Shape-from-polarisation: a nonlinear least squares approach

Ye Yu*, Dizhong Zhu* and William A. P. Smith
Department of Computer Science, University of York, UK
{yy1571,dz761,william.smith}@york.ac.uk

Abstract

In this paper we present a new type of approach for estimating surface height from polarimetric data, i.e. a sequence of images in which a linear polarising filter is rotated in front of a camera. In contrast to all previous shape-from-polarisation methods, we do not first transform the observed data into a polarisation image. Instead, we minimise the sum of squared residuals between predicted and observed intensities over all pixels and polariser angles. This is a nonlinear least squares optimisation problem in which the unknown is the surface height. The forward prediction is a series of transformations for which we provide analytical derivatives allowing the overall problem to be efficiently optimised using Gauss-Newton type methods with an analytical Jacobian matrix. The method is very general and can incorporate any (differentiable) illumination, reflectance or polarisation model. We also propose a variant of the method which uses image ratios to remove dependence on illumination and albedo. We demonstrate our methods on glossy objects, including with albedo variations, and provide comparison to a state of the art approach.

1. Introduction

The polarisation state of light reflected from a dielectric (i.e. non-metallic) object conveys information about both the material properties and shape of the object [18]. The reason for this phenomenon is that unpolarised light becomes partially polarised when it is reflected specularly [13] or diffusely via subsurface scattering [1]. The degree of polarisation and the orientation of the polarisation are related to local surface orientation, the refractive index of the material and whether the reflection was diffuse or specular. Usually, this information is not visible in an image captured by a conventional camera. However, using either a custom polarisation camera (based on polarising beamsplitters or micropolarising filters on the sensor) or simply placing a linear polarising filter in front of a conventional camera, this rich source of additional information becomes available.

Shape-from-polarisation (SfPol) has received far less attention than other shape-from-X methods in the last two decades. One of the reasons for this is that polarisation information is an innately ambiguous shape cue. Previous work has therefore either relied on simple priors for disambiguation [1, 10] or combined polarisation with an additional cue such as shading [8,15], photometric stereo [3,12], binocular stereo [2] or a coarse depth map [7]. A recent trend in shape-from-shading [16] and photometric stereo [9] has been to estimate surface height directly rather than surface orientation. Recovering height from orientation requires an additional step of surface integration. Estimating height directly reduces the number of unknowns, guarantees satisfaction of integrability constraints and avoids accumulation of errors through a two-step process. Very recently, the same idea has been considered for SfPol [7, 15].

In this paper, we propose a completely new approach to the SfPol problem. Like [7, 15], we estimate surface height directly. However, unlike all previous methods, we do not decompose the captured data into a *polarisation image* and then estimate shape as an independent second step. This two step approach ignores potential uncertainty in the estimated polarisation image. Instead, we take an energy minimisation approach and optimise a nonlinear least squares cost that directly measures error between the observed data and that predicted from the estimated surface height. Our cost function is justified by a probabilistic interpretation (Section 2.1). Our approach is very general. We can incorporate any illumination, reflectance or polarisation model without any requirement that they be invertable. We only require that they are differentiable and in Section 3 provide all derivatives necessary to compute an analytical Jacobian matrix relating surface height and polarimetric intensities. In addition, in Section 4 we propose a variant of our method which uses image ratios to avoid estimating or assuming reflectance or illumination properties. For practical application, we also propose additional priors and a hierarchical approach allowing our method to be applied to challenging real world data.

*Authors contributed equally

1.1. Related work

All previous SfPol methods begin by estimating a polarisation image (defined precisely in Section 2). This is usually done by linear least squares [17] or nonlinear sinusoid curve-fitting [1] when the rotation angles of the polarising filter are known. If the angles are not known, these can also be estimated as part of the decomposition process [14].

Using only polarisation information, the surface orientation at each pixel is only known up to a binary ambiguity. So, early work focussed on solving the disambiguation problem with less emphasis on dealing with additional sources of noise. Atkinson and Hancock [1] and Huynh *et al*. [5] assumed object convexity by propagating outward facing boundary normals into the object interior. Miyazaki *et al*. [10] employed a similar assumption plus a statistical model on the zenith angle distribution to eliminate errors caused by an inaccurate estimate of the refractive index and weak polarisation in some areas.

Instead of relying on a heuristic or statistical model for disambiguation, another class of methods incorporate an additional shape cue. Methods that use shape-from-shading [8, 15] require no additional observations (since a polarisation image includes an intensity channel). Of particular relevance, Smith *et al*. [15] linearised the problem in surface height and also proposed a method to estimate illumination from a polarisation image. Ngo *et al*. [12] consider the relationship between photometric stereo and polarisation constraints. All these algorithms are based on the Lambertian model assumption or necessitate knowledge of the light source. Kadambi *et al*. [7] achieve very impressive results by using polarisation information to refine a depth map obtained by a depth sensor. Huynh *et al*. [6] extended their previous work to a multispectral setting and enforced an integrability condition during disambiguation. Non-diffuse models have also been considered. Morel *et al*. [11] used a specular polarisation model suitable for analysing metallic surfaces. Rahmann and Canterakis [13] treat depth estimation as an optimisation problem and incorporate multiple polarisation images to estimate depth for specular surfaces.

2. Polarisation image

Unpolarised light becomes partially polarised when it is reflected from a surface. This is true both for specular reflection and diffuse, subsurface reflection. Hence, if reflected light is passed through a linear polarising filter which is rotated through a sequence of orientations, ϑ_j with $j = 1..P$, then the observed intensity varies sinusoidally:

$$i^{\text{mod}}_{\vartheta_j}(i_{\text{un}}, \phi, \rho) = i_{\text{un}}\left(1 + \rho\cos[2\vartheta_j - 2\phi]\right). \qquad (1)$$

The sinusoid has period π and can be characterised by three quantities which together are referred to as a *polarisation image*. The *unpolarised intensity*, i_{un}, is the mean value of the sinusoid. This is the intensity that would have been observed without a polariser being present and hence depends on the reflectance properties of the surface and the illumination in the scene. The *phase angle*, $\phi \in [0, \pi)$, defines the phase shift. The *degree of polarisation* (DOP), $\rho \in [0, 1]$, is the ratio between the amplitude and mean value of the sinusoid. The three components of a polarisation image depend on the local surface geometry at the point of reflection as well as the material properties as will be made explicit in Section 3. A simplified expression can be obtained by taking ratios between different polariser orientations:

$$\frac{i^{\text{mod}}_{\vartheta_j}(i_{\text{un}}, \phi, \rho)}{i^{\text{mod}}_{\vartheta_k}(i_{\text{un}}, \phi, \rho)} = f_{\vartheta_j,\vartheta_k}(\phi, \rho) = \frac{1 + \rho\cos[2\vartheta_j - 2\phi]}{1 + \rho\cos[2\vartheta_k - 2\phi]}. \qquad (2)$$

This has the effect of removing any dependency on i_{un} and hence on any assumed reflectance model, material properties or illumination. Hence, using only this ratio expression eliminates the need to estimate albedo and lighting and to assume an underlying reflectance model.

2.1. Probabilistic polarisation model

We assume that observations are subject to additive Gaussian noise: $i^{\text{obs}}_{\vartheta_j} = i^{\text{mod}}_{\vartheta_j}(i_{\text{un}}, \phi, \rho) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is the unknown variance of the noise. Therefore, $i^{\text{obs}}_{\vartheta_j} \sim \mathcal{N}(i^{\text{mod}}_{\vartheta_j}, \sigma^2)$ is itself a normally distributed random variable. Hence, we can write a probabilistic polarisation model as:

$$p(i^{\text{obs}}_{\vartheta_j} | i_{\text{un}}, \phi, \rho) = C(\sigma^2)\exp\left(-\frac{[i^{\text{obs}}_{\vartheta_j} - i^{\text{mod}}_{\vartheta_j}(i_{\text{un}}, \phi, \rho)]^2}{2\sigma^2}\right), \qquad (3)$$

where $C(\sigma^2)$ is a normalising constant. The maximum likelihood solution to the SfPol problem is therefore the surface that gives rise to model intensities that minimise the error to the observed intensities in a least squares sense. This provides justification for our idea of posing the problem as a nonlinear least squares optimisation over the unknown surface height. Note that all previous work begins by estimating the maximum likelihood *polarisation image* (i.e. ρ, ϕ and i_{un} at each pixel independently) and then computes surface normals [1] or surface height [7, 15] that is in some sense optimal with respect to the polarisation image. The problem with this two stage approach is that polarisation image quantities whose estimate is highly uncertain are relied upon to the same degree as those with high certainty.

Under the same noise assumption, the expression (2) is a ratio between two normally distributed random variables and hence follows a Cauchy distribution. In this case, the solution that minimises (in a least squares sense) the error between the observed ratios and those predicted by our model is the best linear unbiased estimator (BLUE) [4].

3. SfPol as analysis by synthesis

We now show how the components of a polarisation image can be derived from the gradient of the surface height function and provide derivatives for each transformation. Subsequently, this enables us to compute analytical derivatives directly relating surface height and sinusoidal intensity, and hence to minimise residuals between observed and predicted intensities by nonlinear least squares. This provides an analysis by synthesis approach for SfPol. We assume that a surface is being viewed orthographically so that it can be written as a height function $z(x, y)$ where (x, y) is a pixel coordinate. We define the gradient of the surface height at a pixel as the vector $\mathbf{g} \in \mathbb{R}^2$ containing the partial derivatives: $\mathbf{g} = [\partial z/\partial x \;\; \partial z/\partial y]^T$.

3.1. Gradient to surface normal

The function $\mathbf{n} : \mathbb{R}^2 \mapsto \mathbb{R}^3$ transforms the gradient vector into a vector whose direction is normal to the surface. The function and its first derivatives (Jacobian) are given by:

$$\mathbf{n}(\mathbf{g}) = \begin{bmatrix} -\mathbf{g} \\ 1 \end{bmatrix}, \quad \mathbf{J}_{\mathbf{n}}(\mathbf{g}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}. \tag{4}$$

The function $\bar{\mathbf{n}} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ normalises a vector to have unit length such that $\|\bar{\mathbf{n}}\| = 1$:

$$\bar{\mathbf{n}}(\mathbf{n}) = \frac{\mathbf{n}}{\|\mathbf{n}\|}, \quad \mathbf{J}_{\bar{\mathbf{n}}}(\mathbf{n}) = \frac{\mathbf{I}}{\|\mathbf{n}\|} - \frac{\mathbf{n}\mathbf{n}^T}{\|\mathbf{n}\|^3}. \tag{5}$$

3.2. Surface normal to spherical coordinates

It is convenient to transform the surface normal vector $\bar{\mathbf{n}}$ into spherical coordinates (α, θ) in a viewer-centred coordinate system. The azimuth angle is computed by the function $\alpha : \mathbb{R}^3 \mapsto [0, 2\pi)$, defined (along with its gradient) as follows:

$$\alpha(\bar{\mathbf{n}}) = \text{atan2}(\bar{n}_2, \bar{n}_1), \;\; \nabla\alpha(\bar{\mathbf{n}}) = \begin{bmatrix} \frac{-\bar{n}_2}{\bar{n}_1^2+\bar{n}_2^2} & \frac{\bar{n}_1}{\bar{n}_1^2+\bar{n}_2^2} & 0 \end{bmatrix}^T. \tag{6}$$

The zenith angle is computed by the function $\theta : \mathbb{R}^3 \mapsto [0, \pi]$:

$$\theta(\bar{\mathbf{n}}) = \arccos(\bar{n}_3), \quad \nabla\theta(\bar{\mathbf{n}}) = \begin{bmatrix} 0 & 0 & \frac{-1}{\sqrt{1-\bar{n}_3^2}} \end{bmatrix}^T. \tag{7}$$

3.3. Orientation to polarisation image

The surface orientation (expressed as a surface normal or spherical coordinates) can be used to compute the three components of a polarisation image. For diffuse polarisation, the phase angle, given by the function $\phi : [0, 2\pi) \mapsto [0, \pi)$, is simply the azimuth angle modulo π [1]:

$$\phi(\alpha) = \begin{cases} \alpha, & \text{if } \alpha \in [0, \pi) \\ \alpha - \pi, & \text{otherwise} \end{cases}, \quad \frac{\partial\phi}{\partial\alpha} = 1. \tag{8}$$

The unpolarised intensity is assumed to adhere to the Lambertian reflectance model and is modelled by the function $i_{\text{un}} : \mathbb{R}^3 \mapsto \mathbb{R}$:

$$i_{\text{un}}(\bar{\mathbf{n}}) = k_d \bar{\mathbf{n}}^T \mathbf{s}, \quad \nabla i_{\text{un}}(\bar{\mathbf{n}}) = k_d \mathbf{s}. \tag{9}$$

where $\mathbf{s} \in \mathbb{R}^3$ is a point light source and $k_d \in [0, 1]$ the diffuse albedo, both of which are assumed known.

Finally, the DoP can be computed by the function $\rho : [0, \pi] \mapsto [0, 1]$ from the zenith angle and the index of refraction, η, that we again assume is known. Assuming a diffuse polarisation model, the DoP is given by:

$$\rho(\theta) = \frac{\sin(\theta)^2 \left(\eta - \frac{1}{\eta}\right)^2}{4\cos(\theta)\sqrt{\eta^2 - \sin(\theta)^2} - \sin(\theta)^2 \left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2}. \tag{10}$$

The derivative $\frac{\partial\rho}{\partial\theta}$ for the diffuse model is given by:

$$\frac{\partial\rho}{\partial\theta} = \frac{2\cos(\theta)\sin(\theta)\left(\eta - \frac{1}{\eta}\right)^2}{4\cos(\theta)\sqrt{\eta^2 - \sin(\theta)^2} - \sin(\theta)^2\left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2} + \frac{\sin(\theta)^2\left(\eta - \frac{1}{\eta}\right)^2}{\left(4\cos(\theta)\sqrt{\eta^2 - \sin(\theta)^2} - \sin(\theta)^2\left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2\right)^2} \times \left(4\sin(\theta)\sqrt{\eta^2 - \sin(\theta)^2} + 2\cos(\theta)\sin(\theta)\left(\eta + \frac{1}{\eta}\right)^2 + \frac{4\cos(\theta)^2\sin(\theta)}{\sqrt{\eta^2 - \sin(\theta)^2}}\right).$$

3.4. Polarisation image to sinusoidal intensities

Finally, the function $i_\vartheta^{\text{mod}} : \mathbb{R}^3 \mapsto \mathbb{R}$ given in (1) can be used to compute an intensity associated with polariser angle ϑ. The partial derivatives of this function are given by:

$$\begin{aligned} \frac{\partial i_\vartheta^{\text{mod}}}{\partial i_{\text{un}}} &= 1 + \rho\cos(2\vartheta - 2\phi), \\ \frac{\partial i_\vartheta^{\text{mod}}}{\partial \phi} &= -2 i_{\text{un}} \rho \sin(2\phi - 2\vartheta), \\ \frac{\partial i_\vartheta^{\text{mod}}}{\partial \rho} &= i_{\text{un}}\cos(2\phi - 2\vartheta). \end{aligned} \tag{11}$$

3.5. Nonlinear least squares

Assume that the surface height values for an image with N foreground pixels are stored in the vector $\mathbf{z} \in \mathbb{R}^N$. The gradient of the surface height function at every pixel can be approximated using finite differences which can be expressed as a matrix multiplication. We write this as a function $\mathbf{G} : \mathbb{R}^N \mapsto \mathbb{R}^{2N}$:

$$\mathbf{G}(\mathbf{z}) = \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} \mathbf{z}, \quad \mathbf{J}_{\mathbf{G}}(\mathbf{z}) = \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix}, \tag{12}$$

where $\mathbf{D}_x \in \mathbb{R}^{N\times N}$ and $\mathbf{D}_y \in \mathbb{R}^{N\times N}$ evaluate the surface gradient in the horizontal and vertical directions respectively and are sparse (two non-zero entries per row using

forward finite differences). Using the derivation in Sections 3.1 to 3.4 we can construct a function $\mathbf{I}^{\text{mod}}_{\vartheta} : \mathbb{R}^{2N} \mapsto \mathbb{R}^N$ that computes the predicted intensities for all pixels with polariser angle ϑ from the surface gradient at every pixel. The derivatives of this function are stored in the Jacobian matrix $\mathbf{J}_{\mathbf{I}^{\text{mod}}_{\vartheta}}(\mathbf{G}) \in \mathbb{R}^{N \times 2N}$ which is constructed using the chain rule applied to the appropriate sequence of derivatives given previously. We can now compute a vector of residuals $\mathbf{r} \in \mathbb{R}^{NP}$ and the Jacobian of the residual function:

$$\mathbf{r}(\mathbf{z}) = \begin{bmatrix} \mathbf{I}^{\text{obs}}_{\vartheta_1} - \mathbf{I}^{\text{mod}}_{\vartheta_1}(\mathbf{G}(\mathbf{z})) \\ \vdots \\ \mathbf{I}^{\text{obs}}_{\vartheta_P} - \mathbf{I}^{\text{mod}}_{\vartheta_P}(\mathbf{G}(\mathbf{z})) \end{bmatrix}, \quad \mathbf{J}_{\mathbf{r}}(\mathbf{z}) = \begin{bmatrix} \mathbf{J}_{\mathbf{I}^{\text{mod}}_{\vartheta_1}}(\mathbf{G}(\mathbf{z})) \\ \vdots \\ \mathbf{J}_{\mathbf{I}^{\text{mod}}_{\vartheta_P}}(\mathbf{G}(\mathbf{z})) \end{bmatrix} \mathbf{J}_{\mathbf{G}}(\mathbf{z}),$$

where $\mathbf{I}^{\text{obs}}_{\vartheta_j} \in \mathbb{R}^N$ is the vector of observed intensities with the jth polariser orientation. Finally, we can solve the following nonlinear least squares problem:

$$\min_{\mathbf{z}} \mathbf{r}(\mathbf{z})^T \mathbf{r}(\mathbf{z}). \tag{13}$$

We emphasise that, under the assumption of the probabilistic model in Section 2.1, this is the maximum likelihood solution for $\mathbf{z}$ given the observed intensities. In practice, we minimise (13) using the trust-region-reflective algorithm, as implemented in the Matlab `lsqnonlin` function. We initialise with a plane, i.e. $\mathbf{z} = \mathbf{0}$.

3.6. Priors

The basic framework described above can be unstable when to applied to real data (for example introducing spikes into the estimated height map) and sometimes converges on local minima. For this reason, we introduce two additional priors.

Smoothness We compute residuals to measure smoothness via convolution of the height map with a Laplacian of Gaussian filter:

$$\mathbf{r}_{\text{smooth}}(\mathbf{z}) = \sqrt{w_{\text{smooth}}}\mathbf{L}\mathbf{z}, \quad \mathbf{J}_{\mathbf{r}_{\text{smooth}}}(\mathbf{z}) = \sqrt{w_{\text{smooth}}}\mathbf{L}, \tag{14}$$

where $\mathbf{L} \in \mathbb{R}^{M \times N}$ has five non-zero entries per row and M is the number of pixels with 4 neighbours. Each row of $\mathbf{L}$ evaluates the convolution of the LoG kernel with the neighbourhood around one pixel. The residuals are zero for planar regions of the surface. w_{smooth} controls the weight of the smoothness prior.

Convexity To encourage global convexity, we compute residuals between the azimuth angles given by the estimated height and that of outward facing normals along the boundary:

$$\mathbf{r}_{\text{convex}}(\mathbf{z}) = \sqrt{w_{\text{convex}}} \begin{bmatrix} \sin(\mathbf{S}_{\text{boundary}}\boldsymbol{\alpha}(\mathbf{z})) - \sin(\boldsymbol{\alpha}_{\text{boundary}}) \\ \cos(\mathbf{S}_{\text{boundary}}\boldsymbol{\alpha}(\mathbf{z})) - \cos(\boldsymbol{\alpha}_{\text{boundary}}) \end{bmatrix}, \tag{15}$$

where $\mathbf{S}_{\text{boundary}} \in \{0,1\}^{B \times N}$ is a selection matrix that selects the B pixels lying on the boundary of the object, $\boldsymbol{\alpha}(\mathbf{z})$ is a vector of the azimuth angles for all pixels computed by the series of transformations given above and $\boldsymbol{\alpha}_{\text{boundary}} \in \mathbb{R}^B$ is the vector of azimuth angles of the outward facing vectors to the boundary of the foreground mask. We measure the angular difference in Cartesian coordinates to avoid wrap-around issues.

4. Ratio-based formulation

The method described in Section 3 requires known albedo and lighting and assumes diffuse reflectance and diffuse polarisation. By using the ratio formulation in (2) we can avoid these requirements and derive an uncalibrated method. Moreover, we can use the ratio-based formulation as initialisation, use the estimated height map to estimate lighting and albedo and then run the full optimisation to further refine the solution. Note however that (2) depends only on the DOP and phase angle. This means that this information alone could only recover the surface up to a binary convex/concave ambiguity [15]. In practice, we find that convex/concave ambiguities can be inconsistently resolved so we propose a hierarchical scheme and automatically adjust prior weights appropriately.

The derivatives of the ratio function $f_{\vartheta_j,\vartheta_k} : \mathbb{R}^2 \mapsto \mathbb{R}$ in (2) are given by:

$$\nabla f_{\vartheta_j,\vartheta_k}(\phi,\rho) = \begin{bmatrix} \frac{2\rho\sin(2\phi-2\vartheta_k)\left(\rho\cos\left(2\,\phi-2\vartheta_j\right)+1\right)}{(\rho\cos(2\phi-2\vartheta_k)+1)^2} - \frac{2\rho\sin\left(2\phi-2\vartheta_j\right)}{\rho\cos(2\phi-2\vartheta_k)+1} \\ \frac{\cos\left(2\phi-2\vartheta_j\right)-\cos(2\phi-2\vartheta_k)}{(\rho\cos(2\phi-2\vartheta_k)+1)^2} \end{bmatrix} \tag{16}$$

We extend the ratio function to all pixels via the function $\mathbf{F}^{\text{mod}}_{\vartheta_j,\vartheta_k} : \mathbb{R}^{2N} \mapsto \mathbb{R}^N$ that computes the predicted ratios for all pixels from the surface gradient at every pixel. The derivatives of this function $\mathbf{J}_{\mathbf{F}^{\text{mod}}_{\vartheta_j,\vartheta_k}}(\mathbf{G}) \in \mathbb{R}^{N \times 2N}$ can again be computed by the appropriate combination of derivatives from Section 3. We can now compute a vector of residuals $\mathbf{r} \in \mathbb{R}^{N(P-1)}$ by taking ratios between each pair of consecutive polariser angles:

$$\begin{aligned} \mathbf{r}(\mathbf{z}) &= \begin{bmatrix} \mathbf{F}^{\text{obs}}_{\vartheta_1,\vartheta_2} - \mathbf{F}^{\text{mod}}_{\vartheta_1,\vartheta_2}(\mathbf{G}(\mathbf{z})) \\ \vdots \\ \mathbf{F}^{\text{obs}}_{\vartheta_{P-1},\vartheta_P} - \mathbf{F}^{\text{mod}}_{\vartheta_{P-1},\vartheta_P}(\mathbf{G}(\mathbf{z})) \end{bmatrix}, \\ \mathbf{J}_{\mathbf{r}}(\mathbf{z}) &= \begin{bmatrix} \mathbf{J}_{\mathbf{F}^{\text{mod}}_{\vartheta_1,\vartheta_2}}(\mathbf{G}) \\ \vdots \\ \mathbf{J}_{\mathbf{F}^{\text{mod}}_{\vartheta_{P-1},\vartheta_P}}(\mathbf{G}) \end{bmatrix} \mathbf{J}_{\mathbf{G}}(\mathbf{z}), \end{aligned} \tag{17}$$

where $\mathbf{F}^{\text{obs}}_{\vartheta_j,\vartheta_k} = \mathbf{I}^{\text{obs}}_{\vartheta_j} / \mathbf{I}^{\text{obs}}_{\vartheta_k}$ is a vector of ratios between observed intensities with polariser angles ϑ_j and ϑ_k.

4.1. Specular polarisation

The residuals above are invariant to the unpolarised intensity and hence to the illumination, albedo and reflectance model. The only requirement is to select a polarisation model to compute DOP and phase angles. In Section 3 we assumed diffuse polarisation. However, our optimisation framework allows us to switch in any polarisation model. In particular, if we assume (as in previous work [15]) that pixels are labelled as diffuse or specular dominant, then we can use a specular polarisation model for specular pixels without assuming a particular specular reflectance model. The specular DOP is given by:

$$\rho(\theta,\eta) = \frac{2\sin(\theta)^2\,\cos(\theta)\,\sqrt{\eta^2-\sin(\theta)^2}}{\eta^2-\sin(\theta)^2-\eta^2\sin(\theta)^2+2\sin(\theta)^4}. \qquad (18)$$

In addition, phase angles must be shifted by $\frac{\pi}{2}$ [1] for specular pixels. In previous work, specular polarisation introduced an additional ambiguity since the model above cannot be analytically inverted (there are two possible solutions). In our optimisation-based approach, this does not matter since we simply need derivatives:

$$\begin{aligned}\frac{\partial\rho}{\partial\theta} = &\frac{2\sin(\theta)^3\sqrt{\eta^2-\sin(\theta)^2}}{\eta^2\sin(\theta)^2-\eta^2-2\sin(\theta)^4+\sin(\theta)^2}\\ &-\frac{4\cos(\theta)^2\,\sin(\theta)\,\sqrt{\eta^2-\sin(\theta)^2}}{\eta^2\sin(\theta)^2-\eta^2-2\sin(\theta)^4+\sin(\theta)^2}\\ &+\frac{2\cos(\theta)^2\sin(\theta)^3}{\sqrt{\eta^2-\sin(\theta)^2}\left(\eta^2\sin(\theta)^2-\eta^2-2\sin(\theta)^4+\sin(\theta)^2\right)}\\ &+\frac{2\cos(\theta)\,\sin(\theta)^2\sqrt{\eta^2-\sin(\theta)^2}}{\left(\eta^2\sin(\theta)^2-\eta^2-2\sin(\theta)^4+\sin(\theta)^2\right)^2}\\ &\times\left(2\cos(\theta)\,\eta^2\sin(\theta)-8\cos(\theta)\sin(\theta)^3+2\cos(\theta)\sin(\theta)\right).\end{aligned}$$

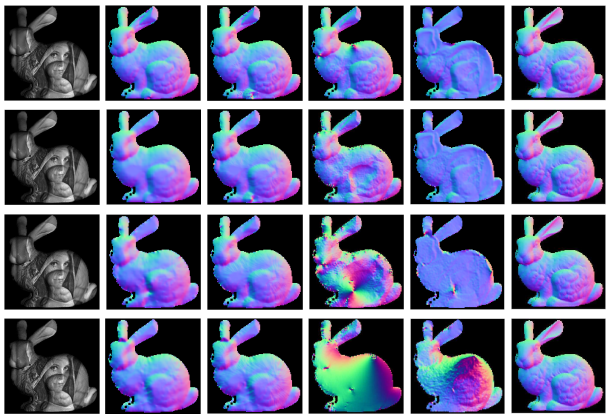
4.2. Hierarchical estimation

To ensure globally consistent resolution of convex/concave ambiguities, we propose to solve the optimisation in a hierarchical setting. Within this setting, we also automatically adjust the weights of the priors such that finescale details can still be recovered at the highest resolution without the smoothness term dominating. From the initial input images, we construct an image pyramid. We initialise at the lowest resolution using a plane and then use the result of each optimisation to initialise the optimisation at the next finer scale by interpolation.

The weights for the two prior constraints are reduced during the optimisation process so that it is initially dominated by the priors and gradually relies more upon the polarisation information. We propose to update w_{smooth} and w_{convex} according to the current ratio model error in the first equation of (17). The initialisation of two weights are calculated by multiplication between two empirically chosen constants and initial mean value of polarisation intensity error vector. Then the weights are updated every 10 iterations during optimisation according to recalculated mean ratio-residual cost.

(a) (b) (c) (d) (e) (f)

Figure 1: Qualitative results on synthetic Blinn-Phong bunny with uniform albedo. See Figure 2 for details.



(a) (b) (c) (d) (e) (f)

Figure 2: Qualitative results on synthetic Blinn-Phong bunny with varying albedo. The four rows are synthetic data with Gaussian noise of standard deviation $\sigma = 0\%, 0.5\%, 1\%, 2\%$ respectively. (a) Input; (b) normal map derived from height recovered by proposed ratio method; (c) normal map from full optimisation method; (d) normal map from [1]; (e) normal map from [15]; (f) ground truth.

5. Experiments

We present experimental results on both synthetic and real data. We compare the two proposed methods (the ratio-based formulation and the full optimisation) to both classical [1] and state-of-the-art [15] methods.

Synthetic data We use the Stanford Bunny height map and render unpolarised intensity images with light source $\mathbf{s} = [\sin(15°), 0, \cos(15°)]^T$ the Blinn-Phong model. We experiment with both uniform albedo and varying albedo (for which we use the Lena image).

We simulate polarisation using (1) and vary the polariser angle from 0° to 180° in 30° increments. Finally, we corrupt the data by adding Gaussian noise with zero mean and varying standard deviation, saturate and quantise to 8 bits. We use these noisy synthetic images as input.

We report the RMS errors of the surface height and mean angular errors of the surface normal in Tab. 1. The ratio-based method offers good performance and is relatively unaffected by varying albedo. The subsequent refinement using the full optimisation further improves performance and always outperforms the comparison methods. Qualitatively, the ratio method sometimes makes convex/concave errors (like flipping the bunny ear to a convex shape) that are partially corrected by the full optimisation.

The visual result of synthetic data is shown in Fig. 1 and 2. Normal maps are visualised as $R = (n_x + 1)/2$, $G = (n_y + 1)/2$ and $B = (n_z + 1)/2$. For comparison method [1], surface normals are estimated directly. For the proposed method and comparison method [15], surface height is estimated and we compute surface normals using finite difference approximations of the gradient of the recovered surface. Comparing the estimated normal maps, we can see that our full optimisation method can recover more fine details than the ratio method and is still able to recover lots of details under significant noise. With varying albedo, the results of both our proposed methods are much better than the two comparison methods [1,15]. With a good initial shape estimation from the ratio method, the full optimisation method can calculate a varying albedo map as long as light source direction is known, and add details beyond the result of the ratio method. The boundary propogation method [1] can handle varying albedo but is extremely sensitive to noise and incorrectly resolves convex/concave interpretations in some places. The linear method [15] degrades less gracefully with noise, with the Laplacian smoothness term dominating and the resulting surface being very flat. Since it is not invariant to albedo, this method fails completely for the varying albedo case.

Real data We show qualitative results on real images in Figures 3, 5 and 7 (zoom for detail). In each case (a) shows an input image, (b)-(d) show estimated depth and normal maps for the ratio-based method, the full optimisation and [15] respectively, (e) and (f) show re-renderings of the surfaces recovered by the ratio-based method and the full optimisation respectively. In general, the results of [15] suffer from flattening in specular regions since they assume the normals in specular pixels all point in the halfway direction. Our ratio method avoids this assumption. Our full optimisation result is initialised by the ratio method and is able to improve fine details. Note particularly in 7(f) that our method is able to recover the fine detail in the writing on the handle of the watergun. The object in Figure 6 contains varying albedo. This causes [15] to fail completely while the ratio method is invariant to these variations and the initialisation of the full optimisation using the albedo and depth estimated by the ratio method remains stable.

Setting	Method	$\sigma = 0\%$		$\sigma = 0.5\%$		$\sigma = 1\%$		$\sigma = 2\%$	
		Height (pix)	Normal (deg)	Height (pix)	Normal (deg)	Height (pix)	Normal (deg)	Height (pix)	Normal (deg)
Uniform albedo	Prop. Ratio	7.89	8.82	8.86	11.16	9.77	12.78	9.89	18.92
	Prop. Full	**7.70**	**7.12**	**7.70**	**7.16**	**7.72**	**7.27**	**7.61**	**7.56**
	[15]	13.47	8.60	8.10	10.18	18.51	16.30	19.00	29.76
	[1]	37.25	42.02	34.56	40.31	36.01	42.47	35.84	44.01
Varying albedo	Prop. Ratio	9.81	13.59	11.90	17.79	10.92	17.58	10.43	21.14
	Prop. Full	**7.61**	**7.31**	**7.62**	**7.41**	**7.59**	**7.77**	**7.60**	**8.69**
	[15]	10.42	15.64	11.17	15.39	13.36	17.27	17.35	22.39
	[1]	36.68	42.14	42.81	43.34	34.96	44.17	42.33	46.52

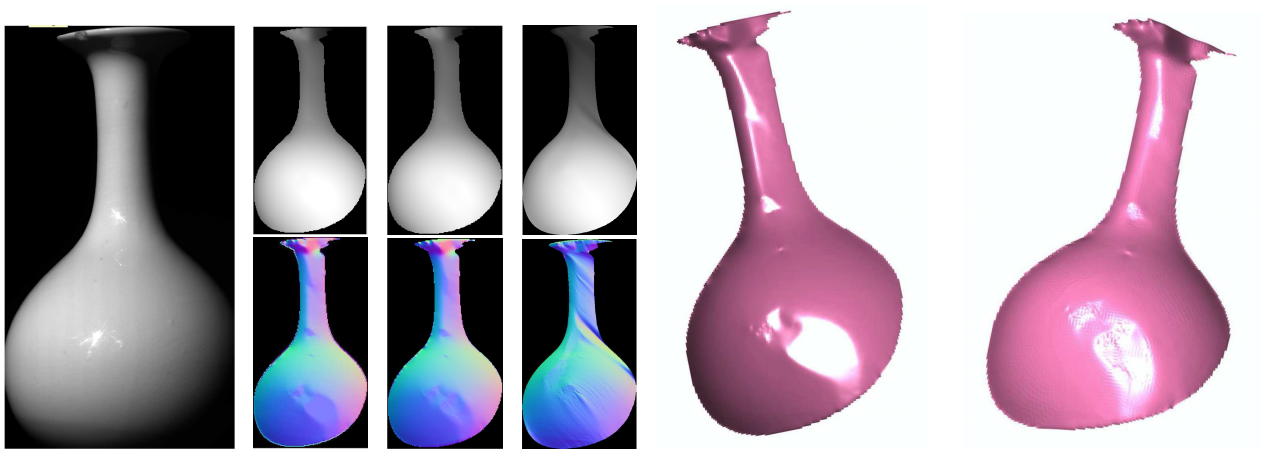
Table 1: Height and surface normal errors on synthetic data. Results shown for proposed ratio and full optimisation method and two comparison methods.

6. Conclusions

We have proposed a new approach to the SfPol problem[1]. Although we use a very general method (nonlinear least squares), the approach is still able to obtain state-of-the-art results from a planar initialisation. In contrast to previous methods, our estimated shape is optimal with respect to an explicit noise model. Previous work (e.g. [1, 6–8, 10, 15]) implicitly assumes Gaussian noise when estimating a polarisation image using least squares. However, uncertainty in the estimated quantities is ignored in the subsequent shape estimation, so the reconstructed shape is not optimal with respect to the assumed noise model.

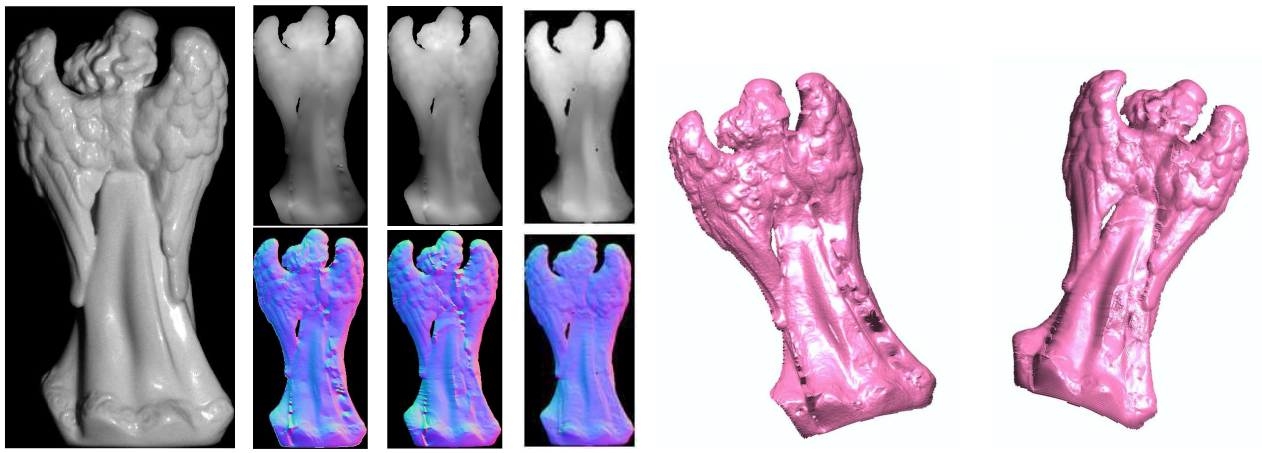
There are many ways in which this work could be extended. A mixed diffuse/specular polarisation model could be used and specular reflectance could be modelled. Albedo and lighting parameters could be estimated as part of the optimisation process, though this would probably necessitate additional priors to regularise the problem. Since we present the cost function as a series of differentiable steps, it would be very easy to incorporate into a convolutional neural network to be trained by backpropagation. For example, a network being trained to regress depth from polarisation measurements could use our transformations as layers to compute a loss between the intensities computed from the estimated depth and the input.

[1]Source code: github.com/waps101/polarisation_optimisation



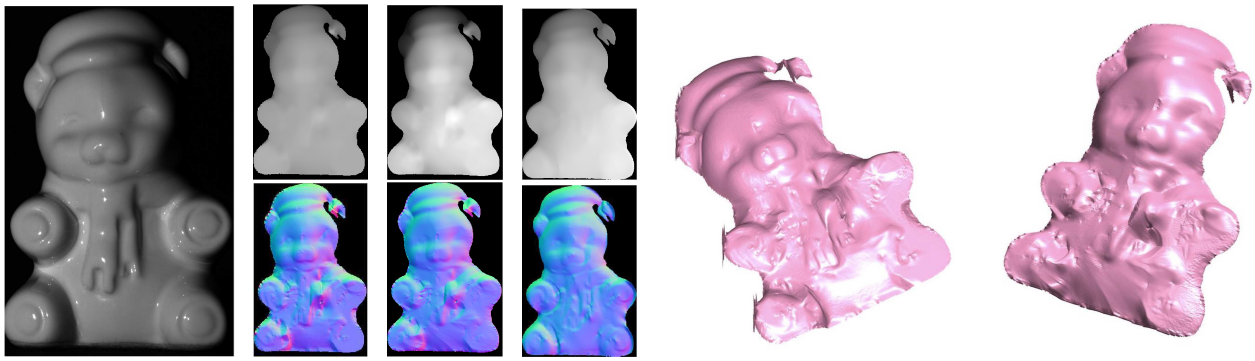
(a) (b) (c) (d) (e) (f)

Figure 3: Qualitative results on porcelain vase. See Fig. 7 caption for details.



(a) (b) (c) (d) (e) (f)

Figure 4: Qualitative results on porcelain angel statue. See Fig. 7 caption for details.

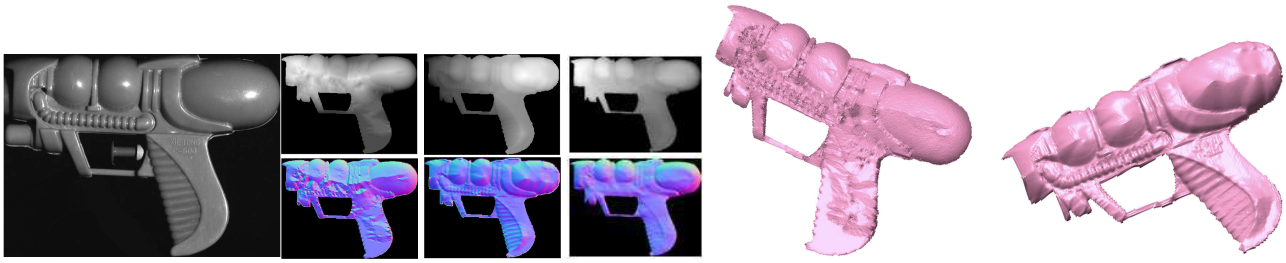


(a) (b) (c) (d) (e) (f)

Figure 5: Qualitative results on porcelain bear. See Fig. 7 caption for details.

(a) (b) (c) (d) (e) (f)

Figure 6: Qualitative results on color porcelain teapot. See Fig. 7 caption for details.



(a) (b) (c) (d) (e) (f)

Figure 7: Qualitative results on plastic watergun: (a) Input grayscale image; (b) Recovered depth map and normal map from ratio method; (c) Recovered depth and normal map by full polarisation model and estimated light source; (d) Recovered depth map and normal map from [15] (e) a new pose of object estimated from ratio method. (f) a new pose of captured object calculated from full polarisation model.

References

[1] G. A. Atkinson and E. R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Trans. Image Process.*, 15(6):1653–1664, 2006. 1, 2, 3, 5, 6

[2] G. A. Atkinson and E. R. Hancock. Shape estimation using polarization and shading from two views. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11), 2007. 1

[3] G. A. Atkinson and E. R. Hancock. Surface reconstruction using polarization and photometric stereo. In *Proc. CAIP*, pages 466–473, 2007. 1

[4] D. V. Hinkley. On the ratio of two correlated normal random variables. *Biometrika*, 56(3):635–639, 1969. 2

[5] C. P. Huynh, A. Robles-Kelly, and E. Hancock. Shape and refractive index recovery from single-view polarisation images. In *Proc. CVPR*, pages 1229–1236, 2010. 2

[6] C. P. Huynh, A. Robles-Kelly, and E. R. Hancock. Shape and refractive index from single-view spectro-polarimetric images. *Int. J. Comput. Vision*, 101(1):64–94, 2013. 2, 6

[7] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proc. ICCV*, pages 3370–3378, 2015. 1, 2, 6

[8] A. H. Mahmoud, M. T. El-Melegy, and A. A. Farag. Direct method for shape recovery from polarization and shading. In *Proc. ICIP*, pages 1769–1772, 2012. 1, 2, 6

[9] R. Mecca, A. Wetzler, A. M. Bruckstein, and R. Kimmel. Near field photometric stereo with point light sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770, 2014. 1

[10] D. Miyazaki, R. T. Tan, K. Hara, and K. Ikeuchi. Polarization-based inverse rendering from a single view. In *Proc. ICCV*, pages 982–987, 2003. 1, 2, 6

[11] O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In *Electronic Imaging*, pages 178–186, 2005. 2

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

[13] S. Rahmann and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proc. CVPR*, volume 1, pages I–I, 2001. 1, 2

[14] Y. Y. Schechner. Self-calibrating imaging polarimetry. In *Proc. ICCP*, pages 1–10, 2015. 2

[15] W. A. Smith, R. Ramamoorthi, and S. Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *Proc. ECCV*, pages 109–125, 2016. 1, 2, 4, 5, 6, 8

[16] S. Tozza and M. Falcone. A comparison of non-lambertian models for the shape-from-shading problem. In *Perspectives in Shape Analysis*, pages 15–42. 2016. 1

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

[18] L. B. Wolff and T. E. Boult. Constraining object features using a polarization reflectance model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):635–657, 1991. 1