Polarimetric Normal Stereo

Yoshiki Fukao   Ryo Kawahara   Shohei Nobuhara   Ko Nishino

Graduate School of Informatics, Kyoto Univesity

https://vision.ist.i.kyoto-u.ac.jp/

Abstract

We introduce a novel method for recovering per-pixel surface normals from a pair of polarization cameras. Unlike past methods that use polarimetric observations as auxiliary features for correspondence matching, we fully integrate them in cost volume construction and filtering to directly recover per-pixel surface normals, not as byproducts of recovered disparities. Our key idea is to introduce a polarimetric cost volume of distance defined on the polarimetric observations and the polarization state computed from the surface normal. We adapt a belief propagation algorithm to filter this cost volume. The filtering algorithm simultaneously estimates the disparities and surface normals as separate entities, while effectively denoising the original noisy polarimetric observations of a quad-Bayer polarization camera. In addition, in contrast to past methods, we model polarimetric light reflection of mesoscopic surface roughness, which is essential to account for its illumination-dependency. We demonstrate the effectiveness of our method on a number of complex, real objects. Our method offers a simple and detailed 3D sensing capability for complex, non-Lambertian surfaces.

1. Introduction

Stereo reconstruction has been a long-standing research topic in computer vision since its inception. Binocular stereo, in particular, has been studied in depth and has been deployed in a wide range of applications. Its simple passive setup which requires minimal calibration, maintenance, and cost has made it a reliable choice for 3D sensing. Even when alternative methods with higher precision are available, binocular stereo is often favored for its dense per-pixel depth that comes with relatively low cost for computation.

Stereo, however, is inherently limited by its underlying reconstruction process, namely matching and triangulation. Correspondence matching fundamentally requires view-independent appearance (color constancy), which translates to limited applicability in terms of target surface materials. Despite the large body of work including those that train deep neural networks to establish matching metrics, departing from this Lambertian surface requirement remains challenging. Triangulating the resulting correspondences also only recovers surface depth. For most cases, due to the fragility of this matching and triangulation, spatial regularization and quantization are employed. As a result, the geometry recovered by stereo, albeit useful for many applications, is often a crude measurement of the true surface.

Can we make stereo, in particular, simple binocular stereo recover detailed geometry of real-world surfaces that are composed of arbitrary materials? Can we match non-Lambertian surface points, but recover the geometry without relying solely on their geometric triangulation? In this paper, we show that we can achieve these by exploiting polarization of light reflected from real-world surfaces.

Catapulted by the introduction of quad-Bayer polarization cameras, polarization cues have started to see adoption in a wide range of computer vision methods. Geometry reconstruction is no exception (see Sec. 2). These past methods, however, use polarization as auxiliary cues for matching and proceeds with regular triangulation of depth. Surface normals are only computed from the recovered depth. That is, they are byproducts of the depth and not measurements of the actual surface normals. They also ignore the complex polarimetric reflection properties and assume purely Lambertian or mirror reflection, which ostracizes a broad range of real-world materials and lighting conditions.

In this paper, we show that we can establish surface point correspondences in a polarimetric stereo pair and recover per-pixel surface normals from the two polarimetric observations. We also integrate a full polarimetric BRDF model to handle complex lighting-dependent polarimetric reflection. To the best of our knowledge, our work is the first to show that surface normals can be directly, not as a byproduct of depth, recovered from binocular polarimetric stereo for a wide range of surfaces from matte, glossy, to mirrored.

Our key idea is to formulate simultaneous estimation of per-pixel depth-independent normal and albedo as RGB-polarimetric cost volume filtering. In addition to a regular RGB cost volume as a function of pixel disparities, we also construct a polarimetric cost volume that stores Stokes

vector differences for different surface normals and albedo values. These surface normals are computed directly from corresponding Stokes vectors in the two stereo views given hypothesized disparity values. Our goal is to filter these cost volumes to arrive at optimal disparities that give pixel correspondences, which in turn enables computation of surface normals and albedo values from the two Stokes vectors.

We achieve this cost volume filtering using belief propagation with three distinct characteristics. First, filtering of the two cost volumes are seamlessly integrating by multiplication of their beliefs. Second, the beliefs encode surface normals and use them to propagate disparities according to them. This leads to depth estimates that respect the surface normals measured in their view-dependent polarimetric observations. Finally, the surface normals themselves are also propagated, which effectively denoises the surface normals on the hypothesized surface. This is essential for using quad-Bayer polarization cameras as they are inherently noisy. These updated surface normals are then fed back into the polarimetric cost volume, i.e., the Stokes vectors are updated to match the surface normals, and the whole process is iterated to convergence. We also fully model the diffuse, specular lobe, and specular spike reflection with a microfacet-based polarimetric BRDF model [3]. Accounting for illumination-dependent polarization by glossy reflection in this way, which past methods ignored, is crucial for practical polarimetric 3D reconstruction.

We experimentally validate our method on a number of objects captured in a variety of lighting conditions. The results demonstrate the accuracy of the recovered surface normals and the method's effectiveness in practical real-world situations. With the advent of polarization cameras, we believe these results have implications in a wide range of areas, including autonomous driving, robotics, VR/AR, and medicine owing to its passive reconstruction of detailed geometry from a simple setup.

2. Related works

The majority of past stereo algorithms are disparity-based, which computes surface normals as gradients of recovered depth. These methods tend to result in overly smooth surface normals. Patch-based reconstruction methods [8, 5] explicitly estimate the surface normal to deform the texture matching window, but cannot estimate per-pixel surface normals as they still rely on window matching.

Light reflection by a surface changes its state of polarization, *i.e.*, the polarization state implicitly encodes the surface normal direction. The mapping between the normal and the polarization state is, however, not bijective. Various methods have been proposed that make different assumptions on the surface orientation while utilizing initial shape reconstruction from non-polarization methods [10, 7, 4, 21] and illumination [9]. Others assume differ-

ent polarimetric reflection properties such as diffuse only [1, 13, 15, 2], mirror dominant [19], and a combination of them [11, 22, 3, 17].

Kadambi *et al*. [10] use polarization cues to refine surface normals of geometry captured with conventional depth sensing. Cui *et al*. [7] disambiguate possible surface normals computed from polarization using depth recovered by conventional stereo. Berger *et al*. [4] combines color-based cost with a polarization-based cost function to aid correspondence search in non-Lambertian areas. Zhao *et al*. [21] refine depth estimates using multiview angle-of-polarization images. These methods fundamentally rely on depth estimates from regular stereo reconstruction or other 3D sensing techniques and polarimetric observations are auxiliary information for regular stereo matching and triangulation, not a source of direct surface normal recovery.

Wolff and Boult [19] directly recover surface normals from polarimetric observations by intersecting specular planes of incidence defined by the polarizer angles at each view. Atkinson and Hancock [2] proposed binocular stereo with polarizers. They assume pure diffuse polarization to estimate normal zenith from the degree of polarization.

Zhu and Smith [22] classify surface points into either pure diffuse or mirror reflection to estimate their surface normals. Smith *et al*. [17] introduce a linear solution for single-image reconstruction. Yu *et al*. [20] propose an analysis by synthesis approach. Miyazaki *et al*. [12] and Chen *et al*. [6] estimate surface normals as the intersection of plane-of-reflections defined by the angle of polarizer at each viewpoint. All these methods use the same diffuse or mirror binary classification, which is inherently limiting as real surfaces are always a combination of them at a pixel, not a spatial binary map of either. Moreover, they ignore specular lobe (glossy) reflection which is essential for handling lighting-dependent polarization of real surfaces.

Baek *et al*. [3] recently introduced a method that estimates the surface normal for full polarimetric reflection consisting of diffuse and specular lobe (not merely mirror). The method, however, requires an active stereo system to obtain the accurate 3D shape by structured lighting and fundamentally relies on a co-axial imaging setup.

In contrast to these past methods, our polarimetric normal stereo is completely passive, does not require initial estimates of depth, and recovers per-pixel surface normals for combined diffuse and specular reflection directly from polarimetric observations.

3. Polarimetric Reflection

Let us first review polarization in general and then polarimetric reflection and its BRDF model.

3.1. Polarization

Light is a composition of transverse waves of electric and magnetic fields that are always perpendicular to each other. The "orientation" of light can be defined as the angle the electric plane wave makes in the plane perpendicular to the traverse direction. Within a non-zero finite time of observation, this orientation can be randomly distributed. We call such light unpolarized. In contrast, light can be oriented in a single direction, which we refer to as linearly polarized light. This orientation can also be rotating as a function of time. Such light is called circular polarized. In this paper, we only consider linear polarization as surface reflection primarily causes it, but not circular polarization unless with water.

Within the temporal span of an observation (i.e., camera exposure), the observed light can consist of a collection of linearly polarized light of varying magnitudes. This results in an elliptic distribution of linear polarization. If we observe such partially linearly polarized light with a camera equipped with a polarization filter on the image plane (or lens plane), the observed intensity will be a function of the filter angle ϕ_c

$$
\begin{aligned}
I(\phi_c) &= I_{\max} \cos^2(\phi_c - \phi) + I_{\min} \sin^2(\phi_c - \phi) \\
&= \overline{I} + \rho \overline{I} \cos(2\phi_c - 2\phi),
\end{aligned} \tag{1}
$$

where I_{max} and I_{min} are the light intensities in the major and minor axes of the ellipse and $\overline{I}$ is the average intensity ($= \frac{I_{\max} + I_{\min}}{2}$). The scalar $\rho = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$ is referred to as the degree of linear polarization (DoLP) and represents how strongly the light is linearly polarized (i.e., how elongated the ellipse is). The angle ϕ is called the angle of linear polarization (AoLP) and represents the major linear polarization angle. The observed intensity $I(\phi_c)$ becomes a sinusoidal wave of ϕ_c which takes on its maximum value at $\phi_c = \phi$.

To recover the polarization state of a linearly polarized light from its intensity, we need at least three observations at three different filter angles (i.e., three angular samples of the polarization ellipse). Quad-Bayer polarization cameras use four filters of different angles laid out on each pixel. Intensity observations at these four filter angles of $\frac{\pi}{4}$ increments can be expressed with the Stokes vector

$$
\boldsymbol{s} = \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} I(0) + I(\frac{\pi}{2}) \\ I(0) - I(\frac{\pi}{2}) \\ I(\frac{\pi}{4}) - I(\frac{3\pi}{4}) \\ 0 \end{bmatrix} = \begin{bmatrix} 2\overline{I} \\ 2\rho\overline{I} \cos(2\phi) \\ 2\rho\overline{I} \sin(2\phi) \\ 0 \end{bmatrix}. \tag{2}
$$

The polarization state can easily be extracted from the Stokes vector

$$
\overline{I} = \frac{s_0}{2}, \quad \rho = \frac{\sqrt{s_1^2 + s_2^2}}{s_0}, \quad \phi = \frac{1}{2} \tan^{-1}\left(\frac{s_2}{s_1}\right). \tag{3}
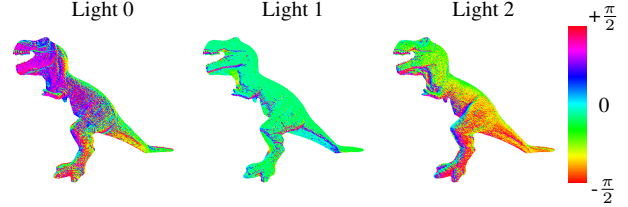$$

Figure 1. When a surface is illuminated from different directions (light 0: left behind camera, 1: above camera, and 2: right behind camera), the angle of polarization changes. This phenomenon cannot be explained with the polarization characteristics of diffuse and mirror reflection, the latter of which is often referred to as specular reflection in past methods. It requires modeling of the microgeometry of the surface projected in each pixel.

3.2. Polarimetric Microfacet BRDF

Polarimetric light reflection by an object surface can be characterized with two processes. When the incident light strikes the interface, part of the light gets reflected in the perfect mirror direction where the incident, surface normal, and viewing directions span the reflection plane. This mirror reflection, regardless of the polarization state of the incident light, linearly polarizes the light in the direction perpendicular to the reflection plane (s-polarized). In contrast, the light that transmits into the subsurface is polarized in the direction parallel to the reflection (refraction) plane (p-polarized), gets depolarized due to scattering, and then becomes p-polarized again when reemitted back into air. Past methods for polarimetric 3D reconstruction have assumed this combination of diffuse plus mirror reflection, often referring to the latter as "specular" reflection. This, however, is incomplete and does not explain an important property of polarization of surface reflection.

Fig. 1 shows images of the AoLP of a real scene computed from polarimetric observations captured with a quad-Bayer polarimetric camera from a fixed view point but with a different light source direction for each image. If the surface reflection was really a linear combination of diffuse reflection and mirror reflection, the AoLP at each surface point should have stayed the same regardless of the lighting. Fig. 1 shows otherwise; the AoLP distribution clearly changes together with the light source direction. This is because, as the light source direction changes, the surface normals that contribute to "specular" reflection (light reflected at the interface of the surface) actually varies. That is, the mesoscopic surface contains a variety of surface normals in the projected area of a pixel, and a different set of them that lies on the plane spanned by the normal and viewing directions are observed via mirror reflection. As a result, this mesoscopic surface roughness introduces illumination-dependent polarization. This means that, just like regular radiometric surface reflection modeling [14], we must account for the polarimetric properties of glossy reflection

(specular lobe) as depicted in Fig. 2(a).

We model the polarimetric light reflection as a linear combination of diffuse and specular lobe reflections. Note that the polarization properties of specular spike reflection are included in the polarimetric specular lobe reflection which we, from now on, refer to as specular reflection. The mesoscopic surface can be modeled as a collection of microfacet mirrors whose polarimetric reflection can be derived similarly to a radiometric microfacet bidirectional reflection distribution function (BRDF). Baek *et al.* [3] introduce such a polarimetric microfacet BRDF. Instead of expressing the polarization state in Stokes vector parametrization, here we review this model in terms of AoLP and DoLP. This formulation is more suitable for surface normal estimation in our setting.

The radiometric microfacet BRDF model can be expressed as a linear combination of diffuse reflection and specular reflection

$$I = (\ell \cdot n)\left(f_d(\ell, n, v_c) + f_s(\ell, n, v_c, \sigma)\right) L, \quad (4)$$

where I is the observed radiance, L is the source radiance, σ is the surface roughness, and f_d and f_s are the diffuse and specular reflectance, respectively. The diffuse reflectance is a function of the incident light direction ℓ, surface normal n, and the viewing direction v_c. In contrast, specular reflectance is also a function of the surface roughness σ.

Diffuse reflectance is that of the light transmitted into the subsurface that is scattered and transmitted back into the viewing direction

$$f_d(\ell, n, v_c) = k_d T(n, v_c) T(\ell, n), \quad (5)$$

where T is Fresnel transmittance and k_d is the diffuse albedo.

For specular reflectance that models the specular lobe and spike, we adopt the microfacet model by Walter *et al.* [18]

$$f_s(\ell, n, v_c) = k_s W(\ell, n, v_c, \sigma) R(h, v_c), \quad (6)$$

where

$$W(\ell, n, v_c, \sigma) = \frac{D(n, h, \sigma) G(\ell, n, v_c, \sigma)}{4|\ell \cdot n||n \cdot v_c|}, . \quad (7)$$

Here $D(n, h, \sigma)$ is the surface normal distribution of the microfacets, where h is the half vector of the viewing and incident light directions, and $G(\ell, n, v_c, \sigma)$ is the geometric attenuation term.

The Fresnel reflection R and transmittance T at polarization filter angle ϕ_c on the image plane becomes

$$\begin{aligned} R(\phi_c) &= R_s \cos^2(\phi_c - \phi_r) + R_p \sin^2(\phi_c - \phi_r) \\ &= \frac{R_s + R_p}{2} + \frac{R_s - R_p}{2} \cos(2\phi_c - 2\phi_r) \\ &= \overline{R} + \rho_r \overline{R} \cos(2\phi_c - 2\phi_r), \quad (8) \end{aligned}$$
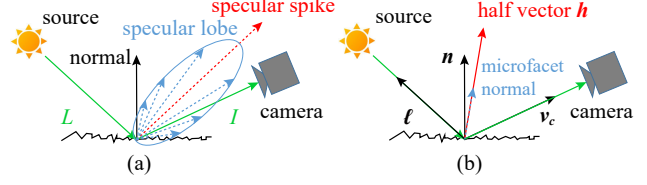

Figure 2. (a) Polarization of specular lobe (gloss) reflection is essential to account for the illumination-dependent polarization of light reflection on real-world surfaces. Past methods have only modeled the specular spike (mirror) reflection as "specular" reflection. (b) We model the specular lobe with a microfacet orientation distribution using the halfway vector.

and

$$\begin{aligned} T(\phi_c) &= \frac{T_p + T_s}{2} + \rho_t \frac{T_p - T_s}{2} \cos(2\phi_c - 2\phi_t) \\ &= \overline{T} + \rho_t \overline{T} \cos(2\phi_c - 2\phi_t), \quad (9) \end{aligned}$$

where the subscripts s and p denote the perpendicular and parallel components to the reflection plane, ρ_r and ρ_t are the degree of linear polarization of reflection and transmittance, respectively, and ϕ_r and ϕ_t are the angle of polarization of reflection and transmittance, respectively. We have dropped dependency on the halfway vector, normal, and light source and viewing directions for brevity. Note that light transmitted into the surface is depolarized before reemitted to air, which is why Fresnel transmittance into the subsurface $T(\ell, n)$ is not a function of ϕ_c.

The observed radiance at polarization filter angle ϕ_c can be written as

$$\begin{aligned} I(\phi_c) = (\ell \cdot n)\big(&k_d T(\ell, n) T(n, v_c, \phi_c) \\ &+ k_s W(\ell, n, v_c, \sigma) R(h, v_c, \phi_c)\big) L. \quad (10) \end{aligned}$$

From Eqs. 10, 9, 8, and 2, the Stokes vector of a surface point with surface normal n, diffuse albedo k_d, and specular albedo k_s can be computed from its polarimetric observations $I(\phi_c)$

$$\widetilde{s}(n, k_d, k_s) =$$
$$\begin{bmatrix} 2(\ell \cdot n)\left(\widetilde{f}_d + \widetilde{f}_s\right) L \\ 2(\ell \cdot n)\left(\widetilde{f}_d \rho_t \cos(2\phi_t) + \widetilde{f}_s \rho_r \cos(2\phi_r)\right) L \\ 2(\ell \cdot n)\left(\widetilde{f}_d \rho_t \sin(2\phi_t) + \widetilde{f}_s \rho_r \sin(2\phi_r)\right) L \\ 0 \end{bmatrix}, \quad (11)$$

where

$$\widetilde{f}_d = k_d T(\ell, n) \overline{T} \quad (12)$$
$$\widetilde{f}_s = k_s W(\ell, n, v_c, \sigma) \overline{R}. \quad (13)$$

4. Polarimetric Normal Stereo

As depicted in Fig. 3, our method directly recovers surface normals from polarimetric stereo pairs by cost volume
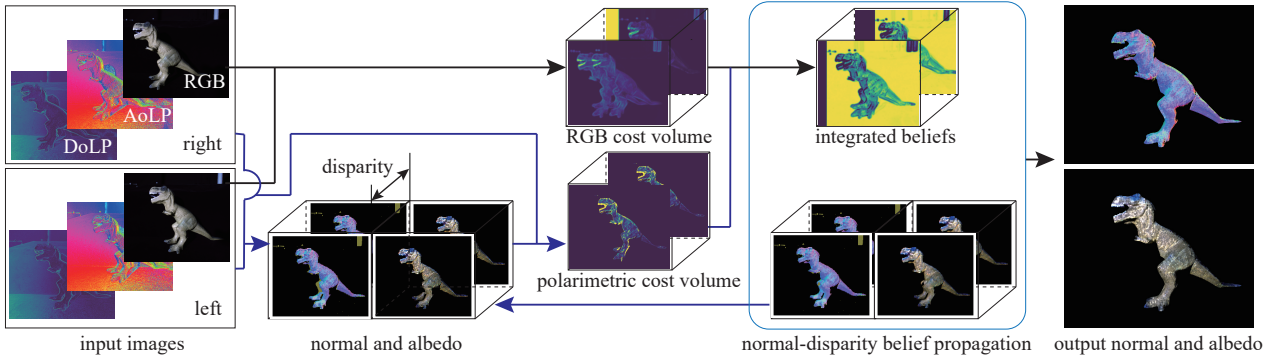
Figure 3. Overall framework of polarimetric normal stereo. From a pair of polarimetric images from which AoLP and DoLP can be computed for each pixel, we construct a polarimetric cost volume, in addition to a regular RGB cost volume, that measures the Stokes vector discrepancy between that computed from the surface normal estimate and the two observations for a given disparity. We filter these cost volumes, while effectively denoising the input polarimetric observations to estimate the surface normal as well as diffuse and specular albedo values at each pixel.

construction and filtering that fully leverages the polarimetric observations with polarimetric cost volume construction, surface normal propagation, and iterative updating.

4.1. Polarimetric Cost Volume

A regular RGB cost volume is constructed by evaluating the RGB color difference for a given set of discrete disparity values at each pixel

$$
\begin{aligned}
C_{RGB}(p, d_p) = &|\overline{I}_L(u, v) - \overline{I}_R(u - d_p, v)| \\
&+ |\nabla \overline{I}_L(u, v) - \nabla \overline{I}_R(u - d_p, v)|,
\end{aligned}
\tag{14}
$$

where d_p denotes the disparity at pixel $p = (u, v)$, $\overline{I}_L$, $\overline{I}_R$ are the left and right DC component of intensity, respectively, and $C_{RGB}(d)$ denotes the cost at (u, v) for a disparity d. The second term encodes the difference in the intensity gradients.

In addition to this regular RGB cost volume, we leverage the polarimetric observations by constructing and filtering a polarimetric cost volume. We first define the polarimetric distance between a Stokes vector $\widetilde{s}$ computed from surface parameter estimates consisting of the surface normal, diffuse albedo and specular albedo and the observed Stokes vectors s in the two views

$$
\begin{aligned}
&L_{\boldsymbol{s}}(p, d_p, \boldsymbol{n}_p, k_{d,p}, k_{s,p}) = \\
&|\boldsymbol{s}_{\{L,R\}}(u, v) - \widetilde{\boldsymbol{s}}_{\{L,R\}}(u, v, \boldsymbol{n}_p, k_{d,p}, k_{s,p})|,
\end{aligned}
\tag{15}
$$

where we add the cost for left and right views $\{L, R\}$. We can estimate the surface normal and albedo values at a pixel as those that minimize this cost

$$
\boldsymbol{n}_p^\star, k_{d,p}^\star, k_{s,p}^\star = \underset{\boldsymbol{n}_p, k_{d,p}, k_{s,p}}{\arg\min} \ L_{\boldsymbol{s}}(p, d_p, \boldsymbol{n}_p, k_{d,p}, k_{s,p}).
\tag{16}
$$

In our implementation, we achieve this with regular gradient descent. This optimization is over-constrained with effectively 10 constraints for 8 unknowns. As the surface normal is shared among the color channels, the Stokes vector will only differ in the first element, which reduces the apparent 18 constraints down to 10. Note that, for a single view, this means there are only 5 constraints, and thus single-view surface normal recovery is not possible. The optimization has a unique solution because the binocular observation provides an AoLP for each view and their intersection resolves the π- and $\pi/2$- ambiguities. This, in other words, means that given two polarimetric observations (i.e., a hypothesized disparity value d), we can estimate the surface normal and albedo values that best explain them.

For any hypothesized disparity value for a given pixel, we can solve for the surface normal and albedo values at the corresponding surface point of the pixel in interest from Eq. 16 and evaluate the goodness of that disparity value with the polarimetric distance (Eq. 15)

$$
C_{\boldsymbol{s}}(p, d_p) = L_{\boldsymbol{s}}(p, d_p, \boldsymbol{n}_p^\star, k_{d,p}^\star, k_{s,p}^\star).
\tag{17}
$$

We refer to this as the polarimetric cost volume. Note that the disparity values parameterize this cost volume but the surface normals and albedo values are used to evaluate the polarimetric distance and that the surface normals are computed from the polarimetric observations, not the disparity.

4.2. Normal-Disparity Belief Propagation

Unlike conventional binocular stereo, our goal is not to estimate the disparity but the surface normal at each pixel. The disparity, however, gives us pixel correspondence which is necessary to obtain two polarimetric observations to estimate the surface normal. We have constructed two cost volumes both parameterized by the dis-
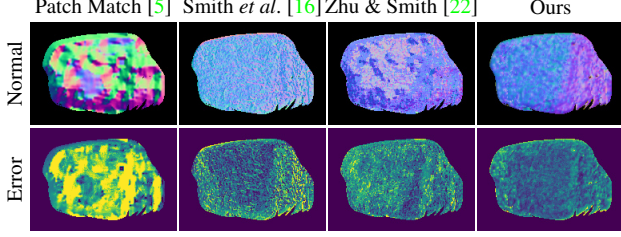
Figure 4. Surface normal estimates and their error maps compared with ground truth. Our method can recover fine surface geometry as per-pixel surface normals independent of depth and regardless of the lighting and surface roughness.

	[5]	[16]	[22]	Ours⋆1	Ours⋆2
(a)	33.3, 23.6	39.3, 34.2	27.8, 18.6	21.8, **17.4**	**21.8**, 17.6
	807.9	646.1	561.2	252.2	**228.7**
(b)	58.5, 52.5	50.8, 45.7	28.9, **15.8**	**20.9**, 16.7	21.3, 17.0
	825.6	872.5	843.3	292.4	**264.6**
(c)	**22.3, 14.3**	46.0, 43.4	25.6, 15.7	23.7, 18.7	23.6, 18.8
	532.5	257.0	420.2	260.9	**243.0**
(d)	**29.2, 25.8**	40.3, 38.3	35.7, 34.0	32.2, 28.4	32.8, 29.3
	336.9	405.1	337.2	412.8	**333.6**
(e)	70.2, 67.9	37.5, 34.4	41.1, 40.5	31.6, **29.2**	**31.5**, 29.8
	883.3	446.3	280.6	308.3	**267.6**

Table 1. Angular errors in degrees for (a) Pig, (b) Lemon, (c) Book, (d) Dinosaur, (e) Stone. The three numbers for each result report the mean, median, and the variance of the error. Our method consistently achieves higher accuracy compared with past methods. Propagating the normals with BP (Ours⋆2) results in less variance than without it (Ours⋆1).
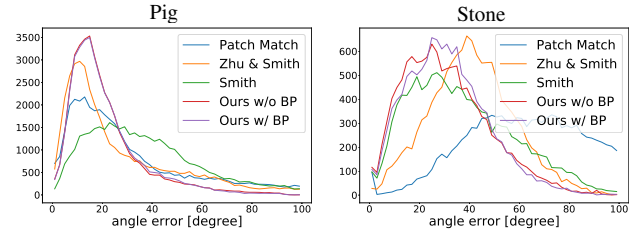

Figure 5. Histograms of angular errors in degrees for Pig and Stone. Our method has fewer pixels with large angular errors.

parity. We filter these cost volumes simultaneously with belief propagation by defining beliefs that encode the uncertainties of these costs and propagate them together with the surface normals. By also propagating the surface normals computed from corresponding polarimetric observations at pixels with disparity values of high certainty, we can effectively denoise the otherwise noisy polarimetric observations, which is critical to use quad-Bayer polarization cameras. These propagated surface normals are reflected in the polarimetric distance used to construct the polarimetric cost volume, and these cost volume reconstruction and filtering is iterated till convergence.

We define the energy potential to maximize as

$$\Psi(\boldsymbol{d}) = \exp\left[-E(\boldsymbol{d})\right], \quad (18)$$

where we define the energy

$$E(\boldsymbol{d}) = \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p} C_V(p, d_p, d_q) + \sum_{p \in \mathcal{P}} (C_{RGB}(p, d_p) + C_{\boldsymbol{s}}(p, d_p)), \quad (19)$$

where $\boldsymbol{d}$ is a vector of disparity values for all pixels. Here we have denoted the set of all pixels with $\mathcal{P}$ and the disparity at pixel $p \in \mathcal{P}$ with $d_p \in \mathcal{D}$, where $\mathcal{D}$ is a discrete set of possible disparity values. $\mathcal{N}_p$ denotes the four pixels horizontally and vertically adjacent to pixel p.

The pairwise cost $C_V(d_p, d_q)$ is defined as

$$C_V(p, d_p, d_q) = \begin{cases} 0 & (|d_p - \hat{d}_q| < 1) \\ P_1 & (1 < |d_p - \hat{d}_q| < 2) \\ P_2 & \text{otherwise} \end{cases}, \quad (20)$$

where $\hat{d}_q$ is the disparity value taking into account the surface normal $\boldsymbol{n}$ at p, which is defined as $\hat{d}_q = d_q + \Delta\hat{d}$. For a smooth surface area, $\Delta\hat{d}$ can be expressed using the surface gradient as

$$\Delta\hat{d} = d_p f \left(f + \frac{n_{z,p}}{n_{x,p}}\Delta u + \frac{n_{z,p}}{n_{y,p}}\Delta v \right)^{-1} - d_p, \quad (21)$$

where f is the focal length of the camera and $\Delta u, \Delta v$ are the horizontal and vertical differences of the pixel location $q - p$. P_1 and P_2 are penalties for discontinuities.

We find the disparity values $\boldsymbol{d}$ and simultaneously the surface normals and albedo values $\{(d_p, k_{d,p}, k_{s,p} : p \in \mathcal{P}\}$ that maximize the energy potential with belief propagation that integrates the beliefs from both the RGB and polarimetric cost volumes. Although each cost volume has non-negative values, in order to make them valid probabilistic uncertainties we define their potentials

$$B_{RGB}(p, d_p) = \exp\left[-C_{RGB}(p, d_p)\right] \quad (22)$$
$$B_{\boldsymbol{s}}(p, d_p) = \exp\left[-C_{\boldsymbol{s}}(p, d_p)\right]. \quad (23)$$

The uncertainty for a given disparity is then computed as their joint probability

$$B(p, d) = B_{RGB}(p, d_p) \times B_{\boldsymbol{s}}(p, d_p). \quad (24)$$

We can now define the message from pixel p to its neighbor pixel q

$$m_{p \to q}(d_q) = \sum_{d_p=0}^{d} \left(\exp\left[-C_V(p, d_p, d_q)\right] B(p, d_p)\right) \prod_{k \in \mathcal{N}_p \setminus q} m_{k \to p}(d_p). \quad (25)$$
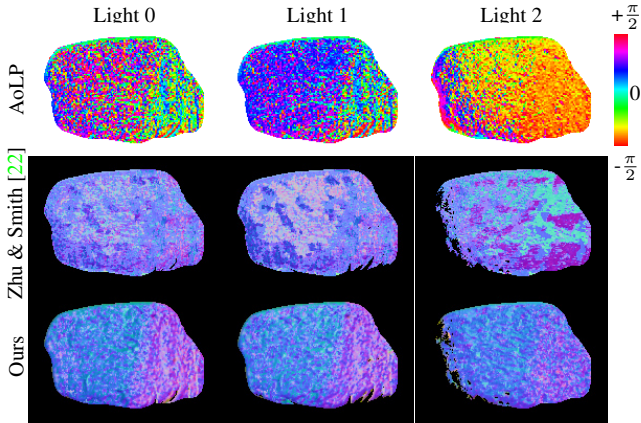
Figure 6. Surface normal estimates for polarimetric stereo pairs of the same object captured under different light source directions (light 0: left behind camera, 1: above light 0, 2: right behind camera). Despite the change in polarimetric observations, our method correctly recovers consistent per-pixel surface normals, while the baseline method suffers from inconsistency and fails to recover the two sides of the rock.

As we pass these messages from pixel to pixel, we also update the surface normal and albedo values at each pixel according to the uncertainty of its disparity value. By updating these quantities as a weighted linear combination of the current normal and albedo estimates at a pixel and its neighbors using the messages (uncertainties) as the weights

$$\boldsymbol{n}_q^\star = (1 - m_{p \to q}(d_q))\boldsymbol{n}_q + m_{p \to q}(d_q)\boldsymbol{n}_p \qquad (26)$$

(same for $k_{d,q}^\star$ and $k_{s,q}^\star$) we are able to denoise the raw polarimetric observations, effectively, while estimating the disparity, normal, and albedo values at each pixel.

As we propagate more certain surface normals and albedo values from neighbors, the polarimetric cost volume computed from the raw polarimetric observations should be updated to reflect the new normals and albedo values by substituting $n_p^\star$, $k_{d,p}^\star$, and $k_{s,p}^\star$ with those computed in Eq. 26. We then go back to running belief propagation on this updated polarimetric cost volume and the original RGB cost volume and iterate this process till convergence.

5. Experimental Results

We experimentally evaluate the effectiveness of our polarimetric normal stereo method on a number of real polarimetric images. We use two commercial color polarization cameras (Lucid TRI050S-QC) that use quad-Bayer polarization filter chips (Sony IMX250MYR) and calibrate them with conventional stereo calibration methods.

5.1. Surface Normal Estimation

We first evaluate the accuracy of recovered surface normals and compare it with past relevant methods. We consider three methods for comparison. The two representative

Figure 7. Diffuse albedo estimates of past methods suffer from residual shading as they do not model the full polarimetric reflection. Our method, in contrast, does not suffer from such artifacts.

shape from polarization methods, Zhu and Smith [22] and Smith et al. [16], model polarimetric reflection of only diffuse and mirror and essentially conduct binary classification on the surface[1]. In contrast, we model the full polarimetric BRDF including glossy specular reflection. Note that Zhu and Smith [22] assume known point source similar to our method. Smith et al. [16] can handle an unknown point source direction, but only when the object surface has uniform albedo and it has to be known, like in our method, for spatially varying albedo. Although we leave as future work, since the cost volume construction and filtering are clean separate steps in our method, we believe we can incorporate light source estimation as an alternating minimization where we iteratively update the point source direction used to construct the cost volumes. We also compare with surface normals computed by differentiating depth estimates reconstructed with PatchMatch Stereo [5] as a baseline.

Fig. 4 shows the surface normal estimates using our method and other methods as well as ground truth computed from photometric stereo. The results clearly show that our surface normal estimates capture the detailed geometry of the complex objects and match the ground truth well. They are also more accurate than other methods. For instance, the results show that the surface normals computed from recovered depth [5] do not capture fine surface geometry. Both methods by [22] and [16] result in large surface regions with inaccurate surface normals as they cannot take into account the illumination-dependency of polarimetric appearance. In sharp contrast, our polarimetric normal stereo is able to restore detailed surface geometry regardless of the depth and light source directions.

Table 1 shows mean and median angular errors of the surface normal estimates of all objects for all methods. The results show that our method achieves the highest accuracy for all objects. PatchMatch stereo [5] cannot leverage polarimetric information and suffers from textureless appearance especially of objects like the stone and the lemon. The stereo method by Zhu and Smith [22] only uses polarimetric information for matching and does not account for glossy reflection. These results demonstrate that directly computing surface normals from polarimetric information is essential to recover accurate fine geometry from polarimetric ob-

---

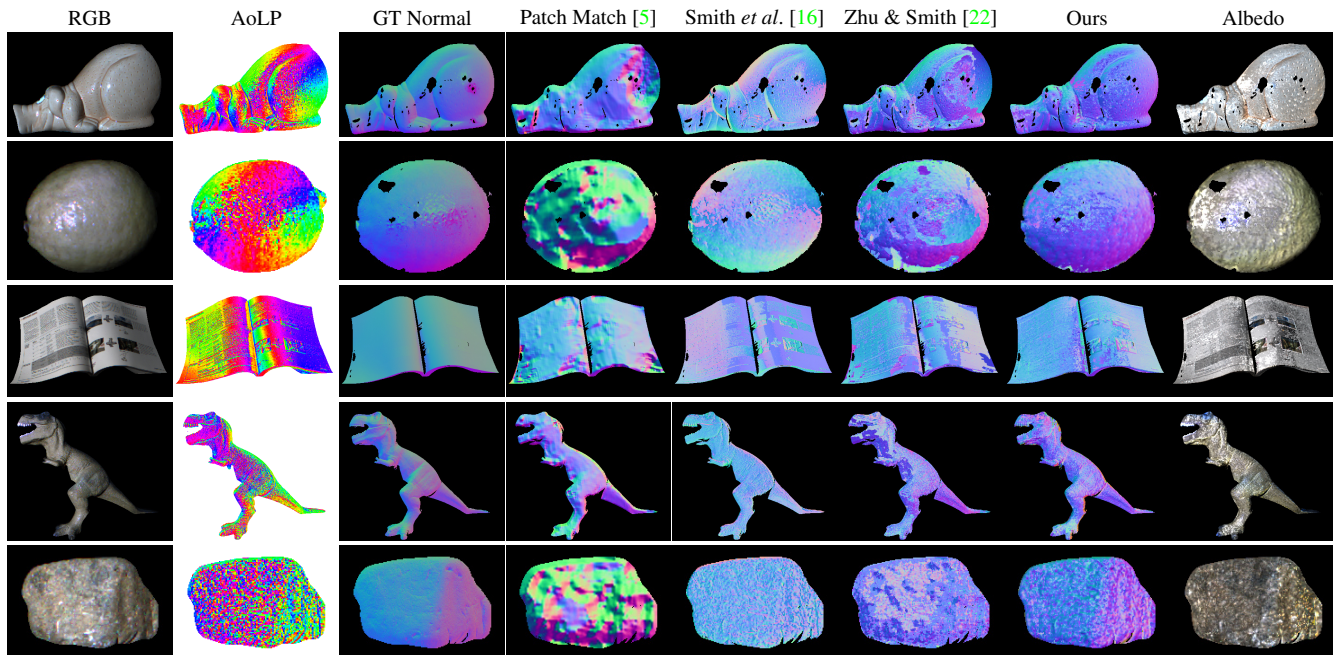[1]We used implementations provided by the paper authors.

Figure 8. Surface normal, albedo, and surface roughness recovery of various complex real objects. The results demonstrate the accuracy of polarimetric normal stereo. Patch Match Stereo [5] cannot estimate the surface normal for objects with homogeneous textures. The height recovery method by Smith *et al.* [16] and the stereo method bu Zhu & Smith [22] cannot accurately resolve the π-ambiguity.

servations. The height recovery method by Smith *et al.* [16] also cannot handle glossy reflection and results in large errors, especially for rough surfaces. Our method achieves state-of-the-art accuracy on these complex, real objects.

Fig. 5 shows histograms of angular errors for Pig and Stone. The results show that our method has fewer pixels with large angular errors than past methods.

5.2. Lighting Invariance and Albedo Estimation

Fig. 6 shows surface normal estimates for three different polarimetric stereo pairs of the same object taken under different light source directions. Note how the input AoLP changes for different lighting. Our method is able to recover consistent surface normals regardless of the lighting.

Fig. 7 shows diffuse albedo estimates. The albedo estimates by [22] and [16] suffer from residual shading as they model shading on the DC component of the intensity which actually includes the specular lobe. In contrast, our method is able to accurately estimate the spatially varying albedo without remaining shading, except for some irregularities in small saturated spots.

5.3. Complex Objects

Fig. 8 shows our results on various objects with complex reflection and geometry. The results demonstrate that our method is able to recover the fine geometry of these objects accurately regardless of material composition. As the input AoLP images show, the polarimetric observations are

quite noisy. Our method is able to robustly recover the surface normals and albedo values thanks to the denoising integrated in cost volume filtering. The black holes in images from 3rd through 8th columns of Fig. 8 correspond to pixels where photometric stereo for ground truth capture failed due to saturation. These holes are not identical to the highlights in the RGB images since the images for photometric stereo were captured under different lighting conditions.

6. Conclusion

We introduced a novel binocular stereo method that leverages polarimetric observations to recover fine geometry of objects with complex non-Lambertian reflectance properties. Our method models the lighting-dependent polarimetric appearance and directly recovers per-pixel surface normal and albedo from pairs of polarimetric observations. We achieved this by introducing a novel polarimetric cost volume and an iterative filtering method based on belief propagation that also denoises raw polarimetric observations. We believe this polarimetric normal stereo method significantly extends the reach of binocular stereo by enabling fine geometry reconstruction while retaining its simplicity and passiveness.

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] Gary A. Atkinson and Edwin R. Hancock. Shape estimation using polarization and shading from two views. *TPAMI*, 29, 2007. 2

[3] Seung-Hwan Baek, Daniel S. Jeon, Xin Tong, and Min H. Kim. Simultaneous acquisition of polarimetric svbrdf and normals. *ACM Transactions on Graphics (TOG)*, 37:1 – 15, 2018. 2, 4

[4] Kai Berger, Randolph Voorhies, and Larry H Matthies. Depth from stereo polarization in specular scenes for urban robotics. In *Proc. ICRA*, pages 1966–1973. IEEE, 2017. 2

[5] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In *Proc. BMVC*, 2011. 2, 6, 7, 8

[6] Lixiong Chen, Yinqiang Zheng, Art Subpa-asa, and Imari Sato. Polarimetric three-view geometry. In *Proc. ECCV*, 2018. 2

[7] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *Proc. CVPR*, July 2017. 2

[8] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *TPAMI*, 32(8):1362–1376, 2009. 2

[9] Graham Fyffe and Paul Debevec. Single-shot reflectance measurement from polarized color gradient illumination. In *Proc. ICCP*, pages 1–10, 2015. 2

[10] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proc. ICCV*, 2015. 2

[11] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul E Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 2007(9):10, 2007. 2

[12] Daisuke Miyazaki, Takuya Shigetomi, Masashi Baba, Ryo Furukawa, Shinsaku Hiura, and Naoki Asada. Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering*, 56(4):1 – 17, 2016. 2

[13] Daisuke Miyazaki, Robby T Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *Proc. ICCV*, page 982, 2003. 2

[14] Shree K. Nayar, Katsushi Ikeuchi, and Takeo Kanade. Surface reflection: Physical and geometrical perspectives. *TPAMI*, 13(7):611–634, 1991. 3

[15] Trung Ngo Thanh, Hajime Nagahara, and Rin-ichiro Taniguchi. Shape and light directions from shading and polarization. In *Proc. CVPR*, pages 2310–2318, 2015. 2

[16] William AP Smith, Ravi Ramamoorthi, and Silvia Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *Proc. ECCV*, pages 109–125, 2016. 6, 7, 8

[17] William A. P. Smith, Ravi Ramamoorthi, and Silvia Tozza. Height-from-polarisation with unknown lighting or albedo. *TPAMI*, 41(12):2875–2888, 2019. 2

[18] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, pages 195–206. Eurographics Association, 2007. 4

[19] Lawrence B. Wolff and Terrance E. Boult. Constraining object features using a polarization reflectance model. *TPAMI*, 13(7):635–657, 1991. 2

[20] Ye Yu, Dizhong Zhu, and William AP Smith. Shape-from-polarisation: a nonlinear least squares approach. In *Procóf ICCV Workshops*, pages 2969–2976, 2017. 2

[21] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. In *Proc. ECCV*, 2020. 2

[22] Dizhong Zhu and William A. P. Smith. Depth from a polarisation + rgb stereo pair. In *Proc. CVPR*, 2019. 2, 6, 7, 8