Transparent Shape from a Single View Polarization Image

Supplement Material

Mingqi Shao Chongkun Xia Zhendong Yang Junnan Huang Xueqian Wang*
Tsinghua Shenzhen International Graduate School
{smq21, yangzd21, hjn21}@mails.tsinghua.edu.cn {xiachongkun, wang.xq}@sz.tsinghua.edu.cn

In the supplementary material, we discuss the periodicity of physics-based prior confidence and the robustness to the refractive index in Addition Results. Then we provide a detailed analysis of the noise in the AoLP map as a supplement to the main text. We also demonstrate the calculation of the physics-based prior used in our paper. Finally, we provide details about our dataset and network architecture.

1. Additional Results

1.1. Abalation Study on the Periodicity of Physics-based Confidence

As mentioned in the main text, the physics-based confidence is not only low in the area with high transmittance, but also at the junction of 0 and π in the AoLP map. This situation is caused by the periodicity of AoLP and it is allowed since the lower confidence of the physics-based prior in such area is helpful to avoid the network being misguided by the jumped physics-based prior. Here we conduct the ablation study on this to support our argument.

To keep the confidence of the areas with junction from being suppressed, we re-calculate the confidence map on the AoLP value φ' after the following periodic modulation:

$$\varphi' = \begin{cases} 2\varphi & 0 \leq \varphi < \frac{\pi}{2} \\ 2(\pi - \varphi) & \frac{\pi}{2} \leq \varphi \leq \pi \end{cases} \tag{1}$$

where φ is the original captured AoLP value.

Fig.1 shows the confidence maps under different definitions. It can be seen that the confidence value will not be suppressed in the areas with junction between 0 and π after periodic modulation. We conduct the experiment on the two confidence maps. As shown in Fig.2, allowing the areas with junction have larger confidence will result in the worse performance. The normals provided by the physics-based prior are discontinuous in such areas, and the jumping information brings difficulty for the neural network which is inclined to learn low-frequency information. This argument is also supported by the quantitative results in Table 1.

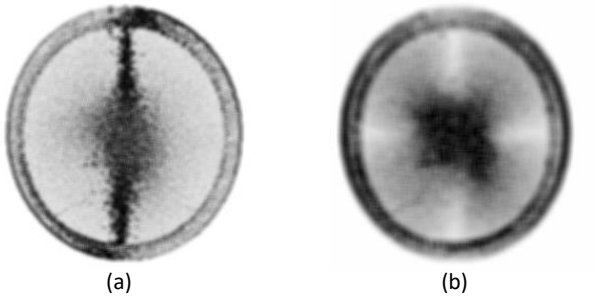

Figure 1. **Physics-based prior confidence maps under different definitions.** (a) confidence map without period definition(used in our main text). (b) confidence map with period definition

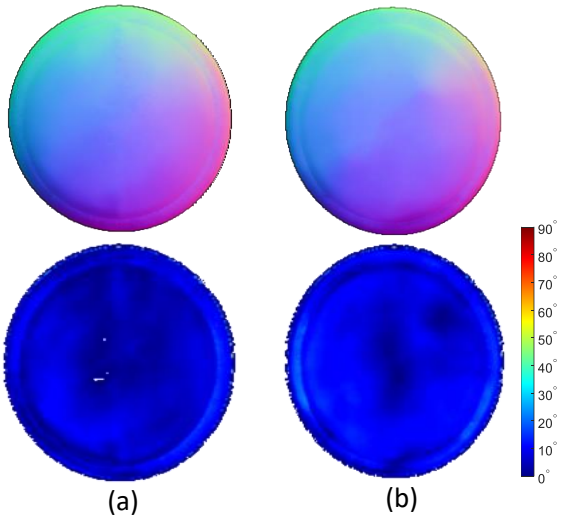

Figure 2. **Network performance with different physics-based prior confidence denitions** (a) network performance without period definition. (b) network performance with period definition

*Corresponding author

Table 1. Network performance with different physics-based prior confidence denitions

Studied Module	Angular Error		Accuracy		
	Mean	Median	11.25	22.5	30
w periodicity	16.69°	13.13°	47.78%	82.17%	92.45%
w/o periodicity(Ours)	16.29°	12.85°	48.31%	83.20%	93.90%

1.2. Robustness to Refractive Index

In our paper, we use the assumption that the refractive index is 1.52 to facilitate offline calculation of physics-based prior. To verify the impact of this assumption on the estimation performance, we conduct the ablation study on the refractive index. The quantitative results are listed in Table 2, from which we can see that our method's performance is robust to the refractive index used in the physics-based prior calculation.

Table 2. Performance of our method with different refractive index

Refractive Index	1.4	1.5	1.52	1.6	1.7
Mean Error	16.54°	16.34°	16.29°	16.62°	16.58°
Median Error	13.21°	13.03°	12.85°	13.25°	13.32°

In Fig.4, we show curves of Eq.2 under different refractive indexes. The difference between them is small, only when ρ_s is close to 1, the zenith error with different refractive indexes is relatively large and the maximum zenith error is limited to single-digit degrees. Therefore, the refractive index can not have a significant impact on the physics-based prior, which is the source of our network's robustness to the refractive index.

2. Analysis of Noise in AoLP Map

2.1. $\frac{\pi}{2}$-ambiguity

According to the dichromatic reflectance model[1], the reflected light from the surface of an object is the superposition of specular reflection and diffuse reflection as shown in Fig.3.

Specular reflection is the direct reflection of the surface. Following the Fresnel equations, the degree of linear polarization(DoLP) and angle of linear polarization(AoLP) can be written as follows by given azimuth ϕ, zenith angle θ and relative refractive index η:

$$\rho_s = \frac{2\sin^2\theta\cos\theta\sqrt{\eta^2-\sin^2\theta}}{\eta^2-\sin^2\theta-\eta^2\sin^2\theta+2\sin^4\theta} \tag{2}$$

$$\varphi_s = \phi \pm \frac{\pi}{2} \tag{3}$$

Different from the specular reflection, the incident light is re-emitted into the air after it enters and scatters on the medium. The re-emitted light is diffuse reflection light.

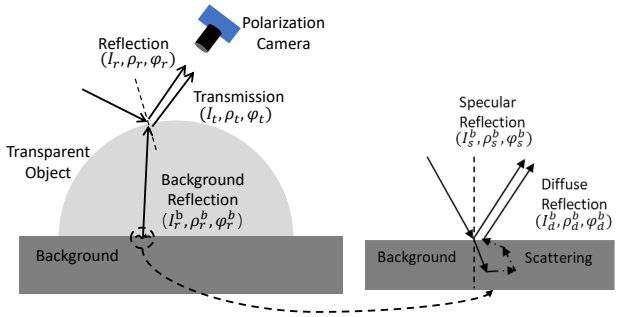

Figure 3. **Different components captured by polarization camera.** The captured ray consists of reflection and transmission components, where the transmission component is composed of two kinds of reflections from the rough background

Similarly, the DoLP ρ_d and AoLP φ_d can be obtained by Fresnel equation:

$$\rho_d = \frac{(\eta-1/\eta)^2\sin^2\theta}{2+2\eta^2-(\eta+1/\eta)^2\sin^2\theta+4\cos\theta\sqrt{\eta^2-\sin^2\theta}} \tag{4}$$

$$\varphi_d = \phi \pm \pi \tag{5}$$

Eq.3 and Eq.5 indicate that there is a phase shift of $\pi/2$ between φ_s and φ_d, i.e., the $\frac{\pi}{2}$**-ambiguity**. This ambiguity occurs on the surface that the dominance of specular reflection and diffusion reflection can not be determined. In this paper, the surface of transparent object is smooth and the diffuse reflection can be ignored. Therefore, only Eq.2 and Eq.3 are adopted for physics-based prior calculation. For the rough background surface, $\frac{\pi}{2}$**-ambiguity** may exist, which is described in detail below.

2.2. Noise from Background in the AoLP Map

Since the background is a rough surface, its surface reflected light has both specular and diffuse components. Suppose the intensity functions of specular and diffuse reflection have the following forms:

$$I_s^b(\theta_{pol}) = I_s^b[1+\rho_s^b\cos(2\alpha_{pol}-2\varphi_s^b)] \tag{6}$$

$$I_d^b(\theta_{pol}) = I_d^b[1+\rho_d^b\cos(2\alpha_{pol}-2\varphi_d^b)] \tag{7}$$

where α_{pol} is the angle of the polarizer in front of the camera. Therefore the intensity function of each point of the background can be obtained by superposition $I_s^b(\alpha_{pol})$ and $I_d^b(\alpha_{pol})$:

$$
\begin{aligned}
I_r^b(\alpha_{pol}) &= I_s^b[1+\rho_s\cos(2\alpha-2\varphi_s^b)] + I_d^b[1+\rho_d^b\cos(2\alpha-2\varphi_d^b)] \\
&= I_s^b + I_d^b + I_s^b\rho_s^b\cos(2\alpha-2\varphi_s^b) + I_d^b\rho_d^b\cos(2\alpha-2\varphi_s^b\mp\pi) \\
&= I_s^b + I_d^b + |I_s^b\rho_s^b - I_d^b\rho_d^b|\cos[2\alpha-2(\varphi_s^b - \frac{1-\mathrm{sign}(\mathrm{I_s^b}\rho_s^b - \mathrm{I_d^b}\rho_d^b)}{2}\pi)] \\
&= I_{r0}^b[1+\rho_r^b\cos(2\alpha-2\varphi_r^b)]
\end{aligned}
\tag{8}
$$

Therefore, the aolp of each point φ_r^b in the background can be written as:

$$
\varphi_r^b = \varphi_s^b - \frac{1-\mathrm{sign}(\mathrm{I_s^b}\rho_s^b - \mathrm{I_d^b}\rho_d^b)}{2}\pi \tag{9}
$$

The $\frac{\pi}{2}$ phase shift occurs when the dominance(signs of $I_s\rho_s^b - I_d\rho_d^b$) of specular and diffuse reflection components between adjacent pixels changes, which is the source of the noise from the background. It is mentioned in the main text that although the polarization state of I_r^b will change when the ray passes through the transparent object, the noise characteristic of φ_t is still consistent with φ_r^b.

3. Calculation of Physics-based Prior

According to Eq.2 and Eq.3, the corresponding zenith angle and azimuth angle can be solved from the observed DoLP and AoLP, but the solution is not unique. It can be easily seen that an AoLP φ corresponds to two azimuth angles ϕ_1, ϕ_2. For DoLP ρ, as shown in Figure 3, each DoLP value also corresponds to two zenith angles θ_1, θ_2.

By combining the zenith and azimuth angles, we can get four normals:

$$
N_{phy0} = (\sin\theta_1\cos\phi_1, \sin\theta_1\sin\phi_1, \cos\theta_1) \tag{10}
$$

$$
N_{phy1} = (\sin\theta_2\cos\phi_1, \sin\theta_2\sin\phi_1, \cos\theta_2) \tag{11}
$$

$$
N_{phy2} = (\sin\theta_1\cos\phi_2, \sin\theta_1\sin\phi_2, \cos\theta_1) \tag{12}
$$

$$
N_{phy3} = (\sin\theta_2\cos\phi_2, \sin\theta_2\sin\phi_2, \cos\theta_2) \tag{13}
$$

These four normal maps contain the possible orientations of each point on the surface. Because of the physical polarization theory information contained in them, we regard and call them as the **physics-based prior**.

Since Eq.2 is a highly nonlinear function, it is difficult to obtain the analytical expression of $\theta(\rho_s)$. We adopt the iterative method to solve the approximate solutions of θ_1, θ_2, which is a time-consuming process and hence we calculate the physics-based prior offline and input it to the network instead of replacing it with some layers in the network.

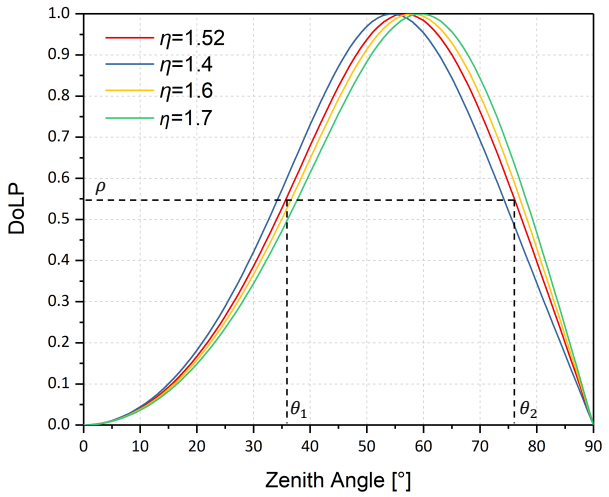

Figure 4. **Relationship between DoLP and zenith angle in specular reflection.** We draw the relationship between DoLP and zenith angle under different refractive indexes, and each DoLP corresponds to two zenith angles, which means there is ambiguity in the zenith angle calculation

4. Details of Our Dataset and Network

4.1. Dataset Details

Our dataset consists of real-world and synthetic collections and their acquisition setup and method have been described in our main submission. As shown in Fig.5, the real-world collection has 486 samples of 10 objects and the synthetic part has 936 samples of 13 objects, respectively.

Each sample includes the following information: $I_{0°}$, $I_{45°}$, $I_{90°}$, $I_{135°}$, Intensity map(I), DoLP map(ρ), AoLP map(φ), mask map, four physics-based prior normal maps, and ground-truth normal map. Where $I_{0°}$, $I_{45°}$, $I_{90°}$, $I_{135°}$ is the original intensity maps captured by our polarization camera. The polarization state (I, ρ, φ) can be calculated from the four intensity maps:

$$
I = \frac{I_{0°} + I_{45°} + I_{90°} + I_{135°}}{4} \tag{14}
$$

$$
\rho = \frac{\sqrt{(I_{0°} - I_{90°})^2 + (I_{45°} - I_{135°})^2}}{I_{0°} + I_{90°}} \tag{15}
$$

$$
\varphi = \frac{1}{2}\arctan\frac{I_{45°} - I_{135°}}{I_{0°} - I_{90°}} \tag{16}
$$

Then the polarization state can be used to calculate physics-based prior by exploiting Eq.2 and Eq.3. Only specular reflection model is used here since the diffuse reflection on smooth transparent surface is tiny. The mask map and ground-truth normal map are generated after manually aligning the ground-truth 3D model with the captured polarization image.

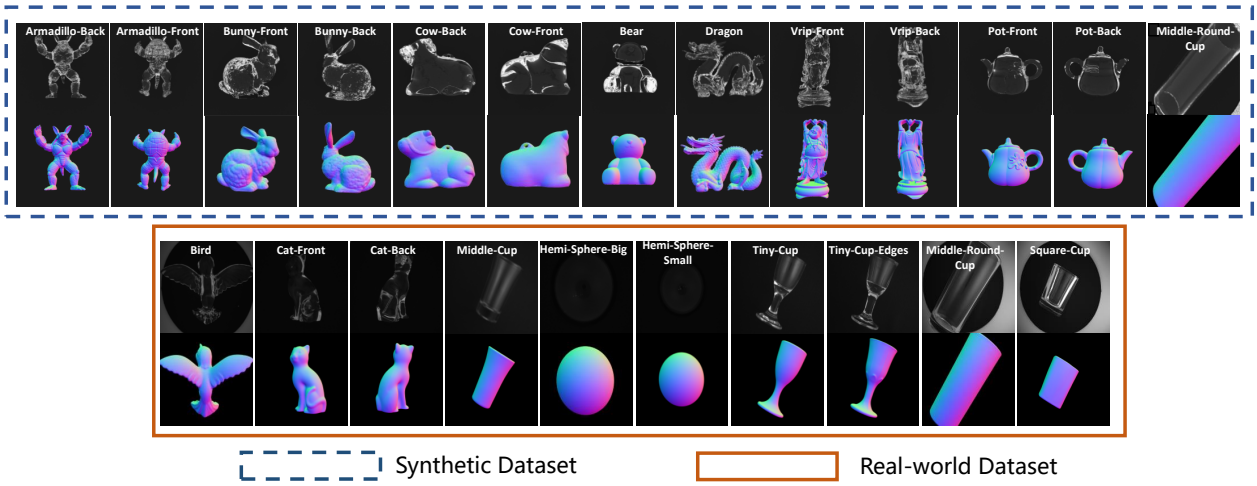

Figure 5. **TransSfP Dataset.** TransSfP dataset consists of two parts: synthetic dataset and real-world dataset. The synthetic dataset consists of 936 samples of 13 objects (including the front and back of the same object) at different angles. The real dataset consists of 486 samples of 10 objects at different angles

4.2. Detailed Network Architecture

The overview of our network has been given in Fig.3 of the main submission. The main structure of the encoder is ResNet50[3] with EPSABlock[6]. Table 3 lists the structure and parameters of the modules used in our network. Resnet50 with EPSABolck is used in Down1,2,3,4, and its structure is so large that can not list here, which detailed structure is provided in the code files in the attachment, as true is ASPP module. The structures of Up 1,2,3,4 and Fusion 1,2,3,4 are similar, the only difference is the number of channels in the convolution layers.

Table 3. **Detailed description of each module in our network architecture.** The **BN** and **IN** in this table represent batch normalization[4] and instance normalization[5], respectively

Module	Block
Down 0	$\begin{bmatrix} 1\times1, 64 \\ \text{BN, ReLU} \\ 7\times7, 64, \text{stride } 2 \\ \text{BN, ReLU} \\ 3\times3 \text{ max pool stride } 2 \end{bmatrix}$
Down 1	ResNet50 Stage 1 with EPSABlock[6]
Down 2	ResNet50 Stage 2 with EPSABlock
Down 3	ResNet50 Stage 3 with EPSABlock
Down 4	ResNet50 Stage 4 with EPSABlock
ASPP	ASPP module[2] with dialations of $[1, 12, 24, 36]$
Up 4(3,2,1)	$\begin{bmatrix} \text{Fusion Module } 4(3,2,1) \\ \text{Bilinear Upsample} \\ 3\times3, 512(256,128,64) \\ \text{IN, LeakyReLU} \\ 1\times1, 256(128,64,32,16) \\ \text{IN, LeakyReLU} \end{bmatrix}$
Fusion 4(3,2,1)	$\begin{bmatrix} \text{Bilinear Upsample} \\ 3\times3, 1024(512,256,64) \\ \text{IN, LeakyReLU} \\ 1\times1, 256(128,64,32) \\ \text{IN, LeakyReLU} \end{bmatrix}$
Fusion 0	$\begin{bmatrix} \text{Bilinear Upsample} \\ 3\times3, 64 \\ \text{IN, LeakyReLU} \\ 1\times1, 16 \\ \text{IN, LeakyReLU} \end{bmatrix}$
Up 0	$\begin{bmatrix} 3\times3, 32 \\ \text{IN, LeakyReLU} \\ 1\times1, 3 \end{bmatrix}$

References

[1] Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *IEEE transactions on image processing*, 15(6):1653–1664, 2006. 2

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 4

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[4] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 4

[5] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4

[6] Hu Zhang, Keke Zu, Jian Lu, Yuru Zou, and Deyu Meng. Epsanet: An efficient pyramid squeeze attention block on convolutional neural network. *arXiv preprint arXiv:2105.14447*, 2021. 4