Depth from a polarisation + RGB stereo pair

Dizhong Zhu and William A. P. Smith
University of York, York, UK
{dz761,william.smith}@york.ac.uk

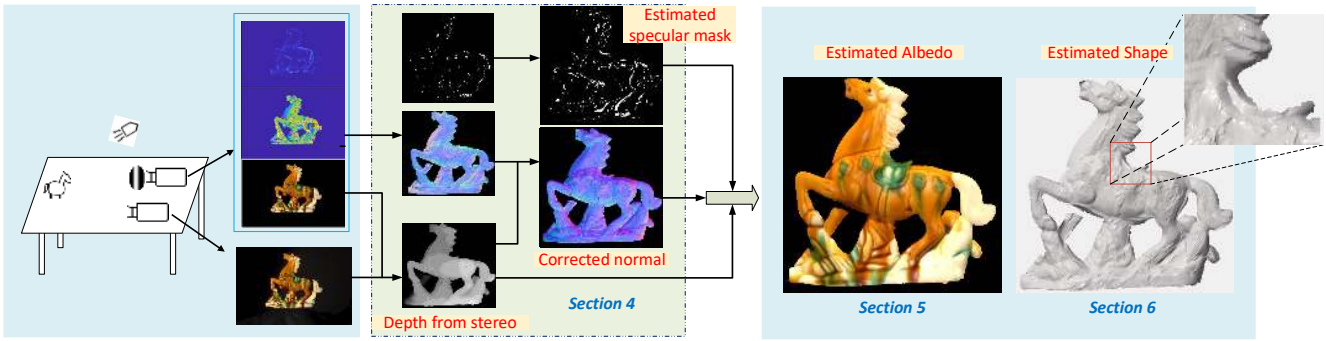

Figure 1: Overview: From a stereo pair of one polarisation image and one RGB image (a) we merge stereo depth with polarisation normals using a higher order graphical model (b) before estimating an albedo map and the final geometry (c).

Abstract

In this paper, we propose a hybrid depth imaging system in which a polarisation camera is augmented by a second image from a standard digital camera. For this modest increase in equipment complexity over conventional shape-from-polarisation, we obtain a number of benefits that enable us to overcome longstanding problems with the polarisation shape cue. The stereo cue provides a depth map which, although coarse, is metrically accurate. This is used as a guide surface for disambiguation of the polarisation surface normal estimates using a higher order graphical model. In turn, these are used to estimate diffuse albedo. By extending a previous shape-from-polarisation method to the perspective case, we show how to compute dense, detailed maps of absolute depth, while retaining a linear formulation. We show that our hybrid method is able to recover dense 3D geometry that is superior to state-of-the-art shape-from-polarisation or two view stereo alone.

1. Introduction

Surface reflection changes the polarisation state of light. By measuring the polarisation state of reflected light, we are able to infer information about the material properties and geometry of the surface. Polarisation is a particularly attractive shape estimation cue because it is dense (surface orientation information is available at every pixel), can be applied to smooth, featureless, glossy surfaces (on which multiview methods would fail to find correspondences) and it can be captured in a single shot (using a polarisation camera). For this reason, the shape-from-polarisation cue has recently been rediscovered and significant progress has been made in the past three years [2, 7, 9, 15, 16, 18, 24, 28, 29, 34].

Recent work has posed shape-from-polarisation in terms of direct estimation of orthographic surface height [27–29]. This is attractive because it halves the degrees of freedom (one height value per pixel rather than two values to represent surface orientation) and avoids the two step process of surface orientation estimation followed by surface integration to obtain a height map. However, polarisation cues do not provide any direct constraints on metric depth, only on local surface orientation. Hence, the surfaces recovered by these methods are globally inaccurate and subject to low frequency distortion. Moreover, the orthographic assumption is practically limiting.

For this reason, in this paper we consider a hybrid setup in which a single polarisation image is augmented by a second image from a standard RGB camera. This provides us with a conventional stereo cue from which we can compute coarse but metrically accurate depth estimates. This serves a number of purposes. First, this provides coarse guide nor-

mals that can be used for initial disambiguation of the polarisation cue. Second, it is used to regularise the final reconstruction, resolving scale ambiguity and reducing low frequency bias. We make a number of novel contributions:

1. Use a higher order graphical model to capture integrability constraints during disambiguation
2. Show how to automatically label pixels as diffuse or specular dominant via our graphical model
3. Show how to incorporate gradient-consistency constraints into albedo estimation
4. Extend the linear formulation of Smith *et al*. [28] to the perspective case, retaining linearity and also including the stereo depth map as a guide surface

Our approach has a number of practical advantages over recent state-of-the-art. Unlike Smith *et al*. [28] we do not assume uniform albedo. Unlike Kadambi *et al*. [15, 16], we do not use a depth (kinect) camera and so our capture environment is not restricted. We compare to these and other relevant state-of-the-art methods and obtain better reconstructions. Compared to [7–9, 33], we only require a single polarisation image.

1.1. Related work

Shape-from-polarisation. Both Miyazaki *et al*. [22] and Atkinson and Hancock [3] used a diffuse polarisation model to estimate surface normals from the phase angle and degree of polarisation. They use a local, greeedy method that propagates from the object boundary assuming global convexity. This is very sensitive to noise, limits applicability to objects with a visible occluding boundary and does not consider integrability. Morel *et al*. [23] took a similar approach but used a specular polarisation model suitable for metallic surfaces. Huynh *et al*. [13] also assumed convexity to disambiguate the polarisation normals.

Polarisation and X. A variety of work seeks to augment polarisation with an additional shape-from-X cue. Huynh *et al*. [14] extended their earlier work to use multispectral measurements to estimate both shape and refractive index. Drbohlav and Sara [10] showed how the Bas-relief ambiguity [6] in uncalibrated photometric stereo could be resolved using polarisation. However, this approach requires a polarised light source. Coarse geometry obtained by multiview space carving [20, 21] has been used to resolve polarisation ambiguities. Kadambi *et al*. [15, 16] combine a single polarisation image with a depth map obtained by an RGBD camera. The depth map is used to disambiguate the normals and provide a base surface for integration. Our approach uses a simpler setup in that it does not require a depth camera. Mahmoud *et al*. [17] and Smith *et al*. [28] augment polarisation with a shape-from-shading cue. The later shows how to solve directly for surface height (i.e. relative depth) by solving a large, sparse linear system of equations. However, they assume constant albedo and orthographic projection - all assumptions that we avoid. Follow-up work showed how to estimate albedo independently [27]. Yu *et al*. [34] take a similar approach but avoid linearising the objective function, instead directly minimising the true nonlinear objective. This allows the use of reflectance and polarisation models of arbitrary complexity. Ngo *et al*. [24] derived constraints that allowed surface normals, light directions and refractive index to be estimated from polarisation images under varying lighting. However, this approach requires at least 4 light directions. Atkinson [2] combine calibrated two source photometric stereo with polarisation phase and resolve ambiguities via a region growing process. Tozza *et al*. [29] generalised [28] to consider two source photo-polarimetric shape estimation. Subsequently, Mecca *et al*. [18] also proposed a differential formulation with a well-posed solution for two light sources.

Multiview Polarisation. Some of the earliest work on polarisation vision used a stereo pair of polarisation measurements to determine the orientation of a plane [30]. Rahmann and Canterakis [26] combined a specular polarisation model with stereo cues. Similarly, Atkinson and Hancock [5] used polarisation normals to segment an object into patches, simplifying stereo matching. Note however that this method is restricted to the case of an object rotating on a turntable with known angle. Stereo polarisation cues have also been used for transparent surface modelling [19]. Berger *et al*. [7] used polarisation stereo for depth estimation of specular scenes. Cui *et al*. [9] incorporate a polarisation phase angle cue into multiview stereo enabling recovery of surface shape in featureless regions. Chen *et al*. [8] provide a theoretical treament of constraints arising from three view polarisation. Yang *et al*. [33] propose a variant of monocular SLAM using polarisation video. All of these methods require multiple polarisation images whereas our proposed approach uses only a single polarisation image augmented by a standard RGB image from a second view.

2. Problem formulation

In this section we list our assumptions and introduce notations, the perspective surface depth representation and basic polarisation theory.

2.1. Assumptions

Our method makes the following assumptions:

- Intrinsic parameters of both cameras known
- Dielectric material with known refractive index
- Distant point light source with known direction
- Diffuse reflectance follows Lambert's law
- Object is smooth, i.e. C^2-continuous (integrable)

These assumptions are all common to previous work. We draw attention to the fact that we do not assume ortho-

graphic projection, known albedo or that pixels have been labelled as diffuse or specular dominant, making our approach more general than previous work.

2.2. Perspective depth representation

Our setup consists of a polarisation camera and an RGB camera. We work in the coordinate system of the polarisation camera and parameterise the surface by the unknown depth function $Z(\mathbf{u})$, where $\mathbf{u} = (x, y)$ is a location in the polarisation image. The 3D coordinate at $\mathbf{u}$ is given by:

$$P(\mathbf{u}) = \begin{bmatrix} \frac{x-x_0}{f} Z(\mathbf{u}) \\ \frac{y-y_0}{f} Z(\mathbf{u}) \\ Z(\mathbf{u}) \end{bmatrix}, \tag{1}$$

where f is the focal length of the polarisation camera in the x and y directions and (x_0, y_0) is the principal point. The direction of the outward pointing surface normal is defined as the cross product of the partial derivatives with respect to x and y [11]:

$$\mathbf{n}(\mathbf{u}) = \begin{bmatrix} -\frac{Z(\mathbf{u}) \cdot Z_x(\mathbf{u})}{f_y} \\ -\frac{Z(\mathbf{u}) \cdot Z_y(\mathbf{u})}{f_x} \\ \frac{x-x_0}{f_x} \frac{Z(\mathbf{u}) \cdot Z_x(\mathbf{u})}{f_y} + \frac{y-y_0}{f_y} \frac{Z(\mathbf{u}) \cdot Z_y(\mathbf{u})}{f_x} + \frac{Z(\mathbf{u})^2}{f_x f_y} \end{bmatrix} \tag{2}$$

where Z_x, Z_y denotes the partial derivative of $Z(\mathbf{u})$ w.r.t. x and y. Note that the magnitude of $\mathbf{n}(\mathbf{u})$ is arbitrary, only its direction is important. For this reason, we can cancel any common factors. In particular, we can divide through by $Z(\mathbf{u})$ to remove quadratic terms and multiply through by $f_x f_y$ to avoid numerical instability caused by division by $f_x f_y$ (which is potentially very large):

$$\mathbf{n}(\mathbf{u}) = \begin{bmatrix} -f_y Z_x(\mathbf{u}) \\ -f_x Z_y(\mathbf{u}) \\ (x - x_0) Z_x(\mathbf{u}) + (y - y_0) Z_y(\mathbf{u}) + Z(\mathbf{u}) \end{bmatrix} \tag{3}$$

We denote by $\bar{\mathbf{n}}(\mathbf{u}) = \mathbf{n}(\mathbf{u}) / \|\mathbf{n}(\mathbf{u})\|$, the unit length surface normal.

The vector pointing towards the viewer from a point on the surface is given by:

$$\mathbf{v}(\mathbf{u}) = -\begin{bmatrix} \frac{x-x_0}{f_x} & \frac{y-y_0}{f_y} & 1 \end{bmatrix}^T / \left\| \begin{bmatrix} \frac{x-x_0}{f_x} & \frac{y-y_0}{f_y} & 1 \end{bmatrix} \right\|. \tag{4}$$

Note that this is independent of surface depth.

2.3. Polarisation theory

When unpolarised light is reflected by a surface it becomes partially polarised [31]. The polarisation information can be estimated by capturing a sequence of images in which a linear polarising filter mounting on camera lens is rotated through a sequence of $P \geq 3$ different angles ϑ_j, $j \in \{1, \ldots, P\}$. The measured intensity at a pixel varies sinusoidally with the polariser angle, it can be written as:

$$i_{\vartheta_j}(\mathbf{u}) = i_{\text{un}}(\mathbf{u}) \left(1 + \rho(\mathbf{u}) \cos(2\vartheta_j - 2\phi(\mathbf{u}))\right). \tag{5}$$

The polarisation image is thus obtained by decomposing the sinusoid at every pixel location into three quantities [31]: the *phase angle*, $\phi(\mathbf{u})$, the *degree of polarisation*, $\rho(\mathbf{u})$, and the *unpolarised intensity*, $i_{\text{un}}(\mathbf{u})$. The parameters of the sinusoid can be estimated from the captured image sequence using non-linear least squares [4], linear methods [13] or via a closed form solution [31] for the specific case of $P = 3$, $\vartheta \in \{0°, 45°, 90°\}$.

A polarisation image provides a constraint on the surface normal direction at each pixel. The exact nature of the constraint depends on the polarisation model used. In this paper we will consider diffuse polarisation, due to subsurface scattering (see [4] for more details), and specular polarisation due to direct reflection.

Degree of polarisation constraint. The degree of diffuse polarisation $\rho_d(\mathbf{u})$ at each point $\mathbf{u}$ can be expressed in terms of the refractive index η and, in the perspective case, the viewing angle $\theta(\mathbf{u}) = \arccos[\bar{\mathbf{n}}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u})] \in [0, \frac{\pi}{2}]$ as follows (Cf. [4]):

$$\rho_d(\mathbf{u}) = \frac{(\eta - 1/\eta)^2 \sin^2\theta(\mathbf{u})}{2 + 2\eta^2 - (\eta + 1/\eta)^2 \sin^2\theta(\mathbf{u}) + 4\cos\theta(\mathbf{u})\sqrt{\eta^2 - \sin^2\theta(\mathbf{u})}}. \tag{6}$$

This expression can be inverted. From the measured degree of polarisation, the viewing angle $\theta(\mathbf{u})$ (and hence one degree of freedom of the surface normal) can be estimated by rewriting (6) [28]. This relates the cosine of the viewing angle to a function, $f(\rho(\mathbf{u}), \eta)$, that depends on the measured degree of polarisation and the refractive index:

$$\cos\theta(\mathbf{u}) = \mathbf{n}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u}) = f(\rho(\mathbf{u}), \eta) = \sqrt{\frac{\eta^4(1-\rho_d^2) + 2\eta^2(2\rho_d^2 + \rho_d - 1) + \rho_d^2 + 2\rho_d - 4\eta^3\rho_d\sqrt{1-\rho_d^2} + 1}{(\rho_d + 1)^2(\eta^4 + 1) + 2\eta^2(3\rho_d^2 + 2\rho_d - 1)}} \tag{7}$$

where we drop the dependency of ρ_d on $(\mathbf{u})$ for brevity. Similarly, the degree of polarisation of a specular reflection is given by:

$$\rho_s(\mathbf{u}) = \frac{2\sin^2\theta(\mathbf{u})\cos\theta(\mathbf{u})\sqrt{\eta^2 - \sin^2\theta(\mathbf{u})}}{\eta^2 - \sin^2\theta(\mathbf{u}) - \eta^2\sin^2\theta(\mathbf{u}) + 2\sin^4\theta(\mathbf{u})}. \tag{8}$$

This expression has two solutions possible solutions for $\theta(\mathbf{u})$ given a measured degree of specular polarisation.

Phase angle constraint The phase angle determines the azimuth angle of the surface normal $\alpha(\mathbf{u}) \in [0, 2\pi]$ up to a 180° ambiguity. For diffuse dominant reflectance this is given by:

$$\alpha(\mathbf{u}) = \phi(\mathbf{u}) \text{ or } (\phi(\mathbf{u}) + \pi), \tag{9}$$

and for specular dominant reflectance by:

$$\alpha(\mathbf{u}) = \phi(\mathbf{u}) \pm \frac{\pi}{2}. \tag{10}$$

Diffuse shading constraint Under the assumption of perfect diffuse reflectance, the unpolarised intensity for diffuse dominant pixels follows Lambert's law:

$$i_d(\mathbf{u}) = \frac{a(\mathbf{u})\mathbf{n}(\mathbf{u}) \cdot \mathbf{s}}{\|\mathbf{n}\|}, \tag{11}$$

where $\mathbf{s} \in \mathbb{R}^3$ is the known distant point source direction and $a(\mathbf{u}) \in [0, 1]$ the diffuse albedo at pixel $\mathbf{u}$.

Diffuse/specular dominance We assume that total reflectance is a mixture of subsurface diffuse reflectance, i_d, and specular surface reflection, i_s (for which we do not assume any particular reflectance model). This means that observed sinusoid is a sum of two sinusoids with a phase difference of $\pi/2$. The resulting sinusoid will be in phase with either the diffuse or specular sinusoid depending on which reflectance "dominates". Concretely, if $i_d\rho_d > i_s\rho_s$ then the pixel is diffuse dominant and we neglect specular reflectance, i.e. we assume $i_{\text{un}} = i_d$.

3. Overview of method

Our proposed method comprises the following steps:

1. Estimate the disparity from stereo images and reconstruct a coarse depth map by known camera matrix.
2. Compute guide surface normals by taking the gradient of the coarse depth map.
3. Use guide surface normal to disambiguate the polarisation normals via a higher order graphical model.
4. Estimate diffuse albedo from disambiguated polarisation normals.
5. Linearly estimate perspective depth from polarisation using coarse depth map as a constraint.

Our pipeline is illustrated in Fig. 1 and each step is described in detail in the following sections.

4. Integrability-based disambiguation with a higher order graphical model

The constraints in Section 2.3 restrict the surface normal at a pixel to six possible directions. If the pixel is diffuse dominant, then the viewing angle is uniquely determined by the degree of polarisation and the azimuth angle restricted to two possibilities by the phase angle, leading to two possible normal directions. If the pixel is specular dominant, the degree of polarisation restricts the viewing angle to two possibilities, with the azimuth again also restricted to two, given four possible normal directions in total. Previous work [15, 28] assumes that the labelling of pixels as specular or diffuse dominant is known in advance. We do not assume that the labels are known and propose an initial resolution of this six-way ambiguity using a higher order graphical model. The motivation for using a higher order model is that a ternary potential can measure deviation from integrability.

We set up an energy cost function to be mimised w.r.t. the surface normal as follows:

$$\begin{aligned} E(\mathbf{n}(\mathbf{u})) = &\sum_{\mathbf{u}\in\nu} \Phi(\mathbf{n}(\mathbf{u})) + \sum_{(\mathbf{u},\mathbf{v})\in\mathcal{N}} \varphi(L(\mathbf{u}), L(\mathbf{v})) \\ &+ \sum_{(\mathbf{u},\mathbf{v},\mathbf{w})\in\mathcal{T}} \Psi(\mathbf{n}(\mathbf{u}), \mathbf{n}(\mathbf{v}), \mathbf{n}(\mathbf{w})) \end{aligned} \tag{12}$$

Here ν corresponds to all foreground pixels, $\mathcal{N}$ is the set of adjacent pixels and $\mathcal{T}$ is the set of pixel triplets $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ where $\mathbf{u} = (x, y)$, $\mathbf{v} = (x + 1, y)$ and $\mathbf{w} = (x, y + 1)$. Before further explaining the energy terms, let us clarify two important elements that will be used in following. **1)**. The stereo setup produces a coarse depth map by computing the disparity from the camera pair. We use the semi-global matching method [12] to compute the disparity and reconstruct a depth map with the camera matrices, as displayed in Figure 2(a). Thus its surface normal can be computed by simply taking the forward difference on the coarse depth map. We denote these surface normal by $\hat{\mathbf{n}}$ where they are noisy as shown in Figure 2(b). **2)**. We make a rough initial estimate of the specular/diffuse dominant pixel labelling, L. We simply set $L(\mathbf{u}) = 1$ if the measured intensity is saturated (Figure 2(c)). L will be subsequently updated (Figure 2(f)).

Unary cost The unary term aims to minimise the angle between $\mathbf{n}(\mathbf{u})$ and $\hat{\mathbf{n}}(\mathbf{u})$, where $\mathbf{n}(\mathbf{u})$ has up to six solutions. We denote the first two solutions from diffuse component in $\mathcal{D}$ and the rest from specular component in $\mathcal{S}$. We also take account the initial specular mask L i.e. Where the diffuse normal will be assigned to low probability if its corresponding specular mask equal to one. The unary cost can be written as

$$\Phi(\mathbf{n}(\mathbf{u})) = \begin{cases} k \cdot f(\mathbf{u}) & \text{if } (L(\mathbf{u}) = 1, \mathbf{n}(\mathbf{u}) \in \mathcal{D}) \text{ or } (L(\mathbf{u}) = 0, \mathbf{n}(\mathbf{u}) \in \mathcal{S}) \\ f(\mathbf{u}) & \text{if } (L(\mathbf{u}) = 0, \mathbf{n}(\mathbf{u}) \in \mathcal{D}) \text{ or } (L(\mathbf{u}) = 1, \mathbf{n}(\mathbf{u}) \in \mathcal{S}) \end{cases}$$

where $f(\mathbf{u})$ depends on the cosine of the angle between $\mathbf{n}(\mathbf{u})$ and $\hat{\mathbf{n}}(\mathbf{u})$ and is defined as

$$f(\mathbf{u}) = \exp(-\mathbf{n}(\mathbf{u}) \cdot \hat{\mathbf{n}}(\mathbf{u})). \tag{13}$$

The parameter $k < 1$ penalises surface normal disambiguations that are not consistent with the corresponding specular mask. We set $k = 0.1$ in our experiments.

Pairwise cost We encourage pairwise pixels in $\mathcal{N}$ to have similar diffuse or specular labels and penalise where the labels changed. We define

$$\varphi(L(\mathbf{u}), L(\mathbf{v})) = |L(\mathbf{u}) - L(\mathbf{v})|. \tag{14}$$

Ternary cost In order to encourage the disambiguated surface normals to satisfy the integrability constraint, we use a ternary cost to measure deviation from integrability. For an integrable surface, the mixed second order partial derivatives on the gradient field should be equal [25]. Specifically, $\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}$. Where p, q are the partial derivatives in the x and y direction respectively. The surface gradient is directly linked to the surface normal by

$$p(\mathbf{u}) = -n_x(\mathbf{u})/n_z(\mathbf{u}) \quad \text{and} \quad q(\mathbf{u}) = -n_y(\mathbf{u})/n_z(\mathbf{u})$$

We take three-pixel neighbourhoods $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ to compute the gradient of p, q, where

$$\frac{\partial p(\mathbf{u})}{\partial y} = p(\mathbf{w}) - p(\mathbf{u}) \, , \, \frac{\partial q(\mathbf{u})}{\partial x} = q(\mathbf{v}) - q(\mathbf{u})$$

In reality, due to noise and the discretisation to the pixel grid, the gradient field may not have exactly zero curl, but we seek the surface normals that give minimum curl values. Hence, the ternary cost is defined by:

$$\Psi(\mathbf{n}(\mathbf{u}), \mathbf{n}(\mathbf{v}), \mathbf{n}(\mathbf{w})) = \|p(\mathbf{w}) - p(\mathbf{u}) - (q(\mathbf{v}) - q(\mathbf{u}))\| \, .$$

Graphical model optimisation We use higher order belief-propagation to minimise (12) as implemented in the OpenGM toolbox [1]. The optimum surface normal $\mathbf{n}'$ will be labeled as one of the six possible disambiguations and we update our specular mask L according to:

$$L(\mathbf{u}) = \begin{cases} 0 & \text{if } \mathbf{n}(\mathbf{u}) \in \mathcal{D} \\ 1 & \text{if } \mathbf{n}(\mathbf{u}) \in \mathcal{S} \end{cases} .$$

The surface normals that result from this disambiguation process are still noisy (they use only local information) and may be subject to low frequency bias meaning that integrating them into a depth map does not yield good results. Hence, in Section 6 we solve globally for depth, using the stereo depth map as a guide to remove low frequency bias.

5. Albedo estimation with gradient consistency

We now use the surface normals estimated by the graphical model optimisation to compute an albedo map. In principal, the albedo can be computed from these normals and the unpolarised intensity simply by rearranging (11). However, this purely local estimation is unstable and noise in the normals leads to artefacts in the estimated albedo map. We propose a simple but very effective regularisation to resolve this problem. We encourage the gradient of the estimated albedo map to be similar to the gradient of the unpolarised intensities at points where the intensity gradient is above a threshold and zero elsewhere. In other words, we encourage the albedo gradients to be sparse and hence the albedo piecewise uniform.

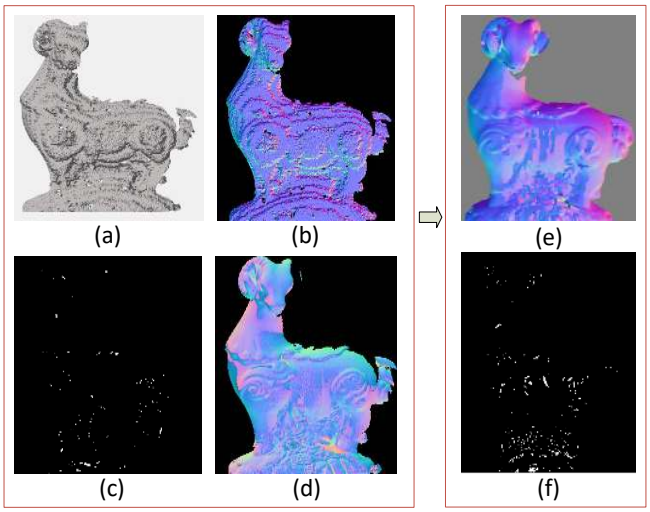

Figure 2: (a) Depth map from disparity map. (b) Guide surface normal from stereo depth map. (c) Preset specular mask. (d) One possible polarisation normal. (e) The corrected normal via our graphical model. (f) The updated specular mask via graphical model.

The estimated albedo minimises the following energy function

$$E(\mathbf{u}) = E_{Lamb}(\mathbf{u}) + \lambda_I E_{smooth}(\mathbf{u}). \tag{15}$$

The first term penalises the difference between rendered Lambertian intensity and estimated unpolarised intensity:

$$E_{Lamb}(\mathbf{u}) = \left\| a(\mathbf{u})\mathbf{n}' \cdot (\mathbf{u})\mathbf{s} - I_d(\mathbf{u}) \right\|_2^2 \tag{16}$$

where I_d is diffuse dominant pixels from the estimated unpolarisation intensity, α represents a pixel-wise albedo map, $\mathbf{n}'$ is the optimum surface normal map from the previous section and $\mathbf{s}$ the light source. We can easily choose the diffuse pixels by excluding the specular mask where $L(\mathbf{u}) = 1$.

The second term penalises the difference between the estimated albedo gradient and the sparsified unpolarised intensity gradient. We denote the neighbour of $\mathbf{u}$ in x direction with v and y direction with w, thus the smooth term can be written as

$$\begin{aligned} E_{smooth}(\mathbf{u}) = &\|a(\mathbf{u}) - a(\mathbf{v}) - g(I_d(\mathbf{u}) - I_d(\mathbf{v}))\| \\ &+ \|a(\mathbf{u}) - a(\mathbf{w}) - g(I_d(\mathbf{u}) - I_d(\mathbf{w}))\| \end{aligned} \tag{17}$$

where $g(.)$ is a threshold function that returns 0 if the input is $< t$, otherwise it returns the input albedo map only contains values on the diffuse pixels, we fill the hole on specular pixels with nearest neighbour method. In Figure 3 we see how the smoothness term affects the estimated albedo map and depth.

6. Linear perspective depth from polarisation

Finally, with albedo known and coarse depth values from two view stereo, we are ready to estimate dense depth from polarisation. We generalise a perspective camera model from Smith *et al.* [28], note that it differs via the use of the coarse depth values

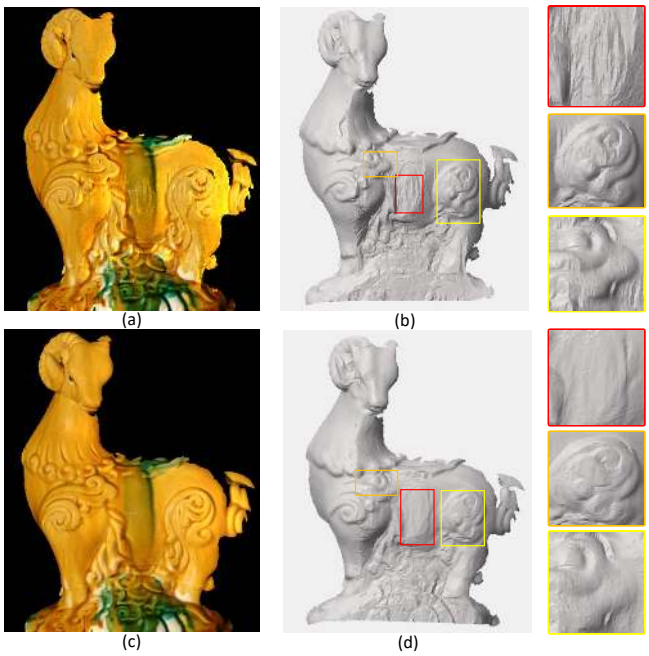

Figure 3: (a)/(c) Estimated albedo (b)/(d) Estimated geometry. First row: $\lambda_I = 0$, second row: $\lambda_I = 3$. Comparing (a) and (c), the albedo map becomes smoother. Comparing (b) and (d), the red rectangle region becomes smoother but while fine detailis largely preserved.

and optimum normal from section 4. The fact that we estimate metric depth rather than relative height. As in [28], we express polarisation and shading constraints in the form of a large, sparse linear system in the unknown depth values, meaning the method is very efficient and guaranteed to attain the globally optimal solution.

Phase angle constraint. The first constraint encourages the recovered surface normal to satisfy equation (10). Following [28], the projection of the surface normal into the image plane (n_x, n_y) should be collinear with the phase angle vector. We seperate pixels into diffuse dominant and specular dominant with the help of specular mask L. The phase angle constraint for diffuse dominant pixels and specular dominant pixels are represented in first row and second row respectively in this matrix form:

$$\begin{bmatrix} \cos(\phi(\mathbf{u})) & -\sin(\phi(\mathbf{u})) & 0 \\ \cos(\phi(\mathbf{u}) + \frac{\pi}{2}) & -\sin(\phi(\mathbf{u}) + \frac{\pi}{2}) & 0 \end{bmatrix} \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = 0 \quad (18)$$

Shading/polarisation ratio constraint. Recall that the viewing angle is the angle between the surface normal and the viewer direction. Making the normalisation factor of the surface normal explicit, we can write $\cos(\theta_r(\mathbf{u})) = \frac{\mathbf{n}(\mathbf{u})\cdot\mathbf{v}(\mathbf{u})}{\|\mathbf{n}(\mathbf{u})\|}$. By isolating the normalisation factor we arrive at:

$$\|\mathbf{n}(\mathbf{u})\| = \frac{\mathbf{n}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u})}{\cos(\theta_r(\mathbf{u}))}. \quad (19)$$

Substituting this into (11) we obtain:

$$\frac{\mathbf{n}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u})}{\cos(\theta_r(\mathbf{u}))} = \frac{a(\mathbf{u})\mathbf{n}(\mathbf{u}) \cdot \mathbf{s}}{i_{\text{un}}(\mathbf{u})} \quad (20)$$

Notice that our shading constraint only submit on the diffuse pixels. So we choose the pixels $\mathbf{u} \in \mathcal{D}$ where $L(\mathbf{u}) = 0$. Unlike [28], the perspective model means that the view vectors depend on pixel locations. Now we can reformulate the equation into a compact matrix form with respect to the surface normal:

$$\begin{bmatrix} s_x \cdot a(\mathbf{u}) \cos\theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u}) v_x(\mathbf{u}) \\ s_y \cdot a(\mathbf{u}) \cos\theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u}) v_y(\mathbf{u}) \\ s_z \cdot a(\mathbf{u}) \cos\theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u}) v_z(\mathbf{u}) \end{bmatrix}^T \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = 0 \quad (21)$$

Surface normal constraint. We also encourage our recovered surface normal should co-linear with the optimised normal n' from section 4 where their cross product is a zero vector. It can be formalised in following manner

$$\begin{bmatrix} 0 & -n'_z(\mathbf{u}) & n'_y(\mathbf{u}) \\ n'_z(\mathbf{u}) & 0 & -n'_x(\mathbf{u}) \\ -n'_y(\mathbf{u}) & n'_x(\mathbf{u}) & 0 \end{bmatrix} \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (22)$$

Global linear depth estimation. The relationship between the surface normal and depth under perspective viewing is given by (3). We can arrive at a linear relationship between the constraints described above and the unknown depth.

We first extend (3) to the whole image. Consider an image with N foreground pixels whose unknown depth values are vectorised in $\mathbf{Z} \in \mathbb{R}^N$. The surface normal direction (unnormalised) can be computed for all pixels with:

$$\mathbf{NZ} = \begin{bmatrix} n_x(\mathbf{u}_1) \\ \dots \\ n_x(\mathbf{u}_N) \\ n_y(\mathbf{u}_1) \\ \dots \\ n_y(\mathbf{u}_N) \\ n_z(\mathbf{u}_1) \\ \dots \\ n_z(\mathbf{u}_N) \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} -f_y\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -f_x\mathbf{I} & \mathbf{0} \\ \mathbf{X} & \mathbf{Y} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{I} \end{bmatrix} \quad (23)$$

where $\mathbf{X} = \text{diag}(x_1 - x_0, \dots, x_N - x_0)$ and $\mathbf{Y} = \text{diag}(y_1 - y_0, \dots, y_N - y_0)$. $\mathbf{D}_x, \mathbf{D}_y \in \mathbb{R}^{N \times N}$ compute finite difference approximations to the derivative of Z in the x and y directions respectively. In practice, we use smoothed central difference approximations where possible, reverting to central or forward/backward differences where all neighbours are not available. Hence $\mathbf{D}_x, \mathbf{D}_y$ have at most six non-zero values per row.

Combining (23) with (18), (21) and (22) leads to equations that are linear in depth. We now combine these equations into a large linear system of equations for the whole image. Of the N foreground pixels we divide these into diffuse and specular pixels according to the mask L. We denote the number of diffuse pixels with N_D and specular

	Coarse depth	Shading	Polarisation
Stereo [12]	✓		
Smith-2016 [28]		✓	✓
Smith-2018 [27]		✓	✓
Polarised 3D [15]	✓		✓
Wu-2014 [32]	✓	✓	
Proposed	✓	✓	✓

Table 1: Summary of the different method

with N_S. We now form a linear system in the vector of unknown depth values, $\mathbf{Z}$:

$$\begin{bmatrix} \lambda \mathbf{AN} \\ \mathbf{W} \end{bmatrix} \mathbf{Z} = \begin{bmatrix} \mathbf{0}_{4N+N_D} \\ Z_{\text{guide}}(\mathbf{u}_1) \\ \vdots \\ Z_{\text{guide}}(\mathbf{u}_N) \end{bmatrix} \tag{24}$$

where $Z_{\text{guide}}(\mathbf{u}_i)$ are the stereo depth values from Section 4 and $\mathbf{W} \in \mathbb{R}^{K \times N}$ performs a sparse indices matrix of $\boldsymbol{Z}$ at positions $(x_1, y_1), \ldots, (x_K, y_K)$. $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathbf{0}_{4N+N_D}$ is the zero vector of length $4N+N_D$. $\mathbf{A}$ has $4N+N_D$ rows, $3N$ columns and is sparse. Each row evaluates one equation of the form of (18), (21) and (22). $\lambda > 0$ is a weight which trades off the influence of the guide depth values against satisfaction of the polarisation constraints. We then solve (24) in a least squares sense using sparse linear least squares.

7. Experimental results

We present experimental results on both synthetic and real data. We compare our method against [12, 15, 27, 28, 32], the differences are summarised in Table 1. We set $\lambda_I = 1, \lambda = 1$ and $t = 0.01$ through our experiments. Note that the source code for [15] is not available so we are only able to compare against a single result provided by the authors. Similarly, real image results for [32] were provided by the author running the implementation for us. Whereas [12,27,28] are open sourced and we compare quantitatively. For synthetic data, we render images of the Stanford bunny with Blinn-Phong reflectance with varying albedo texture using the pinhole camera model, as shown in Figure 4 (left). The texture map is from [35]. We simulate the effect of polarisation according to (5) by setting refractive index value to 1.4 and corrupt the polarisation image and second camera intensity by adding Gaussian noise with zero mean and standard deviation σ. The metric ground truth of the depth map is range between 72.33mm to 90.09mm.

In Figure 4 we show the estimated albedo map of the synthetic data and compare with [27]. In Table 2 we show the mean absolute error in the surface depth (in millimetre) and mean angular error (in degrees) in the surface normals. We include comparison with the initial stereo depth [12]

Method	$\sigma = 0\%$		$\sigma = 0.5\%$		$\sigma = 1\%$	
	Depth (mm)	Normal (deg)	Depth (mm)	Normal (deg)	Depth (mm)	Normal (deg)
[12]	0.49	38.151	0.49	39.78	0.49	39.67
[28]	10.68	30.38	85.91	29.966	113.80	32.03
[27]	12.02	22.53	36.08	26.54	40.88	28.54
Prop	0.29	9.799	0.30	9.86	0.31	14.03

Table 2: Mean absolute difference in depth and mean angular surface normal errors on synthetic data. For [27, 28] methods reconstructed the depth up to scale we compute the optimum scale to align with the ground truth depth map.

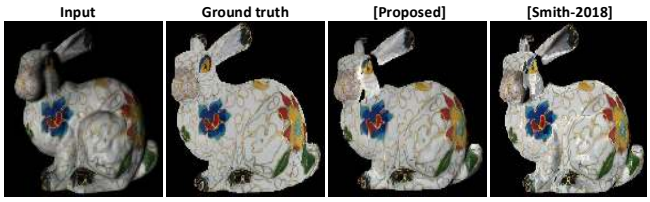

Figure 4: Albedo estimates on synthetic data.

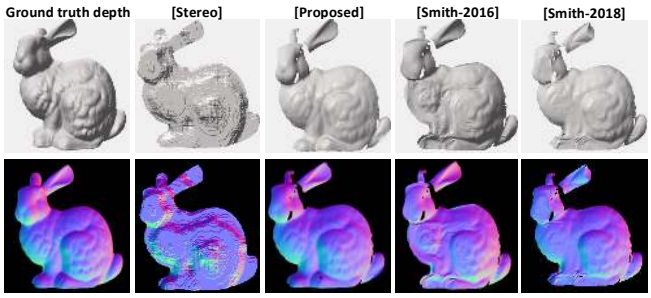

Figure 5: Qualitative shape estimation results on synthetic data with comparison with [28]

and state-of-the-art polarisation methods [27,28]. In Figure 5 we display the qualitative results of this experiment.

Next we show results on a dataset of real images. The first dataset is from [15]. Although the depth here is provided by a Kinect sensor, not stereo, our graphical model optimisation in Section 4 can take any source of depth map. In this case we replace the depth map with the Kinect one and keep the rest of the process identical when we evaluate the data. The comparison can be viewed in Figure 7 where we show that our proposed result can give more details on the reconstruction. In this experiment, we estimate the light source direction using [28].

We then show results on our own collected data. We place the polarisation and RGB cameras with parallel image planes and the RGB camera shifted 5cm along the x axis relative to the polarisation camera as illustrated in Figure 1. We compare our method with [32] directly performed by the author. In Figure 6 we show qualitative results for three objects with glossy reflectance and varying albedo.

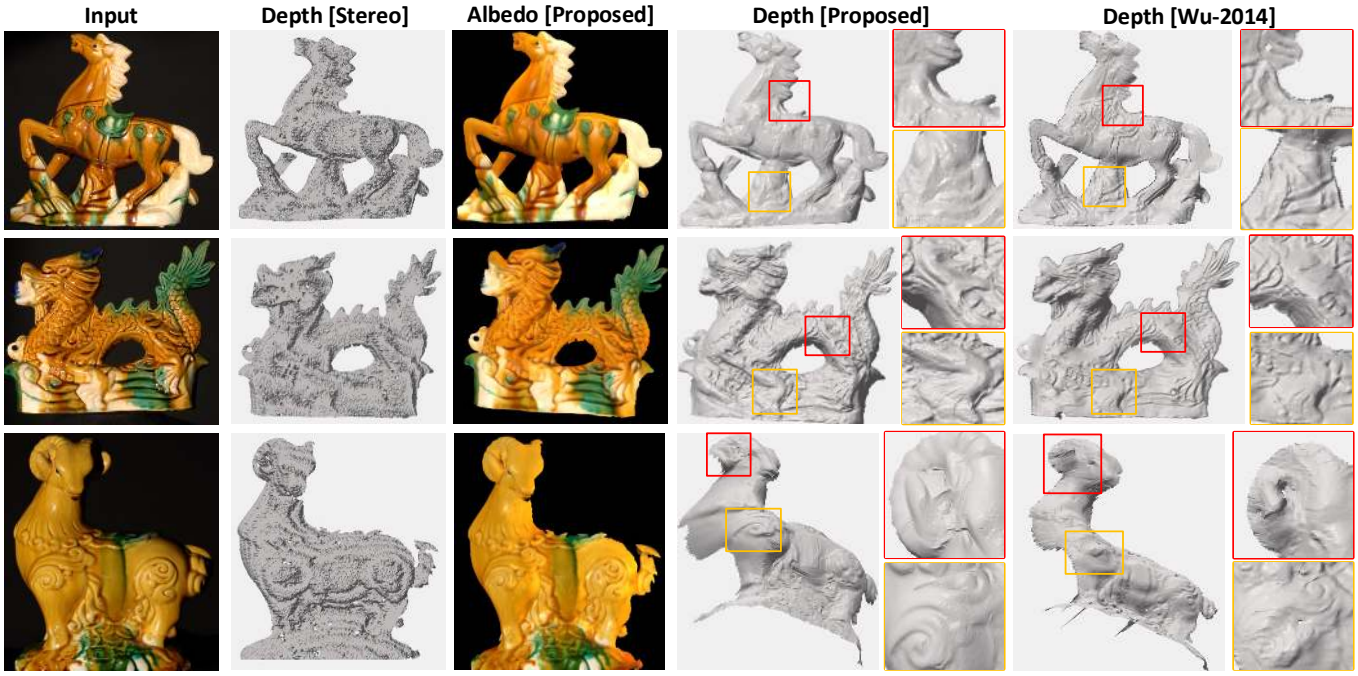

Figure 6: We show our results on complex object. From left to right we show an image from the input sequence; Depth from stereo reconstruction [12]; Our proposed estimated albedo map and the estimated depth. Depth estimation by [32].

Our method gives improved detail (see insets) but also more stable overall shape (see third row). Notice that in this experiment we calibrated the light source in advance with a uniform albedo sphere using method in [28].

8. Conclusions

In this paper we have proposed a method for estimating dense depth and albedo maps for glossy, dielectric objects with varying albedo. We do so using a hybrid imaging system in which a polarisation image is augmented by a second view from a standard RGB camera. We avoid assumptions common to recent methods (constant albedo, orthographic projection) and reduce low frequency distortion in the recovered depth maps through the stereo cue.

Since we rely on stereo, our method does not work well on textureless objects. However, note that our method works equally well with a Kinect depth map as the result shows in Figure 7. We also assume the refractive index is known in our framework. It could be potentially measured given a sufficiently accurate guide depth map. Although our stereo setup cannot provide this, it could potentially be provided by photometric stereo or multiview stereo. There are many exciting possibilities for extending this work. The lighting, reflectance and polarisation models could be generalised. In particular, a more comprehensive model of mixed specular/diffuse reflectance and polarisation would be beneficial. Our linear approach is efficient and does not require initialisation, but it may be useful to subsequently perform a nonlinear optimisation over all unknowns (depth, albedo, refractive index) simultaneously such that the true underlying objective function can be minimised (taking inspiration from [34]).

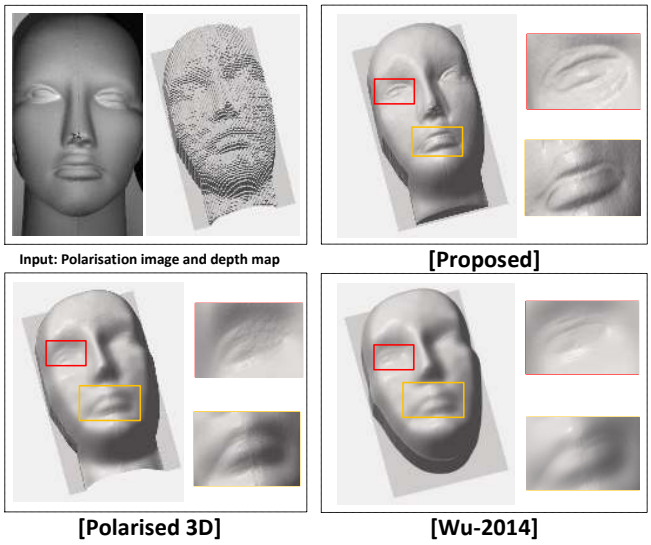

Figure 7: Comparison on [15] dataset. Top-left: One of the polarisation intensity images and Kinect depth map. Top-right: our result. Bottom-Left: [15]. Bottom-Right: [32].

References

[1] Bjoern Andres, Thorsten Beier, and Jörg H Kappes. Opengm: A c++ library for discrete graphical models. *arXiv preprint arXiv:1206.0111*, 2012. 5

[2] Gary A Atkinson. Polarisation photometric stereo. *Comput. Vis. Image Underst.*, 2017. 1, 2

[3] 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

[4] Gary A. Atkinson and Edwin R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image processing*, 15(6):1653–1664, 2006. 3

[5] Gary A Atkinson and Edwin R Hancock. Shape estimation using polarization and shading from two views. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11):2001–2017, 2007. 2

[6] P. N. Belhumeur, D. J. Kriegman, and A.L. Yuille. The Bas-relief ambiguity. *Int. J. Comput. Vision*, 35(1):33–44, 1999. 2

[7] K. Berger, R. Voorhies, and L. H. Matthies. Depth from stereo polarization in specular scenes for urban robotics. In *Proc. ICRA*, pages 1966–1973, 2017. 1, 2

[8] Lixiong Chen, Yinqiang Zheng, Art Subpa-asa, and Imari Sato. Polarimetric three-view geometry. In *Proc. ECCV*, pages 20–36, 2018. 2

[9] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *Proc. CVPR*, pages 1558–1567, 2017. 1, 2

[10] Ondřej Drbohlav and Radim Šára. Unambiguous determination of shape from photometric stereo with unknown light sources. In *Proc. ICCV*, pages 581–586, 2001. 2

[11] Gottfried Graber, Jonathan Balzer, Stefano Soatto, and Thomas Pock. Efficient minimal-surface regularization of perspective depth maps in variational stereo. In *Proc. CVPR*, pages 511–520, 2015. 3

[12] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. CVPR*, volume 2, pages 807–814. IEEE, 2005. 4, 7, 8

[13] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin Hancock. Shape and refractive index recovery from single-view polarisation images. In *Proc. CVPR*, pages 1229–1236, 2010. 2, 3

[14] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin R Hancock. Shape and refractive index from single-view spectro-polarimetric images. *Int. J. Comput. Vision*, 101(1):64–94, 2013. 2

[15] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3D: High-quality depth sensing with polarization cues. In *Proc. ICCV*, 2015. 1, 2, 4, 7, 8

[16] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Depth sensing using geometrically constrained polarization normals. *Int. J. Comput. Vision*, 2017. 1, 2

[17] Ali H Mahmoud, Moumen T El-Melegy, and Aly A Farag. Direct method for shape recovery from polarization and shading. In *Proc. ICIP*, pages 1769–1772, 2012. 2

[18] Roberto Mecca, Fotios Logothetis, and Roberto Cipolla. A differential approach to shape from polarization. In *Proc. BMVC*, 2017. 1, 2

[19] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):73–82, 2004. 2

[20] Daisuke Miyazaki, Takuya Shigetomi, Masashi Baba, Ryo Furukawa, Shinsaku Hiura, and Naoki Asada. Polarization-based surface normal estimation of black specular objects from multiple viewpoints. In *3DIMPVT*, pages 104–111, 2012. 2

[21] 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):041303–041303, 2017. 2

[22] Daisuke Miyazaki, Robby T Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *Proc. ICCV*, pages 982–987, 2003. 2

[23] Olivier Morel, Fabrice Meriaudeau, Christophe Stolz, and Patrick Gorria. Polarization imaging applied to 3D reconstruction of specular metallic surfaces. In *Proc. EI 2005*, pages 178–186, 2005. 2

[24] T. T. Ngo, H. Nagahara, and R. Taniguchi. Shape and light directions from shading and polarization. In *Proc. CVPR*, pages 2310–2318, 2015. 1, 2

[25] Nemanja Petrovic, Ira Cohen, Brendan J Frey, Ralf Koetter, and Thomas S Huang. Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models. In *Proc. CVPR*, volume 1, pages I–I. IEEE, 2001. 5

[26] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proc. CVPR*, 2001. 2

[27] William Smith, Ravi Ramamoorthi, and Silvia Tozza. Height-from-polarisation with unknown lighting or albedo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2019. 1, 2, 7

[28] William AP Smith, Ravi Ramamoorthi, and Silvia Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *European Conference on Computer Vision*, pages 109–125. Springer, 2016. 1, 2, 3, 4, 5, 6, 7, 8

[29] Silvia Tozza, William AP Smith, Dizhong Zhu, Ravi Ramamoorthi, and Edwin R Hancock. Linear differential constraints for photo-polarimetric height estimation. In *Proc. ICCV*, 2017. 1, 2

[30] Lawrence B Wolff. Surface orientation from two camera stereo with polarizers. In *Proc. SPIE Conf. Optics, Illumination, and Image Sensing for Machine Vision IV*, volume 1194, pages 287–298, 1990. 2

[31] L. B. Wolff. Polarization vision: a new sensory approach to image understanding. *Image Vision Comput.*, 15(2):81–93, 1997. 3

[32] Chenglei Wu, Michael Zollhöfer, Matthias Nießner, Marc Stamminger, Shahram Izadi, and Christian Theobalt. Real-time shading-based refinement for consumer depth cameras. *ACM Transactions on Graphics (ToG)*, 33(6):200, 2014. 7, 8

[33] Luwei Yang, Feitong Tan, Ao Li, Zhaopeng Cui, Yasutaka Furukawa, and Ping Tan. Polarimetric dense monocular slam. In *Proc. CVPR*, pages 3857–3866, 2018. 2

[34] Y. Yu, D. Zhu, and W. A. P. Smith. Shape-from-polarisation: a nonlinear least squares approach. In *Proc. ICCV Workshop on Color and Photometry in Computer Vision*, 2017. 1, 2, 8

[35] Kun Zhou, Xi Wang, Yiying Tong, Mathieu Desbrun, Baining Guo, and Heung-Yeung Shum. Texturemontage: Seamless texturing of arbitrary surfaces from multiple images. *ACM Transactions on Graphics*, 24(3):1148–1155, 2005. 7