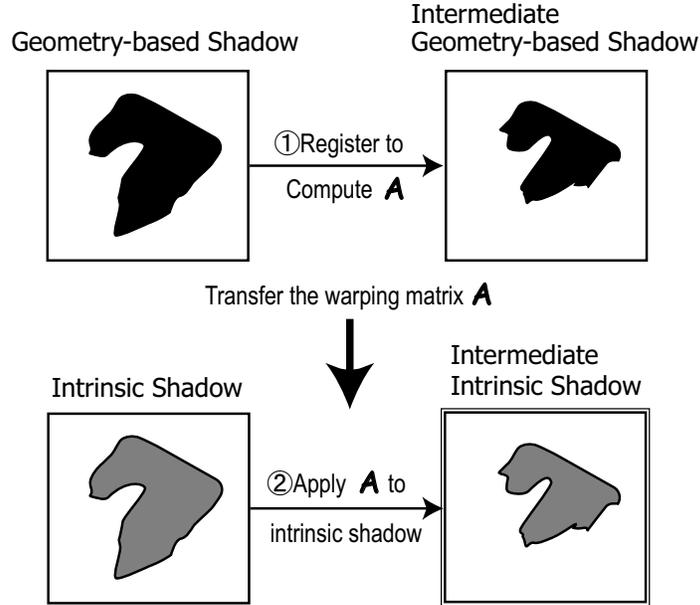


Fig. 6. After computing transformation matrix \mathbf{A} of the geometric-based shadow by subimage registration, the transform \mathbf{A} is then applied to the corresponding intrinsic shadow to generate intermediate shadow.

to compute the region-based shadow transformation matrices. By computing those matrices, the changes in geometric-based shadow shape from one sampled light position to another can be used to guide the transformations of shadows computed from intrinsic images. In Figure 5, transformation matrix \mathbf{A}_i^{+j} corresponds to warping of the shadow blob from base image i to intermediate image j . Since those shadow blobs do not have texture in them, the nearest shadow blobs are assumed to be the corresponding shadow blobs. We assumed that geometric distortion of the geometric-based shadow blobs can be described by linear 2D geometric transformations as long as they are densely computed. Thus, we model the transform as 2D affine and the transform \mathbf{A} is described by a 3×3 matrix.

This is done by attaching the intrinsic image shadows to the geometric-based shadows, and as the geometric shadows are morphed from one sampled lighting to another, the intrinsic shadows are morphed correspondingly as shown in Figure 6. Attachment is done by applying an *AND* operation on the corresponding shadow regions. In this operation, correspondences between the intrinsic shadows and the geometric-based shadows are estimated by checking overlapping regions between them. Since the process is done in the image coordinate frame, finding overlapping shadow regions can be accomplished by pixel-wise test. Figure 7 shows an example of geometric-based shadow warping applied to intrinsic shadow masks. Once the shadow masks are predicted at intermediate illumination conditions, the view can

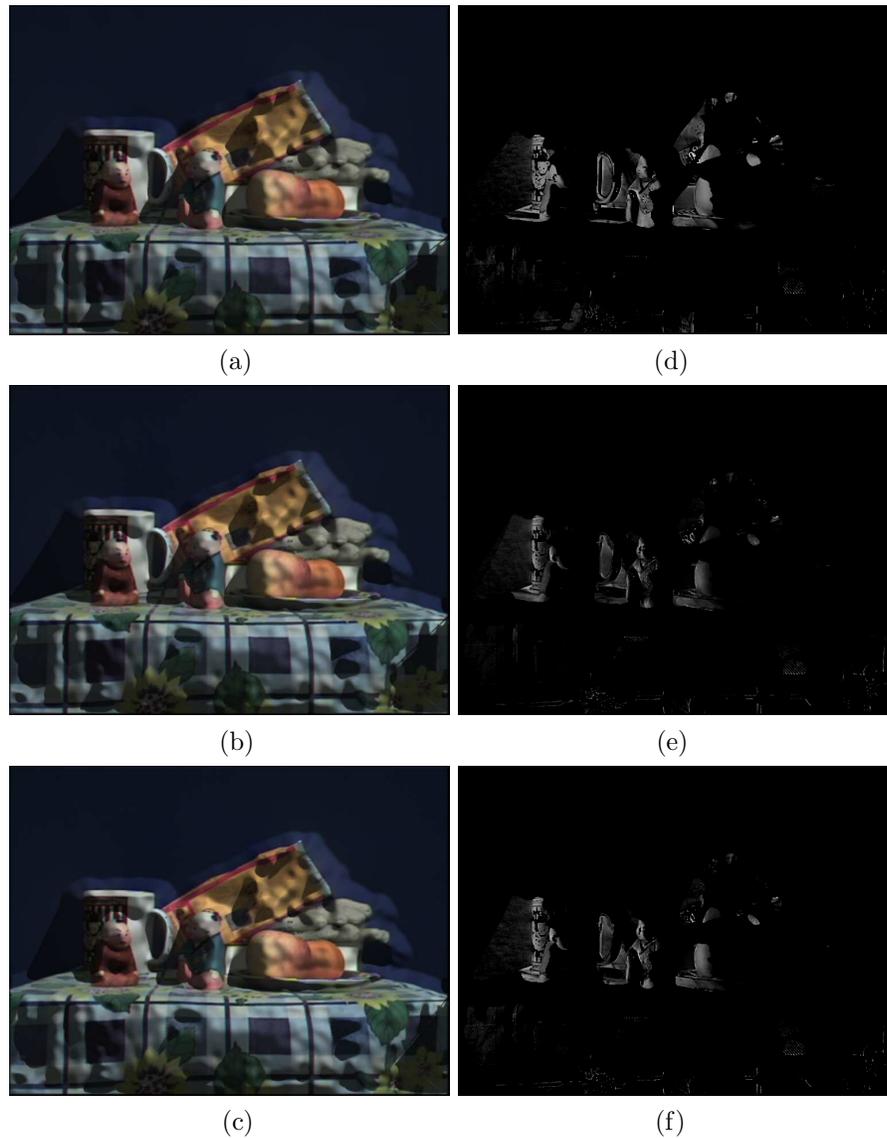


Fig. 7. Illustration of applying the transformation of geometric shadows to intrinsic shadows. (a) Geometric shadows at L1, (b) Geometric shadows between L1 and L2, (c) Geometric shadows at L2, (d) Shadow masks at L1, (e) Shadow masks between L1 and L2 (after applying the geometric-based warping), (f) Shadow masks at L2. L1 and L2 are sampled illumination conditions.

then be synthesized. This synthesis is computed by removing shadows in the sampled images using their intrinsic images, linearly interpolating diffuse reflections, then computing shadows from the intrinsic lumigraphs.

To validate our shadow warping technique, we prepared a simple CG scene

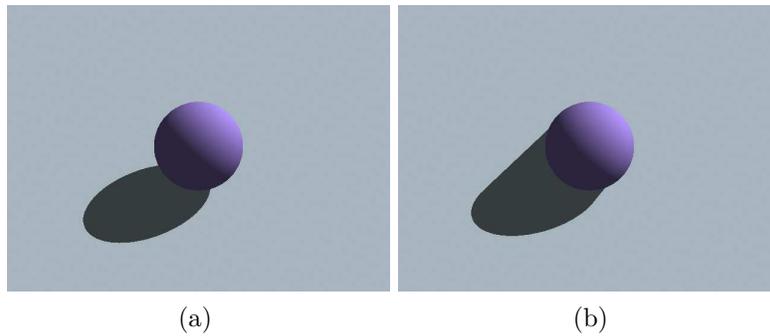
12 *Y. Matsushita, S. B. Kang, S. Lin, H-Y. Shum and X. Tong*

Fig. 8. A CG scene. (a) Original scene, (b) illuminated scene with view-dependent geometry.

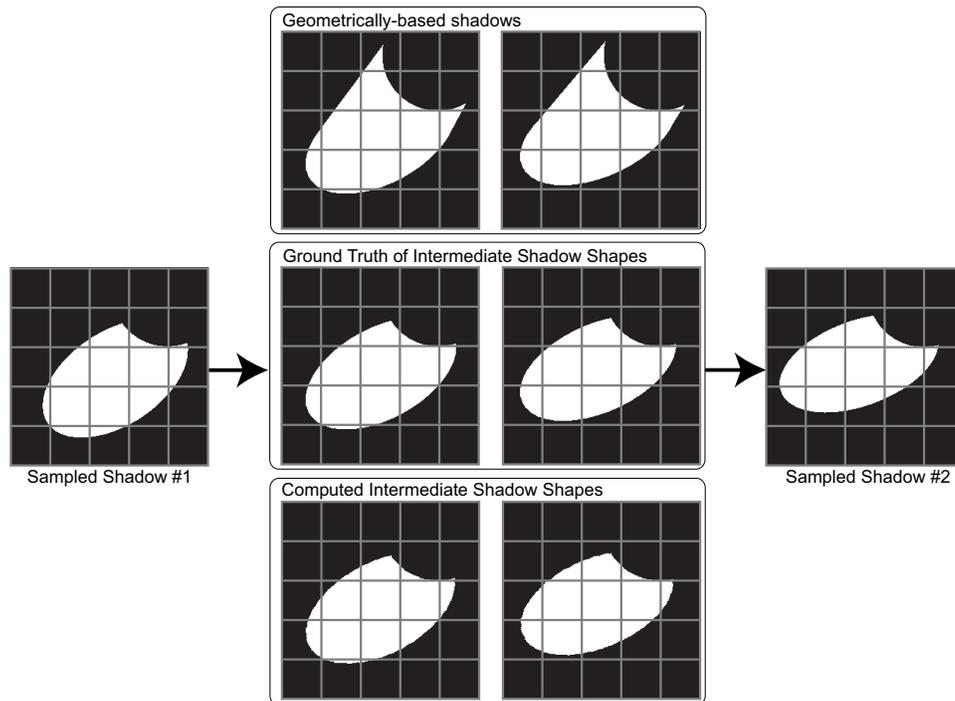


Fig. 9. Intermediate shadow shapes. A grid is overlaid for better visualization.

shown in Figure 8(a).

Since the view-dependent geometries are essentially $2\frac{1}{2}$ representations, they are not able to include occluded surfaces such as the back of the sphere. As a result, their derived shadows may be incorrect, as evidenced in Figure 8(b). Despite this, such shadows provide useful cues on general shadow distortions between sampled il-

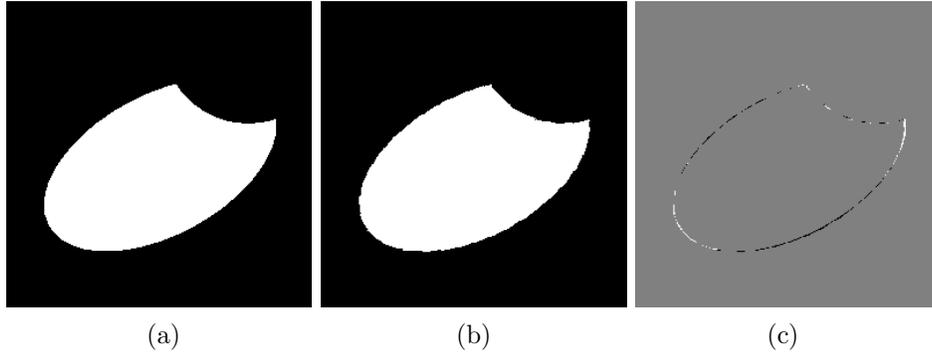


Fig. 10. Comparison between ground truth (a) and the computed intermediate shadow shape (b). In the difference image (c), i.e. (a) - (b), white pixels indicate positive error while black pixels show negative error. Gray pixels represent no error.

lumination conditions. The intermediate geometrically-based shadow shapes can be computed at an arbitrary dense sampling rate, and each deformation is obtained by subimage registration. Finally, those deformation matrices are attached to sampled shadow shapes to obtain the intermediate shadow shapes. Figure 9 depicts the intermediate geometrically-based shadows, ground truth of intermediate shadow shapes that are not used for this computation, and computed intermediate shadow shapes. These computed intermediate shadow shapes are reasonably correct as confirmed in Figure 10.

5. Results

In this section, we show results of lighting interpolation for two real scenes. Both scenes are captured under the roughly parameterized directional light sources. In this experiment, we used 7 directional light sources covering a hemicycle surrounding the scene.

5.1. Toy Scene

Figures 11(a,b) show examples of interpolating the lighting condition of a toy scene with our shadow warping technique and with direct linear interpolation, respectively. For this scene, we captured seven light fields with different lighting conditions, where each light field is composed by 17×17 images. We can clearly see the difference between the results of our method and those of linear interpolation, especially on the cast shadow of the toy on the left. This is more evident by comparing the leftmost two images in Figure 13. The direct linear interpolation resulted in significantly softer shadows, which is less consistent with the original sampled images. And furthermore, a comparison among the ground truth, the result of our method, and that of linear interpolation is shown in Figure 14. As we can see, our method produced a more realistic-looking shadow, compared to linear interpolation.



Fig. 11. (a) Lighting interpolation examples for the toy indoor scene. (b) Lighting interpolation using direct interpolation for the toy indoor scene.

To quantitatively compare the results of our method and simple interpolation method, the difference between the results and the ground truth is computed pixel-by-pixel. In Figure 14, (a) is the result of our method, (b) is the ground truth, and (c) is the result of simple interpolation. The image difference between (a) and (b) is shown in (d), and between (b) and (c) is shown in (e). The image differencing is done by summing up the RGB components' distance. As is shown in a colorbar in the figure, larger differences are colored in red while smaller differences are colored in blue. We can clearly see that a better result is obtained by our method.

5.2. Portrait Scene

Figures 12(a,b) show the results of lighting interpolation of a scene containing a portrait. We captured ten light fields under different illumination conditions for this scene. Each light field is composed by 16×16 images. While cast shadows in Figure 12(b) are blurred and exhibit abrupt movements for linear interpolation, cast shadows warped by our method look more natural in Figure 12(a), and they move more smoothly. This is more evident by comparing the rightmost two images in Figure 13. Again, the direct linear interpolation method resulted in softer shadows, unlike those in the original input images.

6. Conclusions and Future Work

We have described an approach for lighting interpolation of a scene without the need for accurate physically-based rendering or detailed 3D geometry. It uses only light fields captured under different, sparsely sampled, illumination conditions. Our approach uses intrinsic images and local view-dependent depths computed from stereo in order to predict shadows at intermediate illumination conditions, which add significantly to the realism of the synthesized view. The limitation of the method is that the method requires the scene to be largely diffuse, since the reflectance image $R(x, y)$ in Weiss's framework of intrinsic images is assumed to be view-independent. We are working to derive time-varying reflectance images $R(x, y, t)$ and corresponding illumination images $L(x, y, t)$ to overcome its limitation.

In future work, we would like to perform object manipulation such as object insertion and removal, all while enabling realistic scene lighting interpolation. Moreover, we would like to address the more difficult issue of lighting interpolation of outdoor scenes. This has the added difficulty of not being able to capture light fields with a set of consistent illumination conditions, because of the time elapsed between successive camera snapshots during light field capture.

References

1. H. G. Barrow, and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. In *Computer Vision Systems, A. Hanson and E. Riseman (Eds.)*, pages 3–26, 1978.



Fig. 12. (a) Lighting interpolation examples for the portrait indoor scene. (b) Lighting interpolation using direct interpolation for the portrait indoor scene.



Fig. 13. Closeup views. The left of each pair is generated using our method while the other is computed using direct interpolation.

2. S. Boivin, and A. Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. *Computer Graphics (SIGGRAPH)*, pages 107–116, Aug. 2001.
3. P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Computer Graphics (SIGGRAPH)*, pages 189–198, July 1998.
4. A. Fournier, A. S. Gunawan, and C. Romanzin. Common illumination between real and computer generated scenes. In *Graphics Interface*, pages 254–262, May 1993.
5. W. T. Freeman, and P. A. Viola. Bayesian model of surface perception. *Advances in Neural Information Processing Systems (NIPS)*, Vol. 10, pages 787–793, Dec. 1998.
6. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. *Computer Graphics (SIGGRAPH)*, pages 43–54, Aug. 1996.
7. S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in multi-view stereo. In *Conf. on Computer Vision and Patt. Recog.*, Vol. I, pages 103–110, Dec. 2001.
8. E. P. Lafortune, S. C. Foo, K. E. Torrance, and D. P. Greenberg. Non-linear approximation of reflectance functions. *Computer Graphics (SIGGRAPH)*, pages 117–126, Aug. 1997.
9. E. H. Land, and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. of America*, 61(1):1–11, 1971.
10. M. Levoy, and P. Hanrahan. Light field rendering. *Computer Graphics (SIGGRAPH)*, pages 31–42, Aug. 1996.

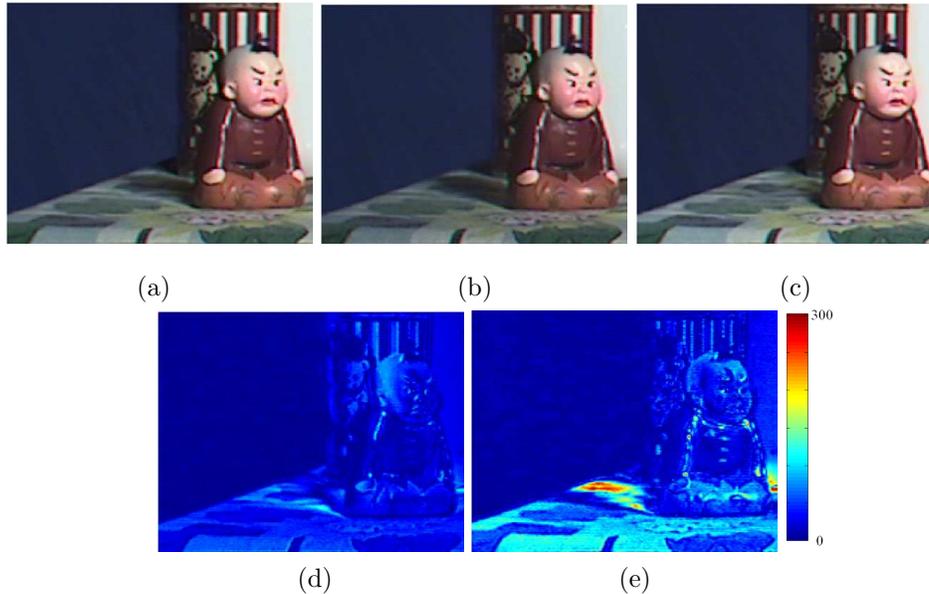


Fig. 14. Comparison with the ground truth. (a) Interpolated result of our method, (b) the ground truth, (c) simple pixel-wise interpolation, (d) difference between our result(a) and the ground truth(b), (e) difference between simple interpolation(c) and the ground truth(b).

11. C. Loscos, G. Drettakis, and L. Robert. Interactive virtual relighting of real scenes. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 6, No.4, pages 289–305, 2000.
12. T. Malzbender, D. Gelb, and H. Wolters. Polynomial texture maps. *Computer Graphics (SIGGRAPH)*, pages 519–528, Aug. 2001.
13. S. R. Marschner, and D. P. Greenberg. Inverse lighting for photography. In *IS&T/SID 5th Color Imaging Conf*, pages 262–265, Nov. 1997.
14. J. Nimeroff, E. Simoncelli, and J. Dorsey. Efficient re-rendering of naturally illuminated environments. *5th Eurographics Workshop on Rendering*, pages 359–373, June 1994.
15. N. Otsu. A Threshold Selection Method From Gray Level Histograms. *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-9, pages 62–66, 1979.
16. I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from brightness in shadows: Adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *Int'l Conf. on Computer Vision*, pages 875–883, Sep. 1999.
17. I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from shadows. In *Conf. on Computer Vision and Pattern Recog.*, volume I, pages 306–312, Nov. 1999.
18. Y. Sato, M. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. *Computer Graphics (SIGGRAPH)*, pages 379–387, Aug. 1997.
19. G. J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH)*, pages 265–272, July 1992.
20. Y. Weiss. Deriving intrinsic images from image sequences. In *Int'l Conf. on Computer Vision*, volume II, pages 68–75, July 2001.
21. T. T. Wong, P. A. Heng, S. H. Or, and W. Y. Ng. Image-based rendering with control-

- lable illumination. In *8th Eurographics Workshop on Rendering*, pages 13–22, June 1997.
22. D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, and D. H. Salesin, and W. Stuetzle. Surface light fields for 3D photography. *Computer Graphics (SIG-GRAPH)*, pages 287–296, July 2000.
 23. Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. *Computer Graphics (SIG-GRAPH)*, pages 215–224, Aug. 1999.
 24. S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstructing Surfaces with Arbitrary BRDFs. In *Int'l Conf. on Computer Vision*, pages 391–398, 2001.
 25. M. Koudelka, S. Magda, P. Belhumeur, and D. Kriegman. Image-based Modeling and Rendering of Surfaces with Arbitrary BRDFs. In *Conf. on Computer Vision and Pattern Recog.*, pages 568–575, 2001.
 26. K. Nishino, Y. Sato, and K. Ikeuchi. Eigen-texture method: Appearance compression based on 3D model. In *Computer Vision and Pattern Recog.* vol. 1, pages 618–624, 1999.
 27. R. Furukawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Appearance Based Object Modeling using Texture Database: Acquisition Compression and Rendering. In *Thirteenth Eurographics Workshop on Rendering*, pages 257–266, 2002.
 28. K. Deng, L. Wang, Z. Lin, T. Feng, and Z. Deng. Correction and rectification of light fields. In *Computers & Graphics.*, 27(2), pages 169–177, 2003.



Yasuyuki Matsushita received his B.Eng., M.Eng. and Ph.D degrees in electrical engineering from the University of Tokyo in 1998, 2000 and 2003 respectively. Currently, he is an associate researcher at Microsoft Research Asia. His research interests include photometric methods in computer vision, image-based rendering and modeling from images.



Sing Bing Kang received his B.Eng. and M.Eng. degrees in electrical engineering from National University of Singapore (NUS) in 1987 and 1990, respectively, and his M.Sc. and Ph.D. degrees in robotics from Carnegie Mellon University (CMU) in 1992 and 1994, respectively. His doctoral work was in enabling robot systems to observe, recognize, and replicate grasping tasks. He was a member of research staff at Compaq's Cambridge Research Laboratory from 1995 to 1999, where he worked on 3D modeling from image sequences and image-based rendering. He is currently a researcher at Microsoft Corporation working on environment modeling from images. His paper on the Complex Extended Gaussian Image had won the IEEE Computer Society Outstanding Paper award at the 1991 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). His IEEE Transactions on Robotics and Automation paper on human to robot hand mapping had been awarded the 1997

King-Sun Fu Memorial Best Transaction Paper award. Dr. Kang is a senior member of IEEE.

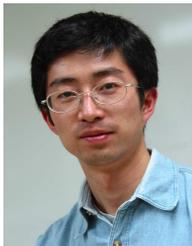


Stephen Lin is a researcher in the Visual Computing Group of Microsoft Research, Asia. His current research interests include reflectance and illumination, physics-based computer vision, and image-based rendering. Steve received his B.S.E. degree in electrical engineering from Princeton University in 1992, and his Ph.D. in Computer Science and Engineering from the University of Michigan in 2000.



Heung-Yeung Shum received his Ph.D. in robotics from the School of Computer Science, Carnegie Mellon University in 1996. He worked as a researcher for three years in the vision technology group at Microsoft Research Redmond. In 1999, he moved to Microsoft Research Asia where he is currently a senior researcher and the assistant managing director. His research interests include computer vision, computer graphics, human computer interaction, pattern recognition, statistical learning and robotics.

He is on the editorial boards for IEEE Transactions on Circuit System Video Technology (CSVT), IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and Graphics Models. He is the General Co-Chair of Tenth International Conference on Computer Vision (ICCV 2005 Beijing).



Xin Tong received his B.S. Degree and Master Degree in Computer Science from Zhejiang University in 1993 and 1996 respectively. He received his Ph.D. in Computer Graphics from Tsinghua University, Beijing in July 1999. He is now a researcher of Microsoft Research Asia. His research interests include image-based rendering and modeling, realistic rendering, texture synthesis and mesh compression.