Linear depth estimation from an uncalibrated, monocular polarisation image

William A. P. Smith[1], Ravi Ramamoorthi[2] and Silvia Tozza[3]

[1]University of York   [2]UC San Diego   [3]Sapienza - Università di Roma

william.smith@york.ac.uk   ravir@cs.ucsd.edu   tozza@mat.uniroma1.it

Abstract. We present a method for estimating surface height directly from a single polarisation image simply by solving a large, sparse system of linear equations. To do so, we show how to express polarisation constraints as equations that are linear in the unknown depth. The ambiguity in the surface normal azimuth angle is resolved globally when the optimal surface height is reconstructed. Our method is applicable to objects with uniform albedo exhibiting diffuse and specular reflectance. We extend it to an uncalibrated scenario by demonstrating that the illumination (point source or first/second order spherical harmonics) can be estimated from the polarisation image, up to a binary convex/concave ambiguity. We believe that our method is the first monocular, passive shape-from-x technique that enables well-posed depth estimation with only a single, uncalibrated illumination condition. We present results on glossy objects, including in uncontrolled, outdoor illumination.

Keywords: polarisation, shape-from-x, Bas-relief ambiguity

1 Introduction

When unpolarised light is reflected by a surface it becomes partially polarised [1]. The degree to which the reflected light is polarised conveys information about the surface orientation and, therefore, provides a cue for shape recovery. There are a number of attractive properties to this 'shape-from-polarisation' (SfP) cue. It requires only a single viewpoint and illumination environment, it is invariant to illumination and surface albedo and it provides information about both the zenith and azimuth angle of the surface normal. Like photometric stereo, shape estimates are dense (the surface normal is estimated at every pixel so resolution is limited only by the sensor) and, since it does not rely on detecting or matching features, it is applicable to smooth, featureless surfaces.

However, there are a number of drawbacks to using SfP in a practical setting. The polarisation cue alone provides only ambiguous estimates of surface orientation. Hence, previous work focuses on developing heuristics to locally disambiguate the surface normals. Even having done so, surface orientation is only a 2.5D shape cue and so the estimated normal field must be integrated in order to recover surface depth [2] or used to refine a depth map captured using other cues [3]. This two step approach of disambiguation followed by integration means
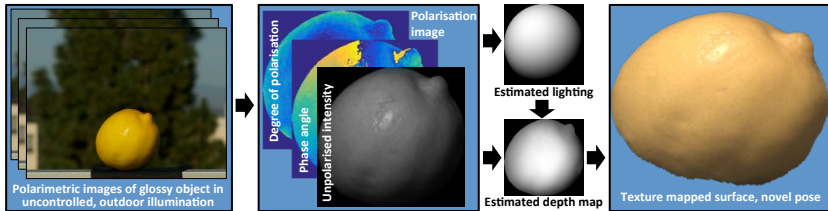
Fig. 1: Overview of method: from a single polarisation image of a homogenous, glossy object in uncontrolled (possibly outdoor) illumination, we estimate lighting and compute depth directly.

that the surface integrability constraint is not enforced during disambiguation and also that errors accumulate over the two steps. In this paper, we propose a SfP method (see Fig. 1 for an overview) with the following novel ingredients:

1. In contrast to prior work, we compute SfP in the depth, as opposed to the surface normal, domain. Instead of disambiguating the polarisation normals, we defer resolution of the ambiguity until surface height is computed. To do so, we express the azimuthal ambiguity as a collinearity condition that is satisfied by either interpretation of the polarisation measurements.
2. We express polarisation and shading constraints as linear equations in the unknown depth enabling efficient and globally optimal depth estimation.
3. We use a novel hybrid diffuse/specular polarisation and shading model, allowing us to handle glossy surfaces.
4. We show that illumination can be determined from the ambiguous normals and unpolarised intensity up to a binary ambiguity (a particular generalised Bas-relief [4] transformation: the convex/concave ambiguity). This means that our method can be applied in an uncalibrated scenario and we consider both point source and 1st/2nd order spherical harmonic (SH) illumination.

1.1 Related work

Previous SfP methods can be categorised into two groups, those that: 1. use only a single polarisation image, and 2. combine a polarisation image with additional cues. The former group (of which our method is a member) can be considered 'single shot' methods (single shot capture devices exist using polarising beamsplitters [25] or CMOS sensors with micropolarising filters [26]). More commonly, a polarisation image is obtained by capturing a sequence of images in which a linear polarising filter is rotated in front of the camera (possibly with unknown rotation angles [5]). SfP methods can also be classified according to the polarisation model used (dielectric versus metal, diffuse, specular or hybrid models) and if they compute shape in the surface normal or surface height domain.

Single polarisation image The earliest work focussed on capture, decomposition and visualisation of polarisation images [6]. Both Miyazaki et al. [2] and Atkinson and Hancock [7] used a diffuse polarisation model and, under an assumption of object convexity, propagate disambiguation of the surface normal inwards from the boundary. This greedy approach will not produce globally

optimal results, limits application to objects with a visible occluding boundary and does not consider integrability. Morel et al. [8] took a similar approach but used a specular polarisation model suitable for metallic surfaces. Huynh et al. [9] also assumed convexity to disambiguate the polarisation normals; however, their approach also estimates refractive index. As in our method, Mahmoud et al. [10] exploited the unpolarised intensity via a shading cue. Assuming Lambertian reflectance and known lighting direction and albedo, the surface normal ambiguity can be resolved. We avoid all of these assumptions and, by strictly enforcing integrability, improve robustness to noise.

Polarisation with additional cues Rahmann and Canterakis [11] combined a specular polarisation model with stereo cues. Similarly, Atkinson and Hancock [12] used polarisation normals to segment an object into patches, simplifying stereo matching. Huynh et al. [13] extended their earlier work to use multispectral measurements to estimate both shape and refractive index. There have been a number of attempts to augment polarisation cues with calibrated, Lambertian photometric stereo, e.g. [14]. Drbohlav and Sara [15] showed how the Bas-relief ambiguity [4] in uncalibrated photometric stereo could be resolved using polarisation. However, this approach requires a polarised light source. Recently, Ngo et al. [16] 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 in contrast to the single direction required by our method. Very recently, Kadambi et al. [3] proposed an interesting approach in which a single polarisation image is combined 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.

2 Problem formulation and polarisation theory

We make the following assumptions (more general than much previous work in the area): **1.** Dielectric (i.e. non-metallic) material with uniform (but unknown) albedo. **2.** Orthographic projection. **3.** The refractive index of the surface is known, though dependency on this quantity is weak and we fix it to a constant value for all of our experiments. **4.** Pixels can be classified as either diffuse dominant or specular dominant. **5.** The object surface is smooth (i.e. C^2 continuous).
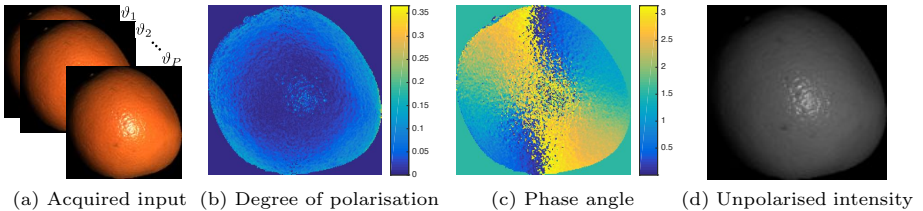
We parameterise surface height by the function $z(\mathbf{u})$, where $\mathbf{u} = (x, y)$ is an image point. Foreground pixels belonging to the surface are represented by the set $\mathcal{F}$, $|\mathcal{F}| = K$. The unit surface normal can be expressed in spherical world coordinates as:

$$\mathbf{n}(\mathbf{u}) = [n_x(\mathbf{u})\ n_y(\mathbf{u})\ n_z(\mathbf{u})]^T = [\sin\alpha(\mathbf{u})\sin\theta(\mathbf{u})\ \cos\alpha(\mathbf{u})\sin\theta(\mathbf{u})\ \cos\theta(\mathbf{u})]^T, \quad (1)$$

and formulated via the surface gradient as follows

$$\mathbf{n}(\mathbf{u}) = \frac{[-p(\mathbf{u})\ -q(\mathbf{u})\ 1]^T}{\sqrt{p(\mathbf{u})^2 + q(\mathbf{u})^2 + 1}}, \quad (2)$$

where $p(\mathbf{u}) = \partial_x z(\mathbf{u})$ and $q(\mathbf{u}) = \partial_y z(\mathbf{u})$, so that $\nabla z(\mathbf{u}) = [p(\mathbf{u})\ q(\mathbf{u})]^T$.

(a) Acquired input (b) Degree of polarisation   (c) Phase angle   (d) Unpolarised intensity

Fig. 2: Polarimetric capture (a) and decomposition to polarisation image (b-d).

2.1   Polarisation image

When unpolarised light is reflected from a surface, it becomes partially polarised. There are a number of mechanisms by which this process occurs. The two models that we use are described in Sect. 2.2 and are suitable for dielectric materials. A *polarisation image* (Fig. 2b-d) can be estimated by capturing a sequence of images (Fig. 2a) in which a linear polarising filter in front of the camera is rotated through a sequence of $P \geq 3$ different angles ϑ_j, $j \in \{1, \ldots, P\}$. The intensity at a pixel varies sinusoidally between $I_{\min}$ and $I_{\max}$ with the polariser angle:

$$i_{\vartheta_j}(\mathbf{u}) = \frac{I_{\max}(\mathbf{u}) + I_{\min}(\mathbf{u})}{2} + \frac{I_{\max}(\mathbf{u}) - I_{\min}(\mathbf{u})}{2} \cos[2\vartheta_j - 2\phi(\mathbf{u})]. \quad (3)$$

The polarisation image is obtained by decomposing the sinusoid at every pixel into three quantities [6]. These are the *phase angle*, $\phi(\mathbf{u})$, the *degree of polarisation*, $\rho(\mathbf{u})$, and the *unpolarised intensity*, $i_{\mathrm{un}}(\mathbf{u})$, where:

$$\rho(\mathbf{u}) = \frac{I_{\max}(\mathbf{u}) - I_{\min}(\mathbf{u})}{I_{\max}(\mathbf{u}) + I_{\min}(\mathbf{u})} \quad \text{and} \quad i_{\mathrm{un}}(\mathbf{u}) = \frac{I_{\max}(\mathbf{u}) + I_{\min}(\mathbf{u})}{2}. \quad (4)$$

The parameters of the sinusoid can be estimated from the captured image sequence using nonlinear least squares [7], linear methods [9] or via a closed form solution [6] for the specific case of $P = 3$, $\vartheta \in \{0°, 45°, 90°\}$. See supplementary material for details of our sinusoid fitting scheme.

2.2   Polarisation models

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. We assume that the object under study is composed of a dielectric material exhibiting both diffuse reflection (due to subsurface scattering) and specular reflection (due to direct reflection at the air/surface interface). We make use of both types of reflection. This model is particularly suitable for smooth, glossy materials such as porcelain, skin, plastic and surfaces finished with gloss paint. We follow recent works [3,17] and assume that reflection from a point can be classified as diffuse dominant or specular dominant (see supplementary material for our classification scheme). Hence, a pixel $\mathbf{u}$ belongs either to the set of diffuse pixels, $\mathcal{D}$, or the set of specular pixels, $\mathcal{S}$, with $\mathcal{F} = \mathcal{D} \cup \mathcal{S}$.
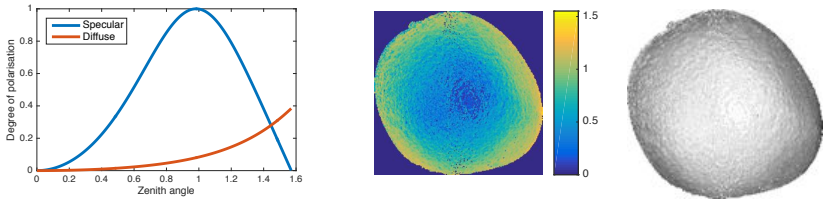
(a) Degree of polarisation dependency   (b) Estimated zenith angle   (c) $\cos(\theta)$

Fig. 3: (a) Relationship between degree of polarisation and zenith angle, for specular and diffuse dielectric reflectance with $\eta = 1.5$. (b) Estimated zenith angle from degree of polarisation in Fig. 2b. (c) Visualisation of estimated zenith angle.

Diffuse polarisation model   For diffuse reflection, the degree of polarisation is related (Fig. 3a, red curve) to the zenith angle $\theta(\mathbf{u}) \in [0, \frac{\pi}{2}]$ of the normal in viewer-centred coordinates (i.e. the angle between the normal and viewer):

$$\rho(\mathbf{u}) = \frac{\sin(\theta(\mathbf{u}))^2 \left(\eta - \frac{1}{\eta}\right)^2}{4 \cos(\theta(\mathbf{u})) \sqrt{\eta^2 - \sin(\theta(\mathbf{u}))^2} - \sin(\theta(\mathbf{u}))^2 \left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2}, \quad (5)$$

where η is the refractive index. The dependency on η is weak and typical values for dielectrics range between 1.4 and 1.6. We assume $\eta = 1.5$ for the rest of this paper. This expression can be rearranged to give a closed form solution for the zenith angle in terms of a function, $f(\rho(\mathbf{u}), \eta)$, that depends on the measured degree of polarisation and the refractive index:

$$\cos(\theta(\mathbf{u})) = \mathbf{n}(\mathbf{u}) \cdot \mathbf{v} = f(\rho(\mathbf{u}), \eta) = \qquad (6)$$

$$\sqrt{\frac{2\rho + 2\eta^2 \rho - 2\eta^2 + \eta^4 + \rho^2 + 4\eta^2 \rho^2 - \eta^4 \rho^2 - 4\eta^3 \rho \sqrt{-(\rho-1)(\rho+1)} + 1}{\eta^4 \rho^2 + 2\eta^4 \rho + \eta^4 + 6\eta^2 \rho^2 + 4\eta^2 \rho - 2\eta^2 + \rho^2 + 2\rho + 1}}$$

where we drop the dependency of ρ on $\mathbf{u}$ for brevity. Since we work in a viewer-centred coordinate system, the viewing direction is $\mathbf{v} = [0 \ 0 \ 1]^T$ and we have simply: $n_z(\mathbf{u}) = f(\rho(\mathbf{u}), \eta)$, or, in terms of the surface gradient,

$$\frac{1}{\sqrt{p(\mathbf{u})^2 + q(\mathbf{u})^2 + 1}} = f(\rho(\mathbf{u}), \eta). \quad (7)$$

The phase angle determines the azimuth angle of the surface normal $\alpha(\mathbf{u}) \in [0, 2\pi]$ up to a 180° ambiguity: $\mathbf{u} \in \mathcal{D} \Rightarrow \alpha(\mathbf{u}) = \phi(\mathbf{u})$ or $(\phi(\mathbf{u}) + \pi)$. Hence, for a diffuse pixel $\mathbf{u} \in \mathcal{D}$, this means that the surface normal is given (up to an ambiguity) by either $\mathbf{n}(\mathbf{u}) = \bar{\mathbf{n}}(\mathbf{u})$ or $\mathbf{n}(\mathbf{u}) = \mathbf{T}\bar{\mathbf{n}}(\mathbf{u})$ where

$$\bar{\mathbf{n}}(\mathbf{u}) = \begin{bmatrix} \sin\phi(\mathbf{u})\sin\theta(\mathbf{u}) \\ \cos\phi(\mathbf{u})\sin\theta(\mathbf{u}) \\ \cos\theta(\mathbf{u}) \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \mathbf{R}_z(180°) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

Specular polarisation model   For specular reflection, the degree of polarisation is again related to the zenith angle (Fig. 3a, blue curve):

$$\rho_s(\mathbf{u}) = \frac{2\sin(\theta(\mathbf{u}))^2\cos(\theta(\mathbf{u}))\sqrt{\eta^2 - \sin(\theta(\mathbf{u}))^2}}{\eta^2 - \sin(\theta(\mathbf{u}))^2 - \eta^2\sin(\theta(\mathbf{u}))^2 + 2\sin(\theta(\mathbf{u}))^4}. \tag{9}$$

This expression is problematic for two reasons: 1. it cannot be analytically inverted to solve for zenith angle, 2. there are two solutions. The first problem is overcome simply by using a lookup table and interpolation. The second problem is not an issue in practice. Specular reflections occur when the surface normal is approximately halfway between the viewer and light source directions. We assume that the light source $\mathbf{s}$ is positioned in the same hemisphere as the viewer, i.e. $\mathbf{v} \cdot \mathbf{s} > 0$. In this configuration, specular pixels will never have a zenith angle $> \sim 45°$. Hence, we can restrict (9) to this range and, therefore, a single solution.

In contrast to diffuse reflection, the azimuth angle of the surface normal is perpendicular to the phase of the specular polarisation [18] leading to a $\frac{\pi}{2}$ shift: $\mathbf{u} \in \mathcal{S} \Rightarrow \alpha(\mathbf{u}) = (\phi(\mathbf{u}) - \pi/2)$ or $(\phi(\mathbf{u}) + \pi/2)$.

Figure 3b shows zenith angle estimates using the diffuse/specular model on $\mathcal{D}/\mathcal{S}$ respectively. In Fig. 3c we show the cosine of the estimated zenith angle, a visualisation corresponding to a Lambertian rendering with frontal lighting.

2.3   Shading constraint

The unpolarised intensity provides an additional constraint on the surface normal direction via an appropriate reflectance model. We assume that, for diffuse-labelled pixels, light is reflected according to the Lambertian model. We also assume that albedo is uniform and factor it into the light source vector $\mathbf{s}$. Hence, unpolarised intensity is related to the surface normal by:

$$\mathbf{u} \in \mathcal{D} \Rightarrow i_{\mathrm{un}}(\mathbf{u}) = \cos(\theta_i(\mathbf{u})) = \mathbf{n}(\mathbf{u}) \cdot \mathbf{s}, \tag{10}$$

where $\theta_i(\mathbf{u})$ is the angle of incidence (angle between light source and surface normal). In terms of the surface gradient, this becomes:

$$i_{\mathrm{un}}(\mathbf{u}) = \frac{-p(\mathbf{u})s_x - q(\mathbf{u})s_y + s_z}{\sqrt{p(\mathbf{u})^2 + q(\mathbf{u})^2 + 1}}. \tag{11}$$

Note that if the light source and viewer direction coincide (a configuration that is physically impossible to achieve precisely) then this equation provides no more information than the degree of polarisation. Hence, we assume that the light source direction is different from the viewing direction, i.e. $\mathbf{s} \neq \mathbf{v}$.

For specular pixels, we do not use the unpolarised intensity directly (though it is used in the labelling of specular pixels - see supplementary material). Instead, we assume simply that the normal is approximately equal to the halfway vector:

$$\mathbf{u} \in \mathcal{S} \Rightarrow \mathbf{n}(\mathbf{u}) \approx \mathbf{h} = (\mathbf{s} + \mathbf{v})/\|\mathbf{s} + \mathbf{v}\|. \tag{12}$$

3   Linear depth estimation with known illumination

We now show that the polarisation shape cues can be expressed as per pixel equations that are linear in terms of the surface gradient. By using finite difference approximations to the surface gradient, this allows us to write the problem of depth estimation in terms of a large system of linear equations. This means that depth estimation is both efficient and certain to obtain the global optimum. In this section we assume that the lighting and albedo are known. However, in the following section we describe how they can be estimated from the polarisation image, allowing depth recovery with uncalibrated illumination.

3.1   Polarisation constraints as linear equations

First, we note that the phase angle constraint can be written as a collinearity condition. This condition is satisfied by either of the two possible azimuth angles implied by the phase angle measurement. Writing it in this way is advantageous because it means we do not have to disambiguate the surface normals explicitly. Instead, when we solve the linear system for depth, the azimuthal ambiguities are resolved in a globally optimal way. Specifically, for diffuse pixels we require the projection of the surface normal into the x-y plane, $[n_x \; n_y]$, and a vector in the image plane pointing in the phase angle direction, $[\sin(\phi) \; \cos(\phi)]$, to be collinear. These two vectors are collinear when the following condition is satisfied:

$$\mathbf{n}(\mathbf{u}) \cdot [\cos(\phi(\mathbf{u})) \; -\sin(\phi(\mathbf{u})) \; 0]^T = 0. \tag{13}$$

Substituting (2) into (13) and noting that the nonlinear term in (2) is always > 0 we obtain the first linear equation in the surface gradient:

$$- p(\mathbf{u}) \cos(\phi(\mathbf{u})) + q(\mathbf{u}) \sin(\phi(\mathbf{u})) = 0. \tag{14}$$

A similar expression can be obtained for specular pixels, substituting in the $\frac{\pi}{2}$-shifted phase angles. This condition exhibits a natural weighting that is useful in practice. The phase angle estimates are more reliable when the zenith angle is large (i.e. when the degree of polarisation is high and so the signal to noise ratio is high). When the zenith angle is large, the magnitude of the surface gradient is large, meaning that disagreement with the estimated phase angle is penalised more heavily than for a small zenith angle where the gradient magnitude is small.

The second linear constraint has two different forms for diffuse and specular pixels. The diffuse constraint is obtained by combining the expressions for the unpolarised intensity and the degree of polarisation. To do so, we take a ratio between (11) and (7) which cancels the nonlinear normalisation factor:

$$\frac{i_{\mathrm{un}}(\mathbf{u})}{f(\rho(\mathbf{u}), \eta)} = -p(\mathbf{u})s_x - q(\mathbf{u})s_y + s_z, \tag{15}$$

yielding our second linear equation in the surface gradient. For specular pixels, we express (12) in terms of the surface gradient as:

$$p(\mathbf{u}) = -h_x/h_z \quad \text{and} \quad q(\mathbf{u}) = -h_y/h_z. \tag{16}$$

3.2   Linear height recovery

The surface gradient in (2) can be approximated numerically from the discretised surface height function using finite differences. To reduce sensitivity to noise and improve robustness, where possible we use a smoothed central difference approximation. Such an approximation is obtained by convolving the surface height function with Sobel operators $\mathbf{G}_x, \mathbf{G}_y \in \mathbb{R}^{3\times3}$: $\partial_x z \approx z * \mathbf{G}_x$ and $\partial_y z \approx z * \mathbf{G}_y$. At the boundary of the image or the foreground mask, not all neighbours may be available for a given pixel. In this case, we use unsmoothed central differences (where both horizontal or both vertical neighbours are available) or, where only a single neighbour is available, single forward/backward differences.

Substituting these finite differences into (14), (15) and (16) therefore leads to linear equations with between 3 and 8 unknown values of z (depending on which combination of numerical gradient approximations are used). Of course, the surface height function is unknown. So, we seek the surface height function whose finite difference gradients solve the system of linear equations over all pixels. Due to noise, we do not expect an exact solution. Hence, for an image with K foreground pixels, we can solve in a least squares sense the system of $2K$ linear equations in the K unknown height values. In order to resolve the unknown constant of integration (i.e. applying an arbitrary offset to z does not affect its orthographic images), we add an additional linear equation to set the height of one pixel to zero. We end up with the linear least squares problem $\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|^2$, where $\mathbf{A}$ has $2K + 1$ rows, K columns and is sparse (each row has at most 8 non-zero values). This can be solved efficiently. Note: this is a system of linear equations in depth. It is not a partial differential equation. Hence, we do not require boundary conditions to be specified.

We also find it advantageous (though not essential) to include two priors on the surface height: 1. Laplacian smoothness, 2. convexity. Both are expressed as linear equations in the surface height. See supplementary material for details.

4   Illumination estimation from a polarisation image

The method described above enables linear depth recovery from a single polarisation image when the illumination direction is known. In this section, we describe how to use the polarisation image to estimate illumination, prior to depth estimation, so that the method above can be applied in an uncalibrated scenario. First, we show that the problem of light source estimation is ambiguous. Second, we derive a method to compute the light source direction (up to the binary ambiguity) from ambiguous normals using the minimum possible number of observations. Third, we extend this to an efficient least squares optimisation that uses the whole image and is applicable to noisy data. Finally, we relax the lighting assumptions to allow more flexible 1st and 2nd order SH illumination.

We consider only diffuse pixels for illumination estimation, since specular pixels are sparse and we wish to avoid estimating the parameters of a particular assumed specular reflectance model. Hence, the unpolarised intensity is assumed
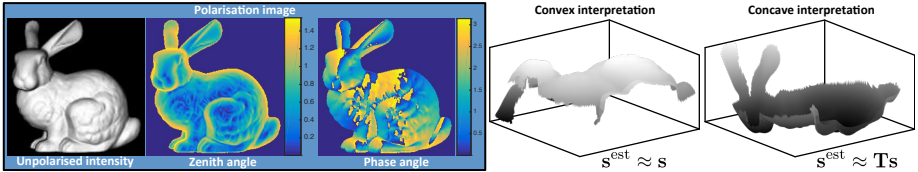
Fig. 4: A polarisation image of a diffuse object enables uncalibrated surface reconstruction up to a convex/concave ambiguity. Both interpretations are consistent with the polarisation image.

to follow a Lambertian model with uniform albedo, as in (11). For the true $\mathbf{s}$, $i_{\text{un}}(\mathbf{u}) = \bar{\mathbf{n}}(\mathbf{u})^T\mathbf{s} \vee i_{\text{un}}(\mathbf{u}) = (\mathbf{T}\bar{\mathbf{n}}(\mathbf{u}))^T\mathbf{s}$. Hence, a single pixel restricts the light source to two planes.

4.1 Relationship to the Bas-relief ambiguity

For an image with K diffuse pixels, there are 2^K possible disambiguations of the polarisation normals. Suppose that we know the correct disambiguation of the normals and that we stack them to form the matrix $\mathbf{N}_{\text{true}} \in \mathbb{R}^{K\times 3}$ and stack the unpolarised intensities in the vector $\mathbf{i} = [i_{\text{un}}(\mathbf{u}_1) \ \dots \ i_{\text{un}}(\mathbf{u}_K)]^T$. In this case, the light source $\mathbf{s}$ that satisfies $\mathbf{N}_{\text{true}}\mathbf{s} = \mathbf{i}$ is given by the pseudo-inverse:

$$\mathbf{s} = \mathbf{N}_{\text{true}}^{+}\mathbf{i}. \tag{17}$$

However, for any invertible 3×3 linear transform $\mathbf{A} \in GL(3)$, it is also true that $\mathbf{N}_{\text{true}}\mathbf{A}^{-1}\mathbf{A}\mathbf{s} = \mathbf{i}$, and so $\mathbf{A}\mathbf{s}$ is also a solution using the transformed normals $\mathbf{N}_{\text{true}}\mathbf{A}^{-1}$. However, the only such $\mathbf{A}$ where $\mathbf{N}_{\text{true}}\mathbf{A}^{-1}$ is consistent with the polarisation image is $\mathbf{A} = \mathbf{T}$, i.e. where the azimuth angle of each normal is shifted by π. Hence, if $\mathbf{s}$ is a solution with normals $\mathbf{N}_{\text{true}}$ then $\mathbf{T}\mathbf{s}$ is also a solution with normals $\mathbf{N}_{\text{true}}\mathbf{T}$. Note that $\mathbf{T}$ is a generalised Bas-relief (GBR) transformation [4] with parameters $\mu = 0$, $\nu = 0$ and $\lambda = \pm 1$, i.e. the binary convex/concave ambiguity. Hence, from a polarisation image with unknown lighting, we will be unable to distinguish the true normals and lighting from those transformed by $\mathbf{T}$. Since $\mathbf{T}$ is a GBR transformation, the transformed normals remain integrable and correspond to the true surface negated in depth (see Fig. 4).

4.2 Minimal solutions

Suppose that $\mathbf{N} \in \mathbb{R}^{K\times 3}$ contains one of the 2^K possible disambiguations of the K surface normals, i.e. $\mathbf{N}_j = \bar{\mathbf{n}}(\mathbf{u}_j)$ or $\mathbf{N}_j = \mathbf{T}\bar{\mathbf{n}}(\mathbf{u}_j)$. If $\mathbf{N}$ is a valid disambiguation, then (with no noise) we expect: $\mathbf{N}\mathbf{N}^{+}\mathbf{i} = \mathbf{i}$. We can see in a straightforward way that three pixels will be insufficient to distinguish a valid from an invalid disambiguation. When $K = 3$, $\mathbf{N}^{+} = \mathbf{N}^{-1}$ and so $\mathbf{N}\mathbf{N}^{+} = \mathbf{I}$ and hence the condition is satisfied by any combination of disambiguations. The reason for this is that, apart from degenerate cases, any three planes will intersect at a point so any combination of transformed or untransformed normals will allow an $\mathbf{s}$ to be found that satisfies all three equations.

However, the problem becomes well-posed for $K > 3$. The system of linear equations must be consistent and have a unique solution. If some, but not all, of the normals are transformed from their true direction then the system of equations will be inconsistent. By the Rouché–Capelli theorem [19], consistency and uniqueness require $\text{rank}(\mathbf{N}) = \text{rank}([\mathbf{N} \ \mathbf{i}]) = 3$. So, we could try each possible combination of disambiguated normals and check whether the rank condition is satisfied. Note that we only need consider half of the possible disambiguations. We can divide the 2^K disambiguations into 2^{K-1} pairs differing by a global transformation and only need consider one of each of the pairs. So, for the minimal case of $K = 4$, we construct the 8 possible normal matrices $\mathbf{N}$, with the first row fixed to $\mathbf{N}_1 = \bar{\mathbf{n}}(\mathbf{u}_1)$, and find the one satisfying the rank condition. For this $\mathbf{N}$ we find $\mathbf{s}$ by (17) and the solution is either $(\mathbf{N}, \mathbf{s})$ or $(\mathbf{NT}, \mathbf{Ts})$.

4.3   Alternating optimisation

In practice, we expect the ambiguous normals and unpolarised intensities to be noisy. Therefore, a least squares solution over all observed pixels is preferable. Since the unknown illumination is only 3D and we have a polarisation observation for every pixel, the problem is highly overconstrained. Following the combinatorial approach above, we could build all 2^K possible systems of linear equations, solve them in a least squares sense and take the one with minimal residual as the solution. However, this is NP-hard and impractical for any non-trivial value of K. Instead, we can write an optimisation problem to find $\mathbf{s}$:

$$\mathbf{s}^* = \underset{\mathbf{s} \in \mathbb{R}^3}{\arg\min} \sum_{j \in \mathcal{D}} \min \left([\bar{\mathbf{n}}(\mathbf{u}_j) \cdot \mathbf{s} - i_{\text{un}}(\mathbf{u}_j)]^2, [\mathbf{T}\bar{\mathbf{n}}(\mathbf{u}_j) \cdot \mathbf{s} - i_{\text{un}}(\mathbf{u}_j)]^2 \right). \quad (18)$$

This is non-convex since the minimum of two convex functions is not convex [20]. However, (18) can be efficiently optimised using alternating assignment and optimisation. In practice, we find that this almost always converges to the global minimum even with a random initialisation. In the assignment step, given an estimate for the light source at iteration t, $\mathbf{s}^{(t)}$, we choose from each ambiguous pair of normals the one that yields minimal error under illumination $\mathbf{s}^{(t)}$:

$$\mathbf{N}_j^{(t)} := \begin{cases} \bar{\mathbf{n}}(\mathbf{u}_j) & \text{if } \left[\bar{\mathbf{n}}(\mathbf{u}_j)\cdot\mathbf{s}^{(t)} - i_{\text{un}}(\mathbf{u}_j)\right]^2 < \left[\mathbf{T}\bar{\mathbf{n}}(\mathbf{u}_j)\cdot\mathbf{s}^{(t)} - i_{\text{un}}(\mathbf{u}_j)\right]^2 \\ \mathbf{T}\bar{\mathbf{n}}(\mathbf{u}_j) & \text{otherwise} \end{cases}. \quad (19)$$

At the optimisation step, we use the selected normals to compute the new light source by solving the linear least squares system: $\mathbf{s}^{(t+1)} := (\mathbf{N}^{(t)})^+\mathbf{i}$. These two steps are iterated to convergence. In all our experiments, this process converged in fewer than 10 iterations. To resolve the ambiguity in our experimental results, we always take the light source estimate that gives the maximal surface.

4.4   Extension to 1st and 2nd order spherical harmonic lighting

Using a first or second order SH diffuse lighting model [21,22], the binary ambiguity in the surface normal leads to a binary ambiguity in the SH basis vector
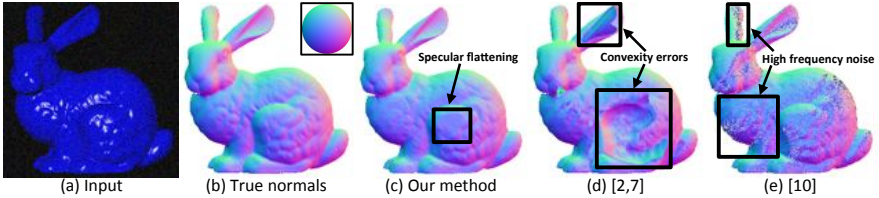
Fig. 5: Typical surface normal estimates (c-e) from noisy synthetic data (a). The inset sphere in (b) shows how surface orientation is visualised as a colour.

at each pixel. Specifically, a first order SH lighting model introduces a constant term: $i_{un}(\mathbf{u}) = \mathbf{b}_4(\mathbf{u})^T \mathbf{s}_4$ with basis vector $\mathbf{b}_4(\mathbf{u}) = \begin{bmatrix} n_x(\mathbf{u}) & n_y(\mathbf{u}) & n_z(\mathbf{u}) & 1 \end{bmatrix}^T$. With ambiguous normals, the basis vector is known up to a binary ambiguity: $\mathbf{b}_4(\mathbf{u}) = \bar{\mathbf{b}}_4(\mathbf{u})$ or $\mathbf{b}_4(\mathbf{u}) = \mathbf{T}_4 \bar{\mathbf{b}}_4(\mathbf{u})$ with $\bar{\mathbf{b}}_4(\mathbf{u}) = \begin{bmatrix} \bar{n}_x(\mathbf{u}) & \bar{n}_y(\mathbf{u}) & \bar{n}_z(\mathbf{u}) & 1 \end{bmatrix}^T$ and the transformation given by: $\mathbf{T}_4 = \text{diag}(-1, -1, 1, 1)$. Solving for $\mathbf{s}_4$ is the same problem as solving for a point source, leading to the same ambiguity. If $\mathbf{s}_4$ is a solution with minimal residual then $\mathbf{T}_4 \mathbf{s}_4$ is also an optimal solution and the transformation of the normals corresponds to a GBR convex/concave transformation. Similarly, a second order SH lighting model: $i_{un}(\mathbf{u}) = \mathbf{b}_9(\mathbf{u})^T \mathbf{s}_9$ with basis vector $\mathbf{b}_9 = \begin{bmatrix} 1 & n_x & n_y & n_z & 3n_z^2-1 & n_x n_y & n_x n_z & n_y n_z & n_x^2-n_y^2 \end{bmatrix}^T$, can be handled in exactly the same way with the appropriate transformation matrix given by: $\mathbf{T}_9 = \text{diag}(1, -1, -1, 1, 1, 1, -1, -1, 1)$. For shape estimation, we compute the 4D or 9D lighting vector, subtract from the diffuse intensity the zeroth and second order appearance contributions and then run the same algorithm as for point source illumination using only the first order appearance.

5   Experimental results

We begin with a quantitative evaluation on synthetic data. We render images of the Stanford bunny with Blinn-Phong reflectance under point source illumination (Fig. 5a). We simulate polarisation according to (3), (5) and (9) with varying polariser angle, add Gaussian noise of standard deviation σ and quantise to 8 bits. We vary light source direction over $\theta_l \in \{15°, 30°, 60°\}$ and $\alpha_l \in \{0°, 90°, 180°, 270°\}$. We estimate a polarisation image for each $(\sigma, \theta_l, \alpha_l)$ and use this as input. For comparison, we implemented the only previous methods applicable to a single polarisation image: 1. boundary propagation [2,7] and 2. Lambertian shading disambiguation [10]. The second method requires known lighting and albedo. For both this and our method, we provide results with ground truth lighting/albedo (superscript "gt") and lighting/albedo estimated using the method in Sect. 4.3 (superscript "est"). For the comparison methods, we compute a depth map using least squares integration, as in [23]. For our method, we compute surface normals using a bicubic fit to the estimated depth.

We show typical results in Fig. 5c-e and quantitative results in Tab. 1 (RMS depth error and mean angular surface normal error averaged over α_l and 100 repeats for each setting; best result for each setting emboldened). The boundary
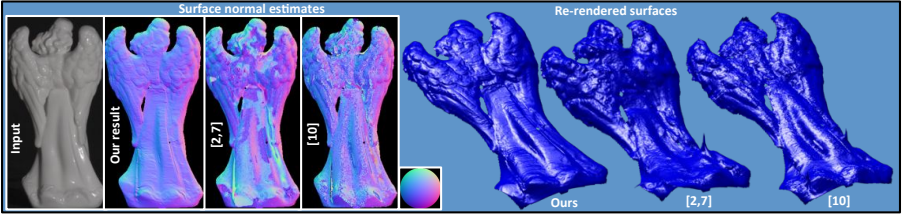
Fig. 6: Qualitative comparison on real world data. Light source direction = [2 0 7].

propagation method [2,7] assumes convexity, meaning that internal concavities are incorrectly recovered. The Lambertian method [10] exhibits high frequency noise since solutions are purely local. Both methods also contain errors in specular regions and propagate errors from normal estimation into the integrated surface. Our solution is smoother and more stable in specular regions yet still recovers fine surface detail. Note however that the simple constraint in (12) encourages all specular normals to point in the same direction, leading to over-flattening of specular regions. Quantitatively, our method offers the best performance across all settings. In many cases, the result with estimated lighting is better than with ground truth. We believe that this is because it enables the method to partially compensate for noise. In Tab. 2 we show the quantitative accuracy of our lighting estimate. We use the same point source directions as above. When the lighting is within 15° of the viewing direction, the error is less than 1°. For order 1 and 2 SH lighting, we use the same order 1 components as the point source directions and randomly generate the order 0 and 2 components.

In order to evaluate our method on real world images, we capture two datasets using a Canon EOS-1D X with an Edmund Optics glass linear polarising filter. The first dataset is captured in a dark room using a Lowel Prolight to approximate a point source. We experiment with both known and unknown lighting. For known lighting, the approximate position of the light source is measured and to calibrate for unknown light source intensity and surface albedo, we use the method in Sect. 4.3 to compute the length of the light source vector, fixing its direction to the measured one. The second dataset is captured outdoors on a sunny day using natural illumination. We use an order 1 SH lighting model.

We show a qualitative comparison between our method and the two reference methods in Fig. 6 using known lighting (see supplementary material for more comparative results). The comparison methods exhibit the same artefacts as on synthetic data. Some of the noise in the normals is removed by the smoothing effect of surface integration but concave/convex errors in [2,7] grossly distort the overall shape, while the surface details of the wings are lost by [10]. In Fig. 7 we show qualitative results of our method on a range of material types, under a variety of known or estimated illumination conditions (both indoor point source and outdoor uncontrolled). Note that the recovered surface of the angel remains stable even with estimated illumination (compared to known illumination in Fig. 6). Note also that our method is able to recover the fine surface detail of the skin of the lemon and orange under both point source and natural illumination.

Table 1: Depth and surface normal estimation errors on synthetic data.

θ_l	Method	$\sigma = 0\%$ Depth (pixels)	Normal (degrees)	$\sigma = 0.5\%$ Depth (pixels)	Normal (degrees)	$\sigma = 1\%$ Depth (pixels)	Normal (degrees)	$\sigma = 2\%$ Depth (pixels)	Normal (degrees)
15°	Ours^{gt}	**3.65**	**3.30**	5.68	5.39	16.09	9.59	16.96	16.19
	Ours^{est}	3.75	3.36	**5.60**	**5.35**	**15.77**	**9.44**	**16.80**	**16.01**
	[10]^{gt}	11.46	4.43	51.78	8.07	173.89	13.37	92.73	22.78
	[10]^{est}	11.64	4.86	51.69	8.10	173.83	13.34	92.08	22.71
	[2,7]	17.96	12.19	55.51	9.69	180.02	13.20	90.64	17.59
30°	Ours^{gt}	**3.67**	**4.68**	15.42	7.00	10.20	10.06	16.84	16.14
	Ours^{est}	6.07	7.57	**14.89**	**6.83**	**9.43**	**9.75**	**16.25**	**15.67**
	[10]^{gt}	14.22	6.00	780.40	8.69	203.68	13.19	286.65	21.29
	[10]^{est}	16.34	10.90	780.55	9.21	203.76	13.97	286.15	22.23
	[2,7]	20.68	15.96	798.36	15.90	208.59	18.92	286.51	25.80
60°	Ours^{gt}	**7.57**	**11.05**	13.70	**14.22**	**67.62**	**17.81**	21.62	**23.25**
	Ours^{est}	12.49	13.91	**12.06**	14.89	76.82	19.22	**20.94**	24.96
	[10]^{gt}	17.50	14.83	1355.47	20.88	7028.37	26.68	940.74	34.30
	[10]^{est}	20.90	20.82	1356.00	24.37	7014.33	30.12	936.88	37.53
	[2,7]	22.80	23.74	1376.83	30.25	6468.54	35.32	956.13	41.13

Table 2: Quantitative light source estimation results on synthetic data.

| θ_l	Point source err $= \arccos(\mathbf{s} \cdot \mathbf{s}^{est})$ $\sigma = 0\%$	$\sigma = 0.5\%$	$\sigma = 1\%$	$\sigma = 2\%$	Order 1 SH err $= \|\mathbf{s} - \mathbf{s}^{est}\|$ $\sigma = 0\%$	$\sigma = 0.5\%$	$\sigma = 1\%$	$\sigma = 2\%$	Order 2 SH err $= \|\mathbf{s} - \mathbf{s}^{est}\|$ $\sigma = 0\%$	$\sigma = 0.5\%$	$\sigma = 1\%$	$\sigma = 2\%$
15°	0.045°	0.069°	0.20°	0.56°	0.0040	0.0040	0.0031	0.0041	0.0007	0.0013	0.0024	0.0035
30°	0.084°	0.33°	0.88°	2.42°	0.0046	0.0047	0.0059	0.0041	0.0006	0.0036	0.0025	0.013
60°	0.81°	3.44°	7.83°	15.97°	0.0062	0.0084	0.0060	0.0060	0.0012	0.0025	0.0091	0.0052

6   Conclusions

We have presented the first SfP technique in which polarisation constraints are expressed directly in terms of surface depth. Moreover, through careful construction of these equations, we ensure that they are linear and so depth estimation is simply a linear least squares problem. The SfP cue is often described as being locally ambiguous. We have shown that, in fact, even with unknown lighting the diffuse unpolarised intensity image restricts the uncertainty to a global convex/concave ambiguity. Our method is practically useful, enabling monocular, passive depth estimation even in outdoor lighting. For reproducibility, we make a full implementation of our method and the two comparison methods available[1].

In future, we would like to relax the assumptions in Sect. 2. From a practical perspective, the most useful would be to allow spatially-varying albedo. Rather than assuming that pixels are specular or diffuse dominant, we would also like to allow mixtures of the two polarisation models and to exploit specular shading. To do so would require an assumption of a specular BRDF model. An alternative would be to fit a data-driven BRDF model [24] directly to the ambiguous polarisation normals, potentially allowing single shot BRDF and shape estimation.

---
[1] https://github.com/waps101/depth-from-polarisation
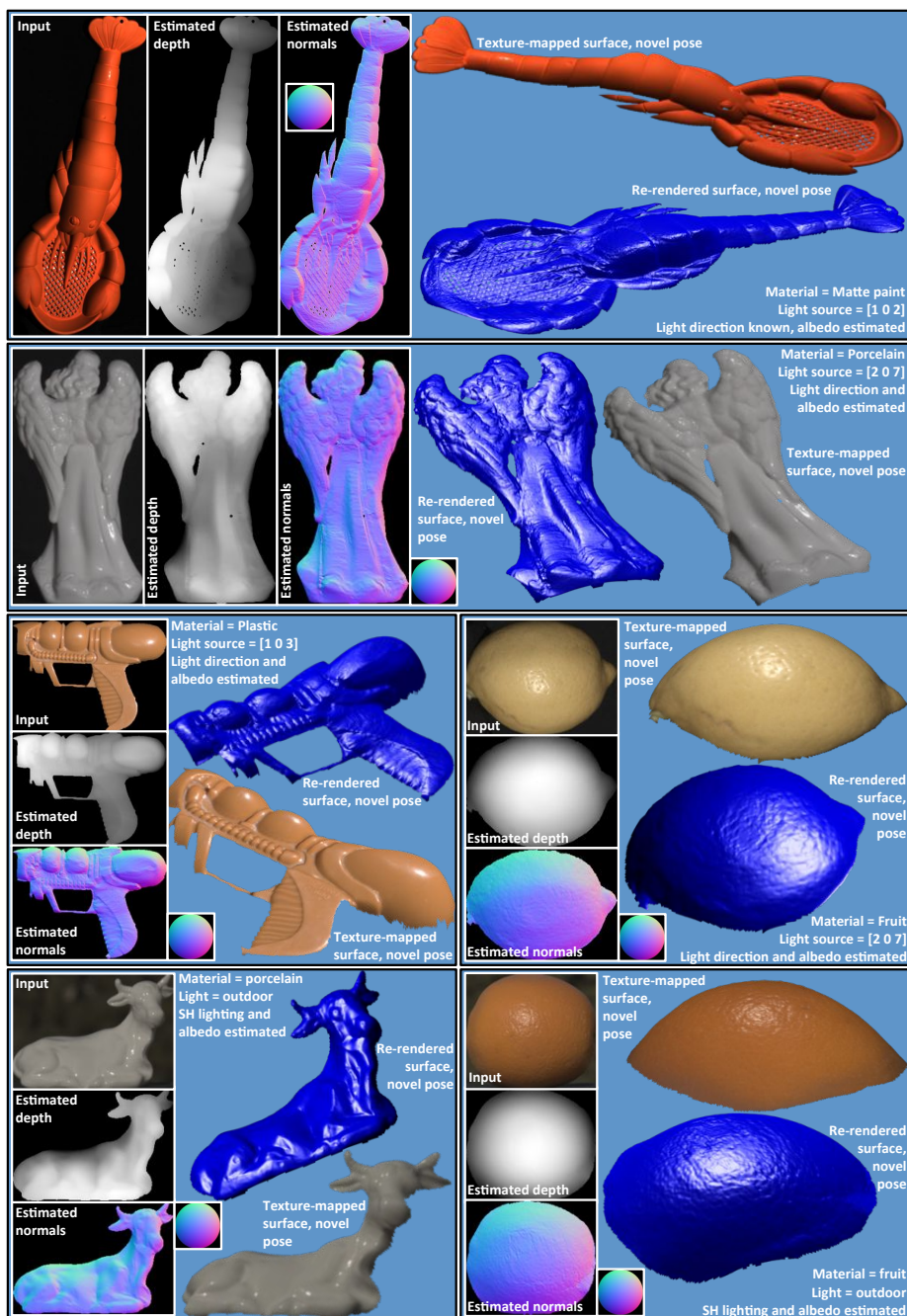
Fig. 7: Qualitative results on a variety of material types. The first three rows show results captured in dark room conditions with a point light source. The two panels in the final row show results in outdoor, uncontrolled illumination. Depth maps are encoded as brighter = closer. The first row shows a result with known lighting direction, all others are estimated.

References

1. Wolff, L.B., Boult, T.E.: Constraining object features using a polarization reflectance model. IEEE Trans. Pattern Anal. Mach. Intell. **13**(7) (1991) 635—657
2. Miyazaki, D., Tan, R.T., Hara, K., Ikeuchi, K.: Polarization-based inverse rendering from a single view. In: Proc. ICCV. (2003) 982–987
3. Kadambi, A., Taamazyan, V., Shi, B., Raskar, R.: Polarized 3D: High-quality depth sensing with polarization cues. In: Proc. ICCV. (2015)
4. Belhumeur, P.N., Kriegman, D.J., Yuille, A.: The Bas-relief ambiguity. Int. J. Comput. Vision **35**(1) (1999) 33–44
5. Schechner, Y.Y.: Self-calibrating imaging polarimetry. In: Proc. ICCP. (2015)
6. Wolff, L.B.: Polarization vision: a new sensory approach to image understanding. Image Vision Comput. **15**(2) (1997) 81–93
7. Atkinson, G.A., Hancock, E.R.: Recovery of surface orientation from diffuse polarization. IEEE Trans. Image Process. **15**(6) (2006) 1653–1664
8. Morel, O., Meriaudeau, F., Stolz, C., Gorria, P.: Polarization imaging applied to 3D reconstruction of specular metallic surfaces. In: Proc. EI 2005. (2005) 178–186
9. Huynh, C.P., Robles-Kelly, A., Hancock, E.: Shape and refractive index recovery from single-view polarisation images. In: Proc. CVPR. (2010) 1229–1236
10. Mahmoud, A.H., El-Melegy, M.T., Farag, A.A.: Direct method for shape recovery from polarization and shading. In: Proc. ICIP. (2012) 1769–1772
11. Rahmann, S., Canterakis, N.: Reconstruction of specular surfaces using polarization imaging. In: Proc. CVPR. (2001)
12. Atkinson, G.A., Hancock, E.R.: Shape estimation using polarization and shading from two views. IEEE Trans. Pattern Anal. Mach. Intell. **29**(11) (2007) 2001–2017
13. Huynh, C.P., Robles-Kelly, A., Hancock, E.R.: Shape and refractive index from single-view spectro-polarimetric images. Int. J. Comput. Vision **101**(1) (2013)
14. Atkinson, G.A., Hancock, E.R.: Surface reconstruction using polarization and photometric stereo. In: Proc. CAIP. (2007) 466–473
15. Drbohlav, O., Šára, R.: Unambiguous determination of shape from photometric stereo with unknown light sources. In: Proc. ICCV. (2001) 581–586
16. Ngo, T.T., Nagahara, H., Taniguchi, R.: Shape and light directions from shading and polarization. In: Proc. CVPR. (2015) 2310–2318
17. Tozza, S., Mecca, R., Duocastella, M., Del Bue, A.: Direct differential photometric stereo shape recovery of diffuse and specular surfaces. JMIV **56**(1) (2016) 57–76
18. Robles-Kelly, A., Huynh, C.P. In: Imaging Spectroscopy for Scene Analysis. Springer-Verlag (2013) 219
19. Carpinteri, A. In: Structural mechanics. Taylor and Francis (1997) 74
20. Grant, M., Boyd, S., Ye, Y.: Disciplined convex programming. In: Global Optimization: From Theory to Implementation. Springer (2006) 155–210
21. Ramamoorthi, R., Hanrahan, P.: On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. JOSA A **18**(10) (2001) 2448–2459
22. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. **25**(2) (2003) 218–233
23. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3D geometry. ACM Trans. Graphic. **24**(3) (2005)
24. Nielsen, J.B., Jensen, H.W., Ramamoorthi, R.: On optimal, minimal BRDF sampling for reflectance acquisition. ACM Trans. Graphic. **34**(6) (2015)
25. http://www.fluxdata.com/imaging-polarimeters
26. https://www.ricoh.com/technology/tech/051_polarization.html