Supplementary Material for PISR: Polarimetric Neural Implicit Surface Reconstruction for Textureless and Specular Objects

Guangcheng Chen[1,2], Yicheng He[1,2], Li He[1,2], and Hong Zhang⋆[1,2]

[1] Southern University of Science and Technology, Shenzhen, China
[2] Shenzhen Key Laboratory of Robotics and Computer Vision
{chengc2023,12332136}@mail.sustech.edu.cn, {hel, hzhang}@sustech.edu.cn

1 Overview

In this supplementary material, we provide the following elements:

– Basics on DoFP polarization camera (Sec. 2).
– Additional details of our dataset (Sec. 3).
– Signed error maps and normal maps by PISR-ON and PISR-PN (Sec. 4).
– Additional qualitative results by PISR-PN (Sec. 5).

We also provide a video that shows additional qualitative results, including sequences of rendered RGB images and normal maps.

2 Basics on DoFP polarization camera

A division-of-focal-plane (DoFP) polarization camera [8] can measure the intensities at four different polarizer angles in one shot, which is achieved by assigning each pixel a polarizer. As shown in Fig. 1, every 2×2 adjacent pixels have linear polarizers at four different angles 0°,45°,90° and 135° and each 4×4 superpixel can measure intensities in 12 channels including RGB channels. Since there is no circular polarizer on the sensor, the polarization camera can only measure linear polarized light.

From the perspective of a camera, linear polarized light can be described in terms of intensity I, angle of polarization (AoP) φ and degree of polarization (DoP) ρ. For each pixel, the measured intensity I_ϕ of polarized light is a function of the polarizer angle ϕ:

$$\begin{aligned} I_\phi = f_{\text{pol}}(\phi) =& \frac{1}{2}[I + \rho I \cos(2\varphi - 2\phi)] \\ =& \frac{1}{2}(I + \rho I \cos 2\varphi \cos 2\phi + \rho I \sin 2\varphi \sin 2\phi), \end{aligned} \tag{1}$$

⋆ Corresponding author.
The source code is available at https://github.com/GCChen97/PISR

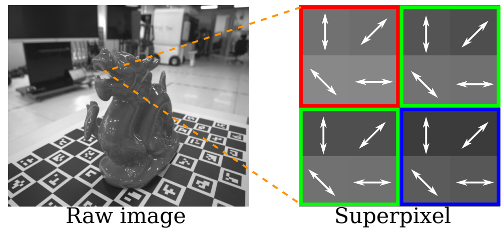

Fig. 1: Polarization image.

where the constant $1/2$ is the attenuation caused by the polarizer [2]. Let $s_0 = I$, $s_1 = \rho I \cos 2\varphi$ and $s_2 = \rho I \sin 2\varphi$, then we can get the Stokes vector $\mathbf{s} = [s_0, s_1, s_2]^T$ at the camera frame for representing linear polarized light.

Denote the four images captured by the polarization camera as I_0, I_{45}, I_{90} and I_{135}. With the four intensity values, we can calculate $\mathbf{s}$ according to Eq. (1) as follows:

$$s_0 = \frac{1}{2}(I_0+I_{45}+I_{90}+I_{135}) = \frac{1}{2}[f_{pol}(0)+f_{pol}(\frac{1}{4}\pi)+f_{pol}(\frac{2}{4}\pi)+f_{pol}(\frac{3}{4}\pi)] = I \quad (2)$$

$$s_1 = I_0 - I_{90} = f_{pol}(0) - f_{pol}(\frac{2}{4}\pi) = \rho I \cos 2\varphi \quad (3)$$

$$s_2 = I_{45} - I_{135} = f_{pol}(\frac{1}{4}\pi) - f_{pol}(\frac{3}{4}\pi) = \rho I \sin 2\varphi \quad (4)$$

Therefore, we can easily calculate φ and ρ with $\mathbf{s}$ according to the above equations:

$$\varphi = \frac{1}{2}\text{arctan2}(s_2, s_1), \quad \rho = \frac{\sqrt{s_1^2 + s_2^2}}{s_0}. \quad (5)$$

3 Additional details of our dataset

We capture images of each object at two (three for Figure) different heights with a color polarization camera [8]. At each height, the camera is moved around the object to capture images from 20 different viewpoints, resulting in 40 (60 for Figure) polarization images for each object. We show images of two viewpoints of each object in Fig. 3. Each object is placed on a table indoors under uncontrolled lighting conditions.

Since geometry reconstruction is more sensitive to the accuracy of camera calibration than novel view synthesis, we calibrate the camera and use SuperGlue [6] and COLMAP [7] to achieve robust and accurate sparse reconstruction given camera intrinsics. We additionally use a ChAruco board to provide key points for feature extraction and matching. An example of a sparse reconstruction result is shown in Fig. 2.

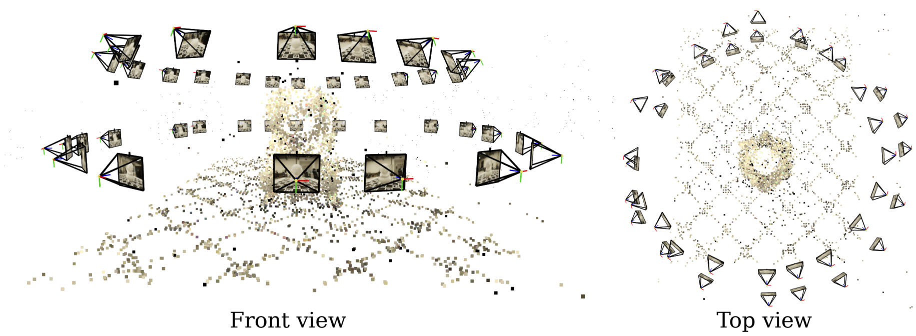

Fig. 2: The sparse reconstruction result of Standing Rabbit

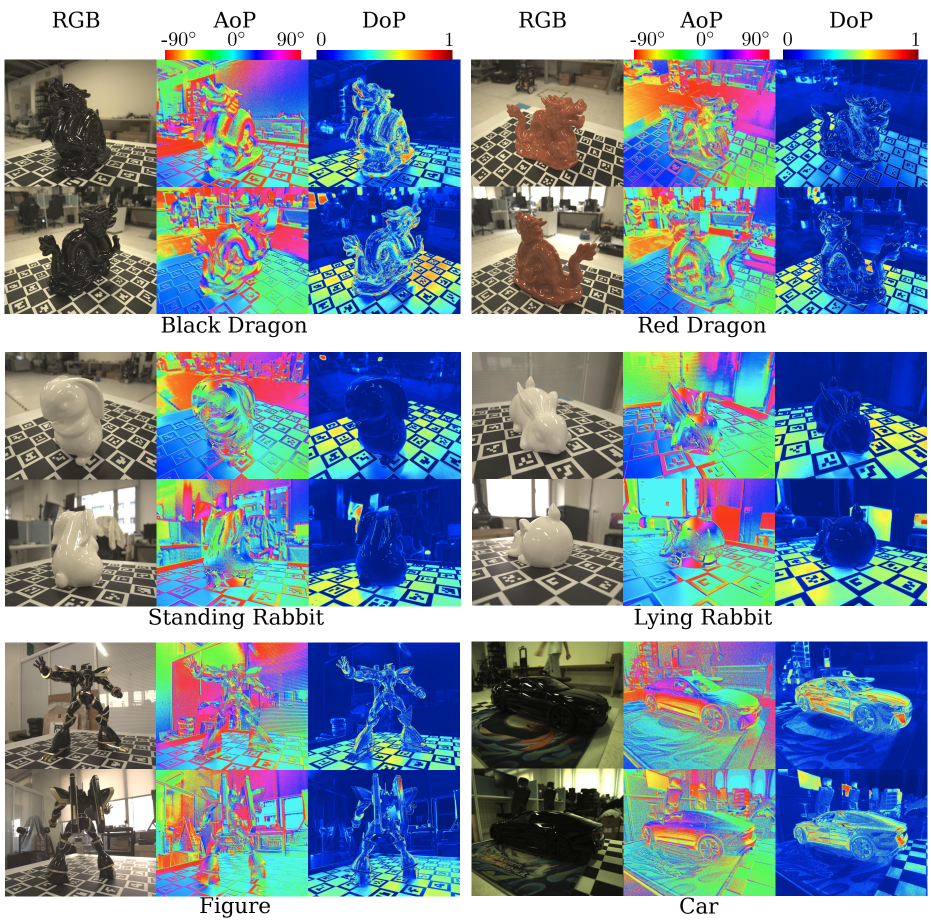

Fig. 3: Examples of RGB images, AoP maps and DoP maps of our dataset

4 Signed error maps and normal maps

As shown in Fig. 4, there are more artifacts on the reconstructed meshes by PISR-ON. PISR-ON and PISR-PN are based on the orthographic and perspective polarimetric constraints, respectively. The orthographic constraint is based on the assumption of orthographic projection, which is inconsistent with the commonly used pinhole camera model, and thus can lead to errors in the reconstruction results.

The signed error maps present the directed distances from the reconstructed points to the ground-truth points. Let $\mathcal{R}$ be the vertex set of the reconstructed mesh and $\mathcal{G}$ be the ground truth. For a vertex $\mathbf{r} \in \mathcal{R}$, its signed error is defined as:

$$e_{r \to \mathcal{G}} = \operatorname{sgn}\Big((\mathbf{r} - \mathbf{g})^T \mathbf{n}_\mathbf{g}\Big) \cdot \min_{\mathbf{g} \in \mathcal{G}} \|\mathbf{r} - \mathbf{g}\| \tag{6}$$

where $\operatorname{sgn}(\cdot)$ is the sign function that gets the sign of its argument and $\mathbf{n}_\mathbf{g}$ is the ground-truth normal.

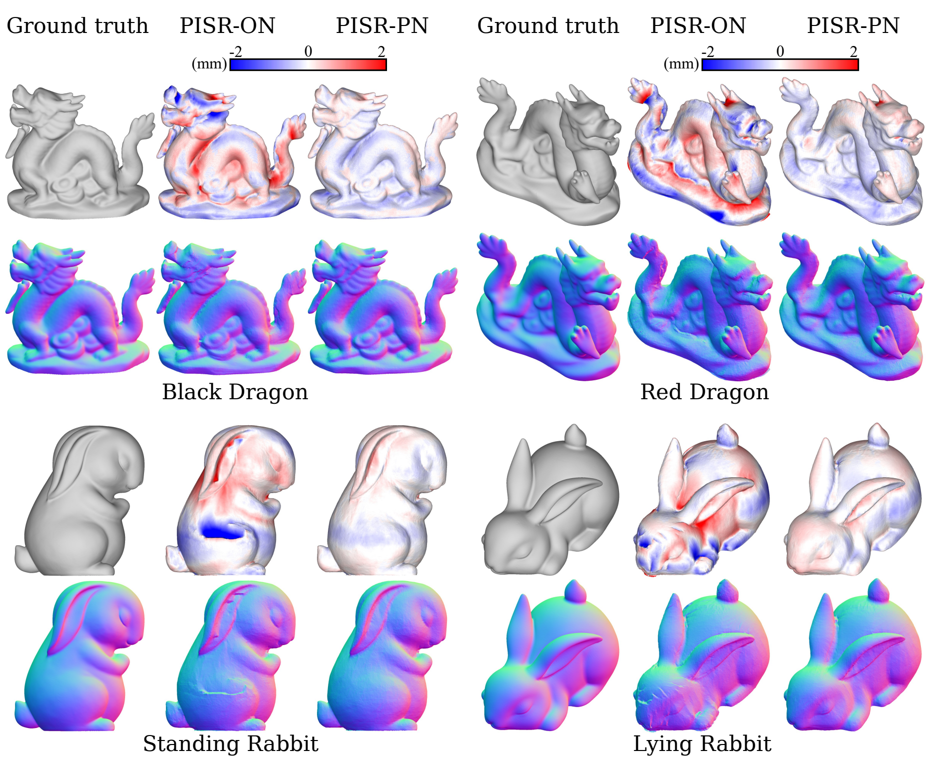

Fig. 4: Signed error maps and normal maps of the reconstruction results by PISR-ON and PISR-PN (zoom in for a better view). For the signed error maps in the first row of each object, higher color saturation means higher errors and the errors are truncated to within ± 2 mm.

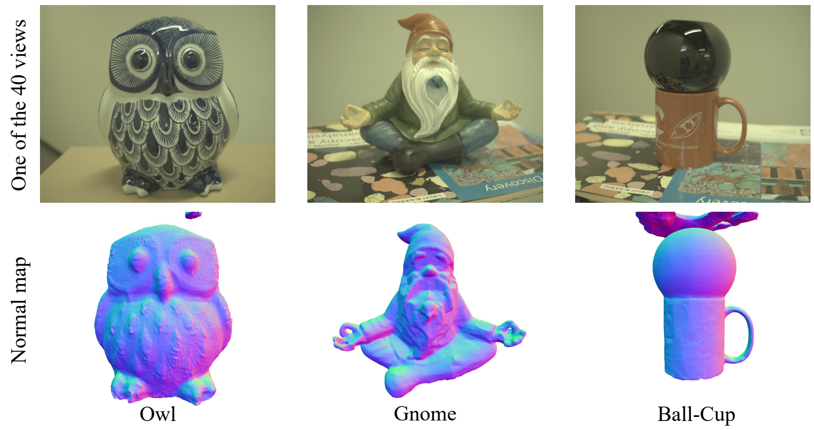

Fig. 5: Qualitative results of the PANDORA dataset.

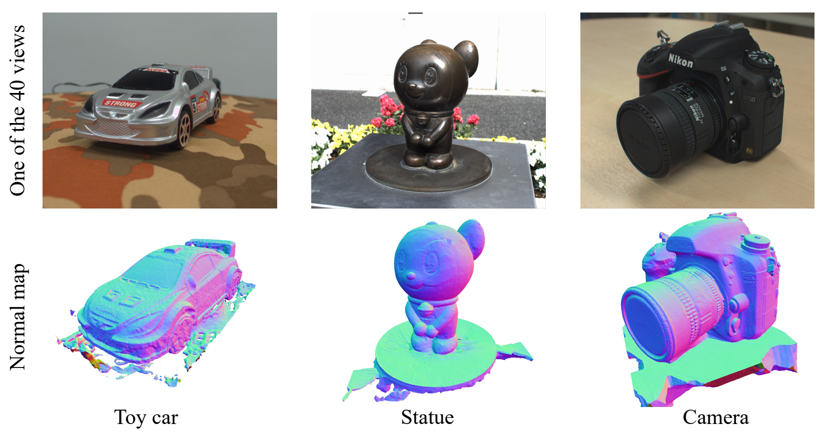

Fig. 6: Qualitative results of the PMVIR dataset.

5 Additional qualitative results

We present qualitative results on the PANDORA [1] and PMVIR [9] datasets in Fig. Fig. 5 and Fig. Fig. 6. Due to the imbalanced convergence between the foreground and background [4], PISR could fail, particularly when the camera's focal length is longer, as is the case with these two datasets. Therefore, we add 2D mask loss as work [1] to help PISR separate the foreground and background.

Since PISR uses demosaiced RGB images and AoP maps for shape estimation, the noise on texture edges caused by the demosaicing process could lead to incorrect bumpiness on the reconstructed surfaces of objects that are supposed to be smooth. Using raw sensor data as RawNeRF [5] and NeISF [3] without demosaicing might bypass this problem.

References

1. Dave, A., Zhao, Y., Veeraraghavan, A.: Pandora: Polarization-aided neural decomposition of radiance. In: European Conference on Computer Vision. pp. 538–556. Springer (2022)
2. Jakob, W., Speierer, S., Roussel, N., Nimier-David, M., Vicini, D., Zeltner, T., Nicolet, B., Crespo, M., Leroy, V., Zhang, Z.: Mitsuba 3 renderer (2022), https://mitsuba-renderer.org
3. Li, C., Ono, T., Uemori, T., Mihara, H., Gatto, A., Nagahara, H., Moriuchi, Y.: Neisf: Neural incident stokes field for geometry and material estimation. arXiv preprint arXiv:2311.13187 (2023)
4. Liu, Y., Wang, P., Lin, C., Long, X., Wang, J., Liu, L., Komura, T., Wang, W.: Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. In: SIGGRAPH (2023)
5. Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., Barron, J.T.: NeRF in the dark: High dynamic range view synthesis from noisy raw images. CVPR (2022)
6. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: Learning feature matching with graph neural networks. In: CVPR (2020)
7. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
8. Sony: Polarization image sensor with four-directional on-chip polarizer and global shutter function, https://www.sony-semicon.co.jp/e/products/IS/industry/product/polarization.html, accessed 2024-02-04
9. Zhao, J., Monno, Y., Okutomi, M.: Polarimetric multi-view inverse rendering. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)