

Polarisation photometric stereo

Gary A. Atkinson

Bristol Robotics Laboratory, Department of Engineering Design and Mathematics, University of the West of England, Bristol, BS16 1QY, UK



ARTICLE INFO

Article history:

Received 30 December 2016

Revised 1 March 2017

Accepted 30 April 2017

Available online 1 May 2017

Keywords:

Polarisation

Photometric stereo

3D Reconstruction

ABSTRACT

This paper concerns a novel approach to fuse two-source photometric stereo (PS) data with polarisation information for complete surface normal recovery for smooth or slightly rough surfaces. PS is a well-established method but is limited in application by its need for three or more well-spaced and known illumination sources and Lambertian reflectance. Polarisation methods are less studied but have shown promise for smooth surfaces under highly controlled capture conditions. However, such methods suffer from inherent ambiguities and the depolarising effects of surface roughness. The method presented in this paper goes some way to overcome these limitations by fusing the most reliable information from PS and polarisation. PS is used with only two sources to deduce a constrained mapping of the surface normal at each point onto a 2D plane. Phase information from polarisation is used to deduce a mapping onto a different plane. The paper then shows how the full surface normal can be obtained from the two mappings. The method is tested on a range of real-world images to demonstrate the advantages over standalone applications of PS or polarisation methods.

Crown Copyright © 2017 Published by Elsevier Inc. All rights reserved.

1. Introduction

Three-dimensional surface reconstruction/analysis has been a major goal of computer vision for many years. There have been a plethora of methods proposed in the last few decades, each with their unique pros and cons. For example, binocular stereo (Brown and Hager, 2003) applies at any range and is inexpensive but inappropriate for featureless surfaces. By contrast, time-of-flight cameras (Foix and Alenyà, 2011) are known to operate on a variety of scenes at high speed but suffer at long range. This paper uses a fusion of photometric stereo (PS) and polarisation vision aimed at featureless surfaces at close range: a goal that typically requires expensive methods that use lasers or active illumination, else only operate at low resolution.

The proposed method initially applies a form of two-source PS to estimate part of the surface normal. This is much more applicable to real world applications than most PS methods that require three or more sources on different illumination planes. Polarisation information is then used in conjunction with the PS-based estimates in order to fully constrain the normals at each pixel. The paper therefore represents a significant step forward towards commercial robotics applications able to exploit a largely untapped area of light: that of its polarisation state.

This paper starts by reviewing the state-of-the-art in three-dimensional surface estimation using polarisation. Then, in

Section 2, the key background theory from Fresnel Theory and polarisation imaging is provided. The novel reconstruction algorithm is described in detail in Section 3. Experimental results and conclusions are provided in Sections 4 and 5 respectively.

1.1. Related work

One of the early proposed methods for surface shape estimation was that of shape-from-shading, which aims to acquire geometry from an image using intensity information only (Drouot et al., 2008). Due to its single-view and passive nature, it is an attractive goal and does not suffer from the limitations of the above examples (binocular stereo and time of flight). Unfortunately, the method is known to be ill-posed due to the bas-relief ambiguity and other reasons (Belhumeur et al., 1999). PS was proposed as a means to overcome the ambiguities at the expense of the need for three or more light sources, of known direction, illuminated separately with an image captured for each. Only relatively recently however has this method become useful in many applications due to the relative ease of high speed camera–illumination synchronisation (Hansen et al., 2010). However, the applications are still highly limited due to the need for at least three sources. For example, for many land-based robotics applications, it is easy to include two light sources (to the far-left and far-right of the robot), but it is difficult to position a source high above the ground. That said, Zhang et al. (2016) apply geometric assumptions to force a solution for smooth surfaces with two-source PS.

E-mail address: gary.atkinson@uwe.ac.uk

A less-studied approach to 3D vision problems is that of *polarisation vision*. That is, the use of the polarisation state of reflected light to deduce surface information (Wolff and Boulton, 1991). As with the field of hyperspectral imaging, the method taps into an entire set of information about the incoming light (polarisation state/spectral composition) that a standard greyscale or RGB camera is unable to. Thus, one may argue that a great deal of potentially useful information about the incoming light is lost in standard vision techniques. Of course, the additional information available from polarisation and hyperspectral imaging systems comes at the cost of more expensive and often slower capture hardware.

Most previous computer vision research using polarisation uses Fresnel theory applied to *specular* reflection (Hecht, 1998). The theory tells us that initially unpolarised illumination undergoes a partial linear polarisation process upon reflection from surfaces. The specific properties of the polarisation correlate to the relationship between the surface orientation and the viewing direction (Saito et al., 1999; Wolff and Boulton, 1991). Unfortunately, there are inherent ambiguities present (see Section 2) and the refractive index of the target is typically required. Further, different equations are required to model reflection if a *diffuse* component is present.

Atkinson and Hancock (2007); Miyazaki et al. (2004) and Berger et al. (2016) all use multiple viewpoints to overcome the ambiguity issue. In Miyazaki et al. (2004), specular reflection is used on transparent objects – a class of material that causes difficulty for most computer vision methods. In Atkinson and Hancock (2007), a patch matching approach is used for diffuse surfaces to find stereo correspondence and, hence, 3D data. Results are of comparable quality to this paper, but cannot easily handle specularities/inter-reflections. In Berger et al. (2016), an energy functional for a regularization-based stereo vision is applied. Taamazyan et al. (2016) use a combination of multiple views and a physics-based reflection model to simultaneously separate specular and diffuse reflection and estimate shape (Miyazaki et al., 2003 also separate reflection components but only as a pre-processing step).

Drbohlav and Šára (2001, 2002) use PS (as in this paper) but with linearly polarized incident illumination. Atkinson and Hancock (2008) also use PS but with more light sources and less resilience to inter-reflections than the method of this paper. Ngo et al. (2015) propose a similar combined polarisation/two-source PS method to the one presented in this paper with optimal results reported to be of the order 2° error from ground truth (this paper manages $3-4^\circ$). However, the Ngo paper focuses on estimating the lighting directions rather than addressing complications due to specularities/inter-reflections (which was a key motivation here).

Garcia et al. (2015) use a circularly polarised light source to overcome the ambiguities while Morel et al. (2006) extend the methods of polarisation to metallic surfaces by allowing for a complex index of refraction. Smith et al. (2016) assume a known constant refractive index and albedo but are able to pose the problem as a system of linear equations. This allows to directly deduce depth up to a concave/convex ambiguity in the presence of both diffuse and specular reflection. Furthermore, the method is successfully applied to conditions of uncontrolled illumination and only relies on a single view.

In an effort to overcome the need for the refractive index, Miyazaki et al. (2003) constrain the histogram of surface normals while Rahmann (2003) and Rahmann and Canterakis (2001) use multiple views and the orientation of the plane of polarisation for correspondence. Huynh et al. (2010) actually estimate the refractive index using spectral information. Kadambi et al. (2015) propose a method to allow polarisation information to be combined with alternative depth capture methods resulting in overall improved reconstructions.

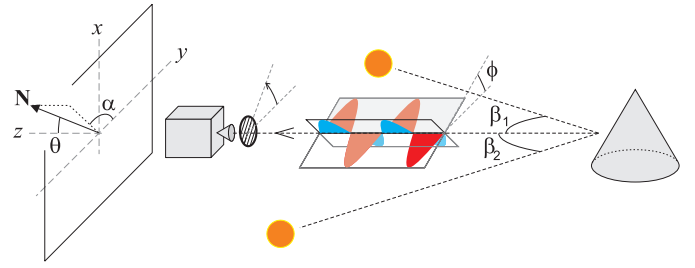


Fig. 1. Arrangement and definitions. Two polarisation images are acquired using the camera and rotating polariser. Each image has a different source illuminated. The surface normal, \mathbf{N} , is defined by its zenith angle, θ , and azimuth angle, α .

1.2. Contribution

The motivation for this paper relates to the following weaknesses of existing methods in related areas: firstly the need for three light sources in PS; secondly the multiple ambiguities in polarisation; and thirdly the complexities due to specular inter-reflections and diffuse reflection in both PS and polarisation. The contributions are as follows:

- Exploit the advantages of PS methods but using only two sources.
- Reduce the dependency of PS on light source knowledge.
- Use a combination of PS and polarisation data with a novel region growing algorithm to deduce the full 3D surface normals of an object in the presence of specular inter-reflections (assuming that the majority of the surface is not specularly reflecting). The method is also somewhat robust to direct specularities, depending on their relative intensity compared to the diffuse component.
- Avoid the need to know or estimate the refractive index.

As stated earlier, it is thought that overcoming these limitations allows a significant step towards the usage of polarisation information in real-world robotics applications.

2. Polarisation vision

This paper is based on the premise that light undergoes a partial polarisation process upon reflection from smooth surfaces. Consider the experimental arrangement shown in Fig. 1, which is used for all of this work. Fresnel reflectance theory (Hecht, 1998) is able to quantify the polarisation process that occurs when initially unpolarised incident light is reflected towards the camera. The reflected light can be parametrised by three values. First is the intensity, I , of the light. Second is the phase angle, ϕ , which defines the principle angle of the electric field component of the light as shown in the figure. Finally, the degree of polarisation, ρ , indicates the level of polarisation from 0 (unpolarised) to 1 (completely linearly polarised) (Atkinson and Hancock, 2006).

As explained elsewhere in the literature (Atkinson and Hancock, 2006; Wolff and Boulton, 1991), for specularly reflected light the phase angle aligns perpendicularly to the projection of the surface normal onto the image plane. Assuming that the Wolff sub-surface scattering model (Wolff, 1994) applies, diffuse reflection causes parallel alignment. Since there is no distinction in phase angle shifts of π radians, there are therefore four possible surface azimuthal angles (defined in Fig. 1) for an unknown reflectance type:

$$\alpha \in \left[\phi, \phi + \frac{\pi}{2}, \phi + \pi, \phi + \frac{3\pi}{4} \right] \quad \text{where } 0 \leq \phi < \pi \quad (1)$$

The degree of polarisation contains information principally related to the zenith angle of the surface. Unfortunately, the

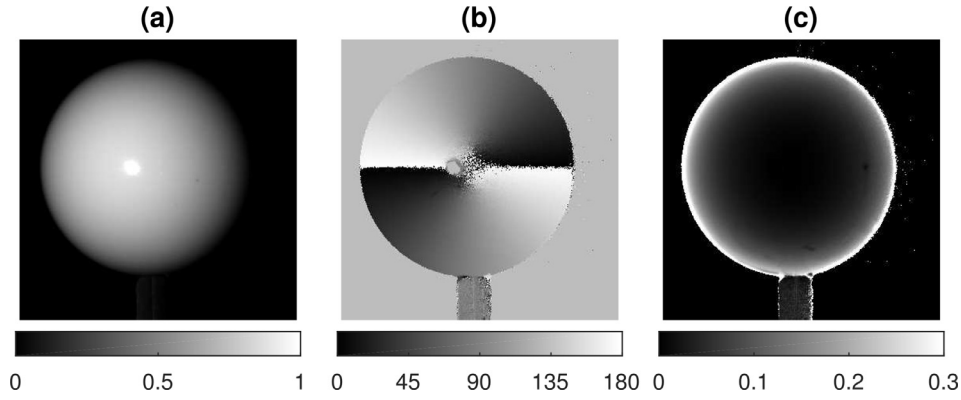


Fig. 2. Example polarisation image of a white snooker ball with no inter-reflections and only one specularity (which is saturated in the image). (a) Intensity I , (b) phase angle ϕ , (c) Degree of polarisation, ρ .

relationship between zenith angle and degree of polarisation varies substantially between specular and diffuse reflection and depends on the refractive index of the reflecting surface, which is typically unknown (Atkinson and Hancock, 2006). For this reason, its use for surface normal estimation is avoided in this paper. The method does however use the degree of polarisation as a measure of reliability of the phase data: any phase data with a corresponding degree of polarisation below a threshold (fixed at 1%) is deemed unreliable due to their correspondingly high associated noise levels.

The method described in the next section requires two polarisation images of an object for input (a single polarisation image comprises a separate intensity, phase and degree of polarisation value for each pixel). There are now several commercially available cameras that capture such data directly such as the Fraunhofer “POLKA” (Fraunhofer Institute for Integrated Circuits, 2017). The POLKA has a sensor divided into 2×2 cells, with each element sensitive to light of a given electric field orientation (0° , 45° , 90° and 135°). If the intensities measured for these angles are called I_0 , I_{45} , I_{90} and I_{135} respectively, then the polarisation data for each cell is calculated via the Stoke’s parameters S_0 , S_1 and S_2 (assuming no circular polarisation is present) (Hecht, 1998):

Stoke’s parameters:

$$S_0 = \frac{I_0 + I_{45} + I_{90} + I_{135}}{2} \quad (2)$$

$$S_1 = I_0 - I_{90} \quad (3)$$

$$S_2 = I_{45} - I_{135} \quad (4)$$

Polarisation image data:

$$I = S_0 \quad (5)$$

$$\phi = \frac{1}{2} \arctan_2(S_2, S_1) \quad (6)$$

$$\rho = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \quad (7)$$

where \arctan_2 is the four quadrant inverse tangent (Burger and Burge, 2010).

Raw data for this paper were captured using a Dalsa Genie HM1400 camera with a motor-controlled circularly polarising filter in front of the lens. The motor rotated the filter to angles of 0° , 45° , 90° and 135° between each image capture, after which (2–7) were applied to calculate the polarisation image; following a similar method to the POLKA. A circular filter was used, rather than linear, in order to minimise effects of any inter-reflections in the lens

(Schneider KMP-IR Xenoplan 23/1,4-M30,5) (Karpel and Schechner, 2004). The rotating polariser method is clearly much slower (but cheaper) than using the POLKA (or similar products) and thus restricts application to stationary objects. However, the rest of the paper is independent of the capture method and so could easily be applied to moving objects if such cameras were used with this method in the future.

Examples of the three components of the polarisation image for a white snooker ball are shown in Figs. 2 to 4. Note that the intensity is normalised to the range $[0,1]$. For Fig. 2, the capture conditions were ideal, with the ball illuminated by a single small white LED placed close to the camera and blackout curtains surrounding the ball to diminish all inter-reflections from the environment. The reflection is therefore of diffuse type throughout the surface and so the first or third solutions to (1) must be true for all pixels.

For Fig. 3, a plain matte white board was placed behind the ball. A sudden phase shift is present near the occluding contours of the ball due to an inter-reflection from the board. Since this region is effectively undergoing a specular inter-reflection, the second or fourth solutions to (1) are true. By contrast, the first or third solutions are true for the rest of the ball, which is exhibiting diffuse reflection. Note also, that the degree of polarisation is higher near the inter-reflection on the right-hand side of the ball. This is also predicted by the reflection theory but is of less significance to this paper (Atkinson and Hancock, 2006). The noise-ridden background to Fig. 3 is due to the very low degree of polarisation for the board and the shadow cast by the ball; neither of which are of significance here.

The polarisation image in Fig. 4 includes an extended light source located to the upper-right of the ball. The phase shift is clearly present here, demonstrating the specular nature of the direct reflection. Careful examination of the degree of polarisation demonstrates some inter-reflections from the laboratory are present just below the main specularity. Since the intensity from these inter-reflections is very low, the dominant reflection type is diffuse there.

3. Method

The approach can be broadly divided into the following steps assuming that the starting point is two full polarisation images corresponding to the two light source locations shown in Fig. 1.

1. Extract estimates of 2D (y - z) surface normals from two-source PS.
2.
 - (a) Extract estimates of 2D (x - y) surface normals from polarisation aided by PS.

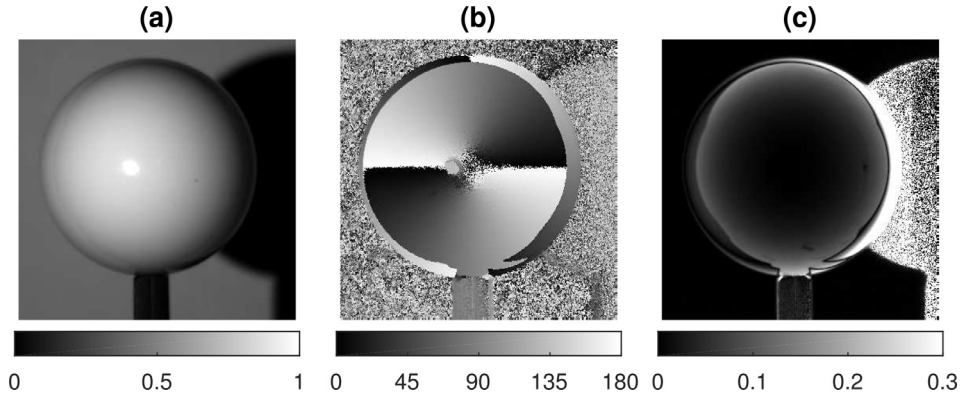


Fig. 3. Example polarisation image of a white snooker ball with specular inter-reflections from a white board behind the ball. Otherwise, conditions were the same as in Fig. 2. (a) Intensity I , (b) phase angle ϕ , (c) Degree of polarisation ρ . The degree of polarisation values above 0.3 (right side of ball) are all shown white here to improve the clarity of the rest of the figure.

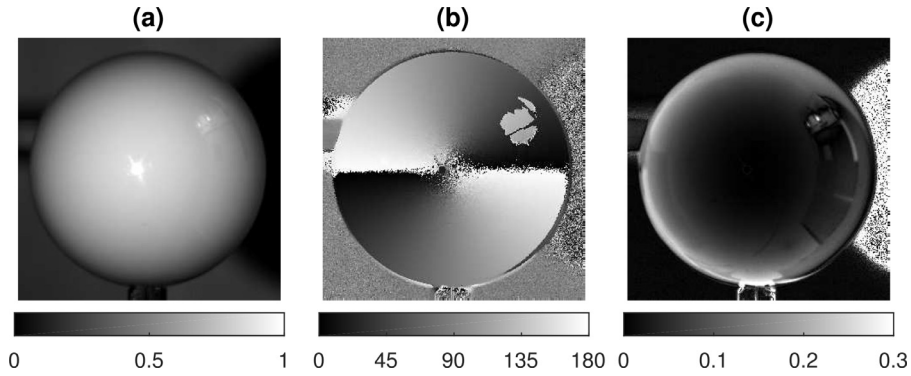


Fig. 4. Example polarisation image of a white snooker ball with a direct specularity from an extended source and a sequence of uncontrolled environmental inter-reflections on the right-hand side. (a) Intensity I , (b) phase angle ϕ , (c) Degree of polarisation ρ .

(b) Disambiguate spurious data or specular inter-reflections (or, potentially, direct specularities) using region growing.

3. Combine data from each modality to obtain 3D (x - y - z) surface normals.

The first step essentially applies a 2D version of PS to obtain a 2D surface normal at each pixel. This is not merely a projection of the 3D normal onto the y - z plane but rather a vector representing one particular degree of freedom of the orientation of the surface at each point. The second step uses the polarisation phase angle of the incoming light and assumes diffuse reflection (for now) to estimate a 2D version of the surface normal in the x - y plane. PS data is applied to disambiguate most of these vectors but region growing is needed where specularities occur or where previous disambiguation is deemed incorrect (as defined in Section 3.2.2). Finally, the information above is combined to form a 3D surface normal map.

3.1. Application of PS

The aim here is to use the method of PS to estimate a 2D normal in the y - z plane at each point. To do this, the original methodology of Woodham (1980) is adapted to two images in order to obtain a set of n normals as follows:

$$\left\{ \mathbf{N}_i^{(ps)} = \begin{bmatrix} N_{y,i}^{(ps)} \\ N_{z,i}^{(ps)} \end{bmatrix} \right\}_{i=1}^n \leftarrow \begin{bmatrix} -\cos(\beta_R) & \cos(\beta_L) \\ \sin(\beta_R) & -\sin(\beta_L) \end{bmatrix} \begin{bmatrix} I_{L,i} \\ I_{R,i} \end{bmatrix} \forall i \quad (8)$$

where the angles β_L and β_R are defined in Fig. 1 and the suffix “ps” is a reminder that these estimates are from photometric stereo. The “ \leftarrow ” symbol is used throughout this paper to refer to

variable assignment. The vectors are then normalised into unit vectors in preparation for fusion with the polarisation data.

One of the aims of this paper is to ensure that the method is robust to disruptions in illumination. For example, standard PS yields poor results if one source appears much brighter than the other or if the distance to the target object varies (as this affects the light source vector). The results for this paper were obtained with accurately positioned illumination so this potential issue does not manifest. However, it is possible to use the phase information from polarisation to calibrate the light source intensities assuming a linear (or at least known) camera response profile. This is done by scaling the intensities such that areas where $\phi = 90^\circ$ or 270° have equal I values. Further, it is worth noting at this point that the distance-to-target issue is less pronounced here than standard PS since only the z -component of the normal is affected by incorrect source vectors.

3.2. Application of polarisation vision

This section describes the method to estimate 2D normals in the x - y plane. It assumes that the method for polarisation image acquisition described in Section 2 has been applied to arrive at corresponding intensity, phase and degree of polarisation values for all n pixels: $\{I_i, \phi_i, \rho_i\}_{i=1}^n$. This section also makes use of $\{\mathbf{N}_i^{(ps)}\}_{i=1}^n$. This part of the algorithm is in two parts, the code for which is summarised in Algorithm 1.

The raw data yields two estimates of polarisation values for each pixel (one corresponding to each light source direction). In theory, the data should be identical between each polarisation image, except for the locations of specularities and shadows. The proposed algorithm first forms a new polarisation image corre-

Algorithm 1 Input: $\{\phi_i, \rho_i, N_i^{(ps)}\}_{i=1}^n$ Output: $\{N_i^{(po)}\}_{i=1}^n$.

```

1:  $\{\phi_i\} \leftarrow \text{sharpen}(\{\phi_i\})$ 
2:  $\{\alpha_i\} \leftarrow \begin{cases} \phi_i & \forall i | (N_{y,i}^{(ps)} < 0 \wedge \phi_i > \frac{\pi}{2}) \vee (N_{y,i}^{(ps)} > 0 \wedge \phi_i < \frac{\pi}{2}) \\ \phi_i + \pi & \text{otherwise} \end{cases}$ 
3:  $\{N_i^{(po)} = \begin{bmatrix} N_{x,i}^{(po)} \\ N_{y,i}^{(po)} \end{bmatrix}\}_{i=1}^n \leftarrow \begin{bmatrix} -\sin \alpha_i \\ \cos \alpha_i \end{bmatrix} \forall i$ 
4:  $\{R_i\}_{i=1}^n \leftarrow \begin{cases} 1 & \forall i | \text{isBorderPixel}(i) \vee \rho_i < 0.01 \\ 0 & \text{otherwise} \end{cases}$ 
5: while [More seed points?] do
6:    $j \leftarrow \text{Get seed point}$ 
7:    $\{N^{(po)}\} \leftarrow \text{localAlign}(\{N^{(po)}, R\}, j)$ 

```

sponding to what is deemed the most reliable data for each pixel. Two methods were considered for this: (1) use the polarisation data from the image corresponding to the highest intensity at each point, (2) use the polarisation data from the image corresponding to smallest deviation between raw data and the transmitted radiance sinusoid (Atkinson and Hancock, 2006). While the latter approach is perhaps more technically sound, it was experimentally determined that the final quality of results were very similar between the two approaches. For this reason, the former more computationally efficient approach was used for the remainder of the research.

3.2.1. Initial azimuth estimation

This is covered by the first three lines of Algorithm 1. The first line applies a sharpening operator to the phase data. The reason for this that the subsequent specular detection/disambiguation algorithm (“localAlign”) is more reliable in the presence of sharp transitions between specularities and diffuse regions. An alternative to this is to increase the definition of a neighbourhood in the subsequent function, allowing the method to bridge over any gradual transitions.

The method makes the initial assumption that all surface normal projections onto the x-y (image) plane are aligned parallel to, or anti-parallel to, the phase angle, as predicted by the theory for diffuse reflection covered in Section 2. Line 2 sets the azimuthal angle, α , of each point accordingly. Whether said projections are parallel or anti-parallel depend on the best match to the estimated 2D normal from PS. Line 3 simply generates a set of 2D surface normals on the x-y plane using the calculated azimuth angles. The suffix “po” reminds us that these estimates are primarily from polarisation.

3.2.2. Final azimuth estimation

This is covered by the last four lines of Algorithm 1. It is expected that most of the normals will be correct at this point. However, it is likely that some regions of the image will be incorrect due to one of the following reasons:

- Diffusely reflecting regions with incorrect disambiguation. This is typically where $N_y^{(ps)} \approx 0$, meaning the initial disambiguation is not robust. These regions have an azimuth error of π radians.
- Specular (direct or inter-reflective) regions. In these areas the azimuth angle error is $\pi/2$ radians, as predicted by the theory described in Section 2.

Lines 4 to 7 of Algorithm 1 are intended to address these possibilities, while enforcing (1). First, a set of pixels, $\{R_i\}_{i=1}^n$ are deter-

mined that should *not* be further considered by the algorithm. In the first instance, these pixels correspond to:

- Image border pixels: purely to avoid slowing down the algorithm with numerous conditional statements.
- Those with very low (<1%) degree of polarisation: since these areas are deemed to have too much noise to be of use.

Next, a seed point is chosen. This can easily be done either manually or randomly using a point of high confidence (i.e. where the degree of polarisation is high and either $N_y^{(ps)} \ll 0$ or $N_y^{(ps)} \gg 0$). One weakness here however, is that this point must be of diffuse reflection so may benefit from some heuristics-based selection in future work. Assume for now that only one seed point is needed, but note that the code permits more if necessary (e.g. for complicated shapes). The rest of the process involves a recursive call to a region growing function, which progressively aligns spurious normals according to the constraints of (1).

Algorithm 2 “localAlign”.

Input: $\{N_i^{(po)}, R_i\}_{i=1}^n, j$ Output: $\{N_i^{(po)}\}_{i=1}^n$.

```

1: if [ $R_j = 1$ ] then
2:   return  $\{N^{(po)}, R\}$ 
3:  $R_j \leftarrow 1$ 
4: for [ $k \leftarrow \text{neighboursOf}(j)$ ] do
5:    $\Delta\alpha \leftarrow \arccos(N_j^{(po)} \cdot N_k^{(po)})$ 
6:   if [small ( $\Delta\alpha$ )] then
7:      $\{N^{(po)}\} \leftarrow \text{localAlign}(\{N^{(po)}, R\}, k)$ 
8:   if [small ( $\pi - \Delta\alpha$ )] then
9:      $N_k^{(po)} \leftarrow -N_k^{(po)}$ 
10:     $\{N^{(po)}\} \leftarrow \text{localAlign}(\{N^{(po)}, R\}, k)$ 
11:   if [small ( $\frac{\pi}{2} - \Delta\alpha$ )] then
12:      $N_{\text{rot}} \leftarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot N_k^{(po)}$ 
13:     if [ $\arccos(N_j^{(po)} \cdot N_{\text{rot}}) < \frac{\pi}{2}$ ] then
14:        $N_k^{(po)} \leftarrow N_{\text{rot}}$ 
15:     else
16:        $N_k^{(po)} \leftarrow -N_{\text{rot}}$ 
17:      $\{N^{(po)}\} \leftarrow \text{localAlign}(\{N^{(po)}, R\}, k)$ 

```

The region growing works as follows (see Algorithm 2). First, the function checks whether the point in question j should be considered by reference to $\{R_i\}$ (lines 1 and 2). On occasions where $R_j = 1$, the function call terminates. Where this is not the case, point j is added to $\{R_i\}$, i.e. so it is not considered again (line 3). The remainder of the function is completed for each neighbour of the point (line 4). The neighbourhood can be defined in multiple ways but a simple 4-connected region is used for this paper.

Next, the angle between neighbouring x-y surface normals is calculated, $\Delta\alpha$ (line 5). There are four possibilities for $\Delta\alpha$:

- $\Delta\alpha$ is close to zero: assume both azimuth angles are correct, move from j to neighbouring point, k , and continue the process (lines 6 and 7).
- $\Delta\alpha$ is close to π : as above but assume azimuth disambiguation was incorrect so rotate by π (lines 8 to 10).

- $\Delta\alpha$ is close to $\pi/2$: rotate azimuth by $\pi/2$ since this region is likely following specular reflection. The direction of rotation is chosen to minimise the angle between the normals at j and k . Again, move from j to k and continue the process thereafter (lines 11 to 17).
- Otherwise: there could be a boundary of orientation so there is no reason to believe a the azimuth estimate is erroneous.

The definition of “close” for this recursive function remains open for now but only minor effects on results were observed as the threshold for closeness was varied between 5° and 10° .

3.3. Fusion of data into full 3D surface normals

The methods from the previous two sections give two sets of 2D surface normals: $\{\mathbf{N}^{(ps)}\}$ on the y - z plane and $\{\mathbf{N}^{(po)}\}$ on the x - y plane. These can be combined into a single 3D vector. First assume that the 2D vectors are normalised such that the following is true for each point:

$$N_x^{(po)^2} + N_y^{(po)^2} = 1 = N_y^{(ps)^2} + N_z^{(ps)^2} \quad (9)$$

where the subscripts i are omitted for the sake of brevity.

Denote the 3D surface normal at a particular point $\mathbf{N} = [N_x, N_y, N_z]^T$. Further, choose the components estimated by polarisation for x and y so there is only one unknown, N_z :

$$\mathbf{N} = [N_x, N_y, N_z]^T = [N_x^{(po)}, N_y^{(po)}, N_z]^T \quad (10)$$

Normalising this vector to unit length gives:

$$\mathbf{n} = \frac{[N_x^{(po)}, N_y^{(po)}, N_z]^T}{\sqrt{1 + N_z^2}} = [n_x^{(po)}, n_y^{(po)}, n_z]^T \quad (11)$$

This can then be re-normalised such that $n_y^{(po)^2} + n_z^2 = 1$, matching the form of the 2D estimate from PS as in (9). For the z component specifically:

$$\frac{n_z}{\sqrt{n_y^{(po)^2} + n_z^2}} = N_z^{(ps)} \quad (12)$$

Substituting the components of \mathbf{n} from (11) and simplifying:

$$N_z^{(ps)} = \frac{N_z}{\sqrt{N_y^{(po)^2} + N_z^2}} \quad (13)$$

Rearranging:

$$N_z = \left| N_y^{(po)} \right| \left(\frac{1}{N_z^{(ps)^2}} - 1 \right)^{-1/2} \quad (14)$$

where the $|\cdot|$ sign is used since N_z must be positive. This means that the only unknown in (10) is resolved and so the surface normal is fully determined.

As stated earlier, the regions of the images where the degree of polarisation is less than 1% are not used in the algorithm to ensure robustness. For the results in the next section, bi-cubic interpolation was used to estimate the normals for these areas. It was also found that improvements could be made by interpolating over regions of very low N_y , although this was kept to a minimum. It is acknowledged however, that more sophisticated methods from the shape-from-shading literature (Durou et al., 2008) may be more appropriate for future work.

The normals can be converted to depth using any of a number of integration methods (Saracchini et al., 2012). Since it is not the aim of this paper to develop a new integrator, the well-established

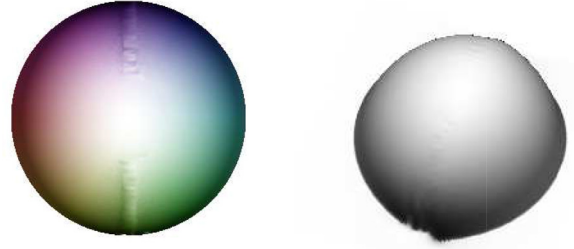


Fig. 5. Surface normals and depth estimated from the polarisation image shown in Fig. 2 (and its right-illuminated counterpart). Surface normals are encoded by colour (azimuth) and saturation (zenith).

Frankot–Chellappa method (Frankot and Chellappa, 1988) is primarily used for this step. The method is fast and highly robust to noise, but can suffer from over-smoothing. An alternative method that does not over-smooth is briefly considered in Section 4.3.

4. Results

This section contains a detailed analysis of the performance of the method for a range of target objects. In the first instance, a thorough numerical evaluation for the surface normals and height estimations are carried out for a white snooker ball under varying conditions and compared to ground truth and baseline methods. Subsequently, a range of more challenging shapes and textures are considered.

4.1. Spherical objects

Consider Fig. 2, the polarisation image for a white snooker ball captured in ideal conditions as described in Section 2. The angle between the camera and light sources was $\beta_L = \beta_R = 19.6 \pm 0.5^\circ$. One hundred images were captured at each of the four polariser angles (0° , 45° , 90° and 135° , aligned with resolution to the nearest degree) and the mean intensity used at each pixel to minimise noise. As expected, the phase angle directly relates to the surface azimuth up to a 180° ambiguity and the degree of polarisation is highest near the occluding contours (Atkinson and Hancock, 2006).

As mentioned earlier in this paper, and elsewhere in the literature, among the issues facing polarisation-based computer vision algorithms are the complications caused by the simultaneous presence of specular and diffuse reflection. To test the robustness of the method proposed here, a second polarisation image was captured in exactly the same conditions as for Fig. 2 but with a large planar matte white surface placed at approximately 128 mm behind the target object. The resulting image is shown in Fig. 3, as also discussed in Section 2. The specular inter-reflection can only just be seen in the intensity image, yet the 90° phase shift and increased degree of polarisation make it clear in those two components of the polarisation image.

The results of applying the surface normal estimation algorithm to the polarisation images shown in Figs. 2 and 3 are shown in Figs. 5 and 6, respectively. The results are qualitatively good although the region of interpolation (described in Section 3.3) near the top-centre and bottom-centre of the normal maps is clearly apparent. Note that the data has been cropped here so only the object itself is being integrated.

Fig. 7 shows a comparison of the profile of the reconstructions from Figs. 5 and 6 to ground truth data. To aid comparison, the profiles are aligned such that the tops of each curve are touching. Very similar results were found for pink and yellow balls, while blue and green results were slightly poorer due to their lower brightness causing higher noise levels.

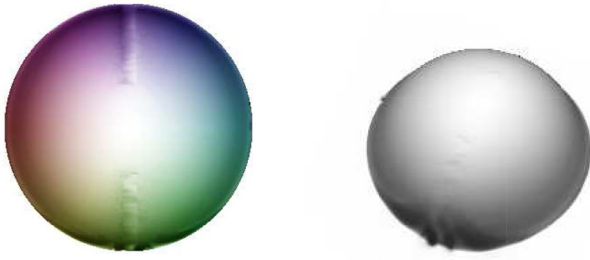


Fig. 6. Surface normals and depth estimated from the polarisation image shown in Fig. 3 (and its right-illuminated counterpart). Surface normals are encoded by colour (azimuth) and saturation (zenith).

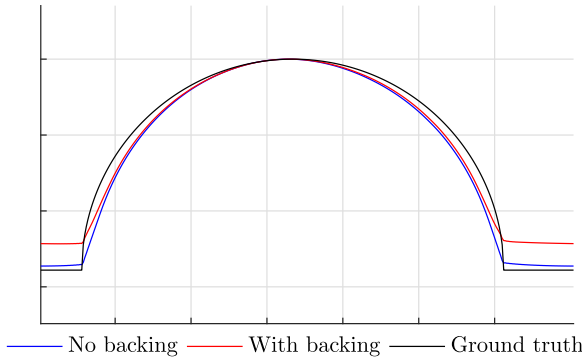


Fig. 7. Comparison between the estimated depth profile of the snooker ball and ground truth for the data represented in Figs. 5 and 6.

In order to optimise the surface normal estimates with regards to the light source separation, both simulated and real-world data were analysed for varying light source angles. For the simulated data, images were generated assuming no inter-reflections/specularities were present and a range of illumination angles, β , between 5° and 25° . Gaussian noise was also added with standard deviations ranging from 0 (no noise) to 0.015 (typical for the camera settings used). The results are shown in Fig. 8, where the performance metric used is the median angular error, ε , between the method and ground truth. This metric has been chosen since the more obvious choice of height residuals would also be

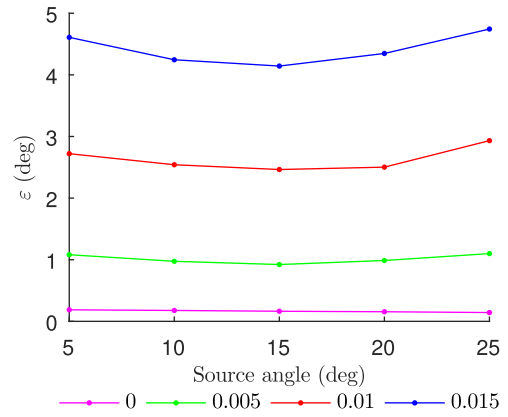


Fig. 8. Surface normal errors using synthetic data with varying levels of Gaussian noise added (standard deviations shown in legend).

assessing the integration method used, which is not the contribution of the current paper.

The results to the analysis of simulated data show the following:

- The surface normals are not perfectly estimated even with no noise present as a result of small surface regions being interpolated over. There may also be some edge-effects degrading results slightly.
- There is relatively little variation in performance with illumination angle for most noise levels.
- There is an optimum illumination angle of about 15° for higher noise levels. This is due to the fact that (1) at very small angles the difference between intensity images is very small which is poor conditioning for PS, and (2) at large angles major areas of shadow are present.

Experimental results for varying the illumination angle are shown in Fig. 9. Errors in both surface angle, ε , and root-mean-square height are shown although the focus here is on the former due to reasons already mentioned. Unsurprisingly, the magnitudes of the errors are slightly higher than for simulated data for most cases. Surprisingly however, the trend towards an optimum separation is reversed for the real data. It is not entirely clear why this should be the case but repeated experiments have yielded similar trends. It was therefore decided to use 20° illumination angles

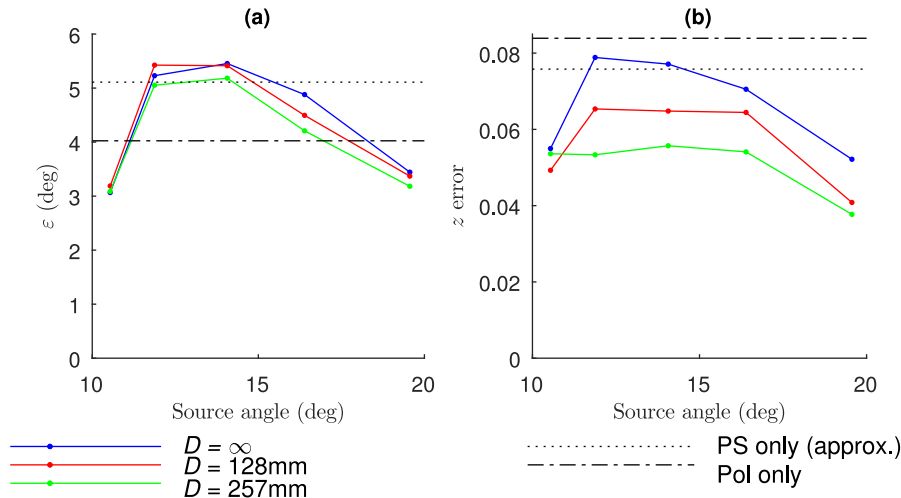


Fig. 9. (a) Angular errors and (b) height errors for the white snooker ball with a white board at various distances behind the ball, D . Height values are in units such that a value of 1 corresponds to a distance of one radius. Best case results for PS-only and polarisation-only are also indicated.

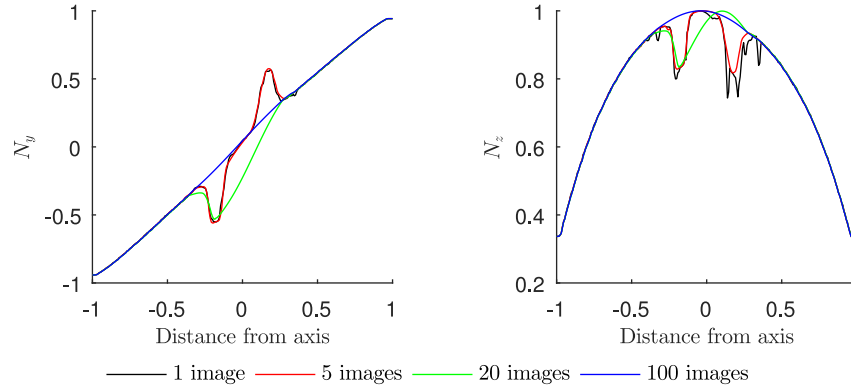


Fig. 10. Comparison between surface normal estimates for varying number of images per polariser angle for a cross-section of the snooker ball.

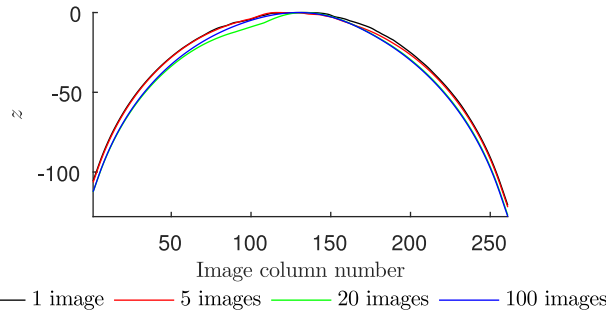


Fig. 11. Comparison between surface depth estimates for varying number of images per polariser angle for a cross-section of the snooker ball.



Fig. 13. Photograph of the multicoloured cylinder used for Fig. 12.

for the remainder of the research (higher values than this were neither practical for the experimental rig being used nor for most real-world applications).

Fig. 9 also shows results with the white board re-introduced at several distances and the baseline results for standard photometric stereo and single-source polarisation. The results with the board present show a strong performance by the method at overcoming specular inter-reflections, which is one of the key strengths of the method. Indeed, there is evidence that in some cases the reconstructions may be even better when strong specular behaviour is present. For standard photometric stereo, four sources were used at the illumination angle that gave the best surface normal estimates for that method (which was experimentally determined to be 25°). For the single-source polarisation baseline, a single polarisation image was used with the illumination source as close to the camera as possible ($\beta_L = \beta_R \approx 5^\circ$ in practice) in order to minimise

shadowed regions. Regions of the image were manually selected in order to disambiguate the azimuth angles and the zenith angles were calculated according to the Wolff diffuse reflection model with a refractive index of 1.5 (Wolff and Boult, 1991). At first sight, the baseline polarisation method appears comparable to the novel method, but it is reiterated that the novel method is superior as it does not require manual disambiguation nor knowledge of refractive index.

For the case with no board present behind the ball, a test on partially simulated data was conducted in order to analyse the relative effect of the PS and polarisation elements of the method. Average angular errors for (1) entirely real; (2) simulated polarisation (ϕ and ρ); and (3) simulated intensity (I) data were: 7.6° , 6.4° and 2.9° respectively. This is rather striking as it shows the PS component has by far the largest associated error. This is presumably due to the resilience of polarisation phase data against reflectance properties compared to PS and the lesser need for carefully positioned light sources. It should be noted however, that such a breakdown of experimental error may change for certain target objects,

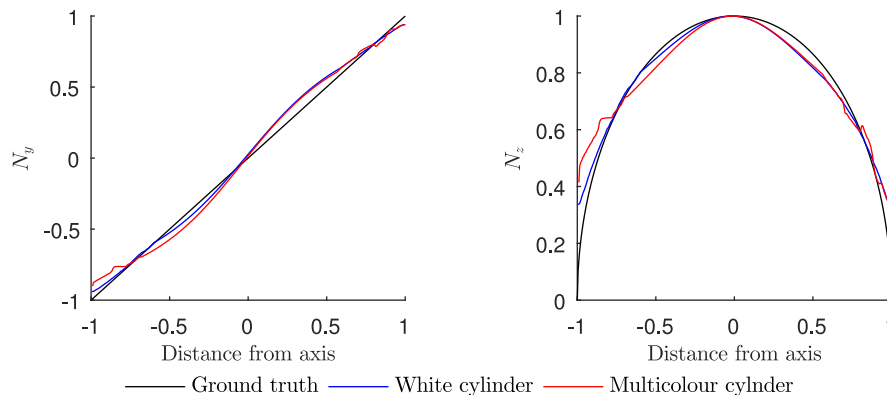


Fig. 12. Comparison between surface normal estimates and ground truth for a white porcelain and multicoloured porcelain cylinder.

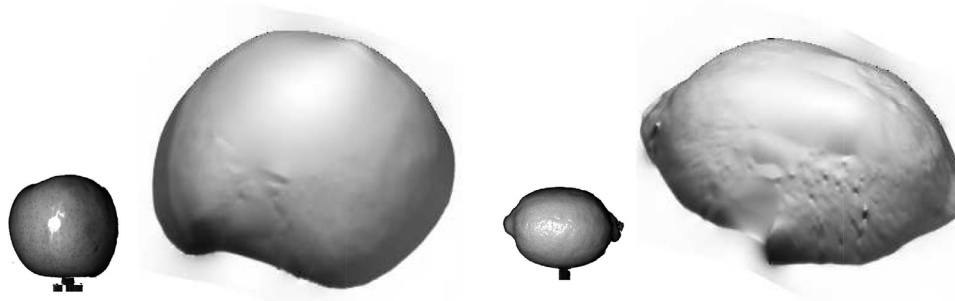


Fig. 14. Surface reconstructions of an apple and lemon.

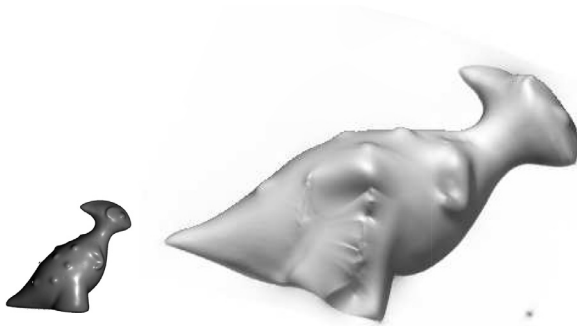


Fig. 15. Surface reconstruction of a model dinosaur.

such as where large areas of the surface have very low degree of polarisation.

These areas of low degree of polarisation have associated high noise levels and may be a potential weakness of the method. Results so far in this paper overcame this in a rather expensive manner by capturing 100 images in very rapid succession for each of the four polariser angles. The effects of using fewer images is illustrated by Fig. 10 which shows the estimated surface normal (y- and z-components only for compactness) using 1, 5, 20 and 100 images. At first sight, the quality of results is poor when only few images are used. However, the increased noise has a relatively low impact on surface height reconstruction, as shown in the Fig. 11. This is due to the facts that (1) the Frankot–Chellappa surface integrator is smoothing over the noise, and (2) the higher noise regions correspond to areas of low zenith angle which have smaller effects on the integration than those of larger zenith angle.

4.2. Other objects

Surface normal estimates for two cylinders are shown in Fig. 12. For this case, the cylinders are both porcelain: one pure white and one multicoloured, as shown in Fig. 13. The results clearly show a close match to theory; with the main discrepancy being at the occluding contours. For the case shown, the axes of the cylinders

were oriented parallel to the x-axis (see Fig. 1). Unfortunately, the method completely fails for the case where the cylinder axis is parallel to y-axis. This is due to the fact that all pixels on the surface in such cases appear equally bright from each light source, making seed selection and region growing nearly impossible: a clear weakness of the method which will be addressed in future work.

Tests on objects of more complicated geometry are shown in Figs. 14 and 15. The general geometries of the fruits in Fig. 14 have clearly been reproduced, albeit with over-smoothing in places. The model dinosaur in Fig. 15 also has reasonable shape, despite a few areas of incorrect disambiguation on the foot and head. It is hoped that such issues might be resolved in future work by optimising the location and number of seed points used.

4.3. An alternative integrator

The choice of the Frankot–Chellappa integration method for conversion from normals to depth was made since (1) it is well-established in the field and (2) it operates at high speed (much less than one second to reconstruct a VGA frame). However, the method is not very effective in certain circumstances. The most known of its weaknesses is its tendency to over-smooth surfaces as a result of its basis functions being unsuited to discontinuities.

Another weakness, which is more specific to the method of this paper, relates to the fact that the integrator applies equal weighting to all points on the surface. While this is sensible for most reconstruction methods, the approach here may benefit from weighting certain pixels more than others. Specifically, it is beneficial to give a higher weight to pixels of higher degree of polarisation since those are less prone to noise/distortion.

An alternative integrator that is much less affected by discontinuities and allows for variable pixel weighting is that of the M-estimator (Agrawal et al., 2006). Fig. 16 demonstrates the advantage of the M-Estimator approach using the underside of a porcelain bowl as a test object. The figure first shows the result of using the Frankot–Chellappa method. The result is highly distorted at the base of the bowl since the degree of polarisation is very low here over a wide area. In an effort to mitigate these effects, the M-Estimator was used with the weight set to zero wherever the

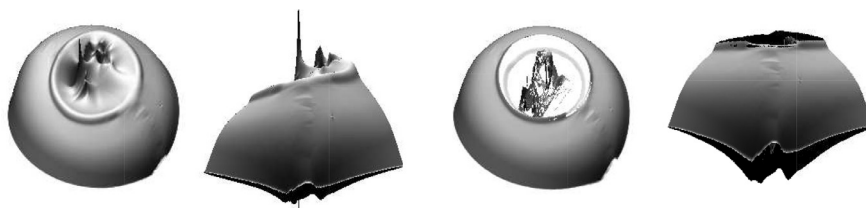


Fig. 16. Reconstructions of the underside of a porcelain bowl, including the side view. The left reconstructions were based on the Frankot–Chellappa method, while those to the right use the M-Estimator.

degree of polarisation is less than 1%. As shown in the figure, this significantly improves results for most of the surface¹. The obvious disadvantage of this approach was that the integration took much longer (≈ 40 s compared to ≈ 0.1 s).

5. Conclusion

Results in this paper demonstrate the potential for use of polarisation information for scene understanding in the presence of both diffuse reflection and specular inter-reflections. The results show reconstruction quality that is improved or comparable with the baseline methods while needing only two illumination sources, which is competitive to many other methods in the field. At present, the data is captured using a rotating polariser and switching light sources meaning the data takes several seconds to capture. As stated in Section 2 however, the method can be applied to data captured from a polarisation camera such as the POLKA. This means that the only significant limitation in capture time relates to the light source switching, which has easily been shown to operate in the order of milliseconds (Hansen et al., 2010). The processing time for this method is of the order five seconds at VGA resolution on an average personal computer.

The method has clearly demonstrated the use of polarisation information to reduce the number of light sources needed for photometric stereo to two. Unlike with many alternative methods in polarisation vision, this advantage does not come at the expense of needing to know a refractive index or making strong assumptions on reflectance properties. One major weakness of many PS methods is that of shadowing. While the novel method still suffers from reduced information at shadowed regions; only the zenith angle is affected. It is therefore predicted that the estimated azimuth angles for such regions may be combined with intensity information from one source only to fully recover shadowed surfaces in follow-on work. The dependence of the method on the reflectance properties of the surface is also deemed superior to many PS approaches since polarisation (at least phase angle) has little relationship to other reflectance parameters. That said, it is acknowledged that the proposed approach is still somewhat affected by reflectance properties due to the use of (8) for the z-component of the normal.

One limitation that should be addressed in future work relates to the open question of how to recover the geometry of an entire scene (as opposed to a single object). It may be that optimisation of seed point generation and weighting inter-object pixels in the integrator may reduce this limitation. Another limitation relates to the complications caused by very rough surfaces. The main problem caused by roughness for many polarisation-based methods is the depolarising effect of microscopic inter-reflections on the surface. This typically has a direct degrading effect on the reconstruction because the zenith angle is inherently dependent on degree of polarisation. For the method proposed in this paper however, the zenith angle is entirely determined by PS, diminishing the effects of roughness. Nevertheless, the increased noise present due to lower degrees of polarisation from rough surfaces will still clearly impact results of the method to an extent.

References

- Agrawal, A., Raskar, R., Chellappa, R., 2006. What is the range of surface reconstructions from a gradient field? In: *Proc. ECCV*, pp. 578–591.
- Atkinson, G.A., Hancock, E.R., 2006. Recovery of surface orientation from diffuse polarization. *IEEE Trans. Im. Proc.* 15, 1653–1664.
- Atkinson, G.A., Hancock, E.R., 2007. Shape estimation using polarization and shading from two views. *IEEE Trans. Patt. Anal. Mach. Intell.* 29, 2001–2017.
- Atkinson, G.A., Hancock, E.R., 2008. Two-dimensional BRDF estimation from polarization. *Comp. Vis. Im. Understanding* 111, 126–141.
- Belhumeur, P.N., Kriegman, D.J., Yuille, A.L., 1999. The bas-relief ambiguity. *Intl. J. Comp. Vis.* 35, 33–44.
- Berger, K., Voorhies, R., Matthies, L., 2016. Incorporating polarization in stereo vision-based 3D perception of non-Lambertian scenes. In: *Proc. SPIE*, 9837.
- Brown, M.Z., Hager, G.D., 2003. Advances in computational stereo. *IEEE Trans. Patt. Anal. Mach. Intell.* 25, 993–1008.
- Burger, W., Burge, M.J., 2010. *Principles of Digital Image Processing: Fundamental Techniques*. Springer.
- Drbohlav, O., Šára, R., 2001. Unambiguous determination of shape from photometric stereo with unknown light sources. In: *Proc. ICCV*, pp. 581–586.
- Drbohlav, O., Šára, R., 2002. Specularities reduce ambiguity of uncalibrated photometric stereo. In: *Proc. ECCV*, pp. 46–62.
- Durou, J.D., Falcone, M., Sagnoni, M., 2008. Numerical methods for shape-from-shading: a new survey with benchmarks. *Comp. Vis. Im. Understanding* 109, 22–43.
- Foix, S., Alenyà, G., 2011. Lock-in time-of-flight (ToF) cameras: a survey. *IEEE Sensors J.* 11, 1917–1926.
- Frankot, R.T., Chellappa, R., 1988. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. Patt. Anal. Mach. Intell.* 10, 439–451.
- Fraunhofer Institute for Integrated Circuits, <http://www.iis.fraunhofer.de/en/ff/bsy/tech/kameratechnik/polarisationskamera.html>. Accessed: 1 March, 2017.
- Garcia, N.M., de Erausquin, I., Edmiston, C., Gruev, V., 2015. Surface normal reconstruction using circularly polarized light. *Opt. Express* 23, 14391–14406.
- Hansen, M.F., Atkinson, G.A., Smith, L.N., Smith, M.L., 2010. 3D face reconstructions from photometric stereo using near infrared and visible light. *Comp. Vis. Im. Understanding* 114, 942–951.
- Hecht, E., 1998. *Optics*, third edition Addison Wesley Longman.
- Huynh, C.P., Robles-Kelly, A., Hancock, E., 2010. Shape and refractive index recovery from single-view polarisation images. In: *Proc. CVPR*, pp. 1229–1236.
- Kadambi, A., Taamazyan, V., Shi, B., Raskar, R., 2015. Polarized 3D: High-quality depth sensing with polarization cues. In: *Proc. ICCV*.
- Karpel, N., Schechner, Y., 2004. Portable polarimetric underwater imaging system with a linear response. In: *Proc. SPIE*, 5432.
- Miyazaki, D., Kagesawa, M., Ikeuchi, K., 2004. Transparent surface modelling from a pair of polarization images. *IEEE Trans. Patt. Anal. Mach. Intell.* 26, 73–82.
- Miyazaki, D., Tan, R.T., Hara, K., Ikeuchi, K., 2003. Polarization-based inverse rendering from a single view. In: *Proc. ICCV*, 2, pp. 982–987.
- Morel, O., Stolz, C., Meriaudeau, F., Gorria, P., 2006. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Appl. Opt.* 45, 4062–4068.
- Ngo, T., Nagahara, H., Taniguchi, R., 2015. Shape and light directions from shading and polarization. In: *Proc. CVPR*, pp. 2310–2318.
- Rahmann, S., 2003. Reconstruction of quadrics from two polarization views. In: *Proc. IbPRIA*, pp. 810–820.
- Rahmann, S., Canterakis, N., 2001. Reconstruction of specular surfaces using polarization imaging. In: *Proc. CVPR*, pp. 149–155.
- Saito, M., Sato, Y., Ikeuchi, K., Kashiwagi, H., 1999. Measurement of surface orientations of transparent objects using polarization in highlight. In: *Proc. CVPR*, 1, pp. 381–386.
- Saracchini, R.F., Stolfi, J., Leitão, H.C.G., Atkinson, G.A., Smith, M.L., 2012. A robust multi-scale integration method to obtain the depth from gradient maps. *Comp. Vis. Im. Understanding* 116, 882–895.
- Smith, W., Ramamoorthi, R., Tozza, S., 2016. Linear depth estimation from an uncalibrated, monocular polarisation image. In: *Proc. ECCV*, pp. 109–125.
- Taamazyan, V., Kadambi, A., Raskar, R., 2016. Shape from mixed polarization. *arXiv:1605.02066*.
- Wolff, L.B., 1994. Diffuse-reflectance model for smooth dielectric surfaces. *J. Opt. Soc. Am. A* 11, 2956–2968.
- Wolff, L.B., Boult, T.E., 1991. Constraining object features using a polarisation reflectance model. *IEEE Trans. Patt. Anal. Mach. Intell.* 13, 635–657.
- Woodham, R.J., 1980. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* 19, 139–144.
- Zhang, W., Smith, M.L., Smith, L.N., Farooq, A., 2016. Gender recognition from facial images: two or three dimensions? *J. Opt. Soc. Am. A* 33, 333–344.

¹ The seemingly random pattern near the middle of the M-Estimator reconstruction is due to text printed on the surface.