Deep Shape from Polarization

Yunhao Ba[1], Alex Gilbert[1⋆], Franklin Wang[1⋆], Jinfa Yang[2], Rui Chen[2], Yiqin Wang[1], Lei Yan[2], Boxin Shi[2⋆⋆], and Achuta Kadambi[1⋆⋆]

[1] University of California, Los Angeles
{yhba, alexrgilbert}@ucla.edu, franklinxzw@gmail.com, achuta@ee.ucla.edu
[2] Peking University
{jinfayang, shiboxin}@pku.edu.cn

Supplementary Contents

This supplement is organized as follows:

- Section 1 elaborates on the physics used to calculate the priors described in the main paper.
- Section 2.1 discusses runtime details (cross-referenced to Section 1 of the main paper).
- Section 3.1 expands on texture copy (cross-referenced to Section 5.3 of the main paper).
- Section 3.2 includes further comparisons for lighting invariance (cross-referenced to Section 5.2 of the main paper).
- Section 3.3 includes additional results that did not fit in the main paper (cross-referenced to Section 5.6 of the main paper).

1 Physics Priors

In section 3.1 of the main paper, we present the image formation model, which relates the measured intensity of polarized light at a given pixel to two constants describing polarization: the degree of polarization (DoP), ρ, and the polarization phase angle, ϕ. These constants are in turn related to the azimuth and zenith angle of the surface normal. While this relationship depends on the reflection type (i.e., diffuse dominant or specular dominant) and contains inherent ambiguities, we show in the paper that giving our network the possible normal estimates as *physics priors* guides it toward more geometrically realistic solutions. Below we present in detail the physics behind our method for obtaining the priors. Ultimately we feed our network three normals at each pixel (concatenated channel-wise): two solutions to the specular model, $\boldsymbol{N}_{spec1}$ and $\boldsymbol{N}_{spec2}$, and one solution to the diffuse model, $\boldsymbol{N}_{diff}$.

⋆ Equal contribution.
⋆⋆ Corresponding authors.

1.1 Solving for DoP and phase angle

Recall from the main paper that for a given pixel

$$I(\phi_{pol}) = \frac{I_{max} + I_{min}}{2} + \frac{I_{max} - I_{min}}{2}\cos(2(\phi_{pol} - \phi)), \tag{1}$$

where ϕ_{pol} is the angle of the polarizer, and I_{max} and I_{min} are the upper and lower bounds on $I(\phi_{pol})$ respectively. We also introduced ρ as

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2}$$

We define the *unpolarized intensity* of the pixel as the average of the upper and lower bounds,

$$I_{un} = \frac{I_{max} + I_{min}}{2}. \tag{3}$$

Using the trigonometric identity $cos(A - B) = cos(A)cos(B) + sin(A)sin(B)$, the cosine term can be expanded, and equation 1 can be rewritten as

$$I(\phi_{pol}) = I_{un} + I_{un}\rho\cos(2\phi_{pol})\cos(2\phi) + I_{un}\rho\sin(2\phi_{pol})\sin(2\phi). \tag{4}$$

Applying a change of variables as done in [6], we can create a per-pixel system of equations linear in ρ and ϕ,

$$\begin{bmatrix} 1 & cos(2\phi_{pol,1}) & sin(2\phi_{pol,1}) \\ & \vdots & \\ 1 & cos(2\phi_{pol,m}) & sin(2\phi_{pol,m}) \end{bmatrix} \begin{bmatrix} I_{un} \\ a \\ b \end{bmatrix} = \begin{bmatrix} I_{\phi_{pol,1}} \\ \vdots \\ I_{\phi_{pol,m}} \end{bmatrix}, \tag{5}$$

where $a = I_{un} \cdot \rho cos(2\phi)$ and $b = I_{un} \cdot \rho sin(2\phi)$, and, without loss of generality, the number of polarizer angles m is set to 4, with $\phi_{pol,i} \in \{0°, 45°, 90°, 135°, \}$. These four polarization angles can be captured in a single shot by our polarization camera [4].

This system of equations can be solved with least squares. We utilize MatLab's built in `mldivide` function to find the least squares solution for I_{un}, a, and b.

Finally we estimate

$$\rho = \frac{\sqrt{a^2 + b^2}}{I_{un}} \qquad \phi = \frac{1}{2}atan2(b, a). \tag{6}$$

Note the *ambiguity problem* here: both ϕ and $\phi + \pi$ satisfy equation 1. We find that using the three solutions for *both* of the possible phase angles (e.g., six total solutions per pixel) provides minimal improvement. We therefore use only the solutions corresponding to one resolution of the ambiguity. We also try to emphasise high frequency detail in the priors by adding a scale factor of two to the ϕ estimates. Both estimates yield similar performance, and we keep the scale factor of 2 in the main paper and supplement.

1.2 Specular Solution

Consider the case where the pixel is dominated by specular reflections (i.e., light from a single external refraction with the surface).

Azimuth angle Let the "incidence plane" be defined as that containing both the light source and the camera, such that the azimuth angle of the surface normal will be equal to the angle of the incidence plane.

The Fresnel reflectance coefficients give the fraction of incident power reflected for light polarized parallel to the incidence plane ($F_\parallel$) and perpendicular to the incidence plane ($F_\perp$). For purely specular reflection, the authors of [3] show that I_{max} and I_{min} can be written in terms of the Fresnel coefficients as

$$I_{max} = \frac{F_\perp}{F_\parallel + F_\perp} I \qquad I_{min} = \frac{F_\parallel}{F_\parallel + F_\perp} I, \tag{7}$$

where I is the measured intensity of specular reflection. By these equations, the minimum intensity occurs parallel to the plane of incidence, i.e. when the polarization angle, ϕ_{pol}, matches the azimuth angle of the surface normal, φ. But by equation 1, intensity is minimized when ϕ_{pol} and the phase angle, ϕ, are $90°$ apart. Therefore, in the specular case

$$\varphi = \phi + \frac{\pi}{2}. \tag{8}$$

Note, as mentioned in the previous section, $\phi - \frac{\pi}{2}$ would also satisfy the equivalence, but we only include one set of solutions as our priors.

Zenith angle Using Snell's law and the Fresnel equations, it is shown in [3] that the Fresnel coefficients can be written in terms of the surface normal's zenith angle, θ, and the refractive index of the surface, n, as

$$F_\parallel = \frac{1 + n^2 - (\frac{n^2+1}{n^2}) sin^2\theta - 2cos\theta\sqrt{n^2 - sin^2\theta}}{1 + n^2 - (\frac{n^2+1}{n^2}) sin^2\theta + 2cos\theta\sqrt{n^2 - sin^2\theta}} \tag{9}$$

and

$$F_\perp = \frac{1 + n^2 - 2sin^2\theta - 2cos\theta\sqrt{n^2 - sin^2\theta}}{1 + n^2 - 2sin^2\theta + 2cos\theta\sqrt{n^2 - sin^2\theta}}. \tag{10}$$

Combining equations 2, 7, 9, and 10, [3] show that

$$\rho = \frac{2\sin^2\theta\cos\theta\sqrt{n^2 - \sin^2\theta}}{n^2 - \sin^2\theta - n^2\sin^2\theta + 2\sin^4\theta}. \tag{11}$$

For our dataset, n is unknown, but according to [1], the dependency of ρ on n is weak and generally falls between 1.3 and 1.6 for most dielectric materials. We use $n = 1.5$, as stated in the main paper.

Equation 11 cannot be inverted analytically, and solving for the zenith angle using numerical interpolation produces two solutions, $\theta_{spec,1}$ and $\theta_{spec,2}$. With the azimuth and zenith angles determined, the two normal solutions, $\boldsymbol{N}_{spec,1}$ and $\boldsymbol{N}_{spec,2}$, are acquired by converting from spherical to Cartesian coordinates with

$$N_{spec,1} = \begin{pmatrix} cos(\varphi)sin(\theta_{spec,1}) \\ sin(\varphi)sin(\theta_{spec,1}) \\ cos(\theta_{spec,1}) \end{pmatrix} \tag{12}$$

and

$$N_{spec,2} = \begin{pmatrix} cos(\varphi)sin(\theta_{spec,2}) \\ sin(\varphi)sin(\theta_{spec,2}) \\ cos(\theta_{spec,2}) \end{pmatrix} \tag{13}$$

1.3 Diffuse Solution

Consider the case where the pixel is dominated by diffuse reflections, arising from light waves which undergo multiple internal refractions beneath the material surface before refracting out.

Azimuth angle Whereas for the specular case the phase angle, ϕ, and azimuth angle, φ, are 90° apart, in the diffuse case

$$\varphi = \phi. \tag{14}$$

The authors explain in [7] that this 90° phase difference between specular and diffuse reflections arises from the fact that they come from complementary processes (reflection and transmission, respectively). While for specular reflection $F_\perp \geq F_\parallel$ causes the maximum transmitted radiance orthogonal to the incidence plane, for specular reflection $(1-F_\perp) \leq (1-F_\parallel)$ implies the maximum transmitted radiance will be parallel to the incidence plane. The azimuth ambiguity also exists in the diffuse case, but we choose a single azimuth solution for calculating the normals.

Zenith angle Due to the complementary nature mentioned above, it is shown in [7] that for the diffuse case

$$I_{max} = \frac{1-F_\parallel}{2-F_\parallel-F_\perp} I \qquad I_{min} = \frac{1-F_\perp}{2-F_\parallel+F_\perp} I. \tag{15}$$

Using equations 2, 15, along with Snell's law and the Fresnel equations, in [1] they show that in the diffuse case, ρ can also be written in terms of n and θ as

$$\rho = \frac{(n-\frac{1}{n})^2 \sin^2\theta}{2+2n^2-(n+\frac{1}{n})^2\sin^2\theta+4\cos\theta\sqrt{n^2-\sin^2\theta}}. \tag{16}$$

Equation 16 can be rearranged to obtain a close-form estimation of the zenith angle, θ.

Finally, we find the normal solution, $\boldsymbol{N}_{diff}$, by converting from spherical to Cartesian coordinates as done in equations 12 and 13.

2 Operational Details

Below we present details of our exact method for estimating surface normals during test-time, as well as a high level analysis of the runtime of our full pipeline. We recognize realtime performance as a direction for future study, and consider some possibilities for optimization in our current method.

2.1 Runtime Analysis

The process of surface normal reconstruction consists of two main steps: solving for the *physics priors*, followed by the use of our network for final inference. During inference, we split each image into 256×256 patches and rejoin the estimated normal patches from our network output. The final prediction is the average of 32 cropped predictions with shifted input to preserve the accuracy at patch boundaries. The process used for calculating the priors is described in detail in section 1. The runtime (averaged across all the objects in our test set) is 17.9 seconds for prior calculation and 8.43 seconds for inference. Prior Calculation times are measured using Matlab's `timeit` function. Network Inference times are measured using the profiler built into PyTorch's `autograd` package.

Tests were conducted on a Linux (Ubuntu 18.04 LTS) machine with an Intel i5-8400 CPU (6 cores, 2.80 GHz), 16GB of RAM, and NVIDIA GeForce RTX 2070 GPU (8 GB of GPU RAM). Prior calculation was performed solely on the CPU, while the network inference step was executed on the GPU. Timing results are averaged over our entire testing set.

Making our approach real time is an interesting avenue for future work. We currently do little optimization of the prior calculation process. Exploiting parallelism, sparse solvers, and GPU computation offer opportunities for vast improvement in execution time.

3 Additional Results

In the following section we present various additional results to further reinforce the claims we make in the main paper. Specifically we address the following points:

- robustness to *texture copy*;
- invariance to changes in lighting conditions; and
- overall qualitative improvement over previous purely physics-based methods.

3.1 Suppression of Texture Copy

In Figure 1, we provide a more comprehensive set of examples demonstrating the ability of priors to encourage more geometrically realistic reconstructions. Specifically we show four more examples with significant specular highlights (the Father Christmas object and the vase under outdoor overcast lighting, and

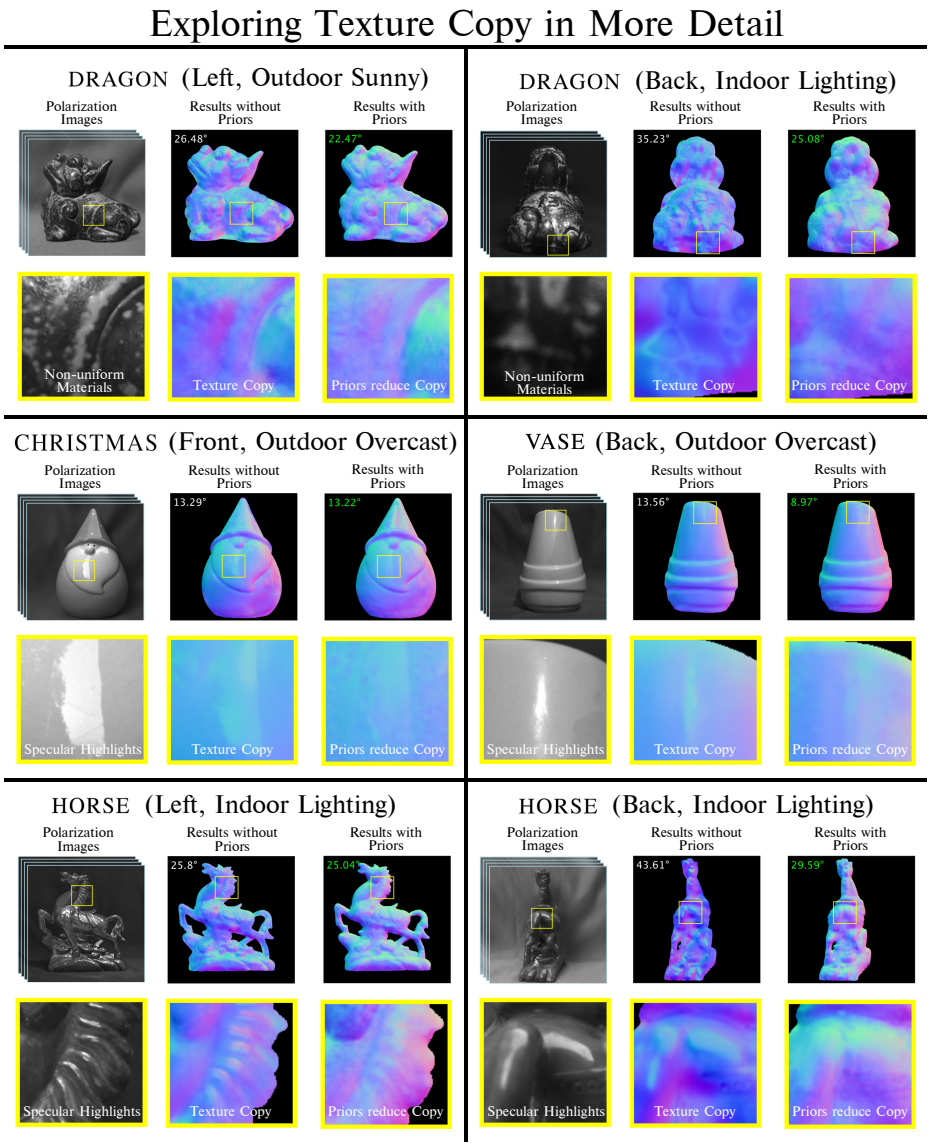

Fig. 1. Additional examples of texture copy suppression. The orientation of the object and lighting conditions are listed next to the name of the object. The number on the top left corner of the whole reconstructions is the MAE. This is further proof that by incorporating the geometry-based information from the physics priors, the proposed method is able to suppress texture copy artifacts related to specular highlights and material variance.

the HORSE under indoor lighting from two different directions), as well as two examples on the DRAGON object, which exhibits sharp material variation. In each example, we focus on a particularly problematic facet, and see that incorporating priors suppresses the copy of image texture (due to either highlights or non-uniform material) in the estimated geometry. Indeed, this improvement is not a one-off phenomena specific to the example provided in the main paper, but is observed throughout our testing set.

The robustness of our model with physics priors to texture copy is still not perfect, and for several cases texture artifacts remain faintly visible. We still consider our model to be strong in this category for the following reasons. Firstly, in the occasional cases when copies are not perfectly erased, they are obviously suppressed, as is apparent in Figure 1 below. Second, our performance in tricky regions is qualitatively better than that of all previous physics-based methods.

3.2 Qualitative Lighting Invariance

In the main paper, we provided quantitative evidence of our method's consistency across different lighting conditions over the test set. In Figure 2 we provide qualitative evidence that reinforces why we consider our method strongly lighting invariant. Furthermore, these qualitative observations expose the variation of other methods under different lightings, which may have been less apparent solely from the mean angular error (MAE) values.

We show two objects in the test set, one simple (CHRISTMAS), and one complex (DRAGON). For these objects, we see that the three physics-based methods [2,3,5] show variations qualitatively in their reconstructions across different lightings. Note, our method requires no known or estimated lighting to achieve this invariance.

3.3 More Qualitative Comparisons with Previous Methods

As described in section 4.1 of the main paper, for each object in our dataset, we capture 12 separate configurations: 4 object orientations under each of 3 lighting conditions. As further evidence of the superior performance of our method relative to previous, purely physics-based approaches, we present another set of results from our test set under various different orientations and lighting.

Figure 3 is organized similarly to Figure 7 in the main paper. The RGB images in (row a) are not used as input, but are shown in the top row of the figure for context about material properties. The input to all the methods shown is four polarization images, shown in (row b) of Figure 7. The ground truth shape is shown in (row c), and corresponding shape reconstructions for the proposed method are shown in (row d). Comparison methods are shown in (row e) through (row g).

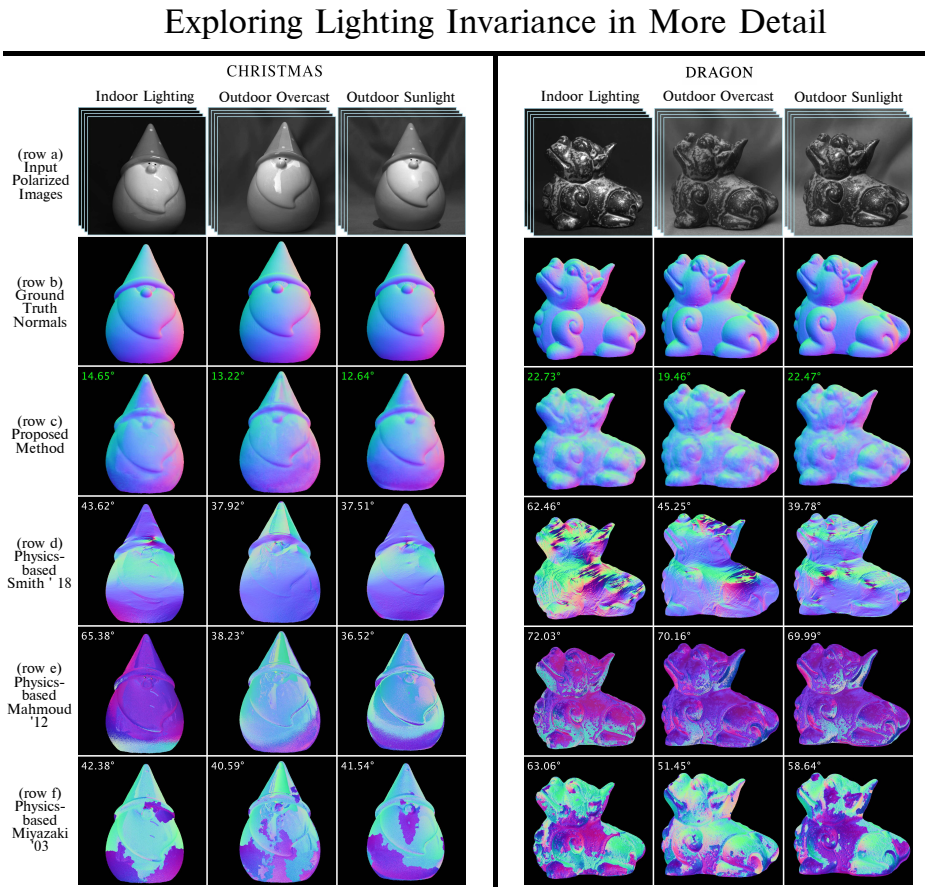

Fig. 2. Qualitative results on lighting invariance show that the proposed method's performance is better. The number shown on the top left corner of the objects is the MAE. We see that for both the CHRISTMAS object and DRAGON objects, the proposed method shows less variations between reconstructions under differing lighting conditions.

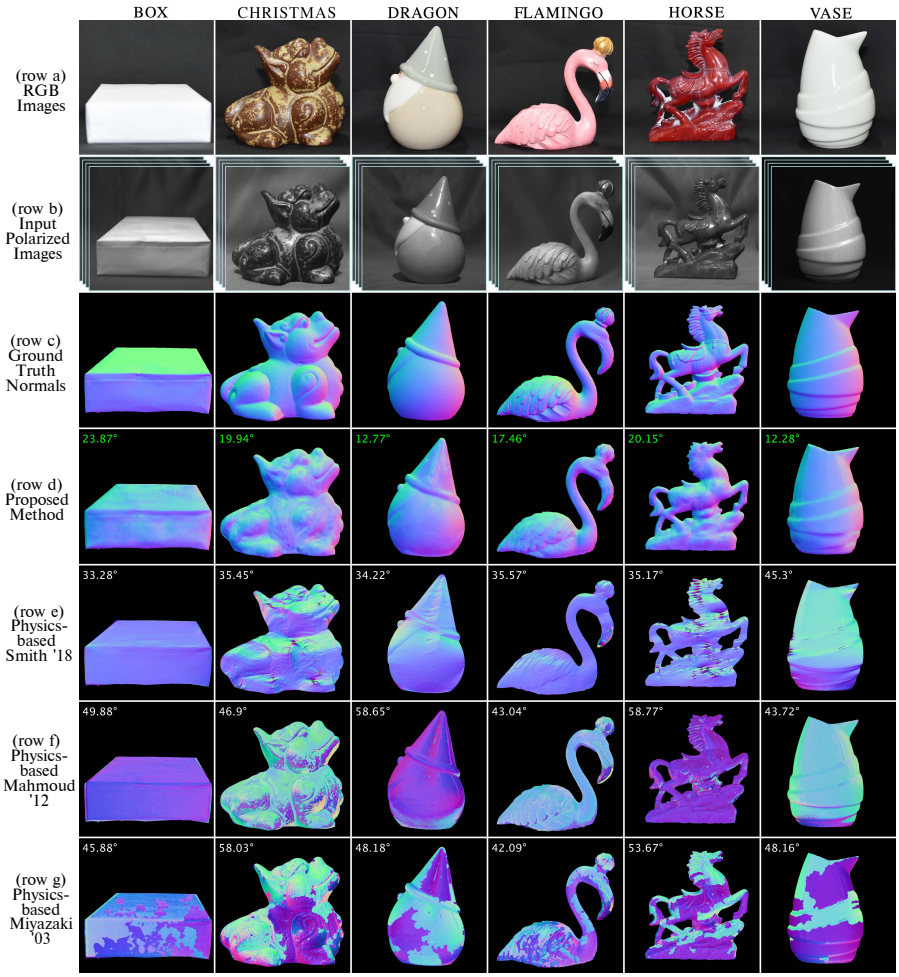

Fig. 3. Proposed method shows qualitative and quantitative improvements for additional lightings and orientations. The corresponding MAE is shown on the top left corner of each reconstruction. The lowest MAE for each object is highlighted in green. Our method (row d) achieves the best performance across all objects both quantitatively and qualitatively.

References

1. Atkinson, G.A., Hancock, E.R.: Recovery of surface orientation from diffuse polarization. IEEE TIP (2006)
2. Mahmoud, A.H., El-Melegy, M.T., Farag, A.A.: Direct method for shape recovery from polarization and shading. ICIP (2012)
3. Miyazaki, D., Tan, R.T., Hara, K., Ikeuchi, K.: Polarization-based inverse rendering from a single view. ICCV (2003)
4. PolarM polarization camera: http://www.4dtechnology.com/products/polarimeters/polarcam/ (2017)
5. Smith, W.A.P., Ramamoorthi, R., Tozza, S.: Height-from-polarisation with unknown lighting or albedo. PAMI (2018)
6. Tozza, S., Smith, W.A.P., Zhu, D., Ramamoorthi, R., Hancock, E.R.: Linear differential constraints for photo-polarimetric height estimation. ICCV (2017)
7. Wolff, L.B., Boult, T.E.: Constraining object features using a polarization reflectance model. IEEE Transactions on Pattern Analysis and Machine Intelligence (1991)