

3D shape reconstruction using a polarisation reflectance model in conjunction with shading and stereo

Dizhong Zhu

Doctor of Philosophy

University of York

Department of Computer Science

January 2021

Abstract

Reconstructing the 3D geometry of objects from images is a fundamental problem in computer vision. This thesis focuses on shape from polarisation where the goal is to reconstruct a dense depth map from a sequence of polarisation images.

Firstly, we propose a linear differential constraints approach to depth estimation from polarisation images. We demonstrate that colour images can deliver more robust polarimetric measurements compared to monochrome images. Then we explore different constraints by taking the polarisation images under two different light conditions with fixed view and show that a dense depth map, albedo map and refractive index can be recovered.

Secondly, we propose a nonlinear method to reconstruct depth by an end-to-end method. We re-parameterise a polarisation reflectance model with respect to the depth map, and predict an optimum depth map by minimising an energy cost function between the prediction from the reflectance model and observed data using nonlinear least squares.

Thirdly, we propose to enhance the polarisation camera with an additional RGB camera in a second view. We construct a higher-order graphical model by utilising an initial rough depth map estimated from the stereo views. The graphical model will correct the surface normal ambiguity which arises from the polarisation reflectance model. We then build a linear system to combine the corrected surface normal, polarimetric information and rough depth map to produce an accurate and dense depth map.

Lastly, we derive a mixed polarisation model that describes specular and diffuse polarisation as well as mixtures of the two. This model is more physically accurate and allows us to decompose specular and diffuse reflectance from multiview images.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Thesis Outline	2
2	Background and Related Work	5
2.1	Background	5
2.1.1	Polarisation phenomenon	5
2.1.2	Polarisation reflectance model	10
2.1.3	Assumptions	14
2.1.4	Specular reflection polarisation model	15
2.1.5	Diffuse polarisation model	17
2.1.6	Polarisation information from polarimetric image data	19
2.1.7	Multi-view stereo	24
2.2	Related work	27
2.2.1	Monocular polarisation	27
2.2.2	Two view Polarisation	29
2.2.3	Multi-view Polarisation	32

2.2.4	Limitations	33
2.3	Conclusion	35
3	Underpinning methods for shape-from-polarisation	36
3.1	Multichannel estimation	37
3.1.1	Solve unpolarised intensities	38
3.1.2	Solve degree of polarisation and phase angle	38
3.1.3	Alternating optimisation	39
3.2	Polarisation image estimation with multiple light sources	41
3.2.1	Solve unpolarised intensity	41
3.2.2	Solve degree of polarisation and phase angle	42
3.3	Least square reconstruction of surface height	43
3.3.1	Linear least squares height-from-normals	44
3.3.2	Linear equations	44
3.3.3	Discrete formulation	45
3.3.4	Numerical differentiation kernels	46
3.3.5	2D Savitzky-Golay filters	47
3.3.6	K-nearest pixels kernel	48
3.3.7	3D K-nearest neighbours kernel	49
3.3.8	Implementation	50
3.4	Summary	50
4	Monocular shape-from-polarisation	51
4.1	Shape-from-polarisation by nonlinear least squares	52

4.1.1	Shape-from-polarisation as analysis by synthesis	53
4.1.2	Nonlinear least squares	56
4.1.3	Priors	56
4.1.4	Hierarchical estimation	57
4.1.5	Ratio-based formulation	58
4.2	Photo-polarimetric stereo	59
4.2.1	Photo-polarimetric height constraints	60
4.2.2	A unified PDE formulation	63
4.2.3	Height estimation via linear least squares	66
4.2.4	Refractive index estimation	67
4.2.5	Two source lighting estimation	67
4.3	Experiments and evaluation	68
4.3.1	Evaluating nonlinear square method	69
4.3.2	Evaluating photo-polarimetric stereo method	73
4.4	Summary	79
5	Depth from a polarisation + RGB stereo pair	81
5.1	Perspective depth representation	82
5.2	Integrability-based disambiguation with a higher order graphical model	83
5.2.1	Unary cost	84
5.2.2	Pairwise cost	85
5.2.3	Ternary cost	85
5.2.4	Graphical model optimisation	85

5.3	Albedo estimation with gradient consistency	86
5.4	Linear perspective depth from polarisation	87
5.4.1	Phase angle constraint.	87
5.4.2	Shading/polarisation ratio constraint.	89
5.4.3	Surface normal constraint.	89
5.4.4	Global linear depth estimation.	89
5.5	Experimental results	91
5.6	Summary	95
6	Mixed polarisation model with Multi-view stereo	96
6.1	Mixed Polarisation Model	96
6.2	Multi-view constraint	99
6.2.1	Polarisation model under camera coordinate	100
6.2.2	Reformulation under Multi-view stereo	100
6.3	Estimation of mixed polarisation model	101
6.3.1	Extract correspondences in image space	101
6.3.2	Optimisation strategy	102
6.4	Summary	103
7	Conclusion and Future work	106
7.1	Summary	106
7.2	Overarching conclusions	107
7.3	Future work	109

List of Figures

1.1	(a) left image shows a scene under glare light without wearing sun glasses, right image shows a similar scene wearing sun glasses. (b) left images shows the photo captured without polarising filter, right image shows the photo captured with a polarising filter.	3
2.1	Double refraction occurs in calcite because it is an anisotropic crystal and the o-ray wave front propagates as a sphere, whereas the e-ray propagates as an ellipsoid [18].	6
2.2	Illustration of two orthogonal waves, assume the plane of our paper is incident plane, then E_y is <i>p-polarised</i> component while E_x is <i>s-polarised</i> component.	6
2.3	An illustration of some polarisation states. Imagine that the instantaneous state of a light wave will move along the arrow on the ellipse.	7
2.4	[18] Classical method to measure the Stoke parameters, imagine the E_x, E_y of optical ray align with x, y axis respectively.	11
2.5	The degenerate polarisation states and their corresponding Stokes vectors. The linearly polarised lights are degenerates case of the ellipse equation.	12
2.6	An optical ray with direction \vec{k}_i propagates from medium 1 to medium 2, which produces a reflected ray \vec{k}_r and a transmitted ray \vec{k}_t at the interface between medium 1 and medium 2. The reflected angle $\theta_r = \theta_i$ and the transmitted angle θ_t is determined by the Snell's law. E, B present the amplitudes of electronic field and magnetic field of optical ray respectively. And direction of E is perpendicular to the incident plane while direction of B is coplanar with incident plane.	14

2.7	An experiment setup to capture polarimetric images. Compare to Figure 2.4, we remove the phase retarder so that only linear polarisation information will be captured.	15
2.8	The incident ray is coplanar with surface normal, while the most energy of reflected ray is <i>s-polarised</i> which is perpendicular to the incident plane. The surface normal is oriented at ϕ_i when mapped to the image plane, so the angle between polariser and reflected ray is $\vartheta - \phi_i + \frac{\pi}{2}$	16
2.9	The optical ray emitted from a light source will bounce multiple times inside the object which makes the optical ray remain unpolarised. Then the optical ray is transmitted out from the object and received by the detector.	18
2.10	The incident ray is coplanar with surface normal and refracted ray, for the most energy of refractive ray is <i>p-polarised</i> which is coplanar to the incident plane. The surface normal is oriented at ϕ_i when mapped to the image plane, so the angle between polariser and refracted ray is $\vartheta - \phi_i$	18
2.11	Sony's Polarsens 4 Pixel Block Polarizer Design. source: https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/	21
2.12	The upper row shows input images that are captured by a camera with a mounted linear polariser. The upper right numbers show the degree of angle rotated by the polariser. We plot the samples by red cross and can see they are not a perfect sinusoidal curve due to noise. But we show using more images is more robust to noise.	22
2.13	The 3 images show I_{un}, ρ, φ respectively solved by (2.40).	22
2.14	A plot shows the function of degree of polarisation w.r.t the zenith angle (θ) with given refractive indices 1.3 and 1.6. (a) Left image presents the specular component, where we can see there are two θ map to the same value of DoP (marked with red circles) except the peak. (b) Right image presents the diffuse component, whose θ uniquely maps to one value of degree of polarisation.	24

2.15	[35] Left image. The 3D point \mathbf{X} and the corresponding image positions are denoted by x and x' where their camera centres are C and C' . \mathbf{X}, x, x', C and C' lay on the same plane called the epipolar plane. Right image. When we do not know the 3D position of point \mathbf{X} and only know its pixel location x in left image. The 3D position \mathbf{X} must lie on the ray that connects camera center C and x . The correspondences in the right image lie on this epipolar line l'	25
2.16	[24] "Winner-takes-all" procedure, pick the depth value which has the highest normalized cross correlation (NCC) Score.	26
2.17	[16] The NCC Peak are potential depth values (denoted by blue triangles), and we are looking for an optimum depth to minimise the cost built by (2.45). After optimisation, the depth value in the green box will be set as optimum.	27
2.18	[70] Two ambiguity depth with estimated light source	28
2.19	[52] The derivative of DoP gives a clue to disambiguate θ	30
2.20	[42] (a) The coarse depth map captured by Kinect. (b) Input polarimetric images. (c) Recovered shape from polarisation images. (d) Using coarse depth map to disambiguate result of (c). (e) Enhanced coarse depth map.	32
2.21	a).The top left illustrate the setup of multi-view polarisation. b). The top right shows the estimated azimuth angle of the car. c). The bottom left shows the reconstruction from multi-view stereo. d). The bottom right shows the reconstruction result of multi-view polarisation. [20]	33
3.1	Multichannel polarisation image estimation. Top row: the input image sequence; Second row left: degree of polarisation (ρ) and right: phase angle (φ) estimated from a single channel; Third row left: degree of polarisation (ρ) and right: phase angle (φ) estimated from three colour channels and two light source directions. .	40
3.2	An example shows multichannel estimation on snooker ball images.	43

3.3	An example of computing 2D Savitzky-Golay filters on an arbitrary domain. In this example, we use a 3×3 kernel. For point A we can use the default square kernel. The order two Savitzky-Golay filter for the horizontal derivative is shown below as \mathbf{K}_A . For point B we use the 3^2 nearest pixels and build a custom order two Savitzky-Golay filter shown below as \mathbf{K}_B . In practice, higher order kernels provide better performance.	49
4.1	An illustration on how the hierarchical estimation works. We start from the lowest resolution with a plane as initialisation. The result from each layer will rescale to a proper resolution as an initialisation to the next optimisation. The final shape estimation will produce by the last layer which keep the original resolution. . . .	58
4.2	The zenith angle θ is just the angle between the surface normal and the viewer. For an orthographic camera, v is constant across the image.	60
4.3	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 [6]; (e) normal map from [70]; (f) ground truth.	70
4.4	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 [6]; (e) normal map from [70]; (f) ground truth.	71
4.5	Qualitative results on porcelain vase. See Fig. 4.9 caption for details.	72
4.6	Qualitative results on porcelain angel statue. See Fig. 4.9 caption for details. . .	72
4.7	Qualitative results on porcelain bear. See Fig. 4.9 caption for details.	72
4.8	Qualitative results on color porcelain teapot. See Fig. 4.9 caption for details. . .	73

4.9	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 [70] (e) a new pose of object estimated from ratio method. (f) a new pose of captured object calculated from full polarisation model.	73
4.10	Estimation on synthetic data of 3DRFE dataset	74
4.11	We show our depth reconstruction and surface normal error with our proposed method, Mecca2017 [51], SIRFS [10] and Smith2018 [69]. And the light estimation has been showned in Table 4.3	75
4.12	We show albedo estimation comparison with different methods, from left to right which are proposed, Mecca2017 [10], and Smith2018 [69].	76
4.13	Qualitative results on real objects with varying albedo obtained by using Prop. 1+3 and comparison to [71] (zoom for detail).	78
4.14	Albedo estimation on real data with different methods from left to right which are proposed, Mecca2017 [51], and Smith2018 [69].	79
5.1	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).	82
5.2	(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.	86
5.3	(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.	88
5.4	Albedo estimates on synthetic data.	92
5.5	Qualitative shape estimation results on synthetic data with comparison with [70]	92

5.6	We show our results on complex object. From left to right we show an image from the input sequence; Depth from stereo reconstruction [37]; Our proposed estimated albedo map and the estimated depth. Depth estimation by [85].	93
5.7	Comparison on [42] dataset. Top-left: One of the polarisation intensity images and Kinect depth map. Top-right: our result. Bottom-Left: [42]. Bottom-Right: [85].	94
6.1	An experiment setup to capture the polarimetric images. Compare to Figure 2.4, we remove the phase retarder so that only linearly polarisation information will be captured.	97
6.2	We set $I_d = 0.5$ with two different specular intensity of the mixed polarisation model. (a) The curves shows the value of $\frac{I_d\rho_d - I_s\rho_s}{I_d + I_s}$.(b) The curves shows the value of $ \frac{I_d\rho_d - I_s\rho_s}{I_d + I_s} $	99
6.3	We simulate the Stanford bunny in two different views, we add noise to the ground truth mesh which provides a initialisation for θ and ϕ . The first row shows two view images. The second row shows our diffuse intensity, and specular intensities in two views(from left to right). The third row shows the ground truth of the diffuse and specular intensities.	104
6.4	We shows real data images that capture under natural environment, we capture each images in four different views. Column a).Original images captured under nature illumination. b).Diffuse intensity per vertex. c).Specular intensity per vertex.	105

List of Tables

2.1	Each proposed method relaxes some of the limitations in previous work. (a) Non-uniform albedo indicates the method is able to apply on the object where the albedo is non-uniform. (b) Refractive index indicates the method do not need the pre-knowledge of the refractive index value of the object, and the proposed method is able to estimate the refractive index value instead. (c) Metric reconstruction indicates the method is able to reconstruct a metrically accurate depth instead of height map up to unknown scale. (d) Mixed reflectance indicates the method is able to deal with the pixel intensity mixed with diffuse and specular reflectance. We do not need to make an assumption that the pixel is either diffuse or specular dominant.	34
4.1	Summary of the different formulations	63
4.2	Height and surface normal errors on synthetic data. Results shown for proposed ratio and full optimisation method and two comparison methods.	70
4.3	Light estimation on snooker ball image and 3DRFE image. While the ground truth are $[-0.51, 0, 0.86]$, $[0, -0.51, 0.86]$ respectively.	77
4.4	We compare albedo absolute mean error + standard deviation and compare shape mean angular error + standard deviation for Figure 4.11 with different method.	77
4.5	We estimate the refractive index of different objects that in our dataset.	79
5.1	Summary of the different method	91

5.2	Mean absolute difference in depth and mean angular surface normal errors on synthetic data. For [69, 70] methods reconstructed the depth up to scale we compute the optimum scale to align with the ground truth depth map.	91
-----	---	----

Acknowledgement

I would like to give great thank to everyone who gave me any kind of helps during my four years PhD. Because of their help and support I can finish my PhD at end.

My supervisor Dr William A. P. Smith deserve a great thank during my PhD. He guides me in the research and make me clear about my goal. His generous and knowledge helps me to overcome obstacles during the research. And the way he thinks encourage me to dig out the fundamentals of the problems. Also, I would like to give thank to Prof Edwin Hancock who gives me feedback during my whole PhD.

Thanks to my friends Dr Jiang zhe, Dr Chao Zhang, Dr Ye Yu, Dr Xiaotian Dai. Dr Viswadeep Sarangi. I really enjoyed the time that spent with all of them. I am not able to mention all your names but I do really appreciate everyone in the computer science department.

At last, I would like to give my warmest thank to my families. My wife Yiwen Li, who always helps me and supports me whole-heartily. With her love and deep understanding, everything becomes joyful. To my parents, I would like to thank their unconditional love and support.

Declaration

I declare that all work contained within this thesis is a result of my own investigations, except where explicit attribution has been given. The content of some of the chapters has already been published within the following publications:

- *Tozza, Silvia, William AP Smith, **Dizhong Zhu**, Ravi Ramamoorthi, and Edwin R. Hancock. “Linear differential constraints for photo-polarimetric height estimation.” In Proceedings of the IEEE International Conference on Computer Vision, pp. 2279-2287. 2017.*
- *Yu, Ye *, **Dizhong Zhu***, and William AP Smith. “Shape-from-polarisation: a nonlinear least squares approach.” Proceedings of the IEEE International Conference on Computer Vision. 2017.* indicates equal contribution*
- ***Dizhong Zhu**, and William AP Smith. “Depth from a polarisation+ RGB stereo pair.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.*
- ***Dizhong Zhu**, and William Alfred Peter Smith. “Least squares surface reconstruction on arbitrary domains.” 16th European Conference on Computer Vision (ECCV 2020) Proceedings. Springer-Verlag, 2020.*

Chapter 1

Introduction

1.1 Motivation

Retrieving 3D information about our surroundings is one of the fundamental problems in computer vision. Objects with different shapes and appearances can be perceived by human and animals easily, but it is a non-trivial task in computer vision. Over the last few decades, both academic and industrial researchers have been inspired and dedicated to develop algorithms to make machines able to perceive and produce a wide range of applications. For instance, creating robots that can replace humans in tedious and dangerous work such as autonomous driving aims to manoeuvre vehicles safely while perceiving different traffic, road conditions etc. Rescue robots can enter into dangerous and complex environments where humans cannot safely reach. Logistics robots can manage warehouses and transfer goods to save a lot of human resources. All these related tasks involve a fundamental problem: how to receive the data from our 3D world. Several methods and sensors have been developed to tackle this problem.

This thesis focuses on reconstructing 3D shape from polarisation images. A polarisation image records polarisation state at each pixel. Computational polarisation vision seeks to exploit this additional information in order to tackle computer vision tasks. This is a disruptive technology, with the promise of providing new approaches to computer vision problems and broad potential applications. On the other hand, we already see a wide range of applications utilising polarisation in our daily life. You might experience that while walking or driving on a sunny day, the glare illumination will blind your view. By wearing a pair of polarised sunglasses, glare will be removed and your view will become clearer. Or you would like to take a photo of a scene

under the river while you are resting ashore, and the scene under the river has been concealed by the reflected light. By mounting a polarising filter on your camera, adjusting it then you can capture a clear view under the water surface. An example has been shown in Figure 1.1. There are many other critical applications developed by the use of polarisation such as Liquid Crystal Display, 3D stereo movies etc.

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 [82]. The reason for this phenomenon is that unpolarised light becomes partially polarised when it is reflected specularly [64] or diffusely via subsurface scattering [6]. 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 rotating linear polarising filter in front of a conventional camera, this rich source of additional information becomes available.

A polarisation camera captures additional information about the surface orientation of objects. These surface features can greatly enhance methods which just use intensity information alone. (a) It provides surface normal information on featureless regions where stereo matching methods fail. (b) A proper formulation can achieve diffuse and specular separation in natural illumination which will be addressed in this thesis. (c) Detailed reconstruction is possible on pixels with high specularity and inter-reflection while multi-view stereo and shape from shading completely fail.

1.2 Thesis Outline

The remainder of this thesis is organised as follows:

- **Chapter 2. Background and Related work:** We go through the basic theory of polarisation and show the derivation of the polarisation reflectance model which is widely used in computer vision. We then review the related work involved with polarisation.
- **Chapter 3. Underpinning methods for shape-from-polarisation:** In the first part, we propose a multichannel estimation method that utilises chromatic polarimetric intensity images to suppress noise in the polarisation decomposition. This improves the quality of

the polarisation images and will better serve our depth estimation from polarisation images. In the second part, we propose a linear least squares approach to formulate the surface height from gradient. Moreover, we developed a 2D Savitzky-Golay filter which provides a higher order surface derivative approximation that can be applied to arbitrary domains.

Dizhong Zhu, and William AP Smith. "Least squares surface reconstruction on arbitrary domains." Proc. ECCV (2020).

- **Chapter 4. Monocular shape-from-polarisation:** We propose two approaches with active light sources to address the problems of depth estimation and albedo estimation



(a) source: <https://www.misterspex.co.uk/sunglasses-guide/sunglasses-lenses>



(b) source: <http://www.paddling.net/sameboat/archives/sameboat496.html>

Figure 1.1: (a) left image shows a scene under glare light without wearing sun glasses, right image shows a similar scene wearing sun glasses. (b) left images shows the photo captured without polarising filter, right image shows the photo captured with a polarising filter.

under a monocular polarisation setup. (a) We first use a nonlinear approach to estimate the object depth under one view with one light source. We re-parameterise the diffuse polarisation reflectance function w.r.t depth. The optimum depth should minimise the cost between predicted intensity and observed polarimetric images. (b) In the second approach, we utilise two different light sources to build a linear system to reconstruct the depth and achieve albedo estimation.

Tozza, Silvia, William AP Smith, Dizhong Zhu, Ravi Ramamoorthi, and Edwin R. Hancock. "Linear differential constraints for photo-polarimetric height estimation." In Proceedings of the IEEE International Conference on Computer Vision, pp. 2279-2287. 2017.

*Yu, Ye *, Dizhong Zhu*, and William AP Smith. "Shape-from-polarisation: a non-linear least squares approach." Proceedings of the IEEE International Conference on Computer Vision. 2017.* indicates equal contribution*

- **Chapter 5. Depth from a polarisation + RGB stereo pair:** We propose a hybrid depth imaging system in which a polarisation camera is augmented by a second image from a standard digital camera. The method includes the following key steps: (a) A graphical model to solve the normal ambiguity and label each pixel as diffuse or specular dominant by minimising a novel energy. (b) With diffuse label and corrected normal we estimate the albedo of the object. (c) Build a linear system to solve the depth under perspective camera model.

Dizhong Zhu, and William AP Smith. "Depth from a polarisation+ RGB stereo pair." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

- **Chapter 6. Mixed polarisation model with Multi-view stereo:** We propose a comprehensive mixed polarisation model of both specular and diffuse polarised reflectance. We then propose a novel method for fitting this model to multi-view data. We emphasise that this line of work is not yet complete but provides a first attempt in this direction which shows promising results.
- **Chapter 7. Conclusion and Future work:** This chapter summarises the main contributions from this thesis and suggests potential future works based on polarisation.

Chapter 2

Background and Related Work

This chapter reviews the fundamental theory of the polarisation phenomenon and explains how to derive the polarisation reflectance model which is widely used in shape-from-polarisation. We then review some important papers that use polarisation images for different computer vision tasks.

2.1 Background

2.1.1 Polarisation phenomenon

In 1669 *Rasmus Bartholin* discovered double refraction of a light ray. When a single ray of natural incident light propagated through a calcite crystal it caused two rays to emerge, which demonstrated that a single ray of light actually consists of two rays [11] as shown in Figure 2.1. *Christiaan Huygens* further showed that by allowing these two emerged lights to pass through a second calcite crystal, and rotating it, the intensity of one emerged ray was maximised while another vanished. At a 45° rotation degree, the intensities of these two rays were equal. They observed this opposite behaviour of intensity and the two rays were said to be polarised. Until in the early 19th century, *Augustin-Jean Fresnel* proposed his *Fresnel's wave theory* to fully explain this phenomena of light. Furthermore, *Fresnel* and *Arago* experimentally showed that the optical field can be decomposed to two orthogonal components in the plane transverse to the direction of propagation [18, 38]. The component of electric wave perpendicular to the incident plane is called *s-polarised* (or *Transverse Electric*) while the one of magnetic wave perpendicular to the incident plane is called *p-polarised* (or *Transverse Magnetic*) as shown as

E_x, E_y respectively in Figure 2.2. The two optical components are waves that we can present them in following forms:

$$\begin{aligned}\vec{E}_x(\vec{r}, t) &= E_{0x}e^{j(\vec{k}\cdot\vec{r}-\omega t+\delta_x)} \\ \vec{E}_y(\vec{r}, t) &= E_{0y}e^{j(\vec{k}\cdot\vec{r}-\omega t+\delta_y)}\end{aligned}\quad (2.1)$$

where E_{0x}, E_{0y} are the maximum amplitudes of the two waves respectively, \vec{k} is the direction of propagation, \vec{r} is a position vector, δ_x, δ_y are arbitrary phases.

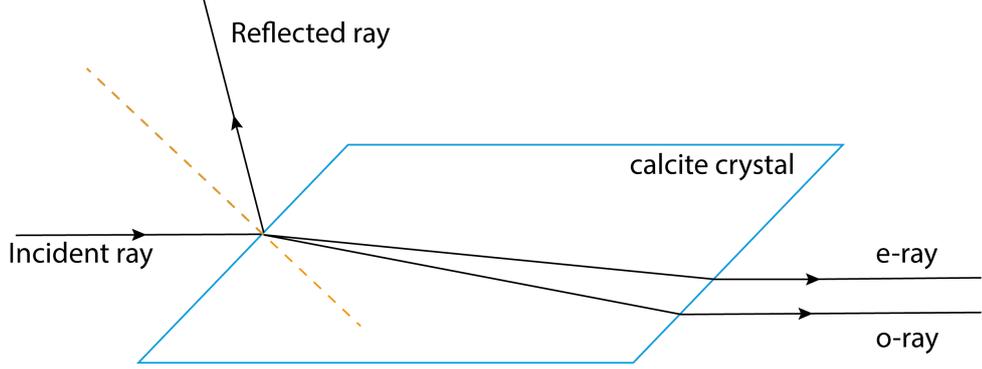


Figure 2.1: Double refraction occurs in calcite because it is an anisotropic crystal and the o-ray wave front propagates as a sphere, whereas the e-ray propagates as an ellipsoid [18].

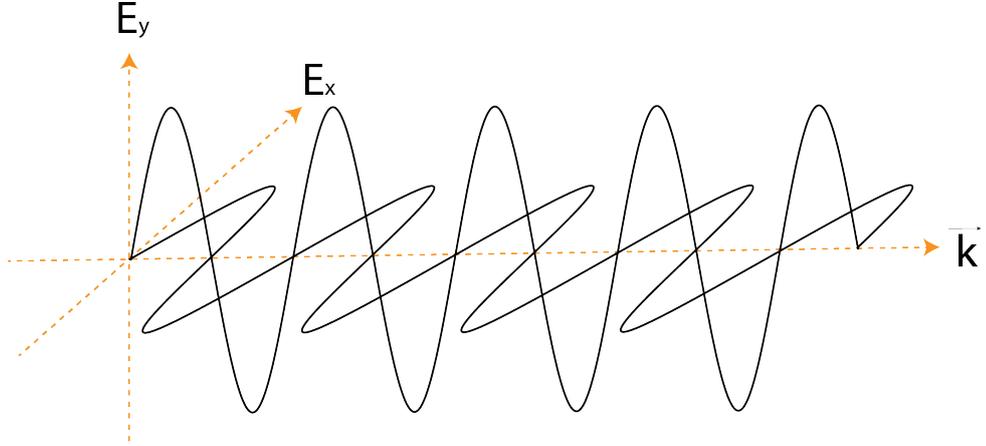


Figure 2.2: Illustration of two orthogonal waves, assume the plane of our paper is incident plane, then E_y is *p-polarised* component while E_x is *s-polarised* component.

Polarisation ellipse. A classical way to describe and visualise the polarisation state is to use the polarisation ellipse [18]. It takes the real part of the equations in (2.1) with certain computation that leads to an equation of an ellipse:

$$\frac{|\vec{E}_x(\vec{r}, t)|^2}{E_{0x}^2} + \frac{|\vec{E}_y(\vec{r}, t)|^2}{E_{0y}^2} - 2\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \cdot \frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \cos(\delta_y - \delta_x) = \sin^2(\delta_y - \delta_x) \quad (2.2)$$

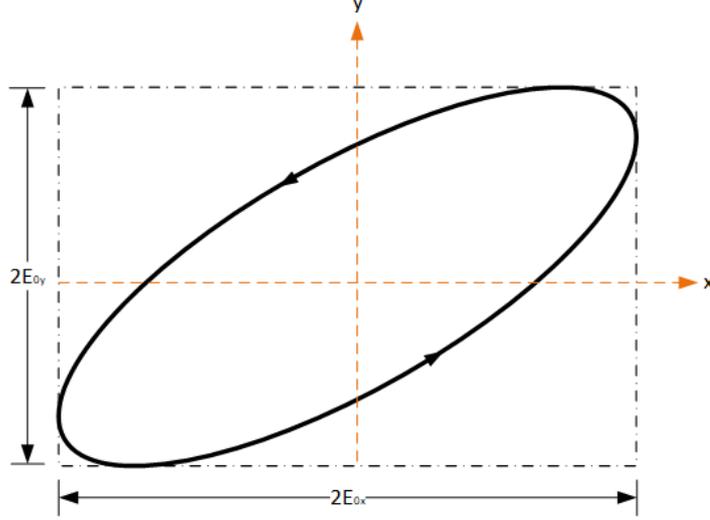


Figure 2.3: An illustration of some polarisation states. Imagine that the instantaneous state of a light wave will move along the arrow on the ellipse.

Proof. In detail, we take the real parts of the equation (2.1) and postulate that $\nu = \vec{k} \cdot \vec{r} - \omega t$ then we have

$$\begin{aligned}\vec{E}_x(\vec{r}, t) &= E_{0x} \cos(\nu + \delta_x) \\ \vec{E}_y(\vec{r}, t) &= E_{0y} \cos(\nu + \delta_y)\end{aligned}$$

- 1) Rearrange the above equations and expand them by trigonometric function, we get

$$\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} = \cos \nu \cos \delta_x - \sin \nu \sin \delta_x \quad (2.3)$$

$$\frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} = \cos \nu \cos \delta_y - \sin \nu \sin \delta_y \quad (2.4)$$

- 2) We multiply $\sin \delta_y$ with equation(2.3) and $\sin \delta_x$ with equation(2.4), we have

$$\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \sin \delta_y = \cos \nu \cos \delta_x \sin \delta_y - \sin \nu \sin \delta_x \sin \delta_y \quad (2.5)$$

$$\frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \sin \delta_x = \cos \nu \cos \delta_y \sin \delta_x - \sin \nu \sin \delta_y \sin \delta_x \quad (2.6)$$

- 3) By subtracting equation(2.5) and equation(2.6), we have

$$\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \sin \delta_y - \frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \sin \delta_x = \cos \nu \sin(\delta_y - \delta_x) \quad (2.7)$$

- 4) Similarly, we multiply $\cos \delta_y$ with equation (2.3) and $\cos \delta_x$ with equation (2.4), then subtracting them we have:

$$\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \cos \delta_y - \frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \cos \delta_x = \sin \nu \sin(\delta_y - \delta_x) \quad (2.8)$$

5) We square equations (2.7), (2.8) respectively and adding them, with some trigonometric transformation we have got:

$$\left(\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \sin \delta_y - \frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \sin \delta_x\right)^2 + \left(\frac{\vec{E}_x(\vec{r}, t)}{E_{0x}} \cos \delta_y - \frac{\vec{E}_y(\vec{r}, t)}{E_{0y}} \cos \delta_x\right)^2 = \sin(\delta_y - \delta_x)$$

which is the same as equation (2.1). ■

The shape of the ellipse reveals the polarisation state of the optical field. There are special combinations of amplitude and phase that are critically important. These are: 1) linear polarisation state of horizontal polarised light (**LHP**) and vertical polarised light (**LVP**). 2) Linearly $\pm 45^\circ$ polarised light (**L+45P, L-45P**). 3) Right and left circularly polarised light (**RCP, LCP**). An example has been shown in Figure 2.5

Stokes parameters. The equation (2.2) of an ellipse only presents an instantaneous state of polarised light, and cannot be measured directly. In order to measure the energy of the polarised light, *Sir George Stokes* introduced Stokes polarisation parameters, by taking a time average of the polarisation ellipse can lead to the following equation [36]

$$s_0^2 = s_1^2 + s_2^2 + s_3^2 \quad (2.9)$$

where

$$\begin{aligned} s_0 &= E_{0x}^2 + E_{0y}^2 \\ s_1 &= E_{0x}^2 - E_{0y}^2 \\ s_2 &= 2E_{0x}E_{0y} \cos(\delta_y - \delta_x) \\ s_3 &= 2E_{0x}E_{0y} \sin(\delta_y - \delta_x) \end{aligned}$$

For convenience, we rewrite the stokes parameters into a **Stokes vector**, and the parameters' descriptions can be concluded as follows

$$\mathbf{s} = \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} \begin{array}{l} \rightarrow \text{Intensity of optical ray} \\ \rightarrow \text{preponderance of LHP light over LVP light} \\ \rightarrow \text{preponderance of L+45P light over L-45P light} \\ \rightarrow \text{preponderance of RCP light over LCP light} \end{array} \quad (2.10)$$

The Degree of Polarisation ρ is defined by the equation

$$\rho = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0}, \quad 0 \leq \rho \leq 1 \quad (2.11)$$

We can now model the change of polarisation state as a function $\mathbf{s}' = \mathbf{F}(\mathbf{s})$, where \mathbf{F} is a series of transformations that change the input polarisation state \mathbf{s} to new state \mathbf{s}' . The transformation function \mathbf{F} is actually a linear transformation that can be written as a 4×4 matrix known as a **Muller Matrix** [36]. We can model the change of polarisation state as follows:

$$\begin{bmatrix} s'_0 \\ s'_1 \\ s'_2 \\ s'_3 \end{bmatrix} = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad (2.12)$$

Measuring polarisation state. With equation (2.12), we can measure the Stokes parameters of incident light by using a phase retarder and a rotated linear polariser [18], which is shown in Figure 2.4. The rotating linear polariser can be decomposed to linear polariser and optic rotator. An ideal linear polariser only allows the light aligned with a certain direction to transmit through it and absorbs the rest. The optic rotator allows the light to propagate through it and rotates the polarisation ellipse. The Muller matrix of the rotated linear polariser is

$$\mathbf{M}_{\text{RPOL}}(\vartheta) = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\vartheta & \sin 2\vartheta & 0 \\ \cos 2\vartheta & \cos^2 2\vartheta & \sin 2\vartheta \cos 2\vartheta & 0 \\ \sin 2\vartheta & \sin 2\vartheta \cos 2\vartheta & \sin^2 2\vartheta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (2.13)$$

where ϑ is the angle of rotation *w.r.t* to \vec{E}_x , and the rotation plane is perpendicular to the incident plane. A phase retarder is to shift phase between \vec{E}_x and \vec{E}_y . Concretely, a phase retarder has a fast axis normally along the x axis, and slow axis along y axis, the light through a retarder will shift phase by φ between the orthogonal components. The Muller matrix of the phase retarder can be written as follows:

$$\mathbf{M}_{\text{PR}}(\varphi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \varphi & -\sin \varphi \\ 0 & 0 & \sin \varphi & \cos \varphi \end{bmatrix}. \quad (2.14)$$

We can model this process by (2.12) that \mathbf{s} is the light state in the light source, and \mathbf{s}' is the light state received in the detector. This yields $\mathbf{s}' = \mathbf{M}(\vartheta, \varphi) \cdot \mathbf{s}$, where $\mathbf{M}(\vartheta, \varphi) = \mathbf{M}_{\text{RPOL}}(\vartheta) \cdot \mathbf{M}_{\text{PR}}(\varphi)$,

in detail:

$$M(\vartheta, \varphi) = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\vartheta & \sin 2\vartheta \cos \varphi & -\sin 2\vartheta \sin \varphi \\ \cos 2\vartheta & \cos^2 2\vartheta & \sin 2\vartheta \cos 2\vartheta \cos \varphi & -\sin 2\vartheta \cos 2\vartheta \sin \varphi \\ \sin 2\vartheta & \sin 2\vartheta \cos 2\vartheta & \sin^2 2\vartheta \cos \varphi & -\sin^2 2\vartheta \sin \varphi \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.15)$$

If the detector only measures the first component of \mathbf{s}' , which is the intensity of light ray, it can be simplified as the first row of the $\mathbf{M}(\vartheta, \varphi)$ multiplied with Stokes vector \mathbf{s} :

$$\begin{aligned} I(\vartheta, \varphi) &= (-M_1(\vartheta, \varphi)-) \cdot \mathbf{s} \\ &= \frac{1}{2}(s_0 + s_1 \cos 2\vartheta + s_2 \sin 2\vartheta \cos \varphi - s_3 \sin 2\vartheta \sin \varphi), \end{aligned} \quad (2.16)$$

where $I(\vartheta, \varphi)$ is the intensity of received light ray. In order to compute the Stokes parameters, we first removed the phase retarder, and record the intensity only by rotating the linear polariser in $\vartheta = 0, \frac{\pi}{4}$ and $\frac{\pi}{2}$ angles which we denote the intensity as $I(0, 0), I(\frac{\pi}{4}, 0), I(\frac{\pi}{2}, 0)$ respectively. Then we insert a phase retarder as Figure 2.4 with $\varphi = \frac{\pi}{2}$ and rotated linear polariser in angle $\frac{\pi}{4}$, where denote as $I(\frac{\pi}{4}, \frac{\pi}{2})$:

$$\begin{aligned} s_0 &= I(0, 0) + I(\frac{\pi}{2}, 0) \\ s_1 &= I(0, 0) - I(\frac{\pi}{2}, 0) \\ s_2 &= 2I(\frac{\pi}{4}, 0) - S_0 \\ s_3 &= S_0 - 2I(\frac{\pi}{4}, \frac{\pi}{2}) \end{aligned} \quad (2.17)$$

We utilise (2.16) to develop a more compact way to represent the (2.17), we write it in a matrix form by stacking different polariser rotated angles and whether inserting a phase retarder as

$$\begin{bmatrix} I(\vartheta_1, 0) \\ \vdots \\ I(\vartheta_n, 0) \\ I(\vartheta_1, \frac{\pi}{2}) \end{bmatrix} = \begin{bmatrix} (-M_1(\vartheta_1, 0)-) \\ \vdots \\ (-M_1(\vartheta_n, 0)-) \\ (-M_1(\vartheta_1, \frac{\pi}{2})-) \end{bmatrix} \cdot \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix}. \quad (2.18)$$

This provides a closed form solution by linear least squares [74] and is more robust to noise when $n \geq 3$.

2.1.2 Polarisation reflectance model

The polarisation state of light will change when an optical ray interacts with a surface. This is due to reflection and refraction caused by different refractive indices on two sides of the interface. Where **refractive index** is defined by a ratio between light speed in vacuum and speed in the

medium. To simplify our explanation, we first consider s-polarised optical ray whose electric field is perpendicular to the incident plane as shown in Figure 2.6, where an optical ray propagates from one medium to another medium. $\vec{k}_i, \vec{k}_r, \vec{k}_t$ denotes the propagation directions of incident ray, reflected ray and transmitted ray respectively. Assume the optical ray \vec{k}_i is oriented at angle θ_i w.r.t the interface normal, the reflected ray will propagate at the same angle θ_i . While the transmitted ray will propagate at angle θ_t which is determined by the refractive index of medium 1 (η_i) and refractive index of medium 2 (η_t), which is known as **Snell's law** [15]:

$$\frac{\sin \theta_i}{\sin \theta_t} = \frac{\eta_t}{\eta_i}. \quad (2.19)$$

As shown in Figure 2.6, an optical ray constitutes an electric field \vec{E} and magnetic field \vec{B} due to electromagnetic property. The oscillation directions of \vec{B}, \vec{E} and propagation direction \vec{k} can be written as

$$\begin{aligned} \vec{E} \perp \vec{B} \perp \vec{k} \\ \vec{E} \times \vec{B} = \vec{k} \end{aligned}$$

The amplitude between electronic field and magnetic field has the relationship as follows

$$E = \frac{c\mu}{\eta} B, \quad (2.20)$$

where refractive index η and magnetic permeability μ depends on the property of the medium. c represents the speed of light [59] in a vacuum. We can also write down the boundary condition of the s-polarised ray in Figure 2.6 as

$$\begin{aligned} E_i + E_r &= E_t \\ B_i \cos \theta_i - B_r \cos \theta_i &= B_t \cos \theta_t \end{aligned} \quad (2.21)$$

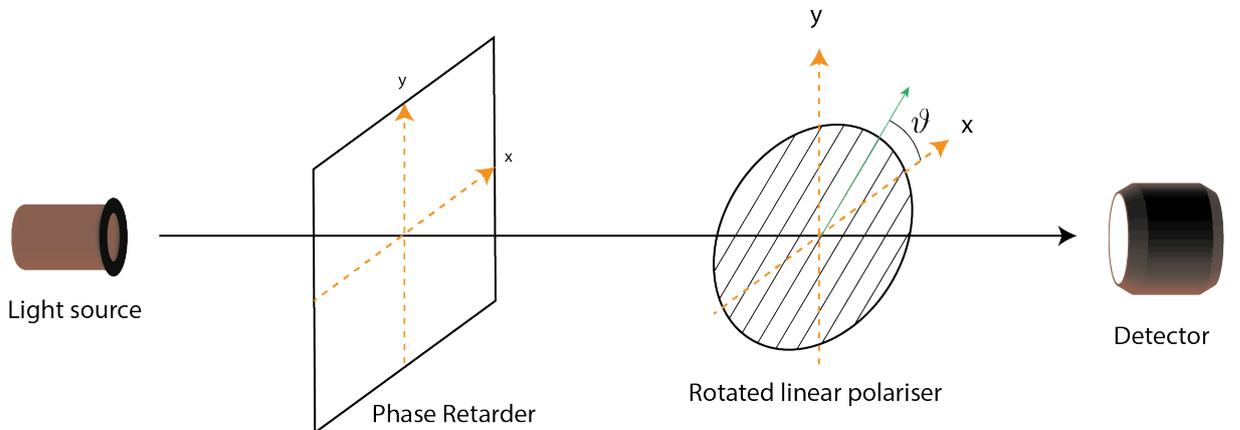


Figure 2.4: [18] Classical method to measure the Stokes parameters, imagine the E_x, E_y of optical ray align with x, y axis respectively.

We substitute equation (2.20) into equation (2.21) to obtain

$$\begin{aligned}
 E_i \frac{\eta_i}{c\mu_i} \cos \theta_i - E_r \frac{\eta_i}{c\mu_i} \cos \theta_i &= E_t \frac{\eta_t}{c\mu_t} \cos \theta_t \\
 \Rightarrow E_i \cos \theta_i - E_r \cos \theta_i &= \frac{\eta_t \mu_i}{\eta_i \mu_t} \cos \theta_t (E_i + E_r) \\
 \Rightarrow \frac{E_r}{E_i} &= \frac{\cos \theta_i - \frac{\eta_t \mu_i}{\eta_i \mu_t} \cos \theta_t}{\cos \theta_i + \frac{\eta_t \mu_i}{\eta_i \mu_t} \cos \theta_t}
 \end{aligned}$$

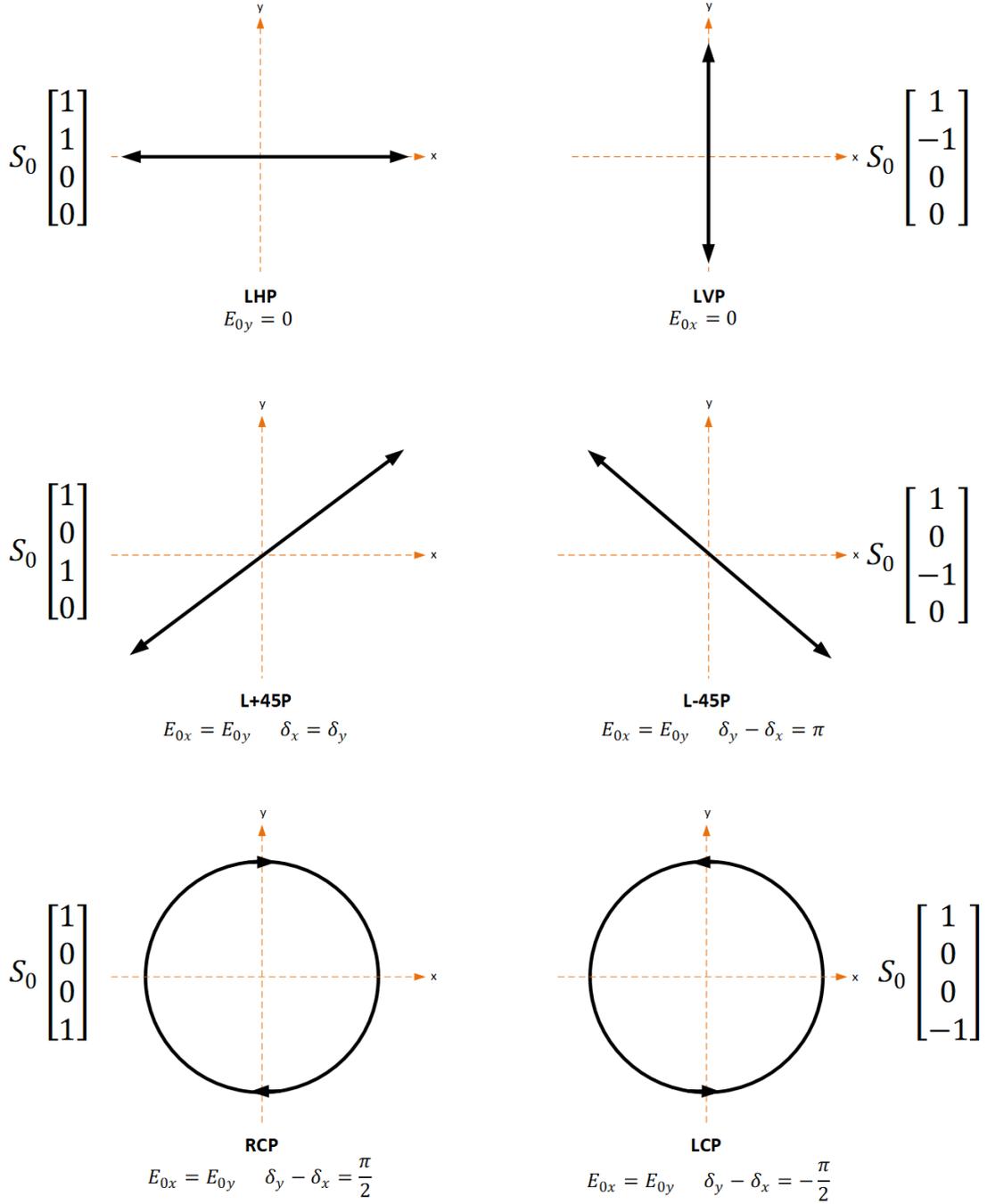


Figure 2.5: The degenerate polarisation states and their corresponding Stokes vectors. The linearly polarised lights are degenerates case of the ellipse equation.

This gives the amplitude ratio of *s-polarised* optical rays between reflected ray and incident ray. In this thesis we only consider non-magnetic medium so that $\mu_i = \mu_t$ approximately, which leads to

$$\frac{E_r}{E_i} = \frac{\cos \theta_i - \frac{\eta_t}{\eta_i} \cos \theta_t}{\cos \theta_i + \frac{\eta_t}{\eta_i} \cos \theta_t}.$$

An amplitude ratio between refractive/incident ray $\frac{E_r}{E_i}$ can be calculated by following the same process, substituting E_t instead of E_r . By substituting a variational form of equation (2.19) that $\cos \theta_t = \sqrt{1 - (\frac{\eta_t}{\eta_i})^2 \sin^2 \theta_i}$, we arrive at the **Fresnel equations** of *s-polarised* ray

$$r_s(\theta_i, \eta) = \frac{E_r}{E_i} = \frac{\cos \theta_i - \sqrt{\eta^2 - \sin^2 \theta_i}}{\cos \theta_i + \sqrt{\eta^2 - \sin^2 \theta_i}} \quad (2.22)$$

$$t_s(\theta_i, \eta) = \frac{E_t}{E_i} = \frac{2 \cos \theta_i}{\cos \theta_i + \sqrt{\eta^2 - \sin^2 \theta_i}}, \quad (2.23)$$

where $\eta = \frac{\eta_t}{\eta_i}$, similarly the *Fresnel equation* of *p-polarised* ray is

$$r_p(\theta_i, \eta) = -\frac{\eta^2 \cos \theta_i - \sqrt{\eta^2 - \sin^2 \theta_i}}{\eta^2 \cos \theta_i + \sqrt{\eta^2 - \sin^2 \theta_i}} \quad (2.24)$$

$$t_p(\theta_i, \eta) = \frac{2\eta \cos \theta_i}{\eta^2 \cos \theta_i + \sqrt{\eta^2 - \sin^2 \theta_i}}. \quad (2.25)$$

Polarisation by reflection and transmission The polarisation state change between incident light and reflected light can be modelled by a Muller matrix as described in (2.12), denoted by $\mathbf{s}_r = \mathbf{M}_R(\theta_i, \eta) \cdot \mathbf{s}_i$ where

$$\mathbf{M}_R(\theta_i, \eta) = \frac{1}{2} \begin{bmatrix} R_s + R_p & R_s - R_p & 0 & 0 \\ R_s - R_p & R_s + R_p & 0 & 0 \\ 0 & 0 & 2\sqrt{R_s R_p} & 0 \\ 0 & 0 & 0 & 2\sqrt{R_s R_p} \end{bmatrix}. \quad (2.26)$$

$$\text{where } R_s = r_s^2, \quad R_p = r_p^2$$

The polarisation state change between incident light and refracted light can be modelled by $\mathbf{s}_t = \mathbf{M}_T(\theta_i, \eta) \cdot \mathbf{s}_i$ where

$$\mathbf{M}_T(\theta_i, \eta) = \frac{1}{2} \begin{bmatrix} T_s + T_p & T_s - T_p & 0 & 0 \\ T_s - T_p & T_s + T_p & 0 & 0 \\ 0 & 0 & 2\sqrt{T_s T_p} & 0 \\ 0 & 0 & 0 & 2\sqrt{T_s T_p} \end{bmatrix}. \quad (2.27)$$

$$\text{where } T_s = t_s^2, \quad T_p = t_p^2$$

2.1.3 Assumptions

We now have all the building blocks to construct the polarisation reflectance model. We made these key assumptions on shape-from-polarisation method through this thesis.

- The incident light is unpolarised light.
- We will not use the circular component when we measure the received optical ray.
- We only consider dielectric material objects.

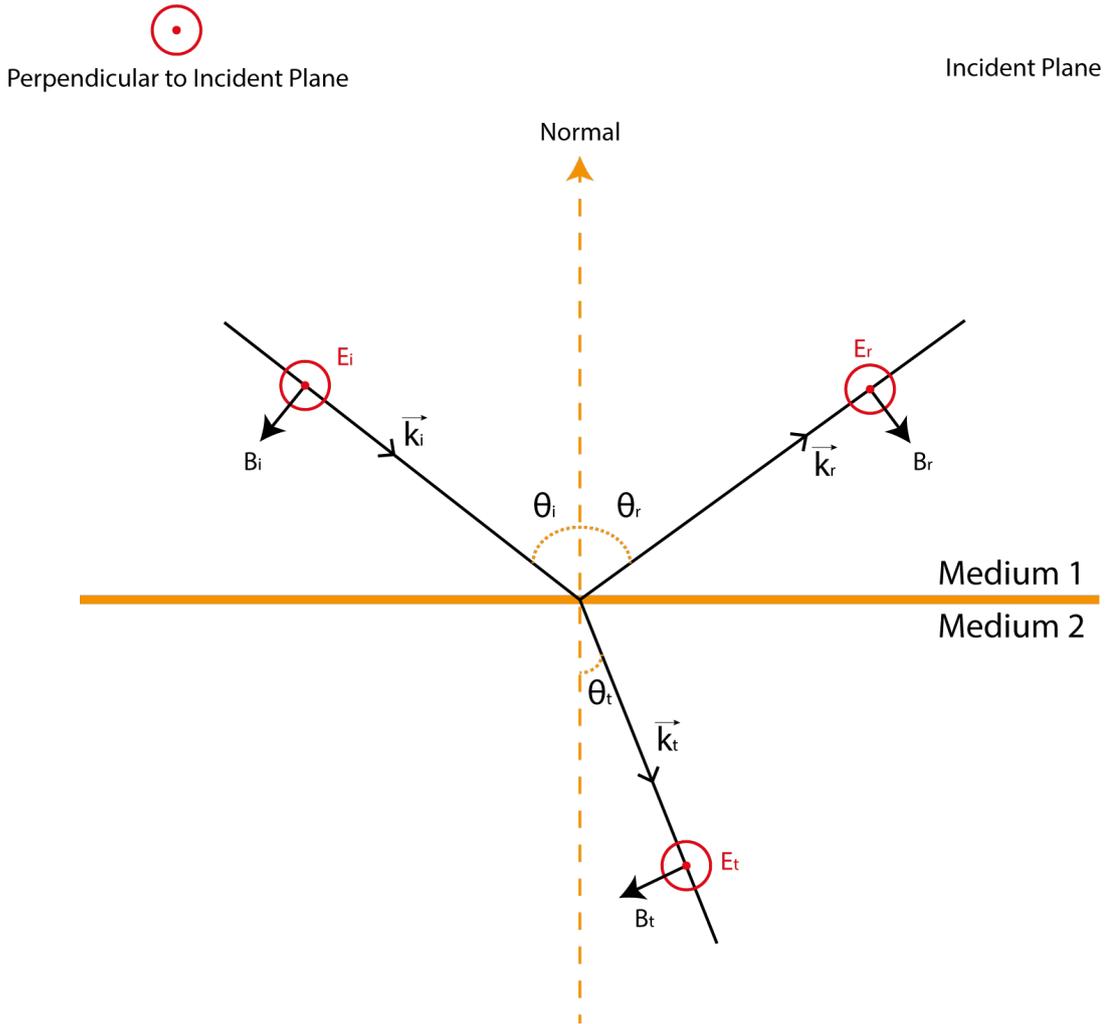


Figure 2.6: An optical ray with direction \vec{k}_i propagates from medium 1 to medium 2, which produces a reflected ray \vec{k}_r and a transmitted ray \vec{k}_t at the interface between medium 1 and medium 2. The reflected angle $\theta_r = \theta_i$ and the transmitted angle θ_t is determined by the Snell's law. E, B present the amplitudes of electronic field and magnetic field of optical ray respectively. And direction of E is perpendicular to the incident plane while direction of B is coplanar with incident plane.

- The object surface is smooth and we do not consider the rough surface which might cause depolarisation effects.
- We do not consider subsurface reflection and refraction.

2.1.4 Specular reflection polarisation model

We first derive the specular reflection component of our polarisation reflectance model. Figure 2.7 shows a typical setup to capture polarimetric intensity images. The polarisation state of the optical ray emitted from light source until to camera follows the path: 1) Unpolarised when emitted from light source. 2) Partially polarised when reflected from object surface. 3) Linearly polarised after passing through a rotated linear polariser filter. This change of polarisation state can be compactly written in Muller matrix with equations (2.13), (2.26) and (2.12) as

$$\mathbf{s}' = \mathbf{M}_{\text{RPOL}}(\vartheta - \phi_i + \frac{\pi}{2})\mathbf{M}_{\text{R}}(\theta_i, \eta)\mathbf{s}, \quad (2.28)$$

where the the unit surface normal is written in spherical coordinates such that the zenith angle is equal to the incident ray angle θ_i , and the azimuth angle is ϕ_i . The rotation angle of the polariser *w.r.t* x axis is denoted by ϑ , and the orientation between polariser *w.r.t* the reflected ray now is $\vartheta - \phi_i + \frac{\pi}{2}$ as shown in Figure 2.8.

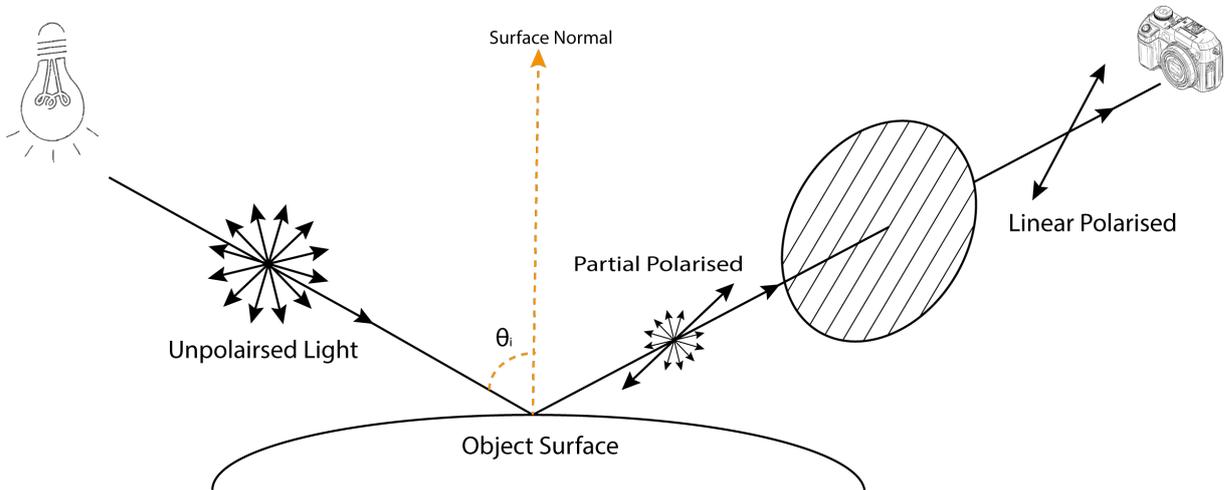


Figure 2.7: An experiment setup to capture polarimetric images. Compare to Figure 2.4, we remove the phase retarder so that only linear polarisation information will be captured.

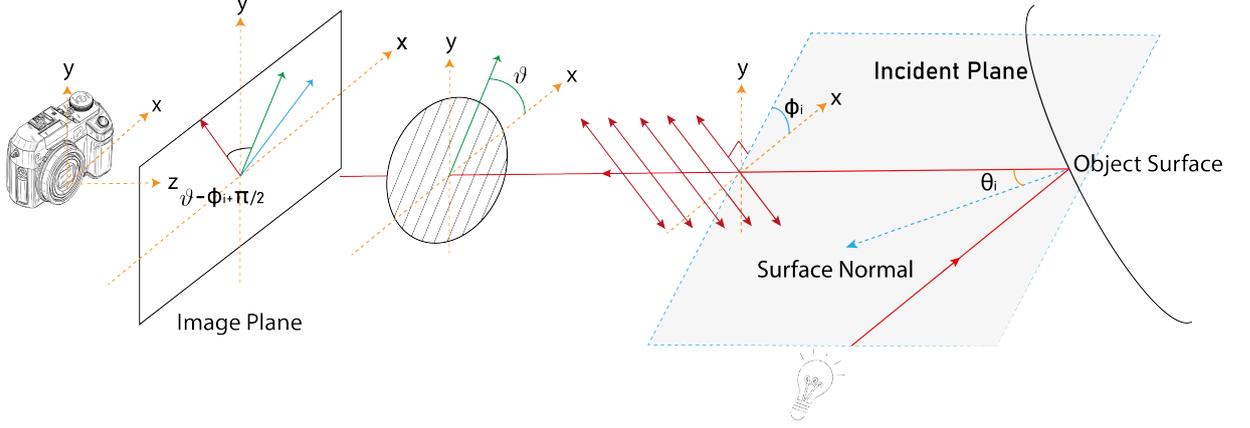


Figure 2.8: The incident ray is coplanar with surface normal, while the most energy of reflected ray is *s-polarised* which is perpendicular to the incident plane. The surface normal is oriented at ϕ_i when mapped to the image plane, so the angle between polariser and reflected ray is $\vartheta - \phi_i + \frac{\pi}{2}$.

If the incident ray is unpolarised, the Stokes vector can be written as $\mathbf{s} = [s_0 \ 0 \ 0 \ 0]^T$. The Stokes vector of reflected ray can be written as follows by substituting (2.26) in

$$\mathbf{s}_r = \mathbf{M}_R \cdot \begin{bmatrix} s_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} s_0(R_p + R_s) \\ s_0(R_s - R_p) \\ 0 \\ 0 \end{bmatrix}. \quad (2.29)$$

In order to measure the Stokes parameters of the reflected ray \mathbf{s}_r , we adapt (2.16) without phase retarder which leads to

$$I(\vartheta) = \frac{1}{4}(s_0(R_p + R_s) + s_0(R_s - R_p) \cos(2\vartheta - 2\phi_i + \pi)). \quad (2.30)$$

Note that the first component of (2.29) is the intensity of the reflected ray by (2.10), and the degree of polarisation is $\rho = \frac{R_s - R_p}{R_s + R_p}$ by (2.11). We substitute equations (2.22) and (2.24) into degree of polarisation such that

$$\rho(\theta_i, \eta) = \frac{2 \sin^2(\theta_i) \cos(\theta_i) \sqrt{\eta^2 - \sin^2(\theta_i)}}{\eta^2 - \sin^2(\theta_i) - \eta^2 \sin^2(\theta_i) + 2 \sin^4(\theta_i)}. \quad (2.31)$$

This is known as the **DoP** of the reflection polarisation reflectance model and we denote with ρ_s . The $\frac{1}{4}s_0(R_p + R_s)$ will be denoted as I_s , it represents the unpolarised intensity when we actually capture by a conventional camera. We finally arrive at a simplified version for reflection polarisation reflectance model that is used in most of the state of the art methods [6, 69, 70, 77, 87, 89]:

$$I(\vartheta) = I_s(1 + \rho_s \cdot \cos(2\vartheta - 2\phi_i + \pi)). \quad (2.32)$$

2.1.5 Diffuse polarisation model

The diffuse polarisation model is modelled by the light scattering inside the object and refracted back out through the object surface. The diffuse polarisation reflectance model follows a similar path as the specular reflection model, the difference is the incident light is emitted from the inside of the object, and the detector is receiving its refracted light as shown in Figure 2.9. We denote the Stokes vector of incident light from the object as \mathbf{s}_o , the change of polarisation state can be described by combining equations (2.13), (2.27) and (2.12) as

$$\mathbf{s}'_o = \mathbf{M}_{\text{RPOL}}(\vartheta - \phi_i) \mathbf{M}_{\text{T}}(\theta_t, \frac{1}{\eta}) \mathbf{s}_o. \quad (2.33)$$

Since the incident ray comes from inside of an object is assumed to be unpolarised, the Stokes vector can be written as $\mathbf{s}_o = [s_{0_o} \ 0 \ 0 \ 0]^T$. The Stokes vector of the refracted ray can be written as follows by substituting (2.27) in

$$\mathbf{s}_t = \mathbf{M}_{\text{T}} \cdot \begin{bmatrix} s_{0_o} \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} s_{0_o}(T_p + T_s) \\ s_{0_o}(T_s - T_p) \\ 0 \\ 0 \end{bmatrix}. \quad (2.34)$$

Notice that the incident ray angle is θ_t and refractive index is the reciprocal of η since the ray is emitted from the medium of object. We can simply convert θ_t to zenith angle θ_i of surface normal by Snell's law (2.19). Compare this with Figure 2.7 in which the incident ray and reflected ray all lay in the medium of air. The orientation between polariser *w.r.t* the refracted ray now is $\vartheta - \phi_i$ as shown in Figure 2.10.

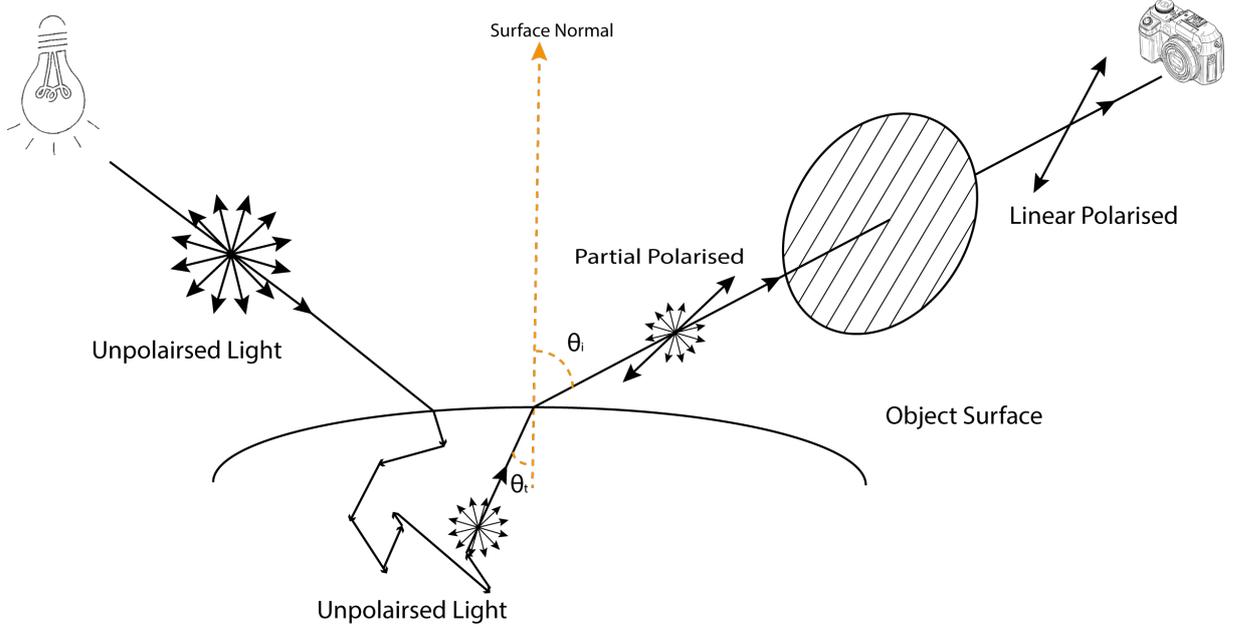


Figure 2.9: The optical ray emitted from a light source will bounce multiple times inside the object which makes the optical ray remain unpolarised. Then the optical ray is transmitted out from the object and received by the detector.

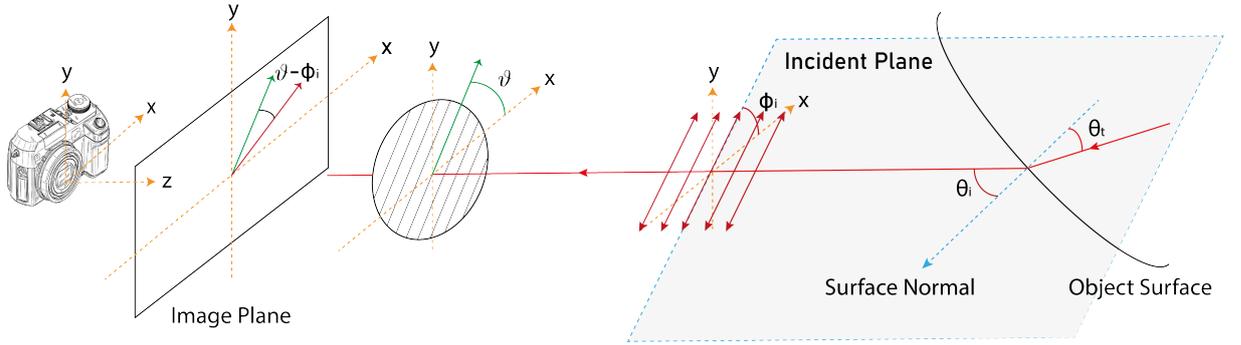


Figure 2.10: The incident ray is coplanar with surface normal and refracted ray, for the most energy of refractive ray is *p-polarised* which is coplanar to the incident plane. The surface normal is oriented at ϕ_i when mapped to the image plane, so the angle between polariser and refracted ray is $\vartheta - \phi_i$.

Combining equations (2.16) (2.27) and (2.33) we get

$$I(\vartheta) = \frac{1}{4}(s_{0o}(T_p + T_s) + s_{0o}(T_s - T_p) \cos(2\vartheta - 2\phi_i)). \quad (2.35)$$

The degree of polarisation is $\rho = \frac{T_s - T_p}{T_s + T_p}$ by (2.11). We convert θ_t to θ_i by (2.19) and substitute

$\frac{1}{\eta}$ with equations (2.23) and (2.25) into degree of polarisation such that

$$\rho(\theta_i, \eta) = \frac{\sin^2(\theta_i)(\eta - \frac{1}{\eta})^2}{4 \cos(\theta_i) \sqrt{\eta^2 - \sin^2(\theta_i)} - \sin^2(\theta_i)(\eta + \frac{1}{\eta})^2 + 2\eta^2 + 2}. \quad (2.36)$$

This is known as the **DoP** of the diffuse polarisation reflectance model and we denote with ρ_d . The $\frac{1}{4}s_0(T_p + T_s)$ will be denoted as I_d , it represents the unpolarised intensity image when we capture with a conventional camera. We finally arrive at a simplified version for the diffuse polarisation reflectance model:

$$I(\vartheta) = I_d(1 + \rho_d \cdot \cos(2\vartheta - 2\phi_i)). \quad (2.37)$$

2.1.6 Polarisation information from polarimetric image data

The first researcher that used a polarimetric approach to estimate shape can be traced back to 1979 by *Koshikawa* [49] who took advantage of the reflected polarised light to estimate the shape of glossy objects. Later Wolff [80, 82] proposed the Fresnel reflectance model that nicely builds a bridge between physics constraints and the task of surface shape estimation in computer vision. With a linear polariser placed in front of a normal camera (Fig 2.7 and Fig 2.9), and by rotating the polariser, the intensity of each pixel in the captured images follows a sinusoidal function *w.r.t* the polarising filter orientation as predicted by equations (2.37) and (2.32). While in experiments we measure the image intensity $I(\vartheta)$ but it remains unknown which model each pixel belongs to. For we have seen both the equations (2.37) and (2.32) form sinusoidal functions. To simplify we denote the unit surface normal as θ, ϕ instead of θ_i, ϕ_i . And we merge these two equations into one for a general presentation:

$$I_\vartheta(I_{\text{un}}, \rho, \varphi) = I_{\text{un}}(1 + \rho \cdot \cos(2\vartheta - 2\varphi)), \quad (2.38)$$

where

- I_{un} is the mean value of the sinusoid or can be presented as the intensity value captured without a polarising filter. It encapsulates the surface reflectance properties of the object and the illumination of the scene.
- ρ represents the degree of polarisation which measures the proportion of how much initially unpolarised optical light ray becomes linearly polarised after reflection/refraction by the surface.
- ϑ represents the polariser orientation *w.r.t* axis of image plane.

- φ represents the phase angle that is related to the azimuth angle ϕ of the surface normal projected to image plane, but the relationship depends on whether specular polarisation model or diffuse polarisation model is used.

Notice that the experimental setup only uses a linear polariser which means the circular component of the optical ray won't be measured. The process is the same as (2.17) shows. Instead of measuring the Stokes parameters directly, we are now interested in unknowns of I_{un} , ρ and φ in (2.38) where we move the known polariser orientation to subscript, and unknowns to the bracket. We expand (2.38) as

$$I_{\vartheta}(I_{\text{un}}, \rho, \varphi) = I_{\text{un}} + I_{\text{un}} \cdot \rho \cdot \cos(2\vartheta) \cos(2\varphi) + I_{\text{un}} \cdot \rho \cdot \sin(2\vartheta) \sin(2\varphi).$$

While the polariser orientation ϑ and intensity I are measured data, we can factor out the unknowns to one side of above equation into a matrix form:

$$I_{\vartheta}(I_{\text{un}}, \rho, \varphi) = \begin{bmatrix} 1 & \cos(2\vartheta) & \sin(2\vartheta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} I_{\text{un}} \\ I_{\text{un}} \cdot \rho \cdot \cos(2\varphi) \\ I_{\text{un}} \cdot \rho \cdot \sin(2\varphi) \end{bmatrix}. \quad (2.39)$$

Theoretically we can solve the x, y, z by capturing N images where $N \geq 3$ whose polariser is oriented at different angles. By stacking (2.40) into matrix form as (2.18) suggests, we have

$$\begin{bmatrix} I_{\vartheta_1} \\ \vdots \\ I_{\vartheta_N} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \cos(2\vartheta_1) & \sin(2\vartheta_1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(2\vartheta_N) & \sin(2\vartheta_N) \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{N \times 3}} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad (2.40)$$

where I_{ϑ_i} represents the intensity under polariser orientation at angle $\vartheta_i | i \in [1 \dots N]$. Due to the noise and model approximation error, the observed images O_{ϑ} are not equal to the value I_{ϑ} that are predicted by the model. We represent the relationship between I_{ϑ} and O_{ϑ} as

$$O_{\vartheta} = I_{\vartheta} + \epsilon. \quad (2.41)$$

To estimate x, y, z , a robust solution is to minimise the following equation:

$$\min_{x,y,z} \|\mathbf{A} \cdot [x, y, z]^T - [O_{\vartheta_1}, \dots, O_{\vartheta_N}]^T\|^2. \quad (2.42)$$

For example we can solve by taking three images with polariser oriented at $0, \frac{\pi}{4}, \frac{\pi}{2}$ suggested by [18, 82]. Although three images are enough to solve the equation, in practice we would like to capture more than three images for robust estimation. The observed images might not perfectly fit the (2.40), but we can solve by linear least squares [74]. An example is shown in Figure 2.12

where seven images are captured at polariser angles in radians of $[0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi]$ which gradually increase the angle by $\frac{\pi}{6}$ as suggested in [87]. Another choice is to use a polarisation camera that embeds the polarising filters at sensor level developed by Sony [72]. This camera has polariser orientations at four different angles on the sensor. In contrast to a normal camera with a mounted polariser, this camera makes the application based on polarisation information easy to deploy. For example in our scenario, we need to take at least three images with different polariser orientations which takes at least three shots. But with the latest camera developed by Sony, we can access polarisation information in a single shot. An illustration of Sony polarisation camera is shown here:

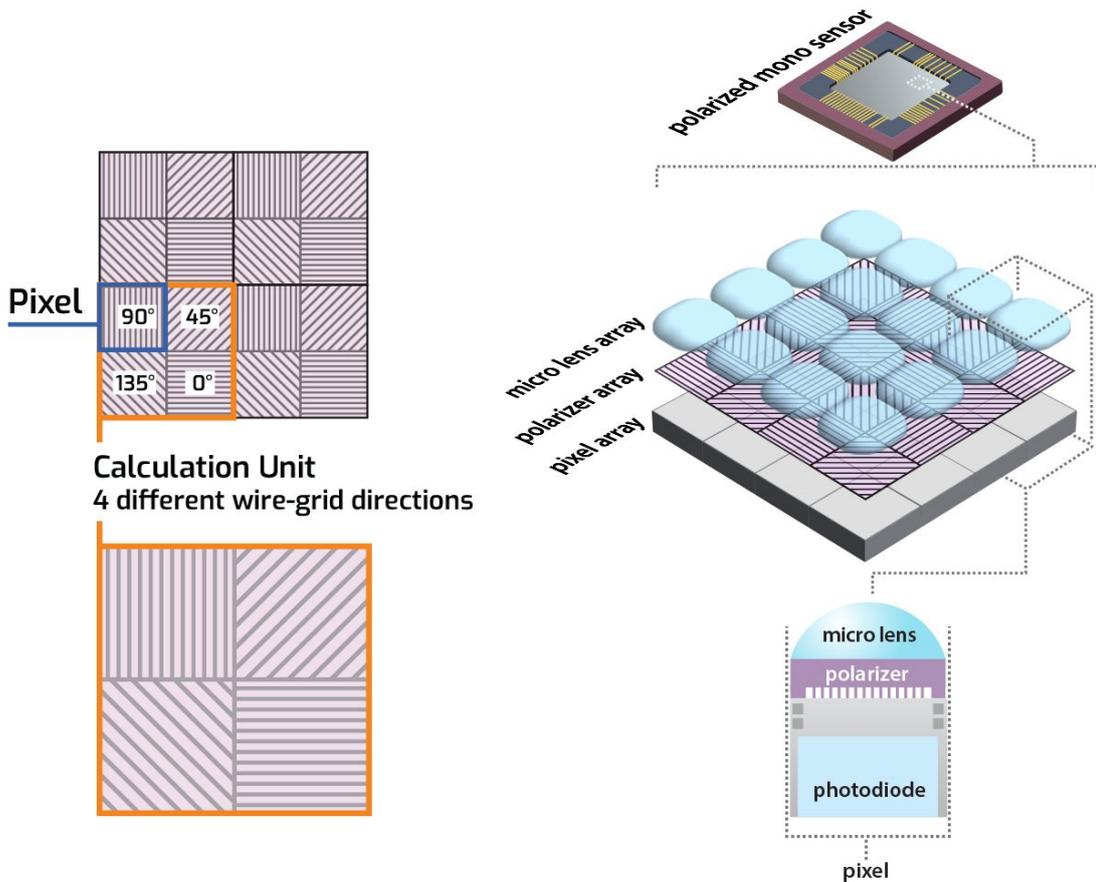


Figure 2.11: Sony's Polarsens 4 Pixel Block Polarizer Design. source: <https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/>

As we see in Figure 2.12, the fitted curve with seven images is closer to the ground truth in contrast to only using the first three images. We finally retrieve the unknowns by $I_{\text{un}} = x$, $\rho = \frac{\sqrt{y^2+z^2}}{x} | \rho \in [0, 1]$ and $\varphi = \frac{1}{2} \tan^{-1} \frac{z}{y}$ as shown in Figure 2.13.

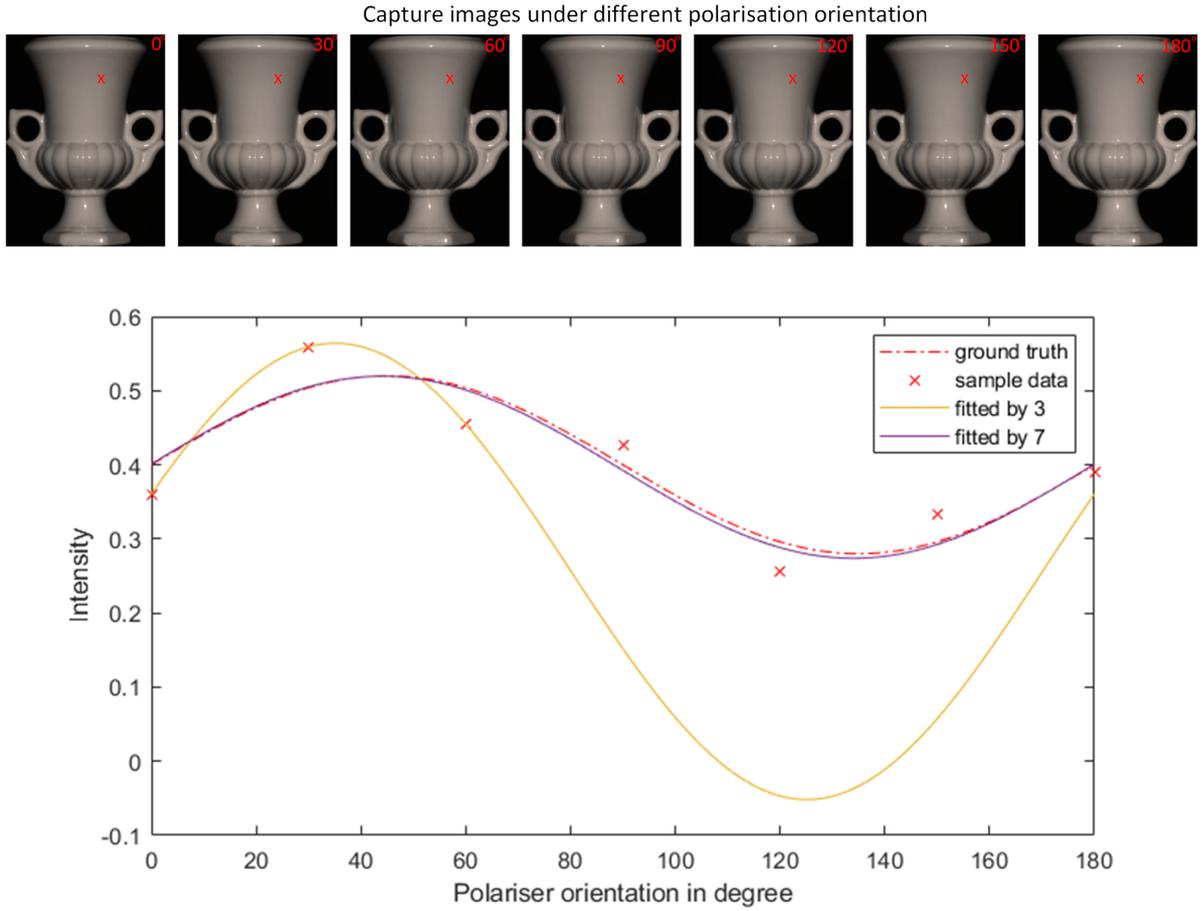


Figure 2.12: The upper row shows input images that are captured by a camera with a mounted linear polariser. The upper right numbers show the degree of angle rotated by the polariser. We plot the samples by red cross and can see they are not a perfect sinusoidal curve due to noise. But we show using more images is more robust to noise.

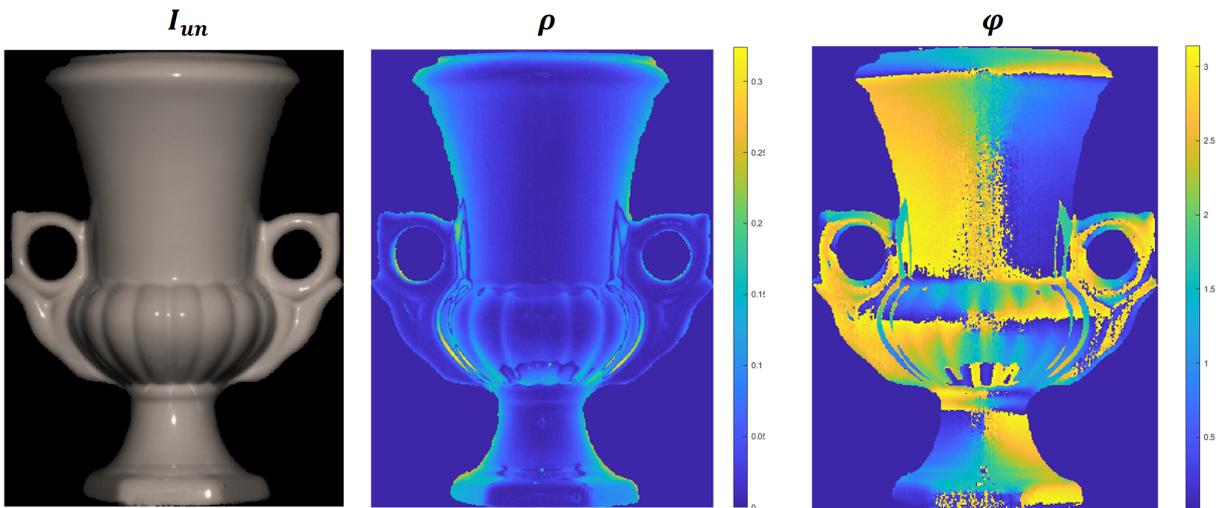


Figure 2.13: The 3 images show I_{un} , ρ , φ respectively solved by (2.40).

We have seen that (2.38) can be interpreted either as specular polarisation reflectance model or diffuse polarisation reflectance model. In this stage, we postulate each pixel can only be either specular dominant or diffuse dominant while in reality it is a superposition between these two and we will model it later. We know that degree of polarisation ρ is dependent on the zenith angle θ of the surface normal and the refractive index η of the material. While the phase angle φ reveals the azimuth angle ϕ of the surface normal with ambiguity shift by π . This is because the linear polariser will allow the incident ray both with azimuth angle ϕ and $\phi + \pi$ to pass through. i.e. in 2.10, the incident plane flipped by π cannot be distinguished from which is not. We have the following relationship for specular polarisation reflectance model

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

$$\phi = \begin{cases} \varphi + \frac{\pi}{2} & \text{or} \\ \varphi - \frac{\pi}{2} \end{cases} \quad (2.43)$$

For the diffuse polarisation reflectance model, we have

$$\rho(\theta, \eta) = \frac{\sin^2(\theta) (\eta - \frac{1}{\eta})^2}{4 \cos(\theta) \sqrt{\eta^2 - \sin^2(\theta)} - \sin^2(\theta) (\eta + \frac{1}{\eta})^2 + 2\eta^2 + 2}$$

$$\phi = \begin{cases} \varphi & \text{or} \\ \varphi - \pi \end{cases} \quad (2.44)$$

We plot the degree of polarisation for each model respectively in 2.14. We constrain θ to lie in $[0, \frac{\pi}{2}]$ otherwise the surface cannot be visible in the image. We are only concerned about dielectric materials in this thesis, whose refractive index lies between $[1.3, 1.6]$ [6]. By observing the curve of specular DoP, there are up to two possible answers for θ , while the diffuse DoP yields a unique answer for θ . And each reflectance model has two possible solutions for ϕ . The total possibility of the surface normal is therefore up to six, where four come from the specular reflection model and two from the diffuse reflection model.

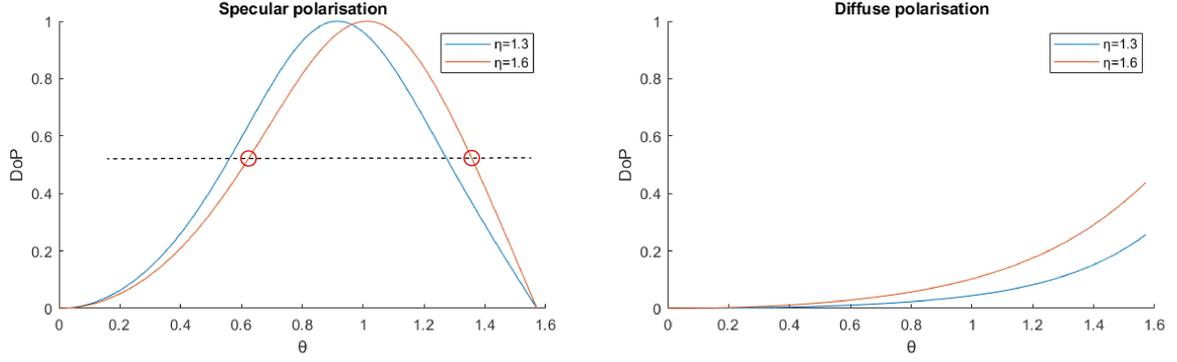


Figure 2.14: A plot shows the function of degree of polarisation w.r.t the zenith angle (θ) with given refractive indices 1.3 and 1.6. (a) Left image presents the specular component, where we can see there are two θ map to the same value of DoP (marked with red circles) except the peak. (b) Right image presents the diffuse component, whose θ uniquely maps to one value of degree of polarisation.

As we discussed previously, in order to recover the shape from polarisation images, we need to address these problems:

- Disambiguate the surface normal from the polarisation model which contains up to six possible solutions. We will get up to four surface normals from the specular polarisation model and up to two from the diffuse polarisation model.
- Retrieve the refractive index η from degree of polarisation.

2.1.7 Multi-view stereo

Multi-view stereo is a classical method to reconstruct the 3D surface of an object. By capturing an object in many different view angles, a dense surface can be reconstructed. The key observation is that some point in the object can be seen by at least two cameras. For simplicity, we assume a point in 3D has been seen by two views. Then the 3D point, its corresponding pixel locations in two images and their camera positions can construct an epipolar plane as shown in Figure 2.15. If we can identify corresponding pixel locations which capture the same point through the images, we can reconstruct its position in 3D (where assume we know the camera parameters) [35].

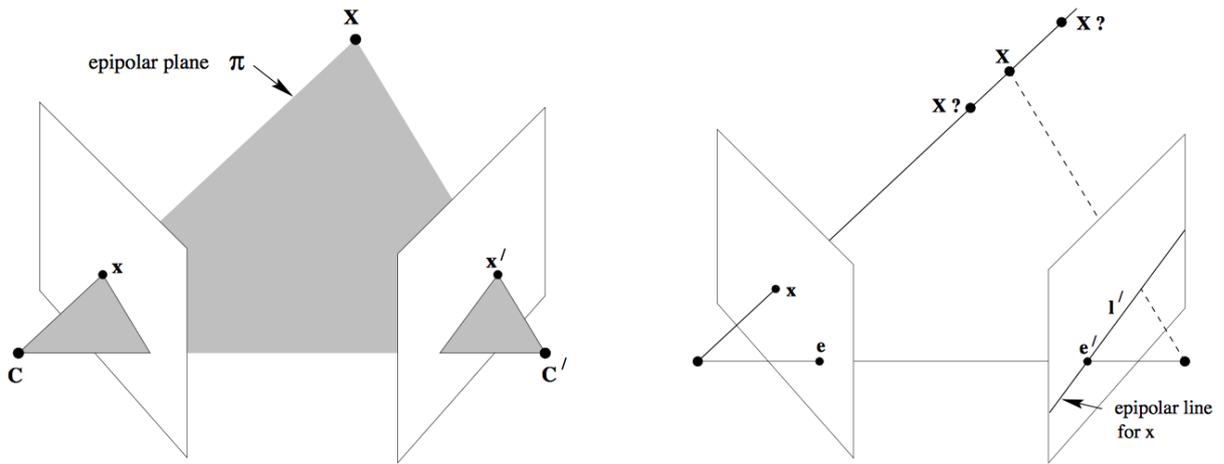


Figure 2.15: [35] **Left image.** The 3D point \mathbf{X} and the corresponding image positions are denoted by x and x' where their camera centres are C and C' . \mathbf{X}, x, x', C and C' lay on the same plane called the epipolar plane. **Right image.** When we do not know the 3D position of point \mathbf{X} and only know its pixel location x in left image. The 3D position \mathbf{X} must lie on the ray that connects camera center C and x . The correspondences in the right image lie on this epipolar line l' .

In [28], the author proposed a window based search method called *Winner takes all*. The algorithm can be described as:

- Choose a reference image and a set of neighbour images. i.e. the neighbour images are defined by the distances between the neighbour cameras and the reference camera smaller than some threshold.
- Choose a pixel location in the reference image which corresponding 3D location can be any depth in a ray that emit from camera centre and connected with the pixel location. Iterate over possible depth values within a given a depth range, and back project the corresponding 3D location to the epipolar line on their neighbor images. The algorithm will try to find out a depth value where the intensity of reference pixel and correspondences are closest, where the intensity similarity is measured by normalised cross correlation. An example can be viewed in Figure 2.16.
- Merge the depth maps and produce a mesh.

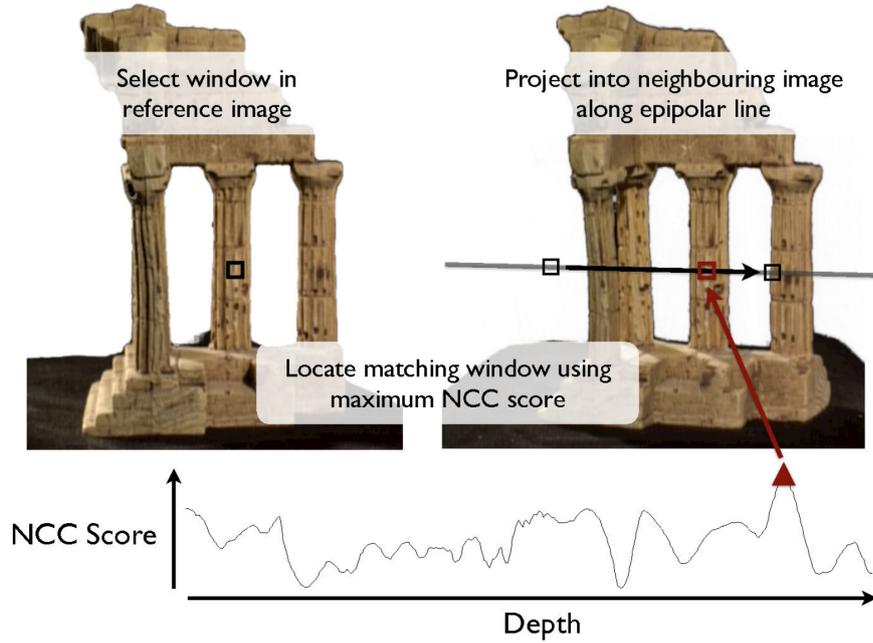


Figure 2.16: [24] "Winner-takes-all" procedure, pick the depth value which has the highest normalized cross correlation (NCC) Score.

A more sophisticated way to get an optimum depth is to use a graphical model. A graphical model uses a graph-based representation [46] as its foundation. It is a powerful model which has been widely and successfully applied in a range fields: robot localisation and mapping [73], speech recognition [26] and neural networks [56] etc. In computer vision, the graphical representation provides an intuitive and compact data structure on images. i.e. the nodes represent image pixels and the edges represent the connection between pixels. The goal of depth reconstruction is to determine a depth value in each pixel, this can be thought as each observed pixel has a hidden state depth [23] that needs to be inferred. In [47], it proposed a Markov Random Field (MRF) depth formulation which can be seen as an optimisation problem. The input depth range is discretised into a finite set of depth values and the problem is to assign a proper depth value z_p from the set of values to each pixel p that minimises the following cost function

$$E(z_p) = \sum_p \Phi(z_p) + \sum_{(p,q) \in N} \Psi(z_p, z_q). \quad (2.45)$$

The term Φ is called a unary potential, it's an energy cost to measure the photo-consistency through all images by a given depth, which means an optimum depth projected into image space should give a similar intensity. The term Ψ is called a pairwise potential that usually enforces spatial regularisation which encourages depths at neighbouring pixels to have similar values [22,39,78,83]. Although the minimisation problem is an NP-hard problem, the existing algorithm

polarisation normals via propagation from the boundary under an assumption of global convexity. Smith et al. [69, 70] utilised the shading cue and phase angle constraint to estimate shape. The shading cue can be presented by the Blinn-Phong reflection model [13], where diffuse pixel is Lambertian model which can be presented as

$$I = \alpha \mathbf{l} \cdot \mathbf{N}, \quad (2.46)$$

where the α represents the albedo of the object, \mathbf{l} presents the light direction and \mathbf{N} represents the surface normal. The specular pixels are represented by half vectors. A precomputed mask is applied to separate specular dominant or diffuse dominant pixels. The phase angle constraint is the projection of the surface normal on the x-y plane which should be collinear with the phase angle, presented as

$$\begin{bmatrix} N_x \\ N_y \end{bmatrix} \times \begin{bmatrix} \cos \phi \\ \sin \phi \end{bmatrix} = 0.$$

Another presentation for surface normal is to use the gradient of depth. Then linear equation about depth can be represented as follows

$$\begin{bmatrix} \sin \phi & -\cos \phi \\ -l_x & -l_y \end{bmatrix} \cdot \begin{bmatrix} D_x \\ D_y \end{bmatrix} \cdot Z = \begin{bmatrix} 0 \\ \frac{I_{un}}{\cos \theta_i} - l_z \end{bmatrix},$$

where depth Z is the depth vector we want to solve with length equal to the number of pixels. $[l_x, l_y, l_z]$ presents the light source direction, D_x, D_y are matrices that approximate the x, y gradients from height by finite difference. When the light source is unknown, it can be estimated by alternative optimisation described in the paper. The final estimated depth will either end up correct or subject to a convex/concave ambiguity. If the light source is known the depth can be determined uniquely.

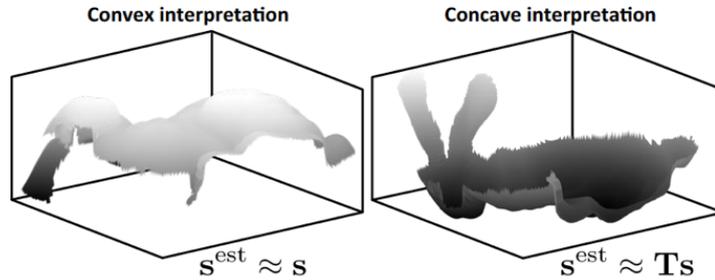


Figure 2.18: [70] Two ambiguity depth with estimated light source

We have seen the equations (2.43) and (2.44) are dependent on refractive index. In [40], the authors proposed a method that recovered shape and refractive index simultaneously. They

capture a sequence of multispectral images which depend on wavelength as well. By introducing Cauchy’s dispersion equation [15], the refractive index can be represented as a function of wavelength as follows

$$\eta(\lambda) = \sum_{k=1}^M C_k \lambda^{-2(k-1)}. \quad (2.47)$$

Each image is captured under certain wavelength optics. i.e. the reflected/refracted rays from the object are different wavelengths that can generate a multispectral image. As shown in (2.20), the Fresnel equation is dependent on wavelength as well. Combined with the above equation, a non-linear cost function can be built which depends on zenith angle θ and coefficients C_k where K is the number of wavelengths. Since digital images captured are in RGB channels, the number of wavelengths can be constrained on RGB only. Then the problem can be solved by trust-region method.

Saman [66] made use of photometric stereo in conjunction with polarisation to estimate refractive index. Photometric stereo was proposed by [84] and is based on Lambertian reflectance model as shown in (2.46). The surface normal contains three unknowns, therefore three images with three different light sources are sufficient to solve the surface normal. In order to get a robust shape estimation, more than three images will be taken to ensure each pixel illuminated at least by three different light sources. Moreover the redundant images make the result more robust to the noise. Equipped with the knowledge of photometric stereo, [66] firstly recovered the normals from photometric stereo. Secondly, computed the zenith angle from the normal. Thirdly, substituted the zenith angle into Equation (2.44) to compute the refractive index.

Similarly, [58] proposed two constraints from photometric stereo and polarisation separately. The photometric stereo constraint took the intensity ratio between two different images with different light sources while the polarisation stereo constraint took the intensity ratio between two different images that under same illumination but with different polariser angles. With these two constraints, a cost function was constructed that tried to find an optimum zenith angle and refractive index pair to minimise it. The cost function was passed to a non-linear solver and refractive index and normal were jointly estimated. Another paper [75] used a non-linear approach to estimate mixed polarisation pixels as well as the refractive index.

2.2.2 Two view Polarisation

Recovering the shape of a transparent object merely by intensity is a challenge. Miyazaki et al. [52] proposed a method to estimate transparent object surfaces by utilising the degree of

polarisation from a polarisation reflection model with knowledge of refractive index. One of the challenges we have seen in Figure 2.14 is that the DoP of the specular model produces two zenith angle values. By observing the derivative of specular DoP shown in Figure 2.19, the author concludes a way to disambiguate the zenith angle θ :

$$\theta \in \begin{cases} 0^\circ < \theta < \theta_B & \rho' > 0 \\ \theta = \theta_B & \rho' = 0 \\ \theta_B < \theta < 90^\circ & \rho' < 0 \end{cases} \quad (2.48)$$

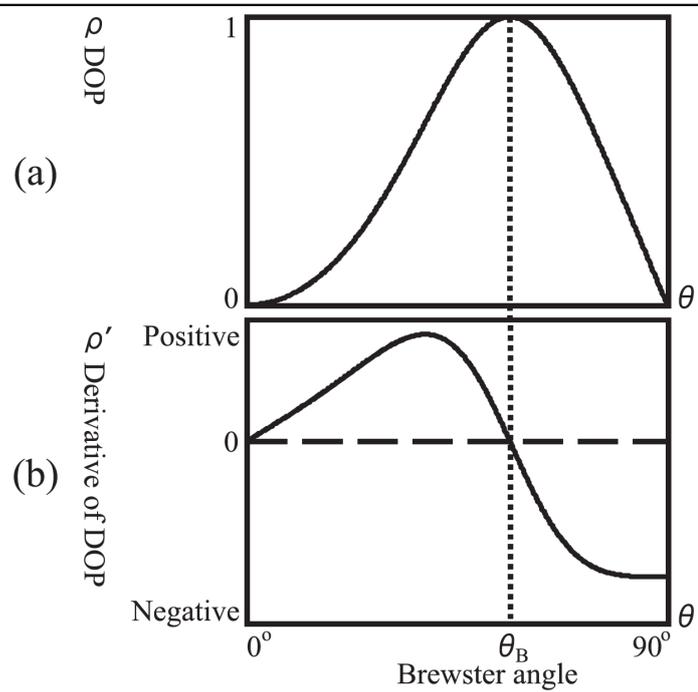


Figure 2.19: [52] The derivative of DoP gives a clue to disambiguate θ .

In order to compute the derivative of DoP, the author build up a two-view environment to capture the same object from two slightly different angles. The DoP of a pixel in the first view is represented by $\rho(\theta)$. And the corresponding pixel in the second view was presented by $\rho(\theta + \Delta\theta)$ who was slightly rotated by angle $\Delta\theta$. So the derivative of polarisation in a certain pixel ρ' can be computed by follows

$$\rho' = \frac{\rho(\theta + \Delta\theta) - \rho(\theta)}{\Delta\theta}.$$

After getting the derivative, the ambiguity of the value θ_i can be identified by the sign of derivative according to (2.48). The phase angle is calculated then disambiguated by propagating from the occluding boundary to the rest of the surface. Where the azimuth angle on the occluding

boundary is assumed pointing vertically to the view direction. A similar boundary propagation method has been used in paper [6], but the authors concentrated on dielectric objects and only used diffuse polarisation reflectance model. By observing that dielectric objects illuminated by a point light are diffuse dominant and their refractive indices normally fall between 1.4 to 1.6, so the zenith angle can be recovered directly from diffuse polarisation component without further ambiguity resolution. From equation (2.44), the azimuth angle on each pixel has two ambiguities, and image with N pixels has 2^N different combinations of azimuth angles. In order to disambiguate it, the paper made assumption that the normals of occluding contours were always pointing away from the surface. By propagating from the occluding contour pixels to their neighbours, the algorithm selected the azimuth angles that helped to preserve the smoothness of the object. After obtaining the normals, a Graph-Spectral [65] method was applied to recover the depth of the object.

In papers [3,5,8], the authors made use of the two-view shading cues and diffuse polarisation to recover the shape of featureless objects. Firstly, taking polarimetric images from two views separately, a unique zenith angle and ambiguous azimuth angles could be obtained. Secondly, segmenting the images to patches which are established for correspondences searching. Thirdly, using patch matching to optimise cost function in order to fully constrain the surface normal by establishing correspondence. Lastly recovering depth from unambiguous normals.

A recent work [42] leveraged the Kinect which is equipped with an RGB-D sensor and combined it with polarimetric information to recover a depth map. (1) The Kinect camera produces a coarse depth map for an object but noisy on surface detail. (2) Polarimetric images will give high resolution in detail per-pixel on normal map but with ambiguity. (3) Using coarse depth map as a guidance to disambiguate the normal map from polarisation. (4) Make use of the spanning tree and depth field constraints to enhance the coarse depth map with the corrected normal map.

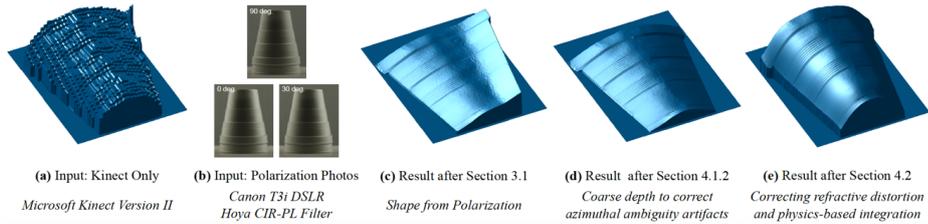


Figure 2.20: [42] (a) The coarse depth map captured by Kinect. (b) Input polarimetric images. (c) Recovered shape from polarisation images. (d) Using coarse depth map to disambiguate result of (c). (e) Enhanced coarse depth map.

2.2.3 Multi-view Polarisation

The limitation of multi-view stereo is it cannot recover featureless regions due to the correspondence search failing in this case. However, polarisation images still has surface orientation information. In paper [19] the authors proposed a polarimetric multi-view stereo that utilises a polarisation model to recover shape in the featureless areas. The depth map estimated from multiveiw stereo gives a strong clue for the surface normal that is utilised as a guidance to disambiguate the normals from polarisation. It observed that the phase angle should be either flipped by π or $\pi/2$ which is unknown. This is a binary label problem which can be modelled by (2.45) and solved. Then the holes in the depth map only contain polarisation information, it will be estimated by propagating the surface normal from known area reconstructed by multi-view stereo. Lastly using the depth map and corrected surface normals jointly to estimate depth map [57]. The paper also mentioned about how to deal with the mixed polarisation pixels: finding out the diffuse dominant pixels and treating the rest of pixels as specular pixels by image intensity.

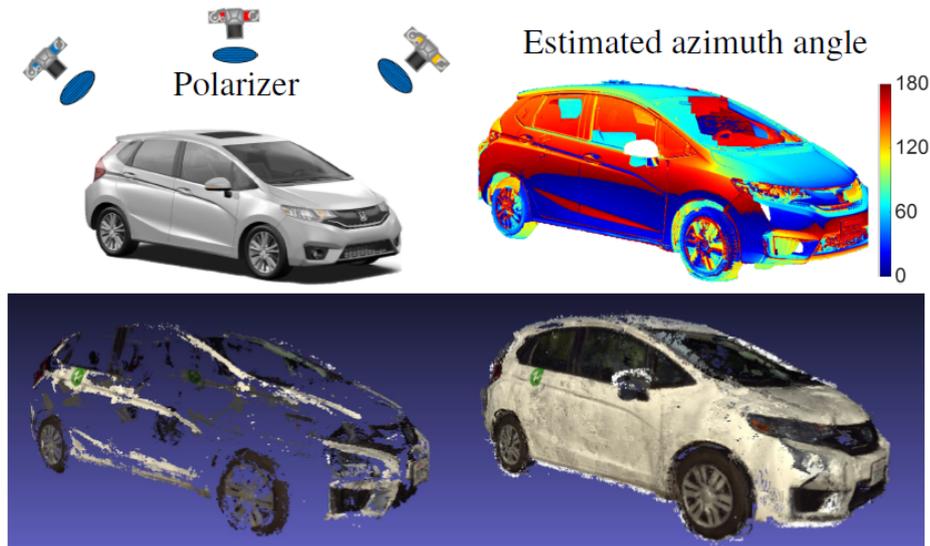


Figure 2.21: a).The top left illustrate the setup of multi-view polarisation. b). The top right shows the estimated azimuth angle of the car. c). The bottom left shows the reconstruction from multi-view stereo. d). The bottom right shows the reconstruction result of multi-view polarisation. [20]

Some of the earliest work on polarisation vision used a stereo pair of polarisation measurements to determine the orientation of a plane [81]. Rahmann and Canterakis [64] combine a specular polarisation model with stereo cues. Similarly, Atkinson and Hancock [8] use 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 [53]. Berger *et al.* [12] use polarisation stereo for depth estimation of specular scenes. Chen *et al.* [17] provide a theoretical treatment of constraints arising from three view polarisation. Yang *et al.* [86] propose a variant of monocular SLAM using polarisation video. Cui *et al.* [21] use polarisation constraints to reduce the number of correspondences required for relative pose estimation. While these methods all require multiple polarisation cameras or a moving camera, we focus on a single viewpoint but with varying illumination direction. The two approaches are likely to be complimentary.

2.2.4 Limitations

We have seen the previous works on shape-from-polarisation methods provides promising results. With only polarisation images, it is possible to reconstruct a dense height map. The ambiguity

was solved by boundary propagation and making an assumption that the reflectance model is diffuse. With shading cues alone, one light source can recover an accurate and dense height map with only concave and convex ambiguity. It provides an easier setup and solution compared to shape from shading or photometric stereo. But these works are limited to uniform albedo object and with prior of known refractive index. The work of polarimetric stereo and multi-view helps to solve the ambiguity problem by using the depth reconstructed from stereo or multi-view as a guidance. But these works did not fully explore the polarisation properties. And all these methods make assumptions that a pixel is either diffuse or specular dominant. To conclude, these methods have one or more of the following limitations:

- The albedo of object is uniform.
- The refractive index of the object is known.
- The shape reconstruction is not metric.
- The pixel is assumed either diffuse or specular dominant.

In this thesis, we take the common assumptions in Section 2.1.3, each method proposed in this thesis are aiming to reconstruct the depth of the object while relaxing some of the limitations above. This can be concluded in following table:

	non-uniform albedo	refractive index	metric reconstruction	Mixed reflectance
Photo-polarimetric stereo	✓			
Non-linear approach	✓	✓		
Polarisation enhanced by second camera	✓		✓	
Mixed polarisation model	✓		✓	✓

Table 2.1: Each proposed method relaxes some of the limitations in previous work. (a) Non-uniform albedo indicates the method is able to apply on the object where the albedo is non-uniform. (b) Refractive index indicates the method do not need the pre-knowledge of the refractive index value of the object, and the proposed method is able to estimate the refractive index value instead. (c) Metric reconstruction indicates the method is able to reconstruct a metrically accurate depth instead of height map up to unknown scale. (d) Mixed reflectance indicates the method is able to deal with the pixel intensity mixed with diffuse and specular reflectance. We do not need to make an assumption that the pixel is either diffuse or specular dominant.

2.3 Conclusion

In the first part of this chapter we reviewed the physics background of the polarisation phenomenon and derived the polarisation reflectance models that are being widely used in the Shape-from-Polarisation methods. In the second part we review some important research works on Shape-from-Polarisation methods. In the next chapters, we start to propose different methods in order to solve the limitations in previous works.

Chapter 3

Underpinning methods for shape-from-polarisation

In this chapter, we propose two underpinning methods that are used throughout the rest of the thesis in our various shape-from-polarisation formulations. We first propose a multichannel estimation method to retrieve unpolarised intensity I_{un} , degree of polarisation ρ and phase angle φ from polarimetric images. We have seen all previous shape-from-polarisation methods begin by estimation of $I_{\text{un}}, \rho, \varphi$. This is usually done by linear least square [70] or nonlinear sinusoid curve-fitting [6]. A major drawback of previous estimation methods [4, 6, 69, 70] is only one colour channel has been taken into account, which does not exploit all possible constraints on ρ, φ . By (2.43) and (2.44), ρ is determined by zenith angle and refractive index, φ is determined by azimuth angle. These are the properties independent on shading. In Section 3.1 we utilise this observation that ρ, φ should be identical through all colour channels in chromatic image (assuming refractive index does not vary with wavelength, because the refractive index changes very small in visible spectrum [45]), a multichannel estimation version is developed. Moreover in Section 3.2, we extend the multichannel estimation to multiple light scenario. For the same object captured under different illumination conditions, ρ, φ should be identical, the only difference is the unpolarised intensity.

In the second part we propose a least square approach to formulate the surface height from gradient. Rather than integrating height from surface normal/gradient, the linear equation *w.r.t* height provides a closed form solution and avoids the integrability problem. This formulation relies on computing the surface gradient from height, hence, numerical surface derivative approximation is critical. In classical approach, only first order accurate (forward or backward

finite difference) are used that make an implicit assumption of surface planarity and are highly susceptible to noise. Occasionally, central difference (second order accurate) [29,62] or smoothed central difference (increased robustness to noise) [57] kernels have been used but the only work to consider kernels accurate to arbitrary order is that of Harker and O’Leary [31–34]. The developed 2D Savitzky-Golay kernel in Section 3.3 provides a higher order and weighted neighbours for surface derivative approximation. Moreover, the Savitzky-Golay kernels can be used as a smoothness regulariser. Unlike planar regularisers, such as a Laplacian filter [69] or zero surface prior [62], we are able to use a high regularisation weight to cope with very significant noise, yet still recover smooth curved surfaces without over flattening. Furthermore, we extend the kernel domain to arbitrary by using K-nearest search to overcome the discontinuity in the domain.

3.1 Multichannel estimation

A polarisation image is usually computed by fitting (2.40) to observed data in a least squares sense. Hence, from more than 3 measurements we estimate I_{un} , ρ and φ . In practice, we may have access to multichannel measurements. For example, we may capture RGB images (3 channels). Since ρ and ϕ depend only on surface geometry (in the case of colour images, the refractive index changes very little in visible spectrum [45], we approximate the value of refractive index is invariant through R, G, B channels), then we expect these quantities to be constant over the channels. On the other hand, I_{un} will vary between channels either because of a shading change caused by the different lighting or because the albedo or light source intensity is different in the different colour channels. Hence, in a multichannel setting with C channels and N polarisation images, we have $C + 2$ unknowns and $C \cdot N$ observations. If we use information across all channels simultaneously, the system is more constrained and the solution will be more robust to noise. Moreover, we do not need to make an arbitrary choice about the channel from which we estimate the polarisation image. This idea shares something in common with that of Narasimhan [55], though their material/shape separation was not in the context of polarisation.

Specifically, we can express the multichannel observations of a single pixel in channel c with polariser angle ϑ_j as

$$I_{\vartheta_j}{}^c = I_{\text{un}}{}^c(1 + \rho \cos(2\vartheta_j - 2\varphi)). \quad (3.1)$$

The system of equations is linear in the unpolarised intensities and, by a change of variables, can be made linear in ρ and φ [40]. Hence, we wish to solve a bilinear system and do so in a least squares sense using interleaved alternating minimisation. Specifically, we **a).** fix ρ and φ and

then solve linearly for the unpolarised intensity in each channel and **b)**, then fix the unpolarised intensities and solve linearly for ρ and φ using all channels simultaneously.

3.1.1 Solve unpolarised intensities

In order solve the unpolarised intensities across channels in a single pixel, we firstly factor out the unknowns from the (3.1) under all N polariser angles. With (2.40) we have:

$$\underbrace{\begin{bmatrix} 1 + \rho \cos(2\vartheta_1 - 2\varphi) \\ \vdots \\ 1 + \rho \cos(2\vartheta_N - 2\varphi) \end{bmatrix}}_{\mathbf{A}_{\rho\varphi} \in \mathbb{R}^N} \cdot \begin{bmatrix} I_{\text{un}}^c \\ \vdots \\ I_{\text{un}}^c \end{bmatrix} = \begin{bmatrix} I_{\vartheta_1}^c \\ \vdots \\ I_{\vartheta_N}^c \end{bmatrix} \quad (3.2)$$

We move the fix term into matrix $\mathbf{A}_{\rho\varphi}$ as denoted in above equation. For we have C channels, then we obtain the unpolarised intensities across channels by solving:

$$\min_{I_{\text{un}}^1, \dots, I_{\text{un}}^C} \left\| \mathbf{C}_A \cdot \begin{bmatrix} I_{\text{un}}^1 & \dots & I_{\text{un}}^C \end{bmatrix}^T - \mathbf{o} \right\|^2, \quad (3.3)$$

where $\mathbf{C}_A \in \mathbb{R}^{CN \times C}$ is given by

$$\mathbf{C}_A = \begin{bmatrix} \mathbf{A}_{\rho\varphi} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{A}_{\rho\varphi} \end{bmatrix}, \quad (3.4)$$

with $\mathbf{A}_{\rho\varphi}$ in diagonal while rest are zero in matrix \mathbf{C}_A . And $\mathbf{o} \in \mathbb{R}^{CN}$ is the observed images that given by

$$\mathbf{o} = \begin{bmatrix} I_{\vartheta_1}^1 \dots I_{\vartheta_N}^1, I_{\vartheta_1}^2 \dots I_{\vartheta_N}^2, \dots, I_{\vartheta_1}^C \dots I_{\vartheta_N}^C \end{bmatrix}^T. \quad (3.5)$$

3.1.2 Solve degree of polarisation and phase angle

We now fix the multichannel unpolarised intensity. In order to solve the degree of polarisation and phase angle, first rearrange the (3.1) as

$$I_{\vartheta_j}^c - I_{\text{un}}^c = I_{\text{un}}^c (\rho \cos(2\vartheta_j - 2\varphi)).$$

we then factor out ρ and φ to achieve:

$$I_{\vartheta_j}^c - I_{\text{un}}^c = \begin{bmatrix} I_{\text{un}}^c \cos(2\vartheta_j) & I_{\text{un}}^c \sin(2\vartheta_j) \end{bmatrix} \cdot \begin{bmatrix} \rho \cos(2\varphi) \\ \rho \sin(2\varphi) \end{bmatrix}$$

With the unpolarised intensities fixed, we solve for ρ and φ using the following linearisation:

$$\min_{a,b} \left\| \mathbf{C}_{\rho\varphi} \begin{bmatrix} a \\ b \end{bmatrix} - \mathbf{d}_{\rho\varphi} \right\|^2, \quad (3.6)$$

where $[a \ b]^T = [\rho \cos(2\varphi), \rho \sin(2\varphi)]^T$, and $\mathbf{C}_{\rho\varphi} \in \mathbb{R}^{CN \times 2}$ is given by

$$\mathbf{C}_{\rho\varphi} = \begin{bmatrix} I_{\text{un}}^1 \cos(2\vartheta_1) & I_{\text{un}}^1 \sin(2\vartheta_1) \\ \vdots & \vdots \\ I_{\text{un}}^1 \cos(2\vartheta_N) & I_{\text{un}}^1 \sin(2\vartheta_N) \\ I_{\text{un}}^2 \cos(2\vartheta_1) & I_{\text{un}}^2 \sin(2\vartheta_1) \\ \vdots & \vdots \\ I_{\text{un}}^C \cos(2\vartheta_N) & I_{\text{un}}^C \sin(2\vartheta_N) \end{bmatrix}, \quad (3.7)$$

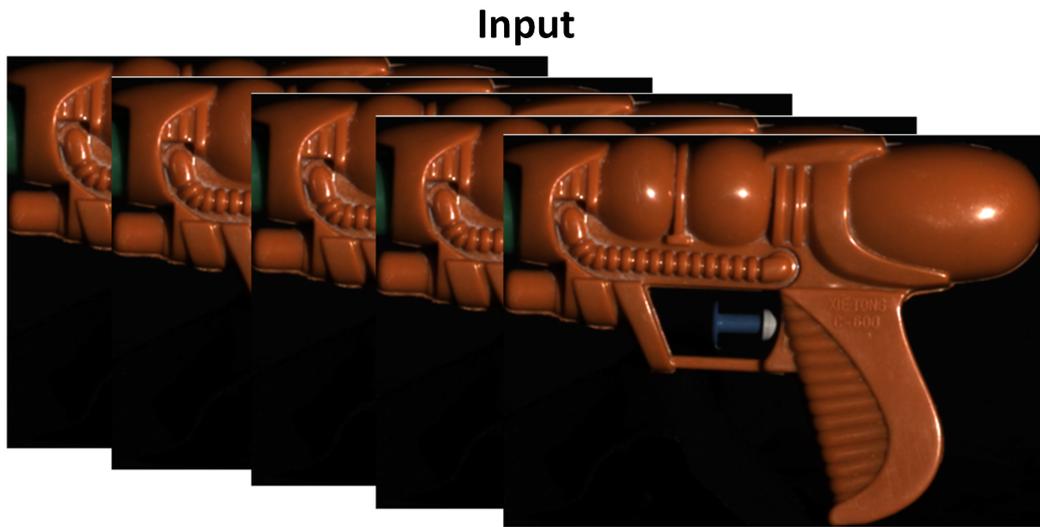
and $\mathbf{d}_{\rho\varphi} \in \mathbb{R}^{CN}$ is given by:

$$\mathbf{d}_{\rho\varphi} = \begin{bmatrix} I_{\vartheta_1}^1 - I_{\text{un}}^1 \\ \vdots \\ I_{\vartheta_N}^1 - I_{\text{un}}^1 \\ I_{\vartheta_1}^2 - I_{\text{un}}^2 \\ \vdots \\ I_{\vartheta_N}^C - I_{\text{un}}^C \end{bmatrix}. \quad (3.8)$$

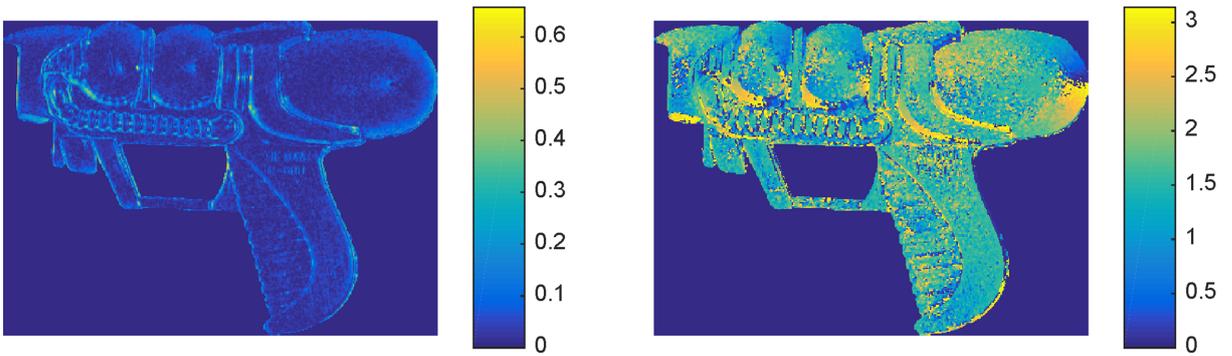
We estimate ρ and φ from the linear parameters using $\varphi = \frac{1}{2} \text{atan2}(b, a)$ and $\rho = \sqrt{a^2 + b^2}$.

3.1.3 Alternating optimisation

We initialise by computing a polarisation image from one channel using linear least squares, as in [40], and then use the estimated ρ and φ to begin alternating interleaved optimisation by solving for the unpolarised intensities across channels. We interleave and alternate the two steps until convergence. In practice, we find that this approach not only dramatically reduces noise in the polarisation images but also removes the ad hoc step of choosing an arbitrary channel to process. We show an example of the results obtained in Figure 3.1. The multichannel result is visibly less noisy than the single channel performance.



Single channel estimation



Multichannel estimation

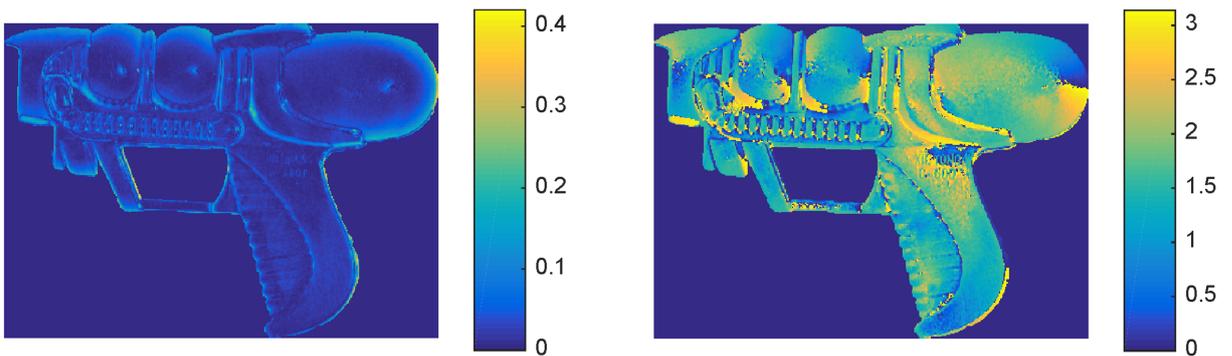


Figure 3.1: Multichannel polarisation image estimation. Top row: the input image sequence; Second row left: degree of polarisation (ρ) and right: phase angle (φ) estimated from a single channel; Third row left: degree of polarisation (ρ) and right: phase angle (φ) estimated from three colour channels and two light source directions.

3.2 Polarisation image estimation with multiple light sources

In the scenario of multiple images, each with a different light source direction, degree of polarisation ρ and phase angle φ depend only on surface geometry, the unpolarised intensity I_{un} might vary due to different light source. **This hold true only while the pixel location through all light sources are either all diffuse dominant or specular dominant.** For example, it cannot be applied when a pixel under first light is diffuse dominant while under second light is specular dominant, for the ρ belongs to different polarisation model as showed in (2.44) and (2.43), phase angle φ will flip by $\frac{\pi}{2}$ as well.

We postulate the pixel through all light sources are under same polarisation model in this method. One way to estimate $I_{\text{un}}, \rho, \varphi$ is to apply multichannel estimation to each light source separately. i.e. estimate $I_{\text{un}1}, \rho_1, \varphi_1$ for first light, and $I_{\text{un}2}, \rho_2, \varphi_2$ for second light which leads to $\rho_1 \neq \rho_2, \varphi_1 \neq \varphi_2$. The drawback of separated estimation is that we might get different value for ρ, φ which theoretically are independent of the light sources. We are aiming to extend the multichannel light estimation in Section 3.1 to multiple light sources while constraining ρ, φ to be the same. We follow the same strategy as multichannel light estimation that **a)**. fix ρ and φ then solve linearly for the unpolarised intensity under different light sources. Assume we have P light sources and the unpolarised intensity under k^{th} light denoted as $I_{\text{un}k}$. **b)**. Constrain the ρ and φ under all light sources then fix the unpolarised intensity $I_{\text{un}1}, \dots, I_{\text{un}P}$ to solve linearly for ρ and φ using all channel simultaneously.

Specifically, we can express the multichannel observations of a single pixel in channel c with polariser angle ϑ_j under light l as

$$I_{\vartheta_j^c} = I_{\text{un}l}^c (1 + \rho \cos(2\vartheta_j - 2\varphi)). \quad (3.9)$$

3.2.1 Solve unpolarised intensity

From (3.2) and (3.3) we have seen how to solve unpolarised intensity under one light source. For multiple light sources, $\mathbf{A}_{\rho\varphi}$ is fixed through all different light sources, to obtain the unpolarised intensities across channels and light sources we can solve:

$$\min_{I_{\text{un}1}^1 \dots I_{\text{un}1}^C, \dots, I_{\text{un}P}^1 \dots I_{\text{un}P}^C} \left\| \mathbf{C}_{AL} \cdot \begin{bmatrix} I_{\text{un}1}^1 & \dots & I_{\text{un}1}^C & \dots & I_{\text{un}P}^1 & \dots & I_{\text{un}P}^C \end{bmatrix}^T - \mathbf{o}_L \right\|^2 \quad (3.10)$$

where $\mathbf{C}_{AL} \in \mathbb{R}^{CNP \times CP}$ is given by

$$\mathbf{C}_{AL} = \begin{bmatrix} \mathbf{C}_A & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{C}_A \end{bmatrix}, \quad (3.11)$$

with \mathbf{C}_A denoted in (3.4) and $\mathbf{o}_L \in \mathbb{R}^{CNP}$ is the observed images that given by

$$\mathbf{o}_L = [\mathbf{o}_1 \dots \mathbf{o}_k \dots \mathbf{o}_P]^T.$$

where \mathbf{o}_k is defined in (3.5) that presents intensities captured under k^{th} light source.

3.2.2 Solve degree of polarisation and phase angle

We now fix the multichannel unpolarised intensity. We can use the same strategy in equation (3.6) to solves ρ and φ . For $\mathbf{C}_{\rho\varphi}$ and $\mathbf{d}_{\rho\varphi}$ contains unpolarised intensity that are light source dependent. We rewrite them as a light source dependent equation as follows

$$\mathbf{C}_{\rho\varphi_l} = \begin{bmatrix} I_{\text{un}l}^1 \cos(2\vartheta_1) & I_{\text{un}l}^1 \sin(2\vartheta_1) \\ \vdots & \vdots \\ I_{\text{un}l}^1 \cos(2\vartheta_N) & I_{\text{un}l}^1 \sin(2\vartheta_N) \\ I_{\text{un}l}^2 \cos(2\vartheta_1) & I_{\text{un}l}^2 \sin(2\vartheta_1) \\ \vdots & \vdots \\ I_{\text{un}l}^C \cos(2\vartheta_N) & I_{\text{un}l}^C \sin(2\vartheta_N) \end{bmatrix}, \quad \mathbf{d}_{\rho\varphi_l} = \begin{bmatrix} I_{\vartheta_1 l}^1 - I_{\text{un}l}^1 \\ \vdots \\ I_{\vartheta_N l}^1 - I_{\text{un}l}^1 \\ I_{\vartheta_1 l}^2 - I_{\text{un}l}^2 \\ \vdots \\ I_{\vartheta_N l}^C - I_{\text{un}l}^C \end{bmatrix}.$$

we solve for ρ and φ using the following linearisation:

$$\min_{a,b} \left\| \mathbf{C}_{\rho\varphi} \begin{bmatrix} a \\ b \end{bmatrix} - \mathbf{D}_{\rho\varphi} \right\|^2, \quad (3.12)$$

where $[a \ b]^T = [\rho \cos(2\varphi), \rho \sin(2\varphi)]^T$, and $\mathbf{C}_{\rho\varphi} \in \mathbb{R}^{CNP \times 2}$ is given by

$$\mathbf{C}_{\rho\varphi} = \left[\mathbf{C}_{\rho\varphi_1} \quad \dots \quad \mathbf{C}_{\rho\varphi_P} \right]^T \quad (3.13)$$

and $\mathbf{D}_{\rho\varphi} \in \mathbb{R}^{CNP}$ is given by

$$\mathbf{D}_{\rho\varphi} = \left[\mathbf{d}_{\rho\varphi_1} \quad \dots \quad \mathbf{d}_{\rho\varphi_P} \right]^T \quad (3.14)$$

We estimate ρ and φ from the linear parameters using $\varphi = \frac{1}{2} \text{atan2}(b, a)$ and $\rho = \sqrt{a^2 + b^2}$. The solving tactic is similar as described in Section 3.1, the initialisation step can be set by computing a polarisation image from one channel under one light source. Figure 3.2 shows an example of using this method under two different light sources.

3.3 Least square reconstruction of surface height

In this section, firstly, we propose a new method for computing numerical derivatives based on 2D Savitzky-Golay filters and K-nearest neighbour kernels. The resulting derivative matrices can be used for least squares surface reconstruction (i.e. (4.40) and (4.19)) over arbitrary (even disconnected) domains in the presence of large noise and allowing for higher order polynomial local surface approximations. Secondly, the Savitzky-Golay filters can be used as a smoothness regulariser as shows in (4.20). Unlike planar regularisers, such as a Laplacian filter [69] or zero surface prior [62], we are able to use a high regularisation weight to cope with very significant noise, yet still recover smooth curved surfaces without over flattening. Thirdly, our least squares surface reconstruction approach is very general, allowing both orthographic and perspective projection (without requiring a nonlinear change of variables [61]), and an optional depth prior. Finally, the proposed method provides an alternate formulation for height-from-normals that uses surface normal components rather than implied surface gradients and is numerically more stable.

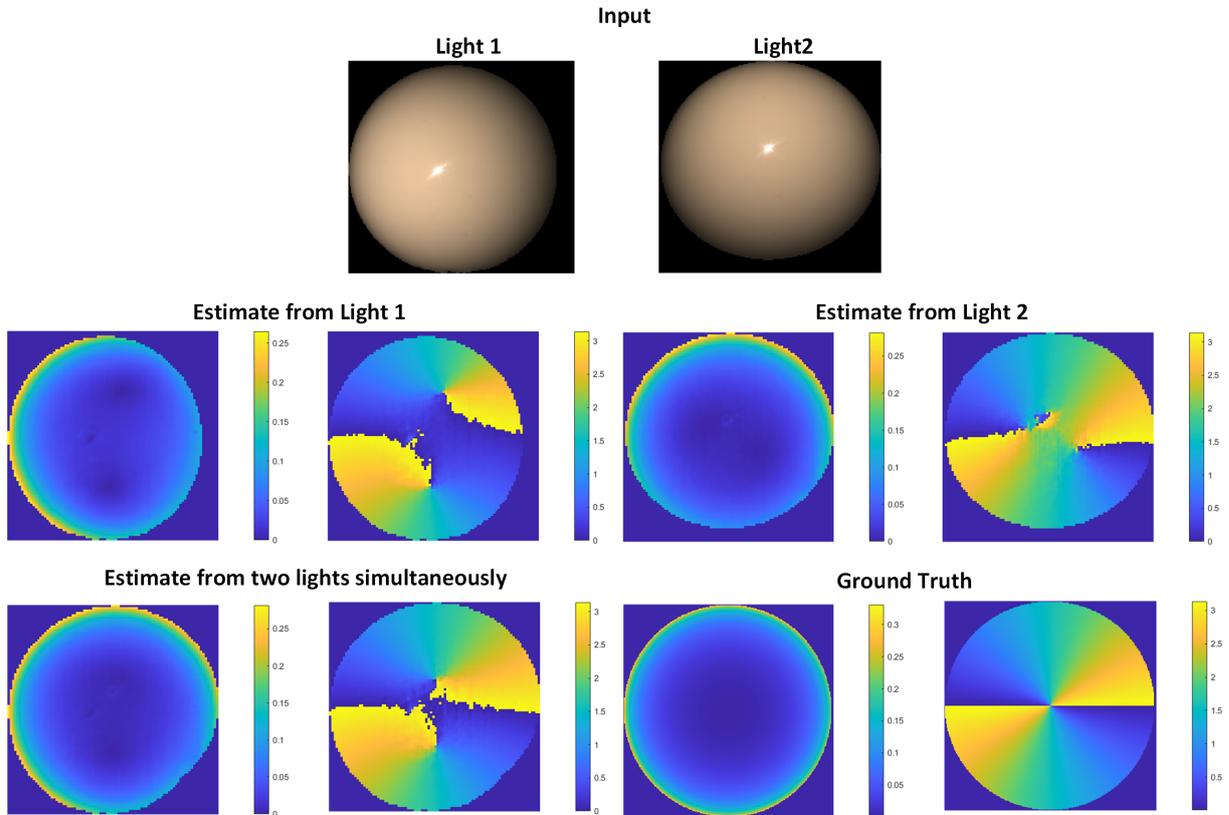


Figure 3.2: An example shows multichannel estimation on snooker ball images.

3.3.1 Linear least squares height-from-normals

We denote an image location in camera units (pixels) as (x, y) such that $\mathbf{u} = (x, y)$ is a pixel location in the image. We parameterise the surface by the height or depth function $z(\mathbf{u})$. In normals-from-depth we are given a noisy observed depth map and wish to estimate the surface normal map $\mathbf{n}(\mathbf{u}) = [n_x(\mathbf{u}), n_y(\mathbf{u}), n_z(\mathbf{u})]^T$ with $\|\mathbf{n}(\mathbf{u})\| = 1$. In surface integration we are given $\mathbf{n}(\mathbf{u})$ and wish to estimate $z(\mathbf{u})$.

To the best of our knowledge, all existing surface integration methods compute height-from-*gradient*, i.e. they transform the given surface normals into estimates of the gradient and solve the following pair of PDEs, usually in a least squares sense:

$$\frac{\partial z(\mathbf{u})}{\partial x} = \frac{-n_x(\mathbf{u})}{n_z(\mathbf{u})}, \quad \frac{\partial z(\mathbf{u})}{\partial y} = \frac{-n_y(\mathbf{u})}{n_z(\mathbf{u})}. \quad (3.15)$$

The problem with this approach is that close to the occluding boundary, n_z gets very small making the gradient very large. The squared errors in these pixels then dominate the least squares solution. We propose an alternative formulation that is more natural, works with both orthographic and perspective projections and, since it uses the components of the normals directly, is best referred to as *height-from-normals*. The idea is that the surface normal should be perpendicular to the tangent vectors. This leads to a pair of PDEs:

$$\frac{\partial \mathbf{p}(\mathbf{u})}{\partial x} \cdot \mathbf{n}(\mathbf{u}) = 0, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial y} \cdot \mathbf{n}(\mathbf{u}) = 0. \quad (3.16)$$

We now consider how to formulate equations of this form in two different cases: orthographic and perspective projection.

3.3.2 Linear equations

Orthographic case The 3D position, $\mathbf{p}(\mathbf{u})$, of the point on the surface that projects to pixel position \mathbf{u} and its derivatives are given by:

$$\mathbf{p}(\mathbf{u}) = \begin{bmatrix} x \\ y \\ z(\mathbf{u}) \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial x} = \begin{bmatrix} 1 \\ 0 \\ \frac{\partial z(\mathbf{u})}{\partial x} \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial y} = \begin{bmatrix} 0 \\ 1 \\ \frac{\partial z(\mathbf{u})}{\partial y} \end{bmatrix}. \quad (3.17)$$

Substituting these derivatives into (3.16) we obtain:

$$\frac{\partial z(\mathbf{u})}{\partial x} n_z(\mathbf{u}) = -n_x(\mathbf{u}), \quad \frac{\partial z(\mathbf{u})}{\partial y} n_z(\mathbf{u}) = -n_y(\mathbf{u}). \quad (3.18)$$

Note that this leads to a simple rearrangement of (3.15) but which avoids division by n_z .

Perspective case In the perspective case, the 3D coordinate corresponding to the surface point at \mathbf{u} and its derivatives are given by:

$$\mathbf{p}(\mathbf{u}) = \begin{bmatrix} \frac{x-c_x}{f_x} z(\mathbf{u}) \\ \frac{y-c_y}{f_y} z(\mathbf{u}) \\ z(\mathbf{u}) \end{bmatrix} \quad (3.19)$$

where f_x, f_y is the focal length of the camera *w.r.t* x, y respectively, and (c_x, c_y) is the principal point. The derivatives are given by:

$$\frac{\partial \mathbf{p}(\mathbf{u})}{\partial x} = \begin{bmatrix} \frac{-1}{f_x} \left((x - c_x) \frac{\partial z(\mathbf{u})}{\partial x} + z(\mathbf{u}) \right) \\ \frac{-1}{f_y} (y - c_y) \frac{\partial z(\mathbf{u})}{\partial x} \\ \frac{\partial z(\mathbf{u})}{\partial x} \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial y} = \begin{bmatrix} \frac{-1}{f_x} (x - c_x) \frac{\partial z(\mathbf{u})}{\partial y} \\ \frac{-1}{f_y} \left((y - c_y) \frac{\partial z(\mathbf{u})}{\partial y} + z(\mathbf{u}) \right) \\ \frac{\partial z(\mathbf{u})}{\partial y} \end{bmatrix} \quad (3.20)$$

Again, these can be substituted into (3.16) to relate the derivatives of z to the surface normal direction.

3.3.3 Discrete formulation

Assume that we are given a foreground mask comprising some subset of the discretised image domain, $\mathcal{F} \subseteq \{1, \dots, W\} \times \{1, \dots, H\}$ with $|\mathcal{F}| = n$. The depth values for the n foreground pixels are stored in a vector $\mathbf{z} \in \mathbb{R}^n$ with arbitrary ordering. We make use of a pair of matrices, $\mathbf{D}_x, \mathbf{D}_y \in \mathbb{R}^{n \times n}$, that compute discrete approximations to the partial derivative in the horizontal and vertical directions respectively. The exact form of these matrices is discussed in the next section. Once these discrete approximations are used, the PDEs in (3.16) become linear systems of equations in \mathbf{z} . This leads to a linear least squares formulation for the height-from-normals problem.

Orthographic case In the orthographic case, we stack equations of the form (3.18):

$$\begin{bmatrix} \text{diag}(\mathbf{n}_z) \mathbf{D}_x \\ \text{diag}(\mathbf{n}_z) \mathbf{D}_y \end{bmatrix} \mathbf{z} = \begin{bmatrix} -\mathbf{n}_x \\ -\mathbf{n}_y \end{bmatrix} \quad (3.21)$$

where

$$\mathbf{n}_x = \begin{bmatrix} n_x(\mathbf{u}_1) \\ \vdots \\ n_x(\mathbf{u}_n) \end{bmatrix}, \quad \mathbf{n}_y = \begin{bmatrix} n_y(\mathbf{u}_1) \\ \vdots \\ n_y(\mathbf{u}_n) \end{bmatrix}, \quad \mathbf{n}_z = \begin{bmatrix} n_z(\mathbf{u}_1) \\ \vdots \\ n_z(\mathbf{u}_n) \end{bmatrix}. \quad (3.22)$$

Note that (3.21) is satisfied by any offset of the true \mathbf{z} , corresponding to the unknown constant of integration. This is reflected in the fact that:

$$\text{rank} \left(\begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} \right) = n - 1. \quad (3.23)$$

So, in the orthographic case, we can only recover \mathbf{z} up to an unknown offset.

Perspective case In the perspective case, we stack equations obtained by substituting (3.20) in (3.16) to obtain:

$$\begin{bmatrix} \mathbf{N}\mathbf{T}_x \\ \mathbf{N}\mathbf{T}_y \end{bmatrix} \mathbf{z} = \mathbf{0}_{2n \times 1}, \quad (3.24)$$

where

$$\mathbf{T}_x = \begin{bmatrix} \frac{-1}{f_x} \mathbf{U} & \frac{-1}{f_x} \mathbf{I} \\ \frac{-1}{f_y} \mathbf{V} & \mathbf{0}_{n \times n} \\ \mathbf{I} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} \mathbf{D}_x \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{T}_y = \begin{bmatrix} \frac{-1}{f_x} \mathbf{U} & \mathbf{0}_{n \times n} \\ \frac{-1}{f_y} \mathbf{V} & \frac{-1}{f_y} \mathbf{I} \\ \mathbf{I} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} \mathbf{D}_y \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} \text{diag}(\mathbf{n}_x) \\ \text{diag}(\mathbf{n}_y) \\ \text{diag}(\mathbf{n}_z) \end{bmatrix}^T, \quad (3.25)$$

$\mathbf{U} = \text{diag}(x_1 - c_x, \dots, x_n - c_x)$ and $\mathbf{V} = \text{diag}(y_1 - c_y, \dots, y_n - c_y)$. Note that (3.24) is a homogeneous linear system. This means that it is also satisfied by any scaling of the true \mathbf{z} . So, in the perspective case, we can only recover \mathbf{z} up to an unknown scaling.

3.3.4 Numerical differentiation kernels

We now consider the precise form of \mathbf{D}_x and \mathbf{D}_y and propose a novel alternative with attractive properties. Since the derivative matrices act linearly on \mathbf{z} they can be viewed as 2D convolutions over $z(x, y)$. Note however that each row of \mathbf{D}_x or \mathbf{D}_y can be different - i.e. different convolution kernels can be used at different spatial locations.

By far the most commonly used numerical differentiation kernels are forward (fw) and backward (bw) difference, shown here for both the horizontal (h) and vertical (v) directions:

$$\mathbf{K}_{\text{fw}}^h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_{\text{fw}}^v = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{K}_{\text{bw}}^h = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_{\text{bw}}^v = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (3.26)$$

As resolution increases and the effective step size decreases, forward and backward differences tend towards the exact derivatives. However, for finite step size they are only exact for order one (planar) surfaces and highly sensitive to noise. Averaging forward and backward yields the central difference (c) approximation, used for example by Quéau *et al.* [62]:

$$\mathbf{K}_c^h = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_c^v = \frac{1}{2} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (3.27)$$

This is order two accurate but still only uses two pixels per derivative and so is sensitive to noise. One way to address this is to first smooth the z values with a smoothing kernel \mathbf{S} and then compute a finite difference approximation. By associativity of the convolution operator we can combine the smoothing and finite difference kernels into a single kernel. For example, the smoothed central difference (sc) approximation, as used by Nehab *et al.* [57] is given by:

$$\mathbf{K}_{\text{sc}}^{\text{h}} = \mathbf{K}_{\text{c}}^{\text{h}} * \mathbf{S} = \frac{1}{12} \begin{bmatrix} -1 & 0 & 1 \\ -4 & 0 & 4 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_{\text{sc}}^{\text{v}} = \mathbf{K}_{\text{c}}^{\text{v}} * \mathbf{S} = \frac{1}{12} \begin{bmatrix} -1 & -4 & -1 \\ 0 & 0 & 0 \\ 1 & 4 & 1 \end{bmatrix}, \quad (3.28)$$

where in this case \mathbf{S} is a rounded approximation of a 3×3 Gaussian filter with standard deviation 0.6. A problem with both smoothed and unsmoothed central difference is that the derivatives and therefore the linear equations for a given pixel do not depend on the height of that pixel. This lack of dependence between adjacent pixels causes a severe “checkerboard” effect that necessitates the use of an additional regulariser, often smoothness. Commonly, this is the discrete Laplacian [69]. However, a smoothness penalty based on this filter is minimised by a planar surface. So, as the regularisation weight is increased, the surface becomes increasingly flattened until it approaches a plane.

With all of these methods alternative kernels must be used at the boundary of the foreground domain. For example, switching from central to backward differences. This means that the numerical derivatives are not based on a consistent assumption.

3.3.5 2D Savitzky-Golay filters

We now show how to overcome the limitations of the common numerical differentiation and smoothing kernels using 2D Savitzky-Golay filters.

The idea of a Savitzky-Golay filter [67,90] is to approximate a function in a local neighbourhood by a polynomial of chosen order. This polynomial is fitted to the observed (noisy) function values in the local neighbourhood by linear least squares. Although the polynomial may be of arbitrarily high order, the fit residuals are linear in the polynomial coefficients and so a closed form solution can be found. This solution depends only on the relative coordinates of the pixels in the local neighbourhood. So, it can be applied (linearly) to any data values meaning that reconstruction with the arbitrary order polynomial can be accomplished with a straightforward (linear) convolution.

The surface around a point (x_0, y_0) is approximated by the order k polynomial $z_{x_0, y_0}(x, y)$:

$\mathbb{R}^2 \mapsto \mathbb{R}$ with coefficients a_{ij} :

$$z_{x_0, y_0}(u, v) = \sum_{i=0}^k \sum_{j=0}^{k-i} a_{ij} (x - x_0)^i (y - y_0)^j. \quad (3.29)$$

Assume we are given a set of pixel locations, $\mathcal{N}_{x_0, y_0} = \{(x_1, y_1), \dots, (x_m, y_m)\}$, forming a neighbourhood around (x_0, y_0) and the corresponding Z values for those pixels. We can form a set of linear equations

$$\begin{bmatrix} 1, & y_1 - y_0, & (y_1 - y_0)^2, & \dots, & (x_1 - x_0)^k \\ & & & \vdots & \\ 1, & y_m - y_0, & (y_m - y_0)^2, & \dots, & (x_m - x_0)^k \end{bmatrix} \mathbf{a} = \mathbf{C}_{\mathcal{N}_{x_0, y_0}} \mathbf{a} = \mathbf{z}_{\mathcal{N}_{x_0, y_0}}, \quad (3.30)$$

where $\mathbf{a} = [a_{00}, a_{01}, a_{02}, \dots, a_{k0}]^T$ and $\mathbf{z}_{\mathcal{N}_{x_0, y_0}} = [z(x_1, y_1), \dots, z(x_m, y_m)]^T$. The least squares solution for \mathbf{a} is given by $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+ \mathbf{z}_{\mathcal{N}_{x_0, y_0}}$ where $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+$ is the pseudo inverse of $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}$. Note that $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+$ depends only on the relative coordinates of the pixels chosen to lie in the neighbourhood of the (x_0, y_0) . Also note that $z_{x_0, y_0}(0, 0)$ is given simply by a_{00} which is the convolution between the first row of $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+$ and the z values. This is a smoothed version of $z(x_0, y_0)$ in which the original surface is locally approximated by a best fit, order k polynomial. Similarly, the first derivative of the fitted polynomial in the horizontal direction is given by a_{10} and in the vertical direction by a_{01} , corresponding to two other rows of $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+$. Note that the order k is limited by the size of the neighbourhood. Specifically, we require at least as many pixels as coefficients, i.e. $k \leq m$.

When \mathcal{N}_{x_0, y_0} is a square neighbourhood centred on (x_0, y_0) then the appropriate row of $\mathbf{C}_{\mathcal{N}_{x_0, y_0}}^+$ can be reshaped into a square convolution kernel. Convoluting this with a $z(x, y)$ map with rectangular domain \mathcal{F} amounts to locally fitting a polynomial of order k and either evaluating the polynomial at the central position, acting as a smoothing kernel, or evaluating the derivative of the polynomial in either vertical or horizontal direction.

3.3.6 K-nearest pixels kernel

In general, the foreground domain will not be rectangular. Often, it corresponds to an object mask or semantic segmentation of a scene. In this case, we need a strategy to deal with pixels that do not have the neighbours required to use the square kernel. 2D Savitzky-Golay filters are ideal for this because the method described above for constructing them can be used for arbitrary local neighbourhoods. We propose to use the K-nearest pixels in \mathcal{F} to a given pixel. In practice, we compute the square $d \times d$ kernel once and use this for all pixels where the required

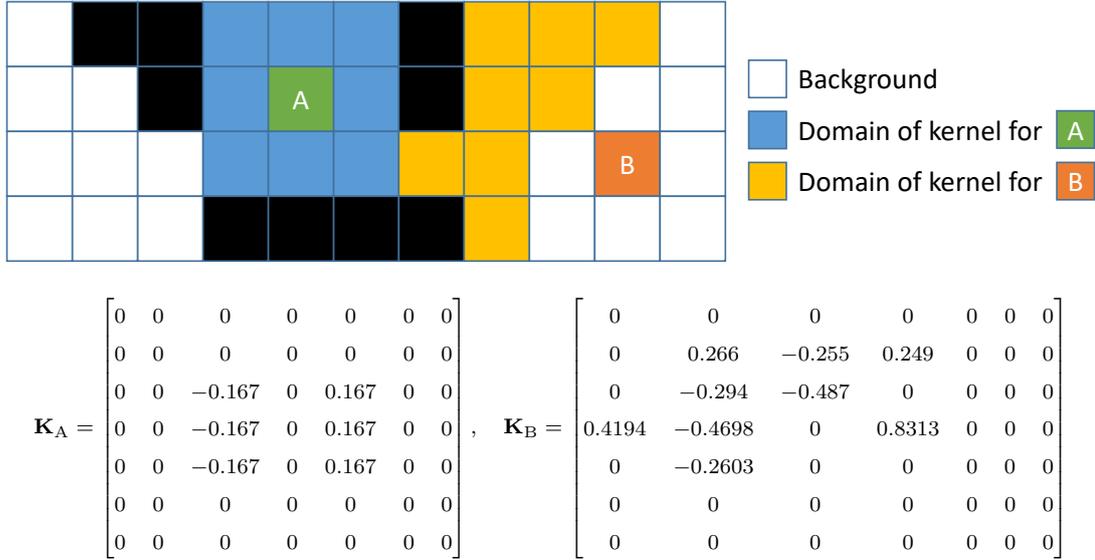


Figure 3.3: An example of computing 2D Savitzky-Golay filters on an arbitrary domain. In this example, we use a 3×3 kernel. For point A we can use the default square kernel. The order two Savitzky-Golay filter for the horizontal derivative is shown below as \mathbf{K}_A . For point B we use the 3^2 nearest pixels and build a custom order two Savitzky-Golay filter shown below as \mathbf{K}_B . In practice, higher order kernels provide better performance.

neighbours lie in \mathcal{F} . For those that do not, we find the d^2 nearest neighbours in \mathcal{F} (one of which will be the pixel itself). Where tie-breaks are needed, we do so randomly, though we observed no significant difference in performance if all tied pixels are included. In Figure 3.3 we show an example of a standard and non-standard case. All non-white pixels lie in \mathcal{F} . Pixel A has the available neighbours to use the square kernel while B does not and uses a custom kernel.

Each element in a kernel for a pixel is copied to the appropriate entries in a row of \mathbf{D}_x or \mathbf{D}_y . We similarly construct a matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ containing the a_{00} kernels, i.e. the smoothing kernel. Each row of these three matrices has d^2 non-zero entries.

3.3.7 3D K-nearest neighbours kernel

For normals-from-depth where a noisy depth map is provided, the K-nearest neighbours kernel idea can be extended to 3D. The idea is to use the depth map with (3.19) to transform pixels to 3D locations, then to perform the KNN search in 3D. The advantage of this is that kernels will avoid sampling across depth discontinuities where the large change in depth will result in adjacent pixels being far apart in 3D distance. This allows us to create large, robust kernels but

without smoothing over depth discontinuities.

3.3.8 Implementation

For an efficient implementation, all pixel coordinates from \mathcal{F} are placed in a KNN search tree so that local neighbourhoods can be found quickly and pixels that can use the square mask are identified by convolution of the mask with a square filter of ones.

To compute normals-from-depth, we use our proposed derivative matrices (with 3D KNN search) to compute the partial derivatives of z , take the cross product between horizontal and vertical derivatives (3.20) and normalise to give the unit surface normal.

To compute height-from-normals, we solve a system of the form of (3.21) (orthographic) or (3.24) (perspective). We augment the system of equations with a smoothness penalty of the form $\lambda(\mathbf{S} - \mathbf{I})\mathbf{z} = \mathbf{0}$, where λ is the regularisation weight. This encourages the difference between the smoothed and reconstructed z values to be zero. For the orthographic system, we resolve the unknown offset by solving for the minimum norm solution - equivalent to forcing the mean z value to zero. For the perspective case, since the system is homogeneous in theory we could solve for the $\|\mathbf{z}\|=1$ solution by solving a minimum direction problem using the sparse SVD. In practice, we find it is faster to add an additional equation forcing the solution at one pixel to unity. Finally, we can optionally include a depth prior simply by adding the linear equation $\omega\mathbf{Iz} = \omega\mathbf{z}_{\text{prior}}$, where ω is the prior weight.

3.4 Summary

In this chapter we first proposed a multichannel estimation to obtain $I_{\text{un}}, \rho, \varphi$. We demonstrate this method on a input of RGB polarimetric images data, the quality of output significantly improved compare with using only single channel. We then further extend this method to a multilight scenario, that fixed a camera, captured the object under different illumination conditions. By constraint ρ, φ are the same, the multilight estimation gives a more robust results. In the second part, we formulate a linear equation of height from normal. And the proposed 2D Savitzky-Golay kernel can provide a numerical derivation approximation and smoothness regulariser. These two method will be used in the following chapters in this thesis.

Chapter 4

Monocular shape-from-polarisation

The intrinsic challenge to recover the shape from polarisation is an innately ambiguous shape cue. In theory, this polarisation information alone restricts the surface normal at each pixel to six possible directions. Previous approaches under one view solve the problem of disambiguating these polarisation normals via propagation from the boundary under an assumption of global convexity [6, 54]. [40] also disambiguate polarisation normals with a global convexity assumption but estimate refractive index in addition. These works rely on diffuse polarisation model and known refractive index, the height map must be integrated from surface normal. A later method [70] showed how to express polarisation and shading constraints directly in terms of surface height, leading to a robust and efficient linear least squares solution. It show how to estimate the illumination, up to a binary ambiguity, making the method uncalibrated. However, they require known or uniform albedo.

In this chapter, we propose an alternative approach to the SfPol problem in Section 4.1. Like [42, 70], we estimate surface height directly. However, unlike all previous methods, we do not decompose the captured data into a *polarisation image* $I_{\text{un}}, \rho, \varphi$ 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.

We then explore the combination of shape-from-polarisation constraints with photometric constraints (i.e. photo polarimetric shape estimation) provided by two light sources in Section 4.2. Photometric stereo with three or more light sources is a very well studied problem with

robust solutions available under a range of different assumptions. Two source photometric stereo is still considered a difficult problem [63] even when the illumination is calibrated and albedo is known. We show that various formulations of two source photo-polarimetric stereo lead to the same general problem (in terms of surface height), that illumination can be estimated and that certain combinations of constraints lead to an albedo invariant formulation. Moreover, the refractive index can be estimated. Hence, with only modest additional data capture requirements (a polarisation image rather than an intensity image), we arrive at an approach for uncalibrated two source photometric stereo.

These two approaches are based on monocular setup and we evaluate these methods both with synthetic and real data in Section 4.3.

4.1 Shape-from-polarisation by nonlinear least squares

The observed intensity captured by polarisation camera follows a sinusoid function *w.r.t* the filter angle ϑ_j as shown in (2.38). In real data the captured polarimetric measurements always carry some noise due to the uncertainties from environments (i.e images contain noise). We assume the observed images are subject to additive Gaussian noise:

$$I_{\vartheta_j}^{\text{obs}} = I_{\vartheta_j}^{\text{pol}}(I_{\text{un}}, \rho, \varphi) + \epsilon$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is the unknown variance of the noise. Therefore, $I_{\vartheta_j}^{\text{obs}} \sim \mathcal{N}(I_{\vartheta_j}^{\text{pol}}, \sigma^2)$ is itself a normally distributed random variable. Hence, we can write a probabilistic polarisation model as:

$$p(I_{\vartheta_j}^{\text{obs}} | I_{\text{un}}, \varphi, \rho) = C(\sigma^2) \exp\left(-\frac{[I_{\vartheta_j}^{\text{obs}} - I_{\vartheta_j}^{\text{pol}}(I_{\text{un}}, \varphi, \rho)]^2}{2\sigma^2}\right), \quad (4.1)$$

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 [6] or surface height [42, 70] 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. A better way is to have an end-to-end method that solve the surface height in one.

4.1.1 Shape-from-polarisation 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 non-linear 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 $\nabla z \in \mathbb{R}^2$ containing the partial derivatives: $\nabla z = [\partial z/\partial x \ \partial z/\partial y]^T$.

From surface height to surface normal Surface normal from a height function $z(x, y)$ can be represented by its gradient. A tangent plane can be constructed at some surface point and the surface normal is the vector which perpendicular to its tangent plane. Hence we take the surface slope in x, y direction at a surface point $P = [x, y, z(x, y)]^T$. We have

$$\frac{\partial P}{\partial x} = \begin{bmatrix} 1 \\ 0 \\ \partial z/\partial x \end{bmatrix}, \quad \frac{\partial P}{\partial y} = \begin{bmatrix} 0 \\ 1 \\ \partial z/\partial y \end{bmatrix}$$

These two surface slopes construct a tangent plane at point P , hence the function $\mathbf{n} : \mathbb{R}^2 \mapsto \mathbb{R}^3$ transform the surface gradient into surface normal vector by taking cross product of these two vectors.

$$\mathbf{n}(\nabla z) = \frac{\partial P}{\partial x} \times \frac{\partial P}{\partial y} = \begin{bmatrix} -\partial z/\partial x \\ -\partial z/\partial y \\ 1 \end{bmatrix} = \begin{bmatrix} -\nabla z \\ 1 \end{bmatrix} \quad (4.2)$$

The derivative of the function \mathbf{n} w.r.t ∇z is a Jacobian matrix which given by

$$\mathbf{J}_{\mathbf{n}}(\nabla z) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \quad (4.3)$$

Surface normal normalisation The function $\tilde{\mathbf{n}} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ normalise the vector to get a unit length that $\|\tilde{\mathbf{n}}\| = 1$, hence

$$\tilde{\mathbf{n}}(\mathbf{n}) = \frac{\mathbf{n}}{\|\mathbf{n}\|} \quad (4.4)$$

The derivative of the function $\tilde{\mathbf{n}}$ w.r.t \mathbf{n} can given by a Jacobian matrix

$$\mathbf{J}_{\tilde{\mathbf{n}}}(\mathbf{n}) = \frac{\mathbf{I}}{\|\mathbf{n}\|} - \frac{\mathbf{nn}^T}{\|\mathbf{n}\|^3} \quad (4.5)$$

Surface normal to spherical coordinates It is convenient to transform the surface normal vector $\bar{\mathbf{n}}$ into spherical coordinates (ϕ, θ) in a viewer-centred coordinate system. We explicitly present the unit surface normal as $\tilde{\mathbf{n}} = [\tilde{n}_1, \tilde{n}_2, \tilde{n}_3]^T$. The azimuth angle function $\phi : \mathbb{R}^3 \mapsto \mathbb{R}^1$ is defined (along with its gradient) as follows:

$$\phi(\tilde{\mathbf{n}}) = \text{atan2}(\tilde{n}_2, \tilde{n}_1) \quad (4.6)$$

The derivative of the function α *w.r.t* $\tilde{\mathbf{n}}$ can be presented as a vector as

$$\nabla\phi(\tilde{\mathbf{n}}) = \begin{bmatrix} \frac{-\tilde{n}_2}{\tilde{n}_1^2 + \tilde{n}_2^2} \\ \frac{\tilde{n}_1}{\tilde{n}_1^2 + \tilde{n}_2^2} \\ 0 \end{bmatrix} \quad (4.7)$$

The zenith angle function $\theta : \mathbb{R}^3 \mapsto \mathbb{R}^1$ is computed by

$$\theta(\tilde{\mathbf{n}}) = \arccos(\tilde{n}_3) \quad (4.8)$$

The derivative of θ *w.r.t* $\tilde{\mathbf{n}}$ is

$$\nabla\theta(\tilde{\mathbf{n}}) = \begin{bmatrix} 0 \\ 0 \\ \frac{-1}{\sqrt{1-\tilde{n}_3^2}} \end{bmatrix} \quad (4.9)$$

Orientation to diffuse polarisation image The three quantities of polarisation images $I_{\text{un}}, \rho, \varphi$ can be computed from the spherical coordinates. For diffuse polarisation, the phase angle function $\phi : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is

$$\varphi(\phi) = \begin{cases} \phi, & \text{if } \alpha \in [0, \pi) \\ \phi - \pi, & \text{otherwise} \end{cases} \quad (4.10)$$

The derivative of function φ *w.r.t* ϕ is

$$\frac{\partial\varphi}{\partial\phi} = 1 \quad (4.11)$$

The unpolarised intensity is assumed to be Lambertian reflectance model that related to the surface albedo k_d , surface normal $\tilde{\mathbf{n}}$ and illumination \mathbf{s} , it can be modelled by function $I_{\text{un}} : \mathbb{R}^3 \mapsto \mathbb{R}^1$

$$I_{\text{un}}(\tilde{\mathbf{n}}) = k_d \tilde{\mathbf{n}}^T \mathbf{s} \quad (4.12)$$

The derivative of function I_{un} *w.r.t* $\tilde{\mathbf{n}}$ is

$$\nabla I_{\text{un}}(\tilde{\mathbf{n}}) = k_d \mathbf{s} \quad (4.13)$$

The degree of polarisation ρ can be computed by zenith angle θ and known refractive index η .

Assuming this as a diffuse polarisation model, the function $\rho : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is given by

$$\rho(\theta) = \frac{\sin^2(\theta)(\eta - \frac{1}{\eta})^2}{4 \cos(\theta) \sqrt{\eta^2 - \sin^2(\theta)} - \sin^2(\theta)(\eta + \frac{1}{\eta})^2 + 2\eta^2 + 2} \quad (4.14)$$

The derivative of function ρ w.r.t θ is given by

$$\begin{aligned} \frac{\partial \rho}{\partial \theta} = & \frac{2 \cos(\theta) \sin(\theta) \left(\eta - \frac{1}{\eta}\right)^2}{4 \cos(\theta) \sqrt{\eta^2 - \sin^2(\theta)} - \sin^2(\theta) \left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2} + \\ & \frac{\sin(\theta)^2 \left(\eta - \frac{1}{\eta}\right)^2}{\left(4 \cos(\theta) \sqrt{\eta^2 - \sin^2(\theta)} - \sin^2(\theta) \left(\eta + \frac{1}{\eta}\right)^2 + 2\eta^2 + 2\right)^2} \cdot \\ & \left(4 \sin(\theta) \sqrt{\eta^2 - \sin^2(\theta)} + 2 \cos(\theta) \sin(\theta) \left(\eta + \frac{1}{\eta}\right)^2 + \frac{4 \cos(\theta)^2 \sin(\theta)}{\sqrt{\eta^2 - \sin^2(\theta)}}\right) \end{aligned}$$

Orientation to specular polarisation image In specular polarisation reflectance model, the three quantities of polarisation images $I_{\text{un}}, \rho, \varphi$ are computed differently. In particular, if we assume (as in previous work [70]) 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 phase angle function $\varphi : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is given by

$$\varphi(\phi) = \begin{cases} \phi + \frac{\pi}{2}, & \text{if } \alpha \in [0, \pi) \\ \phi - \frac{\pi}{2}, & \text{otherwise} \end{cases} \quad (4.15)$$

The derivative of function φ w.r.t ϕ is

$$\frac{\partial \varphi}{\partial \phi} = 1 \quad (4.16)$$

The specular DOP is given by function $\rho : \mathbb{R}^1 \mapsto \mathbb{R}^1$ with known refractive index η :

$$\rho(\theta) = \frac{2 \sin(\theta)^2 \cos(\theta) \sqrt{\eta^2 - \sin^2(\theta)^2}}{\eta^2 - \sin(\theta)^2 - \eta^2 \sin(\theta)^2 + 2 \sin(\theta)^4}. \quad (4.17)$$

The derivative of function ρ w.r.t θ is given by

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

4.1.2 Nonlinear least squares

We can now build an equation of the polarisation reflectance model in terms of surface gradient from above. 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 or higher order kernel described in Section 3.3 which can be expressed as a matrix multiplication. We write this as a function $\mathbf{G} : \mathbb{R}^N \mapsto \mathbb{R}^{2N}$:

$$\mathbf{G}(\mathbf{z}) = \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} \mathbf{z}, \quad \mathbf{J}_{\mathbf{G}}(\mathbf{z}) = \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix}, \quad (4.18)$$

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 as described in Section 3.3.3. Using the derivation in Sections 4.1.1 we can construct a function $\mathbf{I}_{\vartheta}^{\text{pol}} : \mathbb{R}^{2N} \mapsto \mathbb{R}^N$ 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 $\mathbf{J}_{\mathbf{I}_{\vartheta}^{\text{pol}}}(\mathbf{G}) \in \mathbb{R}^{N \times 2N}$ 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:

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

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

$$\min_{\mathbf{z}} \mathbf{r}(\mathbf{z})^T \mathbf{r}(\mathbf{z}). \quad (4.19)$$

We emphasise that, under the assumption of the probabilistic model, this is the maximum likelihood solution for \mathbf{z} given the observed intensities. In practice, we minimise (4.19) using the trust-region-reflective algorithm, as implemented in the Matlab `lsqnonlin` function. We initialise with a plane, i.e. $\mathbf{z} = \mathbf{0}$.

4.1.3 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:

$$\mathbf{r}_{\text{smooth}}(\mathbf{z}) = \sqrt{w_{\text{smooth}}}\mathbf{L}\mathbf{z}, \quad \mathbf{J}_{\mathbf{r}_{\text{smooth}}}(\mathbf{z}) = \sqrt{w_{\text{smooth}}}\mathbf{L}, \quad (4.20)$$

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. If we measure the numerical derivative by Savitzky golay filter described in Section 3.3. We can use the corresponding regulariser to replace the Laplacian filter \mathbf{L} . Which prevents overflat the surface while smooth the noise at same time.

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:

$$\mathbf{r}_{\text{convex}}(\mathbf{z}) = \sqrt{w_{\text{convex}}} \begin{bmatrix} \sin(\mathbf{S}_{\text{boundary}}\boldsymbol{\alpha}(\mathbf{z})) - \sin(\boldsymbol{\alpha}_{\text{boundary}}) \\ \cos(\mathbf{S}_{\text{boundary}}\boldsymbol{\alpha}(\mathbf{z})) - \cos(\boldsymbol{\alpha}_{\text{boundary}}) \end{bmatrix}, \quad (4.21)$$

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.

4.1.4 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 (4.24). 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.

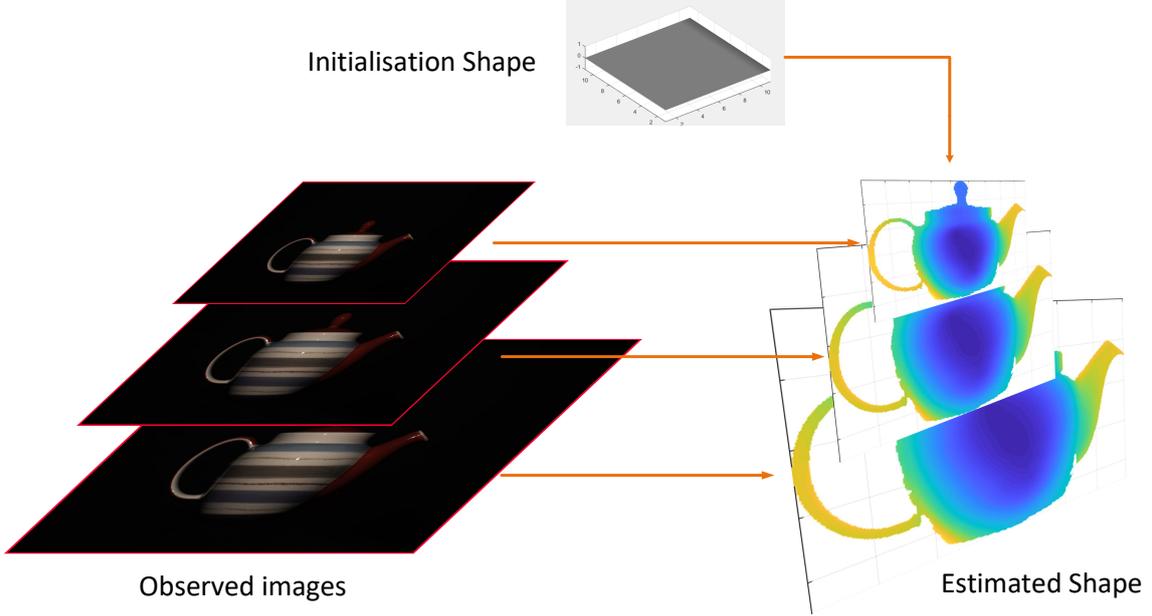


Figure 4.1: An illustration on how the hierarchical estimation works. We start from the lowest resolution with a plane as initialisation. The result from each layer will rescale to a proper resolution as an initialisation to the next optimisation. The final shape estimation will produce by the last layer which keep the original resolution.

4.1.5 Ratio-based formulation

(2.38) requires known albedo and lighting and assumes diffuse reflectance and diffuse polarisation. Alternatively, we proposed an equation by taking ratios between different polariser orientations:

$$\frac{I_{\vartheta_j}^{\text{pol}}(I_{\text{un}}, \varphi, \rho)}{I_{\vartheta_k}^{\text{pol}}(I_{\text{un}}, \varphi, \rho)} = f_{\vartheta_j, \vartheta_k}(\varphi, \rho) = \frac{1 + \rho \cos[2\vartheta_j - 2\varphi]}{1 + \rho \cos[2\vartheta_k - 2\varphi]} \quad (4.22)$$

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. By using this ratio formulation, 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 (4.22) 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 [70]. 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 (4.22) are given by:

$$\nabla f_{\vartheta_j, \vartheta_k}(\varphi, \rho) = \begin{bmatrix} \frac{2\rho \sin(2\varphi - 2\vartheta_k)(\rho \cos(2\varphi - 2\vartheta_j) + 1)}{(\rho \cos(2\varphi - 2\vartheta_k) + 1)^2} - \frac{2\rho \sin(2\varphi - 2\vartheta_j)}{\rho \cos(2\varphi - 2\vartheta_k) + 1} \\ \frac{\cos(2\varphi - 2\vartheta_j) - \cos(2\varphi - 2\vartheta_k)}{(\rho \cos(2\varphi - 2\vartheta_k) + 1)^2} \end{bmatrix} \quad (4.23)$$

We extend the ratio function to all pixels via the function $\mathbf{F}_{\vartheta_j, \vartheta_k}^{\text{mod}} : \mathbb{R}^{2N} \mapsto \mathbb{R}^N$ that computes the predicted ratios for all pixels from the surface gradient at every pixel. The derivatives of this function $\mathbf{J}_{\mathbf{F}_{\vartheta_j, \vartheta_k}^{\text{mod}}}(\mathbf{G}) \in \mathbb{R}^{N \times 2N}$ can again be computed by the appropriate combination of derivatives from Section 4.1.1. 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:

$$\mathbf{r}(\mathbf{z}) = \begin{bmatrix} \mathbf{F}_{\vartheta_1, \vartheta_2}^{\text{obs}} - \mathbf{F}_{\vartheta_1, \vartheta_2}^{\text{mod}}(\mathbf{G}(\mathbf{z})) \\ \vdots \\ \mathbf{F}_{\vartheta_{P-1}, \vartheta_P}^{\text{obs}} - \mathbf{F}_{\vartheta_{P-1}, \vartheta_P}^{\text{mod}}(\mathbf{G}(\mathbf{z})) \end{bmatrix}, \quad (4.24)$$

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

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

4.2 Photo-polarimetric stereo

We now consider a different scenario in which multiple polarisation images are captured with the camera and object fixed but the direction of a point light source being varied. We propose different possible constraints that derive from photo-polarimetric information. We then show how to combine these constraints with polarisation to achieve linear equations *w.r.t* unknown surface height.

4.2.1 Photo-polarimetric height constraints

Degree of polarisation constraint 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 here we will consider diffuse polarisation, due to subsurface scattering (see [7] for more details). The degree of diffuse polarisation ρ_d at a pixel point $\mathbf{u} = (x, y)$ can be expressed in terms of the refractive index η and the surface zenith angle $\theta \in [0, \frac{\pi}{2}]$ as (2.44) (Cf. [7]): Recall that the zenith angle is the angle between the unit surface normal vector $\mathbf{n}(\mathbf{u})$ and the viewing direction \mathbf{v} shows in Figure 4.2. If we know the degree of polarisation $\rho_d(\mathbf{u})$ and the refractive index η (or have good estimates of them at hand), (2.44) can be rewritten *w.r.t* the cosine of the zenith angle, and expressed in terms of the function, $f(\rho_d(\mathbf{u}), \eta)$, that depends on the measured degree of polarisation and the refractive index:

$$\cos(\theta(\mathbf{u})) = \mathbf{n}(\mathbf{u}) \cdot \mathbf{v} = f(\rho_d(\mathbf{u}), \eta) = \sqrt{\frac{\eta^4(1-\rho_d^2) + 2\eta^2(2\rho_d^2 + \rho_d - 1) + \rho_d^2 + 2\rho_d - 4\eta^3\rho_d\sqrt{1-\rho_d^2} + 1}{(\rho_d + 1)^2(\eta^4 + 1) + 2\eta^2(3\rho_d^2 + 2\rho_d - 1)}} \quad (4.25)$$

where we drop the dependency of ρ_d on (\mathbf{u}) for brevity.

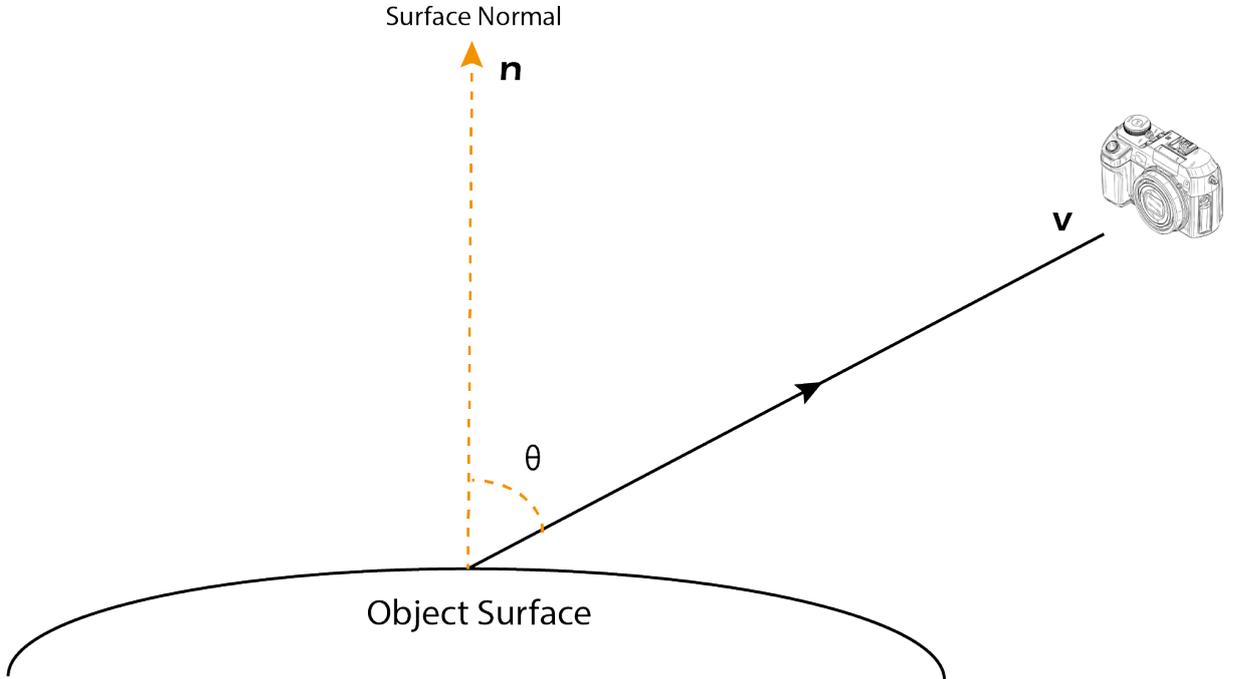


Figure 4.2: The zenith angle θ is just the angle between the surface normal and the viewer. For an orthographic camera, \mathbf{v} is constant across the image.

Shading constraint The unpolarised intensity provides an additional constraint on the surface normal direction via an appropriate reflectance model. We assume that pixels have been labelled as diffuse or specular dominant and restrict consideration to diffuse shading. In practice, we deal with specular pixels in the same way as [71] and simply assume that they point in the direction of the halfway vector between light source \mathbf{s} and view direction \mathbf{v} . For the diffuse pixels, we therefore assume that light is reflected according to the Lambert's law. Hence, the unpolarised intensity is related to the surface normal by:

$$I_{\text{un}}(\mathbf{u}) = \alpha(\mathbf{u}) \cos(\theta_i) = \alpha(\mathbf{u}) \mathbf{n}(\mathbf{u}) \cdot \mathbf{s}, \quad (4.26)$$

where $\alpha(\mathbf{u})$ is the albedo. Writing $\mathbf{n}(\mathbf{u})$ in terms of the gradient of z as reported in equations (4.2) and (4.4) that

$$\mathbf{n}(\mathbf{u}) = \sqrt{1 + |\nabla Z(\mathbf{u})|^2} \begin{bmatrix} \nabla Z(\mathbf{u}) \\ 1 \end{bmatrix} \quad (4.27)$$

then (4.26) can be rewritten as follows:

$$I_{\text{un}}(\mathbf{u}) = \alpha(\mathbf{u}) \frac{-\nabla Z(\mathbf{u}) \cdot \tilde{\mathbf{s}} + s_3}{\sqrt{1 + |\nabla Z(\mathbf{u})|^2}}, \quad (4.28)$$

with $\tilde{\mathbf{s}} = (s_1, s_2)$. This is a non-linear equation, but we will see in Sec. 4.2.1 and 4.2.1 how it is possible to remove the non-linearity by using the ratios technique.

Phase angle constraint An additional constraint comes from the phase angle, which determines the azimuth angle of the surface normal $\phi(\mathbf{u}) \in [0, 2\pi]$ up to a 180° ambiguity. This constraint can be rewritten as a collinearity condition [71], that is satisfied by either of the two possible azimuth angles implied by the phase angle measurement. 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, $[\cos(\varphi) \ \sin(\varphi)]$, to be collinear. This corresponds to requiring a cross product between these two

$$\begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \end{bmatrix} \times \begin{bmatrix} \cos(\varphi(\mathbf{u})) \\ \sin(\varphi(\mathbf{u})) \end{bmatrix} = 0 \quad (4.29)$$

In terms of the surface gradient, using (4.27), it is equivalent to

$$[-\cos(\varphi(\mathbf{u})) \ \sin(\varphi(\mathbf{u}))] \cdot \nabla z(\mathbf{u}) = 0 \quad (4.30)$$

A similar expression can be obtained for specular pixels, substituting in the $\frac{\pi}{2}$ -shifted phase angles. The advantage of doing this will become clear in Sec. 4.2.2.

Degree of polarisation ratio constraint Combining the two constraints illustrated in Sec. 4.2.1 and 4.2.1, we can arrive at a linear equation, that we refer to as the DOP ratio constraint. Recall that $\cos(\theta(\mathbf{u})) = \mathbf{n}(\mathbf{u}) \cdot \mathbf{v}$ and that we can express \mathbf{n} in terms of the gradient of z by using (4.27), then isolating the non-linear term in (4.25) we obtain

$$\sqrt{1 + |\nabla z(\mathbf{u})|^2} = \frac{-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{v}} + v_3}{\cos(\theta(\mathbf{u}))}, \quad (4.31)$$

where $\tilde{\mathbf{v}} = (v_1, v_2)$. On the other hand, considering the shading information contained in (4.28), and again isolating the non-linearity we arrive at the following

$$\sqrt{1 + |\nabla z(\mathbf{u})|^2} = \alpha(\mathbf{u}) \frac{-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{s}} + s_3}{I_{\text{un}}(\mathbf{u})}. \quad (4.32)$$

Note that we are supposing $\mathbf{s} \neq \mathbf{v}$, and $I_{\text{un}}(\mathbf{u}) \neq 0$, $f(\rho_d(\mathbf{u}), \eta) \neq 0$. Inspecting Eqs. (4.31) and (4.32) we obtain

$$\frac{-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{v}} + v_3}{\cos(\theta(\mathbf{u}))} = \alpha(\mathbf{u}) \frac{-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{s}} + s_3}{I_{\text{un}}(\mathbf{u})}. \quad (4.33)$$

We thus arrive at the following partial differential equation (PDE):

$$\mathbf{b}(\mathbf{u}) \cdot \nabla z(\mathbf{u}) = h(\mathbf{u}), \quad (4.34)$$

where

$$\mathbf{b}(\mathbf{u}) := \mathbf{b}^{(f, I_{\text{un}})} = I_{\text{un}}(\mathbf{u}) \tilde{\mathbf{v}} - \alpha(\mathbf{u}) \cos(\theta(\mathbf{u})) \tilde{\mathbf{s}}, \quad (4.35)$$

and

$$h(\mathbf{u}) := h^{(f, I_{\text{un}})} = I_{\text{un}}(\mathbf{u}) v_3 - \alpha(\mathbf{u}) \cos(\theta(\mathbf{u})) s_3. \quad (4.36)$$

Intensity ratio constraint Finally, we construct an intensity ratio constraint by considering two unpolarised images, $I_{\text{un},1}, I_{\text{un},2}$, taken from two different light source directions, \mathbf{s}, \mathbf{t} . We construct our constraint equation by applying (4.26) twice, once for each light source. We can remove the non-linearity as before and take a ratio, arriving at the following equation:

$$I_{\text{un},2}(-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{s}} + s_3) = I_{\text{un},1}(-\nabla z(\mathbf{u}) \cdot \tilde{\mathbf{t}} + t_3). \quad (4.37)$$

The above equation is independent of albedo, light source intensity and non-linear normalisation term. Again as before, we can rewrite (4.37) as a PDE in the form of (4.34) with

$$\mathbf{b}(\mathbf{u}) := \mathbf{b}^{(I_{\text{un},1}, I_{\text{un},2})} = I_{\text{un},2}(\mathbf{u}) \tilde{\mathbf{s}} - I_{\text{un},1}(\mathbf{u}) \tilde{\mathbf{t}}, \quad (4.38)$$

where $\tilde{\mathbf{t}} = (t_1, t_2)$, and

$$h(\mathbf{u}) := h^{(I_{\text{un},1}, I_{\text{un},2})} = I_{\text{un},2}(\mathbf{u}) s_3 - I_{\text{un},1}(\mathbf{u}) t_3. \quad (4.39)$$

Method	Phase angle	DOP ratio	Intensity ratio
[71]	✓	✓	
Proposed 1	✓		✓
Proposed 2		✓	✓
Proposed 3	✓	✓	✓

Table 4.1: Summary of the different formulations

4.2.2 A unified PDE formulation

Commencing from the constraints introduced in Sec. 4.2.1. In this section we show how to solve several different problems in photo-polarimetric shape estimation. The common feature is that these are all linear in the unknown height, and are expressed in a unified formulation in terms of a system of PDEs in the same general form:

$$\mathbf{B}(\mathbf{u})\nabla z(\mathbf{u}) = \mathbf{h}(\mathbf{u}), \quad (4.40)$$

where $\mathbf{B} : \bar{\Omega} \rightarrow \mathbb{R}^{J \times 2}$, $\mathbf{h} : \bar{\Omega} \rightarrow \mathbb{R}^{J \times 1}$, denoting by Ω the reconstruction domain and being $J = 2, 3$ or 4 depending on the cases. (4.40) is a compact and general equation, suitable for describing several cases in a unified differential formulation that solves directly for surface height.

Different combinations of the three constraints that are linear in the surface gradient can be combined in the formulation of (4.40). Each corresponds to different assumptions and have different pros and cons. We explore three variants and show that [71] is a special case of our formulation. We summarise the alternative formulations in Tab. 4.1.

Single light and polarisation formulation This case has been studied in [71]. It uses a single polarisation image, requires known illumination (though [71] show how this can be estimated if unknown) and assumes that albedo is known or uniform. This last assumption is quite restrictive, since it can only be applied to objects with homogeneous surfaces. With just a single illumination condition, only the phase angle and DOP ratio constraints are available. This thus becomes a special case of our general unified formulation (4.40), where \mathbf{B} and \mathbf{h} are defined as

$$\mathbf{B} = \begin{bmatrix} b_1^{(f, I_{\text{un}})} & b_2^{(f, I_{\text{un}})} \\ -\cos \varphi & \sin \varphi \end{bmatrix}, \quad \mathbf{h} = [h^{(f, I_{\text{un}})}, 0]^T, \quad (4.41)$$

with $\mathbf{b}^{(f, I_{\text{un}})}$ and $h^{(f, I_{\text{un}})}$ defined by (4.35) and (4.36), with uniform $\gamma(\mathbf{u})$ and $\mathbf{v} = [0, 0, 1]^T$.

Proposed 1: Albedo invariant formulation Our first proposed method uses the phase angle constraint (4.30) and two unpolarised images, taken from two different light source directions, obtained through (4.28) and combined as in (4.37). In this case the problem studied is described by the system of PDEs (4.40) with

$$\mathbf{B}(\mathbf{u}) = \begin{bmatrix} b_1^{(I_{\text{un},1}, I_{\text{un},2})} & b_2^{(I_{\text{un},1}, I_{\text{un},2})} \\ -\cos \varphi & \sin \varphi \end{bmatrix}, \mathbf{h}(\mathbf{u}) = \begin{bmatrix} h^{(I_{\text{un},1}, I_{\text{un},2})} \\ 0 \end{bmatrix}, \quad (4.42)$$

where $\mathbf{b}^{(I_{\text{un},1}, I_{\text{un},2})}$ and $h^{(I_{\text{un},1}, I_{\text{un},2})}$ defined as in (4.38) and (4.39). The phase angle does not depend on albedo and the intensity ratio constraint is invariant to albedo. As a result, this formulation is particularly powerful because it allows albedo invariant height estimation. Moreover, the light source directions in the two images can be estimated (again, in an albedo invariant manner) using the method in Sec. 4.2.5.

Once surface height has been estimated, we can compute the surface normal at each pixel and it is then straightforward to estimate an albedo map using (4.26). Where we have two diffuse observations, we can compute albedo from two equations of the form of (4.26) in a least squares sense. In real data, where we have specular pixel labels, we use only the diffuse observations at each pixel. To avoid artifacts at the boundary of specular regions, we introduce a gradient consistency term into the albedo estimation. We encourage the gradient of the albedo map to match the gradients of the intensity image for diffuse pixels.

Proposed 2: Phase invariant formulation Our second proposed method uses only the DOP ratio and the intensity ratio constraints. This means that phase angle estimates are not used. The advantage of this is that phase angles are subject to a shift of $\frac{\pi}{2}$ at specular reflections when compared to diffuse reflections. So, the phase angle constraint relies upon having accurate per-pixel specular labels, which classify reflections as either dominantly specular or diffuse (or alternatively use a mixed polarisation model [75] with a four way ambiguity). In this case we need a) two unpolarised intensity images, taken with two different light source directions, \mathbf{s} and \mathbf{t} , obtained through (4.28), b) polarisation information from the function $f(\rho, \eta)$ and c) knowledge of the albedo map. We need $\mathbf{s}, \mathbf{t}, \mathbf{v}$ non-coplanar in order to have the matrix field \mathbf{B} not singular. Note that the function f , obtained from polarization information (as in (4.25)), is the same for the two required images. The reason for this is that it does not depend on the light source directions but only on the viewer direction \mathbf{v} which does not change. This formulation can be deduced starting from (4.37) and (4.33), arriving at a PDE system as in (4.40) with

$$\mathbf{B} = [\mathbf{b}^{(f, I_{\text{un},1})}, \mathbf{b}^{(f, I_{\text{un},2})}, \mathbf{b}^{(I_{\text{un},1}, I_{\text{un},2})}]^T, \quad (4.43)$$

and $\mathbf{h} = [h^{(f, I_{\text{un},1})}, h^{(f, I_{\text{un},2})}, h^{(I_{\text{un},1}, I_{\text{un},2})}]^T$, using (4.35), (4.36), (4.38), (4.39) to define the vector fields \mathbf{b} and the scalar fields h that appear in \mathbf{B} and \mathbf{h} .

Proposed 3: Most constrained formulation Our final proposed method combines all of the previous constraints, leading to a problem of the form (4.40) with

$$\mathbf{B} = \begin{bmatrix} b_1^{(f, I_{\text{un},1})} & b_2^{(f, I_{\text{un},1})} \\ b_1^{(f, I_{\text{un},2})} & b_2^{(f, I_{\text{un},2})} \\ b_1^{(I_{\text{un},1}, I_{\text{un},2})} & b_2^{(I_{\text{un},1}, I_{\text{un},2})} \\ -\cos \varphi & \sin \varphi \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} h^{(f, I_{\text{un},1})} \\ h^{(f, I_{\text{un},2})} \\ h^{(I_{\text{un},1}, I_{\text{un},2})} \\ 0 \end{bmatrix}. \quad (4.44)$$

This formulation uses the most information and so is potentially the most robust method. However, it requires known albedo in order to use the DOP ratio constraint. Nevertheless, it is possible to first apply proposed method 1, estimate the albedo and then re-estimate surface height using the maximally constrained formulation and the estimated albedo map. In fact, the best performance is obtained by iterating these two steps, alternately using the surface height estimate to compute albedo and then using the updated albedo to re-compute surface height.

Extension to colour images We now consider how to extend the above systems of equations when colour information is available. If a surface is lit by a coloured point source, then each pixel provides three equations of the form in (4.26). In principle, this provides no more information than a grayscale observation since the surface normal and light source direction are fixed across colour channels. However, in the presence of noise using all three observations improves robustness. In particular, if the albedo value at a pixel is lower in one colour channel, the signal to noise ratio will be worse in that channel than the others. For a multicoloured object, it is impossible to choose a single colour channel that provides the best signal to noise ratio across the whole object. For this reason, we propose to use information from all colour channels where available.

We already exploit colour information in the estimation of the polarisation image in Sec. 3.1. Hence, the phase angle estimates have already benefited from the improved robustness. Both the DOP ratio and intensity ratio constraints can also exploit colour information by repeating each constraint three times, once for each colour channel. In the case of the intensity ratio, the colour albedo once again cancels if ratios are taken between the same colour channels under different light source directions.

4.2.3 Height estimation via linear least squares

We have seen that each of the variants illustrated in the previous section, each with different advantages, can be written as a PDE system (4.40). Denoting by M the number of pixels, we discretise the gradient in (4.40) via finite differences, arriving at the following linear system in \mathbf{z}

$$\mathbf{A}\mathbf{z} = \bar{\mathbf{h}}, \quad (4.45)$$

where $\mathbf{A} = \bar{\mathbf{B}}\mathbf{G}$, with $\mathbf{G} \in \mathbb{R}^{2M \times M}$ the matrix of derivative described in (4.18). $\bar{\mathbf{B}} \in \mathbb{R}^{JM \times 2M}$ is the discrete per-pixel version of the matrix $\mathbf{B}(\mathbf{x})$, hence $\mathbf{A} \in \mathbb{R}^{JM \times M}$, where J depends on the various proposed cases reported in Sec. 4.2.2 ($J = 2$ for (4.41) and (4.42), $J = 3$ for (4.43) and $J = 4$ for (4.44)). $\bar{\mathbf{h}}$ is the discrete per-pixel version of the function $\mathbf{h}(\mathbf{x})$, $\bar{\mathbf{h}} \in \mathbb{R}^{JM \times 1}$, and $\mathbf{z} \in \mathbb{R}^{M \times 1}$ the vector of the unknown height values. The resulting discrete system is large, since we have JM linear equations in M unknowns, but sparse, since \mathbf{A} has few non-zero values for each row, and has as unknowns the height values. The per-pixel matrix \mathbf{A} is a full-rank matrix, for each choice of $\bar{\mathbf{B}}$ that comes from the proposed formulations in Sec. 4.2.2, under the different assumptions specified for each case. The per-pixel matrix \mathbf{A} related to [71] is full-rank except in one case: when the first two components of the light vector \mathbf{s} are non-zero and $s_1 = -s_2$ and it happens that the phase angle is $\varphi = \pi/4$ at least in one pixel. In that case, the matrix has a rank-deficiency (though in practice φ assuming a value of exactly $\pi/4$, up to numerical tolerance, is unlikely).

We want to find a solution of (4.45) in the least-squares sense, i.e find a vector $\mathbf{z} \in \mathbb{R}^M$ such that

$$\|\mathbf{A}\mathbf{z} - \bar{\mathbf{h}}\|_2^2 \leq \|\mathbf{A}\mathbf{y} - \bar{\mathbf{h}}\|_2^2, \quad \forall \mathbf{y} \in \mathbb{R}^M. \quad (4.46)$$

Considering the associated system of normal equations

$$\mathbf{A}^T(\mathbf{A}\mathbf{z} - \bar{\mathbf{h}}) = 0, \quad (4.47)$$

it is well-known that if there exists $\mathbf{z} \in \mathbb{R}^M$ that satisfies (4.47), then \mathbf{z} is also solution of the least-squares problem, i.e. \mathbf{z} satisfies (4.46). Since \mathbf{A} is a full-rank matrix, then the matrix $\mathbf{A}^T\mathbf{A}$ is not singular, hence there exists a unique solution \mathbf{z} of (4.47) for each data term $\bar{\mathbf{h}}$. Since neither \mathbf{B} nor \mathbf{h} depend on z in (4.40), the solution can be computed only up to an additive constant (which is consistent with the orthographic projection assumption). To resolve the unknown constant, knowledge of z at just one pixel is sufficient. In our implementation, we remove the height of one pixel from the variables and substitute its zero value elsewhere.

4.2.4 Refractive index estimation

Since proposed 2 method only consider the shading cue and phase cue, it is enough to estimate the surface height by the method describe in Section 4.2.3. We go through the equations (4.2), (4.4) and (4.8) to retrieve the zenith angle value from the surface height. And estimated degree of polarisation denoted by ρ_{est} can be computed by the method described in Sec. 3.2. The DoP (2.44) models its relationship with θ and η . So with the known degree of polarisation ρ and zenith angle θ , an optimum refractive index can be estimated by solving the following minimisation problem

$$\min_{\eta} \sum_{\mathbf{u}} \|\rho_{\text{est}}(\mathbf{u}) - \rho_d(\theta(\mathbf{u}), \eta)\|^2. \quad (4.48)$$

Where we assume the refractive index is uniform of target object. The estimated refractive index can be substituted back to proposed 3 method for more accurate height reconstruction.

4.2.5 Two source lighting estimation

Our three proposed shape estimation methods require knowledge of the two light source directions. Previously, Smith *et al.* [71] showed that a single polarisation image can be used to estimate illumination conditions up to a binary ambiguity. However, to do so, they assumed that the albedo was known or uniform, and they also worked only with a single colour channel. In a two source setting, we show that it is possible to estimate both light source directions simultaneously, and do so in an albedo invariant manner. Moreover, we can exploit information across different colour channels to improve robustness to noise. Hence, our three methods can be used in an uncalibrated setting.

The intensity ratio (4.37) provides one equation per pixel relating unpolarised intensities, surface gradient and light source directions. Given two polarisation images with different light directions, we have one such equation per pixel and six unknowns in total. We assume that ambiguous surface gradient estimates are known from ρ and φ , and then use (4.37) to estimate the light source directions.

The intensity ratio (4.37) is homogeneous in \mathbf{s} and \mathbf{t} and so has a trivial solution $\mathbf{s} = \mathbf{t} = [0 \ 0 \ 0]^T$. If we assume that the intensity of the light source remains constant in each colour channel across the two images, then this intensity divides out when taking an intensity ratio and so the length of the light source vectors is arbitrary. We therefore constrain them to unit length (avoiding the trivial solution), and represent them by spherical coordinates

(θ_s, α_s) and (θ_t, α_t) , such that $[s_1, s_2, s_3] = [\cos \alpha_s \sin \theta_s, \sin \alpha_s \sin \theta_s, \cos \theta_s]$ and $[t_1, t_2, t_3] = [\cos \alpha_t \sin \theta_t, \sin \alpha_t \sin \theta_t, \cos \theta_t]$.

This reduces the number of unknowns to four. We can now write the residual at each pixel given an estimate of the light source directions. There are two possible residuals, depending on which of the two ambiguous polarisation normals we use. From the phase angle and the zenith angle estimated from the degree of polarisation using (4.25), we have two possible surface normal directions at each pixel and therefore two possible gradients: $z_x(\mathbf{u}) \approx \pm \cos \phi(\mathbf{u}) \tan \theta(\mathbf{u})$, $z_y(\mathbf{u}) \approx \pm \sin \phi(\mathbf{u}) \tan \theta(\mathbf{u})$. Hence, the residuals at pixel \mathbf{u}_j in channel c are given by either:

$$\begin{aligned} r_{j,c}(\theta_s, \alpha_s, \theta_t, \alpha_t) &= I_{\text{un},1}^c(\mathbf{u}_j)(-z_x(\mathbf{u}_j)t_1 - z_y(\mathbf{u}_j)t_2 + t_3) - \\ &\quad I_{\text{un},2}^c(\mathbf{u}_j)(-z_x(\mathbf{u}_j)s_1 - z_y(\mathbf{u}_j)s_2 + s_3), \\ &\text{or} \\ q_{j,c}(\theta_s, \alpha_s, \theta_t, \alpha_t) &= I_{\text{un},1}^c(\mathbf{u}_j)(z_x(\mathbf{u}_j)t_1 + z_y(\mathbf{u}_j)t_2 + t_3) - \\ &\quad I_{\text{un},2}^c(\mathbf{u}_j)(z_x(\mathbf{u}_j)s_1 + z_y(\mathbf{u}_j)s_2 + s_3). \end{aligned}$$

We can now write a minimisation problem for light source direction estimation by summing the minimum of the two residuals over all pixels and colour channels:

$$\min_{\theta_s, \alpha_s, \theta_t, \alpha_t} \sum_{j,c} \min[r_{j,c}^2(\theta_s, \alpha_s, \theta_t, \alpha_t), q_{j,c}^2(\theta_s, \alpha_s, \theta_t, \alpha_t)].$$

The minimum of two convex functions is not itself convex and so this optimisation is non-convex. However, we find that, even with a random initialisation, it almost always converges to the global minimum. As in [71], the solution is still subject to a binary ambiguity, in that if (\mathbf{s}, \mathbf{t}) is a solution then $(\mathbf{T}\mathbf{s}, \mathbf{T}\mathbf{t})$ is also a solution (with $\mathbf{T} = \text{diag}([-1, -1, 1])$), corresponding to the convex/concave ambiguity. We resolve this simply by choosing the maximal solution when surface height is later recovered.

4.3 Experiments and evaluation

We present experimental results on both synthetic and real data for above two approaches. For nonlinear least square method, we compare the two proposed methods (the ratio-based formulation and the full optimisation) to both classical [6] and state-of-the-art [70] methods. For photo-polarimetric stereo we are focusing on evaluating object with varying albedo, and compare with [70] method.

4.3.1 Evaluating nonlinear square method

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

We simulate polarisation using (2.38) 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. 4.2. 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. 4.3 and 4.4. Normal maps are visualised as $R = (n_x + 1)/2$, $G = (n_y + 1)/2$ and $B = (n_z + 1)/2$. For comparison method [6], surface normals are estimated directly. For the proposed method and comparison method [70], 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 [6, 70]. 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 propagation method [6] can handle varying albedo but is extremely sensitive to noise and incorrectly resolves convex/concave interpretations in some places. The linear method [70] 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.

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
	[70]	13.47	8.60	8.10	10.18	18.51	16.30	19.00	29.76
	[6]	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
	[70]	10.42	15.64	11.17	15.39	13.36	17.27	17.35	22.39
	[6]	36.68	42.14	42.81	43.34	34.96	44.17	42.33	46.52

Table 4.2: Height and surface normal errors on synthetic data. Results shown for proposed ratio and full optimisation method and two comparison methods.

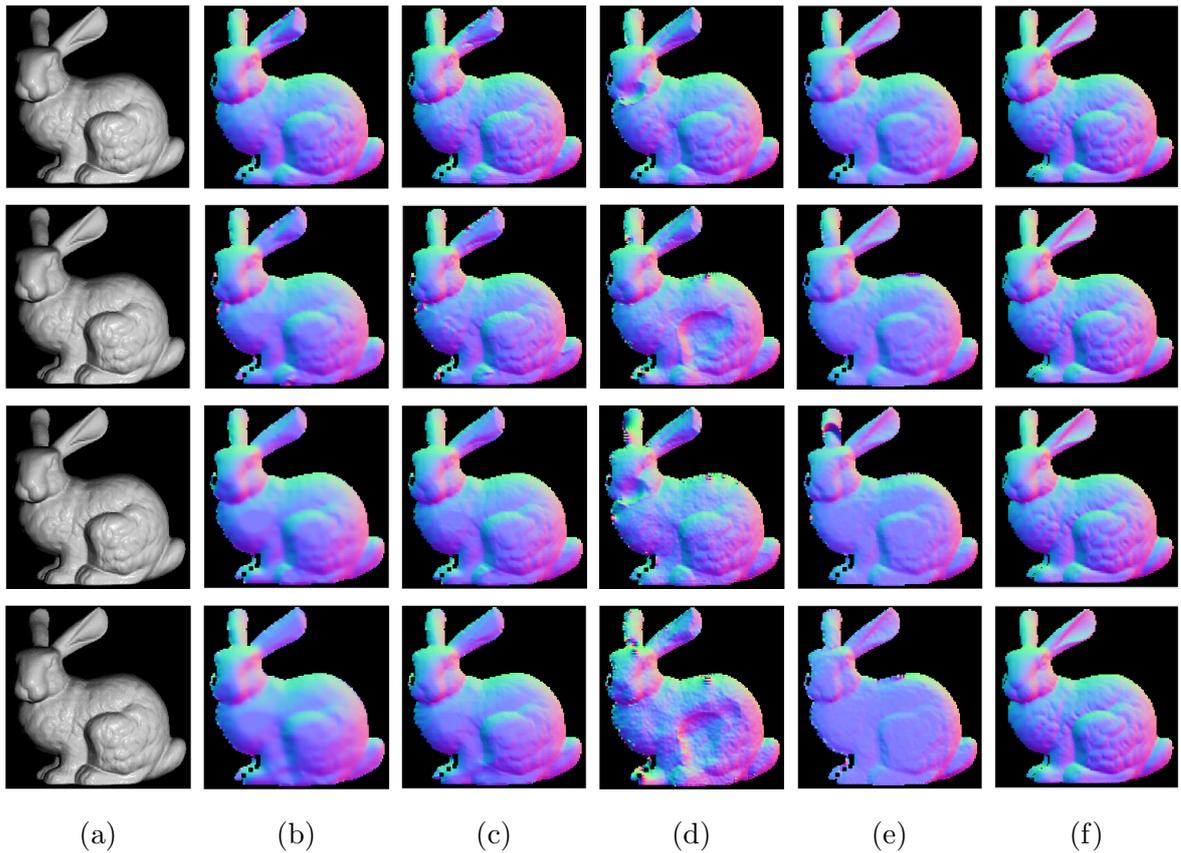


Figure 4.3: Qualitative results on synthetic Blinn-Phong bunny with varying albedo. The four rows are synthetic data with Gaussian noise of standard deviation $\sigma = 0\%, 0.5\%, 1\%, 2\%$ respectively. (a) Input; (b) normal map derived from height recovered by proposed ratio method; (c) normal map from full optimisation method; (d) normal map from [6]; (e) normal map from [70]; (f) ground truth.

Real data We show qualitative results on real images in Figures 4.5, 4.7 and 4.9 (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 [70] 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 [70] 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 4.9(f) that our method is able to recover the fine detail in the writing on the handle of the watergun. The object in Figure 4.8 contains varying albedo. This causes [70] 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.

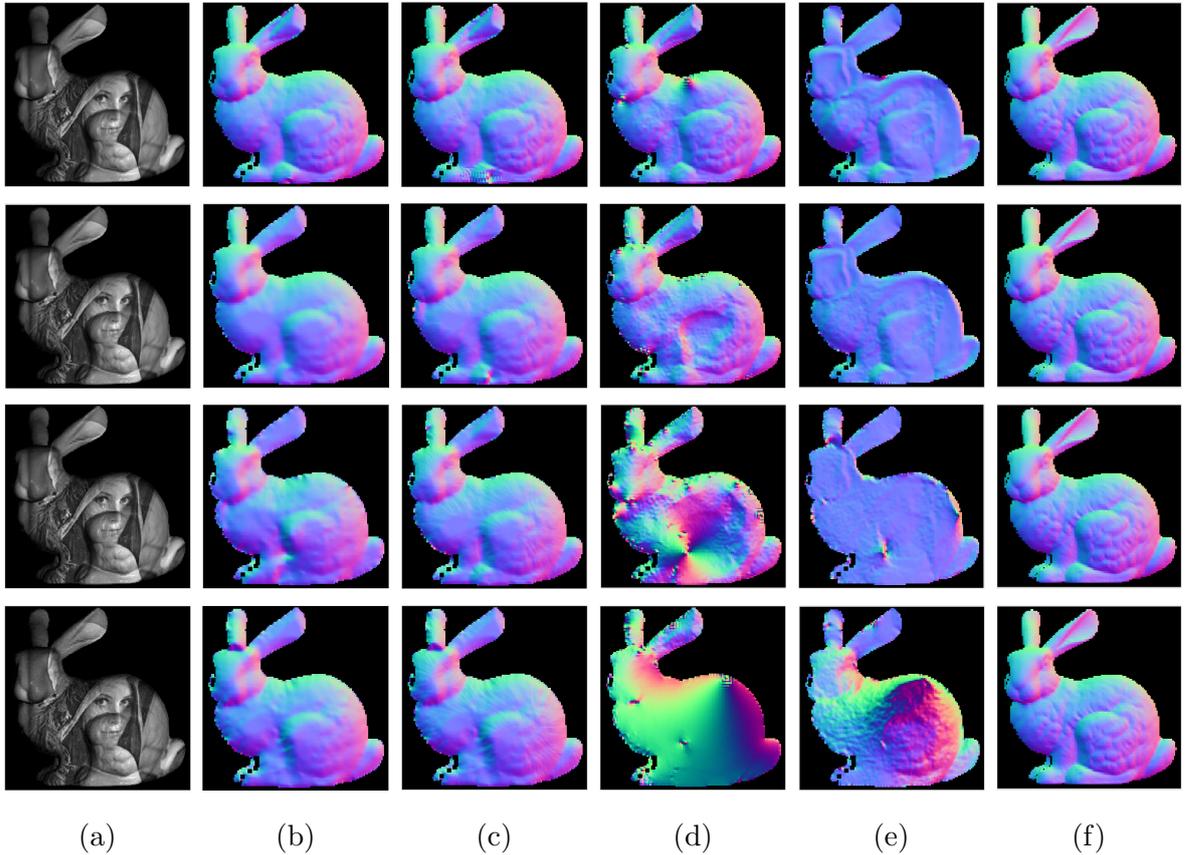


Figure 4.4: Qualitative results on synthetic Blinn-Phong bunny with varying albedo. The four rows are synthetic data with Gaussian noise of standard deviation $\sigma = 0\%, 0.5\%, 1\%, 2\%$ respectively. (a) Input; (b) normal map derived from height recovered by proposed ratio method; (c) normal map from full optimisation method; (d) normal map from [6]; (e) normal map from [70]; (f) ground truth.

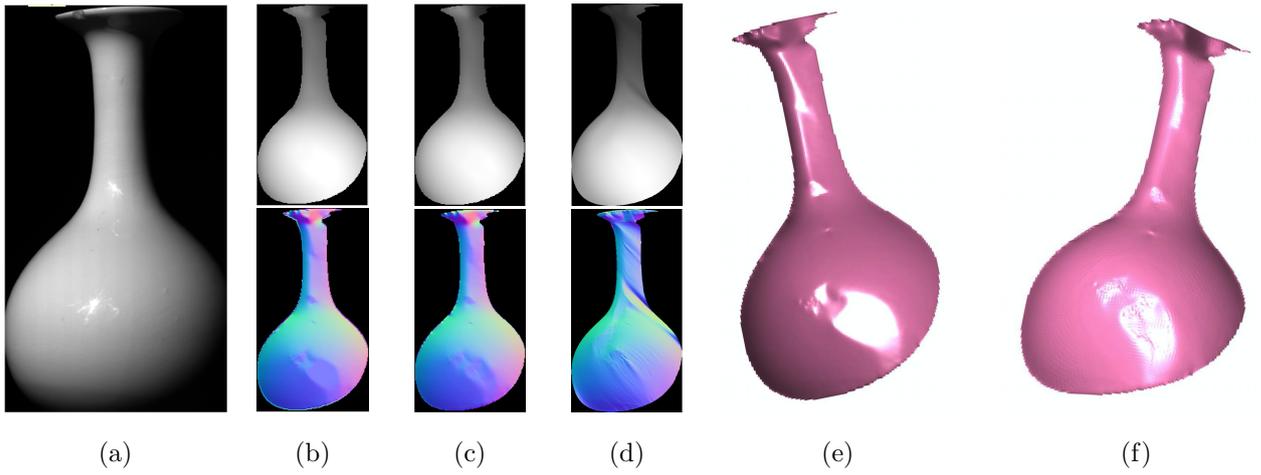


Figure 4.5: Qualitative results on porcelain vase. See Fig. 4.9 caption for details.

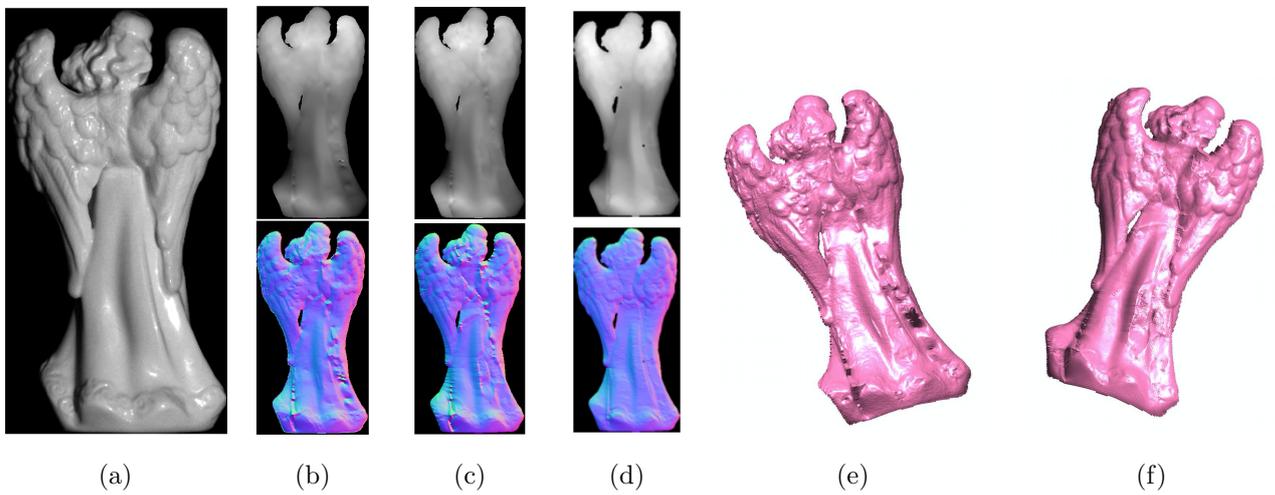


Figure 4.6: Qualitative results on porcelain angel statue. See Fig. 4.9 caption for details.

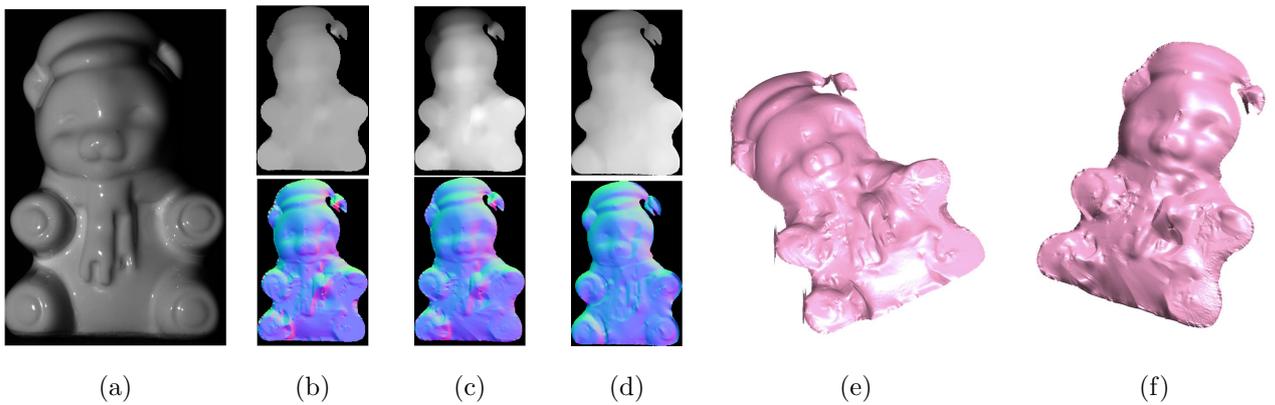


Figure 4.7: Qualitative results on porcelain bear. See Fig. 4.9 caption for details.

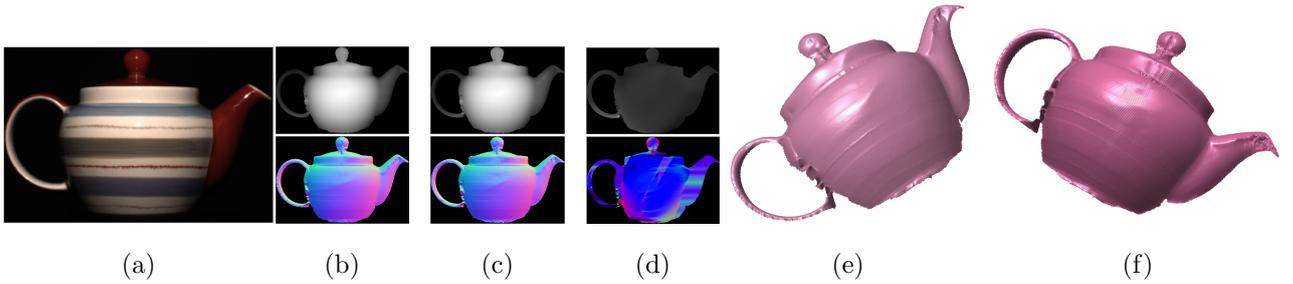


Figure 4.8: Qualitative results on color porcelain teapot. See Fig. 4.9 caption for details.

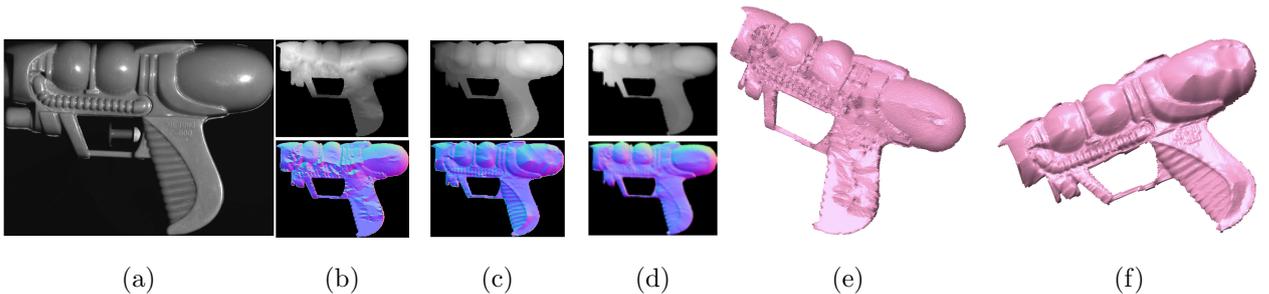


Figure 4.9: Qualitative results on plastic watergun: (a) Input grayscale image; (b) Recovered depth map and normal map from ratio method; (c) Recovered depth and normal map by full polarisation model and estimated light source; (d) Recovered depth map and normal map from [70] (e) a new pose of object estimated from ratio method. (f) a new pose of captured object calculated from full polarisation model.

4.3.2 Evaluating photo-polarimetric stereo method

We begin by using synthetic data generated from the 3DRFE dataset (Fig. 4.10). We differentiate to obtain surface normals and compute unpolarised intensities by rendering the surface using light sources $\mathbf{s} = [-50, 0, 104]^T$ and $\mathbf{t} = [0, -50, 104]^T$ according to (4.26). We simulate the effect of polarisation according to (2.38), varying the polariser angle between 0° and 180° in 10° increments. Next, we corrupt this data by adding Gaussian noise with zero mean and standard deviation σ , saturate and quantise to 8 bits. This noisy data provides the input to our reconstruction. First, we estimate a polarisation image using the method in Sec. 3.1, then apply each of the proposed methods or the state-of-the-art comparison method [10, 51, 69] to recover the height map and light estimation.

Fig. 4.10 show the estimated normal, depth and normal error against ground truth by our proposed method. We also estimate real images of snooker ball where we have a ground truth with same light conditions as 3DRFE dataset. Depth estimation qualitative result shows in

Fig. 4.11 and albedo estimation qualitative result shows in Fig. 4.12. We report the light estimation results by our proposed method, [69] and [10] in Tab. 4.3, and quantitative comparison on depth estimation, albedo estimation in Tab. 4.4. We do show our method outperformed others.

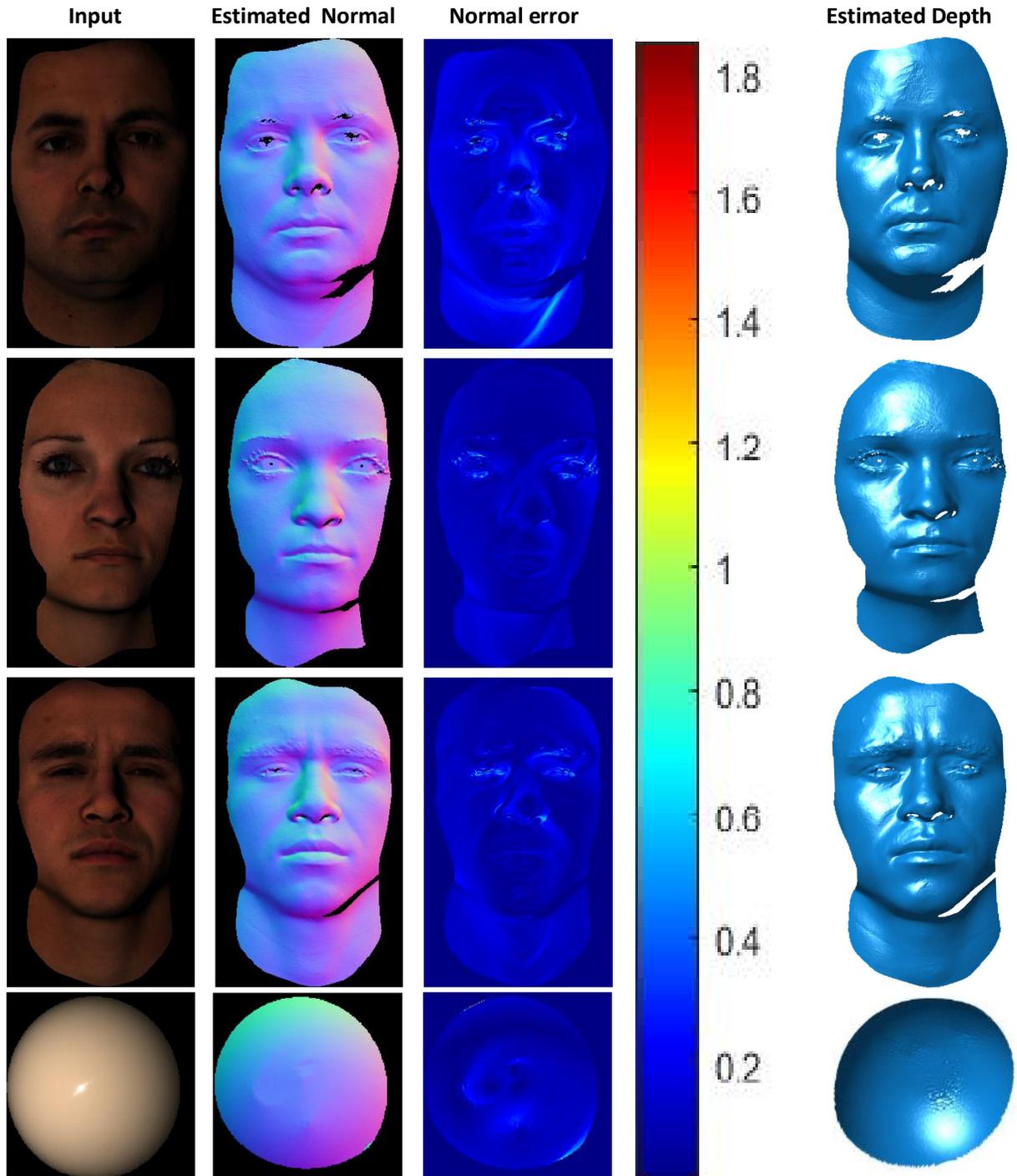


Figure 4.10: Estimation on synthetic data of 3DRFE dataset

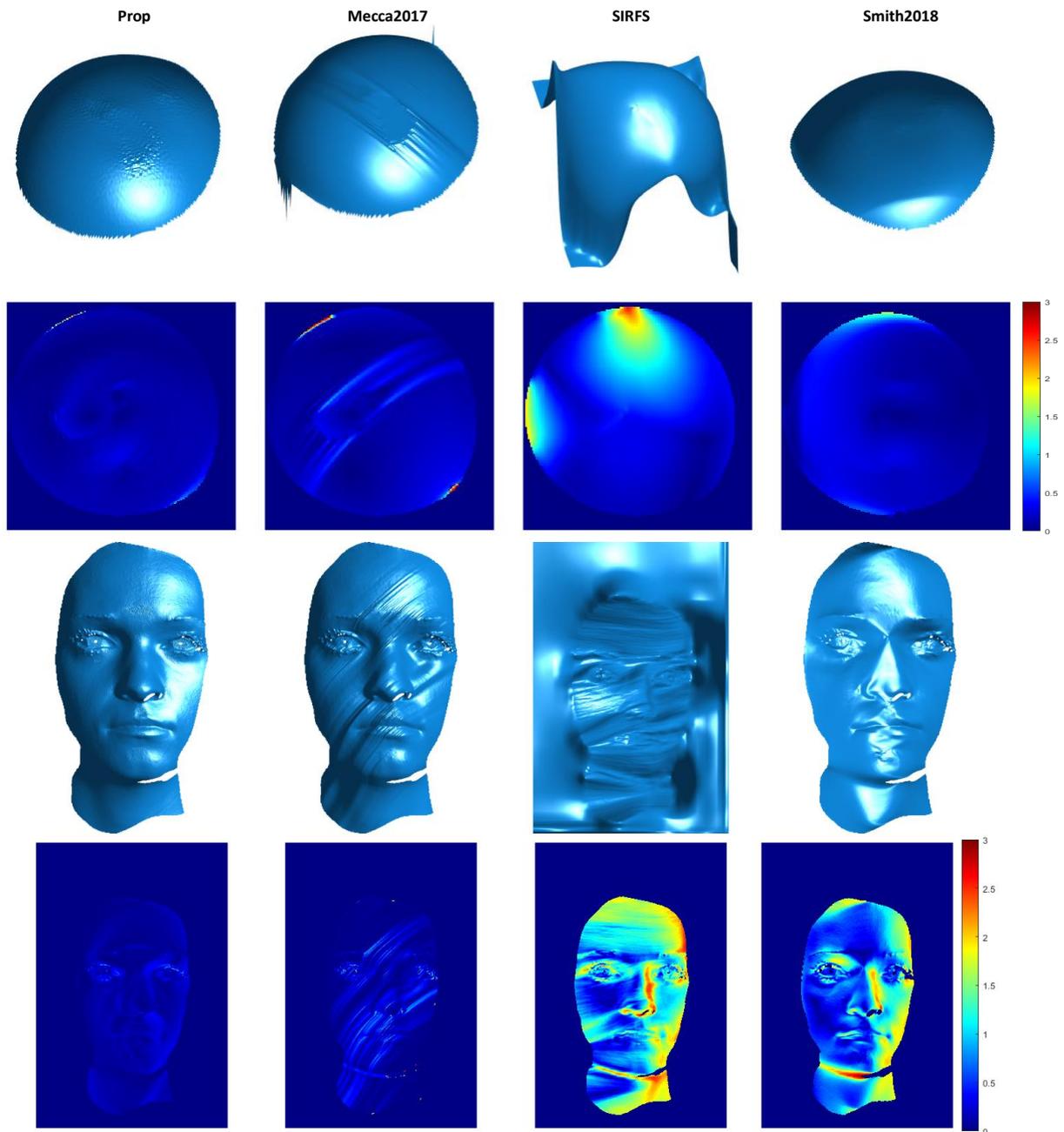


Figure 4.11: We show our depth reconstruction and surface normal error with our proposed method, Mecca2017 [51], SIRFS [10] and Smith2018 [69]. And the light estimation has been showned in Table 4.3

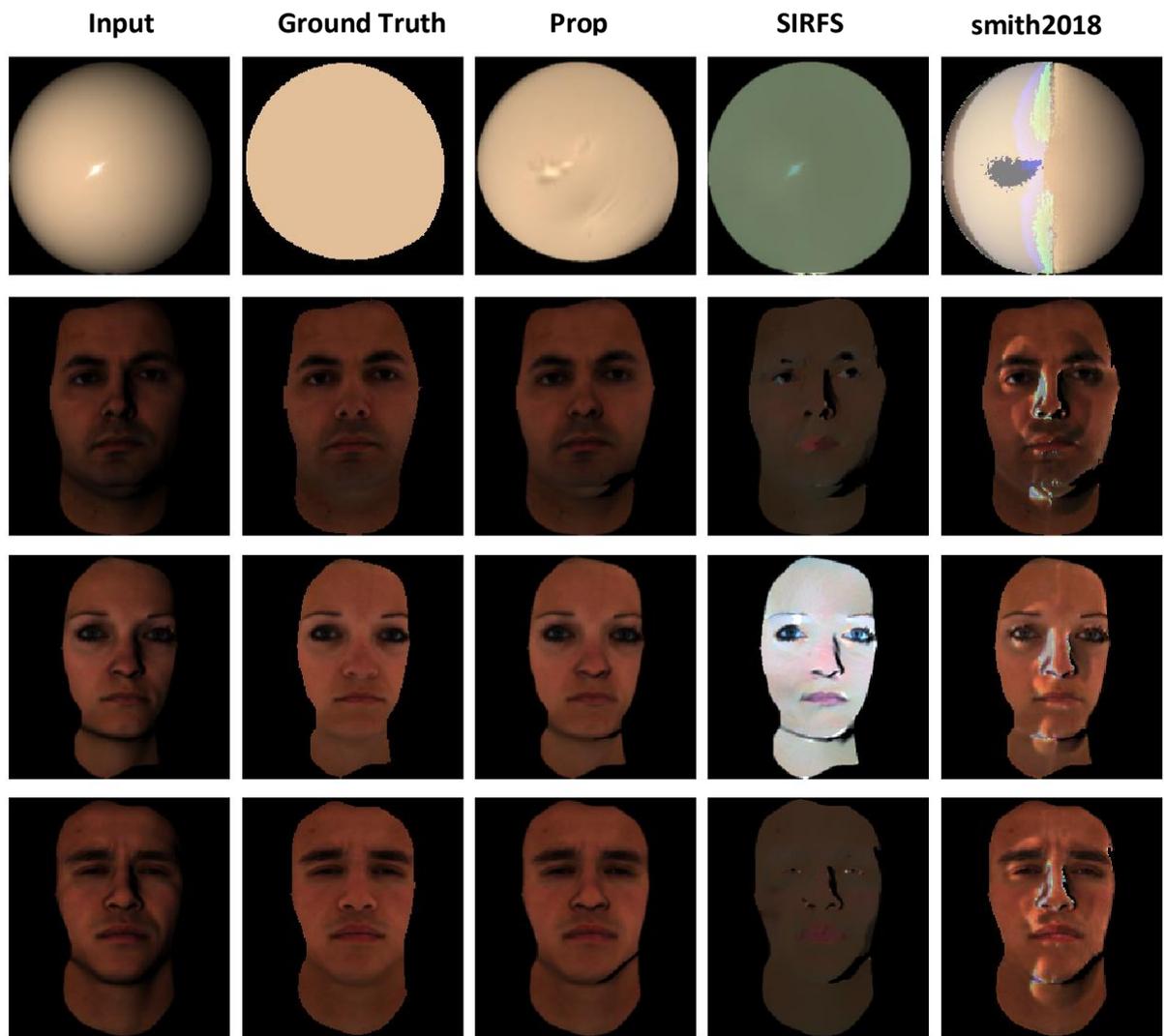


Figure 4.12: We show albedo estimation comparison with different methods, from left to right which are proposed, Mecca2017 [10], and Smith2018 [69].

image	prop		smith2018height [69]		SIRFS [10]	
	Light	RME	Light	RME	Light	RME
snooker 1	-0.42	0.125	-0.40	0.129	-0.67	0.212
	-0.07		-0.03		-0.06	
	0.91		0.92		0.74	
snooker 2	-0.05	0.090	-0.06	0.172	-0.67	0.821
	-0.44		-0.37		-0.06	
	0.90		0.93		0.74	
3DRFE 1	-0.41	0.176	-0.48	0.192	-0.61	0.673
	-0.13		-0.19		0.54	
	0.90		0.86		0.58	
3DRFE 2	-0.04	0.038	-0.02	0.072	-0.62	1.253
	-0.50		-0.57		0.54	
	0.87		0.82		0.57	

Table 4.3: Light estimation on snooker ball image and 3DRFE image. While the ground truth are $[-0.51, 0, 0.86]$, $[0, -0.51, 0.86]$ respectively.

Method	Snooker ball				3DRFE			
	Albedo		Normals		Albedo		Normals	
	AME	STD	MAE	STD	AME	STD	MAE	STD
ours	0.114	0.097	0.094	0.057	0.0367	0.045	0.222	0.151
Mecca2017	N/A	N/A	0.133	0.143	N/A	N/A	0.253	0.309
SIRFS	0.119	0.090	0.500	0.425	0.207	0.198	0.846	0.467
Smith2018	0.258	0.199	0.209	0.183	0.108	0.127	0.850	0.501

Table 4.4: We compare albedo absolute mean error + standard deviation and compare shape mean angular error + standard deviation for Figure 4.11 with different method.

In Fig. 4.13 we show qualitative results on four real objects with spatially varying albedo. From left to right we show: an image from the input sequence; the surface normals of the estimated height map (inset sphere shows how orientation is visualised as colour); the estimated albedo map; a re-rendering of the estimated surface and albedo map under novel lighting with Blinn-Phong reflectance [14]; a rotated view of the estimated surface; and, for comparison, reconstructions of the same surfaces using [71]. The results of [71] are highly distorted in the

presence of varying albedo. Our approach avoids transfer of albedo details into the recovered shape, leading to convincing relighting results. We evaluate performance on refractive index estimation. The synthetic 3DRFE data are rendered with a chosen refractive index of $\eta = 1.5$. The real objects are either made of porcelain (ground truth refractive index reported in [1]) or phenol formaldehyde resin (ground truth refractive index reported in [76]). We show our estimates and the ground truth values in Tab. 4.5.

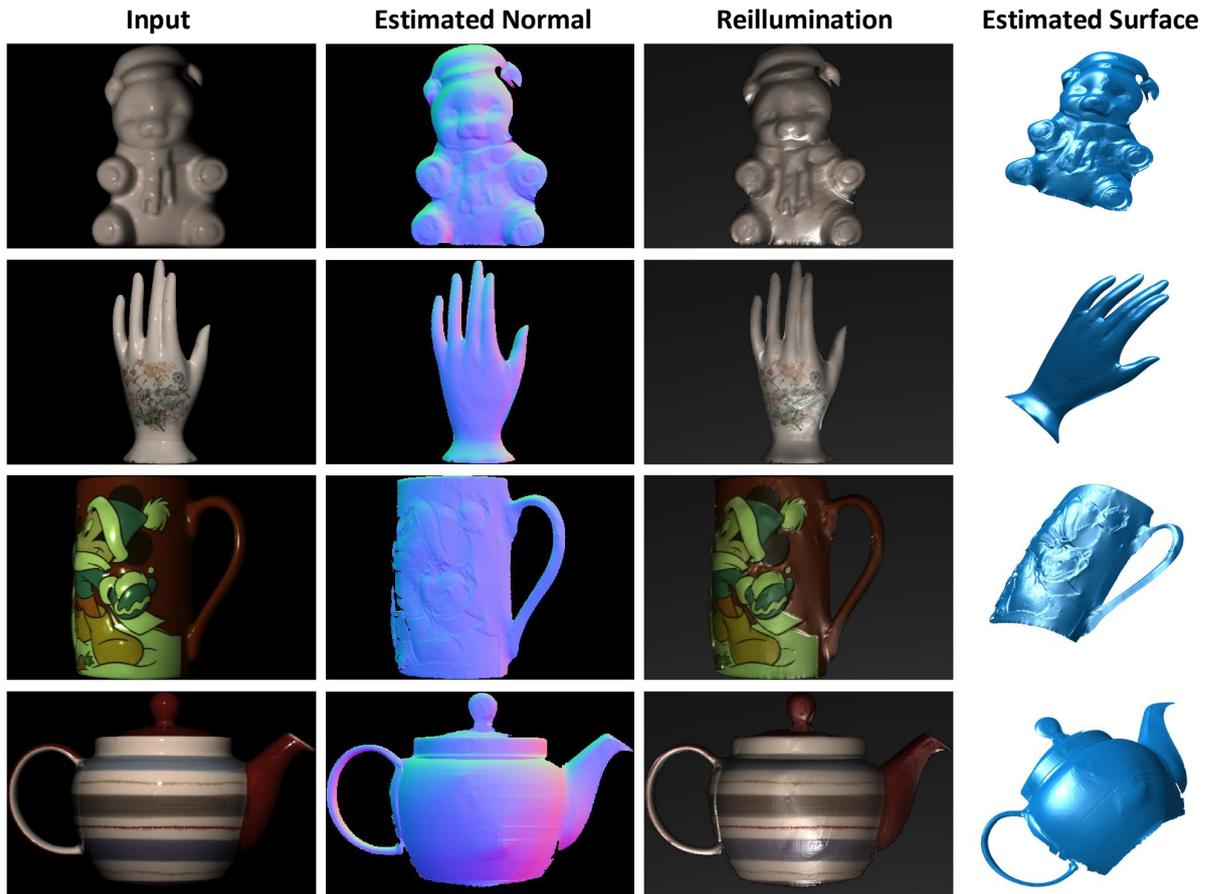


Figure 4.13: Qualitative results on real objects with varying albedo obtained by using Prop. 1+3 and comparison to [71] (zoom for detail).

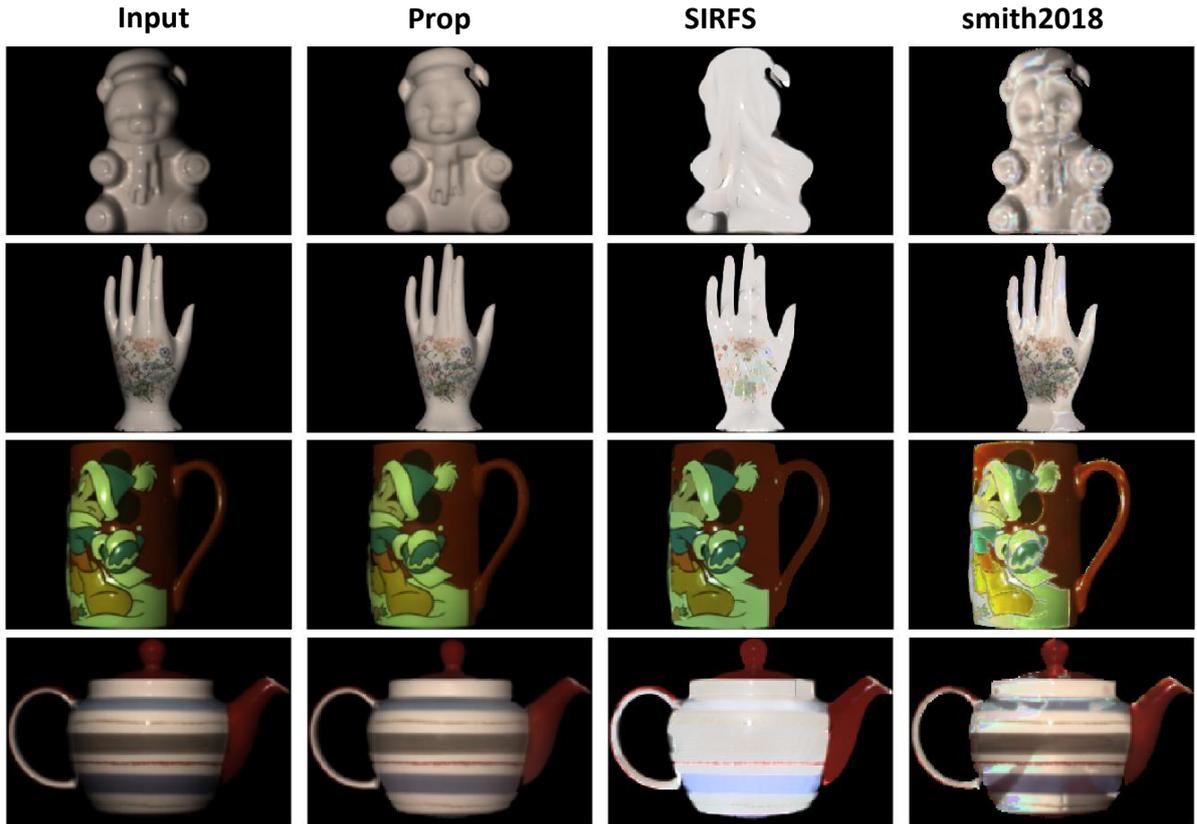


Figure 4.14: Albedo estimation on real data with different methods from left to right which are proposed, Mecca2017 [51], and Smith2018 [69].

	Snooker	Bear	Hand	Cup	Teapot	3DRFE
ours	1.591	1.543	1.423	1.522	1.602	1.583
Ground Truth	1.63 [1]	1.504 [1]				1.5

Table 4.5: We estimate the refractive index of different objects that in our dataset.

4.4 Summary

We first approach a general method that uses nonlinear least squares to obtain the surface height from polarimetric images data. This method is able to obtain a competitive shape from a planar initialisation. Moreover it is optimal *w.r.t* to an explicit noise model. For previous work [6,41,42,50,52,70] 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 *w.r.t* the assumed noise model. The second approach introduced a unifying formulation for recovering height from photo-polarimetric stereo

data and proposed a variety of methods that use different combinations of linear constraints. We also provide uncalibrated, albedo invariant shape estimation with only two light sources. This relieves several constraints that were required in previous works. [6, 70]. Both of the methods recover the surface height directly to avoid explicitly surface normal disambiguation step and integration problems.

However, these two approaches based on polarisation and shading cue do not provide any direct constraints on metric depth, only on local surface orientation. Hence, the surface recovered by these methods are globally inaccurate and subject to low frequency distortion. And the orthographic camera model assumption is practically limiting. Also we manually label the pixel is either diffuse or specular by intensity threshold which is not reliable. "In the next chapter we will introduce a second camera to recover metric depth and automatically differentiate whether the pixel is either diffuse dominant or specular dominant.

Chapter 5

Depth from a polarisation + RGB stereo pair

In this chapter, 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 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.* [70] 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.* [70] we do not assume uniform albedo. Unlike Kadambi *et al.* [42, 43], 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 [12, 17, 20, 86], we only require a single polarisation image.

Overview of method

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. 5.1 and each step is described in detail in the following sections.

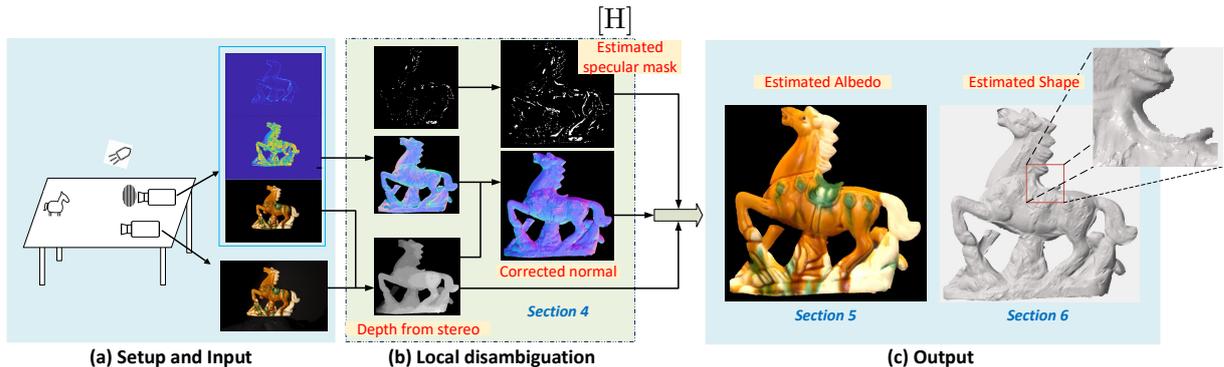


Figure 5.1: Overview: From a stereo pair of one polarisation image and one RGB image (a) we merge stereo depth with polarisation normals using a higher order graphical model (b) before estimating an albedo map and the final geometry (c).

5.1 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 (3.19), and the direction of the outward pointing surface normal is defined as the cross product

of the partial derivatives with respect to x and y in (3.20) [30] that yields:

$$\mathbf{n}(\mathbf{u}) = \begin{bmatrix} -\frac{z(\mathbf{u}) \cdot z_x(\mathbf{u})}{f_y} \\ -\frac{z(\mathbf{u}) \cdot z_y(\mathbf{u})}{f_x} \\ \frac{x-x_0}{f_x} \frac{z(\mathbf{u}) \cdot z_x(\mathbf{u})}{f_y} + \frac{y-y_0}{f_y} \frac{z(\mathbf{u}) \cdot z_y(\mathbf{u})}{f_x} + \frac{z(\mathbf{u})^2}{f_x f_y} \end{bmatrix} \quad (5.1)$$

where $z_x = \frac{\partial z(\mathbf{u})}{\partial x}$, $z_y = \frac{\partial z(\mathbf{u})}{\partial y}$ 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):

$$\mathbf{n}(\mathbf{u}) = \begin{bmatrix} -f_y z_x(\mathbf{u}) \\ -f_x z_y(\mathbf{u}) \\ (x-x_0)z_x(\mathbf{u}) + (y-y_0)z_y(\mathbf{u}) + z(\mathbf{u}) \end{bmatrix} \quad (5.2)$$

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:

$$\mathbf{v}(\mathbf{u}) = - \begin{bmatrix} \frac{x-x_0}{f_x} & \frac{y-y_0}{f_y} & 1 \end{bmatrix}^T / \left\| \begin{bmatrix} \frac{x-x_0}{f_x} & \frac{y-y_0}{f_y} & 1 \end{bmatrix} \right\|. \quad (5.3)$$

Note that this is independent of surface depth.

5.2 Integrability-based disambiguation with a higher order graphical model

Due to the pixel remains unknown to whether diffuse or specular, it restricts 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 [42, 70] 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:

$$E(\mathbf{n}(\mathbf{u})) = \sum_{\mathbf{u} \in \mathcal{V}} \Phi(\mathbf{n}(\mathbf{u})) + \sum_{(\mathbf{u}, \mathbf{v}) \in \mathcal{N}} \varphi(L(\mathbf{u}), L(\mathbf{v})) + \sum_{(\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathcal{T}} \Psi(\mathbf{n}(\mathbf{u}), \mathbf{n}(\mathbf{v}), \mathbf{n}(\mathbf{w})) \quad (5.4)$$

Here \mathcal{V} 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 [37] to compute the disparity and reconstruct a depth map with the camera matrices, as displayed in Figure 5.2(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 $\hat{\mathbf{n}}$ which are noisy as shown in Figure 5.2(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 5.2(c)). L will be subsequently updated (Figure 5.2(f)).

5.2.1 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

$$\Phi(\mathbf{n}(\mathbf{u})) = \begin{cases} k \cdot f(\mathbf{u}) & \text{if } (L(\mathbf{u}) = 1, \mathbf{n}(\mathbf{u}) \in \mathcal{D}) \text{ or } (L(\mathbf{u}) = 0, \mathbf{n}(\mathbf{u}) \in \mathcal{S}) \\ f(\mathbf{u}) & \text{if } (L(\mathbf{u}) = 0, \mathbf{n}(\mathbf{u}) \in \mathcal{D}) \text{ or } (L(\mathbf{u}) = 1, \mathbf{n}(\mathbf{u}) \in \mathcal{S}) \end{cases} \quad (5.5)$$

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

$$f(\mathbf{u}) = \exp(-\mathbf{n}(\mathbf{u}) \cdot \hat{\mathbf{n}}(\mathbf{u})). \quad (5.6)$$

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.

5.2.2 Pairwise cost

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

$$\varphi(L(\mathbf{u}), L(\mathbf{v})) = |L(\mathbf{u}) - L(\mathbf{v})|. \quad (5.7)$$

5.2.3 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 [60]. 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

$$p(\mathbf{u}) = -n_x(\mathbf{u})/n_z(\mathbf{u}) \quad \text{and} \quad q(\mathbf{u}) = -n_y(\mathbf{u})/n_z(\mathbf{u}) \quad (5.8)$$

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

$$\frac{\partial p(\mathbf{u})}{\partial y} = p(\mathbf{w}) - p(\mathbf{u}), \quad \frac{\partial q(\mathbf{u})}{\partial x} = q(\mathbf{v}) - q(\mathbf{u}) \quad (5.9)$$

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:

$$\Psi(\mathbf{n}(\mathbf{u}), \mathbf{n}(\mathbf{v}), \mathbf{n}(\mathbf{w})) = \|p(\mathbf{w}) - p(\mathbf{u}) - (q(\mathbf{v}) - q(\mathbf{u}))\|. \quad (5.10)$$

5.2.4 Graphical model optimisation

We use higher order belief-propagation to minimise (5.4) as implemented in the OpenGM toolbox [2]. 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:

$$L(\mathbf{u}) = \begin{cases} 0 & \text{if } \mathbf{n}(\mathbf{u}) \in \mathcal{D} \\ 1 & \text{if } \mathbf{n}(\mathbf{u}) \in \mathcal{S} \end{cases}. \quad (5.11)$$

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 5.4 we solve globally for depth, using the stereo depth map as a guide to remove low frequency bias.

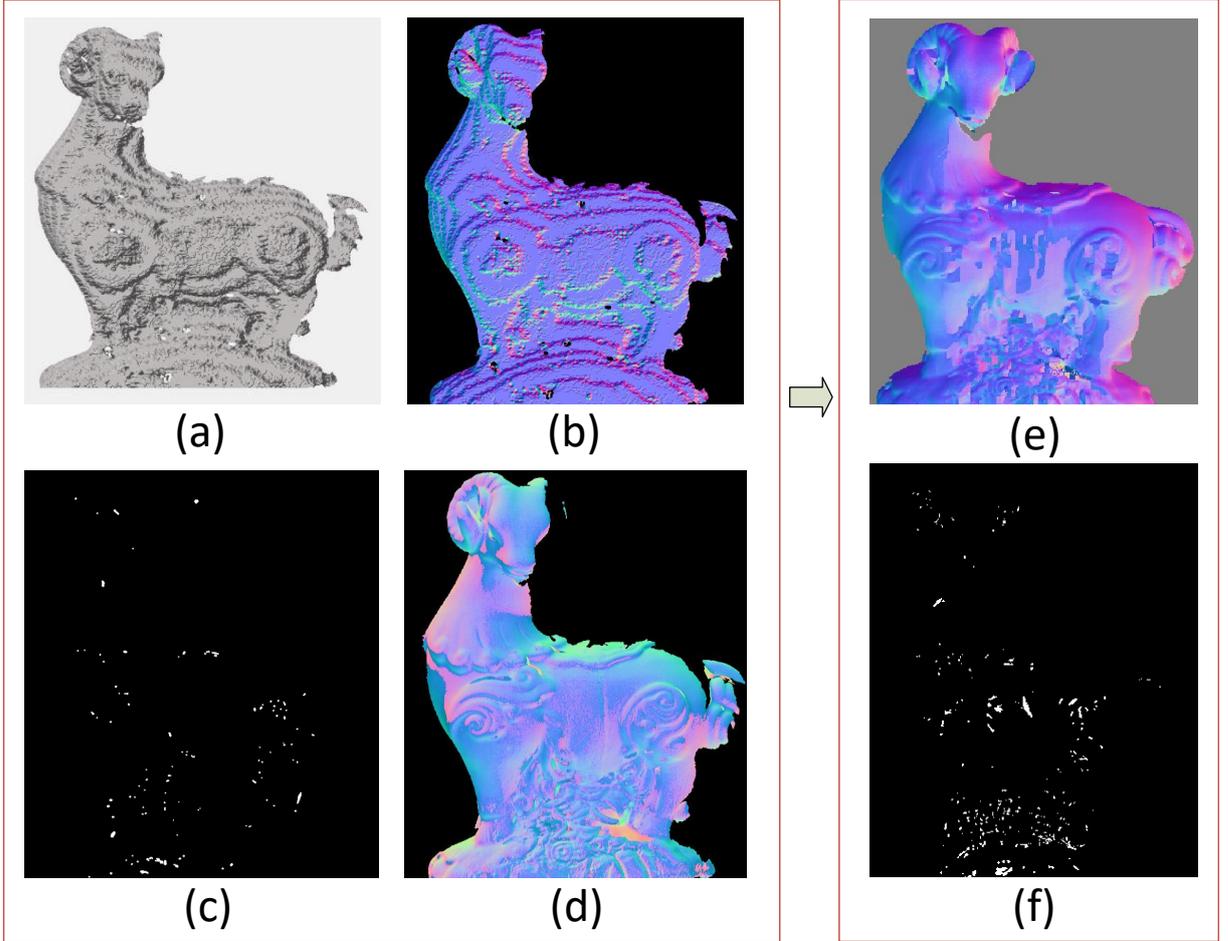


Figure 5.2: (a) Depth map from disparity map. (b) Guide surface normal from stereo depth map. (c) Preset specular mask. (d) One possible polarisation normal. (e) The corrected normal via our graphical model. (f) The updated specular mask via graphical model.

5.3 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 (4.26). 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.

The estimated albedo minimises the following energy function

$$E(\mathbf{u}) = E_{Lamb}(\mathbf{u}) + \lambda_I E_{smooth}(\mathbf{u}). \quad (5.12)$$

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

$$E_{Lamb}(\mathbf{u}) = \|a(\mathbf{u})\mathbf{n}' \cdot (\mathbf{u})\mathbf{s} - I_d(\mathbf{u})\|_2^2 \quad (5.13)$$

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

$$E_{smooth}(\mathbf{u}) = \|a(\mathbf{u}) - a(\mathbf{v}) - g(I_d(\mathbf{u}) - I_d(\mathbf{v}))\| + \|a(\mathbf{u}) - a(\mathbf{w}) - g(I_d(\mathbf{u}) - I_d(\mathbf{w}))\| \quad (5.14)$$

where $g(\cdot)$ 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 5.3 we see how the smoothness term affects the estimated albedo map and depth.

5.4 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.* [70], note that it differs via the use of the coarse depth values and optimum normal from Section 5.2. The fact that we estimate metric depth rather than relative height. As in [70], 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.

5.4.1 Phase angle constraint.

The first constraint encourages the recovered surface normal to satisfy (4.30). Following [70], the projection of the surface normal into the image plane (n_x, n_y) should be collinear with the phase angle vector. We separate 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:

$$\begin{bmatrix} \cos(\phi(\mathbf{u})) & -\sin(\phi(\mathbf{u})) & 0 \\ \cos(\phi(\mathbf{u}) + \frac{\pi}{2}) & -\sin(\phi(\mathbf{u}) + \frac{\pi}{2}) & 0 \end{bmatrix} \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = 0 \quad (5.15)$$

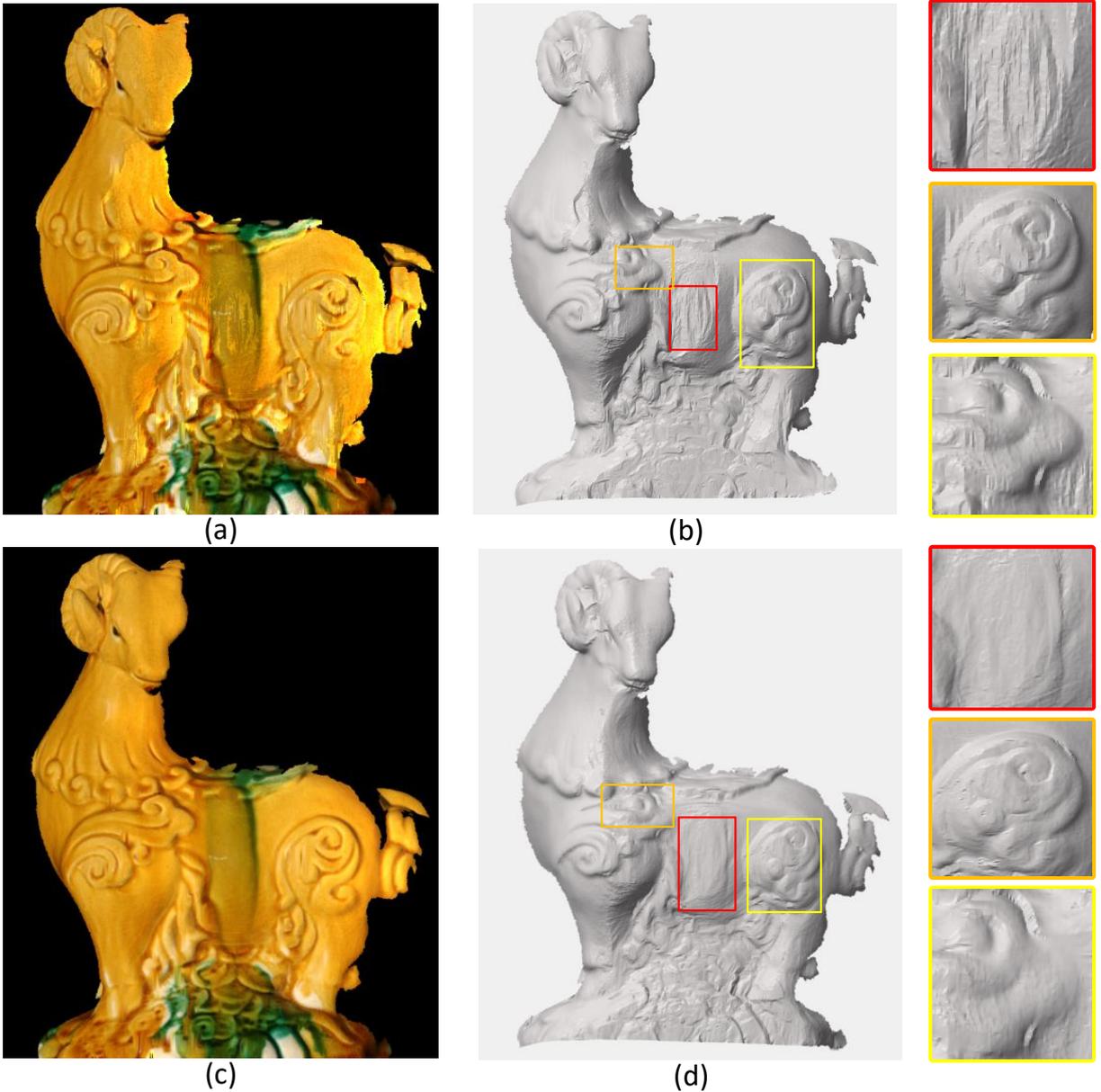


Figure 5.3: (a)/(c) Estimated albedo (b)/(d) Estimated geometry. First row: $\lambda_I = 0$, second row: $\lambda_I = 3$. Comparing (a) and (c), the albedo map becomes smoother. Comparing (b) and (d), the red rectangle region becomes smoother but while fine details are largely preserved.

5.4.2 Shading/polarisation ratio constraint.

Recall that the viewing angle is the angle between the surface normal and the viewer direction describe in (4.33). However the view direction is different in each pixel location. 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:

$$\|\mathbf{n}(\mathbf{u})\| = \frac{\mathbf{n}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u})}{\cos(\theta_r(\mathbf{u}))}. \quad (5.16)$$

Substituting this into (4.26) we obtain:

$$\frac{\mathbf{n}(\mathbf{u}) \cdot \mathbf{v}(\mathbf{u})}{\cos(\theta_r(\mathbf{u}))} = \frac{a(\mathbf{u})\mathbf{n}(\mathbf{u}) \cdot \mathbf{s}}{i_{\text{un}}(\mathbf{u})} \quad (5.17)$$

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 [70], 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:

$$\begin{bmatrix} s_x \cdot a(\mathbf{u}) \cos \theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u})v_x(\mathbf{u}) \\ s_y \cdot a(\mathbf{u}) \cos \theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u})v_y(\mathbf{u}) \\ s_z \cdot a(\mathbf{u}) \cos \theta(\mathbf{u}) - i_{\text{un}}(\mathbf{u})v_z(\mathbf{u}) \end{bmatrix}^T \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = 0 \quad (5.18)$$

5.4.3 Surface normal constraint.

We also encourage our recovered surface normal should be co-linear with the optimised normal n' from Section 5.2 where their cross product is a zero vector. It can be formalised in following manner

$$\begin{bmatrix} 0 & -n'_z(\mathbf{u}) & n'_y(\mathbf{u}) \\ n'_z(\mathbf{u}) & 0 & -n'_x(\mathbf{u}) \\ -n'_y(\mathbf{u}) & n'_x(\mathbf{u}) & 0 \end{bmatrix} \begin{bmatrix} n_x(\mathbf{u}) \\ n_y(\mathbf{u}) \\ n_z(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (5.19)$$

5.4.4 Global linear depth estimation.

The relationship between the surface normal and depth under perspective viewing is given by (5.2). We can arrive at a linear relationship between the constraints described above and the unknown depth.

We first extend (5.2) 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:

$$\mathbf{Nz} = \begin{bmatrix} n_x(\mathbf{u}_1) \\ \dots \\ n_x(\mathbf{u}_N) \\ n_y(\mathbf{u}_1) \\ \dots \\ n_y(\mathbf{u}_N) \\ n_z(\mathbf{u}_1) \\ \dots \\ n_z(\mathbf{u}_N) \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} -f_y \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -f_x \mathbf{I} & \mathbf{0} \\ \mathbf{X} & \mathbf{Y} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{I} \end{bmatrix} \quad (5.20)$$

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 Savitzky-Golay filters to compute the derivative matrix that only search the nearest neighbours where are available. Hence $\mathbf{D}_x, \mathbf{D}_y$ are sparse with values only index by nearest neighbours.

Combining (5.20) with (5.15), (5.18) and (5.19) 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 with N_S . We now form a linear system in the vector of unknown depth values, \mathbf{z} :

$$\begin{bmatrix} \lambda \mathbf{AN} \\ \mathbf{W} \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{0}_{4N+N_D} \\ z_{\text{guide}}(\mathbf{u}_1) \\ \vdots \\ z_{\text{guide}}(\mathbf{u}_N) \end{bmatrix} \quad (5.21)$$

where $z_{\text{guide}}(\mathbf{u}_i)$ are the stereo depth values from Section 5.2 and $\mathbf{W} \in \mathbb{R}^{K \times N}$ performs a sparse indices matrix of \mathbf{z} at positions $(x_1, y_1), \dots, (x_K, y_K)$. $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ 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 (5.15), (5.18) and (5.19). $\lambda > 0$ is a weight which trades off the influence of the guide depth values against satisfaction of the polarisation constraints. We then solve (5.21) in a least squares sense using sparse linear least squares.

5.5 Experimental results

We present experimental results on both synthetic and real data. We compare our method against [37, 42, 69, 70, 85], the differences are summarised in Table 5.1. We set $\lambda_I = 1, \lambda = 1$ and $t = 0.01$ through our experiments. Note that the source code for [42] is not available so we are only able to compare against a single result provided by the authors. Similarly, real image results for [85] were provided by the author running the implementation for us. Whereas [37, 69, 70] 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 5.4 (left). The texture map is from [88]. We simulate the effect of polarisation according to (2.38) 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.

	Coarse depth	Shading	Polarisation
Stereo [37]	✓		
Smith-2016 [70]		✓	✓
Smith-2018 [69]		✓	✓
Polarised 3D [42]	✓		✓
Wu-2014 [85]	✓	✓	
Proposed	✓	✓	✓

Table 5.1: Summary of the different method

Method	$\sigma = 0\%$		$\sigma = 0.5\%$		$\sigma = 1\%$	
	Depth (mm)	Normal (deg)	Depth (mm)	Normal (deg)	Depth (mm)	Normal (deg)
[37]	0.49	38.151	0.49	39.78	0.49	39.67
[70]	10.68	30.38	85.91	29.966	113.80	32.03
[69]	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 5.2: Mean absolute difference in depth and mean angular surface normal errors on synthetic data. For [69, 70] methods reconstructed the depth up to scale we compute the optimum scale to align with the ground truth depth map.

In Figure 5.4 we show the estimated albedo map of the synthetic data and compare with [69]. In Table 5.2 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 [37] and state-of-the-art polarisation methods [69, 70]. In Figure 5.5 we display the qualitative results of this experiment.

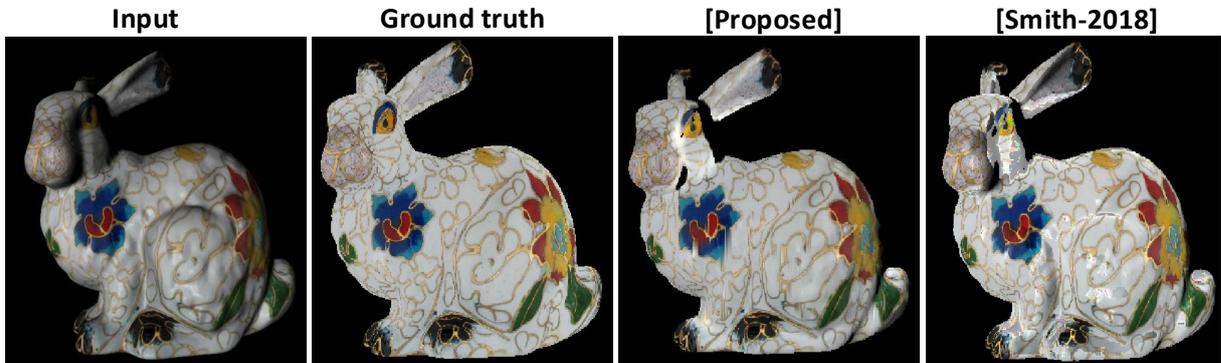


Figure 5.4: Albedo estimates on synthetic data.

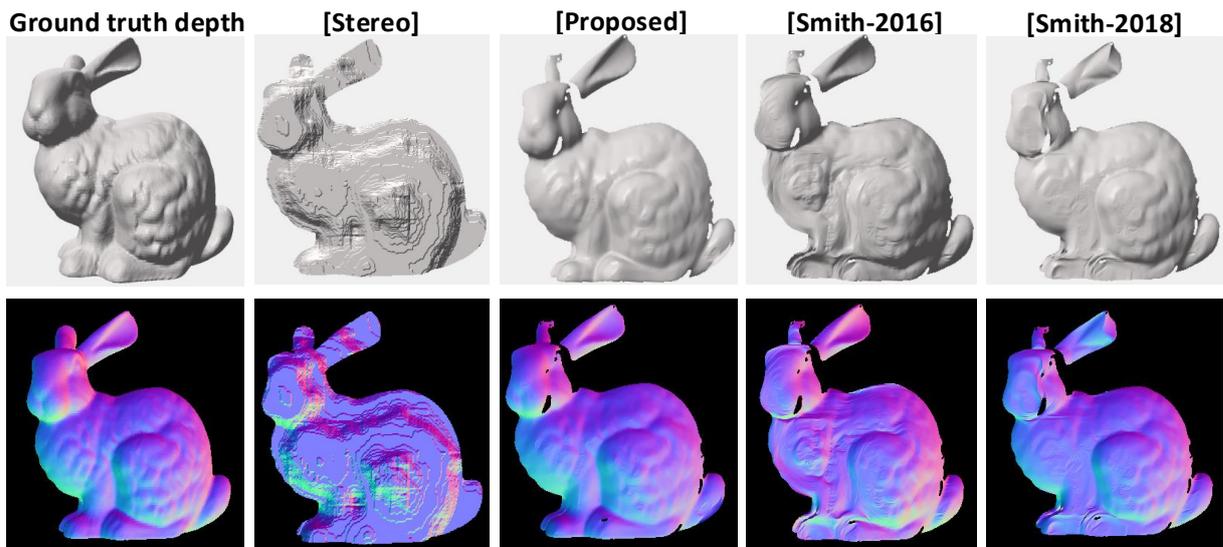


Figure 5.5: Qualitative shape estimation results on synthetic data with comparison with [70]

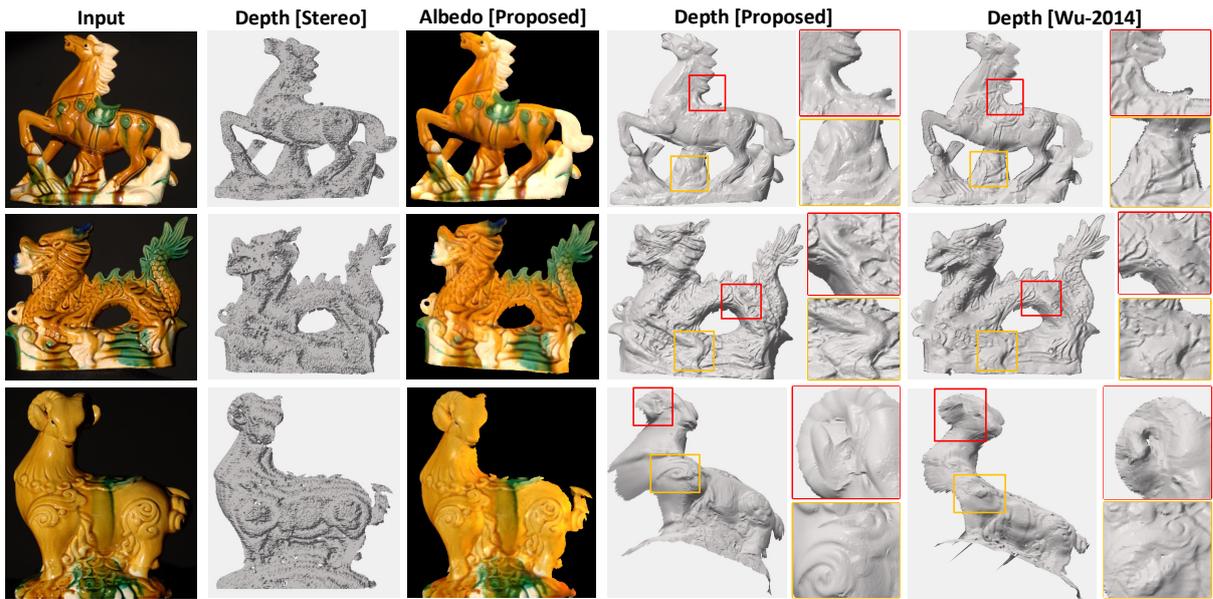


Figure 5.6: We show our results on complex object. From left to right we show an image from the input sequence; Depth from stereo reconstruction [37]; Our proposed estimated albedo map and the estimated depth. Depth estimation by [85].

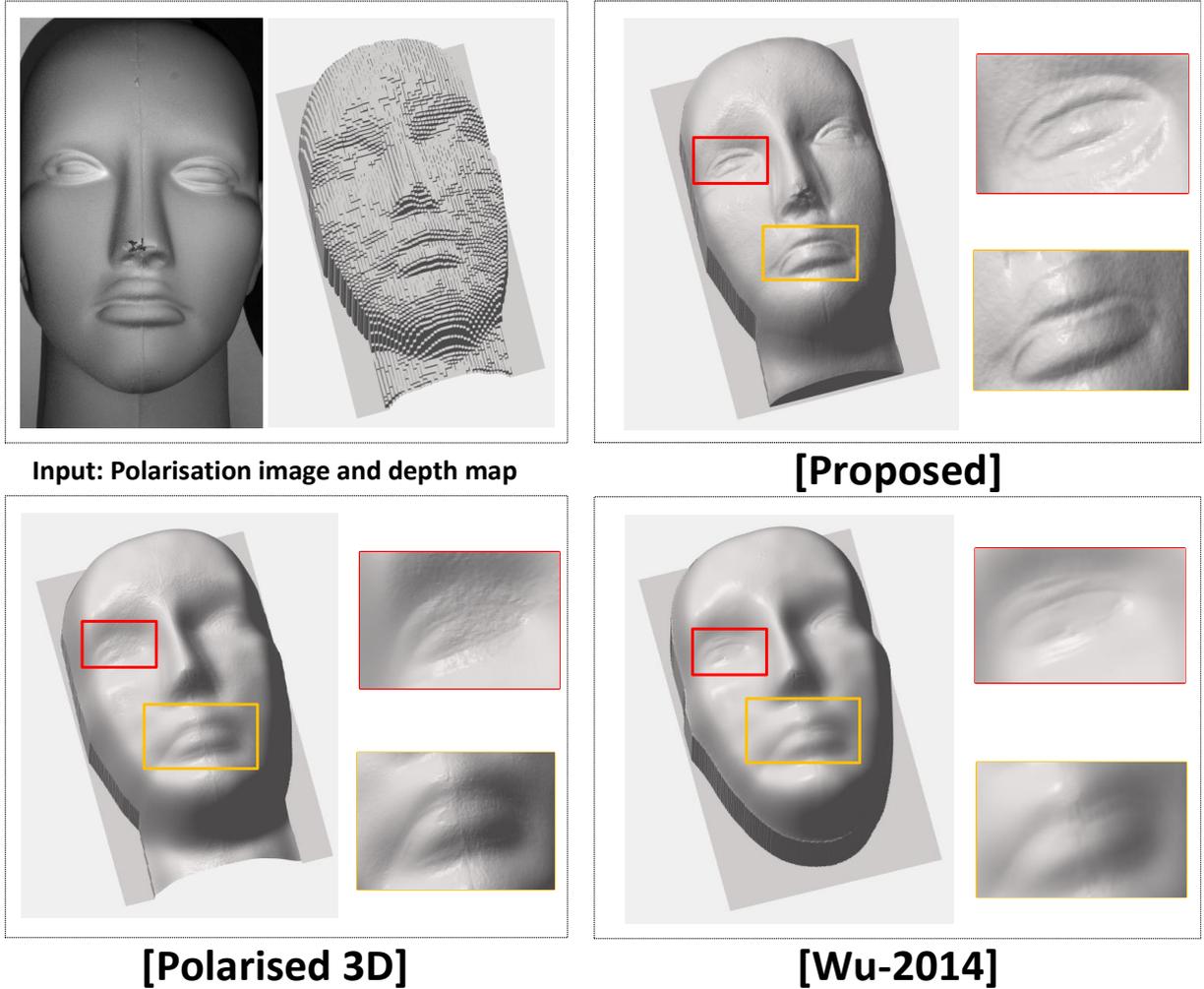


Figure 5.7: Comparison on [42] dataset. Top-left: One of the polarisation intensity images and Kinect depth map. Top-right: our result. Bottom-Left: [42]. Bottom-Right: [85].

Next we show results on a dataset of real images. The first dataset is from [42]. Although the depth here is provided by a Kinect sensor, not stereo, our graphical model optimisation in Section 5.2 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 5.7 where we show that our proposed result can give more details on the reconstruction. In this experiment, we estimate the light source direction using [70].

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 5.1. We compare our method with [85] directly performed by the author. In Figure 5.6 we show qualitative results for three objects with glossy reflectance and varying albedo. 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 [70].

5.6 Summary

We 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. This provides us with a conventional stereo cue from which we can compute coarse but metrically accurate depth estimates. It can be used to disambiguate the surface normal and label the diffuse/specular pixel by graphical model. 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 5.7. We also assume the refractive index is known in our framework. It could be potentially measured given a sufficiently accurate guide depth map as describe in Section 4.2.4. Another limitation is we assume the pixel intensity is either diffuse or specular while this approximation might introduce unknown estimation error. In particular, a more comprehensive model of mixed specular/diffuse reflectance and polarisation would be beneficial. In the next chapter we will introduce mixed polarisation model to bring up a more accurate physical model to approximate the polarisation reflectance model.

Chapter 6

Mixed polarisation model with Multi-view stereo

6.1 Mixed Polarisation Model

In this chapter we take some first steps towards a comprehensive mixed model of both specular and diffuse polarised reflectance. This model is more physically accurate and avoids many assumptions made in previous work and earlier chapters in this thesis. We present both a mixed model and a method for fitting the model to multiview data. We emphasise that this line of work is not yet complete but provides a first attempt in this direction which shows promising results.

The previous works on shape-from-polarisation heavily rely on the assumption that pixels can be classified as either diffuse dominant or specular dominant and one of the two models used per-pixel [20, 42, 70, 77, 89]. This assumption is not accurate due to light emitted from some points on the surface being a mix of reflected and refracted (subsurface scattered) light rays. The emitted ray from the surface can be modelled as a superposition of reflected ray and refracted ray as shown in Figure 6.1. The Stokes vector of reflected ray comes from incident ray \mathbf{s} can be written as (2.29): $\mathbf{s}_r = \mathbf{M}_R \cdot \mathbf{s}$. The Stokes vector of refracted ray comes from incident ray \mathbf{s}_o can be written as (2.34): $\mathbf{s}_o = \mathbf{M}_T \cdot \mathbf{s}$. So the emitted ray from the surface can be written

as $\mathbf{s}_m = \mathbf{s}_r + \mathbf{s}_t$, where

$$\mathbf{s}_m = \frac{1}{2} \begin{bmatrix} s_{0_o}(T_p + T_s) + s_0(R_p + R_s) \\ s_{0_o}(T_s - T_p) + s_0(R_s - R_p) \\ 0 \\ 0 \end{bmatrix}$$

The degree of polarisation now is $\rho = \frac{s_{0_o}(T_s - T_p) + s_0(R_s - R_p)}{s_{0_o}(T_p + T_s) + s_0(R_p + R_s)}$. After transmission through a rotated linear polariser, the Stokes vector of the ray is $\mathbf{M}_{\text{RPOL}} \cdot \mathbf{s}_m$, concretely, we separate it by linear algebra rules [74] that

$$\mathbf{M}_{\text{RPOL}} \cdot \mathbf{s}_m = \frac{1}{2} \mathbf{M}_{\text{RPOL}} \begin{bmatrix} s_0(R_p + R_s) \\ s_0(R_s - R_p) \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2} \mathbf{M}_{\text{RPOL}} \begin{bmatrix} s_{0_o}(T_p + T_s) \\ s_{0_o}(T_s - T_p) \\ 0 \\ 0 \end{bmatrix}$$

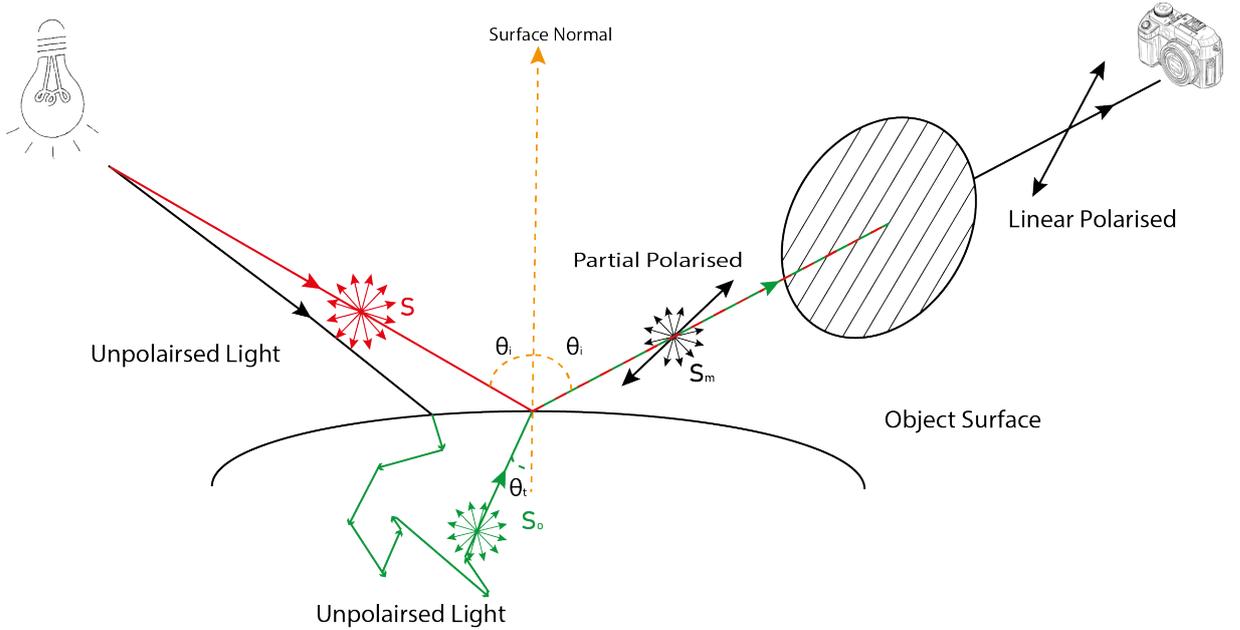


Figure 6.1: An experiment setup to capture the polarimetric images. Compare to Figure 2.4, we remove the phase retarder so that only linearly polarisation information will be captured.

As it shows the intensity captured by the camera is exactly a superposition of (2.37) and (2.32).

$$I_\vartheta = I_d(1 + \rho_d \cdot \cos(2\vartheta - 2\phi)) + I_s(1 + \rho_s \cdot \cos(2\vartheta - 2\phi + \pi))$$

Remove the π in specular component and merge into one:

$$I_\vartheta = (I_d + I_s) + (I_d\rho_d - I_s\rho_s) \cos(2\vartheta - 2\phi) \quad (6.1)$$

We can see the diffuse polarisation model is an approximation of (6.1) where $I_s = 0$, while the specular polarisation is $I_d = 0$.

The ambiguity The (6.1) now contains five unknowns that are $I_d, I_s, \rho_d, \rho_s, \phi$. Compare to (2.38), it still remains a sinusoid function so we can use the same method as described in (2.40), but now the $I_{un} = I_d + I_s$, $\rho = \frac{I_d \rho_d - I_s \rho_s}{I_d + I_s}$ and φ contains four ambiguities as described in equations (2.43), (2.44) depending on whether the pixel is specular dominant or diffuse dominant (defined below). In order to clarify the ambiguity problem in mixed polarisation model, we plot its degree of polarisation in Figure 6.2. We pick two different values for specular intensities.

In the figure (a), when $I_s = 0.01$ which means the specular intensity is small enough to be ignored, this model can be approximated as diffuse polarisation model. Otherwise when $I_s = 0.1$ the ρ can be divided into two region: a specular dominant region is when $I_s \rho_s > I_d \rho_d$, a diffuse dominant region is when $I_s \rho_s < I_d \rho_d$. We can see it in the figure (a) where $\rho < 0$ is the specular dominant region. By definition in (2.11), DoP is in the value between $[0, 1]$, so we take the absolute value from figure (a) to make a figure (b). In figure (b) we can see when the pixel intensity mixed with diffuse and specular, the zenith angle might have up to three solutions (The dashed line cross with the DoP curve in figure (b)). That is intuitive for we have two solutions in specular polarisation model and one from diffuse polarisation model. In figure (b) we can also see if we estimate the zenith angle by approximation to a diffuse polarisation model, the zenith angle estimation will be far away from the ground truth. As the dashed line intersects with the two curves showed in figure (b). Assume the curve $I_s = 0.1$ is the ground truth curve and $I_s = 0.01$ is an approximation of the diffuse polarisation model. We can see the same DoP value interpreted by diffuse polarisation model will lead its estimated zenith angle about 0.8, but the ground truth can only be one of in 0.5, 1.15, 1.25.

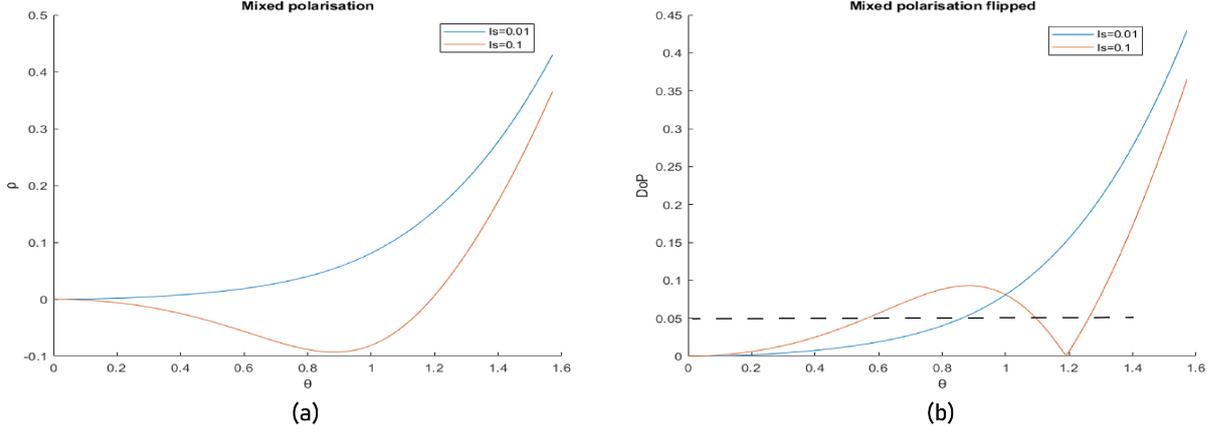


Figure 6.2: We set $I_d = 0.5$ with two different specular intensity of the mixed polarisation model. (a) The curves shows the value of $\frac{I_d \rho_d - I_s \rho_s}{I_d + I_s}$. (b) The curves shows the value of $|\frac{I_d \rho_d - I_s \rho_s}{I_d + I_s}|$.

To conclude, the surface normal ambiguity is determined by whether the pixel is diffuse or specular dominant. 1). Specular dominant is when $I_s \rho_s > I_d \rho_d$, it has four possible normals. 2). Diffuse dominant is when $I_s \rho_s < I_d \rho_d$ which contains two possible normals. This observation follows same rule as described in Section 2.1.6. To determine whether a pixel is diffuse or specular dominant we require estimates of the value of I_d, I_s, ρ_d, ρ_s .

6.2 Multi-view constraint

Although the mixed polarisation model presents the physical model more accurately, to solve the five unknowns directly is challenging. We see the captured pixels' intensity forms a sinusoidal function which only determines three unknowns as the method described in (2.40). In order to tackle this problem, we utilise the multi-view constraint. We have seen in Ch 5, an additional camera provides a coarse depth map that provides a strong constraint on surface normal. The multi-view stereo can provide an even better shape estimation [24, 25, 68] and the surface normal can be retrieved from multi-view stereo shape estimation. Unlike previous chapters that works on a depth map, we work on the 3D space, which mean the unknowns are estimated per vertex. We rewrite the equation 6.5 and adapt the surface normal in to estimate the I_d, I_s with known refractive index. We show this method provides a state-of-art diffuse and specular separation in uncontrolled illumination.

6.2.1 Polarisation model under camera coordinate

We denote a 3D point in world coordinate as $\mathbf{x} = (X, Y, Z)$ and the surface normal in world coordinate of it is denoted as $\mathbf{n}(\mathbf{x}) = [n_x(\mathbf{x}), n_y(\mathbf{x}), n_z(\mathbf{x})]^T$ with $\|\mathbf{n}(\mathbf{x})\| = 1$. For the polarisation model is measured under camera coordinate as show in Figure 2.10 and fig 2.8, we need to determined the θ, ϕ under camera coordinate. To clarify we represent them as θ_l, ϕ_l which mean in local camera coordinate. Under perspective camera model, the surface normal *w.r.t* certain camera is determined by the view direction and orientation of a camera which can be represented by the extrinsic camera parameters [35]. We denote the camera orientation with a 3×3 matrix R and camera position *w.r.t* world coordinate as C . The view direction *w.r.t* a 3D point in world coordinate is given by

$$\mathbf{v}(\mathbf{x}) = C - \mathbf{x} \quad (6.2)$$

θ_l is determined by the surface normal in world coordinate and view direction of a certain camera as show in Figure 4.2 that $\cos(\theta_l(\mathbf{x})) = \frac{\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})}{\|\mathbf{n}(\mathbf{x})\|}$, where

$$\theta_l(\mathbf{x}) = \arccos\left(\frac{\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})}{\|\mathbf{n}(\mathbf{x})\|}\right) \quad (6.3)$$

To compute ϕ_l we should transform the surface normal to camera coordinate by $\mathbf{n}_l(\mathbf{x}) = R \cdot \frac{\mathbf{n}(\mathbf{x})}{\|\mathbf{n}(\mathbf{x})\|}$, then we get $\tan(\phi_l(\mathbf{x})) = \frac{n_{l,y}(\mathbf{x})}{n_{l,x}(\mathbf{x})}$. So we have

$$\phi_l(\mathbf{x}) = \text{atan2}\left(\frac{n_{l,y}(\mathbf{x})}{n_{l,x}(\mathbf{x})}\right) \quad (6.4)$$

In equations (2.44) and (2.43), the degree of polarisation(DoP) depends on surface geometry and refractive index. Here we assume the refractive index is known. Instead using two separate value ρ_s, ρ_d , we can rewrite the diffuse and specular DoP *w.r.t* θ_l as $\rho_s(\theta_l)$ and $\rho_d(\theta_l)$, and the equation 6.1 can be presented as

$$I_{\vartheta} = (I_d + I_s) + (I_d \rho_d(\theta_l) - I_s \rho_s(\theta_l)) \cos(2\varphi - 2\phi_l) \quad (6.5)$$

6.2.2 Reformulation under Multi-view stereo

Assume we have M polarisation images capture by M polarisation cameras with different views. The extrinsic camera parameters (R^k, C^k) denote the orientation and location of the k^{th} camera. We can see the minimum parameters of (6.5) constitutes four unknowns which are: $I_d, I_s, \theta_l, \phi_l$. A sequence of polarimetric image data by one camera can only provide 3 linear independent equations which is not enough to solve 4 unknowns. By adding more views we can add more

equations in order to solve it. Concretely, under multi-view constraint, where the polarimetric images from different view are all captured simultaneously. By the assumption of Lambertian reflectance model as (4.26), the diffuse intensity is independent on view direction, so we expect a vertex in image space will have same diffuse intensity but different specular intensity under different views. Where $I_d^1(\mathbf{x}) = I_d^2(\mathbf{x}) = \dots = I_d^N(\mathbf{x})$ and $I_s^1(\mathbf{x}) \neq I_s^2(\mathbf{x}) \neq \dots \neq I_s^N(\mathbf{x})$. $\theta_l(\mathbf{x}), \phi_l(\mathbf{x})$ are related to surface normal at point \mathbf{x} that transformed to the camera coordinate. Explicitly, for $\theta_l^k(\mathbf{x}), \phi_l^k(\mathbf{x})$ in k^{th} camera,

$$\theta_l^k(\mathbf{x}) = \arccos\left(\frac{\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}^k(\mathbf{x})}{\|\mathbf{n}(\mathbf{x})\|}\right) \quad \text{where } \mathbf{v}^k(\mathbf{x}) = C^k - \mathbf{x} \quad (6.6)$$

$$\phi_l^k(\mathbf{x}) = \arctan\left(\frac{n_{l,y}^k(\mathbf{x})}{n_{l,x}^k(\mathbf{x})}\right) \quad \text{where } \mathbf{n}_l^k(\mathbf{x}) = R^k \cdot \frac{\mathbf{n}(\mathbf{x})}{\|\mathbf{n}(\mathbf{x})\|} \quad (6.7)$$

Rather use local coordinate to present (6.5), we use a global surface normal $\mathbf{n}(\mathbf{x})$ with camera parameters to replace θ_l, ϕ_l . Moreover, the surface normal can be rewritten with spherical coordinate where

$$\mathbf{n}(\mathbf{x}) = \begin{bmatrix} \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \\ \cos(\theta) \end{bmatrix} \quad (6.8)$$

The θ, ϕ are the zenith angle and azimuth angle of the surface normal in world coordinate respectively, so a minimum parametrisation of (6.5) in k^{th} can be written as

$$I_{\theta}^k(I_d, I_s^k, \theta, \phi) = (I_d + I_s^k) + (I_d \rho_d(\theta_l^k(\theta)) - I_s \rho_s(\theta_l^k(\theta))) \cos(2\varphi - 2\phi_l^k(\phi)) \quad (6.9)$$

Where we drop \mathbf{x} for simplicity. With M views we have total $M + 3$ unknowns in the mean while we $3M$ linear independent equations. Once $M > 2$ the unknowns can be solved.

6.3 Estimation of mixed polarisation model

The multi-view stereo method provides a dense shape estimation and camera parameter estimation simultaneously [24, 25, 35]. This provides us a 3D point cloud of the object, surface normal of each vertex and intrinsic/extrinsic parameters of all cameras. We will utilise these information to estimate the mixed polarisation model.

6.3.1 Extract correspondences in image space

A key step to utilise the multi-view constraint is to locate the correspondences in image space across all camera of a vertex \mathbf{x} . Rather search in the image space, we utilise the point cloud

reconstructed by multi-view stereo and back project the point with camera parameters into image space [35]. The intensities of a vertex \mathbf{x} in all cameras can be written as $I_{\vartheta}^1(\mathbf{x}), \dots, I_{\vartheta}^M(\mathbf{x})$, these intensities across image spaces are related to the same vertex in 3D space. The projections may not be to integer positions within the image, in which case we can use bilinear interpolation into the polarimetric images.

6.3.2 Optimisation strategy

The observed image intensity of a certain 3D point \mathbf{x} in k^{th} camera with polariser orientation ϑ_j is denoted as $I_{\vartheta_j}^{\text{obs},k}$. The corresponding intensity from mixed polarisation model is denoted as $I_{\vartheta_j}^{\text{mod},k}$ where we drop \mathbf{x} for simplicity. Now we can compute a vector of residual between the observed intensity and theoretical intensity through all polariser angle and all view, a solution can be computed by nonlinear least squares as described in Section 4.1.2. We find a better way to solve this equation by using (alternating optimisation). Specifically, we a). fix θ, ϕ that solve linear for diffuse intensity and specular intensities for all views. b). Then fix diffuse intensity and specular intensities to find an optimum θ, ϕ that minimise the residual between observed intensity and theoretical intensity compute from mixed polarisation model.

Solve diffuse and specular intensity We first factor out the diffuse intensity and specular intensity from (6.9) that

$$\begin{bmatrix} \mathbf{A}^k & \mathbf{B}^k \end{bmatrix} \begin{bmatrix} I_d \\ I_s^k \end{bmatrix} = \mathbf{I}^k \quad (6.10)$$

where $\mathbf{A}^k \in \mathbb{R}^N, \mathbf{B}^k \in \mathbb{R}^N$ are

$$\mathbf{A}^k = \begin{bmatrix} 1 + \rho_d(\theta_l^k(\theta))\cos(2\vartheta_1 - 2\phi_l^k(\phi)) \\ \vdots \\ 1 + \rho_d(\theta_l^k(\theta))\cos(2\vartheta_N - 2\phi_l^k(\phi)) \end{bmatrix}, \quad \mathbf{B}^k = \begin{bmatrix} 1 - \rho_s(\theta_l^k(\theta))\cos(2\vartheta_1 - 2\phi_l^k(\phi)) \\ \vdots \\ 1 - \rho_s(\theta_l^k(\theta))\cos(2\vartheta_N - 2\phi_l^k(\phi)) \end{bmatrix}$$

$\mathbf{I}^k \in \mathbb{R}^N$ is given by

$$\mathbf{I}^k = \begin{bmatrix} I_{\vartheta_1}^k \\ \vdots \\ I_{\vartheta_N}^k \end{bmatrix}$$

To solve the diffuse intensity and specular intensities through all M views, we now combine all the corresponding image intensity for a certain point the we solve:

$$\min_{I_d, I_s^1, \dots, I_s^M} \|\mathbf{A}\mathbf{B} \cdot \begin{bmatrix} I_d & I_s^1 & \dots & I_s^M \end{bmatrix}^T - \mathbf{o}\|^2 \quad (6.11)$$

where $\mathbf{AB} \in \mathbb{R}^{NM \times (M+1)}$, $\mathbf{o} \in \mathbb{R}^{NM}$ are

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}^1 & \mathbf{B}^1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{A}^2 & \mathbf{0} & \mathbf{B}^2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^M & \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}^M \end{bmatrix}, \quad \mathbf{o} = \begin{bmatrix} \mathbf{I}^1 \\ \mathbf{I}^2 \\ \vdots \\ \mathbf{I}^M \end{bmatrix} \quad (6.12)$$

These equation will be solved with monochromatic images, it can be easily adapt to chromatic images by stack all colour channels.

Solve θ and ϕ We now fix the diffuse intensity and specular intensity, we substitute them to (6.9) and compute a vector of residuals $\mathbf{r}^k \in \mathbb{R}^N$ that

$$\mathbf{r}^k(\theta, \phi) = \begin{bmatrix} I_{\vartheta_1}^{\text{obs},k} - I_{\vartheta_1}^{\text{mod},k}(\theta, \phi) \\ \vdots \\ I_{\vartheta_N}^{\text{obs},k} - I_{\vartheta_N}^{\text{mod},k}(\theta, \phi) \end{bmatrix} \quad (6.13)$$

We can get θ, ϕ by solving the following nonlinear least square problem

$$\min_{\theta, \phi} \sum_{k=1}^M (\mathbf{r}^k(\theta, \phi))^T \mathbf{r}^k(\theta, \phi) \quad (6.14)$$

We initialise θ, ϕ by using the mesh reconstructed from multiple-view stereo and camera parameters as in [25] to estimate the diffuse intensity I_d and specular intensities $I_s^1 \dots I_s^M$, where we put a constraint that $I_d \leq 0, I_s^1 \leq 0, \dots, I_s^M \leq 0$. We then use the estimated diffuse and specular intensities to refine θ, ϕ by nonlinear least square approach described in Section 4.1.2. We interleave and alternate the two steps until convergence. We present experimental results on both synthetic data and real data. For this is ongoing research, we only show diffuse and specular separation results.

6.4 Summary

We propose and derive a mixed polarisation model in this chapter in order to improve the previous method that only consider either diffuse or specular in pixel domain. We utilise the multi-view prior to initialise the mixed polarisation model and achieving diffuse and specular polarisation by interleaving between linear least square and nonlinear least square. We also show the experiments results on synthetic and real captured polarimetric images, it gives a state-of-art result.

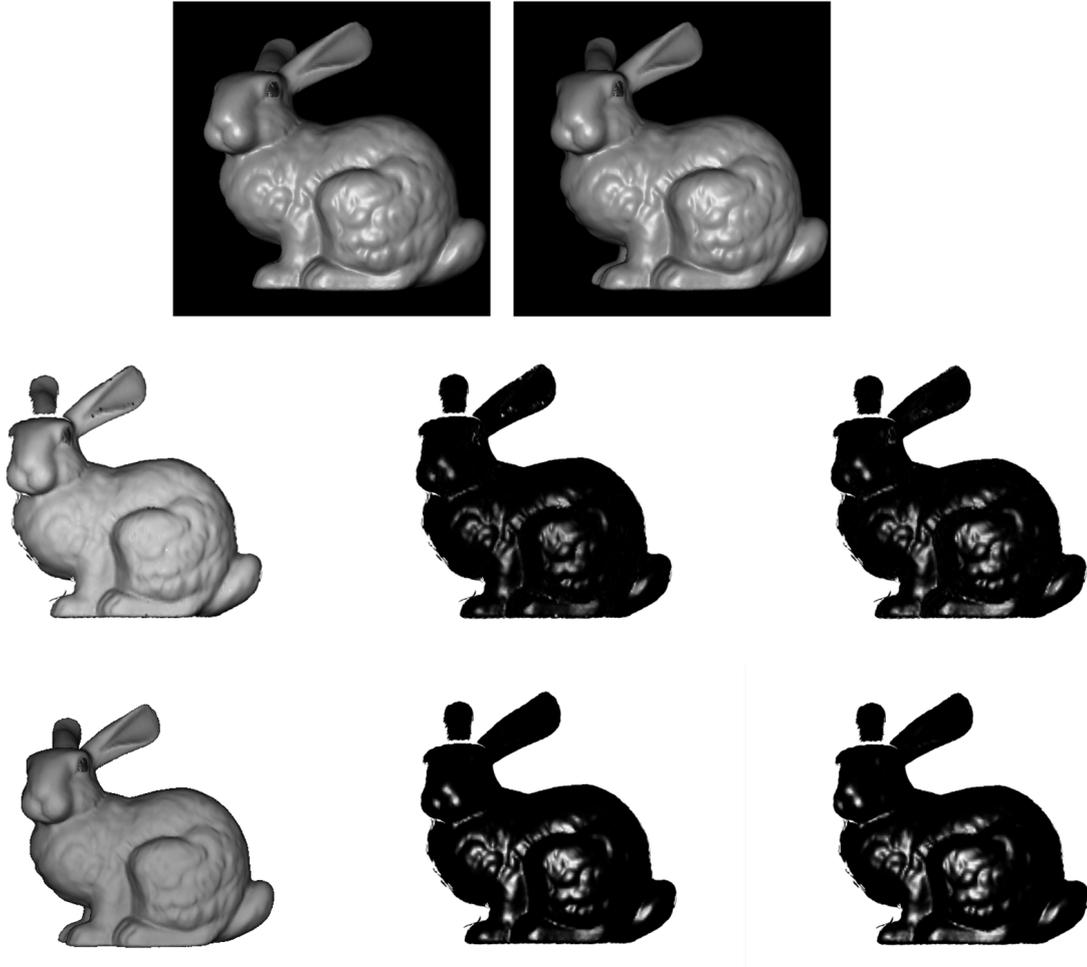


Figure 6.3: We simulate the Stanford bunny in two different views, we add noise to the ground truth mesh which provides a initialisation for θ and ϕ . The first row shows two view images. The second row shows our diffuse intensity, and specular intensities in two views(from left to right). The third row shows the ground truth of the diffuse and specular intensities.

The limitation of this method is we assume the refractive index is known, and multi-view might not work on featureless object. The 3D sensor now are more mature, that we can utilise 3D camera like Kinect instead of relying multi-view stereo, so we can extend our method to featureless object or scene as well. For this is on going research, we need to make more effort to relieve the refractive index and how to refine the surface normal.



Figure 6.4: We shows real data images that capture under natural environment, we capture each images in four different views. Column a).Original images captured under nature illumination. b).Diffuse intensity per vertex. c).Specular intensity per vertex.

Chapter 7

Conclusion and Future work

7.1 Summary

This thesis develops several methods based on shape-from-polarisation. In Chapter 3, we develop two underpinning methods serving for shape-from-polarisation in later chapters. 1) Most of the shape-from-polarisation methods rely on the quality of estimated polarisation images which includes *unpolarised intensity image*, *degree of polarisation image* and *phase angle image*. By assuming the refractive index does not vary with wavelength we can constrain the degree of polarisation and phase angle to be only geometry dependent. We therefore propose a multichannel polarisation images estimation method that fully utilises all colour channels to estimate a polarisation image. This significantly improves the quality of the estimation results. We then extend it to a multi-light environment in Section 3.2 that fixes the camera, the polarimetric images dataset are captured under different illumination conditions. This provides more constraints to improve the estimation results. 2) We explore the height-from-normal method, especially using Savitzky-Golay filters. The proposed 2D Savitzky-Golay kernel can provide a robust numerical derivation approximation. Moreover compared to Laplacian smoothness regulariser, it can de-noise the height reconstruction without over-flattening the surface.

In Chapter 4, we propose two novel shape estimation methods for a monocular polarisation camera. In Section 4.1 we propose an energy minimisation approach and optimise a nonlinear least squares cost that estimate surface height directly. This avoids polarisation image estimation and normal disambiguation steps, allowing us to optimise the depth directly. The second approach in Section 4.2 described a photo-polarimetric stereo method. The key advantage is

that the method works well on non-uniform albedo objects while previous methods are mainly focussed on uniform albedo objects. Moreover, it provides a way to estimate the refractive index of the object while previous methods make an assumption of fixed or known refractive index. We developed different constraints in Section 4.2.1 which allow us to directly estimate the height by linear least squares. These two approaches, limited under the orthographic camera model, do not provide metric depth estimation, only local surface orientation or relative height.

In Chapter 5, we propose a polarisation camera + RGB camera setup. We move on from orthographic camera model to perspective model, which builds a linear system in which metric depth can be solved by linear least squares. The stereo setup allows us to get a coarse depth map that is used for disambiguating the surface normal. We construct a higher-order graphical model to disambiguate the surface normal and label whether the pixel belongs to diffuse or specular dominant reflectance. We then estimate the albedo by the shading cue with corrected normal and finally obtain a metric depth with all these constraints. We show a state-of-art quality results that obtain a dense and accurate metric depth estimation.

Chapter 6 described a mixed polarisation model. The previous polarisation methods only consider the dominant component on the pixel that assume a pixel is either diffuse or specular dominant. We propose a multi-view stereo setup to fit and solve the mixed polarisation model. This delivers a way on separating diffuse, specular intensities and estimation of polarisation images in a single model. This method makes a first step towards a comprehensive mixed model of both specular and diffuse polarised reflectance which shows promising results.

7.2 Overarching conclusions

Taking the work in this thesis as a whole, we can draw a number of quite general, overarching conclusions that may be useful as a guide to future research directions. In this thesis, we propose different approaches to explore the methods of shape-from-polarisation. Especially we enhance the existing methods by polarisation cue. Our photo-polarimetric stereo method using two light source and polarisation cue allow us to reconstruct object with non-uniform albedo and unknown refractive index. While the previous work makes strong assumptions of known refractive index and uniform albedo [6, 70]. We combine stereo and polarisation, find out the stereo can be in used to correct the ambiguity from polarisation cue. On the other hand, enhancing by the polarisation cue which provides a high frequency signal that allow us to reconstruct a dense and detail metric depth. We also find out the our mixed polarisation model and proposed multi-view

solution gives a general approach to solve the polarisation images, while previous methods has to make assumption that each pixel is either diffuse or specular. And we show promising results in real data. We summarise our key conclusions as follows:

- **Polarisation always helps.** In this thesis we have added polarisation information to classical shape-from-x problems including shape-from-shading (Section 4.1), (2 source) photometric stereo (Section 4.2) and binocular stereo (Section 5). In each case, the geometric information we are able to recover is significantly better than using the other cue alone, in some cases making an ill-posed problem well posed. Specifically, polarisation helps recovery of finescale surface detail and resolves local ambiguities due to the phase information.
- **Shape from a single polarisation image alone is saturated.** We may have reached a point where improving the shape we can recover from a single polarisation image is highly challenging, i.e. performance on this task has saturated. Adding a second light source aids polarisation image recovery (Section 3.2) and photo-polarimetric shape estimation (Section 4.2). Adding a second viewpoint aids shape recovery (Section 5) and diffuse/specular separation (Section 6). This may be the most promising direction for future work.
- **We need to look beyond classical optimisation-based approaches.** In this thesis we focussed on expressing shape-from-polarisation as an optimisation problem that can be solved using classical methods. These include linear least squares (Sections 4.2.3, 5.4), nonlinear least squares (Section 4.1) and graphical models (Section 5.2). In all of these approaches, we only make use of the information in a single dataset (monocular or multiview images of a single object) and rely on our physical models to provide sufficient constraint. There is likely to be a significant performance gain by exploiting the power of statistical learning in modern deep learning methods. By training a deep network on large datasets, we can begin to exploit contextual cues that arise from understanding the distribution of commonly encountered local and global shape structures. Recent work [9, 44] shows that problems that are highly ill-posed for a single dataset (e.g. monocular depth estimation, inverse rendering, intrinsic image decomposition etc) can be robustly solved by training on large, representative datasets.

7.3 Future work

Mixed polarisation model We propose a multi-view stereo approach to solve the mixed polarisation model while assuming the refractive index is known. We do show in experimental results that the proposed model can achieve a good separation on diffuse and specular intensity. In the future we will combine 3D camera with multiple polarisation cameras, for depth camera can work on featureless object and provide a coarse metric depth directly. And multiple polarisation cameras can constrain the mixed polarisation model, we will adapt patch-based reconstruction [25] method to refine the shape estimation. We believe such a setup has the possibility to extend state-of-the-art polarised multiview stereo [27] to a single shot setup not requiring polarised light.

Deep neural network with polarisation We have shown a nonlinear approach to obtain the shape from polarimetric image data. Similarly we can build a deep network and use the polarisation reflectance model as constraints to optimise the network that given the polarimetric image data as input and output a predict depth map. The key idea is that the nonlinear optimisation for a single image is highly challenging with many local minima. Exploiting large datasets, the objective can be optimised in aggregate and using stochastic methods such that data-driven statistical priors are implicitly learnt to combine with the polarisation cue. Similar successes on related tasks such as monodepth estimation and inverse rendering suggest that such an approach is likely to be successful.

Loosen the constraints on the incident light Almost all previous works assume the incident light is unpolarised light, and the diffuse, specular polarisation reflectance models are derived from that. But actually the light could be in any polarisation state. We would like to explore more about how different polarised incident light will affect polarisation reflectance model. This will help by using the light stage, that we can control the light source polarisation states. In a similar direction, we have largely assumed a single point light source. Generalising to arbitrary environment lighting conditions would extend our methods into conditions experienced outside of a lab setting.

Bibliography

- [1] *Material index of refraction*, 2004 (accessed November 25, 2020).
- [2] B. Andres, T. Beier, and J. H. Kappes. Opengm: A c++ library for discrete graphical models. *arXiv preprint arXiv:1206.0111*, 2012.
- [3] G. Atkinson and E. R. Hancock. Recovery of surface height using polarization from two views. In *CAIP*, pages 162–170. Springer, 2005.
- [4] G. A. Atkinson. Polarisation photometric stereo. *Comput. Vis. Image Underst.*, 2017.
- [5] G. A. Atkinson and E. R. Hancock. Multi-view surface reconstruction using polarization. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 309–316. IEEE, 2005.
- [6] G. A. Atkinson and E. R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on image Processing*, 15(6):1653–1664, 2006.
- [7] G. A. Atkinson and E. R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image processing*, 15(6):1653–1664, 2006.
- [8] G. A. Atkinson and E. R. Hancock. Shape estimation using polarization and shading from two views. *IEEE transactions on pattern analysis and machine intelligence*, 29(11), 2007.
- [9] Y. Ba, R. Chen, Y. Wang, L. Yan, B. Shi, and A. Kadambi. Physics-based neural networks for shape from polarization. *arXiv preprint arXiv:1903.10210*, 2019.
- [10] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014.
- [11] E. Bartholinus and W. Brandt. *Experiments with the double refracting Iceland crystal: which led to the discovery of a marvelous and strange refraction*. 1959.

- [12] 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.
- [13] J. F. Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198. ACM, 1977.
- [14] J. F. Blinn. Models of light reflection for computer synthesized pictures. *Computer Graphics*, 11(2):192–198, 1977.
- [15] M. Born and E. Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.
- [16] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *European Conference on Computer Vision*, pages 766–779. Springer, 2008.
- [17] L. Chen, Y. Zheng, A. Subpa-asa, and I. Sato. Polarimetric three-view geometry. In *Proc. ECCV*, pages 20–36, 2018.
- [18] E. Collett. *Field guide to polarization*. 2005.
- [19] Z. Cui, J. Gu, B. Shi, P. Tan, and J. Kautz. Polarimetric multi-view stereo.
- [20] Z. Cui, J. Gu, B. Shi, P. Tan, and J. Kautz. Polarimetric multi-view stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1558–1567, 2017.
- [21] Z. Cui, V. Larsson, and M. Pollefeys. Polarimetric relative pose estimation. In *Proc. ICCV*, pages 2671–2680, 2019.
- [22] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.
- [23] S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden markov model: Analysis and applications. *Machine learning*, 32(1):41–62, 1998.
- [24] Y. Furukawa, C. Hernández, et al. Multi-view stereo: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1-2):1–148, 2015.
- [25] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2009.

- [26] M. Gales, S. Young, et al. The application of hidden markov models in speech recognition. *Foundations and Trends® in Signal Processing*, 1(3):195–304, 2008.
- [27] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)*, 30(6):1–10, 2011.
- [28] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2402–2409. IEEE, 2006.
- [29] R. C. Gonzales and R. E. Woods. *Digital image processing*, 2002.
- [30] G. Graber, J. Balzer, S. Soatto, and T. Pock. Efficient minimal-surface regularization of perspective depth maps in variational stereo. In *Proc. CVPR*, pages 511–520, 2015.
- [31] M. Harker and P. O’Leary. Least squares surface reconstruction from measured gradient fields. In *Proc. CVPR*, 2008.
- [32] M. Harker and P. O’Leary. Least squares surface reconstruction from gradients: Direct algebraic methods with spectral, tikhonov, and constrained regularization. In *Proc. CVPR*, pages 2529–2536, 2011.
- [33] M. Harker and P. O’Leary. Direct regularized surface reconstruction from gradients for industrial photometric stereo. *Computers in Industry*, 64(9):1221–1228, 2013.
- [34] M. Harker and P. O’Leary. Regularized reconstruction of a surface from its measured gradient field. *Journal of Mathematical Imaging and Vision*, 51(1):46–70, 2015.
- [35] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [36] E. Hecht. *Optics*. Pearson Education India, 2012.
- [37] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. CVPR*, volume 2, pages 807–814. IEEE, 2005.
- [38] C. Hoffman and R. Driggers. *Encyclopedia of Optical and Photonic Engineering, Second Edition*. CRC Press 2015, 2015.
- [39] X. Hu and P. Mordohai. A quantitative evaluation of confidence measures for stereo vision. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2121–2133, 2012.

- [40] C. P. Huynh, A. Robles-Kelly, and E. Hancock. Shape and refractive index recovery from single-view polarisation images. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1229–1236. IEEE, 2010.
- [41] 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.
- [42] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3370–3378, 2015.
- [43] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Depth sensing using geometrically constrained polarization normals. *Int. J. Comput. Vision*, 2017.
- [44] A. Kalra, V. Taamazyan, S. K. Rao, K. Venkataraman, R. Raskar, and A. Kadambi. Deep polarization cues for transparent object segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8602–8611, 2020.
- [45] R. Kitamura, L. Pilon, and M. Jonasz. Optical constants of silica glass from extreme ultraviolet to far infrared at near room temperature. *Applied optics*, 46(33):8118–8133, 2007.
- [46] D. Koller, N. Friedman, and F. Bach. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [47] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 508–515. IEEE, 2001.
- [48] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2):147–159, 2004.
- [49] K. Koshikawa. A polarimetric approach to shape understanding of glossy objects. In *Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1*, pages 493–495, 1979.
- [50] 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.
- [51] R. Mecca, F. Logothetis, and R. Cipolla. A differential approach to shape from polarization. In *Proc. BMVC*, 2017.

- [52] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Polarization-based transparent surface modeling from two views. In *null*, page 1381. IEEE, 2003.
- [53] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):73–82, 2004.
- [54] D. Miyazaki, R. T. Tan, K. Hara, and K. Ikeuchi. Polarization-based inverse rendering from a single view. In *null*, page 982. IEEE, 2003.
- [55] S. G. Narasimhan, V. Ramesh, and S. Nayar. A class of photometric invariants: Separating material from shape and illumination. In *Proc. ICCV*, volume 2, pages 1387 – 1394, 2003.
- [56] N. M. Nasrabadi. Pattern recognition and machine learning. *Journal of electronic imaging*, 16(4):049901, 2007.
- [57] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *ACM transactions on graphics (TOG)*, volume 24, pages 536–543. ACM, 2005.
- [58] T. Ngo Thanh, H. Nagahara, and R.-i. Taniguchi. Shape and light directions from shading and polarization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2310–2318, 2015.
- [59] J. Peatross and M. Ware. *Physics of light and optics*. Brigham Young University, Department of Physics, 2011.
- [60] N. Petrovic, I. Cohen, B. J. Frey, R. Koetter, and T. S. Huang. Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models. In *Proc. CVPR*, volume 1, pages I–I. IEEE, 2001.
- [61] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 60(4):576–593, 2018.
- [62] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Variational methods for normal integration. *Journal of Mathematical Imaging and Vision*, 60(4):609–632, 2018.
- [63] Y. Quéau, R. Mecca, J.-D. Durou, and X. Descombes. Photometric stereo with only two images: A theoretical study and numerical resolution. *Image and Vision Computing*, 57:175–191, 2017.
- [64] S. Rahmann and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proc. CVPR*, volume 1, pages I–I, 2001.

- [65] A. Robles-Kelly and E. R. Hancock. A graph-spectral method for surface height recovery from needle-maps. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.
- [66] G. Saman and E. Hancock. Refractive index estimation using photometric stereo. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1925–1928. IEEE, 2011.
- [67] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.
- [68] J. L. Schonberger and J.-M. Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016.
- [69] W. Smith, R. Ramamoorthi, and S. Tozza. Height-from-polarisation with unknown lighting or albedo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2019.
- [70] W. A. Smith, R. Ramamoorthi, and S. Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *European Conference on Computer Vision*, pages 109–125. Springer, 2016.
- [71] W. A. P. Smith, R. Ramamoorthi, and S. Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *Computer Vision - ECCV 2016*, volume 9912 of *Lecture Notes in Computer Science*, pages 109–125, 2016.
- [72] Sony. *Image Sensor*, 2020(accessed 22/May/2020). <https://www.sony-semicon.co.jp/e/products/IS/polarization/technology.html>.
- [73] M. W. Spong, S. Hutchinson, M. Vidyasagar, et al. *Robot modeling and control*, volume 3. Wiley New York, 2006.
- [74] G. Strang. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.
- [75] V. Taamazyan, A. Kadambi, and R. Raskar. Shape from mixed polarization. *arXiv preprint arXiv:1605.02066*, 2016.
- [76] F. G. Tickell. *The techniques of sedimentary mineralogy*. Elsevier, 2011.

- [77] S. Tozza, W. A. Smith, D. Zhu, R. Ramamoorthi, and E. R. Hancock. Linear differential constraints for photo-polarimetric height estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2279–2287, 2017.
- [78] G. Vogiatzis, C. H. Esteban, P. H. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2241–2246, 2007.
- [79] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE transactions on information theory*, 51(11):3697–3717, 2005.
- [80] L. B. Wolff. Surface orientation from polarization images. In *Optics, Illumination, and Image Sensing for Machine Vision II*, volume 850, pages 110–121. International Society for Optics and Photonics, 1988.
- [81] L. 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.
- [82] L. B. Wolff and T. E. Boulton. Constraining object features using a polarization reflectance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):635–657, 1991.
- [83] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2115–2128, 2009.
- [84] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980.
- [85] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. *ACM Transactions on Graphics (ToG)*, 33(6):200, 2014.
- [86] L. Yang, F. Tan, A. Li, Z. Cui, Y. Furukawa, and P. Tan. Polarimetric dense monocular slam. In *Proc. CVPR*, pages 3857–3866, 2018.
- [87] 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.

- [88] K. Zhou, X. Wang, Y. Tong, M. Desbrun, B. Guo, and H.-Y. Shum. Texturemontage: Seamless texturing of arbitrary surfaces from multiple images. *ACM Transactions on Graphics*, 24(3):1148–1155, 2005.
- [89] D. Zhu and W. A. Smith. Depth from a polarisation+ rgb stereo pair. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7586–7595, 2019.
- [90] D. Zhu and W. A. P. Smith. Least squares surface reconstruction on arbitrary domains. In *16th European Conference on Computer Vision (ECCV 2020) Proceedings*. Springer-Verlag, 2020.