Specular Object Reconstruction Behind Frosted Glass by Differentiable Rendering

Takafumi Iwaguchi[1], Hiroyuki Kubo[2], and Hiroshi Kawasaki[1]
[1]Kyushu University, Japan iwaguchi@ait.kyushu-u.ac.jp
[2]Chiba University, Japan

Abstract

This paper addresses the problem of reconstructing scenes behind optical diffusers, which is common in applications such as imaging through frosted glass. We propose a new approach that exploits specular reflection to capture sharp light distributions with a point light source, which can be used to detect reflections in low signal-to-noise scenarios. In this paper, we propose a rasterizer-based differentiable renderer to solve this problem by minimizing the difference between the captured and rendered images. Because our method can simultaneously optimize multiple observations for different light source positions, it is confirmed that ambiguities of the scene are efficiently eliminated by increasing the number of observations. Experiments show that the proposed method can reconstruct a scene with several mirror-like objects behind the diffuser in both simulated and real environments.

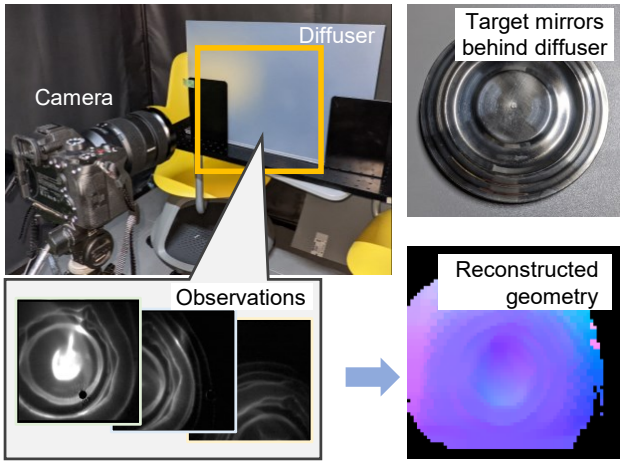

Figure 1. Overview of the proposed method. There are two target mirrors (top right) behind the diffuser (top left); however, under ambient light, these mirrors are hardly observed. We use narrow beams to reconstruct the scene (bottom right) from multiple observations made at different light source locations (bottom left).

1. Introduction

The reconstruction of scenes behind an optical diffuser, such as frosted glass windows or fogging windshields of cars, is desirable for safety and inspection. Examples include inspecting potentially hazardous objects in a space covered with frosted glass for privacy, or inspecting the shape of packaged goods. A similar problem setting can also be found in a study on non-light-of-sight imaging [45]. However, this is a difficult task because the scene is severely blurred by the diffuser and cannot be observed directly. In a typical scenario, an object behind the diffuser is assumed to be a daily commodity with both diffuse and specular reflections. Because specular reflection is usually stronger than diffuse reflection, if the object is illuminated by a point light source, sharp and high-frequency patterns are observed on the diffuser, which are called caustics. Although several algorithms are proposed for "shape from caustics," they mainly focus on designing the shape of mirror or glass to generate a specific target pattern onto the screen or wall [8, 16], and there is no practical solution which can be applied for general conditions, such as recovering a shape of objects behind a diffuser.

The main reason why few solutions have been proposed is that it is theoretically difficult to determine the correspondences between the reflected patterns and light sources. For example, the reflected light of a point light source creates high-frequency patterns not only by high-frequency shapes of mirrors but also by smooth surfaces such as paraboloidal mirrors, which concentrate the light intensity at a single point of the focus of the paraboloid. Moreover, a unique surface cannot be reconstructed from a single observation because of the degenerated nature of reflected light. To solve this problem, we propose a new technique to estimate the shape of an object from the reflection pattern on diffuser images of the reflected light distribution using a differentiable renderer by minimizing the difference between the rendered and several observed images under different

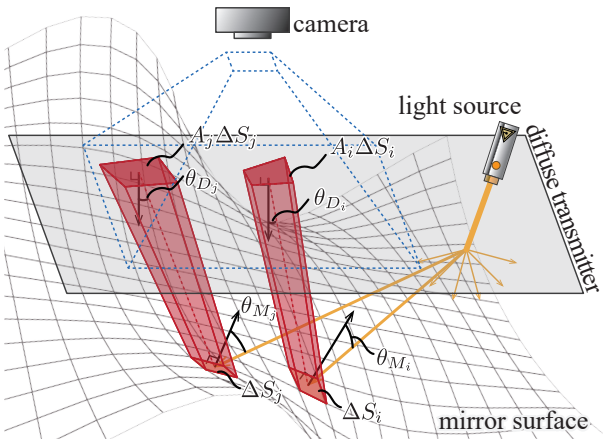

Figure 2. Illustration of light-transport modeling. When a narrow beam strikes the diffuser, the light incident on the i–th and j–th patches is reflected to the diffuser again through the mirror reflection.

lighting conditions. Note that the reflected pattern cannot be reproduced by nonphysics-based differentiable rasterizers [4, 17, 21, 22, 33]. This is a difficult task even for publicly available unidirectional path-tracing based differentiable renders [19, 23, 28, 49] because paths connecting the diffuser, specular surfaces, and point light sources are difficult to sample. Bidirectional path-tracing based differentiable renderers, such as [48], are capable of handling the paths; however, they still suffer from high computational costs and sampling noise, resulting in instability in numerical optimization. Therefore, to make the rendering process of reflected light more efficient, we propose a new type of differentiable renderer that can optimize parameters in a scene with reflection. In the proposed method, reflective objects, such as mirrors, are represented by a polygon mesh, and the light rays emitted from the light sources are reflected on the mesh and finally projected onto the target, that is, a diffuse surface. Considering that the light energy emitted from a light source is conserved until it is projected onto the diffuser, the irradiance on the diffuser can be calculated analytically. The resulting image is rendered using a differentiable rasterizer for sampling the radiance. This two-pass algorithm allows for the rendering of smooth caustics, for which existing physics-based renderers are still poor at. In addition,the rendering cost depends only on the number of mesh vertices and is always constant during the optimization process, independent of the complexity of the reflection pattern. The experimental results confirmed that the proposed method can reconstruct scenes behind a diffuser in both simulated and real environments. In summary, this study makes the following contributions.

- A differentiable renderer considering light transport is proposed, where the renderer efficiently calculates the intensity distribution according to the linear calculation of the direct paths between the light source and scene surface.
- A method of estimating the position and surface normal of a mirror-like object behind a diffuser is proposed, utilizing multiple images of the reflected patterns captured by a camera while a light source is freely moved.
- Real experiments with special data capturing process to avoid noise and direct reflection from a diffuser are conducted to confirm that the proposed algorithm can reconstruct the scene behind a diffuser.

2. Related work

2.1. Differentiable and non-differentiable rendering techniques

2.1.1 Rendering of reflected light

Unidirectional rendering algorithms, such as those involving path and light tracing, are widely used in photorealistic rendering. However, these unidirectional algorithms cannot handle scenes containing ideal pinhole cameras and infinitely small point light sources. Bidirectional path tracing and photon mapping algorithms are reasonable rendering choices for such scenes; however, they still involve large computations and are thus not suitable for our iterative optimization. An efficient method of rendering reflections and refractions by beam tracing is proposed in an early work and is shown to be faster and more accurate than path tracing [5, 7, 10, 14, 37, 43]. This rendering technique is employed for the reflector design in [8] through analysis-by-synthesis via global optimization. We extend the beam-tracing algorithm to differentiable rendering, which allows efficient optimization.

2.1.2 Differentiable rendering

Differentiable rendering has become an important technique for understanding 3D geometry by combining it with machine learning approach [12, 32, 38, 46, 51].

Differentiable rasterizers Rasterization is an operation in mesh rendering, in which 3D polygons are sampled into 2D pixels. The operation involves discrete computation at the boundaries, which prevents differential calculations. The problem is addressed for the shadow region in [27] and for the occluding boundaries in [4, 17, 21, 22]. We follow SoftRas [21] for handling the occluding boundary.

Physics-based differentiable rendering Physics-based differentiable rendering algorithms reproduce complex light transport effects and compute derivatives with respect to the material properties [19, 23, 49]. Methods have been proposed for specific types of light propagation; for example, a method based on photon splatting for caustic rendering [9]

and another for handling the effects of shadows or global illumination [19].

Path-tracing based differentiable rendering algorithms allow us to optimize scene parameters via photorealistic rendering. Most publicly available implementations such as Mitsuba [15, 28] are based on unidirectional Monte Carlo integrals, which are poor at rendering caustics. In general, the light-tracing method performs better in rendering caustics than path tracing. The particle tracer implemented in Mitsuba is an example of a light-tracing method; however, it cannot handle infinitesimal light sources in our scene properly, resulting in the failure shown in Fig. 3.

Path-space differentiable rendering [48], which is based on bidirectional Monte Carlo integration, can render caustics; therefore, it is theoretically applicable in our case. However, a sufficient number of samples per image are required for the rendering results to converge, and the computational cost is multiplied by the number of observations. The proposed method can explore a large solution space for a number of shape parameters at a practical computational cost because the quality of rendering is independent of the number of samplings.

2.2. Shape reconstruction from optical phenomena

The shape reconstruction of an object for a complicated optical property has been intensively researched; e.g., research on the shape reconstruction of a transparent object [16, 18, 30, 44], shape reconstruction for translucent media [6, 26], shape reconstruction from specularity [24, 42] and defocus [39, 50]. A shape from caustics via differentiable rendering is proposed in [16]. This method relies on the refractive properties of the glass volume; therefore, it is not directly applicable to reflective caustics. In [8], the geometry of a reflective surface is constructed using caustic images. This problem setting differs from ours because their goal is to design any surface geometry that generates a target caustic image, whereas our goal is to reconstruct a geometry that generates a caustic pattern. For the technical aspect, beam tracing is employed for rendering, which is based on a mesh-based intensity calculation similar to our method. The shape parameters is estimated via global optimization, which is independent of the rendering algorithm. In contrast, we perform back-propagation through the rendering process for effective optimization. Moreover, we design the scene representation and sampling strategy to be consistent throughout the optimization. Similarly, goal-based caustics also aims to design glass shapes that produce target caustics [25, 31, 36, 47].

Shape reconstruction of non-line-of-sight (NLOS) objects that cannot be directly observed has been widely studied in recent years. Although most approaches adopt time-resolved imaging to reconstruct the shape of an object [3, 11, 13, 29, 40, 41], conventional camera methods have also been proposed [2, 35]. Additionally, object shape reconstruction using a diffuser is proposed [20]. These methods cannot be applied to objects with high specularity, because the material of the object is assumed to be diffuse or retroreflective. We use specularity to recover the shape of many general objects that contain specular reflective components, such as glossy plastics. In [45], specular light paths were exploited for reconstructing an object not in sight. They demonstrated transmissive NLOS imaging utilizing time-resolved measurements, which is similar to our problem setting.

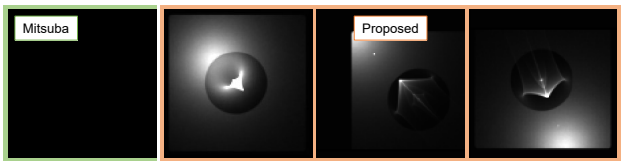

Figure 3. Comparison between Mitsuba3 [15] and the proposed method for LDSDE path, where almost nothing rendered in the scene.

3. Scene parameter optimization with differentiable rendering for object reconstruction

3.1. Scene setting and light transport modeling

We consider a scene with light sources, a mirror or refractive surface, and a diffuse transmitter. The spatial distribution of light is observed using a camera. The light source is assumed to be a narrow beam, as shown in Fig. 2; however, other types of light sources, such as point light and directional light, can be handled in the same manner.

In our scene setting, it is necessary to handle light-diffuse-specular-diffuse-eye (LDSDE) path, which is impossible to be rendered by the path tracing because we suppose infinite thin spot lights. Our insight into the rendering of reflected light is that the energy of the propagating light should be conserved as a consequence of these events. We modeled propagation by considering a small surface area on the reflector. The energy of the incident light over a small area on reflector ΔS should be preserved in the corresponding area $A\Delta S$ in which the reflected light arrives. From the conservation law, the flux is the same on the mirror and on the diffuser, therefore, the irradiance on the mirror E_M and on the diffuser E_D are

$$E_M = \frac{\cos\theta_M \Delta\phi}{\Delta S}, E_D = \frac{\cos\theta_D \Delta\phi}{A\Delta S}, \tag{1}$$

respectively, where a small area on the mirror ΔS receives flux $\Delta\phi$, θ_M is the angle between the mirror normal and incident beam direction, A is the area in which the reflected light arrives, and θ_D is the angle between the diffuser normal and the direction of the reflected beam. Finally, the

irradiance on the diffuser is the inverse of the ratio of the reflected and reflector areas,

$$E_D = \frac{\cos\theta_D}{A\cos\theta_M} E_M. \tag{2}$$

3.2. Differentiable rendering pipeline and optimization

We optimize the scene parameters with respect to the renderer's output by following a common supervised optimization framework using a differentiable renderer. We consider scene parameters comprising (1) light parameters, such as positions and intensities, (2) mirror parameters, such as depth, surface normal maps, and the reflectance maps of mirrors. Errors computed from the measured and rendered images are backpropagated through the differentiable renderer, and the scene parameters are updated directly. We adopt the gradient descent algorithm for the optimization. We consider the data-fidelity term with respect to the output of the renderer. The renderer generates M images that correspond to M light sources. In the following notation, symbols with a hat represent the output of the renderer and those without a hat represent the ground truth. The data-fidelity term for RGB images I_i^{RGB} under the i-th light is $\mathcal{L}_{RGB} = \frac{1}{M}\|\tilde{I_i^{RGB}} - I_i^{RGB}\|_2^2$. The total variation regularization for the normal is added to the loss function to smoothen the estimated normal: $\mathcal{L}_{TV} = \lambda_{TV}\|\tilde{\mathbf{N}}\|_{TV}$. The cost function is formulated as

$$\mathcal{L}_{=}\mathcal{L}_{RGB} + \mathcal{L}_{TV}. \tag{3}$$

Given the interdependence of the scene parameters, it is impossible to determine a unique solution from a single observation. For example, there are ambiguities in reflectance, normal, and distance, even for a flat mirror that does not produce caustics. We input multiple observation images for different light source positions to limit the space of consistent scene parameters in each image.

3.3. Reflection rendering based on the differentiable rasterizer

3.3.1 Mesh representation

The surface of the mirror is represented by a polygon mesh. The reflected light distribution is also represented by a mesh, with each face corresponding to the face of the mirror mesh where reflection occurred. The mesh of the mirror surface comprises a quadrilateral face with four vertices. Normals are defined per face instead of per vertex to constrain the mesh to the shape of the face and prevent divergence of the solution due to its area being zero. Each face normal is sampled from the normal map based on the uv coordinate.

3.3.2 Bidirectional rendering algorithm

The proposed bidirectional rendering algorithm comprises a light pass that assesses light propagation from the light source, and a camera pass that samples the reflected light intensity from the camera.

In the light pass, the distribution of reflected light is computed in a mesh-based calculation; *i.e.*, The reflected light rays that strike the mesh of the mirror are projected onto the diffuse surface, where the mesh irradiance is calculated. First, the direction from the light source to each vertex is calculated. The direction of the reflected ray is calculated as $\mathbf{R} = 2(\mathbf{N}\cdot\mathbf{L})\mathbf{N} - \mathbf{L}$, where $\mathbf{R}$ denotes the reflection vector, $\mathbf{N}$ denotes the normal vector, and $\mathbf{L}$ denotes the lighting vector. The position of the i-th vertex of the mesh of the reflected light P_i is determined by computing the intersection of the reflected ray and diffuser surface. The normal of a vertex is considered the same as that of the face containing the vertex.

The irradiance of the reflected light is calculated by considering the ratio between the area and the corresponding area on the mirror surface following Eq. (2). The received irradiance on the mirror surface is attenuated and is approximated using the average distance E_{Mj} of each vertex of face j. The bidirectional transmittance distribution function (BTDF) is used to account for the diffuser transmittance. We assume a model in which the function depends on the angle of incidence to the diffuser and a parameter α as $F(\omega_i, \alpha) = \cos^{\alpha}\omega_i$. Letting F_i and F_o be BTDF at incidence and emission, respectively, the irradiance on the diffuser is calculated as

$$E_{Dj} = \frac{\mathbf{R}\cdot\mathbf{L}}{A(\mathbf{N}\cdot\mathbf{L})} F_i F_o E_{Mj}. \tag{4}$$

The calculated irradiance is stored as the texture color of each face of the mesh, representing the reflected light uniformly.

In the camera path, the path starting from the camera is connected to the path from the light source. Because the path from the light source is represented as a reflected light mesh, it can be easily sampled at each pixel of the camera by rasterizing the mesh. The intensity of each pixel is the sum of the intensities of the faces of the mesh on that pixel to reproduce the overlap of the reflected light. We calculate the probability map around each triangle by following SoftRas [21] to make the mesh boundary differentiable.

4. Experiment

4.1. Simulation experiment

4.1.1 Reconstruction of scene with multiple mirrors

In this section, we reconstruct the geometry of the multiple mirrors. The experiments are performed on the different combination of the optimization parameters.

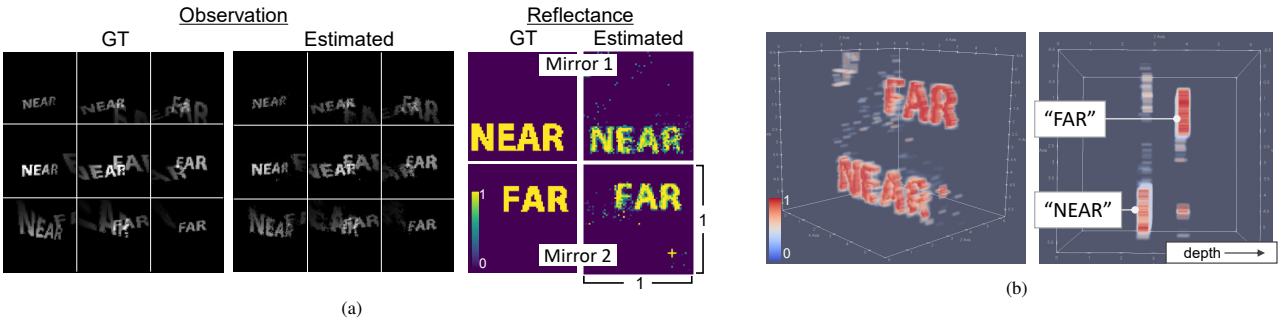

Figure 4. Simultaneous estimation of depth and shape. (a) The reflectance distributions of the two mirrors are correctly decomposed, although the reflected light from the two mirrors overlap in the observational images. (b) Reflectance distribution of 3D volume. Each mirror is estimated at a different depth.

Table 1. Estimated depths of two flat mirrors. The mirror size to be reconstructed is 1 for simulation and 10 cm for the real world.

	Simulation		Real-world	
	estimated	GT	estimated	GT
mirror 1	0.41	0.40	10.37	10.44
mirror 2	0.62	0.60	13.73	14.05

Simultaneous estimation of depth and shape Two flat mirrors with a different boundary shape are placed at different depths. Given a known normal direction, the depth of each mirror and the mirror boundary are reconstructed. There is the ambiguity between the depth and the reflectance, for example, a wrong depth and the mirror with shifted reflectance map also can generate the same observation image. 16 observation images under different light positions are provided. The size of the reconstruction region of the mirror is set to 1×1, and the shape of the mirrors is estimated as the reflectance map. As shown in Fig. 4(a), the light reflected from the two mirrors overlaps in the observed image. The proposed method successfully separates the reflectance distributions of two mirrors. Noise in the reflectance map is thought to be caused by the ambiguity in the region where the reflected light from the two mirrors overlaps. As shown in table 1, the depth is estimated with high accuracy and it is visually confirmed in Fig. 4(b).

Simultaneous estimation of depth and normal direction Two flat mirrors in the scene are placed at different depths, and oriented in different directions. Given a mirror boundary, the depth of each mirror and the normal directions are reconstructed. Nine observation images at different light positions are provided. The initial value for the depth is 0.3, and the normal direction is front-parallel for each mirror. As shown in Fig. 6, the GT and estimated observed images closely matched. The estimated depths and normal directions are estimated with a few error as shown in table 2. The results suggest that providing multiple observations successfully mitigates ambiguity; however, it is still insufficient to determine the exact solution.

Table 2. Numerical evaluation of Simultaneous estimation of depth and surface normal.

	Mirror 1	Mirror 2
Distance (GT)	0.189 (0.200)	0.411 (0.400)
Angle error [degree]	10.41	9.65

4.1.2 Normal reconstruction of an object behind the diffuser

Accuracy of the normal reconstruction is estimated in this experiment. The ground-truth observation images are generated by the physics-based renderer LuxCore [1], whose bidirectional path tracing algorithm can efficiently generate reflective caustics. The input data and results are presented in Fig. 5(b). It is noteworthy that even a simple normal distribution of #1 generates complex sharp reflection patterns that vary according to the incident light positions, as shown in Fig. 3. The normal is reconstructed from 16 observations corresponding to 16 incident light positions arranged on a 4×4 grid with a distance of 0.5 between adjacent sources. The size of the mirror area is 2×2, and the distance between the diffuser and object is 0.4. The resolution of the observation image is 128 x 128 and the resolution of the normal map is 64 x 64. λ_{TV} is set to 0.01.

The results show that all the reconstructed normal maps and rendered images represent the global ground truth. #1 and #2 represent concave and convex shapes, respectively, producing caustics at different locations. #3 consists of two planes of different orientations and wavy region in the middle. #4 shows a letter composed of curves which is raised from the plane and #5 shows the concave relief in the shape of pepper. In all cases, the reconstructed normals reproduce the edges of the original normal; however, we can observe a slight disturbance at the edges of the curves. This is because the caustics rendered from the discretized normals of

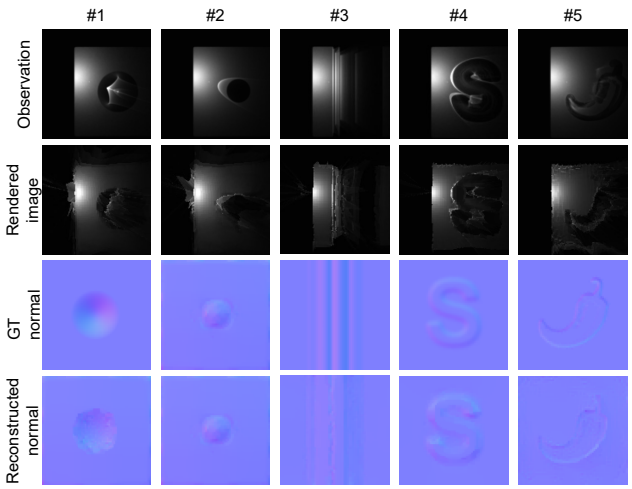

Figure 5. Shape reconstruction results in simulation environment. It is confirmed that the proposed method can estimate normals for multiple shapes including concave and convex surfaces, and straight and curved edges.

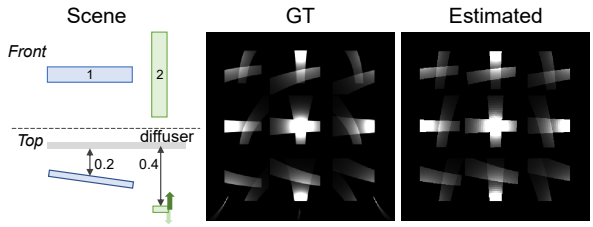

Figure 6. Simultaneous estimation of depth and surface normal of mirrors placed at different depths in different directions (left). The GT (middle) and estimated observations (right) under nine light source positions are closely matched.

the proposed method do not fully reproduce the sharp caustics generated from the smooth surfaces. The mean angular errors (MAE) for #1 to #5 are 1.91, 0.98, 1.06, 1.00, 1.84, and 4.14, respectively.

Evaluation on observation number We investigate the dependence of the reconstruction accuracy on the number of observations. Multiple images corresponding to different light-source configurations are synthesized for the same scene. The number of light sources is one, or 4, 9, 16, or 25 incident points placed on a uniform grid of 1×1 area. The error between the reconstructed normal and the ground-truth normal is averaged over five different scenes. The normal step at which the cost function is minimized is considered as the reconstructed result.

The results are shown in Fig. 7. When there is no noise in the observations (green), the error is large for one observation because of the ambiguity in the normal that generates a caustic pattern. While four observations are sufficient for

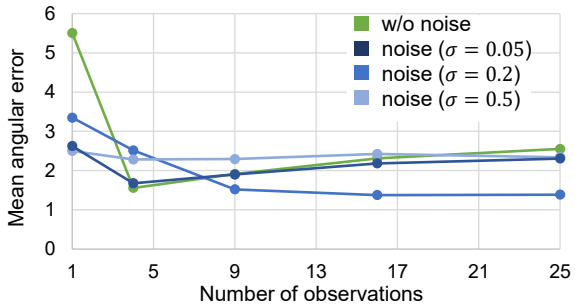

Figure 7. Normal error corresponding to observation number. It is confirmed four input are sufficient for noiseless observations, however, the error increases as the number of observations increases, which may suggest the side-effect of total-variation regularization.

estimating correct normal, for nine or more observations, the error increases as the number of observations increases without noise. This is because the resolution of the image is higher than that of the normal map, and the total variation is included in the cost; thus, the normals are oversmoothed and become closer to flat. The results for different noise levels (blue lines) confirm that for a sufficient number of observations ($n = 4$), the accuracy degrades with the magnitude of the noise. For more observation numbers, noise impacts both the stability of computational optimization and the precision of normal estimation, engendering a trade-off between these aspects. Regarding computational stability, introducing a moderate level of noise can forestall convergence towards local optima. Conversely, excessive noise tends to yield flat normals due to inconsistent observations. We found that at $\sigma = 0.2$, a balance between these two effects was achieved, resulting that the error consistently decreases as the number of observations increases. It can be concluded that there is a possibility to obtain more accurate normals with a larger number of measurements. However, there is a range of appropriate noise levels, and noise must be properly controlled in a real environment.

Comparison with other methods To evaluate the accuracy of the proposed method, experiments are conducted for comparison with the other methods. Because no prior method existed, we compared the normal reconstruction acculacy with a CNN-based baseline method and the modified implementation of [16]. The CNN-based method has a UNet [34] structure in which takes multiple observations that are stacked in the channel direction as input and output normals. The network is trained in a supervised manner using the ADAM optimizer with a learning rate of 0.00001 and an epoch count of 5000. A dataset for training consisting of 1,000 synthesized normals and observation images generated from the corresponding normal. Each synthesized normal represents a single spherical concave surface of random radius, scaled by a factor of 0.1 in x, y direction. The dataset is divided into 800 training data and 200

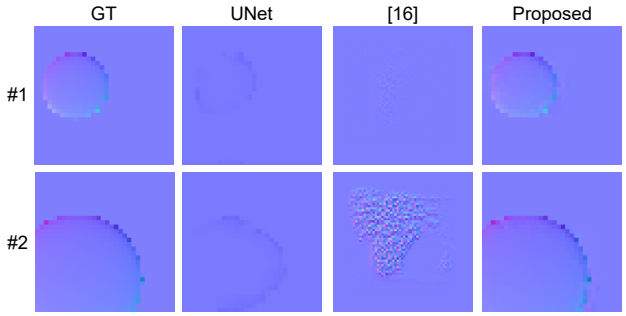

Figure 8. Comparison with baseline methods. While proposed method correctly estimates the normal distribution, CNN-based method fails to estimate the normal. [16] performs poorly for the reflective setup.

test data. For [16] that assumes refractive systems is appropriately modified to accommodate reflective systems. We chose Landweber-pixel method for the reconstruction, and set hyper-parameters $\tau_p = 0.1$ as recommended in the paper. The height map is estimated from a sinle observation and it is converted into normal map for the evaluation. The number of observations is fixed at 16.

The results are shown in Fig. 8. The mean angular errors (MAE) for CNN, [16], proposed are 4.049, 15.23, and 0.487, respectively. While the CNN-based method successfully estimates the sphere's position, the direction and magnitude of the normal deviate from the ground truth. This indicates the difficulty in estimating the normal from the observed images only, without prior physical knowledge. [16] performs poorly for the reflective setup, because it relies physical constraints of the refractive volume which is not available.

4.2. Real-world experiment

4.2.1 Experimental setup

The experimental setup is shown in Fig. 9. The target object is placed on a stage behind optical diffuser made of polyethyleneterephthalate sheets, and a red laser and camera (Sony ILCE-7SM3) is placed in front of the diffuser.

For each incident light position, two images are taken: one with the object and one without the object to extract and remove the region of incident light on the diffuser by calculating the difference between the two images. We assume the light is incident at the brightest pixel. The perspective transformation is performed to obtain the image on the diffuser.

4.2.2 Position and shape estimation of the flat mirror

In this experiment, the placement and shape of the mirrors comprising planes are estimated. Flat mirrors in the shape of the letters 'C' and 'V' are placed in the scene, as shown

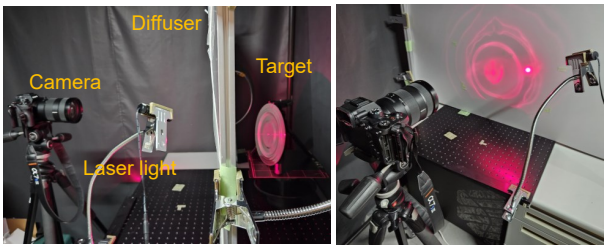

Figure 9. Experimental setup. Because the target object is placed behind the diffuser, it cannot be observed directly.

in Fig. 1. The size of the reconstruction region of the mirror is set to $10cm \times 10cm$. Input are four observation images at different light source positions. The rendered images after convergence and estimated shape are shown in Fig. 10(a) and a volume rendering in (b). It is confirmed that the two mirror-like objects are estimated correctly. An artifact exists at the top of the image, which is considered to be the effect of the remaining backscattered light in the observed image and possibly wrong BTDF parameters which adjusted manually. As shown in table 1, the depth is estimated with error, which is possible because of the ambiguity between the position and surface normal.

4.2.3 Estimation of the normal for arbitrary objects

Multiple measurements are performed for each real object in order to estimate the surface normal. Nine input images are 9 images. The incident point of the light source is adjusted such that the reflected light falls within the plane of the diffuser, and the incident points are selected such that they do not overlap with the reflected light as much as possible. Observations 1 and 2 (first and third columns) in Fig. 11(a) show examples of observation images. The object is represented by a planar polygon, the depth and size of the plane are given, and the normal map and reflectance map (fifth and sixth columns) are the parameters to be optimized. The initial values of the normal map are set uniformly to roughly match the orientation of the object, as measured manually. The observation images generated using the estimated parameters are presented in the second and fourth columns of the figure.

A and B are objects consisting of a single plane. The estimated reflectance represents the shape of the object and the normal map is uniform, indicating that the estimation is successful. The estimated observation image has the same profile as that of the ground-truth image. However, the estimated radiance distribution does not vary significantly, whereas the ground-truth image shows a large gradation range. This may be due to the fact that the given depth and BTDF of the diffuser are different from the actual values.

C consists of two planes. The estimated reflectance represents two separate mirror planes. The normals of each

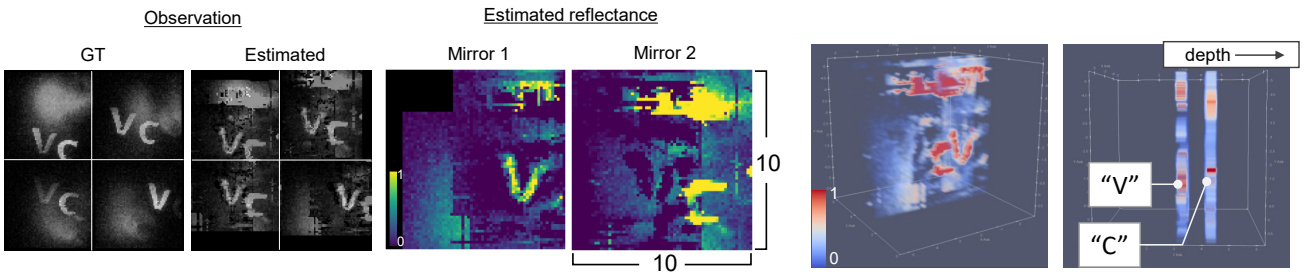

Figure 10. Reconstruction of flat mirrors at different depths. (Top) The estimated depth and reflectance distribution almost reproduce the ground-truth observed image. Artifacts in the reflectance distribution are caused by light reflected from the background in the observed images. (Bottom) Each mirror is estimated at a different depth in 3D volume of the reflectance distribution.

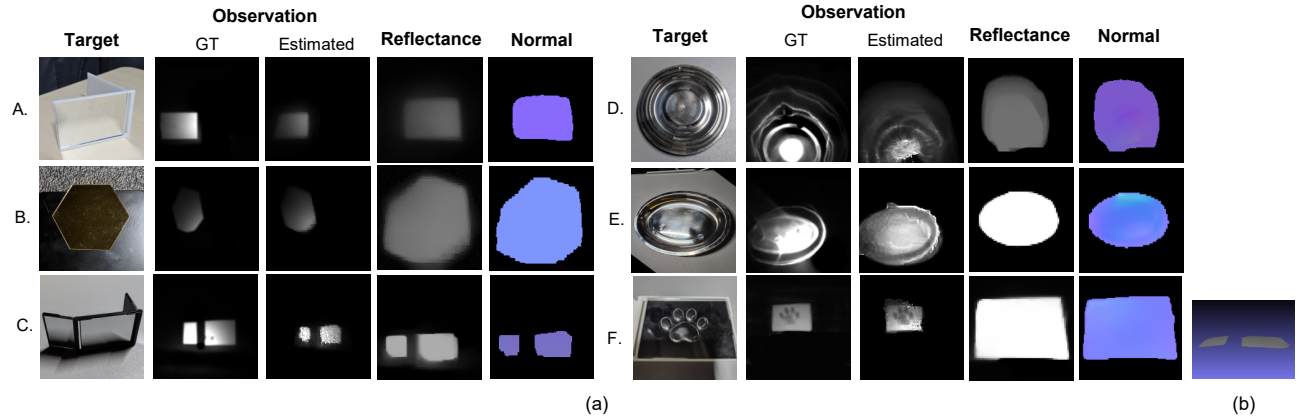

Figure 11. (a) Normal and reflectance reconstruction of real mirror-like objects. Normal and reflectance are simultaneously reconstructed for A–C. Only the normal is reconstructed for D–F. using nine observations. (b)3D reconstruction of shape C.

plane are uniform, and their directions are different, as can be seen from the normal (rightmost column), indicating that the estimation is successful. Figure 11(b) shows 3D reconstructed of this shape using Poisson integration, which correctly represents the angle between two mirrors.

D-F is a complex shape containing multiple curved surfaces. The reflected light produced by the convex shape is weak and cannot be captured by the dynamic range of the camera due to the luminance difference from the caustics. To prevent the reflectance from being estimated to be zero due to the convex shape, reflectance maps are generated using the reflected light from the flat planes at the object boundaries. In the reflectance map, the interior of the object region is filled with uniform values. For D-F, only the normal map is optimized. The pot lid in D has an uneven shape, and caustics can be seen in the observed image. As shown in Fig. 1, the reflection pattern of such a shape varies in distribution and brightness. In the result of D, the normal distribution shows that a shape with a concave center and multiple steps is estimated. In the result of E, the shape of the dish with larger normal variations is estimated, and, in F, smaller normal variations that produce complex caustics are successfully estimated.

5. Conclusion

We demonstrated that our rasterizing-based differentiable rendering method optimizes the parameters of a scene, including specular reflection, and demonstrated measurements for actual and simulated scenes. The proposed differentiable rendering has a clear advantage in efficiently reproducing smooth caustics at a low computational cost, and allows us to optimize the object shape represented by many parameters. One limitation of the proposed method is that it cannot render scenes with more than two bounces. Resolving this issue will be addressed in future research.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Number JP19H04138, JP20H00611, JP21H01457, JP23H03439, JST FOREST (JPMJFR206I), and JST ACT-X (JPMJAX23C2).

References

[1] Luxcorerender. https://luxcorerender.org/. 5

[2] Manel Baradad, Vickie Ye, Adam B. Yedidia, Frédo Durand, William T. Freeman, Gregory W. Wornell, and Antonio Torralba. Inferring light fields from shadows. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[3] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 3

[4] Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson, and Sanja Fidler. *Learning to Predict 3D Objects with an Interpolation-Based Differentiable Renderer*. Curran Associates Inc., Red Hook, NY, USA, 2019. 2

[5] Steven Collins. Adaptive splatting for specular to diffuse light transport. In Georgios Sakas, Stefan Müller, and Peter Shirley, editors, *Photorealistic Rendering Techniques*, pages 121–135, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg. 2

[6] Jack Erdozain, Kazuto Ichimaru, Tomohiro Maeda, Hiroshi Kawasaki, Ramesh Raskar, and Achuta Kadambi. 3d imaging for thermal cameras using structured light. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 2795–2799, 2020. 3

[7] Manfred Ernst, Tomas Akenine-Möller, and Henrik Jensen. Interactive rendering of caustics using interpolated warped volumes. In *Proceedings of Graphics Interface 2005*, GI 2005, pages 87–96, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. Canadian Human-Computer Communications Society. 2

[8] Manuel Finckh, Holger Dammertz, and Hendrik P.A. Lensch. Geometry construction from caustic images. *ECCV*, 6315 LNCS(PART 5):464–477, 2010. 1, 2, 3

[9] Jeppe Revall Frisvad, Lars Schjøth, Kenny Erleben, and Jon Sporring. Photon differential splatting for rendering caustics. *Comput. Graph. Forum*, 33(6):252–263, Sept. 2014. 2

[10] Paul S. Heckbert and Pat Hanrahan. Beam tracing polygonal objects. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, page 119–127, New York, NY, USA, 1984. Association for Computing Machinery. 2

[11] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B. Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 3

[12] Tao Hu, Zhizhong Han, Abhinav Shrivastava, and Matthias Zwicker. Render4completion: Synthesizing multi-view depth maps for 3d shape completion. In *ICCV workshop*, pages 4114–4122, 10 2019. 2

[13] Julian Iseringhausen and Matthias B. Hullin. Non-line-of-sight reconstruction using efficient transient rendering. *ACM Trans. Graph.*, 39(1), jan 2020. 3

[14] K. Iwasaki, Y. Dobashi, and T. Nishita. Efficient rendering of optical effects within water using graphics hardware. In *Proceedings Ninth Pacific Conference on Computer Graphics and Applications. Pacific Graphics 2001*, pages 374–383, 2001. 2

[15] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org. 3

[16] Marc Kassubeck, Florian Burgel, Susana Castillo, Sebastian Stiller, and Marcus Magnor. Shape from caustics: Reconstruction of 3d-printed glass from simulated caustic images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2877–2886, January 2021. 1, 3, 6, 7

[17] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *CVPR*, 2018. 2

[18] K.N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1448–1455 Vol. 2, 2005. 3

[19] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 37(6):222:1–222:11, 2018. 2, 3

[20] David B Lindell and Gordon Wetzstein. Three-dimensional imaging through scattering media based on confocal diffuse tomography. *Nat. Commun.*, 11(4517):1–13, 2020. 3

[21] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. *ICCV*, Oct 2019. 2, 4

[22] Matthew M. Loper and Michael J. Black. Opendr: An approximate differentiable renderer. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *ECCV*, pages 154–169, 2014. 2

[23] Guillaume Loubet, Nicolas Holzschuch, and Wenzel Jakob. Reparameterizing discontinuous integrands for differentiable rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6), Dec. 2019. 2

[24] Satya P Mallick, Todd E Zickler, David Kriegman, and Peter N Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 619–626. Ieee, 2005. 3

[25] Jocelyn Meyron, Quentin Mérigot, and Boris Thibert. Light in power: A general and parameter-free algorithm for caustic design. *ACM TOG*, 37(6), Dec. 2018. 3

[26] Srinivasa G Narasimhan, Shree K Nayar, Bo Sun, and Sanjeev J Koppal. Structured light in scattering media. In *Proc. of TenthICCV*, volume 1, pages 420–427. IEEE, 2005. 3

[27] Thu Nguyen-Phuoc, Chuan Li, Stephen Balaban, and Yong-Liang Yang. Rendernet: A deep convolutional network for differentiable rendering from 3d shapes. In *Advances in Neural Information Processing Systems 31*, 2018. 2

[28] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6), Dec. 2019. 2, 3

[29] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. 3

[30] M. O'Toole, J. Mather, and K. N. Kutulakos. 3D shape and indirect appearance by structured light transport. *IEEE TPAMI*, 38(7):1298–1312, July 2016. 3

[31] Marios Papas, Wojciech Jarosz, Wenzel Jakob, Szymon Rusinkiewicz, Wojciech Matusik, and Tim Weyrich. Goal-based caustics. In *Comput. Graph. Forum*, 2011. 3

[32] Jhony K. Pontes, Chen Kong, Sridha Sridharan, Simon Lucey, Anders Eriksson, and Clinton Fookes. Image2mesh: A learning framework for single image 3d reconstruction. In *ACCV*, 2017. 2

[33] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020. 2

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. 6

[35] Charles Saunders, John Murray-Bruce, and Vivek K Goyal. Computational periscopy with an ordinary digital camera. *Nature*, 565(7740):472–475, 2019. 3

[36] Yuliy Schwartzburg, Romain Testuz, Andrea Tagliasacchi, and Mark Pauly. High-contrast computational caustic design. *ACM TOG*, 33(4), 2014. 3

[37] Musawir A. Shah, Jaakko Konttinen, and Sumanta Pattanaik. Caustics mapping: An image-space technique for real-time caustics. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):272–280, mar 2007. 2

[38] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhöfer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2019. 2

[39] Yuichi Takeda, Shinsaku Hiura, and Kosuke Sato. Fusing depth from defocus and stereo with coded apertures. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 209–216. IEEE, 2013. 3

[40] Chia-Yin Tsai, Aswin C. Sankaranarayanan, and Ioannis Gkioulekas. Beyond volumetric albedo — a surface optimization framework for non-line-of-sight imaging. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1545–1555, 2019. 3

[41] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3(1):1–8, 2012. 3

[42] M. Visentini-Scarzanella, D. Stoyanov, and G.Z. Yang. Metric depth recovery from monocular images using shape-from-shading and specularities. In *ICIP*, pages 25 –28, Orlando, USA, 2012. 3

[43] Mark Watt. Light-water interaction using backward beam tracing. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '90, page 377–385, New York, NY, USA, 1990. Association for Computing Machinery. 2

[44] Bojian Wu, Yang Zhou, Yiming Qian, Minglun Cong, and Hui Huang. Full 3d reconstruction of transparent objects. *ACM Trans. Graph.*, 37(4), jul 2018. 3

[45] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *CVPR*, 2019. 1, 3

[46] Xinchen Yan, Jimei Yang, Ersin Yumer, Yijie Guo, and Honglak Lee. Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 1704–1712, 2016. 2

[47] Yonghao Yue, Kei Iwasaki, Bing Yu Chen, Yoshinori Dobashi, and Tomoyuki Nishita. Poisson-based continuous surface generation for goal-based caustics. *ACM TOG*, 33(3), 2014. 3

[48] Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. Path-space differentiable rendering. *ACM Trans. Graph.*, 39(4), aug 2020. 2, 3

[49] Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. Physics-based differentiable rendering: From theory to implementation. In *ACM SIGGRAPH 2020 Courses*, SIGGRAPH '20, New York, NY, USA, 2020. Association for Computing Machinery. 2

[50] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded aperture pairs for depth from defocus. In *ICCV*, pages 325–332, 2009. 3

[51] Rui Zhu, Hamed Kiani, Chaoyang Wang, and Simon Lucey. Rethinking reprojection: Closing the loop for pose-aware shapereconstruction from a single image. In *ICCV*, 07 2017. 2