Polarimetric Helmholtz Stereopsis
Supplementary Material

Yuqi Ding[1] Yu Ji[2] Mingyuan Zhou[2] Sing Bing Kang[3] Jinwei Ye[1]
[1]Louisiana State University [2] DGene U.S.A. [3] Zillow Group

Contents

1. More Details on Table 1 and Equation 9

Table 1 compares our method with classical 3D reconstruction methods. Readers can find more about these methods in the following references: multiview stereopsis (MVS) [5], photometric stereopsis (PS) [8], structured light (SL) [3], shape-from-polarization (SfP) [6], and Helmholtz stereopsis (HS) [10]. In this table, we assume all classical methods use a traditional camera, whereas our method (polar-HS) uses a polarization camera. It is worth noting that the minimum number of input for SfP is 1 when a polarization camera is used. However, the reconstruction accuracy is low. In our method, the minimum number of input is 4 (one reciprocal pair) because we need to capture two images under 0° and 90° linearly polarized lighting at each camera position in the pair. In addition, note that the accuracy of MVS is only moderate when the number of input images is small. The method is capable of high accuracy when given a large number of images (more than 10) captured from different viewpoints.

We are aware that in recent years, many learning-based methods are proposed for 3D reconstruction from a single and/or multiple images [7, 9]. These methods are usually augmented with some form of data prior by training on a labeled dataset. In contrast, our method relies exclusively on the captured data and uses physical constraints for reconstruction. Therefore, we only compare with the physics-based methods that are more relevant to our approach in Table 1.

In Equation 9, S_a, S_b, and S_l^g are of dimension 4×1 as they are Stokes vectors. The Kronecker product $\otimes$ between S_a (or S_b) and S_l^g results in a matrix of dimension 16×1. ρ_a and ρ_b are two scalars that would not affect the matrix dimension. $\mathbf{n}$, $\mathbf{v}_a$ and $\mathbf{v}_b$ are 3D vectors. The left-hand side of Equation 9 ends up as a 16×1 matrix.

2. Acquisition System Prototype

System Construction. As shown in Fig. 1, our acquisition system consists of a monochrome polarization camera (FLIR Blackfly S Polar-Mono) and a pico projector (Sony MP-CL1A) as point light source. The polarization camera captures four directional polarization images (*i.e.*, I^0, I^{90}, I^{45}, and I^{135}) in one shot as its sensor has on-chip polarizers[1]. The camera uses a 25mm F/5.6 lens. The exposure time we use is 800 ms. With the four directional polarization images captured by the camera, we can compute the Stokes vector as $S(0) = I^0 + I^{90}$, $S(1) = I^0 - I^{90}$, and $S(2) = I^{45} - I^{135}$. Note that since we consider only linear polarization states, $S(3) = 0$.

[1]https://www.sony-semicon.co.jp/e/products/IS/industry/technology/polarization.html

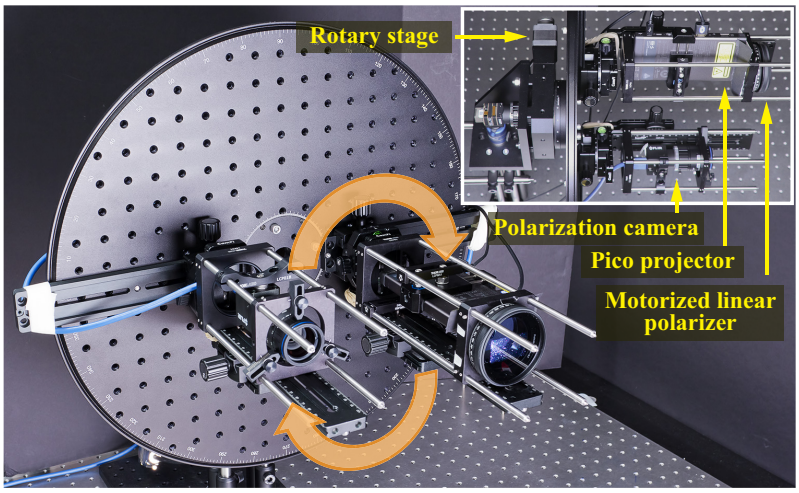

Figure 1. Acquisition system prototype.

We mount a linear polarizer in front of the projector to generate polarized light. The camera and projector are mounted on a rotating wheel so that their positions can be precisely exchanged. The distance between the camera and light source is around 17.5 cm. All moving parts of our system (*i.e.*, the rotating wheel and the light source polarizer) are controlled with motorized rotators; our acquisition procedure is fully automated. The distance between our capture system and the scene is about 50 cm. This distance is chosen based on the camera focal length. The sizes of objects captured are between 5 to 25 cm.

System Calibration. Both the camera and projector are calibrated geometrically and radiometrically. For geometric calibration, we measure the intrinsic and extrinsic parameters of the camera and projector [4] in order to extract their relative positions. For radiometric calibration, we compensate for light anisotropy and camera response function using the method of Jankó *et al*. [2]. We also calibrate the light source polarization state with respect to the camera's. Specifically, we place a mirror in front of the acquisition system to allow the camera to capture an image of the light source. We then turn the polarizer in front of the light source and observe the polarization image captured by the camera. When I^{90} is at its darkest, we consider the light source's angle of polarization to be $0°$. When I^{0} is at its darkest, we consider the light source's angle of polarization as $90°$.

Acquisition Procedure. A reciprocal pair is captured by rotating the wheel $180°$. At each position in the pair, we capture two polarization images under $0°$ and $90°$ polarized light by turning the polarizer in front of the light source. Hence, we have 4 polarization images in a reciprocal pair (captured under two camera positions and there are two lighting conditions in each position). Each polarization image can be decoded into four directional components (*i.e.*, I^{0}, I^{90}, I^{45}, and I^{135}). The total acquisition time of one reciprocal pair is about 2 minutes. Additional reciprocal pairs can be captured by rotating the wheel to a new position and repeating this procedure.

3. Additional Experimental Details and Results

In this section, we provide additional experimental details and surface reconstruction results.

3.1. Additional Experimental Details

All our experiments are run on a laptop computer with Intel Core i7-8750H processor (2.2GHz) and 16GB memory. Our surface reconstruction algorithm is implemented in Matlab without acceleration. The running time of our reconstruction algorithm is about 20 mins for one pair, 23 mins for two pairs, and 26 mins for three pairs. The execution time does not increase significantly with the number of reciprocal pairs. This is because the pair number affects only the dimension of the coefficient matrix $\mathbf{W}$. The complexity of the graph-cut-based depth optimization does not increase when more pairs are used. Instead, the computation overhead on using more numbers of pairs mainly come from accessing data and the SVD decomposition of $\mathbf{W}$.

Here we also clarify the number of samples we use to compute the mean angular error (MAE) of the normal estimation. This number varies among different target objects. Specifically, in the synthetic experiments, the total number of samples is 29,256 for the Stanford bunny scene, 31,550 for the monk scene, 25,519 for the sphere scene, 103,057 for the dragon scene, 100,782 for the Buddha scene, 87,097 for the warrior scene, and 93,146 for the armadillo scene (results for the last four scenes are shown in Fig. 2). In the real experiments, the total number of samples is 79,793 for the ball scene and 63,911 for the corner scene.

3.2. Additional Synthetic Results

Fig. 2 shows our normal and 3D surface reconstruction results on four additional 3D models. We use different materials from the KAIST pBRDF dataset [1]. We show the rendered model in conventional rendering mode, rendered directional polarization images that emulates the data captured by a polarization camera, our recovered normal map and 3D surface. We evaluate the normal estimation with per-pixel angular error and the mean angular error (MAE) in degree. Here the reconstruction results are computed with two reciprocal pairs.

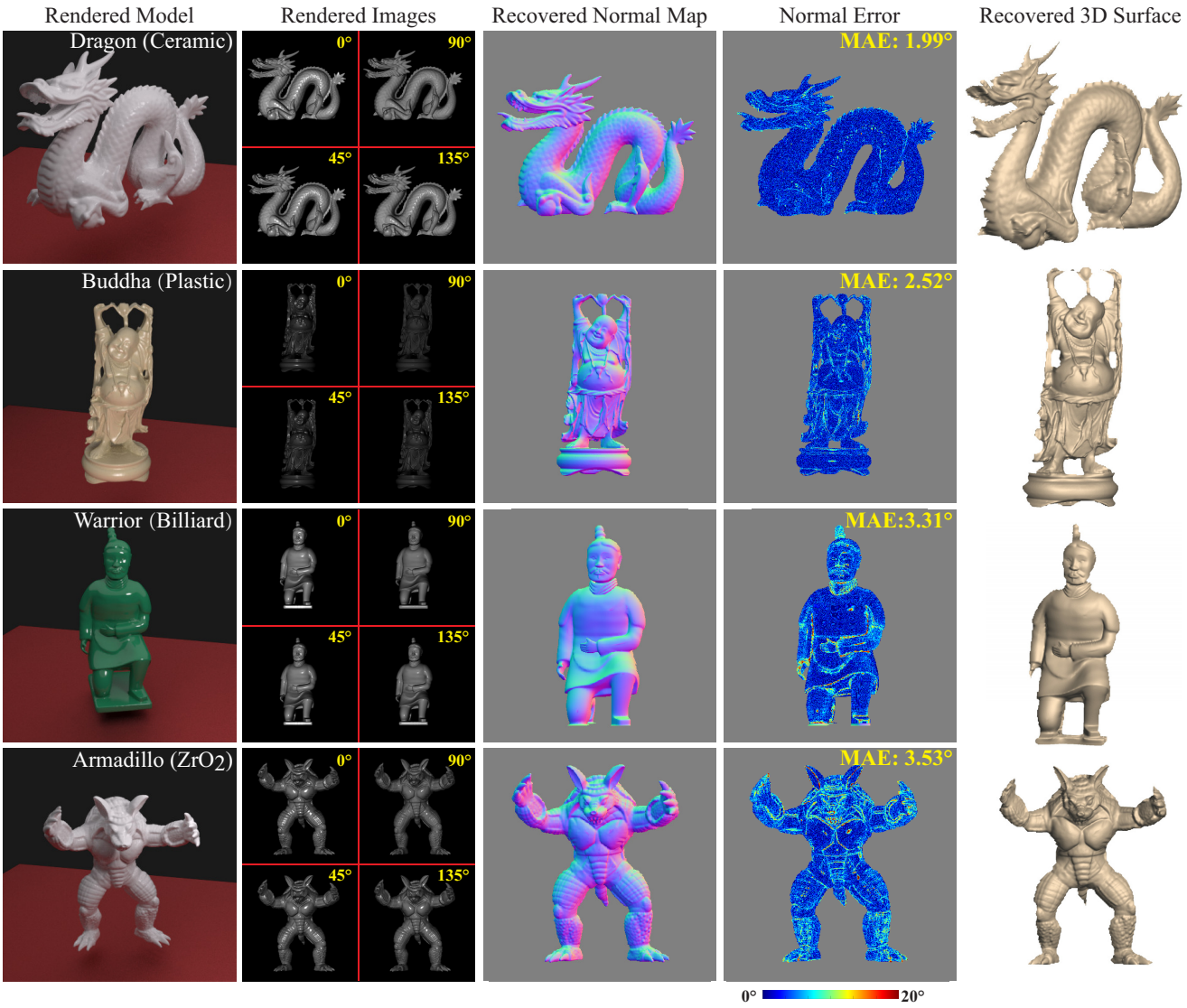

Figure 2. Additional results on synthetic data.

We show additional evaluation of our method with respect to different types of materials in Fig. 3 and Table 1. Here all experiments are performed on a sphere model. We use six materials from the KAIST pBRDF dataset [1]: Spectralon, Chrome,

Gold, Plastic POM, Black Billiard, and Ceramic ZrO_2. We also compare the performance of our method with respect to the number of pairs (one pair vs. two pairs). Fig. 3 shows the cross-sections of our recovered surfaces in comparison with the ground truth. Table 1 reports the MAE of normal estimation for different types of materials.

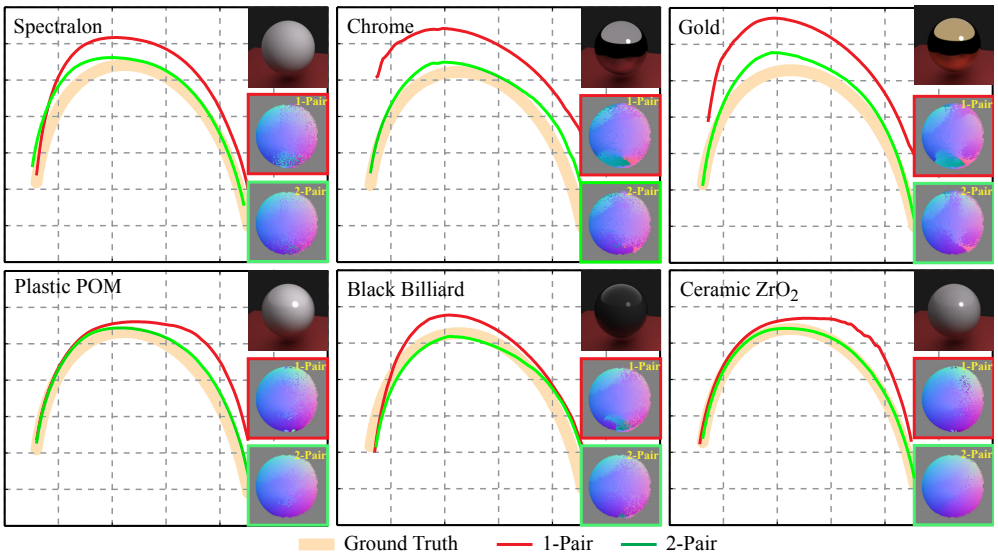

Figure 3. Reconstruction results of our method when using one and two reciprocal pairs for different types of materials. Here we show the cross-sections of reconstructed surfaces in comparison with the ground truth, color images of the sphere rendered with the specific material, and estimated normal maps.

Materials	1-Pair	2-Pair
Spectralon	17.6963	2.2087
Chrome	20.0153	9.4524
Gold	18.0669	7.5412
Black Billiard	18.4281	6.0594
ZrO_2	8.5610	1.0912
Plastic POM	8.8094	0.8978

Table 1. Mean angular error (in degree) of the normal estimations with respect to different types of materials.

3.3. Additional Real Results

Fig. 4 shows our recovered normal and 3D surfaces rendered in two different viewpoints. We also show close-up views of the recovered surfaces to highlight details. The results shown in Fig. 4 are recovered using two reciprocal pairs.

Fig. 5 shows additional examples of our proposed image decomposition. Here we show decomposition results of objects made with different materials. In each example, we show the intensity image (before decomposition), the three composed components (*i.e.*, specular-polarized, diffuse-unpolarized, and diffuse-polarized), as well as the angle of polarization (AoP) computed with the diffuse-polarized component ("diffuse AoP") in comparison with the AoP computed with the image before decomposition ("overall AoP"). We can see that the diffuse AoP computed with diffuse-polarized component better encodes the surface normal. However, this property varies among materials.

Fig. 6 shows comparisons between our one-pair and two-pair reconstruction results. We can see that our one-pair reconstruction results has significant errors when the AoP is close to $90°$; this is where the polarimetric constraint becomes unstable. By comparison, the two-pair results are able to recover fine surface details such as the bird feather and folded paper texture on the dinosaur.

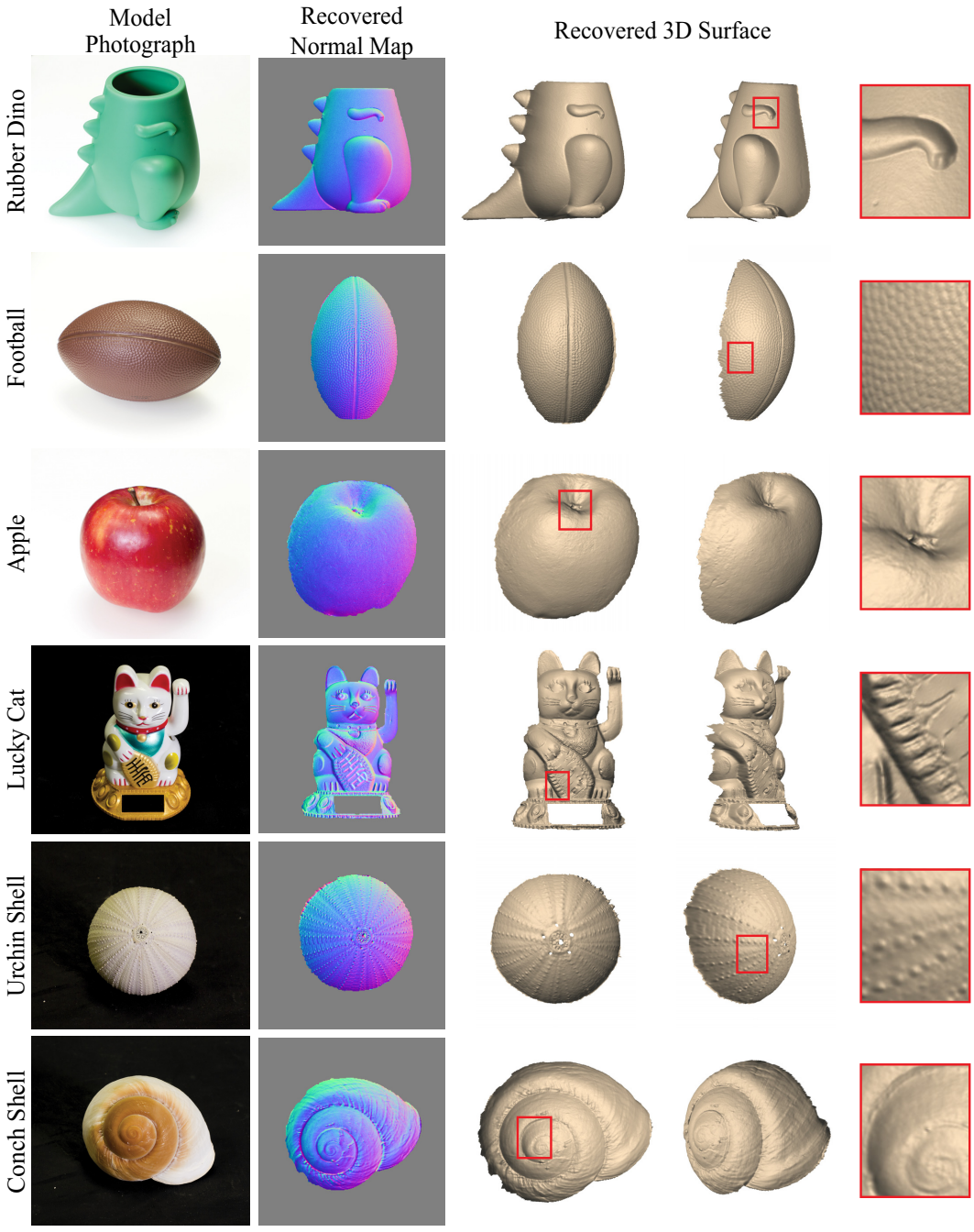

Figure 4. Additional surface reconstruction results on real scenes. Here the results are computed using two reciprocal pairs. From left to right: model photograph, recovered normal map, recovered 3D surface under two different viewpoints, and a close-up view of the recovered 3D surface.

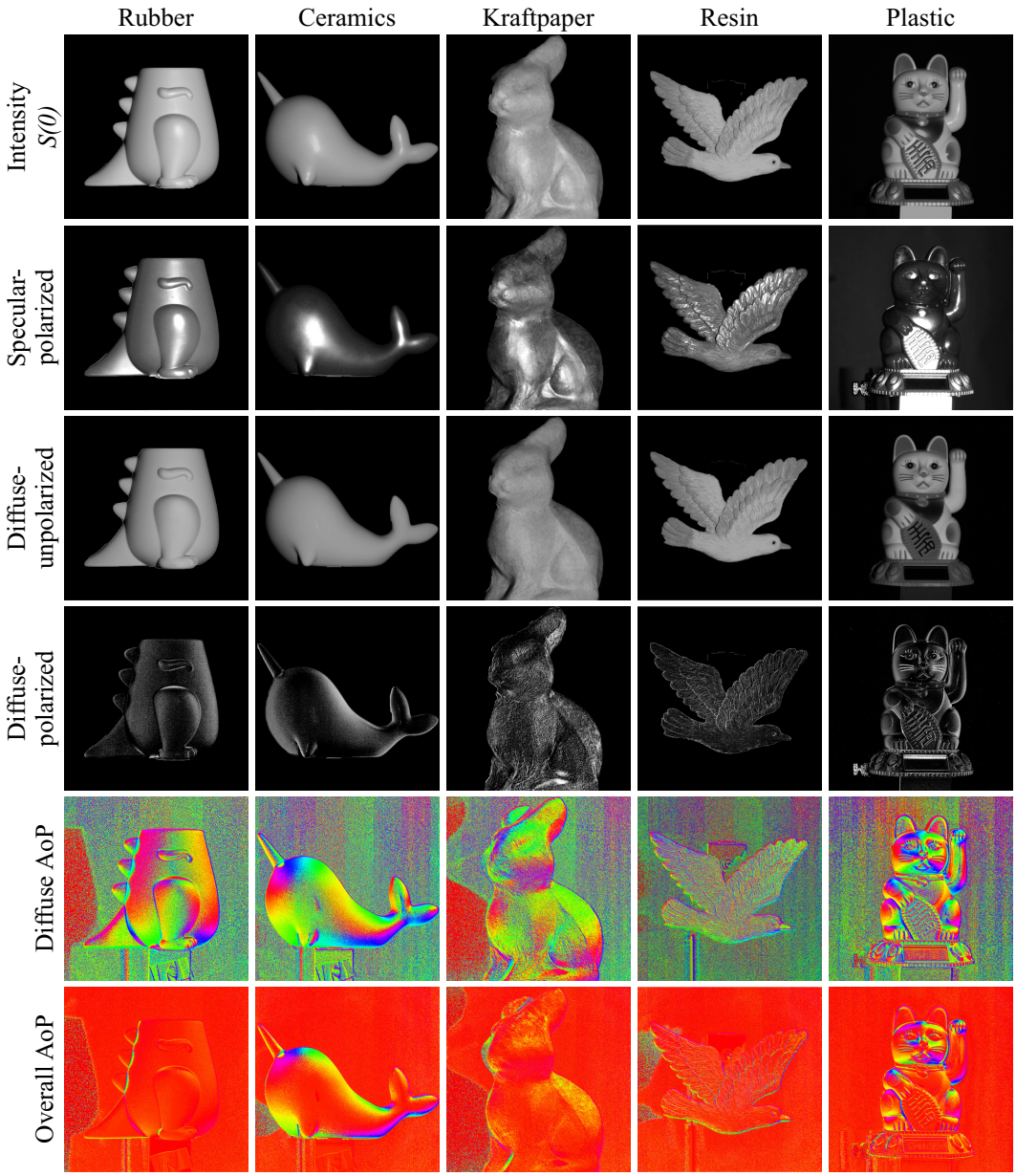

Figure 5. Additional results for our proposed image decomposition.

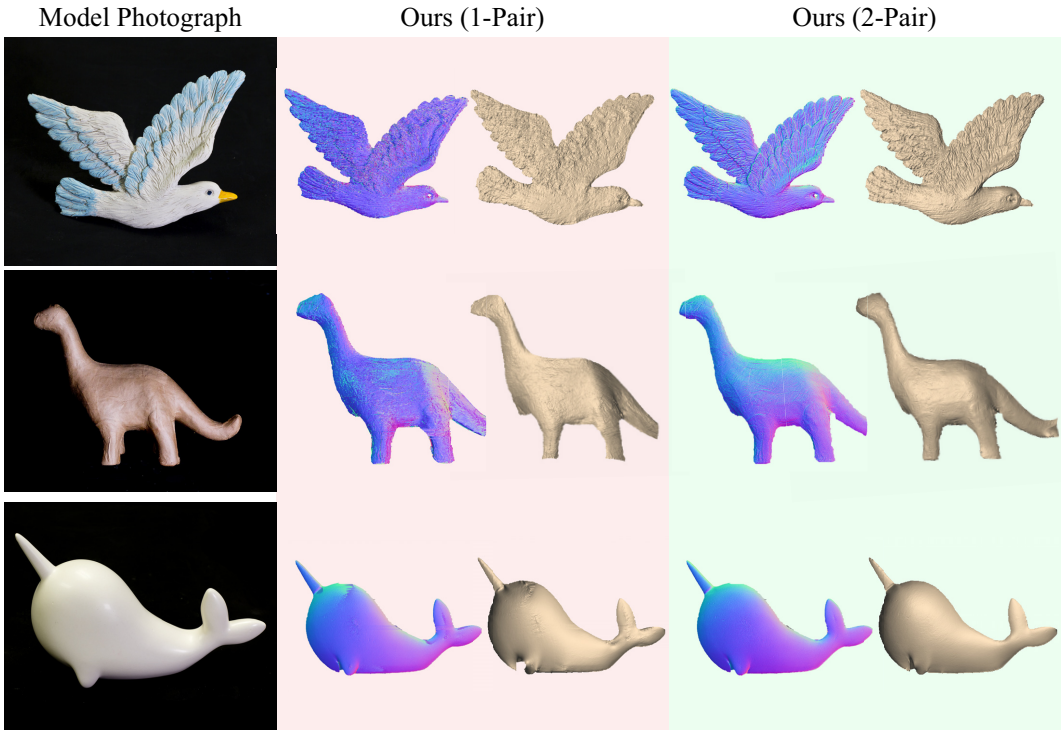

Figure 6. Additional results on comparison between the one-pair and two-pair reconstruction.

4. Failure Examples

Fig. 7 shows two failure examples on transparent and concave surfaces with strong inter-reflection. The toy car windows are made of transparent acrylic glass. Our reconstruction fails at the window glasses because the reflected light is weak. The bowl scene has strong inter-reflection that results in caustic patterns formed on the surface, which our system is unable to handle as well.

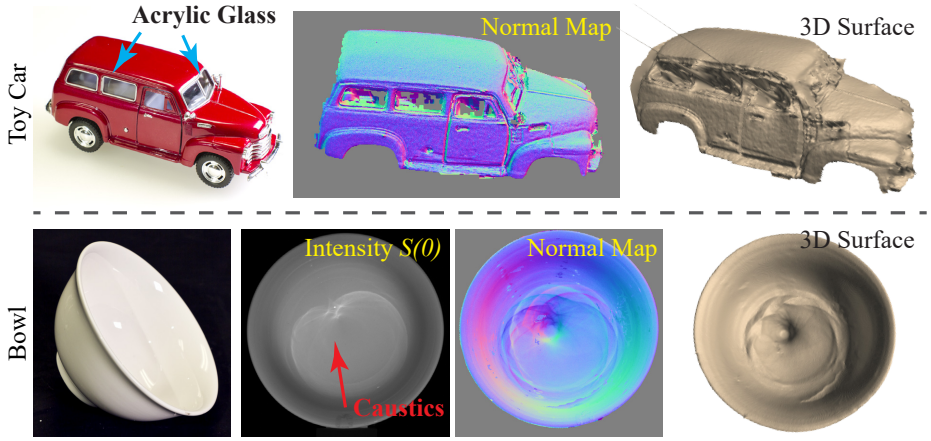

Figure 7. Two failure examples: a toy car (with transparent glass) and a bowl (with strong inter-reflection).

References

[1] Seung-Hwan Baek, Tizian Zeltner, Hyun Jin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H. Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM Transactions on Graphics (Proc SIGGRAPH 2020)*, 39(4), 2020.

[2] Zsolt Jankó, Ondřej Drbohlav, and Radim Šára. Radiometric calibration of a Helmholtz stereo rig. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[3] Raymond A. Morano, Cengizhan Ozturk, Robert Conn, Stephen Dubin, Stanley Zietz, and Jonathan Nissanov. Structured light using pseudorandom codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):322–327, 1998.

[4] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 2012.

[5] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[6] William A. P. Smith, Ravi Ramamoorthi, and Silvia Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *European Conference on Computer Vision (ECCV)*, 2016.

[7] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Multi-view 3d models from single images with a convolutional network. In *European Conference on Computer Vision (ECCV)*, 2016.

[8] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[9] Xiuming Zhang, Zhoutong Zhang, Chengkai Zhang, Joshua B Tenenbaum, William T Freeman, and Jiajun Wu. Learning to Reconstruct Shapes from Unseen Classes. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[10] Todd E. Zickler, Peter N. Belhumeur, and David J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 49(2-3):215–227, 2002.