

Mutual Illumination Photometric Stereo

by
Christopher Powell

Submitted for the degree of Doctor of Philosophy

UNIVERSITY OF EAST ANGLIA

School of Computing Sciences

January 2018

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

“Lok'tar ogar!”

- Garrosh Hellscream

Abstract

Many techniques have been developed in computer vision to recover three-dimensional shape from two-dimensional images. These techniques impose various combinations of assumptions/restrictions of conditions to produce a representation of shape (e.g. surface normals or a height map). Although great progress has been made it is a problem which remains far from solved. In this thesis we propose a new approach to shape recovery - namely 'mutual illumination photometric stereo'. We exploit the presence of colourful mutual illumination in an environment to recover the shape of objects from a single image.

Acknowledgements

Firstly I would like to thank my supervisor, Prof. Graham Finlayson, for his support and understanding, his patience and the guidance offered to me throughout the course of my research. Thank you for dragging me along to the finish line.

I would also like to thank Prof. Mark Drew for sharing his knowledge and for providing me with the opportunity to work in his laboratory at Simon Fraser University, it was a great experience which I will never forget.

My sincere thanks also to Dr. Stephen Laycock for providing support with 3D printing and to Dr. Mike Harris for his assistance with various technical matters and accompaniment in shenanigans.

Lastly I would like to thank my friends and family for keeping me connected with life outside of academia, the escapism has been invaluable. Special thanks to Rachel for her love and support. Thank you, I would not have finished this thesis without it.

Contents

Abstract	ii
Acknowledgements	iii
List of Figures	vi
List of Tables	xiv
Publications	xv
Glossary	xvi
Variables	xvii
1 Introduction	1
1.1 Shape recovery	3
1.2 Mutual illumination	8
1.3 Thesis preview	11
2 Background	15
2.1 Shape recovery overview	15
2.2 Photometric Stereo	23
2.2.1 Shape from shading	25
2.2.2 Classic photometric stereo	28
2.2.2.1 Surface normals to height maps	31
2.2.3 Refined photometric stereo	35
2.2.4 High-end photometric stereo	36
2.2.5 Colour photometric stereo	38
2.2.5.1 High-end colour photometric stereo	42
2.3 Mutual illumination	43
2.4 Implementation improvements	48
2.4.1 Calibration	48
2.4.2 Specularity	58
2.4.3 Photometric stereo in the presence of highlights	60

3	Engineering an environment	66
3.1	Potential environments	67
3.2	Performance metric	69
3.3	Radiosity rendering	71
3.4	Synthesising data	75
3.5	Results	77
3.6	Conclusion	80
4	Data sets	81
4.1	A novel dataset	83
4.1.1	Papercraft Objects	84
4.1.2	3D-Printed objects	86
4.2	Conclusion	90
5	Shape in a box	92
5.1	Calibration	93
5.2	Spherical harmonic rendering	94
5.3	Experiments and results	101
5.3.1	Image capture	101
5.3.2	Comparing to ground truth	103
5.3.3	Surface normal accuracy	107
5.3.4	Conversion to height maps	111
5.3.5	Height map accuracy	113
5.4	Conclusion	115
5.5	Recovered shapes (images)	117
6	Addressing problems	124
6.1	Removal of calibration	124
6.1.1	Exploiting projection	127
6.1.2	Scene geometry	129
6.1.3	Results	131
6.1.4	Conclusion	134
6.2	Specular highlights	136
6.2.1	Pre-processing	136
6.2.1.1	Detection and removal	138
6.2.1.2	Inherent removal	142
6.2.2	Conclusion	146
7	Conclusion	148
7.1	Thesis contributions	148
7.2	Future work	150
7.3	Final remarks	152

Bibliography	153
---------------------	------------

List of Figures

1.1	Top: Image of the author. Bottom: Recovered 3D model using mutual illumination photometric stereo.	2
1.2	A coffee mug. The true key contributor to this thesis.	3
1.3	Pixels y_1 and y_2 correspond to the same point within a scene. Rays projected through the focal points O_1 and O_2 of each camera intersect at point x , the 3D position of the point. Performing this projection for all points gives a full 3D recreation of the scene (Image by Wikipedia user KYN).	4
1.4	At the tail of the arrow is a point on the curved surface. The blue rectangle is the tangent plane to the surface at that point. The ‘surface normal’ is perpendicular to that plane and is indicated by the direction of the arrow. It is essentially, the direction that the surface is facing towards at that point (Image by Wikipedia user Patrick87).	5
1.5	Lambertian reflectance - The strength of observed light is proportional to the angle between the surface normal and the direction of the light source $\cos(\theta)$	6
1.6	An example of a high-end photometric stereo set-up, taken from [1].	7
1.7	Mutual illumination in action - the white walls and ceiling appear pink due to light reflected by the red floor (Note - this is a synthetic image but serves as a good illustration).	8
1.8	Light source \mathbf{e} bounces around the recess before exiting. This leads to the image captured by the camera recording higher intensity values than expected by our imagined simple lighting model.	9
1.9	a) The true shape of the recess. b) The recovered shape. Figure taken from [2].	10
1.10	A mirrored sphere inside our colourful capture environment. Note the abundance of mutual illumination received from the interior walls.	11
1.11	Left: Simple constructed geometric shape (tetrahedron). Right: 3D-printed object with complex shape.	12
1.12	A 3D recovery of the object in Figure 1.11.	12
1.13	Top: Object exhibiting specular highlights. Bottom: Recovered 3D shape.	14

2.1	A structured light shape recovery system. A shows the hardware setup, two cameras and a projector. B and C show two captured images of different patterns projected onto the object. D shows a reconstructed point cloud (Image adapted from work by the Taubin Group at Brown University).	16
2.2	Top: Image of the infrared pattern projected by a Kinect. Bottom: height map of the scene as captured by the Kinect (Image adapted from work by Geospatial Modeling & Visualization).	17
2.3	a) and b) are images captured from different positions. c) shows the resulting disparity/height map and d) is the height map shown in 3D (a) to c) adapted from work by Ugo Capeto). d) is an original mesh produced from c) using Matlab.	18
2.4	Images are captured at different viewpoints and projected onto a grid of voxels. Intersections determine the color for a voxel [3]	19
2.5	Shape reconstruction for a toy dinosaur for different voxel grid resolutions [3].	19
2.6	Adapted from “A Theory of Shape by Space Carving” by Kutulakos and Seitz [4]. Camera views determine a minimum bounding voxel grid (the cuboid) for an object. At each voxel the visible images are tested for photo-consistency. The currently active voxel and camera positions of active images are shown in orange.	20
2.7	Shape recovery of a human hand via space carving using a hundred input images [4].	21
2.8	Time-of-flight systems: Infrared light is emitted. Time taken for the same pulse of light to return gives a measurement of distance. Thus a height map of a scene/object can be estimated (Image taken from Stemmer Imaging).	22
2.9	Left: Colour image. Right: The same scene imaged by a time-of-flight system. (Adapted from http://www.mip.informatik.uni-kiel.de)	22
2.10	Left: A laser scanner attached to a gantry. Right: The 3D shape recovered by the laser scanner (Image adapted from The Digital Michelangelo Project).	23
2.11	Left: Experiment setup used in [5]. Right: Image captured by the camera.	24
2.12	A hemisphere projected into a two-dimensional space (Adapted from [6]).	26
2.13	An example of a reflectance map for a Lambertian sphere (Credit to James O’Shea - University of California).	26
2.14	A reflectance map with $R(x, y) = 0.6$ highlighted (Credit to James O’Shea - University of California).	27
2.15	Left: Greyscale images. Right: Estimated shapes (Credit to James O’Shea - University of California).	28
2.16	A sphere viewed from the same position. As the angle θ between surface normal vector \mathbf{n} and light direction vector \mathbf{e} increases, less light reaches the front of the sphere.	29

2.17	In parts a light sources are labelled as l_1, l_2, l_3 in the scene diagrams, below each diagram is the corresponding captured image of the object. Surface normals in b are displayed using the convention of normal maps in graphics (red = positive x-axis, green = positive y-axis, blue = positive z-axis. The height map in c is in grayscale with white = maximum height, black = minimum height). The same colour coding is used throughout this thesis.	31
2.18	The blue vector n is normal to the red surface. The green vectors a and b lie on the tangent plane (Modified version of the image in Figure 1.4).	33
2.19	A complex shape causes self-occlusion from a red, infinitely distant point light source. Surface normals with identical $[x, y, z]$ components receive different lighting.	35
2.20	Specular spheres used to estimate light parameters in the work of Liao et al. [55].	36
2.21	The dome of cameras and light sources used to reconstruct shape in [1].	37
2.22	In part a light sources are labelled as l_1, l_2, l_3 in the scene diagram, right of the diagram is the corresponding captured image of the object. Parts b and c show recovered surface normals and a height map using the same conventions as Figure 2.17.	39
2.23	Luminance profiles of identical white and black rooms taken with different light intensities. (Figure taken from [7]).	44
2.24	Rendered versions of the white and black rooms in the work by Gilchrist [7]. The renders using only direct illumination are identical, but with mutual illumination added it becomes clear which room is which. (Figure adapted from [8]).	45
2.25	For the room pair on the left one can distinguish between surface colour and illumination, however this is not possible with the room pair on the right [8].	46
2.26	The radiance of shape A is amplified by mutual illumination in comparison with shape B.	47
2.27	In [9] the calibration sphere and objects are painted to possess the same reflectance and the surface normals are recovered.	49
2.28	Example of a calibration step taken from [10]. A sample of object material is placed inside a calibration target (the square in the left image) before capturing video of that object undergoing some deformation (right).	50
2.29	An example of a bas-relief taken from [11]. From a distance the sculpture appears to have full three-dimensional depth (left and centre), however up close it is actually quite flat (right).	51
2.30	From [11] - the same image(left) can be generated from infinite variations of light direction and object shape (right).	51

2.31	Top: Three of twelve input images of a globe. Middle: Shape recovery with the Lambertian reflectance model (Note the bump caused by misinterpretation of highlights). Bottom: Shape recovery from [12] with the Torrance and Sparrow reflectance model.	52
2.32	Demonstrating the effect of interreflections in a non-convex object. The fold in the centre (left) is expected to appear darker than the rest of the shape as it is at a greater distance from the light source, however due to light bouncing around within the crease it actually has greater intensity. The chart on the right shows the intensity profile across a row of the image [13].	53
2.33	Adapted from [13]: a) - d) are input images. e) is the shape recovered by uncalibrated photometric stereo. f) is the shape recovered by Chandraker et al. using mutual illumination to resolve the GBR ambiguity.	54
2.34	The master device remains in place. The slave device is moved and both capture an image using the flash from the slave. Light position and direction can thus be calculated to calibrate the system without using an explicit calibration step [14].	55
2.35	From Lu et al. [15] - pixel intensity profiles given a moving light source. Points A,B and C share the same surface reflectance. A and C also share the same surface normal, thus they share the same intensity profile. D has the same surface normal as A/C but different surface reflectance, thus a different intensity profile (the sharper peaks for D indicate a specular component - see Section 2.4.2).	56
2.36	Left: Colourful illumination cast by a monitor onto a human face. Right: An estimated height map recovered using the approach in [16].	56
2.37	Using a computer screen to provide a known light source - image taken from [17].	57
2.38	Left - An elastomer layer is pressed against a surface (in this case a human hand). Inside the device multi-spectral light sources and a high resolution camera capture an image for colour photometric stereo. Right - The recovered shape of the surface. Images taken from [18].	58
2.39	A diffuse sphere (left) versus a glossy one (right). Note that the highlight is the colour of the light source, this is the case for almost all glossy surfaces.	59
2.40	Here we illustrate the effect that outlying pixel values can have in photometric stereo. The specularity present on the second object clearly causes a large error in the resulting recovery (Images adapted from [19]).	60
2.41	Shiny surfaces reflect light more strongly according to the angle of incidence at a surface normal. If the same point in this diagram is imaged by the two cameras, the corresponding pixel value will have higher intensity in camera 2's image.	62

2.42	Left: Original image with highlights. Middle: Specular pixel mask. Right: Specular free image [20].	64
3.1	Examples of synthesised images. Left: Close-up of a rotated cube. Right: A wide-shot of the cube positioned inside a colourful box. Each face of the cube has the same RGB value.	67
3.2	A white sphere subject to mutual illumination. Red light is causing surface normals which have opposing x-axis components to appear identical. Thus we will be unable to correctly estimate shape in this lighting environment.	69
3.3	A synthesised sphere with Lambertian reflectance subject to colourful mutual illumination. Colour varies smoothly in correspondence with surface normals.	70
3.4	A low polygon (hence the shadows appearing stepped/aliased) environment over successive radiosity rendering passes. The room becomes brighter as the transfer of energy is modelled for successive iterations of the radiosity rendering equation (Image adapted from work by Hugo Elias).	72
3.5	Calculating the energy received at a point x from some other point x' . N refers to the surface normal at a point, d is the distance between the two, θ is the angle between a surface normal and a line connecting the two points. Repeating this for all other x' in the scene will give us the total radiosity at x (Image adapted from work by Hugo Elias).	73
3.6	Calculating a hemispherical view factor (Image by Wikipedia user Jheald).	74
3.7	Views from 3DS Max of the experiment setup. Top - A “daylight system” provides lighting whilst a camera points directly into the box. Bottom - From the perspective of the camera we see a sphere inside the box (this is the view that is rendered to produce our data).	76
3.8	Results of the synthetic experiments on triangular boxes of various dimensions. Four different side lengths were each tested with the same varying heights.	78
3.9	Examples of synthetic shape recoveries performed on spheres in environments with different condition numbers. A better conditioned transformation matrix gives a more accurate result.	79
3.10	The end result of this chapter - a real box designed for mutual illumination photometric stereo.	80
4.1	3D printing offers real objects with almost perfect ground truth shape data. Left: Original 3D model file for the ‘Bust’ object. Right: 3D-printed counterpart.	81
4.2	Images of experiment subjects (3D-printed objects in <i>italics</i> , papercraft in normal type, abbreviated names given in (brackets)). Top row, left to right: <i>Bust</i> , Cone, <i>Face</i> , 4-sided pyramid (Pyr4), 5-sided pyramid (Pyr5). Bottom row, left to right: 6-sided pyramid (Pyr6), <i>Sphere</i> , Star, Tetrahedron (Tetra), <i>Trachea</i> (<i>Trach</i>).	82

4.3	Top row: a simple papercraft object. Bottom: a 3D-printed object. a) Source files, a papercraft template and a 3D model file. b) Captured images. c) left = true normal map, right = recovered normal map. d) left = true height map, right = recovered height map. . . .	84
4.4	Top: Template for a tetrahedron. Each edge of the tetrahedron is the same unit length. Bottom: The assembled counterpart. The surface colour is due to the mutual illumination provided by our box environment.	85
4.5	Top: Template for a four-sided pyramid. Bottom: The assembled counterpart.	85
4.6	Top: Template for a five-sided pyramid. Bottom: The assembled counterpart.	86
4.7	Stereolithography (SLA): A layer of resin is bound together by the laser according to the object design, the platform moves down a layer and the process repeats (Image taken from proto3000).	87
4.8	Fused deposition modelling (FDM): 1 - nozzle applies material, 2 - material has been deposited according to the object design, 3 - movable table (Image by Wikipedia user Zureks).	88
4.9	Typical 3D-prints. Left: Produced by FDM (Taken from Pro Image 3D). Right: Produced by SLA (Image taken from Fibilo).	89
4.10	Left: Original 3D model file for the ‘Face’ object. Right: 3D-printed counterpart.	90
4.11	Left: Original 3D model file for the ‘Trachea’ object. Right: 3D-printed counterpart.	90
5.1	Left: A captured image of a mirrored sphere inside the box environment. Right: The synthesised appearance of a perfect Lambertian sphere in the box environment.	95
5.2	Visualisation of spherical harmonic basis functions and their effect in rendering (Adapted from [21]).	99
5.3	A 3D model rendered using spherical harmonic lighting (Adapted from [21]).	100
5.4	Left: The experiment set-up. Centre: Captured calibration images of a mirrored sphere at different exposure levels. Right: The synthesised appearance of a perfect Lambertian sphere in the same environment.	102
5.5	Illustration of the experiment set-up for Shape From Color.	103
5.6	Left: Segmented captured image. Middle: Generated true height map. Right: Estimated height map.	104
5.7	Left: Segmented captured image. Middle: Transformed true height map. Right: Recovered height map.	107
5.8	Left: Transformed true normal map. Right: Recovered normal map.	107
5.9	Visualisation of the angular error for four recovered sets of surface normals. Error is scaled from black (lower) to white (higher).	110
5.10	Image showing the combined, quantised error for the objects in Figure 5.9. Error is scaled from black (lower) to white (higher).	111

5.11	a) Captured image. b) Recovered normal map. c) Rendering of recovered height map. d) Textured render of recovery.	115
5.12	<i>Bust</i> (3D-printed figurehead)	117
5.13	<i>Cone</i> (Papercraft cone)	118
5.14	<i>Face</i> (3D-printed face)	118
5.15	<i>Pyr4</i> (Papercraft 4-sided pyramid)	119
5.16	<i>Pyr5</i> (Papercraft 5-sided pyramid)	119
5.17	<i>Pyr6</i> (Papercraft 6-sided pyramid)	120
5.18	<i>Sph</i> (3D-printed sphere)	120
5.19	<i>Star</i> (Papercraft star)	121
5.20	<i>Tetra</i> (Papercraft tetrahedron)	121
5.21	<i>Trach</i> (3D-printed human trachea)	122
5.22	<i>Tetra</i> (Papercraft tetrahedron - captured on a rooftop.) Angular error: 3.31° Height accuracy (using Jacobi integration method): 96.02%.	122
5.23	(3D-printed keyring - captured in the author's garden.) Angular error: 9.26° Height accuracy (using Jacobi integration method): 91.03%.	123
6.1	Light rays enter the box and reflect off the coloured box walls. The fragments represent light ray strength, more fragments = weaker.	126
6.2	Two images of the same cube using different projection methods. Left - Perspective projection causes the rear of the cube to appear smaller due to it being further from the camera. Right - In orthographic projection distance does not affect appearance.	127
6.3	If the assumption of orthographic projection were true, we would not be able to see the interior walls of the box in this image. The walls are vertical and thus under orthographic projection should be parallel to the vertical orientation of the camera - we should see only the narrow tops of the box walls.	128
6.4	Representing the 3D space in which the box exists e.g. In a captured image, point A would have $[x; y; z]$ coordinates roughly equal to $[0 : 5; 1 : 0; 1 : 0]$	129
6.5	a) An example of a captured image. b) Segmenting the walls of the box. c) A synthesised Lambertian sphere, rendered using the strength and direction of the light sources estimated from b).	130
6.6	Captured images for all objects and height maps for the <i>Face</i> object.	132
6.7	Preliminary results demonstrating the resilience of our method to less ideal environments. A lampshade illuminated by the indirect illumination from three colourful t-shirts.	135
6.8	Top: Captured image. Bottom: Recovered shape.	137
6.9	Top: Original image. Middle: Specular pixels detected and removed. Bottom: Inpainted image.	139
6.10	Top: Height maps. Bottom: Textured meshes. Left: Without pre-processing. Right: With pre-processing.	139

6.11	Top: Original image. Middle: Specular pixels detected and removed. Bottom: Inpainted image.	140
6.12	Top: Height maps. Bottom: Textured meshes. Left: Without pre-processing. Right: With pre-processing.	140
6.13	A specular lobe consists of the specular and glossy components above. The wider the lobe, the wider the range of surface normals which will produce pixels with specular components (Image taken from Virial International).	141
6.14	Top left: Synthesised image with a specular highlight. Top right: Highlight is removed. Bottom left: Inpainted image. Bottom right: Synthesised image without specular highlight.	142
6.15	Recovered shapes. Left: With specular highlight. Right: After inpainting.	142
6.16	RGB colour space is rotated such that one of the axes aligns with the light colour \mathbf{s} (Image adapted from [22]).	143
6.17	Left: RGB image. Middle: Specular image (S). Right: Diffuse image ($ U + V $).	144
6.18	Shape recovered from the UV channels using a two-dimensional modification of our shape recovery process.	145
6.19	Images of a shiny bowl. Left: No filter. Right: Using an orange lens filter.	146
6.20	Recovered shape of a shiny bowl. Left: Recovery with highlight present. Right: Recovery after highlight removal.	146
7.1	Left: Input image. Right: The room is segmented according to 3D scene understanding [23].	151

List of Tables

3.1	Condition number, $k(F)$, given by triangular boxes with varying dimensions.	77
5.1	Standard deviation of ground truth papercraft height maps.	105
5.2	Standard deviation of ground truth 3D-printed height maps.	106
5.3	Accuracy of estimated surface normals represented as average angular error.	108
5.4	Height map accuracies as percentage values. “Shapelet” is Kovese’s method [24], “FC” is Frankot & Chellappa’s algorithm [25] and “Jacobi” is our Jacobi-type method [26].	114
6.1	Table 1: Angular error of recovered surface normals.	133
6.2	Table 2: Height map accuracies as percentage values.	133

Publications

The following are publications by the author related to this work:

- C. Powell and G. D. Finlayson. Mutual illumination photometric stereo without calibration. Color and Imaging Conference (*CIC*) 2015
- G. D. Finlayson and C. Powell. Shape in a Box. 4th Color and Photometry in Computer Vision Workshop (*CPCV*) - European Conference on Computer Vision (*ECCV*) 2014.
- Work in this thesis was also presented at the Computing Sciences Research Day at the University of East Anglia (UEA) in 2015. Out of a field of 30 papers, this work won the second place prize for best paper.

Glossary

2D	Two-dimensional
3D	Three-dimensional
GBR	Generalised B as R elief
RGB	R ed G reen B lue color space
SFC	S hape F rom C olor
SiaB	S hape in a B ox

Variables

boldfont	=	a vector
<i>CAPITALISED</i>	=	a matrix
α	=	surface albedo
b	=	colour of reflected light
<i>B</i>	=	concatenated colour of reflected light vectors
<i>c</i>	=	an RGB pixel value
$c_{L,S}$	=	Lambertian or Specular component of RGB pixel value
<i>C</i>	=	concatenated RGB pixel values
e	=	normalised light direction
<i>E</i>	=	concatenated normalised light directions
<i>F</i>	=	3 x 3 transform from surface normals to RGB values
<i>G</i>	=	transform from RGB values to surface normals (F^{-1})
n	=	normalised surface normal
$\mathbf{n}_{x,y,z}$	=	x,y or z component of a surface normal
<i>N</i>	=	concatenated normalised surface normals
$N(x, y)$	=	surface normal map
$Z(x, y)$	=	height map

Chapter 1

Introduction

To fully understand how a camera image was produced, we must know everything about the conditions in which it was captured. This includes such information as - the lighting present in a scene, the sensitivities of the capture device and the main focus of this thesis - the geometry of image subjects. Knowing the 3D shape of objects in an image is a key part in solving how the image was formed. Thus 3D shape recovery is of potential relevance to a huge swathe of computational vision problems. Recovery of 3D shape also has many direct applications in work which requires 3D models (e.g. video games and animation). Accordingly, recovery of three-dimensional shape from two-dimensional images has been an active and highly important area of research since the inception of computer vision.

In this thesis we propose a new approach to shape recovery - namely ‘mutual illumination photometric stereo’.



FIGURE 1.1: Top: Image of the author. Bottom: Recovered 3D model using mutual illumination photometric stereo.

Typically shape recovery methods are either monetarily expensive, time consuming, require strict laboratory conditions or some combination of these barriers to entry. We aim to make shape recovery more convenient by requiring only a single image, moving closer towards unrestricted real-world conditions (an area of research which is surprisingly lacking) and avoiding cost prohibitive solutions.

In contrast to almost all previous photometric stereo work, we are not concerned only with direct illumination. We wish to estimate the 3D shape of objects by exploiting the effects of mutual illumination (light that has been reflected from

other surfaces). We begin in this chapter with a brief introduction to the topics of shape recovery and mutual illumination.

1.1 Shape recovery

When we capture an image, the three-dimensional (3D) world within view is compressed into two-dimensional (2D) space - that is, 3D geometry is projected onto a 2D image plane. Shape recovery research is concerned with enabling computers to rebuild the lost dimension.

The human vision system captures stereo images (two eyes) and by using stereo together with other cues can perform well at discerning the shape of objects within a scene. Stereo alone does not work well for smoothly varying surfaces, yet the object in Figure 1.2 is instantly recognisable as having a cylindrical shape. We can determine the object's shape using information such as - shading across its surface, depth cues within the scene (e.g. the shadow cast by the object onto the counter), the context of the scene (a kitchen), referencing against a library of prior visual experiences and (were you standing in the scene rather than viewing an image) the disparity between the images captured by each of your eyes (i.e. stereo vision). We are inherently able to determine the shape and size of an object placed in front of us, in this case a regular coffee mug.



FIGURE 1.2: A coffee mug. The true key contributor to this thesis.

So, how can we enable a computer to achieve the same result? In computer vision, there are two main approaches to estimation of 3D shape. Geometric methods use corresponding points across multiple images to determine their depth within the scene. For example, in Figure 1.3 the position from which the two images (the rectangles) were captured is known. If we find a corresponding point in each image then we can project rays from the camera positions outwards and their intersection tells us the real-world position of that point [27].

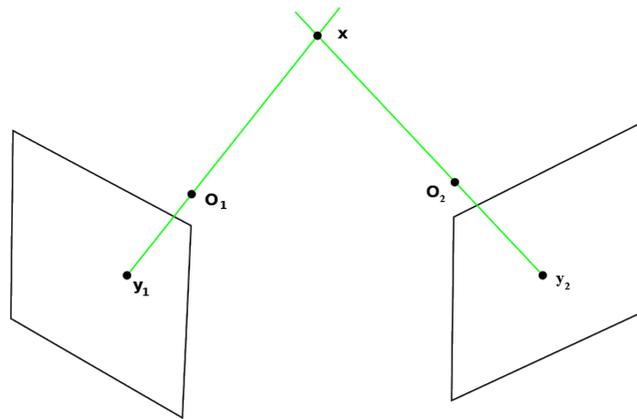


FIGURE 1.3: Pixels y_1 and y_2 correspond to the same point within a scene. Rays projected through the focal points O_1 and O_2 of each camera intersect at point x , the 3D position of the point. Performing this projection for all points gives a full 3D recreation of the scene (Image by [Wikipedia user KYN](#)).

The other main approach to shape recovery is the use of photometric methods. These rely on determining the relationship between lighting, object shape and surface reflectance. In this thesis we are interested in the photometric methods.

There are two important sets of information required to use photometry for shape recovery:

Firstly, the properties of lighting incident to the object - what colour are the light sources in the scene? How powerful are they? Which direction are they facing?

Secondly, we need information about the reflectance characteristics of the object itself - What is its surface colour? How strongly does it reflect light? Does it reflect light equally in all directions?

Essentially, if we image some object then the pixel values captured are equal to some fraction of illumination attenuated by the shape and reflectance of the imaged surface. For now, let us use simple, abstract terms,

$$\textit{Pixels} = \textit{Illumination} * \textit{Shape} * \textit{Reflectance}. \quad (1.1)$$

Thus given an image and information about illumination and reflectance, we can rearrange and solve for the unknown shape (In reality, of course, the process is more complicated but this description suffices for illustrative purposes),

$$\textit{Shape} = \frac{\textit{Pixels}}{\textit{Reflectance} * \textit{Illumination}}. \quad (1.2)$$

In the Lambertian reflectance model surfaces reflect light equally in all directions (i.e. they do not appear to be at all shiny). The strength of observed light is then some fraction of incident light governed solely by the direction of the surface (the ‘surface normal’ - Figure 1.4) as illustrated in Figure 1.5.

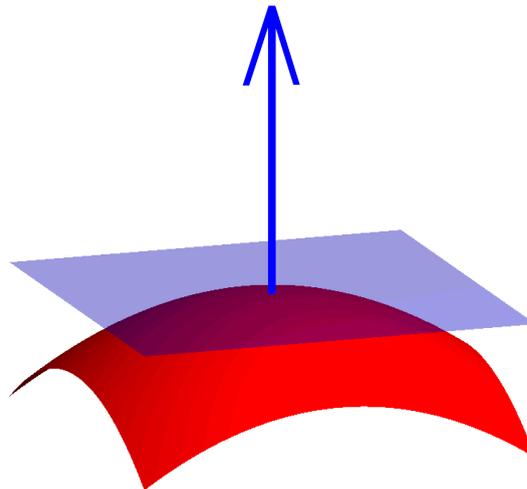


FIGURE 1.4: At the tail of the arrow is a point on the curved surface. The blue rectangle is the tangent plane to the surface at that point. The ‘surface normal’ is perpendicular to that plane and is indicated by the direction of the arrow. It is essentially, the direction that the surface is facing towards at that point (Image by [Wikipedia user Patrick87](#)).

It is common in computer vision and graphics to adopt the assumption that surfaces are Lambertian - though in reality no surface perfectly adheres to the Lambertian model. However, for many purposes it simplifies problems and still provides acceptable results.

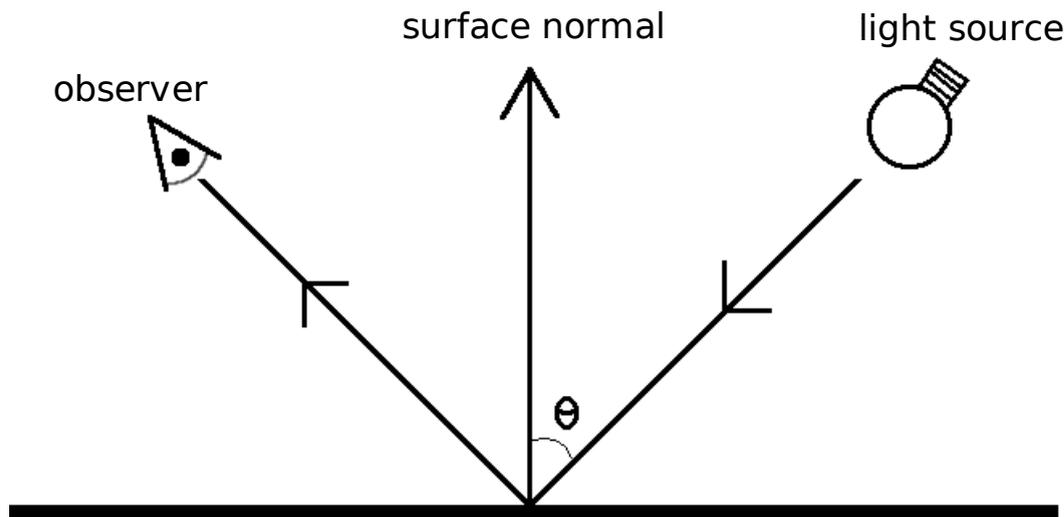


FIGURE 1.5: Lambertian reflectance - The strength of observed light is proportional to the angle between the surface normal and the direction of the light source $\cos(\theta)$.

For example, let us assume the Lambertian reflectance model for our shape recovery problem. We have instantly simplified the task - our *Reflectance* variable is now known, so let us remove it from our abstract equation,

$$Shape = \frac{Pixels}{Illumination}. \quad (1.3)$$

We now have only one unknown quantity remaining in our equation. If we can measure the light sources in the scene (*Illumination*) then we should be able to solve for *Shape*.

This approach to shape recovery was first presented in the seminal works of Horn [28] and Woodham [2]. Using light sources with known directions they were able to estimate the 3D shape of Lambertian objects. Though of course, matters are more complicated than portrayed in our abstract equation. What if we do not have

information about light sources? What if the object is very shiny (and thus clearly not Lambertian)? What if there are unexpected pixel values (e.g. a complex object casts shadows on itself)? Indeed, the early works were quite fragile. As such, many refinements and novel methods for shape recovery have been developed since.

State of the art photometric methods for accurate recovery of object shape (Figure 1.6) use large amounts of data gathered by elaborate arrangements of equipment (sometimes involving hundreds of lights and cameras). However such methods are obviously limited by the requirements of their specialised set-up (e.g. cost of components, space required by system, lack of portability, etc.).

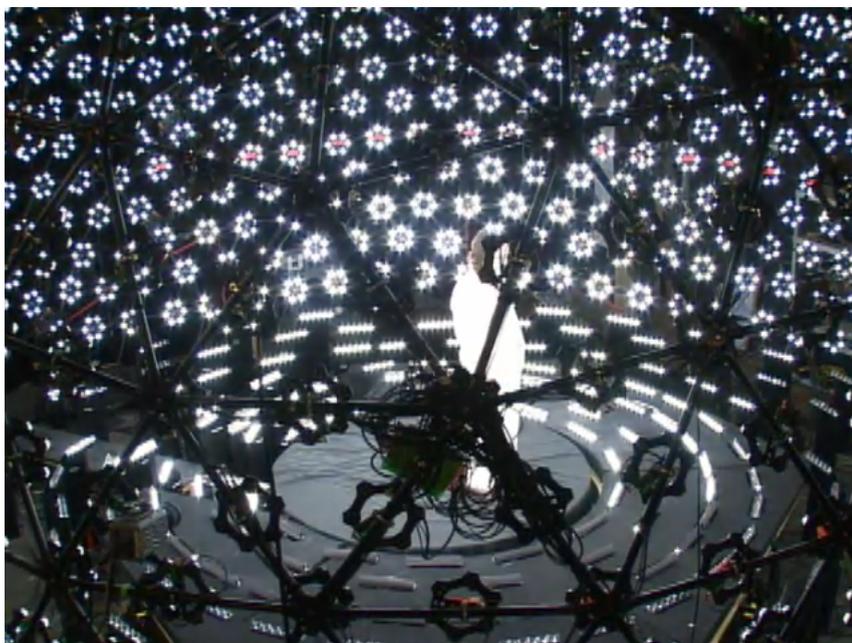


FIGURE 1.6: An example of a high-end photometric stereo set-up, taken from [1].

Throughout our work we focus on performing shape recovery from a single image using a normal digital camera, making our work highly flexible. Typically, shape estimation performed using a single image falls into the category of ‘shape from shading’. However, we classify our work as ‘photometric stereo’. Photometric stereo requires three (or more) images to be captured of an object subject to different light sources. In essence, we capture three grayscale images simultaneously - the red, green and blue colour channels of a single capture. Thus, our method

is a single-image variant of photometric stereo. This is made possible through the exploitation of colourful mutual illumination within a scene.

1.2 Mutual illumination

Mutual illumination (also referred to as ‘indirect lighting’ or ‘global illumination’) is light which has been reflected by one surface onto another (i.e. it does not come directly from a light source). For example, in Figure 1.7 we can clearly see mutual illumination in action - the red floor is influencing the appearance of the white walls.

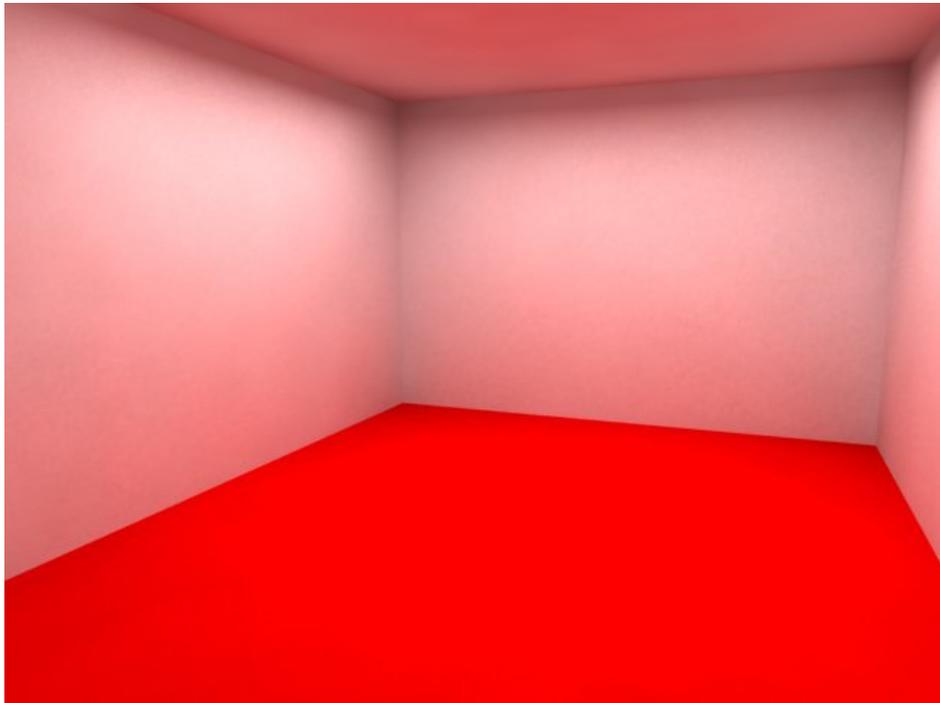


FIGURE 1.7: Mutual illumination in action - the white walls and ceiling appear pink due to light reflected by the red floor (Note - this is a synthetic image but serves as a good illustration).

The effect of mutual illumination can appear to be negligible in scenes with abundant direct light (e.g. outside on a sunny day). Indeed, it has been shown that to a large extent human vision naturally discounts the effect [29] of mutual illumination and instead attempts to perceive the original surface colour.

However, in enclosed spaces (e.g. a room) inter-reflections become much more measurable and can account for a large fraction of light present in the environment. In fact, rather than being discarded; in some circumstances mutual illumination acts as a cue in human determination of surface colour [8]. Yet the vast majority of work in computer vision fails to incorporate the effects of mutual illumination and this can be shown to cause inaccuracies [30].

As an example, let us consider a naive shape recovery algorithm. We want to estimate the 3D height of points of an object in a greyscale image (illustrated in Figure 1.8). We adopt the Lambertian assumption - we assume that shading in the image depends only on the cosine of the angle between the light source direction and the surface normals. For a simple, convex shape this will be true.

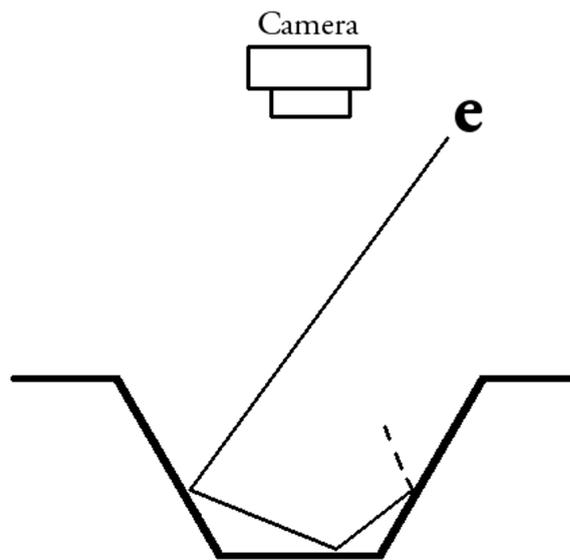


FIGURE 1.8: Light source \mathbf{e} bounces around the recess before exiting. This leads to the image captured by the camera recording higher intensity values than expected by our imagined simple lighting model.

However, the object shown in Figure 1.8 has a concave recess. The true shape at the recess is shown in Figure 1.9 (a) and the actual shape recovered by our algorithm is shown in (b). The shape in (b) is wrong because the algorithm has not accounted for the effect of light bouncing around inside the recess. This inter-reflection of light increases the intensity of recessed pixels [31], thus they do

not obey the Lambertian model and error is introduced into the shape recovery process.



FIGURE 1.9: a) The true shape of the recess. b) The recovered shape. Figure taken from [2].

The starting premise for this thesis is that if mutual illumination is present in an image; then we can use it to further our knowledge of how the image was formed. Rather than ignoring or removing the effects of mutual illumination, we will exploit its presence to provide the information required for our photometric stereo algorithm. In terms of our abstract equation (Equation 1.3) - it will give us the unknown *Illumination* component and thus we can solve for the shape.

Examples of potential final applications for our work could include shape recovery for security situations - The 3D shape of a person walking along an airport corridor with colourful walls could be recovered without the subject even being aware that it has happened. Another possible usage could be fieldwork in remote locations - as long as there is sufficient colourful mutual illumination present (which could be provided by a box as in this thesis) the only other requirement is a camera, thus no power sources are needed.

1.3 Thesis preview

The key contribution of this thesis is a new approach to 3D shape estimation. We demonstrate that the mutual illumination present in a single image can provide sufficient information for shape recovery. To achieve this we capture images of novel ground truth objects, use mutual illumination to estimate their 3D shape and measure the accuracy of our results. This section gives a brief, high-level overview of the thesis contents.

In Chapter 2 we present the existing literature. Our own work begins in Chapter 3. As mentioned previously, the effects of mutual illumination are more measurable in enclosed spaces. To this end we create an environment which is guaranteed to produce favourable conditions - a colourful box (Figure 1.10).

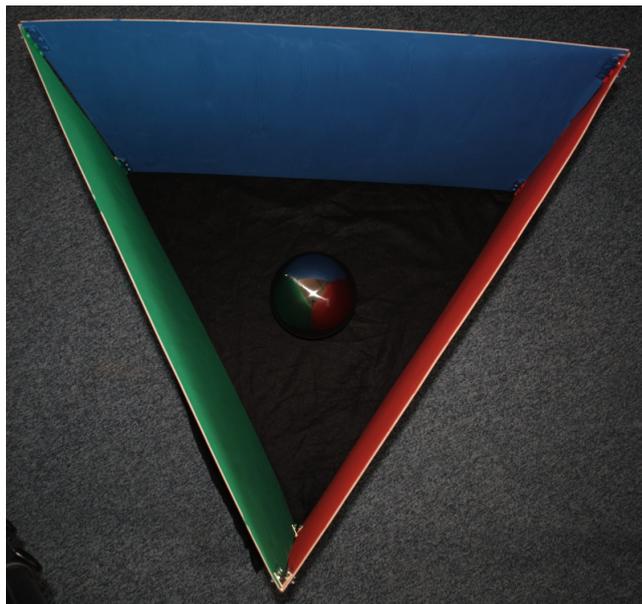


FIGURE 1.10: A mirrored sphere inside our colourful capture environment. Note the abundance of mutual illumination received from the interior walls.

Next we discuss the problem of acquiring ground truth data for shape recovery. In Chapter 4 we propose a novel approach to the ground truth problem. We present objects for which the ground truth shape is known a priori - simple constructed geometric shapes and more complex 3D-printed objects (Figure 1.11).



FIGURE 1.11: Left: Simple constructed geometric shape (tetrahedron). Right: 3D-printed object with complex shape.

In Chapter 5 we detail the process of mutual illumination photometric stereo itself. In brief - we capture an image of a mirrored sphere inside a colourful box (Figure 1.10). This calibrates the system - in terms of our simplistic equation (Equation 1.3) it provides us with the required *Illumination* variable. We then capture images of our objects inside the box. Thus both *Pixels* and *Illumination* are known and as per Equation 1.3 we can solve for the unknown *Shape* variable (Figure 1.12).

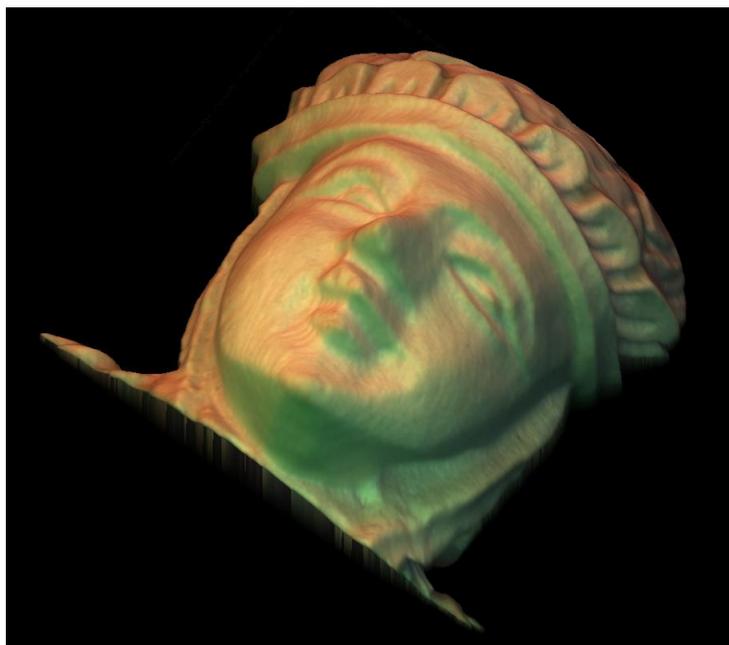


FIGURE 1.12: A 3D recovery of the object in Figure 1.11.

The accuracy of recovered shapes is measured against their ground truth values. We also produce results for the same objects using the established ‘shape from color’ method. We determine that our mutual illumination photometric stereo method recovers shape to the same level of accuracy.

In Chapter 6 we address some common problems in shape recovery with respect to our proposed system. Firstly, we remove the need for calibration (we discuss several problems with this restriction in Chapter 2). That is, we calculate the *Illumination* variable without needing an explicit measurement step.

Secondly, we attempt to recover the shape of objects which violate the Lambertian reflectance assumption. Such objects produce erroneous results as without modification any anomalous pixel values are assumed to be caused by shape (Figure 1.13). To our knowledge there is no previous work which handles objects with specular highlights in single-image photometric stereo systems. We implement two possible approaches and propose the best courses of action for further work in the area.

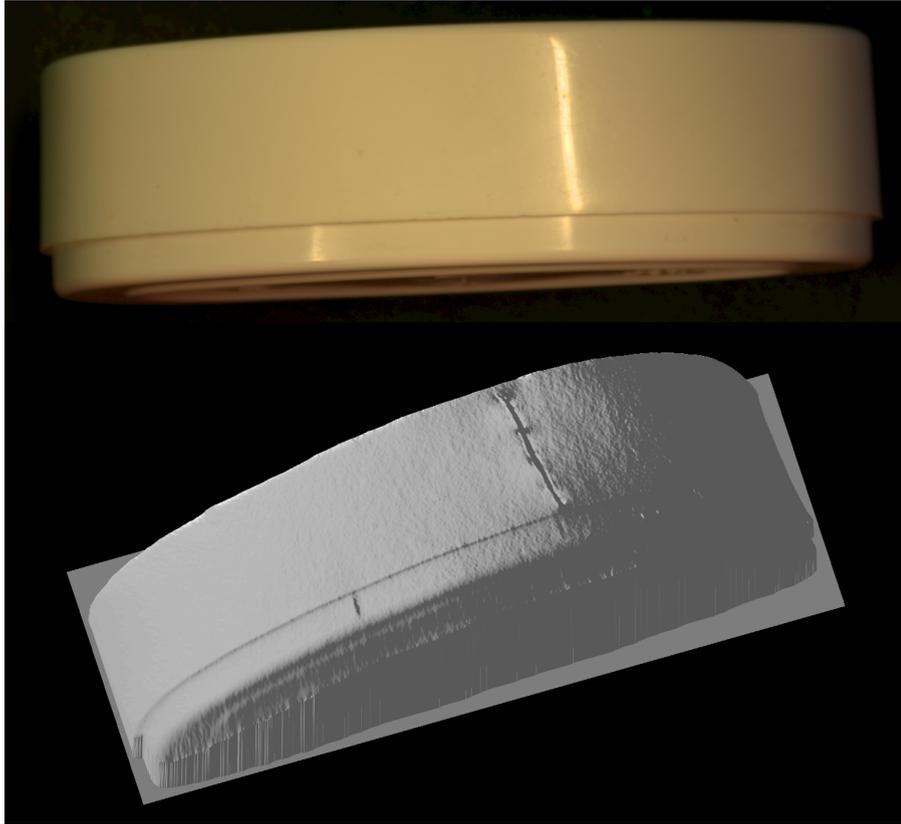


FIGURE 1.13: Top: Object exhibiting specular highlights. Bottom: Recovered 3D shape.

Finally, in Chapter 7 the thesis is summarised and opportunities for further work stemming from the contents are discussed.

Chapter 2

Background

In this chapter we examine related work which precedes our own. We begin with a general overview of shape recovery techniques before focusing on photometric stereo and subsequently colour photometric stereo. This leads into a brief look at the history of mutual illumination in computer vision (an area which has attracted surprisingly little scrutiny). Finally we will introduce the background related to Chapter 6 which addresses some issues in our work.

2.1 Shape recovery overview

In shape recovery, we give as input to an algorithm one or more two-dimensional images and we would like the three-dimensional shape of objects in a scene to be returned. There is a wide range of existing approaches to shape estimation - each impose various assumptions and restrictions on capture conditions. Ultimately, we would like to be able to capture a single image of any object, subject to any scene conditions using readily-available, inexpensive hardware and acquire accurate shape information as a result. However, the amount of variables involved in such a task makes this a very difficult task. In this section we give a brief overview of important shape recovery methods.

Structured light

Structured light systems require one or more cameras and some sort of light-projection hardware. Specific light patterns are cast onto an object. Shape is then recovered according to deformation of the patterns caused by the object geometry [32–34].

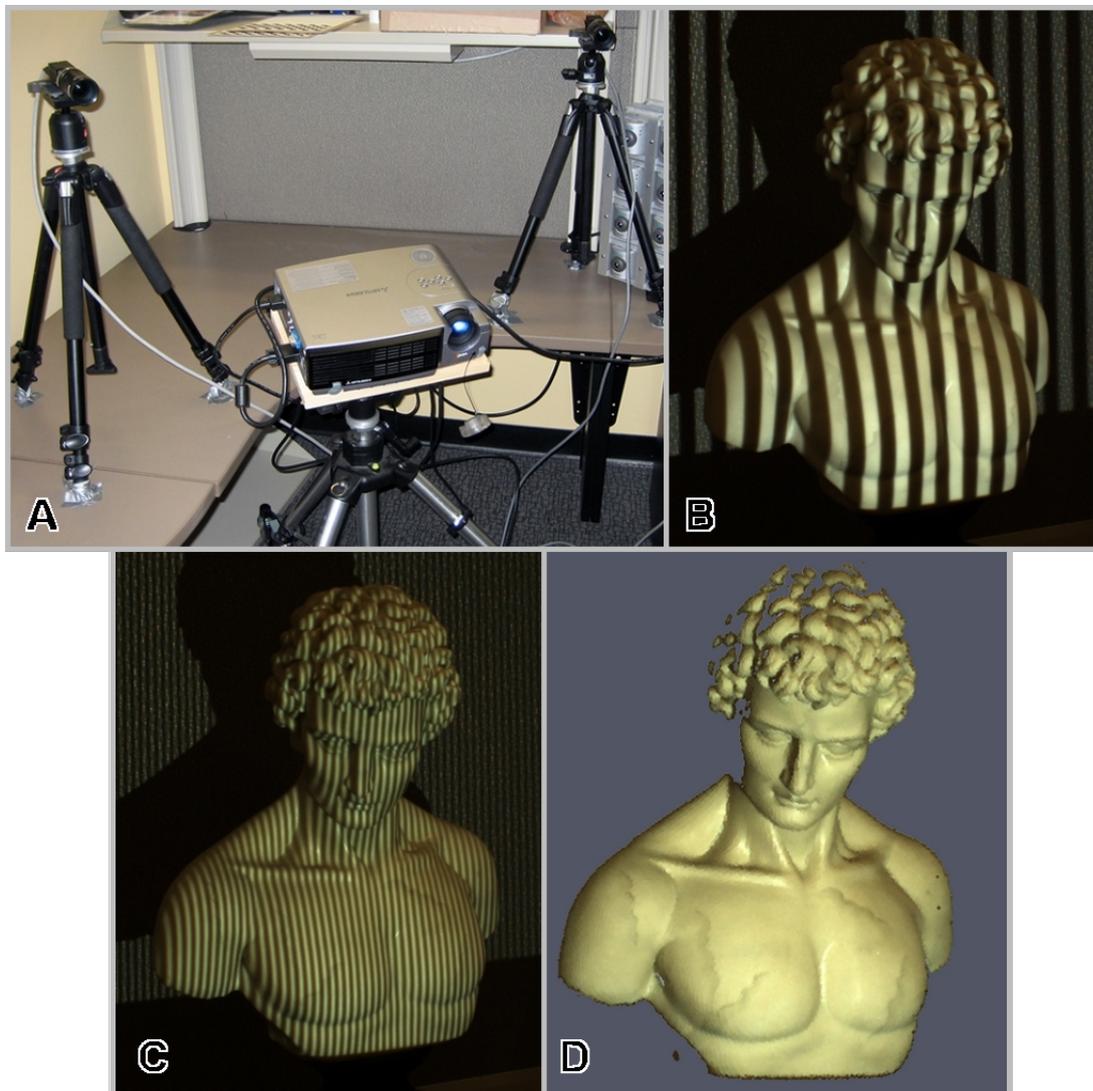


FIGURE 2.1: A structured light shape recovery system. **A** shows the hardware setup, two cameras and a projector. **B** and **C** show two captured images of different patterns projected onto the object. **D** shows a reconstructed point cloud (Image adapted from work by [the Taubin Group at Brown University](#)).

A prominent example of a structured light system can be seen in the Microsoft Kinect (Figure 2.2). This device projects a scattered pattern of infrared dots onto

a scene. The infrared light arrives back at the Kinect's sensor and a height map is calculated from the displacement of the dots - A height map being an image $Z(x, y)$ with each coordinate corresponding to some height value (typically white represents pixels closest to the camera and black those furthest away).

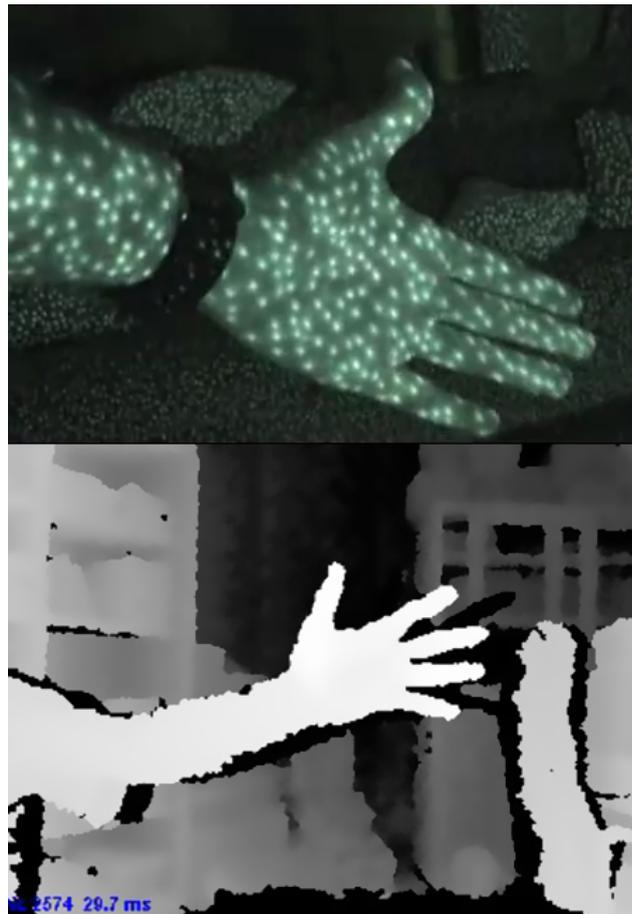


FIGURE 2.2: Top: Image of the infrared pattern projected by a Kinect. Bottom: height map of the scene as captured by the Kinect (Image adapted from work by [Geospatial Modeling & Visualization](#)).

Shape recovery from multiple views

In computer stereo vision [35–38] two images of the same object or scene are captured from different positions (just like in human vision). The disparity (distance) between corresponding pixels is calculated which results in a height map (Figure 2.3). The idea being that points further away from the cameras undergo a larger change in image position than those nearby.

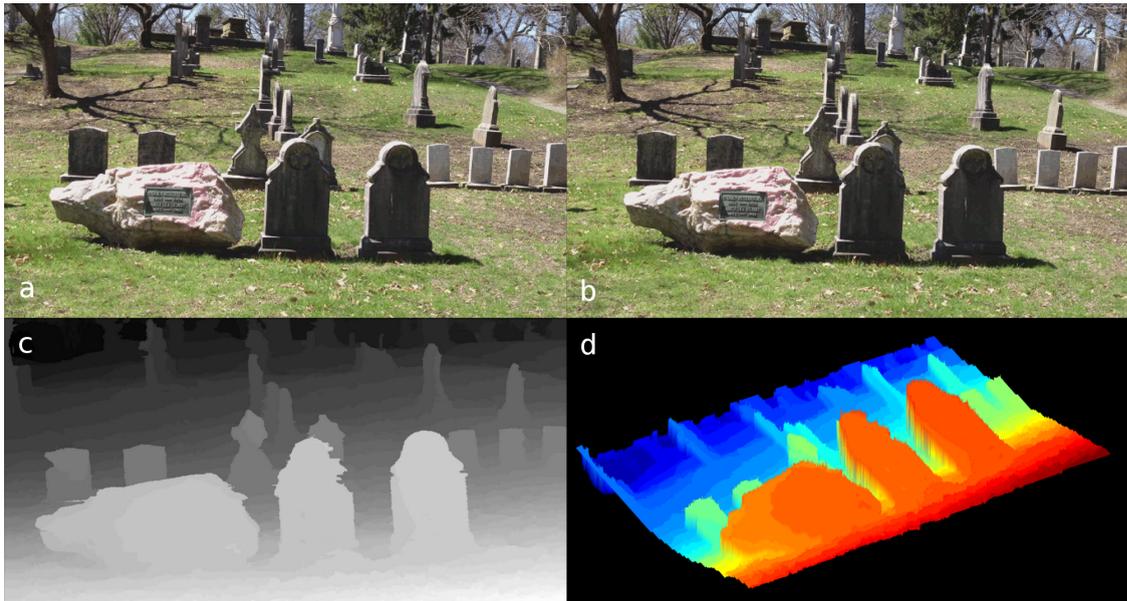


FIGURE 2.3: a) and b) are images captured from different positions. c) shows the resulting disparity/height map and d) is the height map shown in 3D (a) to c) adapted from work by [Ugo Capeto](#)). d) is an original mesh produced from c) using Matlab.

In Figure 2.3 the images a) and b) are captured from slightly different positions - they possess a small baseline. Depth from stereo vision can also be performed using images with a much greater difference between viewpoints - referred to as wide baseline stereo [39].

Recovering shape from multiple image pairs of the same object/scene from different positions is known as ‘Multi-view stereo’ - typically many more than a single pair of images are used in these methods. The data captured by these multiple views is then used to produce some representation of shape via various methods.

For example, one variant of multi-view stereo uses “voxel colouring” [3, 40]. (A voxel is simply a cell of a three dimensional grid - the 3D equivalent of an image pixel in 2D). In Figure 2.4 we can see that two different views (with known positions) of some 3D scene have been captured. The pixels of each image are projected onto a grid of voxels. If two projected pixels with the same colour value intersect at the same voxel, that colour value is assigned to the voxel.

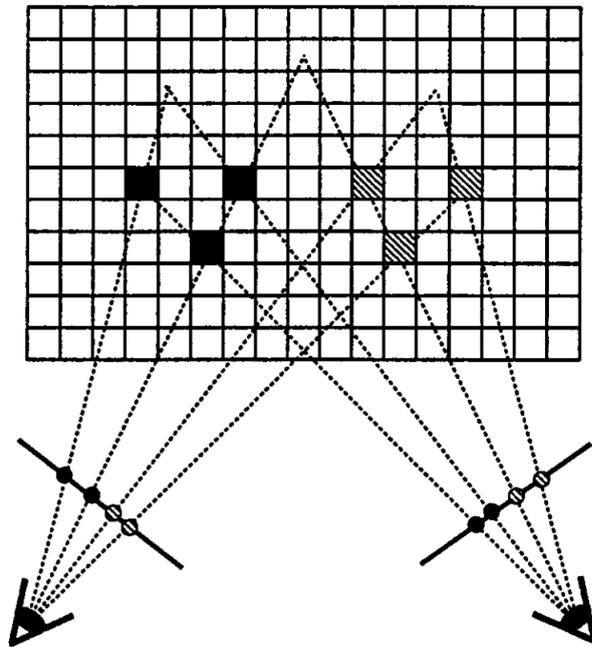


FIGURE 2.4: Images are captured at different viewpoints and projected onto a grid of voxels. Intersections determine the color for a voxel [3]

An example of shape recovered via voxel colouring can be seen in Figure 2.5. The examples shown used 21 images of a toy dinosaur as input, shape estimates are shown for successively finer resolutions of voxel grid.

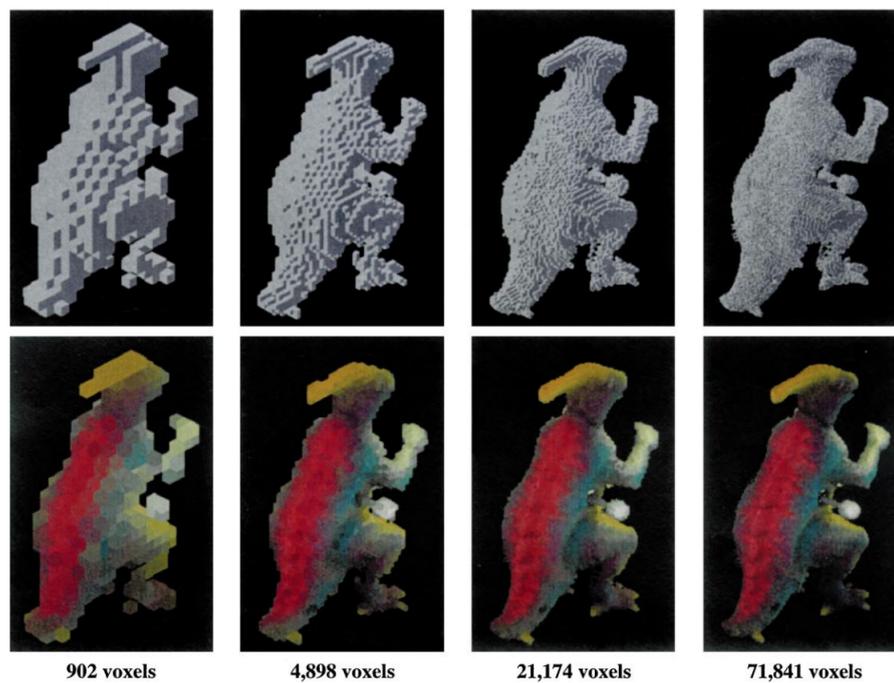


FIGURE 2.5: Shape reconstruction for a toy dinosaur for different voxel grid resolutions [3].

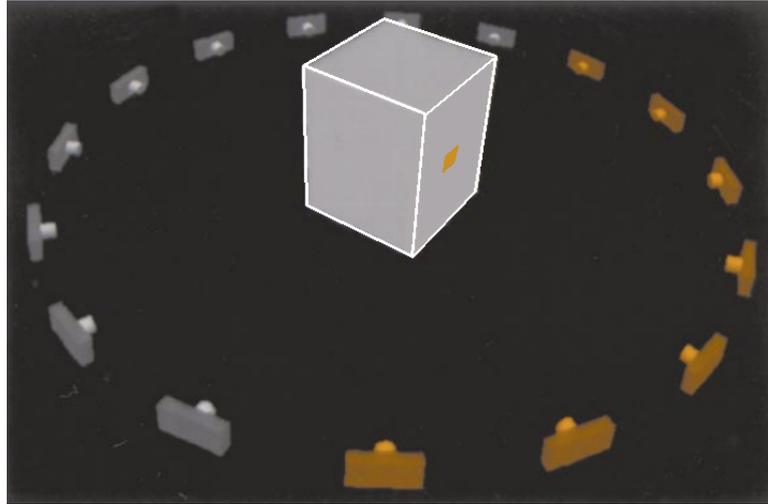


FIGURE 2.6: Adapted from “A Theory of Shape by Space Carving” by Kutulakos and Seitz [4]. Camera views determine a minimum bounding voxel grid (the cuboid) for an object. At each voxel the visible images are tested for photo-consistency. The currently active voxel and camera positions of active images are shown in orange.

A second variant of multiview stereo that uses a voxel grid is “Space Carving” methods. They begin by establishing from many images, a “photo hull” - essentially a voxel grid in which it is determined that the object being recovered must lie. Voxels are then tested for ‘photo-consistency’ when projected outwards to visible images (i.e. in Figure 2.6 the orange cameras represent images which are currently visible to the orange plane). If the appearance of a voxel does not match the pixels onto which it is projected in all of the images, then it is determined as not photo-consistent and is carved away. In Figure 2.7 we can see the shape recovery for a hand using this space carving technique.

For further detail on multiview stereo algorithms we refer the reader to a survey of work in the field conducted by Seitz [41].



FIGURE 2.7: Shape recovery of a human hand via space carving using a hundred input images [4].

Time-of-flight

Outside of traditional photogrammetry there are also different types of specialist hardware designed specifically for shape recovery. One example is time-of-flight systems [42, 43]. These estimate depth by measuring the time that elapses between a signal being emitted and reflected back to the sensor (Figure 2.8). The longer a signal takes to return, the further away the real world point must be from the sensor. Thus a height map of the scene is created (Figure 2.9).

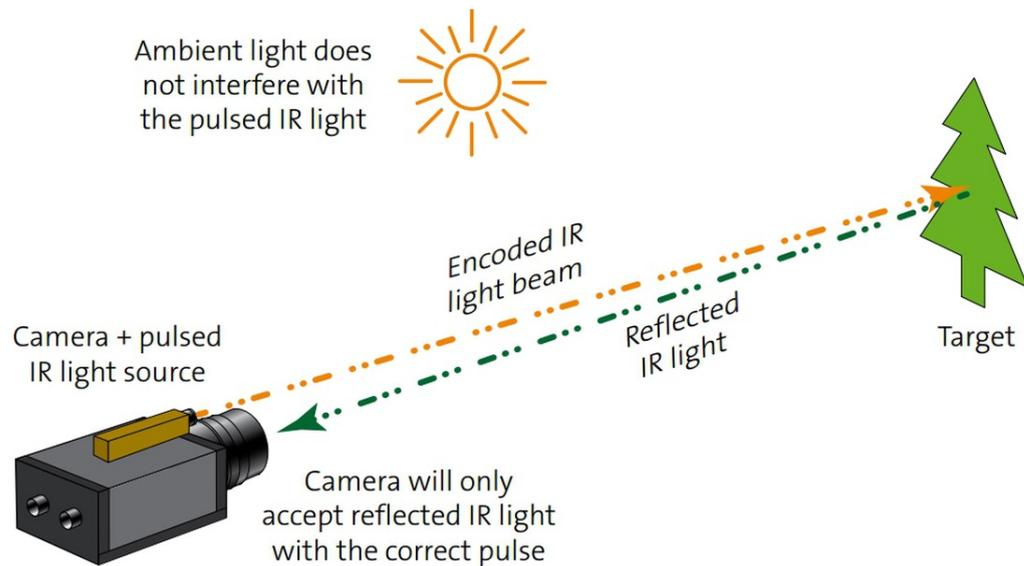


FIGURE 2.8: Time-of-flight systems: Infrared light is emitted. Time taken for the same pulse of light to return gives a measurement of distance. Thus a height map of a scene/object can be estimated (Image taken from [Stemmer Imaging](#)).

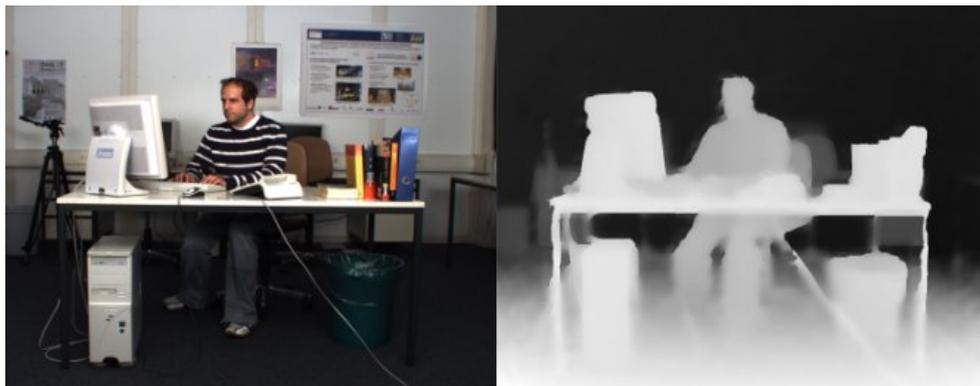


FIGURE 2.9: Left: Colour image. Right: The same scene imaged by a time-of-flight system. (Adapted from <http://www.mip.informatik.uni-kiel.de>)

Laser scanners

Similar to time-of-flight systems, we have laser scanners (Figure 2.10), which are widely held to be capable of returning the most accurate measurements of height [44]. However, it is worth noting that even these can be confounded by lighting conditions or surface reflectance [45].

As with time-of-flight, laser scanning systems project light signals and measure the response. The differences between the two are firstly, the amplification of (typically visible) light strength which is inherent to lasers. Secondly, time-of-flight cameras image entire scenes simultaneously (with comparatively low resolutions of approximately 300×200 pixels being common [46]), whereas laser scanners are much higher resolution but capture via scan lines (and are therefore a lot slower). Thus time-of-flight is better suited to height estimation in dynamic scenes whereas laser scanners are more appropriate for shape recovery of stationary objects (which is what our work focuses on).



FIGURE 2.10: Left: A laser scanner attached to a gantry. Right: The 3D shape recovered by the laser scanner (Image adapted from [The Digital Michelangelo Project](#)).

2.2 Photometric Stereo

Although great progress has been made in shape recovery it is a problem which remains far from being solved. Further, the shape recovery methods that do work for difficult objects do so using costly, inflexible equipment. For example, the laser scanning gantry in Figure 2.10) or the use of non-passive imaging environments.

By ‘non-passive’ we mean that light sources (be they projected patterns, targeted lights or lasers) must be directed towards the recovery target in laboratory conditions.

The motivation for our work originated with Vogiatzis and Hernández [5] where the shape of human faces is recovered using photometric stereo. Whilst the results are impressive, they are gained at the expense of an intrusive, laboratory-controlled set-up. A candidate is required to sit in a dark room for a period of time whilst colourful lights are successively shone directly at their face - hardly a pleasant experience! Even when applied to static objects it is an inconvenience to require the specific lighting setup and the need for a dark room.

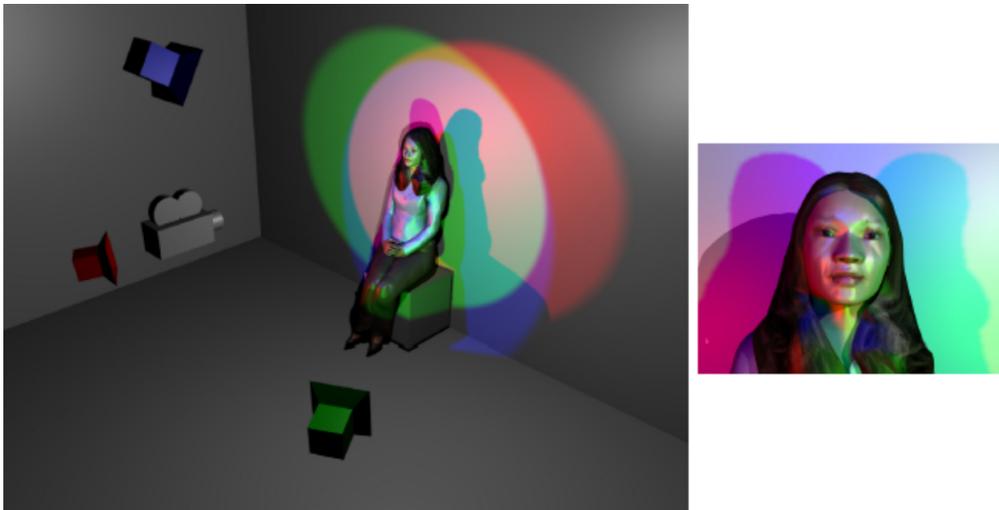


FIGURE 2.11: Left: Experiment setup used in [5]. Right: Image captured by the camera.

The aim of this thesis is to develop a photometric stereo system that moves towards real world conditions, without the need of a special laboratory set-up or multiple lights. Typically photometric stereo approaches take multiple images as input, however throughout this thesis it is demonstrated that by exploiting mutual illumination, one can recover shape from a single image.

In our paper, ‘Shape in a Box’ (SiaB) [47] (or to give the method its more technical moniker - ‘mutual illumination photometric stereo’), we developed on a variant

of photometric stereo known as ‘Shape from Color’ (SFC) [48]. The existing method states that recovery of object shape from a single image is possible if the object is illuminated by three spectrally distinct, direct light sources with linearly independent directions. We remove the restrictions on light sources by instead placing objects in a passive, colourful environment (the titular box).

2.2.1 Shape from shading

Before diving into photometric stereo, it is pertinent to first briefly discuss ‘shape from shading’. First proposed by Horn [28, 49, 50] shape recovery is performed from a single intensity (greyscale) image of an object. This requires known surface reflectance and a single light source for which the intensity and direction is known. Let us look at the work of Ikeuchi and Horn [6] to illustrate the problems of shape from shading.

The possible range of surface orientations for any object can be represented by a sphere. For any image of an object, this is constrained to a hemisphere (we cannot see the back of the sphere). Ikeuchi and Horn [6] thus represent surface orientations by projecting a hemisphere into a 2D ‘orientation space’ (Figure 2.12).

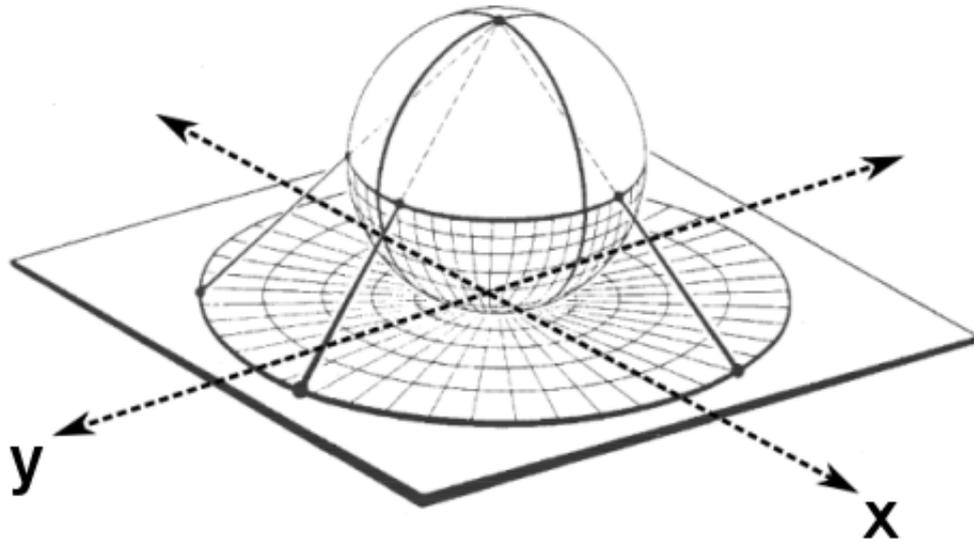


FIGURE 2.12: A hemisphere projected into a two-dimensional space (Adapted from [6]).

Each point in this orientation space $N(x, y)$ therefore corresponds to a 3D surface normal $\mathbf{n} = [n_x \ n_y \ n_z]$ (Figure 1.4). If we have a single light source with known strength and direction then a ‘reflectance map’ $R(x, y)$ can be rendered (Figure 2.13). This reflectance map shows the intensity values generated by the surface orientations in $N(x, y)$. So, given an intensity value observed in an image, one can look up the corresponding surface normal.

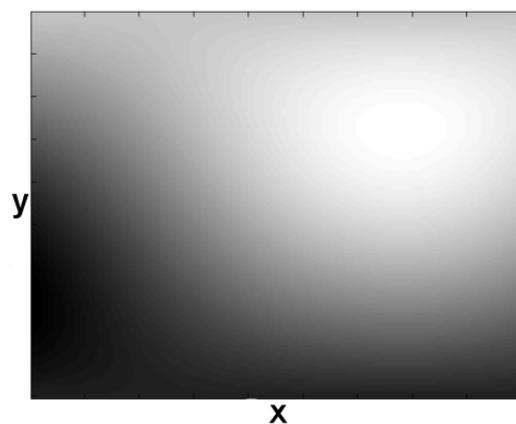


FIGURE 2.13: An example of a reflectance map for a Lambertian sphere (Credit to James O’Shea - University of California).

However, intensity values in $R(x, y)$ are not unique. For example, the reflectance map in Figure 2.14 shows a contour which highlights $R(x, y) = 0.6$ - thus there are many possible surface normals which could have generated the same pixel intensity.

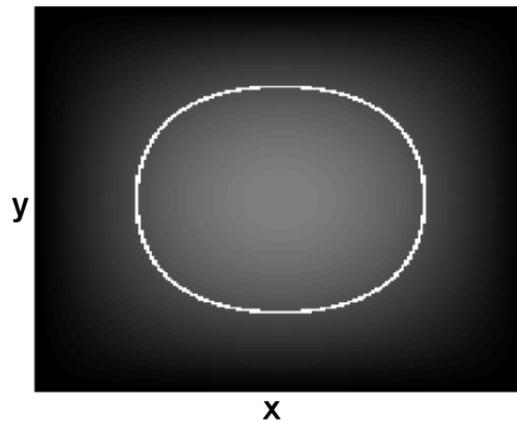


FIGURE 2.14: A reflectance map with $R(x, y) = 0.6$ highlighted (Credit to James O'Shea - University of California).

In the work of Ikeuchi and Horn [6] they address this issue by forcing recovered shapes to be smooth - any estimated surface normal must be similar to its neighbours (this reduces the ambiguity shown in Figure 2.14). This is fine for surfaces with smoothly changing orientations but will otherwise fail (Figure 2.15).

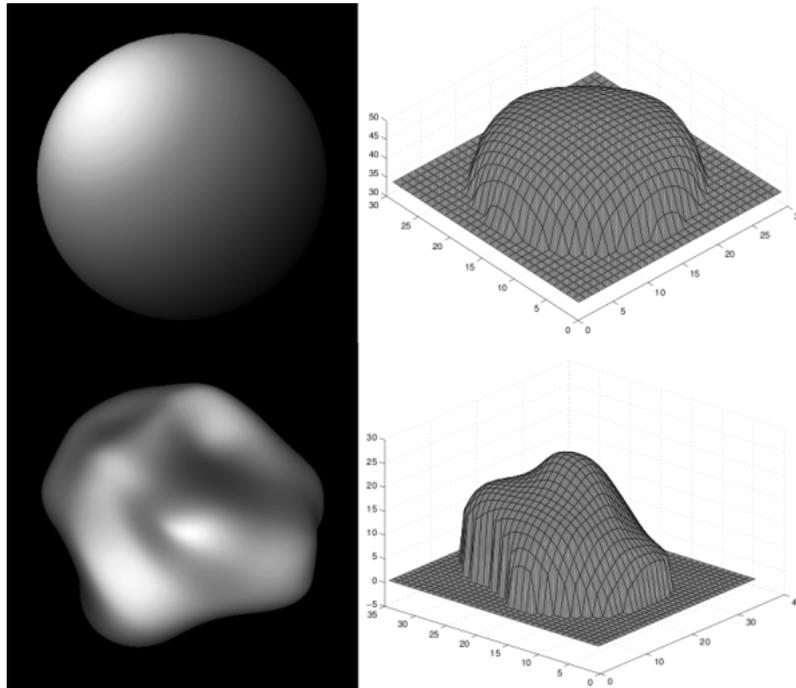


FIGURE 2.15: Left: Greyscale images. Right: Estimated shapes (Credit to James O'Shea - University of California).

In summary, there have been many refinements and adaptations made to shape from shading [51]; but due to inevitable ambiguities in shading it is an inherently ill-posed problem. One approach to reducing these ambiguities is to capture additional images, which leads us on to photometric stereo.

2.2.2 Classic photometric stereo

In classic photometric stereo [2] Woodham proposed that the surface normals of a convex, Lambertian (reflects light equally in all directions) object can be recovered from three images taken for three distant, point light sources with independent directions. Woodham showed that there was a linear relationship between the three pixel values and the surface normals of the object.

To understand how shape can be recovered, let us denote the direction of each light as a unit normal vector \mathbf{e} , we have three (by assumption linearly independent) vectors \mathbf{e}_1 , \mathbf{e}_2 and \mathbf{e}_3 . With respect to Lambert's law, a point on a surface with

unit normal $\mathbf{n} = [n_x \ n_y \ n_z]^t$ illuminated by light source \mathbf{e} results in an image pixel value p ,

$$p = \alpha(\mathbf{e} \cdot \mathbf{n}). \quad (2.1)$$

Both \mathbf{e} and \mathbf{n} are unit vectors, thus $\mathbf{e} \cdot \mathbf{n} = \cos(\theta)$ where θ is the angle between the two vectors (Figure 2.16). The α in Equation 2.1 represents the albedo of the surface - how strongly it reflects light. This value is assumed to simply be a constant scalar value (i.e. the surface has uniform reflectance).

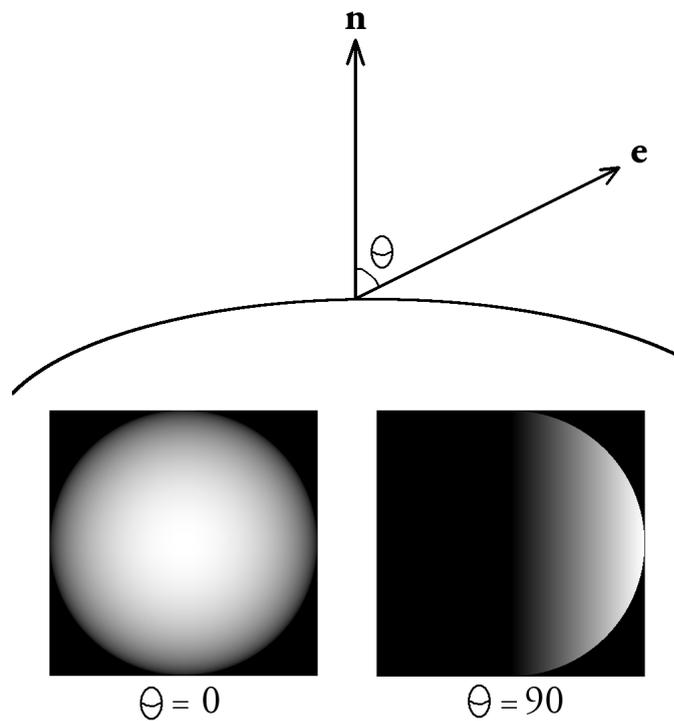


FIGURE 2.16: A sphere viewed from the same position. As the angle θ between surface normal vector \mathbf{n} and light direction vector \mathbf{e} increases, less light reaches the front of the sphere.

Let us use the notation \mathbf{p}_i to denote the i th triple of pixel responses (one for each light) and \mathbf{n}_i denote the corresponding i th scene surface normal. We group the image responses, the lighting directions and the scene surface normals into matrices P , E and N respectively,

$$P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_i \end{bmatrix}, E^t = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix}, N = \begin{bmatrix} \mathbf{n}_1 & \mathbf{n}_2 & \dots & \mathbf{n}_i \end{bmatrix}. \quad (2.2)$$

Under the assumption that the surface in question has uniform, Lambertian reflectance, there exists a linear relationship between the light reflected at each point on the surface (captured pixel values) and the orientation of the surface at each point,

$$P = \alpha E N. \quad (2.3)$$

Since we know E and P , Woodham observed, we can recover N :

$$N = \alpha E^{-1} P. \quad (2.4)$$

We know that the surface normals have unit length. Thus we can determine the albedo, α (we choose α to meet the unit length condition). Many shape recovery methods (photometric stereo included) output estimates of surface normals. However these alone are not a particularly useful representation of shape, thus they are commonly converted into a height map (Figure 2.17). Let us now look at how this conversion is performed.

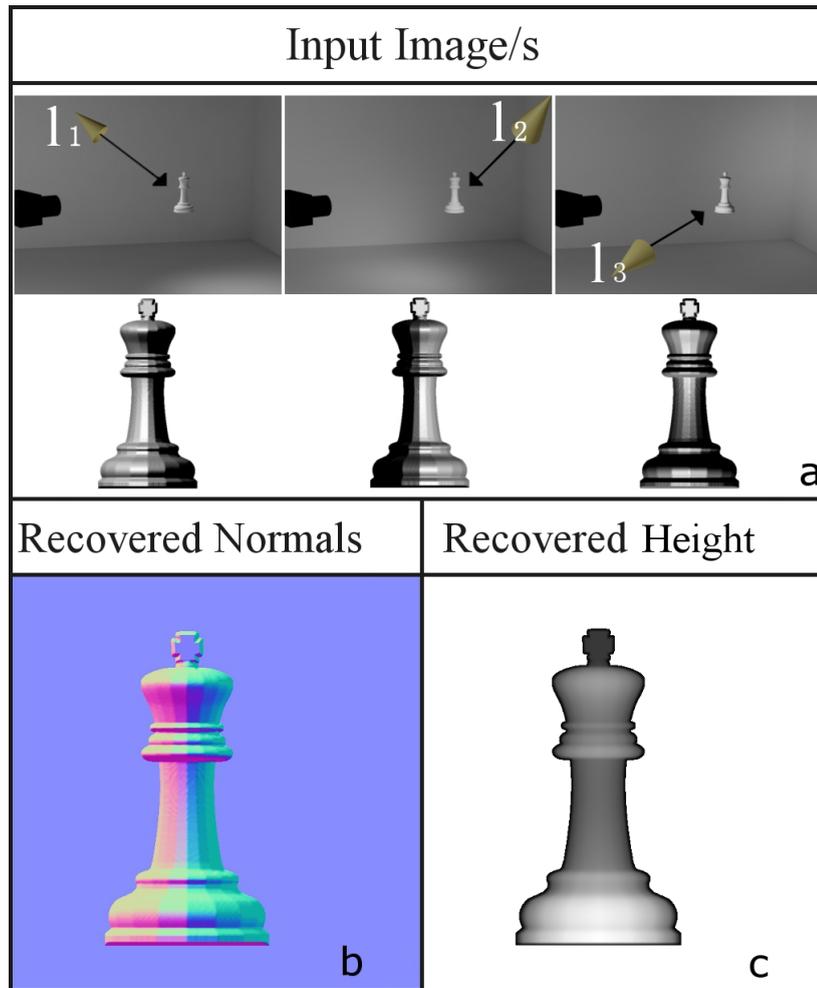


FIGURE 2.17: In parts **a** light sources are labelled as l_1 , l_2 , l_3 in the scene diagrams, below each diagram is the corresponding captured image of the object. Surface normals in **b** are displayed using the convention of normal maps in graphics (red = positive x-axis, green = positive y-axis, blue = positive z-axis. The height map in **c** is in grayscale with white = maximum height, black = minimum height). The same colour coding is used throughout this thesis.

2.2.2.1 Surface normals to height maps

Our shape recovery method has output a vector field of surface normals - essentially a two-dimensional grid N and for every (x, y) coordinate on that grid we have a surface normal \mathbf{n} ,

$$N(x, y) = \mathbf{n}. \quad (2.5)$$

We want to transform this representation of the recovered surface into a height map $Z(x, y)$, with every (x, y) coordinate being equal to some height value z ,

$$Z(x, y) = z. \quad (2.6)$$

The key observation to be made here is that $N(x, y)$ is related to the first order derivative of $Z(x, y)$. To show this it is easiest to think about the problem in reverse.

If we have a height map $Z(x, y)$, how can we calculate the normals to the surface that it represents? In Chapter 1, we showed that a surface normal is perpendicular to the surface (Figure 2.18). Thus if we want to find \mathbf{n} at $N(x, y)$, we can do so by taking the cross product of two vectors which lie on the tangent plane at (x, y) . For each point in $Z(x, y)$ two such vectors can be:

$$\mathbf{a} = [1, 0, \frac{\partial z}{\partial x}] \text{ and } \mathbf{b} = [0, 1, \frac{\partial z}{\partial y}]. \quad (2.7)$$

If we take the cross product of these vectors to find the normal to the tangent plane,

$$\mathbf{a} \times \mathbf{b} = [-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1], \quad (2.8)$$

then dividing the result by its magnitude gives us a unit-length surface normal,

$$N(x, y) = \frac{[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1]}{\|[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1]\|} = \mathbf{n} = [n_x, n_y, n_z]. \quad (2.9)$$

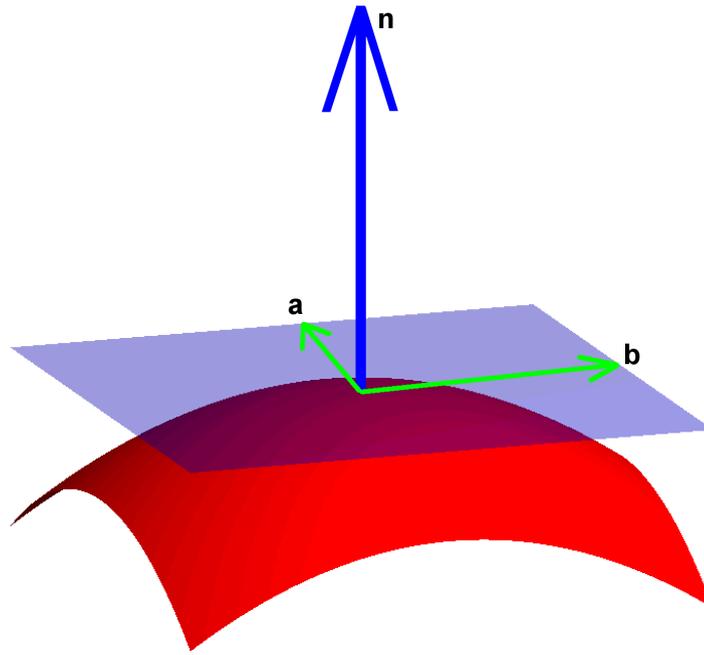


FIGURE 2.18: The blue vector \mathbf{n} is normal to the red surface. The green vectors \mathbf{a} and \mathbf{b} lie on the tangent plane (Modified version of the image in Figure 1.4).

Thus the vector field $N(x, y)$ is simply the first-order derivatives of $Z(x, y)$ scaled by n_z , that is to say,

$$-\frac{\partial Z}{\partial x} = \frac{n_x}{n_z} \text{ and } -\frac{\partial Z}{\partial y} = \frac{n_y}{n_z}. \quad (2.10)$$

So, if we have $\frac{\partial Z}{\partial x}$ and $\frac{\partial Z}{\partial y}$ for a recovered shape, how can we recover the function Z from these derivatives? The first thing to notice is that at every pixel we have two numbers but the function Z has one. Further, because we are finding the derivatives by a circuitous route (by our arguments from photometric stereo), it is unlikely that there exists a $Z(x, y)$ whose derivatives match those that we seek. Thus we would like to minimise,

$$\min(Z) = \int \int \left(\frac{d}{dx} Z(x, y) + \frac{n_x}{n_z} \right)^2 + \left(\frac{d}{dy} Z(x, y) + \frac{n_y}{n_z} \right)^2 dx dy. \quad (2.11)$$

Denoting the computed derivatives at (x, y) as $Z_x(x, y)$ and $Z_y(x, y)$ it turns out that the least squares solution to this problem is solved by integrating,

$$\min(Z) = \int \int \frac{d}{dx} Z_x(x, y) + \frac{d}{dy} Z_y(x, y) dx dy, \quad (2.12)$$

where the integrand is the ‘divergence of the gradient’. If the derivatives are integrable this is the same as the Laplacian of the image. In order to compute the integral we need to know what happens at the imaged object boundaries. Common boundary assumptions are that outside of the shape the height is zero (homogenous Dirichlet assumption) or that the derivative of the shape is zero at the boundary (homogenous Neumann).

So, because the surface normals estimated by shape recovery methods are almost certain to be non-integrable it is necessary to use some reintegration method [24, 25, 52, 53] to recover a height map.

Back to photometric stereo...

When all the underlying assumptions hold, the classic method works perfectly. However, in reality the assumptions are never strictly true. No object exhibits perfect Lambertian reflectance (i.e. zero specular component), interreflections/shadows cause unexpected pixel values and light sources must be carefully configured for maximum accuracy (e.g. across all three images it is necessary for all object pixels to be illuminated by all three light sources for fully accurate recovery).

Object geometry and material properties can easily render the original method useless. For example, as shown in Figure 2.19 complex geometry can cause self-occlusion and thus image captures which do not conform to the relationship expected by classic photometric stereo. We assume that point light sources are

infinitely distant from an object, thus regardless of spatial location a surface normal should receive the same amount of light. However in Figure 2.19 we have identical surface normals which are illuminated completely differently.

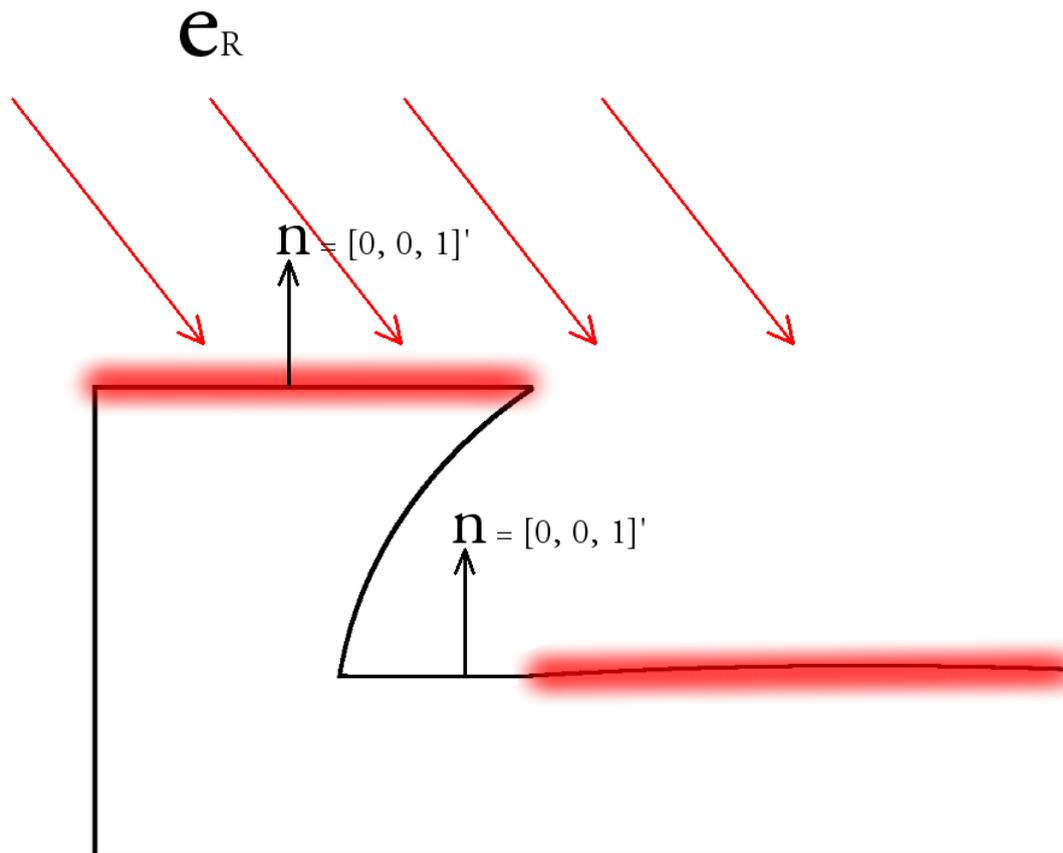


FIGURE 2.19: A complex shape causes self-occlusion from a red, infinitely distant point light source. Surface normals with identical $[x, y, z]$ components receive different lighting.

Another example of the failings of classic photometric stereo can be seen in Figure 2.40 in which an object consisting of a shiny material causes error in shape recovery - this issue is discussed further in Section 2.4.2.

2.2.3 Refined photometric stereo

The classic photometric stereo method has been subject to various modifications for various purposes. Here we briefly cover a couple of interesting works but note

that there is a plethora of existing variations for the interested reader. Murez et al. [54] apply image processing methods (deconvolution and deblurring) to estimate the shape of objects in a scattering medium (e.g. water). Liao et al. [55] perform photometric stereo using near-light sources. That is, the lights are not assumed to be infinitely distant as in the classic approach. This is performed by placing specular spheres in the scene alongside the experiment object which allows them to estimate the direction and distribution of light without an explicit calibration step. Modelling light sources in this way is difficult (hence the need for four measurement spheres in [55]) but should increase the accuracy of photometric stereo methods. The distant light assumption applied by most is unrealistic e.g. in the real world a light source will not project a uniform amount of energy onto every visible point in a scene due to fall-off over distance (imagine pointing a torch at a wall).



FIGURE 2.20: Specular spheres used to estimate light parameters in the work of Liao et al. [55].

2.2.4 High-end photometric stereo

The most advanced photometric stereo systems require laboratory conditions and very specific equipment set-ups, however they are able to recover accurate and highly detailed shape information. Conventional photometric stereo [2] uses only a single light source per captured image which can cause a poor signal-to-noise

ratio (many pixels will be captured with low intensity values) and can easily result in shadowed pixels which will generate incorrect results. These issues can be dealt with by increasing the amount of captured information.

One such example can be seen in the work of Vlasic et al. [1]. A dome equipped with eight cameras and over one thousand light sources provides a greatly increased quality of signal and effectively eliminates self-cast shadows. ‘Self-cast shadows’ refers to shadows which are caused by the target object occluding parts of its own surface from the light source; with a large number of light sources this effect becomes negligible. These capture conditions allow Vlasic et al. to reproduce the shape of dynamic geometry down to the level of a few millimetres. Whilst this method is highly accurate, obviously the requirement for a dome and large amount of equipment is very restrictive - something which we wish to avoid in our own work.

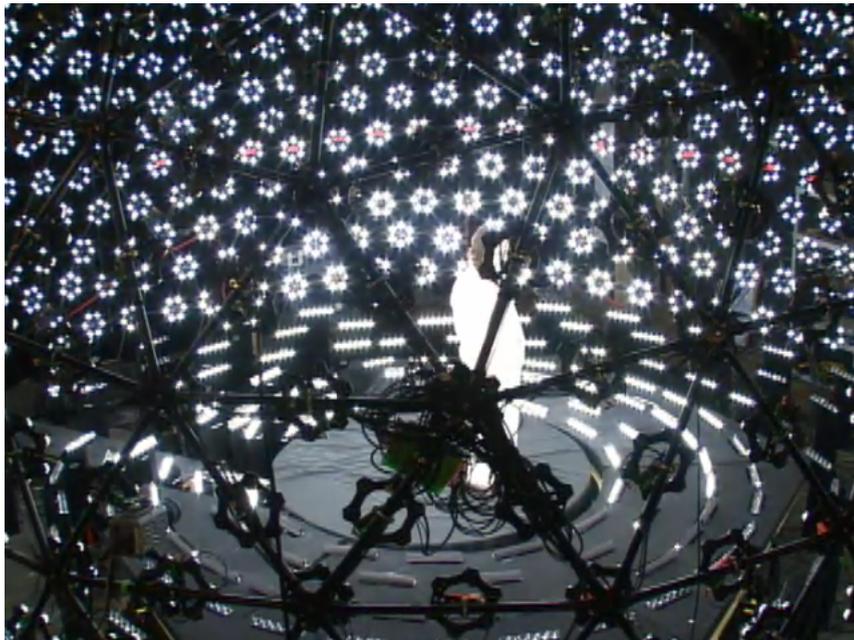


FIGURE 2.21: The dome of cameras and light sources used to reconstruct shape in [1].

2.2.5 Colour photometric stereo

One particularly limiting weakness of the classic photometric stereo approach is the requirement for multiple image captures with only light sources being varied. As everything else (camera, object, scene) must be held constant, this makes it impossible to perform any kind of dynamic shape recovery without considerable modification of the method (as seen in the dome of Vlasic et al. [1]). The need for such stringent conditions is also limiting if we wish to make photometric stereo-based shape recovery accessible to non-experts. To this end we now examine a variant of the classic method which requires only a single image.

In colour photometric stereo (also referred to in the literature as ‘shape from colour’ or ‘multispectral photometric stereo’), we replace the three grayscale images used by Woodham [2] with the three channels (RGB) of a single colour image. ‘Shape from Color’ (SFC) was first proposed by Drew [48], instead of utilising three white light sources; three spectrally distinct lights are employed. The three colour channels provide three distinctly illuminated greyscale images of the same object. See Figure 2.22 for an illustration of the experimental set-up.

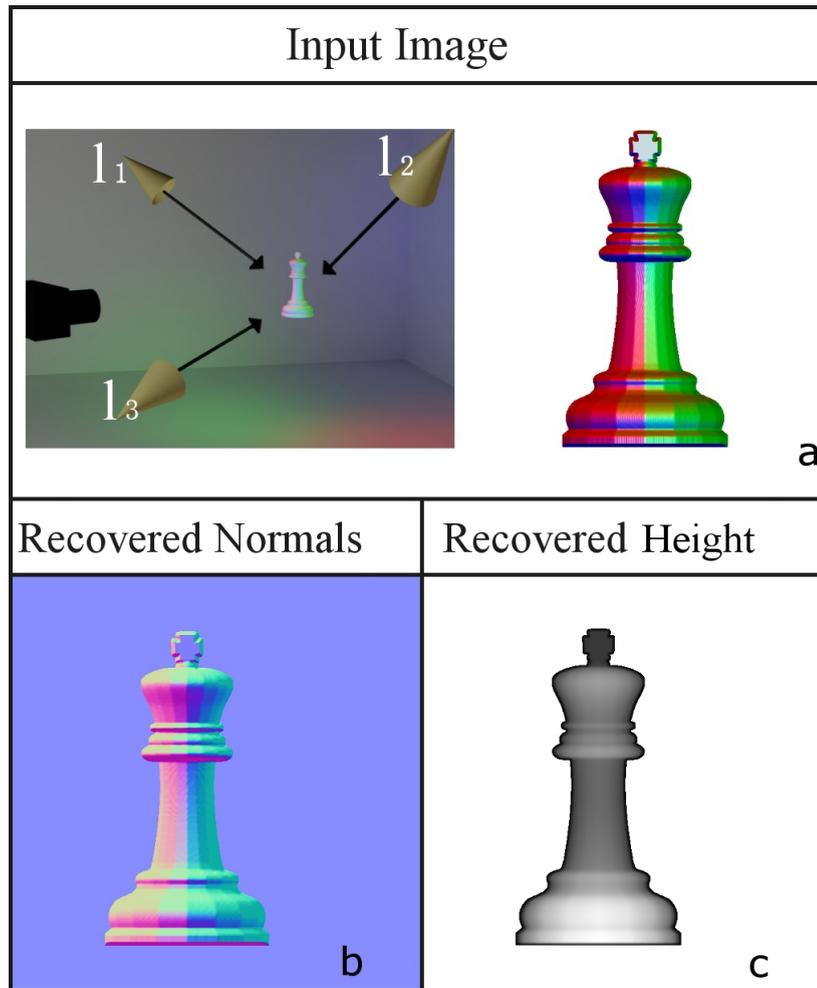


FIGURE 2.22: In part **a** light sources are labelled as l_1 , l_2 , l_3 in the scene diagram, right of the diagram is the corresponding captured image of the object. Parts **b** and **c** show recovered surface normals and a height map using the same conventions as Figure 2.17

Again let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ denote light direction vectors and now let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ denote the colour of the reflected light (in effect \mathbf{b}_i is the RGB of a flat, frontally presented surface with the same albedo as the object to be measured under the i th light) then the RGB camera response \mathbf{c} at a pixel is given by

$$\mathbf{c} = [(\mathbf{e}_1 \cdot \mathbf{n}) \mathbf{b}_1, (\mathbf{e}_2 \cdot \mathbf{n}) \mathbf{b}_2, (\mathbf{e}_3 \cdot \mathbf{n}) \mathbf{b}_3]. \quad (2.13)$$

We can see that this is similar to Equation 2.1 used in the classic approach, only now we are capturing all three images simultaneously and each is scaled by \mathbf{b} . As

in Equation 2.2 we can group our vectors \mathbf{e} and \mathbf{b} into matrices E and B

$$E = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix}^t, B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}. \quad (2.14)$$

Then equation (2.13) can be rewritten as

$$\mathbf{c} = F \mathbf{n} \equiv B E \mathbf{n}. \quad (2.15)$$

In practice, we capture an image of an object with known shape to calibrate the system, thus B and E are not known separately; rather we skip straight to estimating the matrix F through ordinary least squares regression of \mathbf{n} to \mathbf{c} . As in Equation 2.4, it is then a simple matter of inverting the calculated lighting matrix to enable surface normal recovery:

$$\mathbf{n} = G \mathbf{c} \equiv F^{-1} \mathbf{c}. \quad (2.16)$$

However, the calibration target may not have the same albedo as the object which is to be recovered (which will be the case for any useful application of the method). We are relying on F to provide us with the exact linear relationship between RGB pixel values and surface normals. When the recovery object possesses a different albedo, then its measured response B will be different, thus the F established in calibration is incorrect.

Suppose B' denotes the response of a second surface (the first being the calibration object). Both B' and B are 3x3 matrices: each column is the RGB response to a light. Under the Diagonal model of illuminant change the relationship between a pair of surfaces under any illuminants is modelled by a Diagonal matrix. That is if,

$$\mathbf{p}^{1,E} \text{ and } \mathbf{p}^{2,E} \quad (2.17)$$

denote the RGB of a first and second object under a light $E(\lambda)$ then,

$$D\mathbf{p}^{2,E} = \mathbf{p}. \quad (2.18)$$

Further, D is independent of E . It follows then that,

$$B' \approx DB, \quad (2.19)$$

And,

$$\mathbf{c} = D F \mathbf{n}, \quad (2.20)$$

where D is the illumination change. Substituting $F_c = DF$ in Equation 2.16, we would like,

$$\mathbf{n} = F_c^{-1} \mathbf{c} = D^{-1} G \mathbf{n}. \quad (2.21)$$

Because surface normals have unit length:

$$1 = \|D^{-1} G \mathbf{c}\| \equiv \mathbf{c}^T D^{-1} G^T G D^{-1} \mathbf{c}. \quad (2.22)$$

Here we are stating that if a surface normal has been correctly estimated, then it should be of unit length. Deviation from this expected value indicates that some amount of error has occurred. We can minimise this error using the following

expression to estimate D and thus account for surface albedo (where j represents the number of pixel values),

$$\min\left(\sum_{i=1}^j \mathbf{c}_i^T D^{-1} G^T G D^{-1} \mathbf{c}_i - 1\right) \quad (2.23)$$

SFC can estimate shape for smooth objects and the requirement of only a single image means that it can be performed in real-time without strictly requiring processing across the temporal dimension. Although, to achieve reliability in real-world experiments (i.e. not synthetic) it has been necessary to include the propagation of information over time. Examples of such applications can be seen in shape recovery of human faces [5] and video sequences of non-rigid surfaces (e.g. cloth) [10, 56].

2.2.5.1 High-end colour photometric stereo

An example of a high-end system which uses the SFC approach can be seen in the work of Gotardo et al. [57]. This method involves combining colour photometric stereo with multiple other methods to refine the results. Multispectral light sources illuminate the target (as with SFC), however an additional camera is also used to capture two images simultaneously. The cameras continue to record data over time - resulting in colour photometric stereo from two viewpoints extended across the time domain (i.e. two video streams). By performing colour photometric stereo, multiview stereo and optical flow simultaneously they are able to recover highly detailed, dynamic 3D shape.

As with the extended classic photometric stereo method employed in the dome environment of Vlasic et al. [1], the results generated by more the complex setup of Gotardo et al. are detailed and accurate. However, they also share the same issue in that a very specific experimental environment is required. Obviously it is desirable to be able to reconstruct the most high quality 3D shape possible,

but unfortunately such restrictions are at present irreconcilable with our aim of moving photometric stereo out of the laboratory.

2.3 Mutual illumination

Despite it existing to some degree in almost all captured images, mutual illumination (also referred to as ‘indirect illumination’ and ‘interreflection’) is a topic which has seen relatively little attention in computer vision research. Although the influence of mutual illumination can appear negligible in a large number of images, such as those captured in well-lit areas or outdoor scenes; the effects become greatly visibly amplified in more confined surroundings. Further, mutual illumination often appears negligible because our own vision largely discounts it [29], so we are not aware of it (even though physically it is a first order effect).

Gilchrist and Jacobsen [7, 58] carried out early work in mutual illumination and human perception. They showed that the human visual system is capable of exploiting cues provided by mutual illumination to judge the true lightness of a surface.

In their experiments two identical miniature rooms were constructed (Figure 2.24), each containing the same objects of varying shape and size. One room was painted entirely matte white, the other black. In Figure 2.23 we can see three luminance profiles taken from the rooms (luminance being the intensity of light emitted from a unit area of a surface - in this case the walls and objects in the rooms).

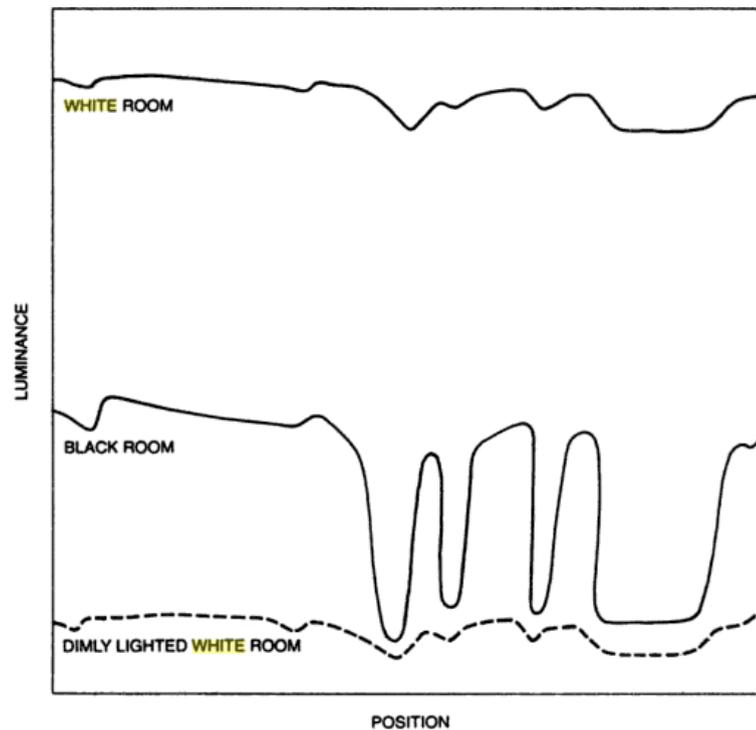


FIGURE 2.23: Luminance profiles of identical white and black rooms taken with different light intensities. (Figure taken from [7]).

In both cases the white room exhibits a smooth profile whereas the black room exhibits a greater range of intensity values. This is because the higher reflectance ratio of the white room causes mutual illumination to have a greater influence - the light bounces more around the enclosed space causing edges to become less defined. However the black room demonstrates very little in the way of mutual illumination effects, causing more sharply defined edges. Human observers were then asked to judge the lightness of surfaces in each room (i.e. informally, a greyscale estimate of the paint applied to the scenes). Even when the white room was dimly lit and produced a lower measure of luminance than the black room, the observers still correctly judged it to possess greater surface lightness. Thus showing that mutual illumination can in fact act as a cue for determining surface lightness, rather than being completely discarded.

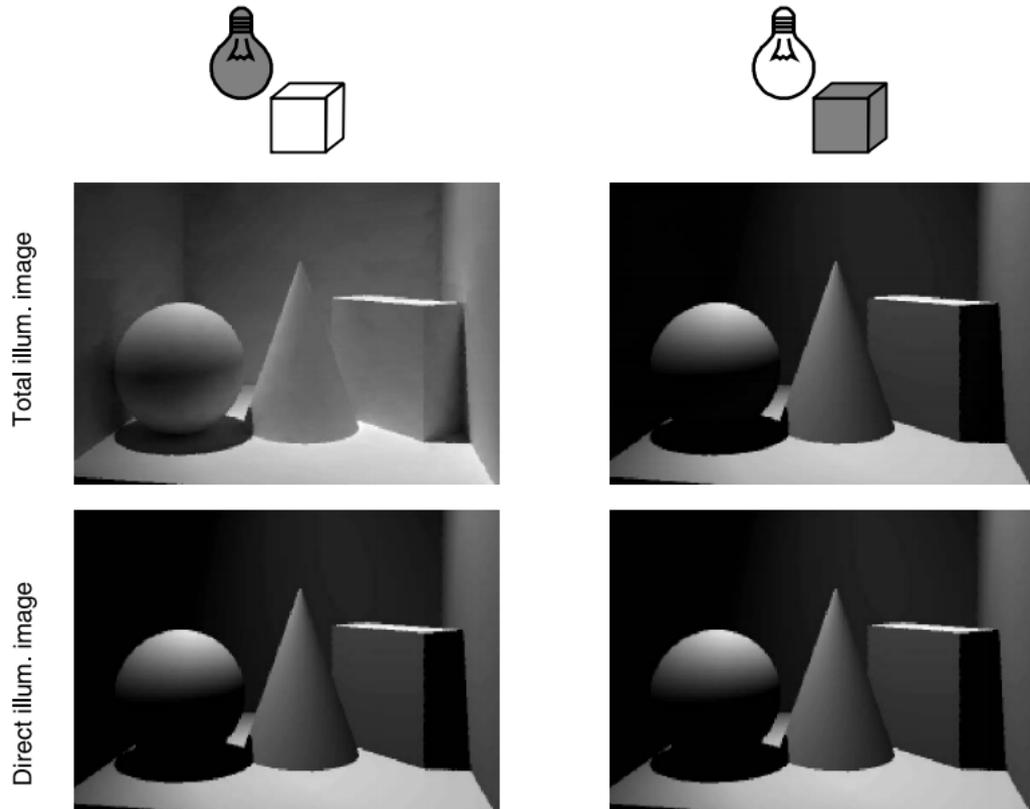


FIGURE 2.24: Rendered versions of the white and black rooms in the work by Gilchrist [7]. The renders using only direct illumination are identical, but with mutual illumination added it becomes clear which room is which. (Figure adapted from [8]).

The research of Gilchrist and Jacobsen [7, 58] was extended to the colour domain by Ruppertsberg and Bloj [8] who tested the hypothesis that observers could discriminate between surface colours in identical rooms, even if the lighting was arranged such that overall an identical colour signal was produced. To help explain their experiment consider how colour signal spectra are created,

$$R(\lambda) = I(\lambda) S(\lambda). \quad (2.24)$$

The colour signal $R(\lambda)$ is defined as the product of illumination $I(\lambda)$ and surface reflectance $S(\lambda)$. As we are simply taking the product of the two components, were their values to be switched an identical colour signal will result. While this symmetry holds for an isolated surface subject only to direct illumination,

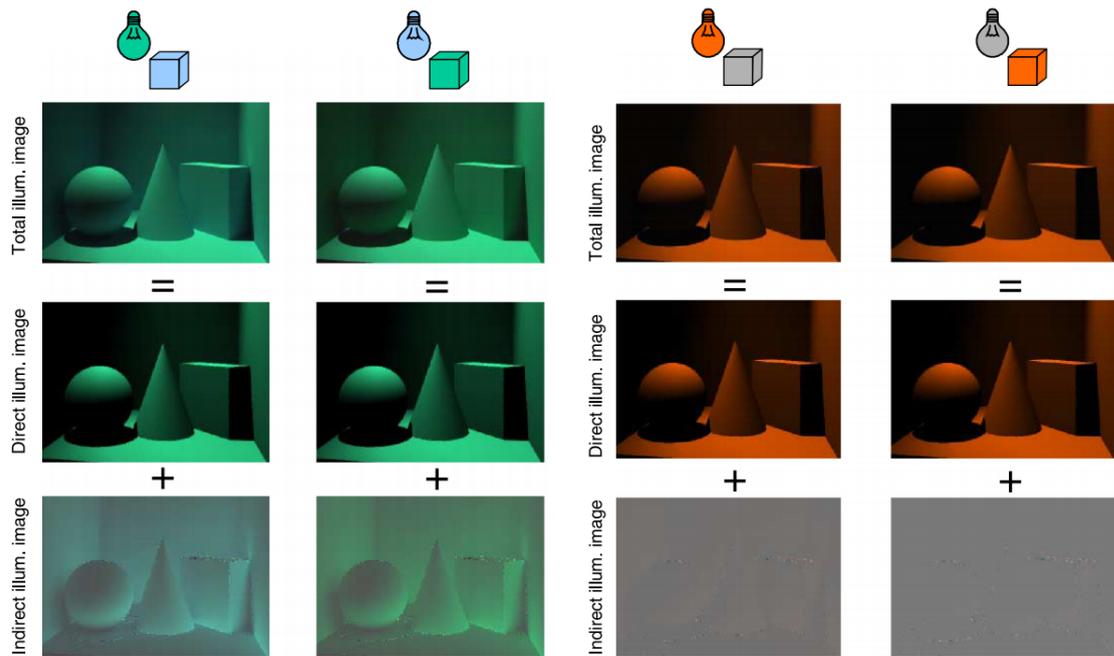


FIGURE 2.25: For the room pair on the left one can distinguish between surface colour and illumination, however this is not possible with the room pair on the right [8].

Ruppertsberg and Bloj wondered if the two would remain indistinguishable with the inclusion of mutual illumination. To test this, they rendered images of ‘room pairs’ with the illumination and surface reflectance switched. They used a color difference measurement [59] to determine that in many cases, the surface colour of the rooms was perceptually distinct due to the additional visual information provided by the indirect illumination. For some combinations of illumination and surface reflectance the resulting luminance was too low for the effect to be noticeable, though a difference in chromaticity could still be measured. In Figure 2.25 examples of both can be seen.

Mutual illumination in the context of shape from shading (shape recovery from a single greyscale image) was investigated by Forsyth and Zisserman [30]. It was shown that failing to account for mutual illumination can introduce error into such a system. This is because it is assumed that shading depends only on direct illumination, ignoring mutual illumination; which leads to incorrect shape recovery. Additionally, in the case of shape from shading further work proposed that it could

actually help provide shape cues [60].

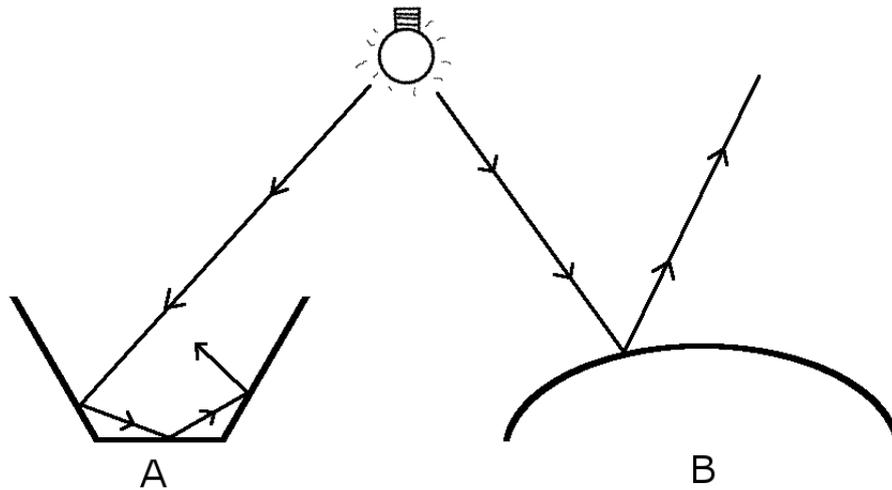


FIGURE 2.26: The radiance of shape A is amplified by mutual illumination in comparison with shape B.

Figure 2.26 illustrates the shape from shading problem. Were both shape A and B subjected to the same infinite point light source in the same direction, points on the surface of A would exhibit greater radiance than points on the surface of B due to the light experiencing multiple bounces before being reflected back out into infinity. This type of mutual illumination is also sometimes referred to as ‘self illumination’. Although they refer to it as the ‘mutual illumination equation’ in their work, Forsyth and Zisserman use what is more commonly known as the equation for ‘radiosity rendering’ to model this reflection of light between surfaces,

$$R(x) = T(x) + \alpha(x) \int_S R(x') \frac{1}{\pi d^2} \cos\theta_x \cos\theta_{x'} \cdot V(x, x'). \quad (2.25)$$

Here $R(x)$ is the total radiance (i.e. the brightness) of a point x , (x' simply refers to some other point in the scene). T is the light emitted directly from the point (in most cases zero), α is the albedo (the fraction of light reflected due to the surface material). We then have an integral across all surfaces in the scene S with d^2 being the distance between the points x and x' (this component represents the

falloff in light intensity over distance), $\theta(x)$ and $\theta(x')$ are the angles between the line joining the two points and their respective surface normals. Finally $V(x, x')$ is simply a visibility function set to either one or zero depending on if the two points can see each other or not.

In practice it is necessary to discretise this function, which is achieved by dividing the scene up into triangular or quadrilateral planar patches. We go into more detail on the specifics of radiosity modelling in Chapter 3 where we use it to engineer a suitable capture environment for mutual illumination based shape recovery.

To summarise, mutual illumination can have a significant, measurable effect on the appearance of images (to the point that it can even affect human interpretation). Despite this it is ignored throughout most of computer vision. Rather than ignore the phenomenon we exploit the presence of mutual illumination to perform recovery of 3D shape from images.

2.4 Implementation improvements

In Chapter 6, we present extensions to our mutual illumination photometric stereo method. Each of these variants has its own background of related and influential work which we briefly examine in this section.

2.4.1 Calibration

The majority of shape recovery methods require some sort of equipment or experimental set-up calibration to function correctly as the underlying algorithms require calibration parameters to be estimated/measured. In Photometric stereo approaches this calibration step typically involves imaging some object of known shape or pattern and calculating the required parameters. However, there are several issues with the requirement for a separate calibration step.

Firstly, after calibration scene conditions may change. For example, in outdoor scenes cloud movement can have a large effect on lighting. Secondly, if the calibration object and experiment objects do not possess the same surface reflectance then this limits the effectiveness of the calibration step. Thirdly, whatever object is used for calibration may itself alter the lighting environment (such as the mirrored ball which we use for calibration in Chapter 5).

Throughout photometric stereo literature there are a variety of approaches towards calibration. A simple and effective example can be seen in the work of Johnson and Adelson [9], where a sphere comprised of the same material as experiment objects is imaged. This establishes a relationship between pixel values and surface normals (due to a sphere being a known shape) which can be applied to unknown items (Figure 2.27). This addresses two of the three issues which we have stated, however Johnson and Adelson were performing shape recovery in natural illumination. Thus the first issue of changes in lighting is still present.



FIGURE 2.27: In [9] the calibration sphere and objects are painted to possess the same reflectance and the surface normals are recovered.

Similarly, Brostow et al. [10] used a specially designed, patterned, planar target containing a material sample which is imaged at multiple orientations. The contained sample is made from the same substance as the shape recovery target. This allows measurements to be taken of how that material appears under the lighting conditions at different angles (The orientation of the target can be established from the deformation of the pattern) before capturing images/video of the test

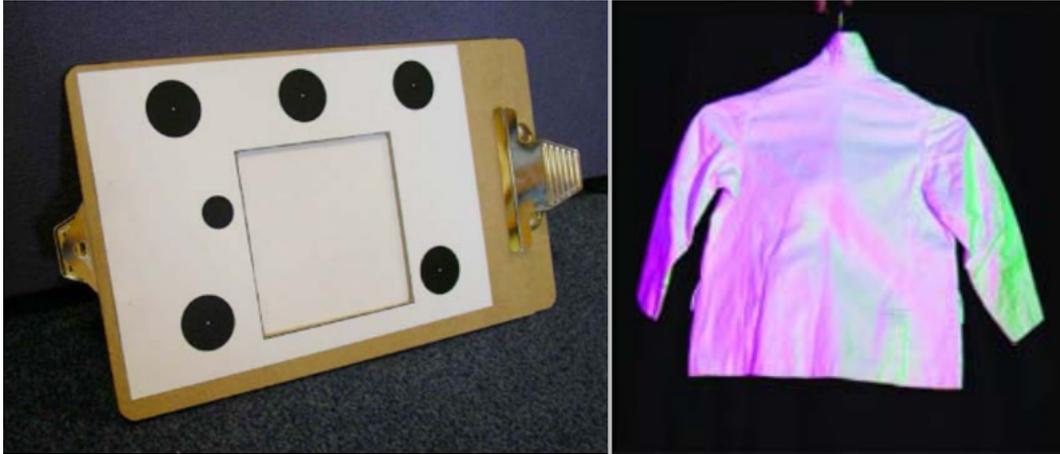


FIGURE 2.28: Example of a calibration step taken from [10]. A sample of object material is placed inside a calibration target (the square in the left image) before capturing video of that object undergoing some deformation (right).

subject - see Figure 2.28). Brostow et al. perform shape recovery in a laboratory environment, thus all three of our issues are addressed; however the requirement for controlled conditions limits the flexibility of the approach.

Calibration can even be as simple as manually measuring the position and angle of light sources [61]. However this is obviously an inconvenient calibration method and again will either require a controlled environment or be subject to sudden illumination changes in real-world conditions.

Thus it is desirable to remove the need for calibration completely. However, without calibration, photometric stereo is a more difficult problem.

The Generalised Bas-Relief Ambiguity

Photometric stereo algorithms can still function without an explicit calibration step [62], however results will be subject to the Generalized Bas-Relief (GBR) ambiguity [11, 63]. This means that it cannot be distinguished whether shading information in an image is due to surface geometry, the direction of light sources or the surface albedo (Figure 2.30). The term GBR ambiguity was inspired by ‘bas-relief’ sculptures, an example of which can be seen in Figure 2.29.



FIGURE 2.29: An example of a bas-relief taken from [11]. From a distance the sculpture appears to have full three-dimensional depth (left and centre), however up close it is actually quite flat (right).

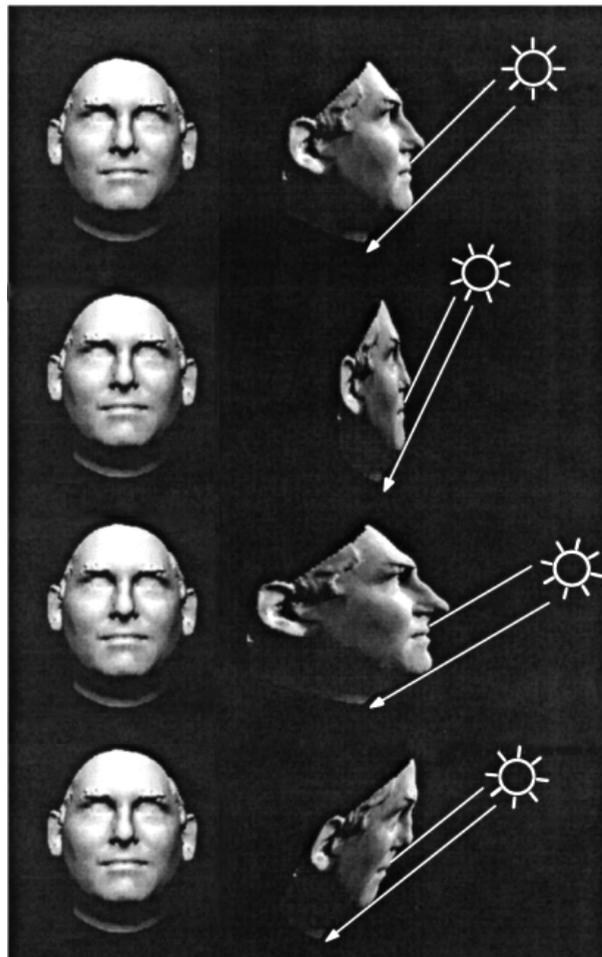


FIGURE 2.30: From [11] - the same image(left) can be generated from infinite variations of light direction and object shape (right).

Some have suggested that this problem can be solved through making assumptions about object properties. In [64] it is proposed that calibration can be performed as long as there are two specular pixels present on the surface of an object, as these can be used to estimate light directions and thus calibrate the system (i.e. the angle of reflection = the angle of incidence, see Figure 2.41). A similar approach is taken by [65], using ‘points where the Lambertian reflectance is maximal’.

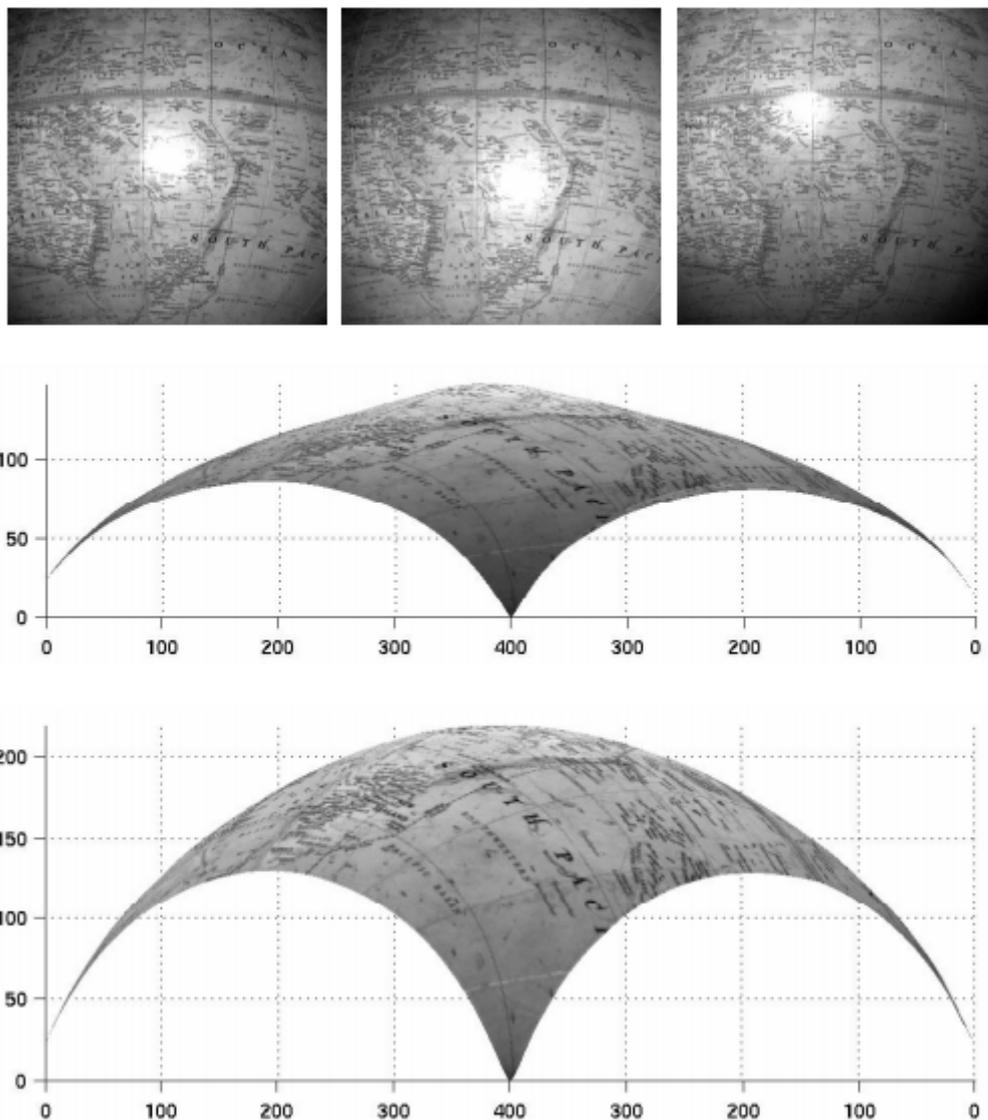


FIGURE 2.31: Top: Three of twelve input images of a globe. Middle: Shape recovery with the Lambertian reflectance model (Note the bump caused by misinterpretation of highlights). Bottom: Shape recovery from [12] with the Torrance and Sparrow reflectance model.

Similar to [64], another attempt at resolving the GBR ambiguity by Georghiadis [12] also exploits surfaces with specular components. Georghiadis fits a Torrance and Sparrow reflectance model to the surface being recovered, which unlike the Lambertian model; accounts for specularities. As in [64], by modelling the specular component of surfaces it is possible to estimate light directions and using many images as input allows for the albedo at each pixel to also be determined - thus resolving the GBR ambiguity (Figure 2.31).

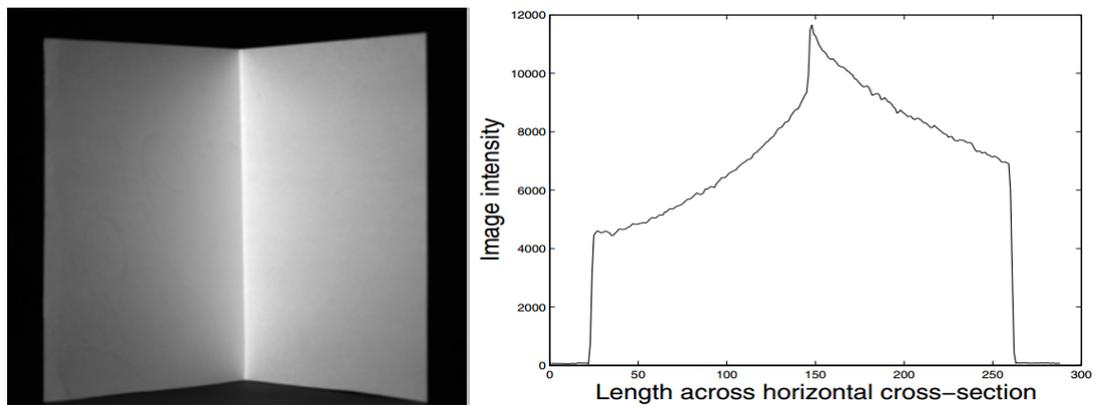


FIGURE 2.32: Demonstrating the effect of interreflections in a non-convex object. The fold in the centre (left) is expected to appear darker than the rest of the shape as it is at a greater distance from the light source, however due to light bouncing around within the crease it actually has greater intensity. The chart on the right shows the intensity profile across a row of the image [13].

In [13] Chandraker et al. propose exploiting mutual illumination to resolve the ambiguity (Figure 2.33). They state that counter-intuitive shading effects produced by the presence of interreflections provide cues to the true surface shape for non-convex objects (i.e. pixel values which do not appear as expected, such as in Figure 2.32).

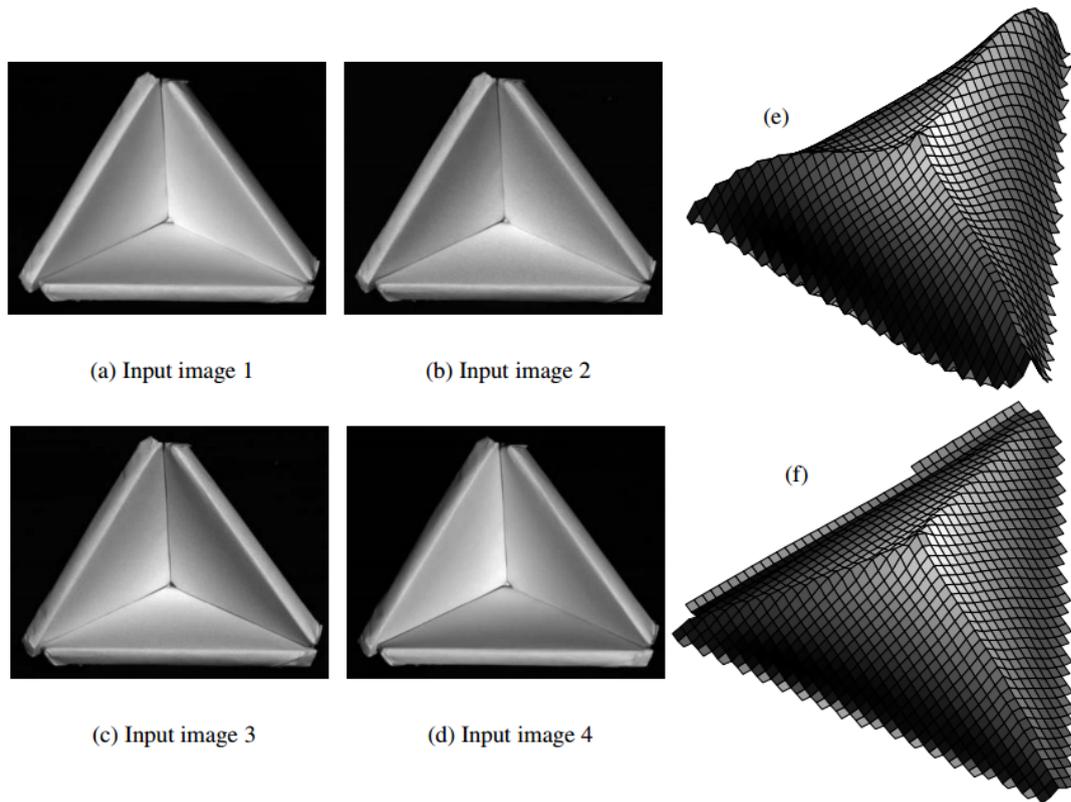


FIGURE 2.33: Adapted from [13]: a) - d) are input images. e) is the shape recovered by uncalibrated photometric stereo. f) is the shape recovered by Chandraker et al. using mutual illumination to resolve the GBR ambiguity.

The term ‘uncalibrated photometric stereo’ tends to refer to literature which focuses on solving the GBR ambiguity, though we use the term in a broader sense to mean any method which does not employ a separate calibration stage. In our work [66] we are not concerned with resolving the GBR ambiguity. We do not perform a calibration step, instead we exploit the environment visible in the image of our target object to estimate light directions, thus avoiding the problem. It should also be noted that all of the above described methods require multiple images, whereas ours - like shape from color - uses only one.

Uncalibrated methods without GBR focus

An interesting example of an uncalibrated method which isn’t concerned with the GBR ambiguity can be seen in [14], wherein a pair of smartphones is used to

capture images. One phone (the master) captures images at a fixed location whilst the other (the slave) is moved around the target, using its flash as a light source. The actual processing of data takes place on the master device, however the slave also captures an image with each flash, allowing light direction to be calculated for each image captured by the master (Figure 2.34).

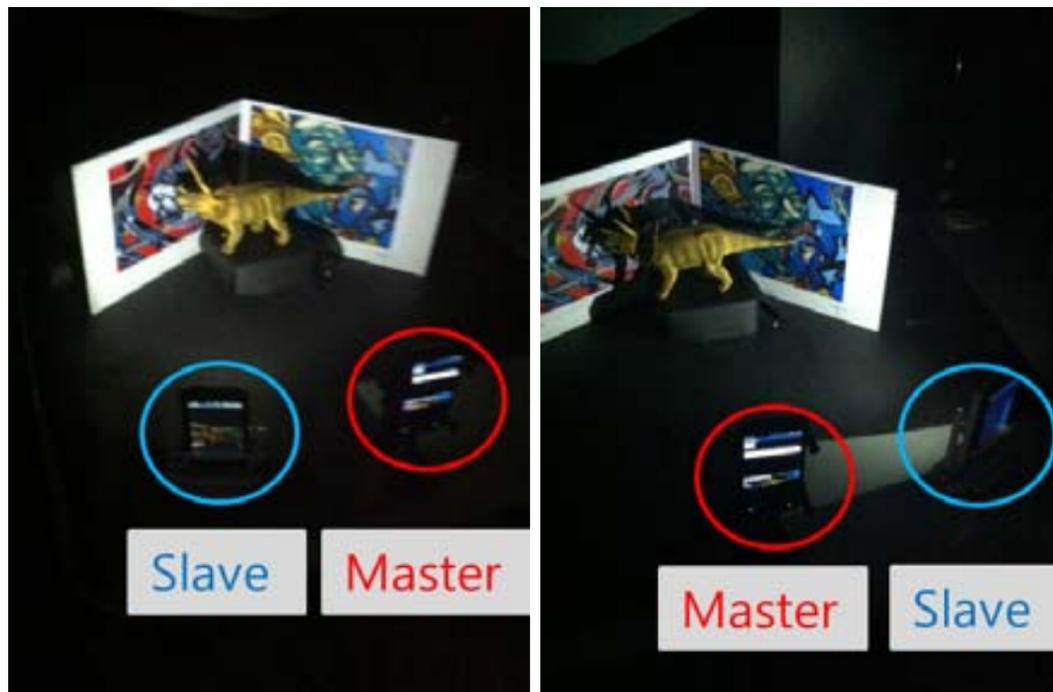


FIGURE 2.34: The master device remains in place. The slave device is moved and both capture an image using the flash from the slave. Light position and direction can thus be calculated to calibrate the system without using an explicit calibration step [14].

Another uncalibrated photometric stereo method is presented in [15]. Here Lu et al. look at the intensity values of pixels captured across multiple images (intensity profiles). Assuming that the object consists of one material then any points with shared intensity profiles should correspond to shared surface normals. In Figure 2.35 there are four points. A, B and C possess the same surface reflectance. A and C also have identical surface normals, thus they have identical intensity profiles. The relationship between the intensity profiles of neighbouring pixels is used by Lu et al. to calculate light directions and thus calibrate their system. Though this

is somewhat naive as mutual illumination would cause unexpected pixel intensities (Figure 2.32).

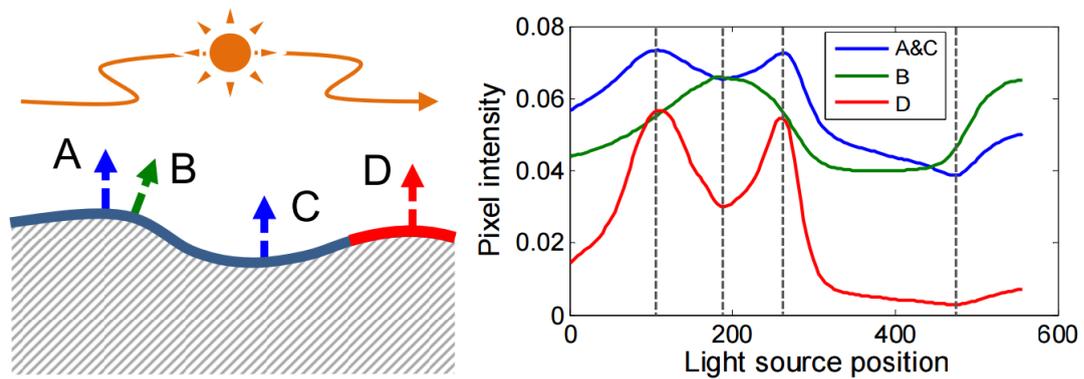


FIGURE 2.35: From Lu et al. [15] - pixel intensity profiles given a moving light source. Points A,B and C share the same surface reflectance. A and C also share the same surface normal, thus they share the same intensity profile. D has the same surface normal as A/C but different surface reflectance, thus a different intensity profile (the sharper peaks for D indicate a specular component - see Section 2.4.2).

More closely related to our own efforts, there are a few existing works which perform colour photometric stereo without an explicit calibration stage. In [16] Schindler recovers shape using only a laptop. The built-in camera captures images whilst the screen displays a multi-spectral pattern. This colourful image acts as the light in the scene, providing enough spectral variation to make shape estimation possible (Figure 2.36).



FIGURE 2.36: Left: Colourful illumination cast by a monitor onto a human face. Right: An estimated height map recovered using the approach in [16].

As the displayed pattern and direction of lighting are known (the screen is parallel to the camera) there is no need for calibration. We, however, are interested in performing shape recovery in more natural environments and the work presented in [16] requires a dark room to function (light provided by the screen is otherwise too indistinct).

A method similar to [16] which requires multiple images is presented in [17]. Face recognition and 3D reconstruction is performed by acquiring images constantly as a white stripe is moved across the screen (acting as a sort of scanning device - see Figure 2.37). Further work demonstrating that computer screens can be used to effectively provide distant point light sources is presented in [67, 68], however these also focus on classic photometric stereo approaches (i.e. not single image) and require some calibration; so are not as relevant. In a similar vein, the work of Papadimitri et al. [69] also deals with uncalibrated near-light sources and takes an iterative optimisation approach to estimate light positions/intensities, surface normals, depth and albedo. However again, this requires multiple images.



FIGURE 2.37: Using a computer screen to provide a known light source - image taken from [17].

Another example of uncalibrated colour photometric stereo is presented by Johnson et al. [18, 70]. They employ a novel approach to force Lambertian reflectance in their image captures. Images are not captured of an object directly, rather the object must be pressed into a layer of elastomer, which effectively lends its own reflectance properties to the surface. This elastomer is then the imaging target and is illuminated by three coloured lights. As the direction and strength of the

lights is always known (they are inside a specially designed, enclosed system - see Figure 2.38) there is no requirement for calibration beyond initial construction of the device. This work is similar to our own in that both employ a constant imaging environment, however the approach taken by [18, 70] is limited to the size of the small elastomer layer and is designed specifically for the reconstruction of micro-geometry of approximately planar objects. Our approach is oriented towards larger objects, passive image capture, can easily be extended to objects of any reasonable size or shape and does not require specialist equipment (i.e. we do not wish it to be necessary for human subjects to rub their face on elastomer to perform facial reconstruction).

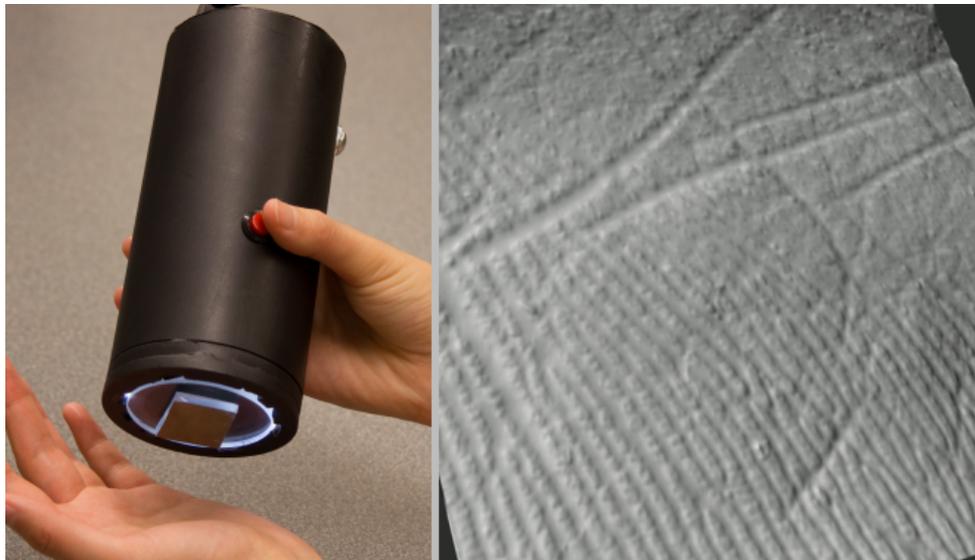


FIGURE 2.38: Left - An elastomer layer is pressed against a surface (in this case a human hand). Inside the device multi-spectral light sources and a high resolution camera capture an image for colour photometric stereo. Right - The recovered shape of the surface. Images taken from [18].

2.4.2 Specularity

The majority of shape recovery algorithms make the assumption that surfaces adhere to the Lambertian model of reflectance [71, 72]. However, in reality there is no perfectly Lambertian surface. For some materials the difference is negligible, but many exhibit noticeable specular properties.

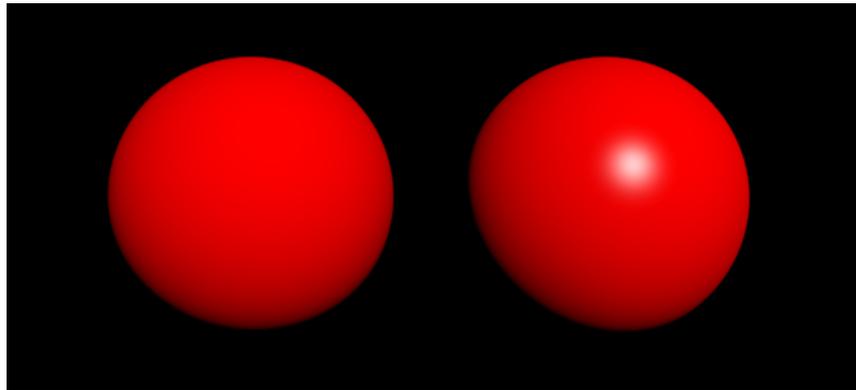


FIGURE 2.39: A diffuse sphere (left) versus a glossy one (right). Note that the highlight is the colour of the light source, this is the case for almost all glossy surfaces.

The presence of specularities on objects confounds Lambertian-based shape recovery methods as the pixel values do not conform to the expected relationship with surface normals (this is also the case when shadows are present). For an illustration of the difference between a diffuse (Lambertian) and glossy (exhibits specularities) surface see Figure 2.39. The diffuse surface appears as expected in the Lambertian lighting model, shading varies smoothly over the surface of the sphere. However the glossy sphere exhibits a specular highlight, something which is not accounted for Lambertian lighting. Thus any shape recovery method which assumes the Lambertian model will not be able to correctly estimate the shape of the glossy sphere.

For many image processing purposes it is sufficient to detect specular pixels and omit them from any further processing. However in our work we require a complete three-dimensional surface to be recovered, thus it is necessary to somehow compensate for these outliers either through replacement or explicit calculation. Let us first look at how specular highlights are typically handled in photometric stereo.

2.4.3 Photometric stereo in the presence of highlights

As discussed, most real objects are in fact not well-suited to the classic Woodham [2] approach due to their surface reflectance characteristics. If the Woodham method is to function correctly, it is expected that measured pixel values conform to the linear relationship represented by the matrix transform E (see Equation 2.3). When pixels include some non-linear component (such as points of specularity on an object) then the recovered surface normals will be incorrect (illustrated in Figure 2.40). As such many variations and modifications to the method have been proposed.

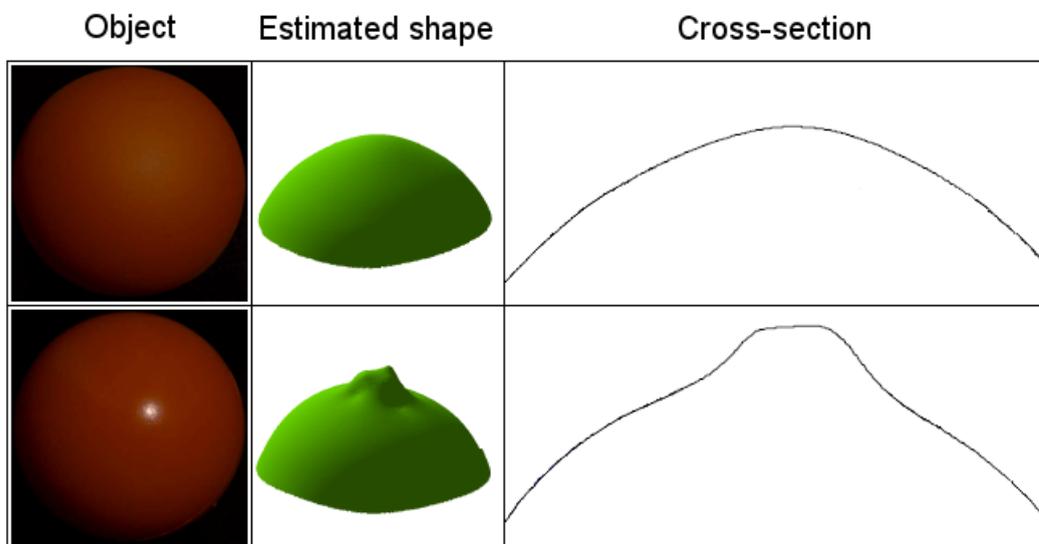


FIGURE 2.40: Here we illustrate the effect that outlying pixel values can have in photometric stereo. The specularity present on the second object clearly causes a large error in the resulting recovery (Images adapted from [19]).

One approach by Coleman and Jain [73] proposes capturing an additional image, giving four greyscale measurements at each object pixel. By capturing four images with four white light sources (again, a different light source is active in each captured image) they are able to detect outlying pixel intensities - i.e. If the pixel intensities vary by a significant amount then this indicates the presence of a highlight. The set of three pixel values which produces the lowest overall intensity are

then used for shape recovery (the fourth measurement is discarded). However no attempt was made to deal with pixels which were subject to shadows.

Barsky and Petrou [74] extend the four-source method to include shadows. They capture four colour images and the detection of outlying pixels is performed by estimating surface chromaticity. Deviation from an estimated surface chromaticity at a pixel across the four captured images indicates that a specular highlight or shadow has occurred. Assuming that a pixel is only subject to a shadow or highlight in one of the captured images, this leaves enough data for shape to be correctly recovered. Barsky and Petrou later published work [75] which gave a detailed performance analysis of their method and offered recommendations for implementing such a system with maximum effectiveness.

Another noteworthy approach to handling shadows in photometric stereo was presented by Hernández et al. [76]. Rather than requiring an additional image to be captured (i.e. they use the standard three images), the method detects pixels in an image which are likely to be subject to shadow by comparison with their local neighbourhood. This data is removed from the algorithm leaving two measurements at the pixel. The ambiguity caused by having only two measurements (surface derivatives cannot be calculated correctly) is addressed by enforcing integrability of the shape (Section 2.2.2.1).

However the obvious issue with these methods is the requirement for multiple images - our single-image approach is therefore incompatible with these methods.

There is a limited selection of literature which deals with single image shape recovery for specular surfaces. Reagheb and Hancock [77] propose estimating the proportion of Lambertian and specular reflectance at each pixel. According to the dichromatic reflection model [78, 79], pixel values are a combination of diffuse (c_L) and specular (c_S) components,

$$\mathbf{c} = m_L c_L + m_S c_S, \quad (2.26)$$

where m_L and m_S are geometric scaling factors which determine the strength of each component. The more closely aligned the surface normal, light direction and viewing direction are, the greater the value of m_S (Figure 2.41).

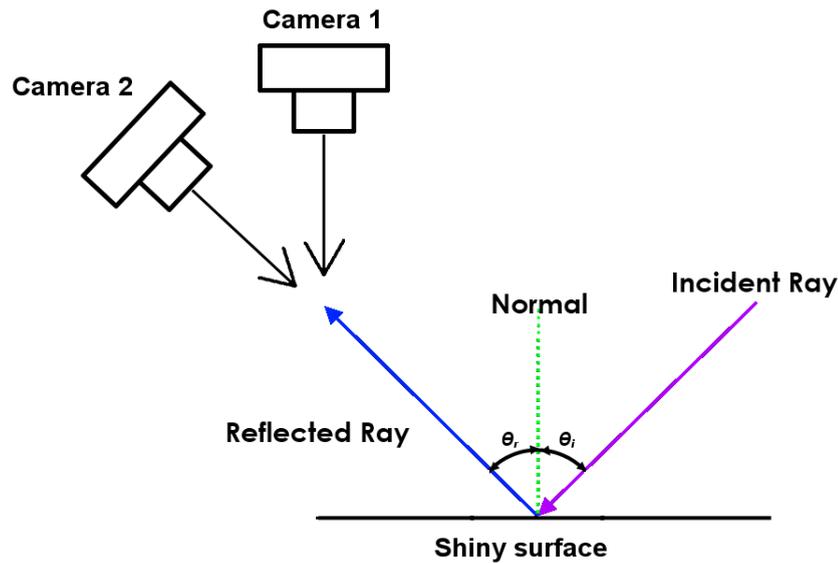


FIGURE 2.41: Shiny surfaces reflect light more strongly according to the angle of incidence at a surface normal. If the same point in this diagram is imaged by the two cameras, the corresponding pixel value will have higher intensity in camera 2's image.

Thus if the specular component of a pixel can be determined, it can be removed and the Lambertian component will remain.

Vogel et al. [80] attempt to solve for the Phong reflectance model (essentially the Lambertian model with additional terms which account for specularities), assuming some distortion of the object caused by perspective projection (we assume orthographic projection - see Chapter 6). Wang et al. [81] also propose use of the Phong model to account for specular highlights but require a single light source which is aligned with the viewing direction.

In Chapter 6 we focus on detecting and removing specular highlights from images as a pre-processing step. Methods for detection and removal of highlights can be broadly separated into two categories. Those which rely on inherent removal of highlights and those which utilise separate detection and removal stages.

Separate detection and removal

Methods with distinct stages of detection and removal of specularities typically rely on some form of inpainting algorithm to attempt correction pixels which are determined to have been affected. One example of this can be seen in the work of Madooei and Drew [82] where specular highlights are detected using matrix factorisation in images assuming a single light source. The offending pixels are removed and an inpainting algorithm is applied to restore the image.

In the work of He et al. [20] an ‘assisting gray image’ is created to provide a fourth image channel (in addition to RGB). They then use independent component analysis (ICA) to detect specular pixels (the image is split roughly into two components, specular and diffuse). This approach is of particular interest to our work as it does not strictly require only a single light source in the scene. He et al. [20] only present results for simple images of foods which possess largely homogeneous surfaces, however we are also concerned with surfaces which have uniform reflectance properties.

Once specular pixels have been detected and removed from an image, the common approach is to then apply an inpainting [83] algorithm to fill the blank areas (Inpainting refers to the use of valid image pixels to ‘paint’ over missing pixels). We note that many different image inpainting algorithms have been proposed but the details are beyond the scope of this thesis. For an overview of the area see the survey paper by Pandya and Limbasiya [84].



FIGURE 2.42: Left: Original image with highlights. Middle: Specular pixel mask. Right: Specular free image [20].

Inherent removal

A common approach to specular removal is the manipulation of colour space to give images which are free of highlights. These are referred to as inherently removing highlights as there is no explicit detection step involved.

In the work of Mallick et al. [22] and Zickler et al. [19] RGB colour space is subject to a rotation such that one axis (e.g. the red axis, R) aligns with the colour vector of the light source present in the scene. This results in a new ‘SUV’ colour space in which the S channel contains any specular highlights, with the U and V channels being entirely diffuse. We present this method in more detail and adapt it to our own work in Chapter 6.

Another colour-space focused approach is presented by Tan and Ikeuchi [85]. In short, they replace specular pixels with diffuse (Lambertian) counterparts by estimating the maximum diffuse chromaticity of every pixel. This is accomplished by iteratively adjusting the chromaticity of a pixel to tend towards the maximum chromaticity present within its neighbourhood. Improved results were shown in the work of Yang et al. [86] by applying a low-pass filter to the maximum fraction of the RGB components in the original image. This adjustment allows diffuse chromaticity information to propagate from existing diffuse pixels present in the

original image, however the use of filtering can cause issues for surfaces with multiple colours; which is avoided in the original work. Both versions possess limitations in that they can only function on chromatic surfaces and that the chromaticity of the single scene light source must be known.

Accounting for specular highlights in shape recovery from a single image is a difficult, underconstrained problem. We implement two different approaches for handling this in Chapter 6 with some limited success.

Chapter 3

Engineering an environment

Work in this Chapter was published in the paper ‘Shape in a Box’ at the 4th Color and Photometry in Computer Vision Workshop (CPCV) - European Conference on Computer Vision (ECCV) 2014

We have established our initial hypothesis - that we should be able to use the effects of mutual illumination within a scene for the purpose of shape recovery. In this chapter we describe our first steps towards developing such a system. We engineer a portable capture environment which is built specifically for performing photometric stereo by exploiting spectrally varying mutual illumination. The end result is a colourful, triangular box.

In Section 3.1, we take existing work into consideration to postulate possible environments which may generate suitable conditions (i.e. enough mutual illumination to facilitate shape recovery). We then proceed to select a metric which will allow us to make a judgement on the effectiveness of our proposals. Next in Section 3.3 we detail the generation of synthetic data. Specifically we describe the mechanics of radiosity rendering, a computer graphics method which can accurately replicate the appearance of mutual illumination within a scene. We then show how we employ this rendering method to synthesise test environments.

Finally, in Section 3.5 we present the results of our investigation and use them to meet our goal of developing a portable capture environment which generates sufficient colour complexity (via mutual illumination) for shape recovery.

3.1 Potential environments

As discussed in Section 2.3, mutual illumination effects can easily become muted in large or brightly lit scenes. In this chapter our aim is to engineer a colourful box into which candidate recovery objects could be placed and imaged, and where there will be enough colour complexity to recover the object's shape. We also want the box to be of a suitable size for portability. If it can be easily picked up and placed anywhere then we are meeting our aim (Chapter 1) of moving shape recovery closer to real-world conditions.

For our shape recovery system to function we need there to be an unambiguous relationship between the imaged pixel values and surface normals of candidate recovery objects. For example, were we to place a cube inside our hypothetical box, each face of the cube should appear to be differently coloured. This is illustrated in Figure 3.1, every pixel of a single face has the same combination of surface normal and rendered RGB value.

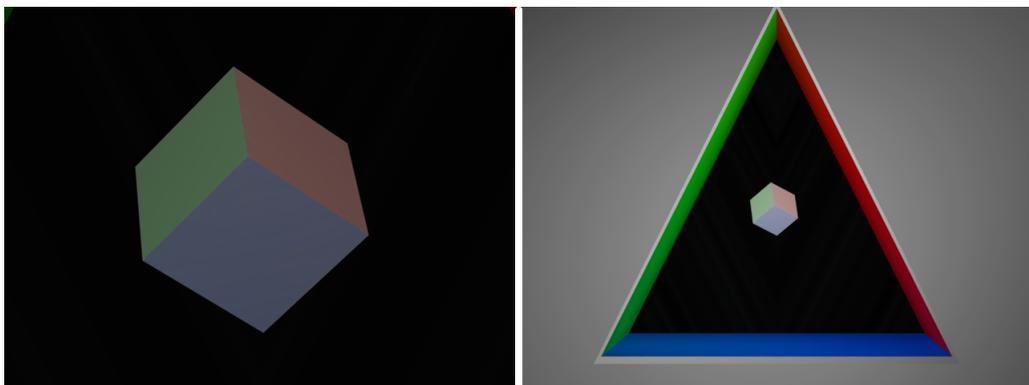


FIGURE 3.1: Examples of synthesised images. Left: Close-up of a rotated cube. Right: A wide-shot of the cube positioned inside a colourful box. Each face of the cube has the same RGB value.

Our hypothesis is that a colourful box should be capable of providing sufficient variation in illumination (i.e. every surface normal corresponds to a different RGB value).

How can we ensure that mutual illumination will be present to a sufficient degree? The indirect light must be the result of some primary light source or sources. So, how can we be sure that the mutual illumination in our box won't be 'drowned out' by direct light? The answer of course, is that in extreme cases (e.g. one or more high intensity lights in close proximity to the box) we cannot guarantee this. We can, however, simulate much more likely conditions, such as an artificially illuminated room.

Light entering through the top of the box will inevitably be of higher intensity than the light reflected from the colourful interior. Thus, objects placed at the bottom need to receive sufficient mutual illumination from the sides and not have that illumination be lost in the strength of the light from the exterior environment. Equally, we do not wish the box to be too deep as in this case the object placed at its bottom could then be too dimly lit.

Based on the work of Drew [48, 87, 88], there is one enclosure design which seems particularly likely to meet our needs. In *Shape from Color* [48], three multispectral, direct light sources provide sufficient variation in colour across the surface of an object to enable shape recovery (see Section 2.2.5). Thus an obvious candidate is a triangular box with three distinctly coloured interior walls. An initial attempt at synthetically engineering such an environment is shown in Figure 3.1. The appearance of the cube meets our expectations. However, is there some way that we can measure the recovery potential for variations on this design and thus establish a box which is likely to succeed?

3.2 Performance metric

For comparison of the conditions generated by different environments, we need some metric to measure the potential for shape recovery performance.

In Shape from Color, surface normals are estimated from captured RGB values via a matrix transform. For convenience, the relevant equation (from Chapter 2) is reproduced here,

$$\mathbf{n} = G \mathbf{c} \equiv F^{-1} \mathbf{c} \quad (3.1)$$

An RGB pixel value \mathbf{c} is transformed by a matrix G to give a surface normal \mathbf{n} , with G being the inverse of F - the matrix which can be measured via calibration of the system (thus relating the surface normals and RGB values).

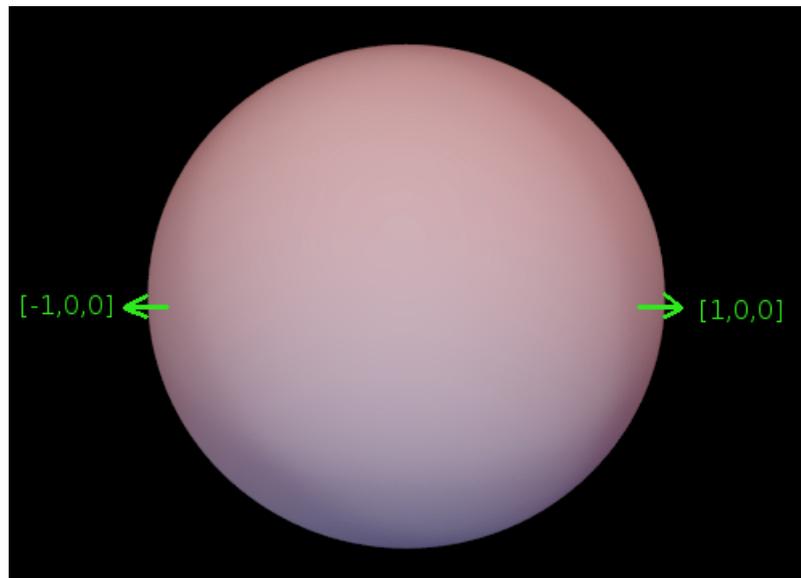


FIGURE 3.2: A white sphere subject to mutual illumination. Red light is causing surface normals which have opposing x-axis components to appear identical. Thus we will be unable to correctly estimate shape in this lighting environment.

We are assuming for now that target recovery surfaces can be modelled using Lambertian reflectance and thus reflected light should vary smoothly over the surface of the object. A small difference between two imaged RGBs should equate

to a similarly small change in their estimated surface normals. Additionally, we do not want there to be ambiguity between an RGB value and its corresponding surface normal (for example, if red light appears on opposing sides of a frontally-presented sphere, we will be unable to tell which surface normals are representative of red pixels - see Figure 3.2). We can express this constraint mathematically as,

$$\hat{\mathbf{n}} = G(\mathbf{c} + \epsilon) \simeq \mathbf{n}. \quad (3.2)$$

That is to say, if we perturb our measured RGB values, \mathbf{c} , by some small amount ϵ , we expect there to be little change in \mathbf{n} . This is illustrated in Figure 3.3 where the smoothly changing surface normals of a sphere result in a corresponding smooth change in pixel colour.

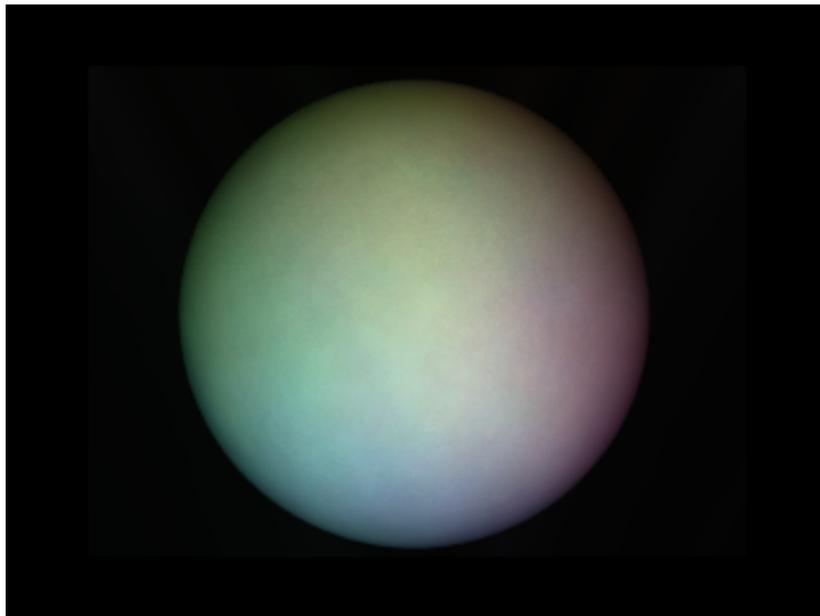


FIGURE 3.3: A synthesised sphere with Lambertian reflectance subject to colourful mutual illumination. Colour varies smoothly in correspondence with surface normals.

We propose measuring the “quality” of a lighting environment by testing the hypothesis that F provides a stable transform between RGBs and normals. Mathematically, F should be well conditioned.

All matrices can be decomposed using a singular value decomposition,

$$F = UDV^T, \quad (3.3)$$

where U and V are 3x3 orthogonal matrices and D is a diagonal matrix. The values, σ_i along the diagonal of D are referred to as the singular values of F . Taking the ratio between the minimum and maximum of these singular values gives the condition number k of F ,

$$k(F) = \frac{\sigma_{max}}{\sigma_{min}}. \quad (3.4)$$

This condition number is a measure of the stability of a matrix inverse. Approximately, if $k(F) \geq 10$, then $\hat{\mathbf{n}}$ in Equation 3.2 can be about 10% different from \mathbf{n} . This is the criteria that we shall choose to build our box. We shall seek an illumination environment that supports an L^2 condition number of no more than 10,

$$\|F\|_2 \cdot \|G\|_2 < 10 \quad (3.5)$$

The final issue to resolve is how exactly can we accurately synthesise the effect of mutual illumination? Simple Lambertian rendering will not suffice as this would only model the effects of direct lighting. It will be necessary to employ a more complex rendering method.

3.3 Radiosity rendering

In the work of Forsyth and Zisserman [30, 60] it was suggested that incorporation of mutual illumination effects into shape recovery methods could help improve

results. In Section 2.3 we presented the equation used by Forsyth and Zisserman to model mutual illumination. The same equation is used in a Computer Graphics method - ‘radiosity rendering’ [89].

Essentially, in radiosity rendering we calculate the spread of energy throughout an environment as it bounces between surfaces. The result is that over successive iterations the environment is flooded by light. This can be seen in Figure 3.4 where a simple, low polygon scene is shown over multiple rendering passes.

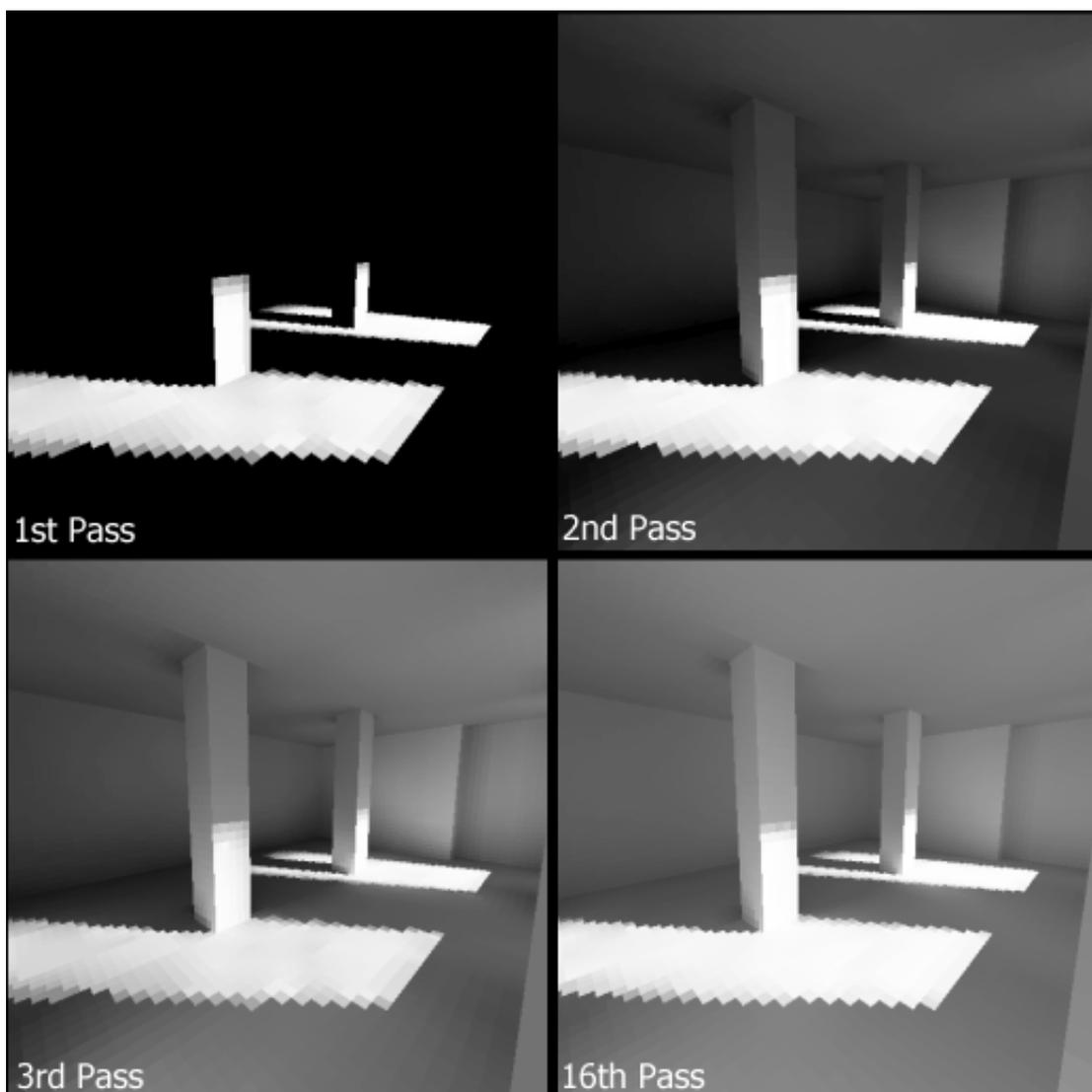


FIGURE 3.4: A low polygon (hence the shadows appearing stepped/aliased) environment over successive radiosity rendering passes. The room becomes brighter as the transfer of energy is modelled for successive iterations of the radiosity rendering equation (Image adapted from work by [Hugo Elias](#)).

The equation which forms the basis of radiosity rendering is reproduced here for the convenience of the reader,

$$R(x) = T(x) + \alpha(x) \int_S R(x') \frac{1}{\pi d^2} \cos\theta_x \cos\theta_{x'} \cdot V(x, x'). \quad (3.6)$$

The term ‘radiosity’ is defined as the energy per unit area leaving a surface in unit time. In Equation 3.6 we calculate $R(x)$, the total radiosity at a point x (x' simply refers to some other point in the scene). T is light emitted directly from the point and α is the albedo of the point. An integral is taken across all surfaces in the scene S with d^2 being the distance between our point of interest x and another point x' . $\theta(x)$ and $\theta(x')$ are the angles between the line joining the two points and their respective surface normals. Finally $V(x, x')$ is simply a visibility function set to either one or zero depending on if the two points can see each other or not. Figure 3.5 visualises this concept. Essentially we are calculating the amount of light which is received by one point in a scene via emittance (direct light sources) or reflectance from other points.

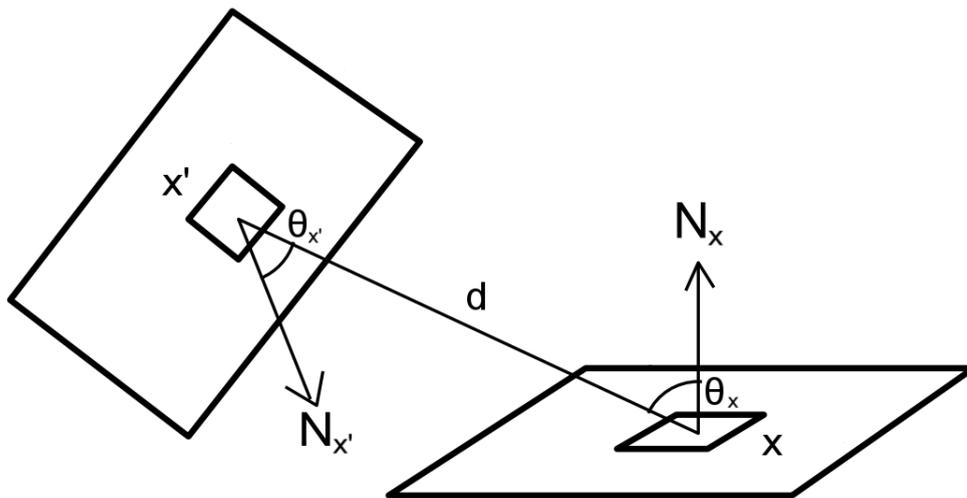


FIGURE 3.5: Calculating the energy received at a point x from some other point x' . N refers to the surface normal at a point, d is the distance between the two, θ is the angle between a surface normal and a line connecting the two points. Repeating this for all other x' in the scene will give us the total radiosity at x (Image adapted from work by [Hugo Elias](#)).

To be of practical use it is necessary to discretise this equation. This is achieved by dividing the contents of a scene into n triangular or quadrilateral planar patches and calculating the irradiance and radiance for each patch ¹

$$R_i = T_i + \alpha_i \sum_{j=1}^n V_{ij} R_j \quad (3.7)$$

Again, R_i is the total radiance at patch i , T_i is emitted light (will be zero unless the patch is a direct light source), α_i is the albedo (or reflectivity) of a point and V_{ij} is a geometrical view factor.

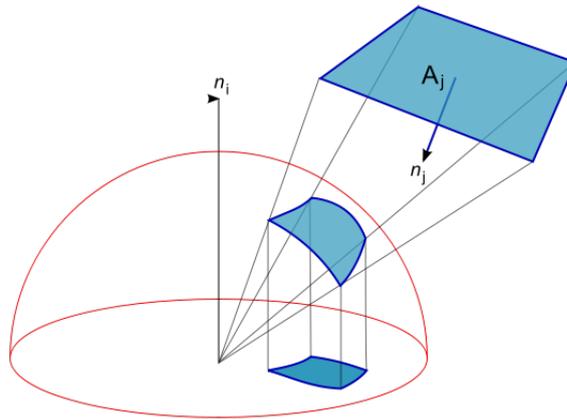


FIGURE 3.6: Calculating a hemispherical view factor (Image by Wikipedia user [Jheald](#)).

There are multiple methods for calculating the view factor between two patches. The most common of these are the use of a unit hemisphere or hemicube centered about the current patch. The hemisphere method is illustrated in Figure 3.6 where the view factor between our current patch i and some other patch A_j is being calculated. A hemisphere has been created around patch i , with the direction $[0, 0, 1]$ on the hemisphere corresponding to the surface normal \mathbf{n} of the patch. A_j

¹Radiance refers to the total energy leaving a surface (in our case the sum of emitted and reflected light at a point) and irradiance refers to the energy which is incident to a surface (here this is the light arriving at a point in the environment from all other points).

is projected along its surface normal \mathbf{n}_j onto the hemisphere and the proportion of this projection which covers the unit circle at the base of the hemisphere is used as the view factor.

The equation for obtaining this hemispherical view factor is given as thus,

$$V_{ij} = \frac{\cos(\theta_i) \cos(\theta_j)}{\pi r_{ij}^2} \text{area}(A_j). \quad (3.8)$$

Again, θ_i and θ_j are the angles between the respective patch's surface normals and a line drawn directly between the centre of the two. $\text{area}(A_j)$ is simply a scaling factor.

In computer graphics the hemicube method is used, as projecting onto a hemicube is much more computationally efficient; but we feel that the hemisphere makes for a more intuitive explanation of the concept. The system of equations for all n scene points can be solved over successive iterations until some minimum change between rendering passes is met.

In our work we want to test whether mutual illumination effects can be used for the purpose of shape recovery. Thus, radiosity rendering provides an excellent way to generate synthetic data which will allow us to investigate the efficacy of such a system.

3.4 Synthesising data

We want to find an enclosure to contain candidate recovery objects that is likely to create sufficient spectrally varied mutual illumination and a suitable level of light intensity across the surface of the object. Thus, before proceeding with any construction of a real-world box, we simulate conditions using 3D modelling software - specifically 3DS Max 2012 - to generate synthetic data (Figure 3.7). Previous

work [90] has demonstrated that 3DS Max is capable of producing sufficiently realistic radiosity-based renderings to justify its use for this purpose.

To establish suitable dimensions for the box we rendered a series of synthetic images of a sphere placed in a box where the length and height of the triangular box walls varied whilst the sphere diameter remained at 15cm (a size appropriate for the objects with which we intend to conduct experiments). We used the ‘mental ray’ renderer, with perfect Lambertian materials assigned to all objects (zero specular component). The light source used in our synthetic scene was the mental ray sunlight system.

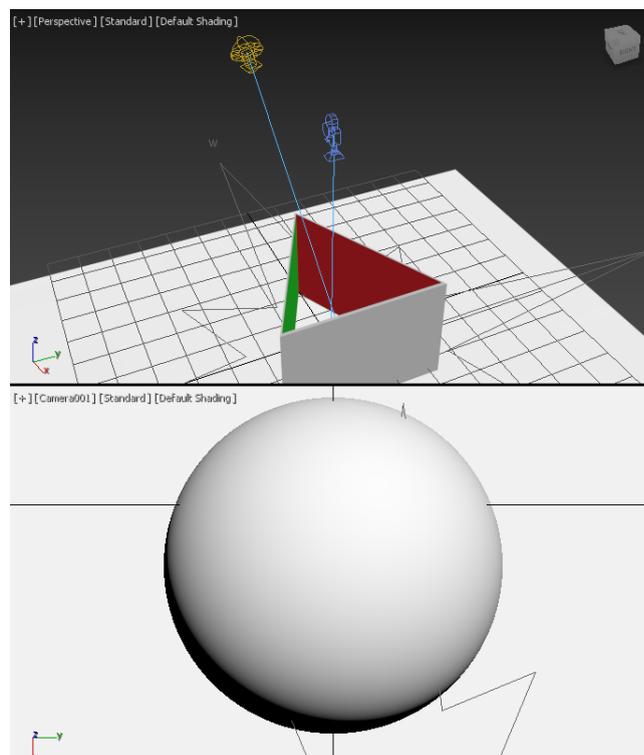


FIGURE 3.7: Views from 3DS Max of the experiment setup. Top - A “daylight system” provides lighting whilst a camera points directly into the box. Bottom - From the perspective of the camera we see a sphere inside the box (this is the view that is rendered to produce our data).

Although a more realistic simulated environment could be built (i.e. a full indoor or outdoor environment) the isolated box and sphere are sufficient for providing proof of concept. Another limitation of our synthetic experiments is any lack of variance over time. In any uncontrolled environment it is inevitable that conditions

will change in at least some small way between imaging of a calibration object and an experiment object (e.g. cloud movement outdoors). Again we ignore this for now though it could also be simulated if desired (mental ray light systems allow for change in the time of day for example).

3.5 Results

After producing a set of synthetic data using radiosity rendering, we can now determine the viability of mutual illumination photometric stereo. We proposed that a condition number, $k(F)$, less than 10 would provide a suitably stable shape recovery system. The condition numbers, $k(F)$, generated from our synthesised boxes are shown in full in Table 3.1 and these results are also illustrated in Figure 3.8.

		Side length (cm)			
		40	60	80	100
Height (cm)	20	21.830	18.570	16.245	14.750
	40	15.738	12.027	9.760	8.719
	45	15.640	11.741	9.425	8.307
	50	15.597	11.594	9.200	8.038
	55	15.578	11.524	8.970	7.849
	60	15.568	11.488	8.920	7.726
	65	15.569	11.472	8.910	7.639
	70	15.569	11.466	8.900	7.585
	75	15.570	11.460	8.876	7.552
	80	15.574	11.457	8.810	7.530
	100	15.590	11.450	8.800	7.500

TABLE 3.1: Condition number, $k(F)$, given by triangular boxes with varying dimensions.

In the table we can see that as the size of the box increases, condition numbers improve. There appears to be a convergence toward some optimal dimensions, this is visualised in Figure 3.8. We expect that this is due to a balancing point between levels of external light and mutual illumination from the enclosure falling across the surface of our calibration sphere.

The minimal condition number results from a side length and height of 100cm and the data indicates that preferable conditions could possibly be obtained from a larger enclosure. However, we must keep in mind the long-term goal of our research - to help move shape recovery into the wild. We wish for the box to be easily portable so that it can be used in any conditions. Thus, as a compromise between practical considerations and condition number we settled on constructing a box with side length 80cm and height 60cm. In Table 3.1 this combination of dimensions gives a condition number of 8.810, which obeys our proposed requirement (condition number of 10) for creating a stable matrix transform. There are more promising results shown in Table 3.1 but the larger dimensions of these will impact on our requirement for portability.

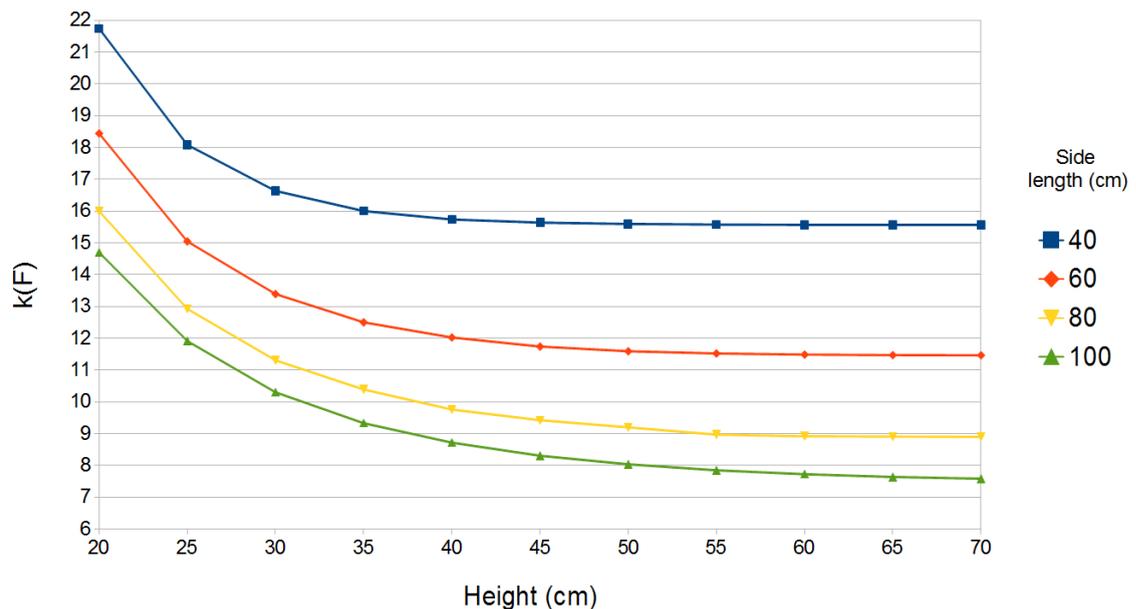


FIGURE 3.8: Results of the synthetic experiments on triangular boxes of various dimensions. Four different side lengths were each tested with the same varying heights.

To ensure we are correct in our assertion that a matrix transform with a condition number less than ten will be capable of producing a decent level of accuracy, we here display the meshes recovered from perfect synthetic spheres for two of our experiments. (Figure 3.9). For details on how this process is performed see Chapter 5).

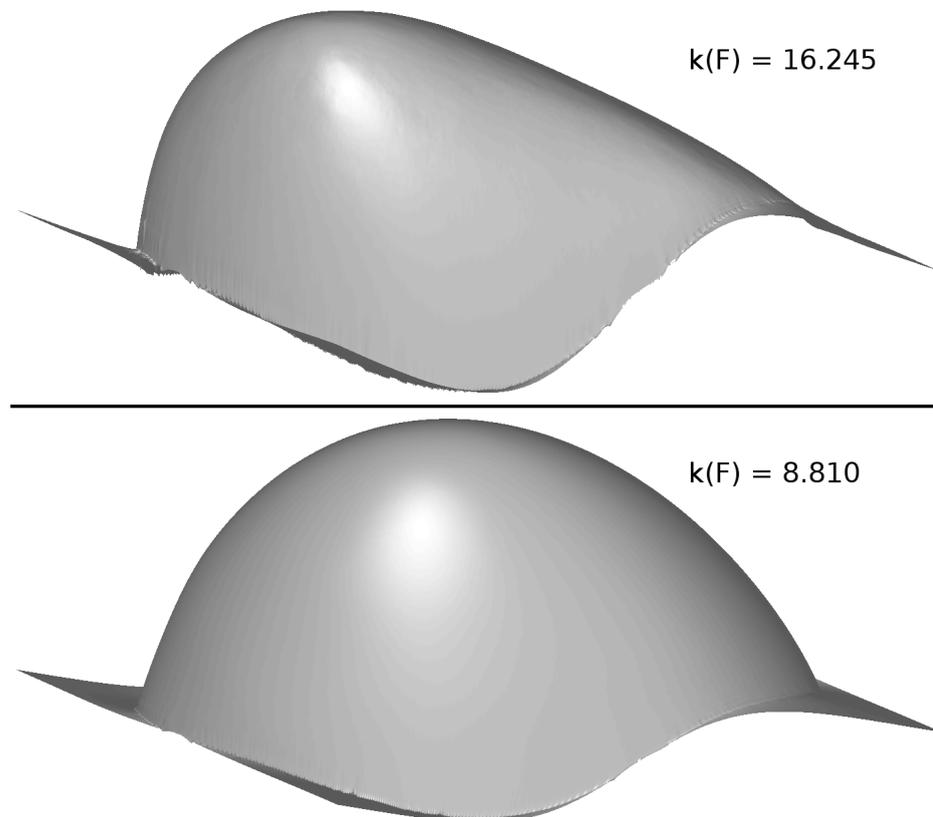


FIGURE 3.9: Examples of synthetic shape recoveries performed on spheres in environments with different condition numbers. A better conditioned transformation matrix gives a more accurate result.

As expected, $k(F) \leq 10$, provides a suitably stable transform from RGB values to surface normals. The meshes in Figure 3.9 were calculated by applying the Fourier-based reintegration method of Frankot and Chellappa [25]. The boundary conditions in their method is what causes the “waviness” around the base of the shape - an explanation of this can be found in Chapter 2.

3.6 Conclusion

In this Chapter we used radiosity rendering to create synthetic data for the purpose of an initial investigation into the viability of mutual illumination photometric stereo. Taking cues from the work of Drew [48, 87, 88] we propose a triangular box with colourful walls.

We have found that, as hypothesised, it is possible to recover shape through exploiting the effects of multispectral mutual illumination. As a result of this work, we have established a suitable design for an enclosure with which to conduct real-world experiments and produced evidence that Forsyth and Zisserman [30, 60] were correct in their proposal that the incorporation of indirect illumination into shape recovery methods can help to improve results.

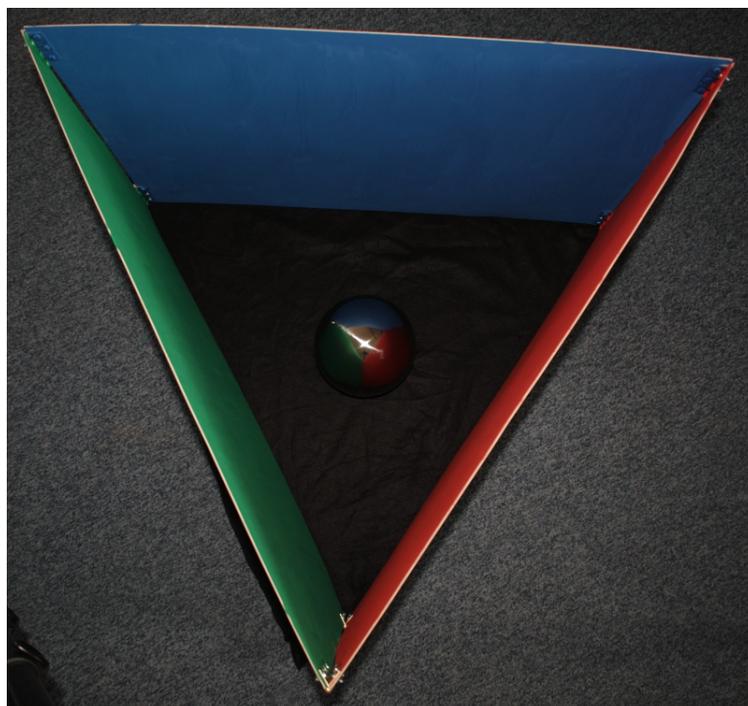


FIGURE 3.10: The end result of this chapter - a real box designed for mutual illumination photometric stereo.

Chapter 4

Data sets

Work in this Chapter was published in the paper ‘Shape in a Box’ at the 4th Color and Photometry in Computer Vision Workshop (CPCV) - European Conference on Computer Vision (ECCV) 2014

An issue that is often glossed over in shape recovery literature is exactly how can we truly determine the accuracy of algorithms? Typically the validity of results is determined through one of three possibilities.



FIGURE 4.1: 3D printing offers real objects with almost perfect ground truth shape data. Left: Original 3D model file for the ‘Bust’ object. Right: 3D-printed counterpart.

One could provide measures of accuracy obtained from synthesised objects. This is the only option which provides a definite ground truth to compare against as the exact original values are known. In photometric stereo for example, we would render an image using synthetic lights and a normal map (which can be calculated from a height map - Chapter 2). We would then attempt to recover the surface normals from the image (and from these the height map) and compare the estimates directly with the original values.

Although synthesised data provides an attractive option, its flaw lies in its perfect nature. A common way to mimic real data is to add noise to synthetic images (e.g. [91]), however there are many more possible issues when algorithms are transferred across to the real-world. For example, real-world images will not contain objects with perfect Lambertian reflectance.

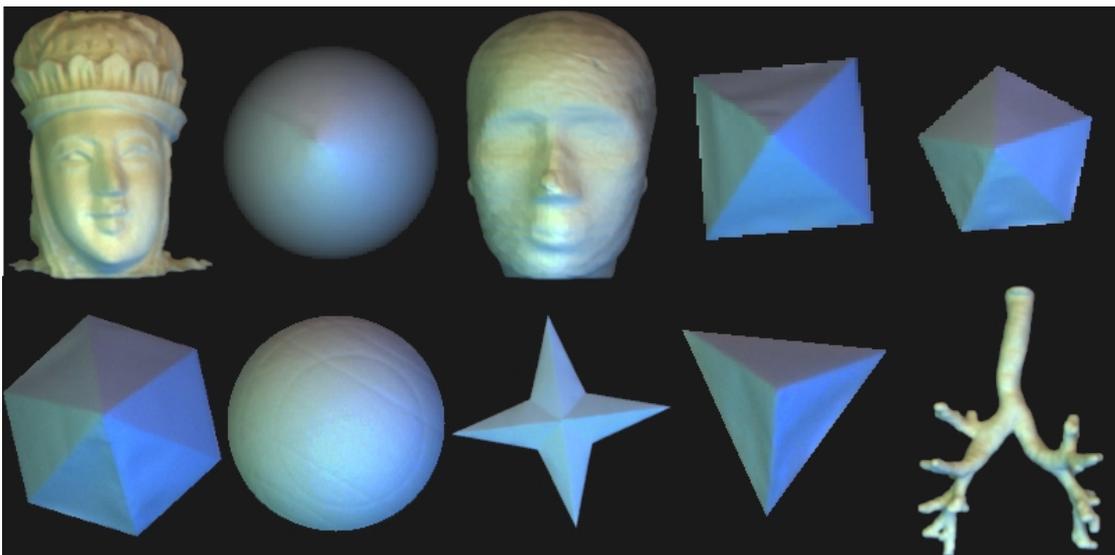


FIGURE 4.2: Images of experiment subjects (3D-printed objects in *italics*, papercraft in normal type, abbreviated names given in (brackets)). Top row, left to right: *Bust*, *Cone*, *Face*, 4-sided pyramid (Pyr4), 5-sided pyramid (Pyr5). Bottom row, left to right: 6-sided pyramid (Pyr6), *Sphere*, *Star*, Tetrahedron (Tetra), *Trachea* (*Trach*).

A second option for measuring algorithm accuracy is to use existing real (i.e. not synthetic) image datasets which provide ground truth for real objects. Comparison to other methods which use the same dataset is common in shape recovery

literature. However, such datasets are not universally compatible and so the scope of comparison is constrained (Indeed, we found no available datasets which were suited to testing our mutual illumination shape recovery).

This also raises the question of what exactly constitutes ground truth shape for real objects? Some form of shape recovery must be used initially to provide the “true” measurement. In the case of photometric stereo we typically rely on methods which report higher levels of accuracy (such as laser scanning) to provide such data.

The final possibility is to judge results based purely on their appearance, i.e. opinions on the visual fidelity of the recovered shape. Whilst this is the least scientific approach it is sufficient for many applications, such as producing 3D models for use in video games or animation - where aesthetics are the primary concern rather than accuracy.

4.1 A novel dataset

In our work, with no suitable prior data available and not wishing to rely purely on synthesised images (such as those used in Chapter 3 for engineering our capture environment); we came up with what is to our knowledge, a novel approach. We perform shape recovery using “papercraft” models. Templates for such objects can be printed onto card or paper and constructed by hand. In addition we also image a number of 3D-printed objects and compare acquired estimates to the original model files.

Figure 4.2 displays many of the objects used in our experiments and a snapshot of the assembly and recovery process for each type of object is shown in Figure 4.3.

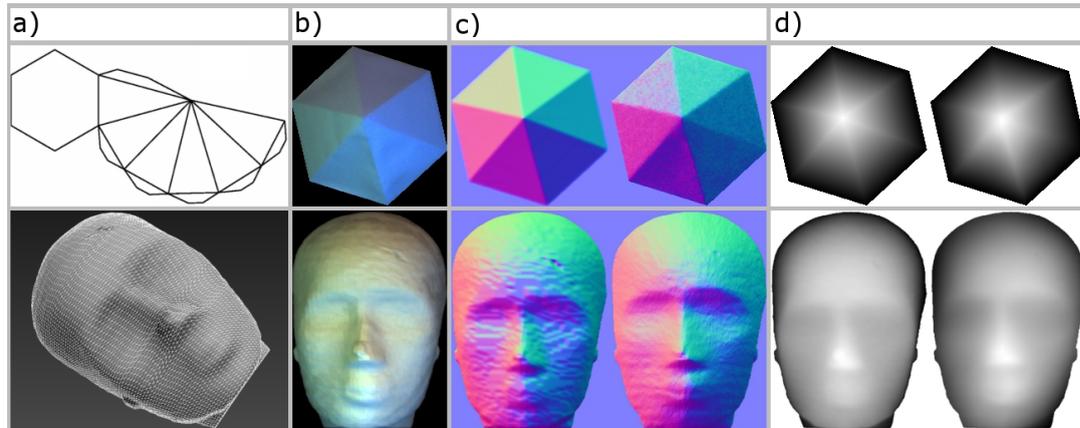


FIGURE 4.3: Top row: a simple papercraft object. Bottom: a 3D-printed object. a) Source files, a papercraft template and a 3D model file. b) Captured images. c) left = true normal map, right = recovered normal map. d) left = true height map, right = recovered height map.

4.1.1 Papercraft Objects

Papercraft models provide cheap, readily accessible experiment subjects. They require only a regular ink printer, scissors and glue. Templates such as those shown in the figures throughout this section (Figures 4.4, 4.5, 4.6) can be printed freely from many sources ¹.

As well as being used for construction, measurement of these templates provides us with ground truth to compare recovered shapes against. For example, in Figure 4.4 all edges of the tetrahedron share the same length; so we can scale the recovered shape to the same unit space for comparison. Further details of this comparison procedure are presented in Chapter 5.

¹e.g. "Paper Models of Polyhedra" - <http://www.korthalsaltes.com>

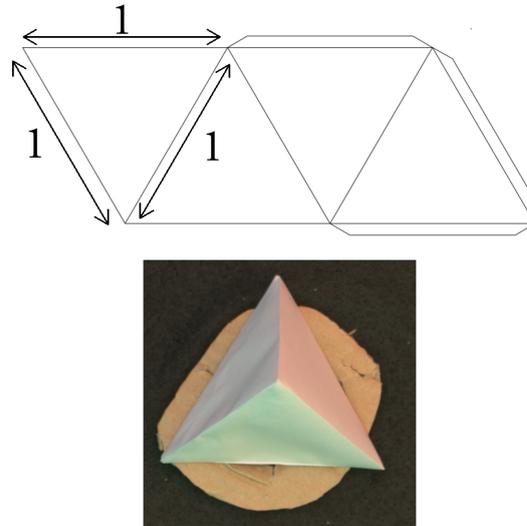


FIGURE 4.4: Top: Template for a tetrahedron. Each edge of the tetrahedron is the same unit length. Bottom: The assembled counterpart. The surface colour is due to the mutual illumination provided by our box environment.

While the perfect 3D shape of the papercraft object is known, it should be noted that there is likely to be some human error involved in fabrication (such as the slight warping of faces which can be seen in Figure 4.4). Nevertheless they are useful as quick and easy to obtain test subjects.

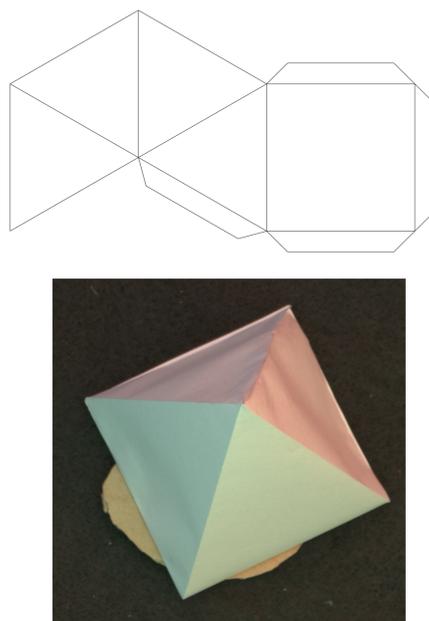


FIGURE 4.5: Top: Template for a four-sided pyramid. Bottom: The assembled counterpart.

In our experiments we used a number of simple papercraft objects. Templates for complex objects are also available, however the additional complexity would be very likely to result in an unacceptable level of error between the template and the constructed object.

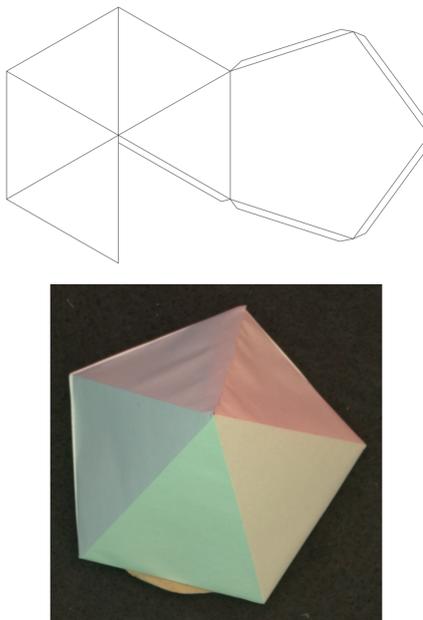


FIGURE 4.6: Top: Template for a five-sided pyramid. Bottom: The assembled counterpart.

4.1.2 3D-Printed objects

3D-printing offers a new approach for creating ground truth data in shape recovery. Any desired shape can be modelled and printed using a wide range of materials² (or even combinations of materials) and the original 3D model file provides us with an exact benchmark to compare recoveries against.

However, as with papercraft objects there is still some risk of fabrication errors causing printed objects to deviate from their original 3D model. In this section we briefly examine current popular 3D-printing methods and justify our choice. We then conclude the chapter by presenting our 3D-printed test subjects.

²One could even print an object using chocolate - <http://chocedge.com>

3D-Printing methods

Accuracy in 3D-printing (i.e. how closely the printed object conforms to its 3D model file) varies depending on the printer, material and the suitability of the model file. For example, currently two of the most popular forms of 3D printer are “stereolithography” (SLA) [92] and “fused deposition modeling” (FDM) [93].

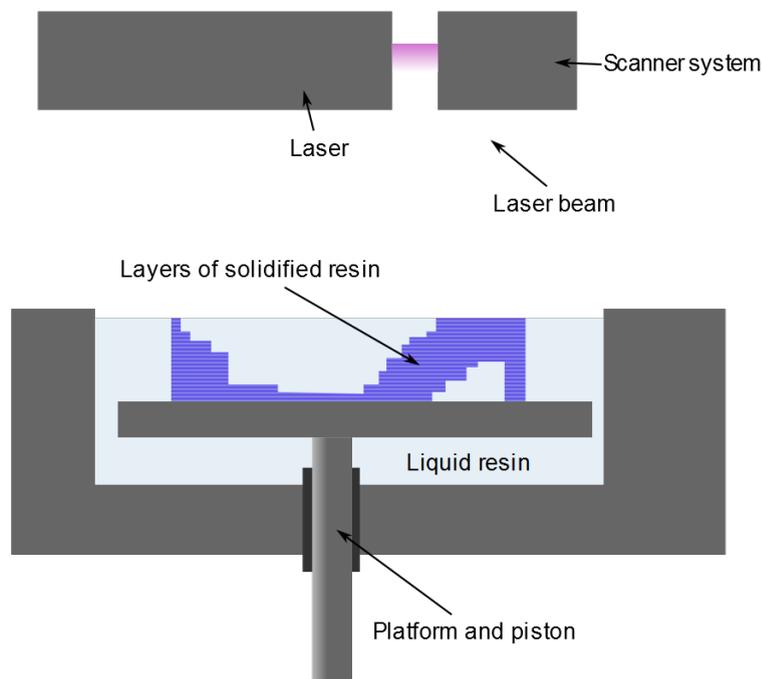


FIGURE 4.7: Stereolithography (SLA): A layer of resin is bound together by the laser according to the object design, the platform moves down a layer and the process repeats (Image taken from [proto3000](https://proto3000.com)).

In SLA systems (Figure 4.7) a laser binds together layers of resin to print an object from the bottom upwards. FDM systems (Figure 4.8) similarly produce prints from the bottom upwards, but by depositing strings of molten material via a nozzle. Typically FDM printers are not as capable of producing fine detail as SLA systems ³.

³<https://all3dp.com/fdm-vs-sla/>

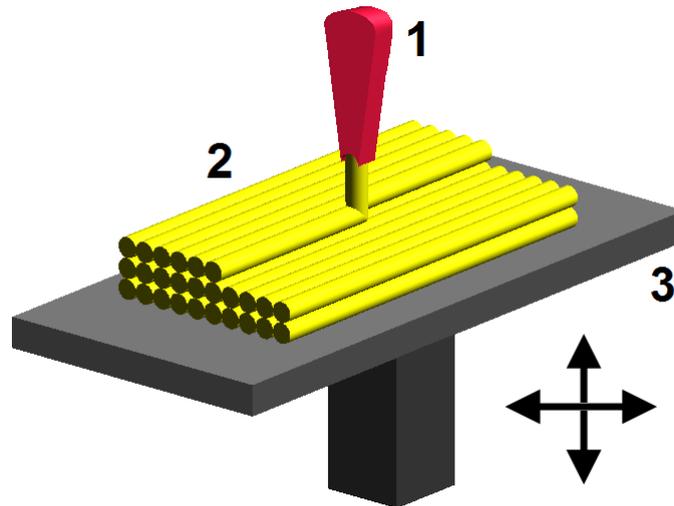


FIGURE 4.8: Fused deposition modelling (FDM): 1 - nozzle applies material, 2 - material has been deposited according to the object design, 3 - movable table (Image by Wikipedia user [Zureks](#)).

The model file itself must also be suited to the chosen printing method. For example, it is common practice to produce hollow 3D-printed objects to minimise the cost of materials. However this can cause issues with structural integrity, thus it may be necessary to add internal support structures to the original 3D model file.

For our purpose it is particularly important that a suitable material is used for printing. We require objects with approximately Lambertian reflectance. As can be seen in Figure 4.9, both of these methods produce glossy objects.



FIGURE 4.9: Typical 3D-prints. Left: Produced by FDM (Taken from [Pro Image 3D](#)). Right: Produced by SLA (Image taken from [Fibilo](#)).

For our objects we used a ‘ZCorp Zprinter 450’. This printer is an example of a “powder bed” or “binder jetting” system, it builds objects by binding together successive layers of a proprietary powder (ZCorp has since been acquired by the 3D Systems Corporation, it is unclear whether the material is still available). Unlike FDM and SLA printers, powder bed systems completely fill each layer of the print area and the appropriate parts of the object cross-section are bound together; meaning that it is necessary to remove excess powder once printing is complete (though this does also serve to support the print during construction).

The powder material used by our chosen printer results in approximately Lambertian reflectance and it is capable of accurately producing fine details. This means that our objects are accurate representations of their original model files. Thus comparison of recovered shapes against the ground truth model files provides an excellent measure of recovery accuracy.

In Figures [4.11](#) and [4.10](#) we show two examples of our experiment subjects alongside their original 3D model files (a further example is shown in Figure [4.1](#)).



FIGURE 4.10: Left: Original 3D model file for the ‘Face’ object. Right: 3D-printed counterpart.

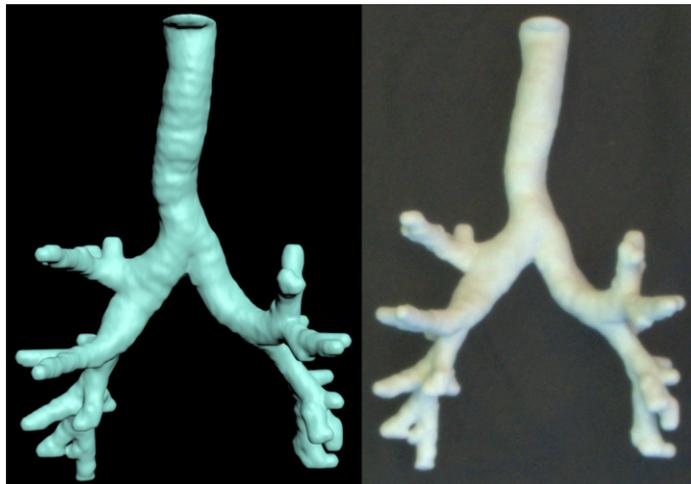


FIGURE 4.11: Left: Original 3D model file for the ‘Trachea’ object. Right: 3D-printed counterpart.

4.2 Conclusion

In this chapter we presented novel sources of real ground truth data - papercraft and 3D-printed objects. We discussed their potential as shape recovery test objects

and presented the objects which are used in our experiments.

This contribution is not limited to the development of our own dataset. These object types could be used for almost universal comparison between shape recovery methods. Currently it is common to evaluate methods using existing image datasets, this limits comparison scope - e.g. one can only compare results with other methods which require similar objects (shape and reflectance) and are subject to the same scene conditions. Our proposed objects could be use as subjects for a wide range of shape recovery methods - reflectance can be adjusted if necessary (print using a different material) and the scene conditions become irrelevant due to the constant ground truth shape.

In the next chapter we use these objects to create our dataset and provide a comparison between the accuracy of our method and an established photometric stereo algorithm [48].

Chapter 5

Shape in a box

Work in this Chapter was published in the paper ‘Shape in a Box’ at the 4th Color and Photometry in Computer Vision Workshop (CPCV) - European Conference on Computer Vision (ECCV) 2014

In our work we aim to progress towards the ‘ultimate goal’ - shape recovery from a single image, in an as unconstrained an environment as possible.

The key contribution of our work is to substitute the multiple lights used in ‘Shape from Color’ (SFC - Chapter 2) and remove the related lab-based restrictions on shape measurement. Instead we use the scene illumination environment directly without any explicit requirements on primary light sources. The motivation for this work is to produce a shape recovery method with the potential to be developed into functioning systems such as the security and remote, low-power requirement situations mentioned in Chapter 1 Section 1.2.

We hypothesise that the mutual illumination provided by a colourful capture environment should provide sufficient variation in colour across the surface of an object for shape recovery to be possible. Our work in Chapter 3 demonstrates this to be true for synthetic data.

In this chapter we detail the mutual illumination photometric stereo recovery process for real images. We measure the accuracy of recovered height and normal maps against real ground truths. We find that our system recovers shape to the same level of accuracy as SFC, despite the relaxed conditions. We verify this by implementing the SFC method ourselves and imaging the same set of objects. Whilst the SFC approach has been adapted and advanced by others (e.g. [5, 16, 56, 57]), comparison to the classic method provides a good benchmark for our own work. Images of recovered shapes can be found at the end of the Chapter. We begin this Chapter with the first step in our process, calibrating for the mutual illumination environment.

5.1 Calibration

If we wish to estimate the 3D shape of unknown objects we must first somehow acquire enough information about the lighting present in our capture environment. In Chapter 2 we described the SFC method. Let us briefly recapitulate. For an RGB pixel value \mathbf{c} we have,

$$\mathbf{c} = F \mathbf{n} \equiv B E \mathbf{n}, \quad (5.1)$$

where \mathbf{n} is the surface normal at that pixel, B is the RGB albedo of the surface and E is the direction of light sources in the scene. If we place an object with known surface normals (e.g. a sphere) in the scene then B and E are not known separately - we skip straight to establishing F which is a 3x3 matrix transform for converting surface normals to pixel values.

If we invert F then we have,

$$\mathbf{n} = G \mathbf{c} \equiv F^{-1} \mathbf{c}, \quad (5.2)$$

where G is a transform for converting pixel values to surface normals. Thus we can subsequently capture an image of an object with unknown surface normals in the same scene and recover the shape. However, in practice there are difficulties with implementing such a system (such as the assumption of Lambertian surface reflectance); particularly when attempting to move towards more natural scene conditions.

In theory calibration is simple. Like Johnson and Adelson [9], one could simply place a known object (e.g. a sphere) comprised of the same material as the object to be recovered in a scene and capture an image. One could then solve for the linear transform which gives the correct surface normals and thus calibrate the system. If both the calibration and recovery objects are of uniform Lambertian reflectance then this should suffice, as difference in the colour albedo of the two can be solved for as shown in Equation 2.23.

However, this approach is not used directly in our work. No surface is perfectly Lambertian and beyond the engineering of an environment which produces colourful mutual illumination we don't want to place any other restrictions on lighting. Thus, unlike other methods (such as SFC), we will not have direct control over primary light directions to - for example - minimise the appearance of specular highlights in the calibration image.

Instead we capture an image of a mirrored sphere to measure the light from all angles. In the next section we show how this data is used to simulate the graphical model of a perfect Lambertian sphere and thus calibrate our system.

5.2 Spherical harmonic rendering

Capturing an image of a mirrored sphere provides an “environment map”. As the name suggests, the image provides a map of the light present in an environment.

Given this image, we in effect know the lighting inside the box (Figure 5.1). Thus we can simulate the appearance of an object inside our environment.

Measuring real illumination in this manner and applying it to synthetic surfaces was introduced in the 1980s [94]. Since then it has been used in films to add realistically illuminated synthetic objects to real scenes (e.g. ‘Flight of the Navigator’ in 1986) and “environment mapping” is also commonly used in computer graphics [95–97] (as it offers a more computationally affordable solution for real-time lighting than ray tracing). Here, we wish to synthesise a perfectly diffuse object (Figure 5.1).

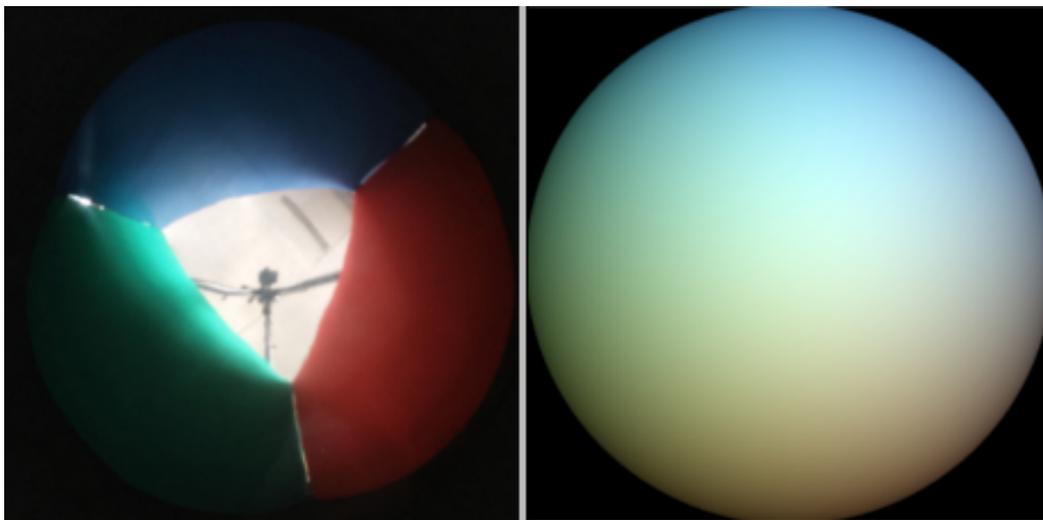


FIGURE 5.1: Left: A captured image of a mirrored sphere inside the box environment. Right: The synthesised appearance of a perfect Lambertian sphere in the box environment.

The mirrored sphere has known shape and if we take it to be perfectly specular¹ then we can calculate light ray directions.

Let us consider an imaged sphere pixel to represent a ray of light, \mathbf{e} which is travelling in some unknown direction towards the sphere before being reflected into the camera. If we take the camera to be distant from the sphere, then the

¹Note - It could of course be questioned, why use a mirrored sphere that isn't perfectly specular to avoid using a calibration object which isn't perfectly diffuse? This way we are at least guaranteed a calibration image which is free of specular effects. Additionally, imaging a mirrored sphere preserves more information about the capture environment. This extra data could be of potential future use.

direction of reflected light \mathbf{r} entering the camera is equal to the viewing direction \mathbf{v} (which we designate as being directly along the z-axis (i.e. $[0, 0, 1]$). The surface normal at each pixel is already known, thus we can solve for \mathbf{e} ,

$$\mathbf{e} = \mathbf{r} - 2 \mathbf{n}(\mathbf{r} \cdot \mathbf{n}). \quad (5.3)$$

If we take the RGB value, \mathbf{c} of calibration pixels as the corresponding strength of each light ray then we have j pairs of light ray directions and strengths (j being equal to the number of imaged sphere pixels). We can render a Lambertian RGB value \mathbf{c}_L for any surface normal,

$$\mathbf{c}_L = \sum_{i=1}^j \mathbf{c}_i (\mathbf{e}_i \cdot \mathbf{n}). \quad (5.4)$$

Applying this “dot product lighting system” to the synthesised surface normals of a perfect sphere provides a diffuse calibration image. However this is an expensive rendering method for such a large number of light rays. Even just taking a sample of points from the mirrored sphere image requires significant computation. Additionally, taking either random or evenly distributed samples risks missing out important values (such as direct light sources) altogether. Thus, we instead apply a different, much faster rendering method to create our synthesised diffuse images.

We use a computer graphics method proposed by Ramamoorthi and Hanrahan [21]. They showed that the reflectance of perfectly diffuse surfaces in a lighting environment can be efficiently stored (with negligible loss of accuracy) using a weighted combination of spherical harmonic basis functions. It is equivalent to, for example, a sound wave being represented as a combination of weighted sine and cosine functions.

Ramamoorthi and Hanrahan determine weightings for 9 basis functions which combine to give a single spherical harmonic representation of the lighting environment. Here we give a practical overview of spherical harmonic lighting (the full mathematical details can be found in [98, 99]).

Let us assume that we have a complete representation of the lighting environment (rather than the j discrete values which we actually have) and denote the environment as L . We can rewrite Equation 5.4 as,

$$C(\mathbf{n}) = \int_{\Omega(\mathbf{n})} L(\mathbf{e})(\mathbf{e} \cdot \mathbf{n}) d\mathbf{n}, \quad (5.5)$$

where $L(\mathbf{e})$ is the strength of light in any direction \mathbf{e} . Thus for any surface normal \mathbf{n} we can calculate an RGB irradiance value C .

Instead of direction vectors we can also refer to points on spheres using horizontal angular coordinates, with θ being the zenith angle and ϕ the azimuth angle i.e. $C(\theta, \phi)$ and $L(\theta, \phi)$. Thus we can replace $C(\mathbf{n})$ with $C(\theta, \phi)$ and $L(\mathbf{n})$ with $L(\theta, \phi)$.

For the purpose of representing irradiance we need only the first three bands of spherical harmonics. Thus we have 9 functions $Y_{l,m}$, with $0 \leq l \leq 2$ and $-l \leq m \leq l$. Any point (θ, ϕ) for each of these functions can be calculated numerically as follows,

$$\begin{aligned} (x, y, z) &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \\ Y_{0,0}(\theta, \phi) &= 0.282095, \\ (Y_{1,1}; Y_{1,0}; Y_{1,-1})(\theta, \phi) &= 0.488603(x; y; z), \\ (Y_{2,1}; Y_{2,-1}; Y_{2,-2})(\theta, \phi) &= 1.092548(xz; yz; xy), \\ (Y_{2,0})(\theta, \phi) &= 0.315392(3z^2 - 1), \\ (Y_{2,2})(\theta, \phi) &= 0.546274(x^2 - y^2). \end{aligned} \quad (5.6)$$

We can now produce the weighted spherical harmonic functions $L_{l,m}$,

$$L_{l,m} = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} L(\theta, \phi) Y_{l,m}(\theta, \phi) \sin\theta \, d\theta d\phi. \quad (5.7)$$

We have now effectively condensed our environment map from the j pixel values on the mirrored sphere to the 9 coefficients of $L_{l,m}$. Though it should be noted that each colour channel must be processed separately thus we actually have 27 values which represent the environment map.

For the purpose of rendering our Lambertian calibration sphere (or any other set of surface normals) we construct a matrix transform H ,

$$H = \begin{bmatrix} c_1 L_{2,2} & c_1 L_{2,-2} & c_1 L_{2,1} & c_2 L_{1,1} \\ c_1 L_{2,-2} & -c_1 L_{2,2} & c_1 L_{2,-1} & c_2 L_{1,-1} \\ c_1 L_{2,1} & c_1 L_{2,-1} & c_3 L_{2,0} & c_2 L_{1,0} \\ c_2 L_{1,1} & c_2 L_{1,-1} & c_2 L_{1,0} & c_4 L_{0,0} - c_5 L_{2,0} \end{bmatrix} \quad (5.8)$$

$$c_1 = 0.429043, \quad c_2 = 0.511664, \quad c_3 = 0.743125,$$

$$c_4 = 0.886227, \quad c_5 = 0.247708.$$

This can be applied to any \mathbf{n} to calculate $C(\mathbf{n})$,

$$C(\mathbf{n}) = \mathbf{n}^t H \mathbf{n}, \quad (5.9)$$

where $\mathbf{n}^t = (x, y, z, 1)$ and H is a 4 x 4 symmetric matrix (again, there is a separate H for each colour channel). It can be helpful to visualise the spherical harmonic basis functions as shown in Figure 5.2:

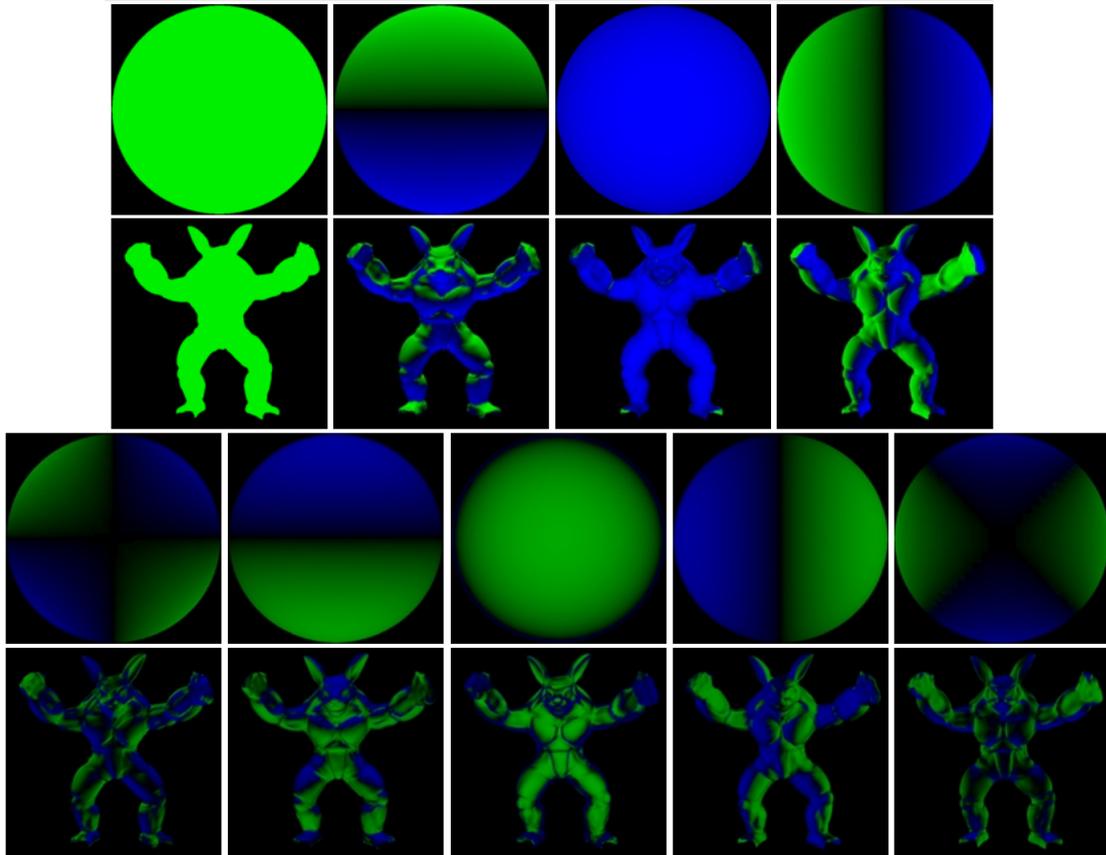


FIGURE 5.2: Visualisation of spherical harmonic basis functions and their effect in rendering (Adapted from [21]).

In the above figure we visualise the 9 spherical harmonic basis functions and below each the effect that it has on rendering of a 3D model. In the figure, green indicates positive values with blue being negative. For the purpose of lighting representation the negative values are clamped to zero (i.e. there is no negative light). Combining the weighted sum of these functions we calculate the matrix transform M .

We can apply H to any Lambertian surface normal and render the corresponding pixel value. For example, the final rendered image of the 3D model in Figure 5.2 would be shaded as in Figure 5.3.



FIGURE 5.3: A 3D model rendered using spherical harmonic lighting (Adapted from [21]).

Essentially the j lights given by the mirror sphere are condensed into a three matrix transforms (one for each colour channel), thus reducing the rendering complexity per normal from $O(j)$ to $O(3)$.

It can be argued that spherical harmonic rendering may not be the best approach to take, as the faster rendering comes at the expense of a small loss in accuracy. It can be shown that the average error will be less than 3% for any physical lighting environment [100]. In the work of Ramamoorthi and Hanrahan [21] they found that the average error was less than 1% for their range of real complex lighting environments.

Whilst accuracy is obviously important we believe that the introduction of such a small amount of error into our system is an acceptable compromise for the sake of greatly improved rendering times. This is especially true if we wish to consider future extension to real-time/video applications or implementation to account for the reduced processing power available in mobile devices.

In summary, we capture an image of a mirrored ball to create an environment map. This measurement of illumination is then processed into weighted spherical harmonic basis functions which produces a lighting matrix. This allows us to rapidly produce a synthetic, perfectly Lambertian calibration sphere from which we can calculate F and subsequently G in Equation 5.2 (the linear transform between pixel values and surface normals).

5.3 Experiments and results

This section details the real-world experiments performed for mutual illumination photometric stereo. We begin by detailing the particulars of image capture. The accuracy of estimated surface normals is reported in section 5.3.3 and compared against the results given by the Shape from Color method. Finally the surface normals are converted to height maps and we present the accuracy for multiple reintegration techniques (Chapter 2).

5.3.1 Image capture

We capture a single image from a fixed viewing position (Note - In early trial runs we captured multiple exposures (Figure 5.4) and blended them into a high dynamic range (HDR) image [101, 102], however it became apparent that this was not necessary). The view is assumed to have orthographic projection (though in reality it is likely to be weakly perspective). From the single image capture, spherical harmonic basis functions [98, 99] are used to simulate the graphical model of a perfectly spherical, perfectly Lambertian reflector with the illumination environment from the mirrored sphere (Section 5.2).

In Figure 5.4, one can see that light from the ambient environment (white in the middle of the sphere) is very bright in comparison to the reflected light inside the

box, risking loss of the mutual illumination effect provided by the walls. It was for this reason that we carried out an investigation into the effect that the dimensions of the box had on the quality (for the purpose of shape recovery) of the lighting environment (Section 3.4).

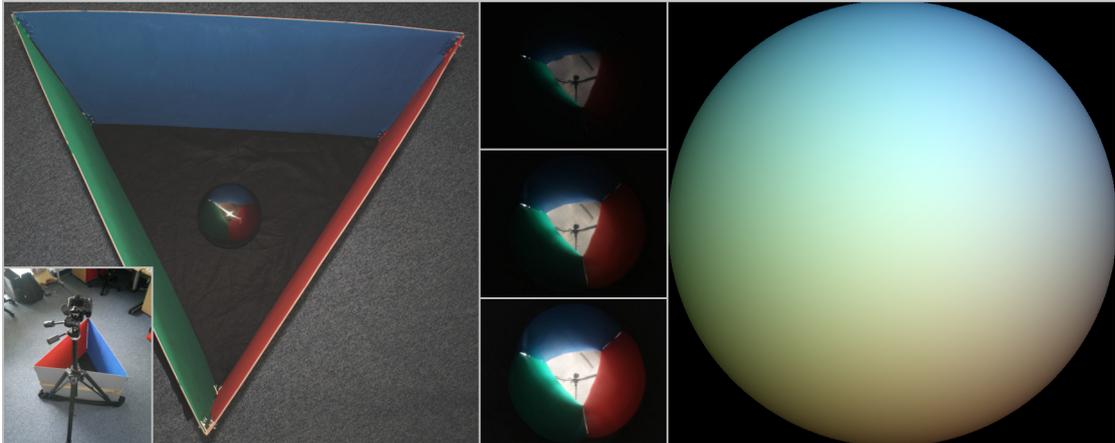


FIGURE 5.4: Left: The experiment set-up. Centre: Captured calibration images of a mirrored sphere at different exposure levels. Right: The synthesised appearance of a perfect Lambertian sphere in the same environment.

All images were captured using a Canon Powershot G11 camera. The raw camera images are used in processing to maintain the linearity of data (the experiment images shown in this thesis have been modified for viewing). Though it should be noted that when attempted we found that processing using rendered images (i.e. the final images produced by the image processing pipeline) actually had very little impact on results. The results given in this chapter are for images captured with the box in an office environment (though test runs conducted in various indoor and outdoor environments produced similar results).

We also present results in this chapter for the same set of objects using the Shape from Color method. That is, we place the objects on a black surface in a dark room. The camera is set up several feet away and behind it we positioned a red, a green and a blue light source (regular lamps using coloured bulbs) directed towards the objects (Figure 5.5). In the original works [48, 87, 88] Drew only presents quantitative results for synthetic objects and qualitative results for real

objects. Our dataset allows us to also present quantitative results for real objects which we use to compare against our own mutual illumination approach. Further details on the Shape from Color method are given in Chapter 2.

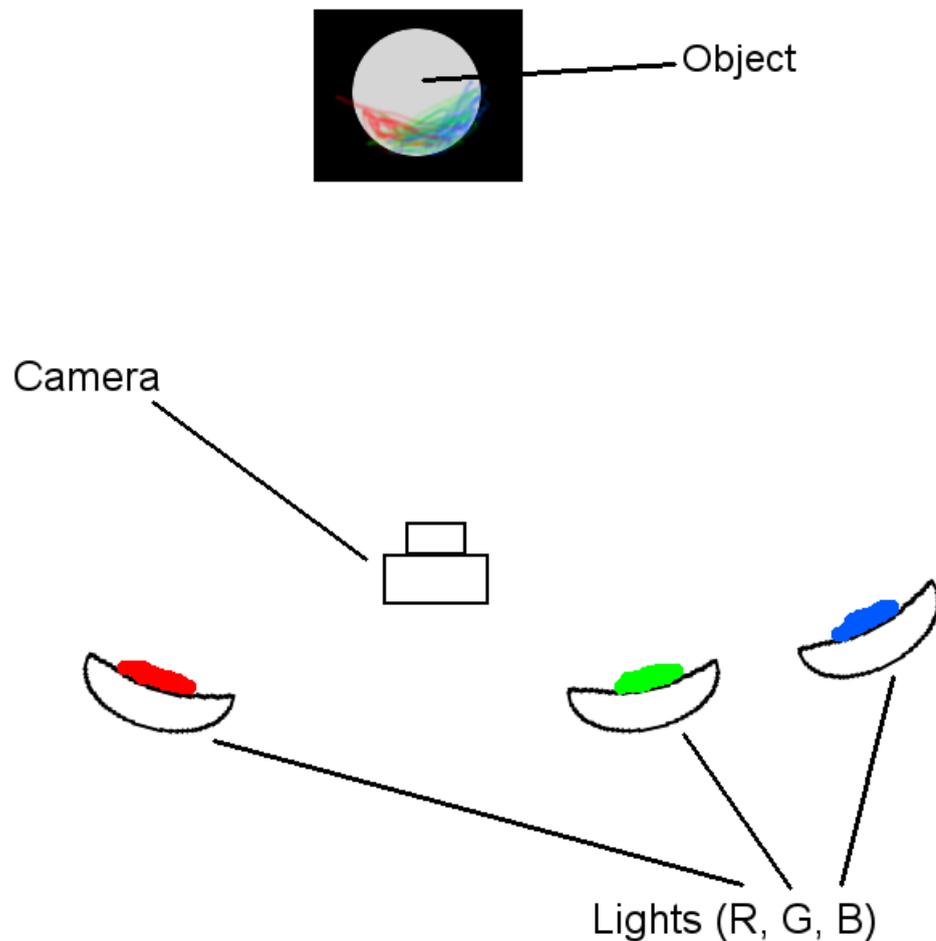


FIGURE 5.5: Illustration of the experiment set-up for Shape From Color.

5.3.2 Comparing to ground truth

To evaluate the accuracy of mutual illumination photometric stereo we capture images of 3D-printed and papercraft objects (Chapter 4). Ground truth for these objects is given respectively by the source files and the templates from which they were created. However there still remains the problem of how exactly to compare

our estimated shape with the true shape, as captured images will not precisely align with the truths.

Papercraft

The process is different for each type of object, first let us detail the comparison for papercraft. After capturing an image the object is manually segmented from its background and we select the vertices by hand. We then generate a perfect height map from these points.

For example, given the Tetrahedron (*Tetra*) in Figure 5.6 we select the three base vertices and the central vertex. The selected points have us $[x, y]$ image coordinates (which are normalised). If we assume that the base vertices exist at height 0 and the central vertex exists at height 1, then for each point we have $[x, y, z]$ coordinates. From this a 3D mesh can be constructed and thus the height at every pixel is known.

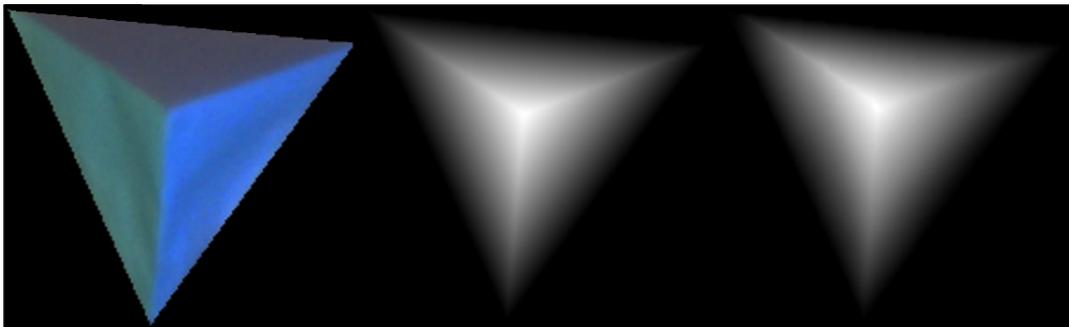


FIGURE 5.6: Left: Segmented captured image. Middle: Generated true height map. Right: Estimated height map.

Calculating the ground truth in this way effectively removes any distortion applied by camera position or orientation from the comparison process (we generate the true height map as it actually appears in the image). As shown in chapter 2 we can then calculate the ground truth surface normals from the height map. However, manually selecting the papercraft vertices in this way has the potential for introducing error to our experiments.

Depending on the points selected the ground truth (and thus the comparison to reconstructed shape) can vary slightly. To validate our approach we have performed the process twenty times for each object. We then took the standard deviation of the twenty height maps produced. The results are shown in Table 5.1:

Object	Standard Deviation of Height
Cone	0.0093
Pyr4	0.0069
Pyr5	0.0073
Pyr6	0.0077
Star	0.0085
Tetra	0.0062
Average	0.0076

TABLE 5.1: Standard deviation of ground truth papercraft height maps.

The generated papercraft ground truth height maps are scaled to unit height (i.e. the highest point on each shape has a value of 1). The average standard deviation of generated height for all papercraft shapes in Table 5.1 is 0.0076, so the manual selection of vertices is very reliable for these simple shapes. Though it is worth noting that the standard deviation increases in line with the number of vertices which must be selected. For example, the *tetrahedron* has three base vertices and a central vertex at its peak; whereas the *six-sided pyramid* possesses a total of seven vertices. More complex papercraft objects may therefore render this method ineffective.

3D Prints

Comparison between estimated and ground truth shape for 3D-printed objects is more problematic. As stated previously, captured images do not precisely align with their corresponding truths. The papercraft objects are all very simple geometric forms for which it is easy to generate true values, but the 3D-prints are much more complex. Our approach to this was as follows.

We view the original source model file from a frontally-facing position with orthographic projection. The height map as visible from this angle is read directly from the OpenGL depth buffer (For the unfamiliar - this means looking at each pixel before it is rendered and recording its distance from the camera) and scaled to possess unit height at the highest point. Again, as shown in Chapter 2 the true surface normals can then be derived (Figure 5.8).

To align the 3D printed ground truths with captured images, corresponding points are manually selected in each. The affine transform which best maps the true shape to an image is calculated, resulting in the two being approximately aligned (see Figure 5.7). Again, the involvement of a manual selection step potentially introduces some uncertainty into our error measures. As with the papercraft objects we performed this procedure twenty times for each object and in Table 5.2 we present the standard deviation of the height maps for each 3D-printed object:

Object	Standard Deviation of Height
Bust	0.0282
Face	0.0223
Sph	0.0091
Trach	0.0355
Average	0.0238

TABLE 5.2: Standard deviation of ground truth 3D-printed height maps.

The results in Table 5.2 show that there is a much larger variation in the height maps of the complex 3D-printed objects (*bust*, *face* and *trachea*) compared to the papercraft objects; though the simpler *sphere* is in line with the latter. Distinct vertices are more difficult for the user to select and so the variation in results is increased. However, the average standard deviation of 3D-printed ground truths is still fairly low. Thus this method still provides a strong indication of accuracy levels and allows us to determine the viability of exploiting mutual illumination in shape recovery.

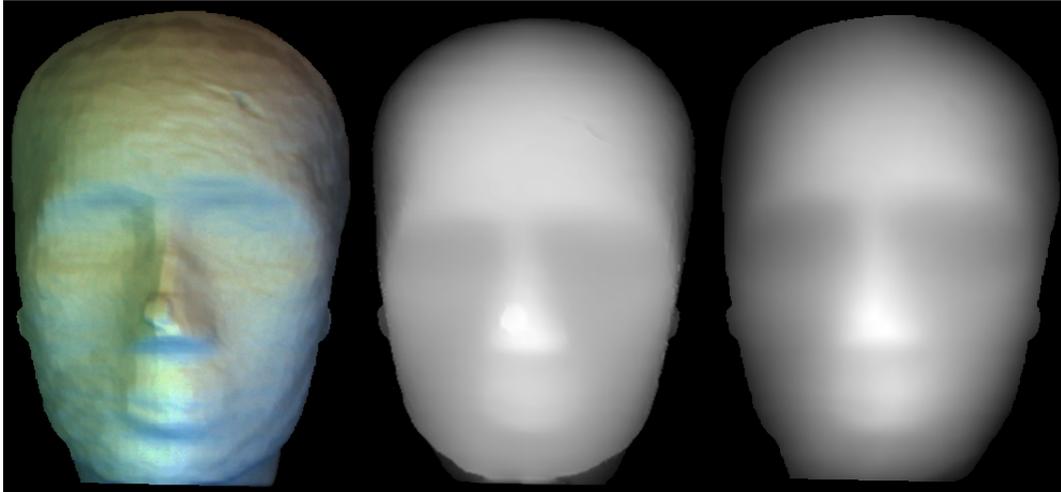


FIGURE 5.7: Left: Segmented captured image. Middle: Transformed true height map. Right: Recovered height map.

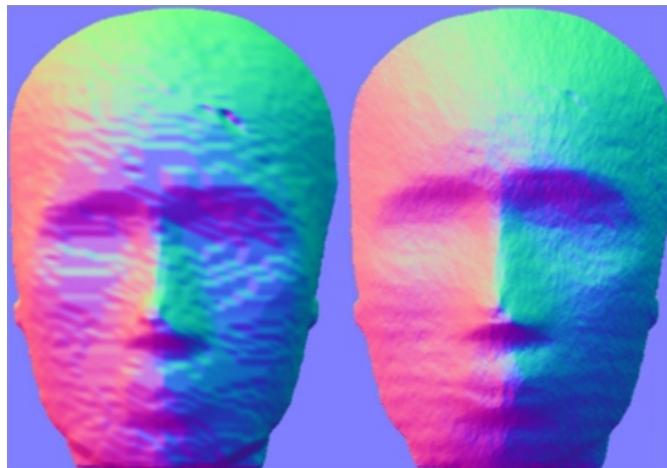


FIGURE 5.8: Left: Transformed true normal map. Right: Recovered normal map.

5.3.3 Surface normal accuracy

We evaluate mutual illumination photometric stereo by capturing images of 3D-printed and papercraft objects. We calculate the average angular error between the estimated shape and the original source files and templates respectively. For comparison, we also conduct experiments using the same objects with the Shape from Color approach. Table 5.3 shows the results from both methods.

Object	Mutual Illumination (°)	Shape from Color (°)
Bust	10.40	10.31
Cone	2.34	2.26
Face	10.41	8.33
Pyr4	3.4	2.07
Pyr5	3.47	3.05
Pyr6	6.85	5.95
Sph	4.89	4.97
Star	17.72	12.03
Tetra	3.45	2.99
Trach	33.97	33.18
Average	9.69	8.51

TABLE 5.3: Accuracy of estimated surface normals represented as average angular error.

Our proposed method recovers shape to almost the same degree of accuracy as SFC (on average there is a 1.17° difference). In Chapter 3 we engineered an environment to produce sufficient mutual illumination for shape recovery. We used the condition number of the matrix transform F (Equation 5.2) to measure the potential for accurate recovery. For the results in Table 5.3 mutual illumination photometric stereo generated $k(F) = 13.85$ whereas the stricter SFC gave $k(F) = 10.93$

This small decrease in accuracy and condition number is not surprising given the more controlled conditions (direct light sources in a dim environment) of SFC. It is surprising, however, that the decrease is so small given the comparative lack of constraints employed in mutual illumination photometric stereo. This demonstrates that using mutual illumination to perform shape recovery is definitely a viable possibility, as well as suggesting that a failure to account for it in existing work (as the overwhelming majority does) is a hindrance.

Predictably, the error for estimated shape of 3D-printed objects is significantly higher than papercraft due to their more complex shape. We are not accounting for any inter-reflection or self-occlusion of light, thus there will be pixel values in the images of 3D-printed objects which are not well defined by the calibrated linear transform. The *Trach* object gives a particularly large error figure as it is a very complex shape with many parts which are occluded from the camera, so it is not possible for us to recover a faithful representation of shape in this case.

One object of particular interest is the *Star* as it returns an unexpectedly high error (more so for mutual illumination). This is likely due to the almost planar structure of the shape, it is very flat in comparison to the other objects and as such does not exhibit as much variation in colour across its surface. This is likely due to a design flaw which is particularly exaggerated in our system. As is common in shape recovery (Chapter 2) we are assuming that all light sources are infinitely distant. Whilst this a reasonable assumption to make for SFC, the enclosed box environment is more likely to produce ambiguous measurements of object surface colour due to interreflection within the box (Figure 6.1).

To further examine the hypothesis that the nature of the box itself has an unwanted influence on results, let us examine the angular error of surface normals with respect to the spatial domain. In Figure 5.9 we show ‘heat maps’ of error for the *tetrahedron*, *sphere*, *four-sided* and *five-sided pyramid* objects.

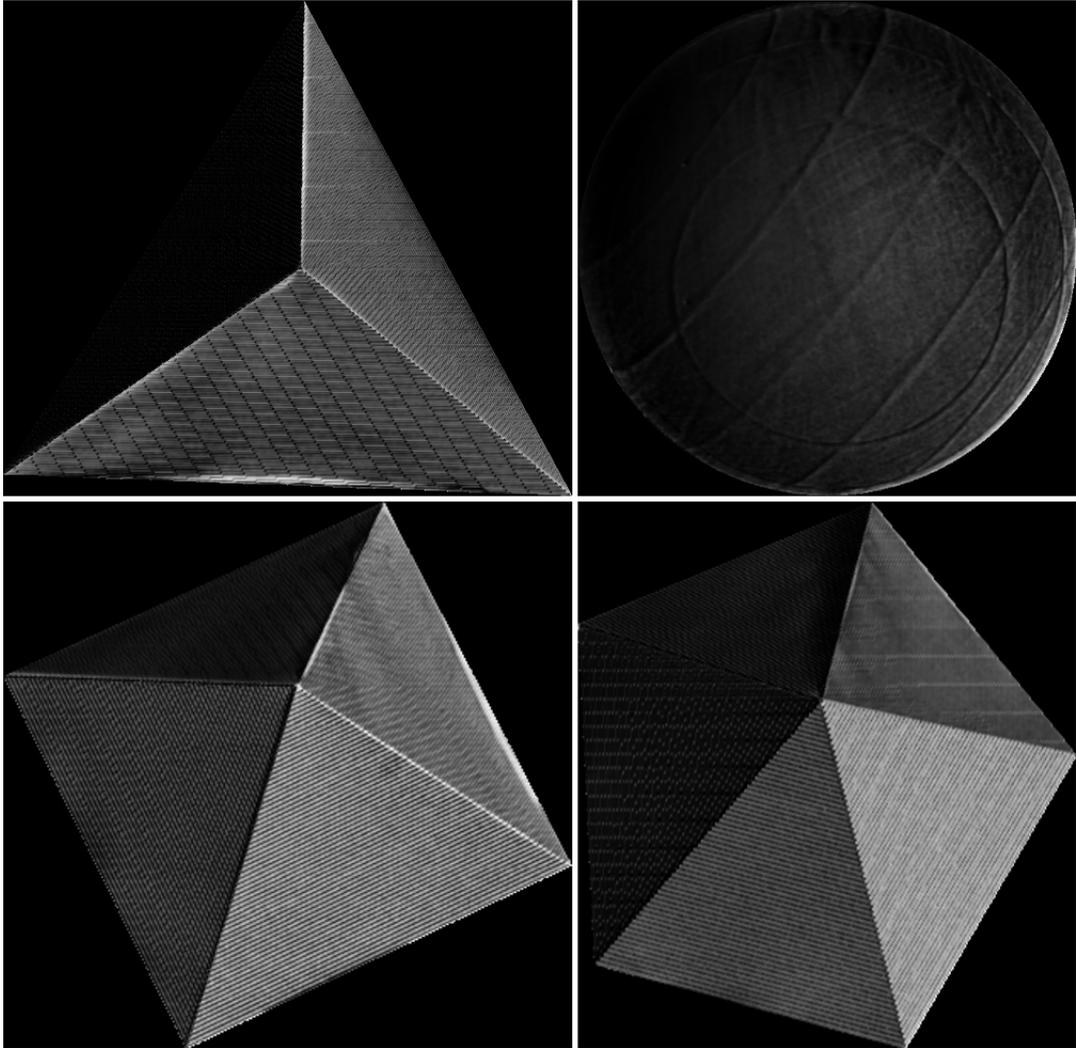


FIGURE 5.9: Visualisation of the angular error for four recovered sets of surface normals. Error is scaled from black (lower) to white (higher).

From the images in Figure 5.9 there seems to be a general trend of there being less error in the top left of each image. To investigate further we scale each image to a common resolution (500px x 500px), split them into a 16x16 grid and average the error present in each square of the grid. We then sum up all four images (Figure 5.10).

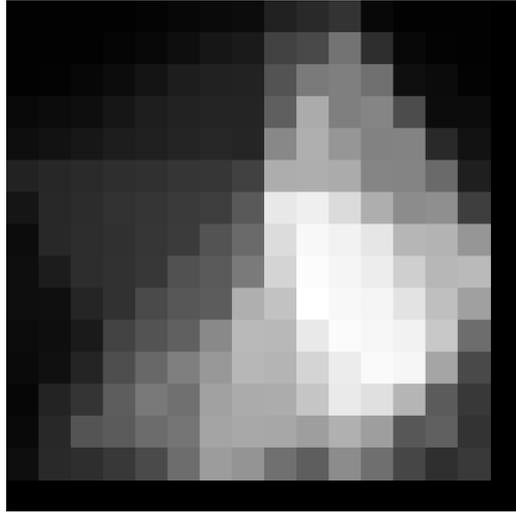


FIGURE 5.10: Image showing the combined, quantised error for the objects in Figure 5.9. Error is scaled from black (lower) to white (higher).

The combined image in Figure 5.10 helps to remove the structure of individual objects and give an overall impression of how error is spatially located within the box environment. We can now see that spatial location within the box environment definitely seems to affect the accuracy of estimated surface normals. We make suggestions for addressing this issue in Chapter 7.

5.3.4 Conversion to height maps

Typically the accuracy of photometric stereo methods is measured using surface normals, which were presented in the previous section. However, normal maps are not a particularly intuitive visualisation of shape and rarely are they the required end product for an application of the technology. Thus surface normals are typically converted into height maps - this gives a much clearer visualisation of results (see Section 5.5). Let us briefly recapitulate from Chapter 2.

A normal map is a vector field, $N(x, y)$. For any (x, y) coordinate in the vector field we have a surface normal \mathbf{n} . Surface normals describe three-dimensional direction, thus each point in the vector field has three components which correspond to the x, y and z axes, that is to say $\mathbf{n} = [n_x, n_y, n_z]$. This vector field can be converted

into a gradient field which corresponds to some surface Z by taking the ratios of the x and y components of the surface normals with their z component [103],

$$\begin{aligned}\frac{\partial Z}{\partial x} &= -\frac{n_x}{n_z}, \\ \frac{\partial Z}{\partial y} &= -\frac{n_y}{n_z}.\end{aligned}\tag{5.10}$$

It is then necessary to integrate over this gradient field to calculate the height at each position. However, for both mutual illumination photometric stereo and Shape from Color, it is almost certain that the underlying assumptions (e.g. perfectly Lambertian reflectance) do not hold. Accordingly the gradients calculated in (5.10) are rarely integrable, meaning that there is no valid height map $Z(x, y)$ which corresponds exactly to the recovered gradients. Thus we seek the integrable surface function $\hat{Z}(x, y)$ which is the closest approximation to $Z(x, y)$, in the sense that the derivatives of $\hat{Z}(x, y)$ are as close to those in equation (5.10) as possible. We can find \hat{Z} by solving Poisson's equation [26],

$$\min(Z) \|\nabla^2 \hat{Z} - \frac{\partial Z}{\partial x} - \frac{\partial Z}{\partial y}\|^2.\tag{5.11}$$

There have been many methods developed to solve this equation and reconstruct height from gradient fields. In the following section we present the results given by three different approaches which we briefly describe here.

Frankot & Chellappa [25] force integrability of the gradient field by projecting it onto Fourier basis functions. Applying the inverse Fourier transform then provides the valid, integrable surface \hat{Z} . Various refinements to this method have been proposed over time. For example, Wei and Klette [104] included additional constraints on surface curvature and area alongside the integrability constraint before minimising the resulting error in the Fourier domain. Similarly, Kovessi's

shapelet correlation approach [24] uses shapelets as basis functions. Results for the Frankot & Chellappa and Kovesi methods are given in the next section.

Also presented are the results of a direct iterative Jacobi integrator. Specifically, since we assumed that an object can be segmented from the background, we knew the occluding contour of the shape. Thus, in effect our reintegration problem has Dirichlet boundary conditions of complex shape. The method iteratively refines height values according to gradient field divergence until the error between successive iterations minimises. This direct method is similar to those presented in [52, 105] and should avoid the over-smoothing typical of the basis function approaches (however it will not be as robust in the presence of noise).

Further detail on reintegration methods is not given here but it is remarked that gradient field reconstruction is an active area of research and we direct the reader to [106] for a more thorough review of the topic.

5.3.5 Height map accuracy

Here we present the percentage accuracy of height maps obtained via three different gradient-to-height methods for both Mutual Illumination Photometric Stereo and Shape from Color. Percentage accuracy was calculated as shown in equation 5.12, where j is the total number of points in a height map. Z is a true height map and \hat{Z} is a recovered height map,

$$accuracy(\hat{Z}) = 100 - 100\sqrt{\frac{\sum_{m=1}^j (Z_{(m)} - \hat{Z}_{(m)})^2}{j}} \quad (5.12)$$

Object	SiaB Accuracy (%)			SFC Accuracy (%)		
	Shapelet	FC	Jacobi	Shapelet	FC	Jacobi
Bust	89.14	88.13	92.94	87.59	89.04	82.25
Cone	84.01	86.00	95.90	92.71	92.01	97.36
Face	92.05	89.79	85.75	88.09	88.17	80.95
Pyr4	81.47	90.85	97.18	94.45	93.19	97.44
Pyr5	85.19	88.72	97.39	93.92	92.80	97.50
Pyr6	88.41	85.25	97.09	94.72	91.06	95.78
Sph	87.51	91.06	89.42	88.41	91.14	84.03
Star	85.76	77.56	94.62	94.74	89.94	94.06
Tetra	85.26	84.23	95.38	96.08	95.03	96.54
Trach	79.63	76.71	81.45	76.84	77.29	60.32
Average	85.84	85.83	92.71	90.76	89.97	88.61

TABLE 5.4: Height map accuracies as percentage values. “Shapelet” is Kovese’s method [24], “FC” is Frankot & Chellappa’s algorithm [25] and “Jacobi” is our Jacobi-type method [26].

The average over all three reintegration methods yielded a figure of 88.13% accurate height recovery from SiaB and 89.78% using the SFC experimental set-up. In the majority of cases, the iterative Jacobi reintegration method achieved the most accurate result of the three. Examples of a recovered height map can be seen in Figures 5.6 and 5.7. Examples of textured shape recoveries for all of our objects are shown in Section 5.5.

Mutual Illumination Photometric Stereo also performs admirably in recovering the shape of human faces. The visual quality of the recovery shown in Figure 5.11 is somewhat surprising as no special processing has been performed to account for the non-uniform surface reflectance - for example - the eyebrows and facial hair shown in figure 5.11 possess completely different surface reflectance properties to that of the skin which forms the majority of the image.

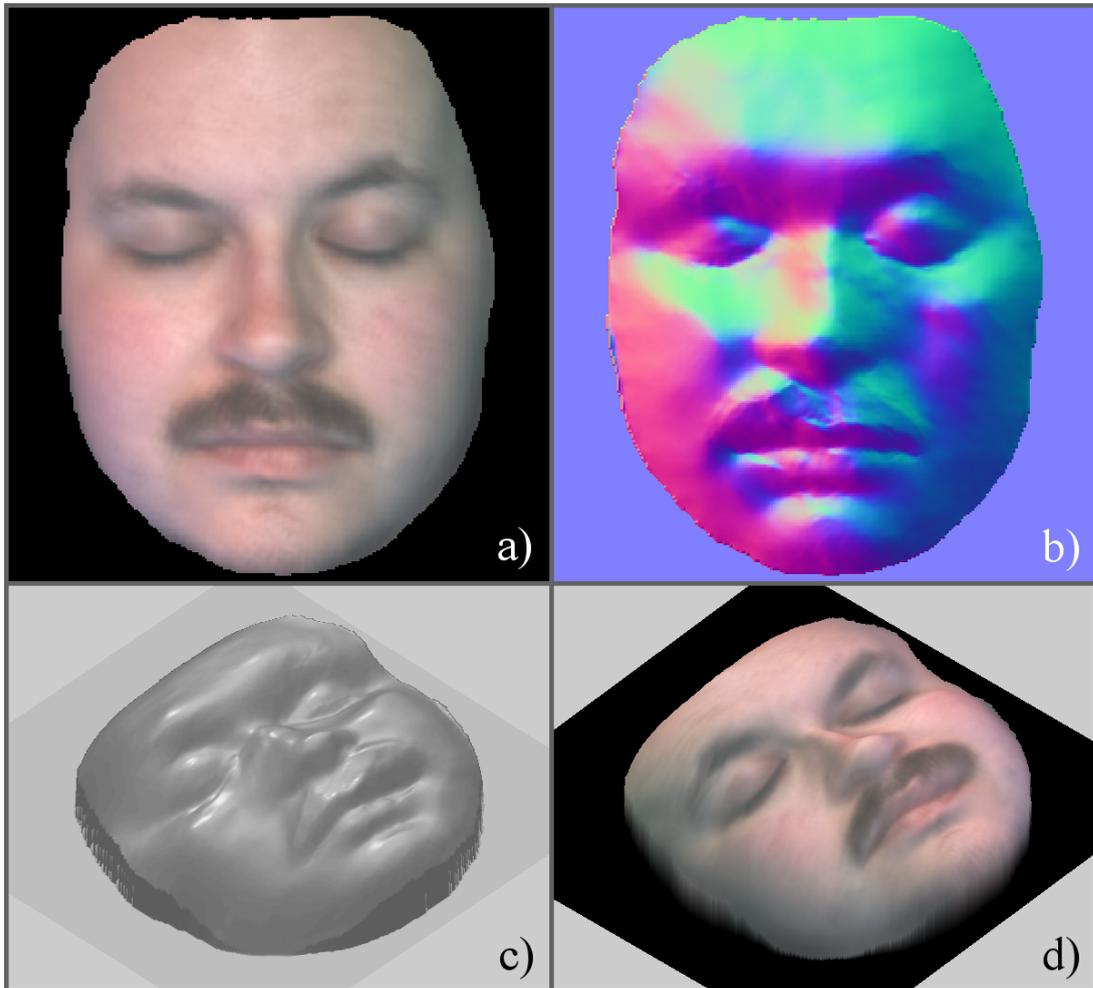


FIGURE 5.11: a) Captured image. b) Recovered normal map. c) Rendering of recovered height map. d) Textured render of recovery.

We posit that the recovery in Figure 5.11 occurs because the least-squares solution (solving the Poisson equation for height map reintegration) forces a smooth result. Nonetheless, it is an encouraging result.

5.4 Conclusion

Mutual illumination photometric stereo successfully recovers 3D shape to almost the same degree of accuracy as the more restrictive Shape from Color method. Despite the absence of a laboratory environment with specified direct light sources

there is very little difference in the average accuracy given by our novel sets of ground truth objects.

Whilst 9.69° of angular error for surface normals is certainly lower than the reported accuracy of many shape recovery methods (e.g. [1]) it remains an impressive result given that we require only a single image and a colourful box.

However, there are several weaknesses in this system. In Chapter 6 we look at two of these - namely the removal of a calibration step and the presence of specular highlights. Other problems which we are yet to address include:

- failure to account for spatial variation of light inside the box - we assume that light sources are infinitely distant - clearly not the case with the walls of the box,
- self-cast mutual illumination and shadows for complex objects - i.e. the recess problem discussed in Chapters 1 and 2,
- absorption of light - light which passes through the surface of the object,
- subsurface scattering (light which passes through the surface and bounces around before again exiting the surface),
- the surprisingly large variance in accuracy produced by different methods of reintegration.
- saturated surface colours cannot be recovered. For example, a yellow surface will not reflect any blue light, thus not enough information is available to solve for surface normals.

Overall though, mutual illumination photometric stereo demonstrates promise and we feel that it warrants further investigation. Tentative results for further works are presented in the next chapter.

5.5 Recovered shapes (images)

Each of these figures shows a textured 3D shape recovered from the corresponding inset image. At the end there are also two examples of shape recovered with the box in outdoor environments.



FIGURE 5.12: *Bust* (3D-printed figurehead)

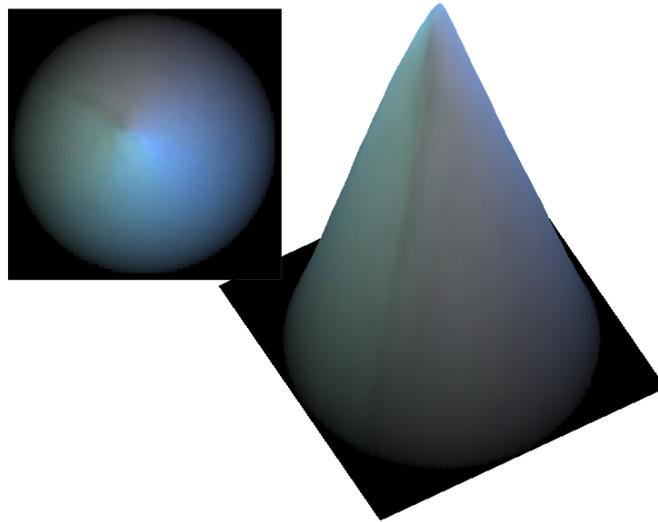
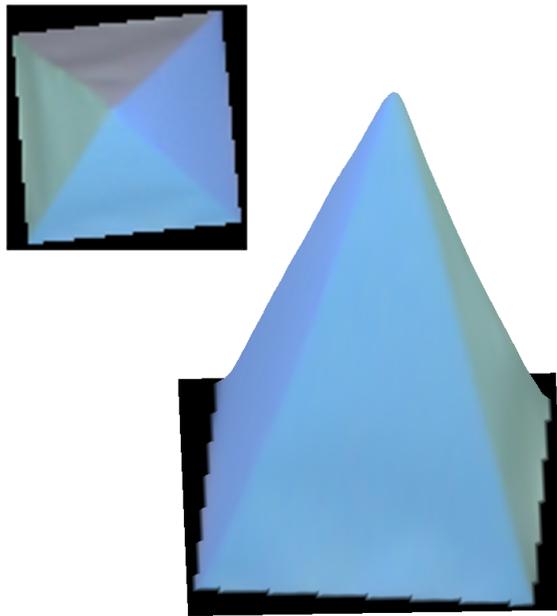
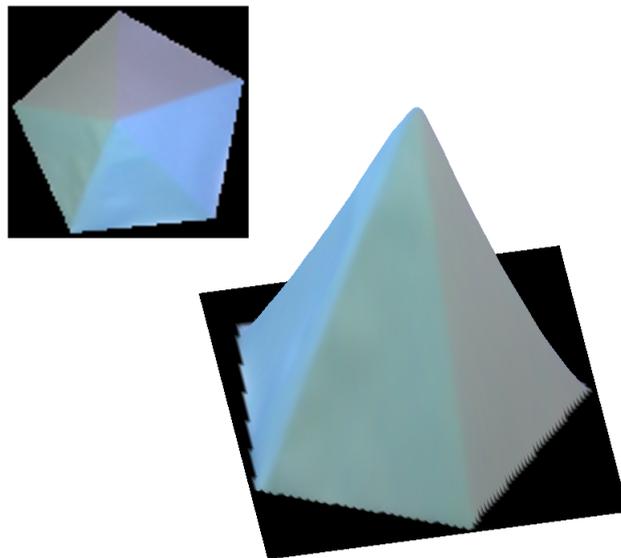
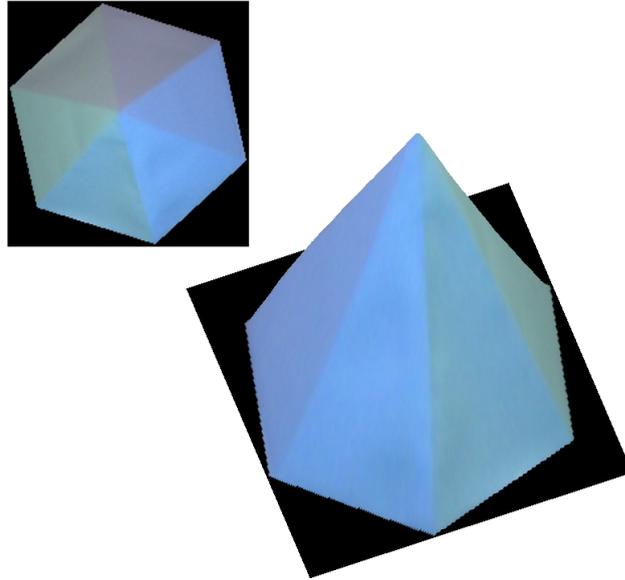
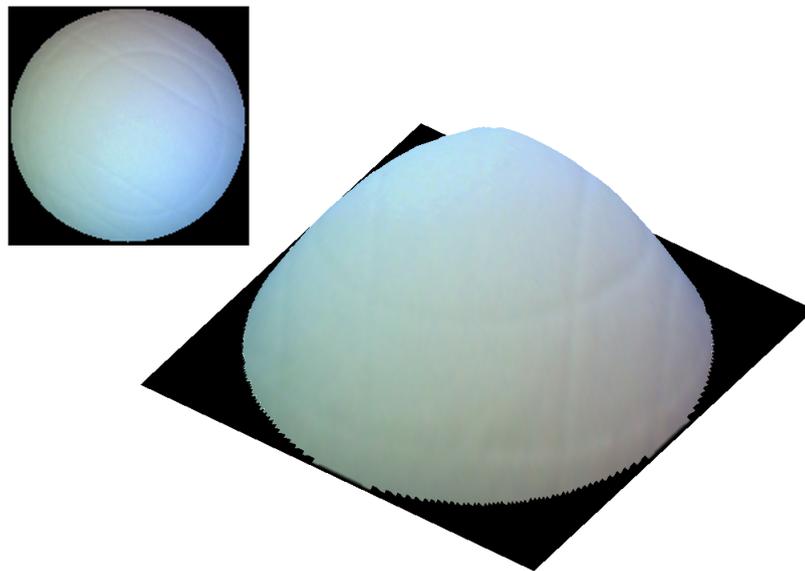


FIGURE 5.13: *Cone* (Papercraft cone)



FIGURE 5.14: *Face* (3D-printed face)

FIGURE 5.15: *Pyr4* (Papercraft 4-sided pyramid)FIGURE 5.16: *Pyr5* (Papercraft 5-sided pyramid)

FIGURE 5.17: *Pyr6* (Papercraft 6-sided pyramid)FIGURE 5.18: *Sph* (3D-printed sphere)

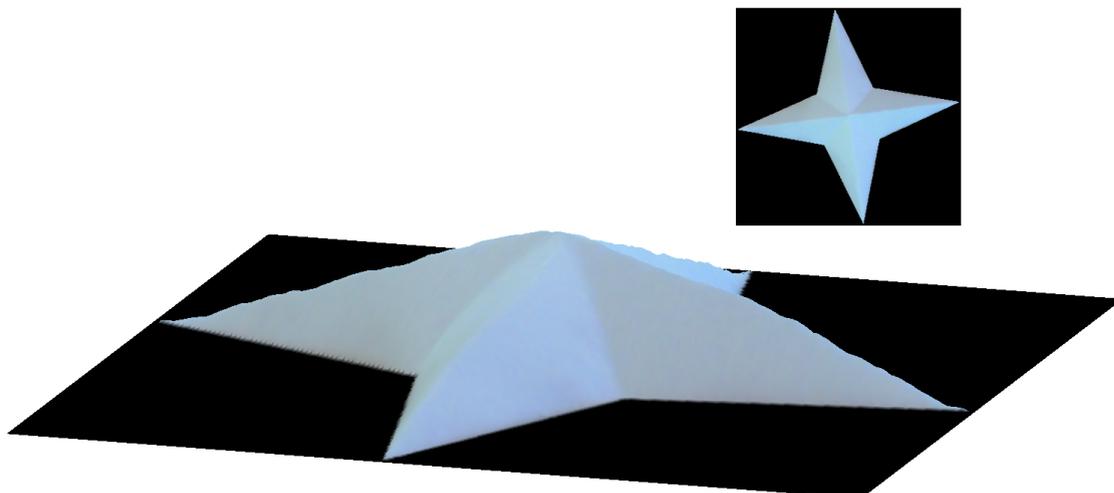


FIGURE 5.19: *Star* (Papercraft star)

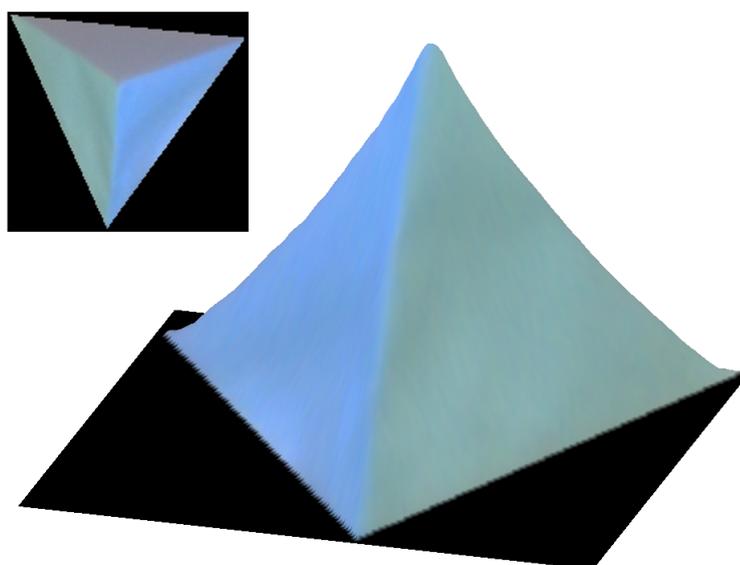
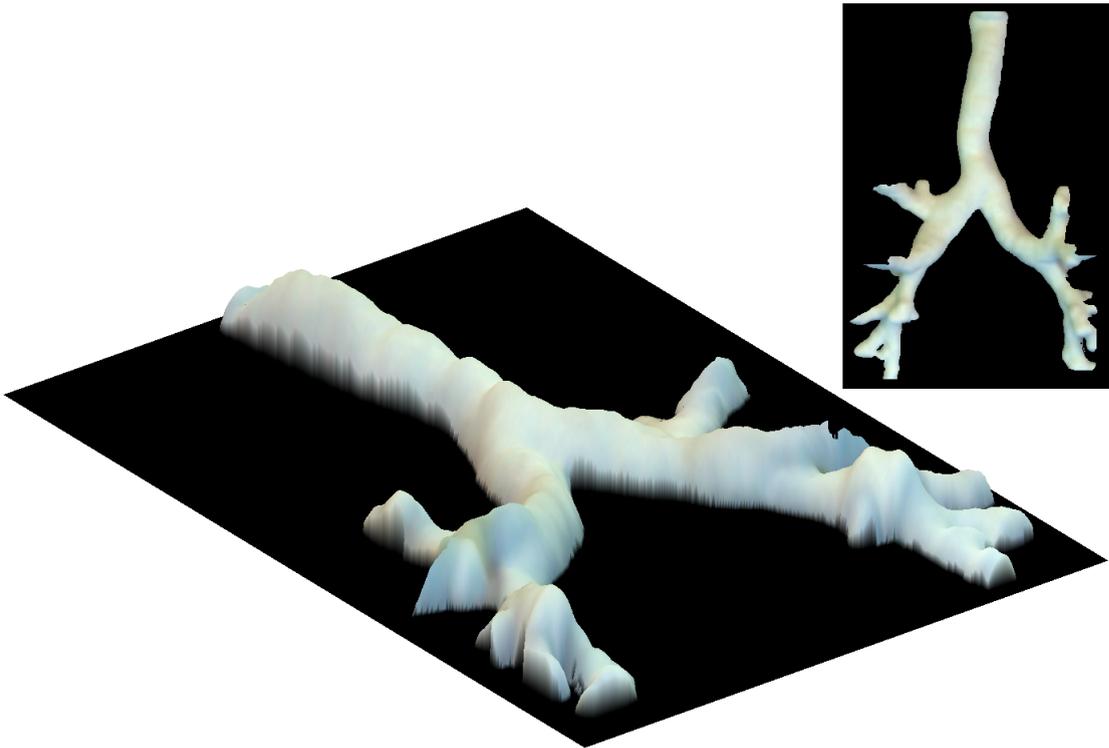


FIGURE 5.20: *Tetra* (Papercraft tetrahedron)

FIGURE 5.21: *Trach* (3D-printed human trachea)

Similar results were also obtained when placing the box in an outdoor environment:

FIGURE 5.22: *Tetra* (Papercraft tetrahedron - captured on a rooftop.) Angular error: 3.31° Height accuracy (using Jacobi integration method): 96.02%.

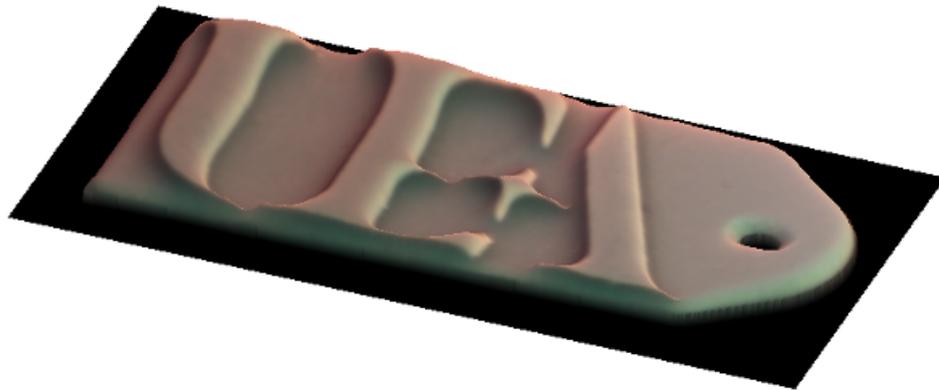
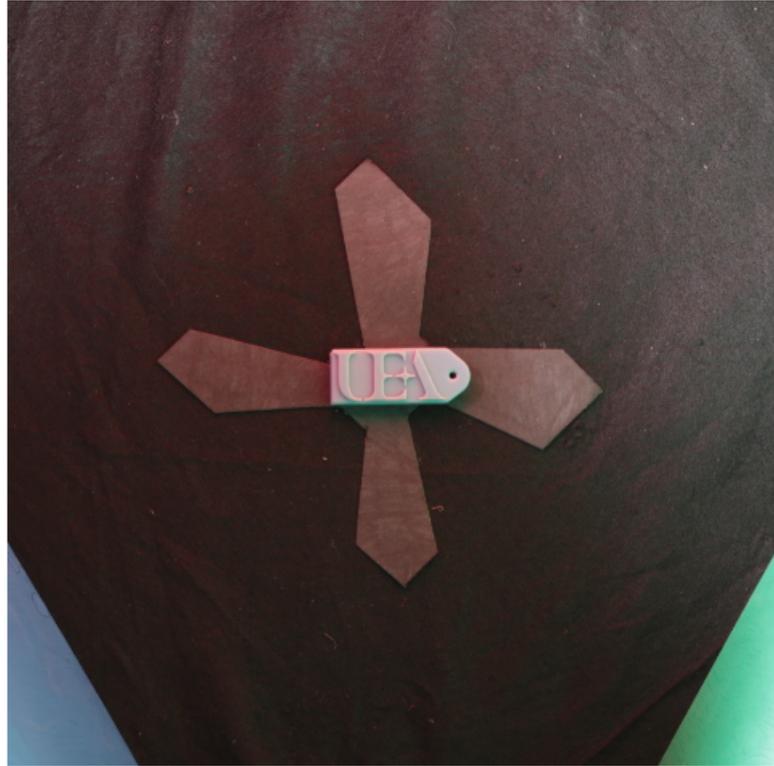


FIGURE 5.23: (3D-printed keyring - captured in the author's garden.) Angular error: 9.26° Height accuracy (using Jacobi integration method): 91.03%.

Chapter 6

Addressing problems

Work in this Section was published in the paper ‘Mutual illumination photometric stereo without calibration’ at the Color and Imaging Conference (CIC) 2015.

In Chapter 5 we established that mutual illumination photometric stereo is a viable approach to shape recovery. Our requirements of only a single image and a colourful environment pose minimal restrictions on capture in comparison with the majority of photometric stereo methods. However our method still suffers from many issues common to shape recovery (Chapter 2). In this chapter we propose solutions for two such problems (though as discussed at the end of the last chapter, there are other issues). Namely that we do not account for specular highlights (Section 6.2) and our requirement of a calibration step (Section 6.1).

6.1 Removal of calibration

By exploiting the mutual illumination present in a colourful environment we can recover the shape of Lambertian objects. Recall that in Chapter 5 we calibrated our system by capturing a representation of lighting within our colourful box

environment using a mirrored sphere. This allowed us to establish a transform from pixel values to surface normals and thus estimate 3D shape.

Although this work moves us closer to our goal of accomplishing single image photometric stereo in general lighting conditions, it is not ideal in design. There are several issues with our calibration step which require addressing.

Firstly, outside of laboratory settings (and to a lesser extent within) it is likely that scene conditions will change in some way before images of objects can be captured. For example, between taking a picture of a calibration object and a second object, the lighting might change (e.g. cloud movement can make a huge difference in lighting conditions).

Secondly, the object used in the calibration might itself fundamentally differ in some aspect of how it reflects light from the object which we wish to measure. This is of particular concern for methods such as our initial approach, which uses a mirrored sphere for calibration. In Chapter 1, we showed that interreflection of light can cause surfaces to appear more brightly illuminated than is expected by the Lambertian model (Figure 1.8). In a similar manner, light is reflected much more strongly by our calibration target than most of our experiment objects. This is illustrated in Figure 6.1. In the Figure the small arrows represent diffuse reflection of light. The Lambertian sphere reflects light equally in all direction. The mirrored sphere reflects light more strongly according to the angle of incidence (Figure 2.41). Thus light entering the mirrored sphere environment will interact with more surfaces before reaching a negligible energy level than it would in the Lambertian sphere environment.

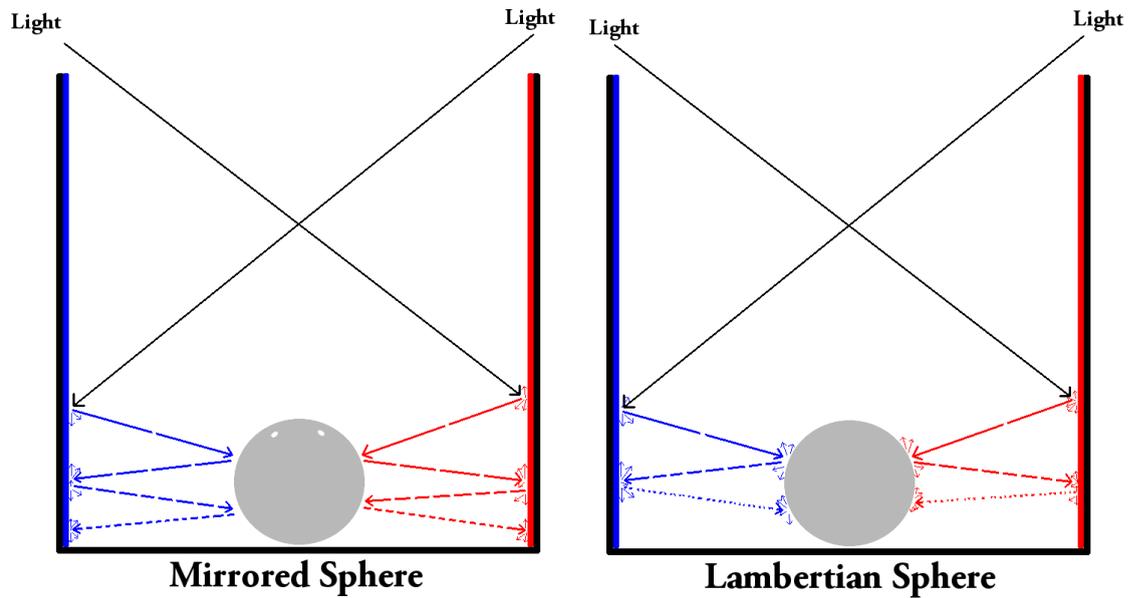


FIGURE 6.1: Light rays enter the box and reflect off the coloured box walls. The fragments represent light ray strength, more fragments = weaker.

The Lambertian sphere reflects light equally in all directions meaning that the reflected rays are weaker and less measurable interreflections with the environment occur. The mirrored sphere is almost perfectly specular so light is reflected strongly in a focused direction. This will cause more measurable interreflections with the environment. To use the terminology of computer graphics, there is a greater number of measurable light bounces in the box containing the mirrored sphere.

Thus the calibration image of a mirrored sphere will not correctly represent the lighting environment experienced by Lambertian objects.

Thirdly, Even if we had identical surface reflectances for both calibration and experiment objects, any deviation in shape would still cause light to be reflected in different directions. Again, this would cause a difference between lighting environments. In most environments such changes are small, however the environment we engineer in Chapter 3 is comparatively small (which is necessary to maximise mutual illumination) and this makes the lighting environment more dependent on the object placed in the scene.

For these reasons we present in this section an initial, naive attempt at removing the need for calibration from our work. The walls of the box are typically visible in object images. We exploit this to calibrate the system for each individual image. Shape recovery proves to still be possible (with a small loss of accuracy) and presents a promising direction for future work.

6.1.1 Exploiting projection

The majority of shape recovery algorithms assume that captured images are taken with orthographic projection. This means that it is assumed that the distance between the camera and objects in the scene does not affect how they appear in captured images. However, in practice real images are almost always subject to a weakly perspective projection (meaning that objects or parts of objects which are further away will appear smaller). Figure 6.2 illustrates the difference between the two projection types.

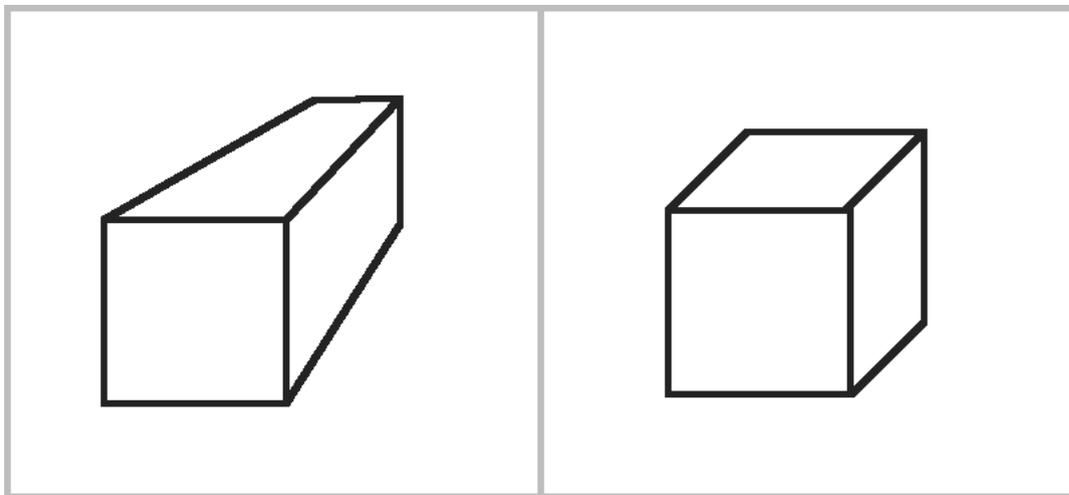


FIGURE 6.2: Two images of the same cube using different projection methods. Left - Perspective projection causes the rear of the cube to appear smaller due to it being further from the camera. Right - In orthographic projection distance does not affect appearance.

There has been some work performed involving perspective projection in photometric stereo. Tankus and Kiryati [107] showed results indicating that switching

to a perspective model could improve the accuracy of photometric stereo. Later Papadimitri and Favaro [108] built on this work and utilised perspective projection to solve the generalised bas-relief (GBR) ambiguity in an uncalibrated system (see Chapter 2 for more on GBR).

Here for the sake of simplicity we naively continue with the assumption that images are captured under orthographic projection whilst exploiting that they in fact are not (see Figure 6.3).

Essentially we simply segment the interior box walls as visible in a captured image and use them to construct light directions and colours. These are then used to calculate a transform between pixel values and surface normals as in our original system, but without the need for calibration.

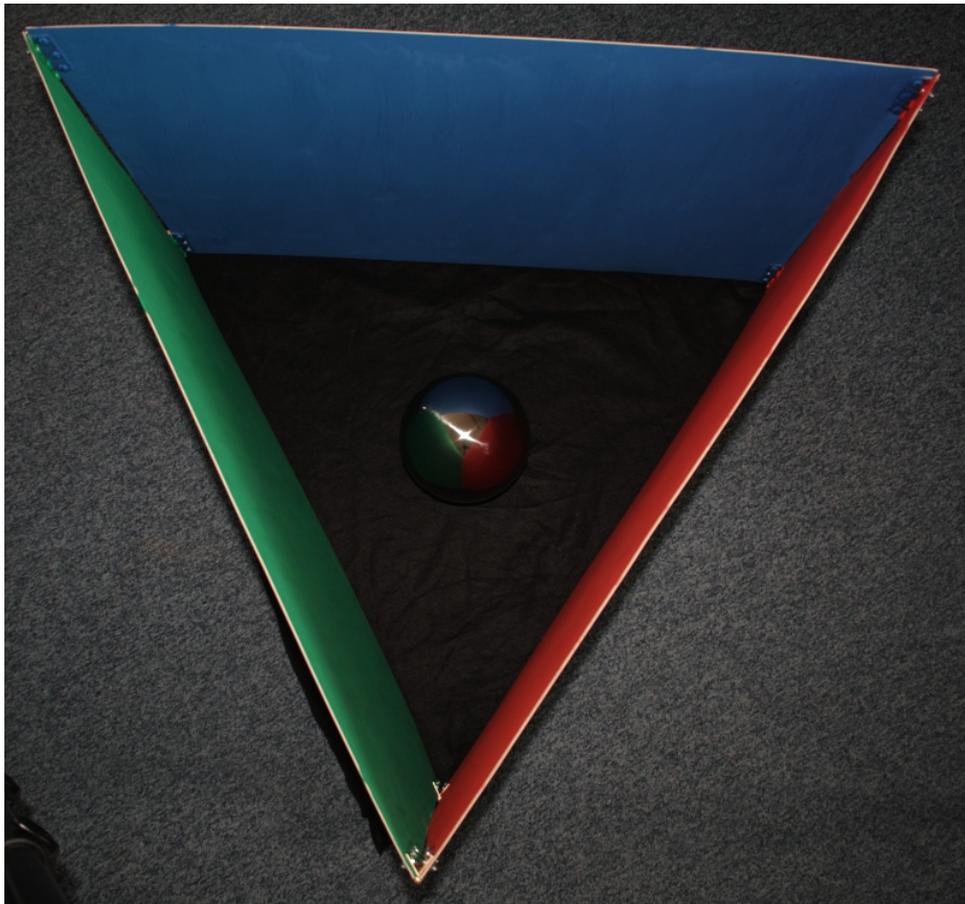


FIGURE 6.3: If the assumption of orthographic projection were true, we would not be able to see the interior walls of the box in this image. The walls are vertical and thus under orthographic projection should be parallel to the vertical orientation of the camera - we should see only the narrow tops of the box walls.

6.1.2 Scene geometry

Before our scene is projected onto a two-dimensional image, it exists in a three-dimensional space. As orthographic projection has been assumed we can represent the three dimensions as being of unit-length. The x and y -axis are given by the width and height of the captured image, with the z -axis consisting of the distance from the bottom of the box ($z = 0$) to the top ($z = 1$). This is illustrated in Figure 6.4.

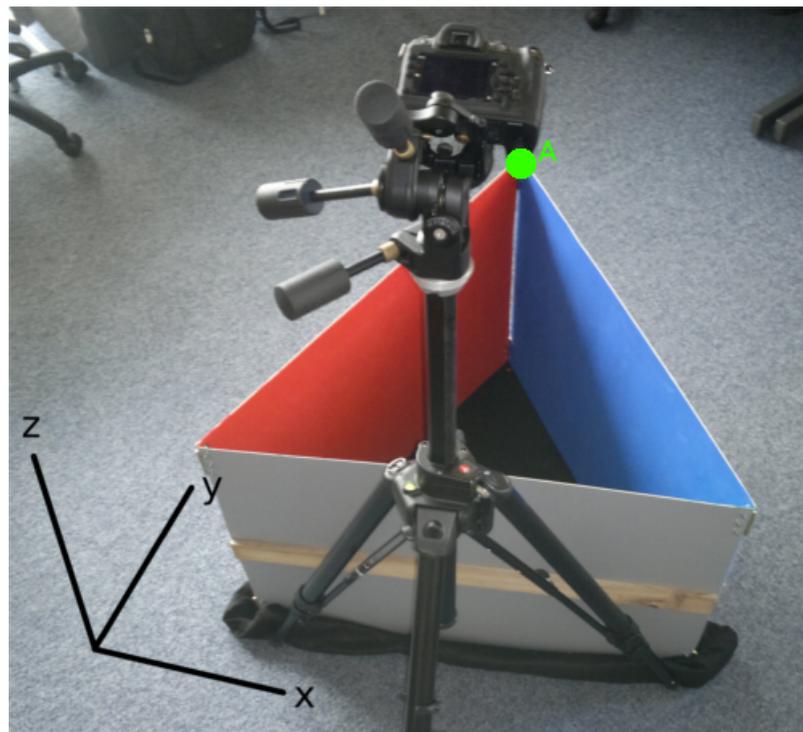


FIGURE 6.4: Representing the 3D space in which the box exists e.g. In a captured image, point A would have $[x; y; z]$ coordinates roughly equal to $[0 : 5; 1 : 0; 1 : 0]$.

As shown in Figure 6.3, the interior walls of the box are visible. If we segment these walls (Figure 6.5) then we have three sets of four points (the corners of each wall) which exist on planes in the three dimensional space of the scene. Given three points $[p_1, p_2, p_3]$ on a plane we can solve a simple cross product to calculate the normal e to each plane,

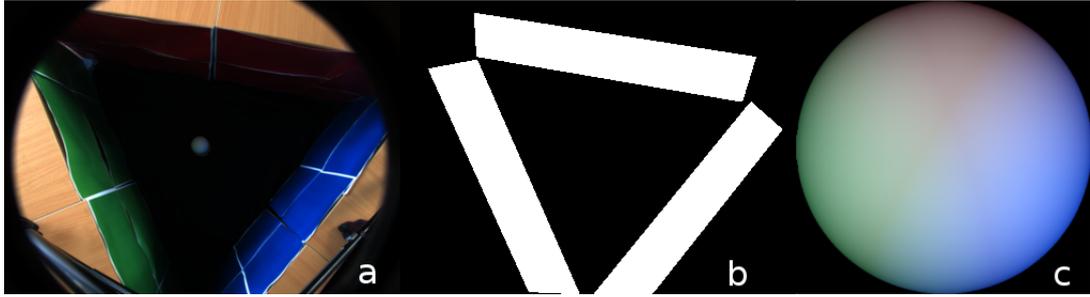


FIGURE 6.5: a) An example of a captured image. b) Segmenting the walls of the box. c) A synthesised Lambertian sphere, rendered using the strength and direction of the light sources estimated from b).

$$\mathbf{e} = (p_2 - p_1) \times (p_3 - p_1). \quad (6.1)$$

These normals are then used to represent the directions of three distant, point light sources. Again this is a naive solution as the box walls act more as close, extended light sources. Though this problem is not addressed here, we suggest that it could be solved by incorporation of a more complex lighting model; such as the radiosity model discussed in Chapter 3. A radiosity model would estimate the transfer of energy between surface patches of different objects. This increased complexity would allow for a representation which is much closer to the reality of the situation.

To calculate the spectral strength of each ‘light source’ \mathbf{s}_i , we simply take the average RGB pixel value for each wall which is easily obtained from the segmentation step. Although segmentation of the walls is performed manually (by selecting four corner points); any reasonable variance in the selection of these points will have a negligible effect on the outcome as averages are being taken over several thousands of pixel values.

In addition we also model the influence of light sources external to the box by including a fourth light source aligned with the z -axis which is white in colour. Thus we have four light source colours and direction vectors.

These light direction vectors and colour values can then be stacked into matrices (see Chapter 2), with the addition of the fourth light,

$$\begin{aligned} S &= \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 & [1, 1, 1] \end{bmatrix}, \\ E^t &= \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & [0, 0, 1] \end{bmatrix}. \end{aligned} \quad (6.2)$$

We can then calculate our matrix transform F which converts surface normals to pixel values (Chapter 5),

$$F = SE. \quad (6.3)$$

We can then apply the inverted transform to the pixel values \mathbf{c} of the object contained in the same image and recover the surface normals \mathbf{n} .

$$\mathbf{n} = F^{-1}\mathbf{c}. \quad (6.4)$$

Although we are making several naive assumptions in this method (which are discussed in the next section), results show that it holds up surprisingly well in practice.

6.1.3 Results

It was shown in Chapter 5 that our mutual illumination photometric stereo method produces results which are roughly equivalent in accuracy to those achieved by the well established Shape from Color (SFC - Chapter 2[48, 87, 88]). Here we produce new results using the same images from Chapter 5 but without the requirement for

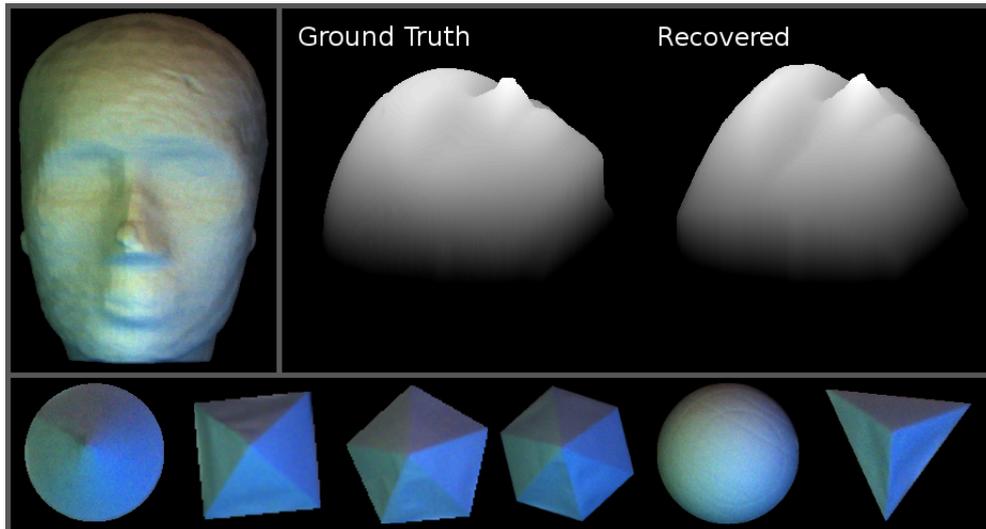


FIGURE 6.6: Captured images for all objects and height maps for the *Face* object.

a calibration step. We compare with the results given by the original calibrated version.

This time we have removed the *Trachea* object from results as its complexity is beyond the capability of a single image system (due to self-occlusion of object geometry). These objects are shown in Figure 6.6 along with a visual comparison between the true and recovered height maps for one of the more complex 3D-printed objects. More details on these objects can be found in Chapter 4.

In Tables 6.1 and 6.2 we show the accuracy of recovered surface normals and height maps obtained from these objects. To obtain these figures we conducted the same comparisons between recoveries and ground truths as detailed in Chapter 5. All images were captured using a Canon Powershot G11.

As shown in Table 6.1, there is an increase in the angular error of surface normals with the removal of the calibration step. This is unsurprising as we are replacing the chrome sphere which in essence measures light sources equal to the number of visible surface pixels (over 100,000 in this case) in the calibration image with just four roughly estimated light sources.

Object	With Calibration	Without Calibration
Bust	10.40	12.02
Cone	2.34	5.16
Face	10.41	11.54
Pyr4	3.4	3.96
Pyr5	3.47	4.24
Pyr6	6.85	2.81
Sphere	4.89	6.95
Tetra	3.45	4.66
Average	4.97	6.42

TABLE 6.1: Table 1: Angular error of recovered surface normals.

Oddly, in one case (*Pyr6*) there is a decrease in angular error with the removal of calibration. This could potentially be caused by the lack of any temporal difference between calibration and object image capture (i.e. one of the reasons why we wish to remove the calibration step).

Object	With Calibration	Without Calibration
Bust	92.94	89.05
Cone	95.9	94.04
Face	85.75	83.19
Pyr4	97.18	96.04
Pyr5	97.39	94.37
Pyr6	97.09	97.9
Sphere	89.42	87.46
Tetra	95.38	95.2
Average	94.02	92.16

TABLE 6.2: Table 2: Height map accuracies as percentage values.

The results given in Table 6.2 were obtained from the estimated surface normals using the Jacobi reintegration method (see Chapter 2). Other than the improvement in percentage accuracy for *Pyr6* compared with its calibrated counterpart there are no oddities here (and given the reduced angular error of surface normals for this shape, improved height accuracy is expected).

6.1.4 Conclusion

This section presents only a tentative initial investigation into how we could remove the need for calibration from our system. We believe that a mere 1.45° increase in average angular error and decrease of 1.86% in height map accuracy merits further work.

There are several potential next steps in this line of research. For example, fully adopting the perspective projection model as in the work of Tankus and Kiryati [107] should provide a reduction in error.

A simple improvement to this work would be the addition of automated segmentation. At present the box walls are segmented through manual selection of corner points and the objects through thresholding of image intensity (as they are on a black background). Given the uniform colour of all surfaces within the box environment (walls, floor and object) robust automated segmentation could be implemented.

There are also several key issues with the modelling of lighting used in our work. Firstly we have naively used each wall of the box to calculate an infinitely distant point light source. However, this is clearly not true and a lot of potentially useful lighting information is being discarded. Additionally, there is no mechanism in place to account for the interreflection of light within the environment (we are effectively assuming a maximum of one bounce before contact with the object). Both of these points could potentially be addressed by a more complex system - e.g. a radiosity based method [109] wherein the environment would be divided into patches and the transfer of light energy between them could be calculated (i.e. the inverse of the rendering method which we used during the engineering of our capture environment in Chapter 3).

There is also another issue with the modelling of lighting in this section. We naively added a single source of light external to the box (the white light source

aligned with the z-axis). For any real environment in which the box is placed, the truth will invariably differ from this. It would be more accurate to allow for multiple external lights, the direction and spectral properties of which could be calculated. This could be performed either by modification of the box (for example, placing strips with known reflectance around the top) or alternatively the gradient of light cast across the walls could be used to provide cues to external conditions [61].

Despite all of these problems it is encouraging that using only a colourful box and a truly singular image capture (truly meaning that there is no separate calibration imaging required) we can recover reasonably accurate models of even quite complex shapes. Indeed, even with less ideal improvised circumstances such as those in Figure 6.7 shape recovery using our mutual illumination method appears to present a promising approach.



FIGURE 6.7: Preliminary results demonstrating the resilience of our method to less ideal environments. A lampshade illuminated by the indirect illumination from three colourful t-shirts.

Note - Figure 6.7 is intended only to provide a rough visual example of the possibility for mutual illumination photometric stereo use in non-engineered environments. We do not have ground truth for the object shown so there are no error measures provided, it was just a quick, improvised experiment.

6.2 Specular highlights

Work in this Section was presented at the Computing Sciences Research Day at the University of East Anglia (UEA) in 2015. Out of a field of 30 papers, this work won the second place prize for best paper.

We make the common assumption that objects in our experiments possess Lambertian reflectance - that is to say that light contacting a surface is reflected equally in all directions. For many objects this assumption holds, however surfaces with significant specular components are problematic (Figure 6.8).

Surfaces with specular properties reflect light more strongly in directions close to the angle of incidence with the causal light source (Figure 2.41). This causes specular highlights to appear which violate the Lambertian assumption and thus cause errors in shape recovery.

Our shape recovery method can proceed without modification if specular highlights are removed from images beforehand. Here we detail two different approaches towards this pre-processing step and provide some qualitative results.

6.2.1 Pre-processing

In Chapter 2 we briefly mentioned the dichromatic reflectance model. Shafer [79] proposed that pixel values consist of the weighted sum of two components - the interface (specular, \mathbf{c}_S) and the body (Lambertian \mathbf{c}_L),

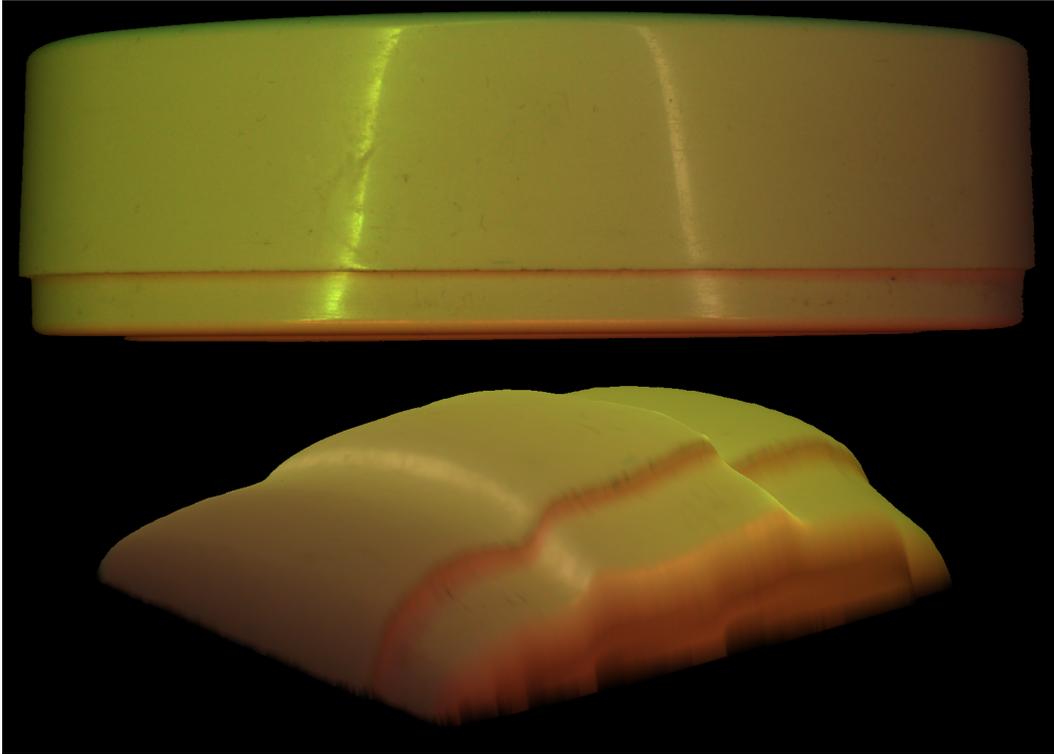


FIGURE 6.8: Top: Captured image. Bottom: Recovered shape.

$$c = m_L c_L + m_S c_S, \quad (6.5)$$

where m_L and m_S are geometric scaling factors which determine the strength of each component. For example, the pixel values in an image of some perfectly Lambertian surface will have $m_S = 0$. This is a key observation which underlies the removal of specular highlights from images - that we can remove the interface and either retain or restore the body.

Methods which remove specular highlights from images fall into two categories - First, those which detect specular pixels and subsequently remove them (i.e. two separate steps, removal of the interface and restoration of the body). Second, those which perform inherent removal without an explicit detection step (i.e. removal of the interface whilst attempting to retain the underlying body). In this section we present results for two methods, one of each type.

Firstly we implement a simple approach to act as a benchmark. Specular pixels are detected via thresholding and removed from the image (set to null). An inpainting [83] algorithm is then applied to create an image which contains only Lambertian pixels (Inpainting refers to using valid image pixels to ‘paint’ over the removed ones. See [84] for more detail).

Our second method involves manipulation of image colour space to remove specular highlights without the need for any explicit detection. However this results in reduced dimensionality of data and thus requires a small modification to our shape recovery method.

6.2.1.1 Detection and removal

To detect specular highlights we simply apply a threshold on pixel intensity values. For each pixel \mathbf{c} we sum the colour channels and test the sum against some threshold value t ,

$$\mathbf{c}_r + \mathbf{c}_g + \mathbf{c}_b > t. \tag{6.6}$$

Pixels which satisfy this criteria are labelled as specular highlights. We remove these and leave blank spaces in the image. The blank pixels are then filled in using an inpainting [83] algorithm. We apply a method developed by Esedoglu and Shen [110]. The results can be seen in Figures 6.9 and 6.11.

We then perform shape recovery as in Chapter 5, using the edited object image as input (Figures 6.10 and 6.12).

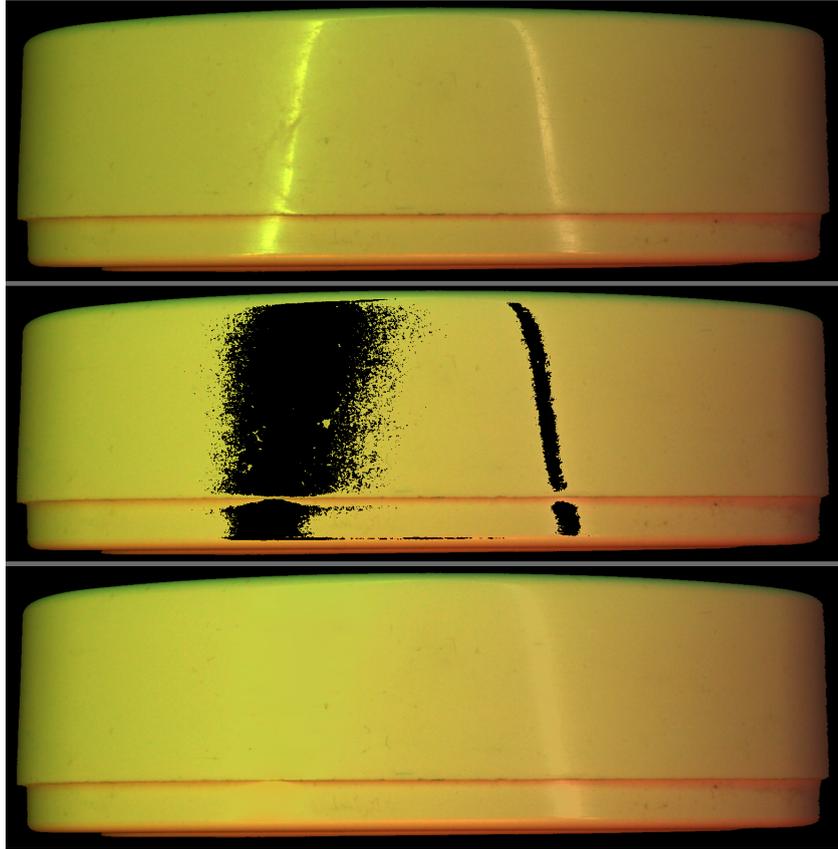


FIGURE 6.9: Top: Original image. Middle: Specular pixels detected and removed. Bottom: Inpainted image.

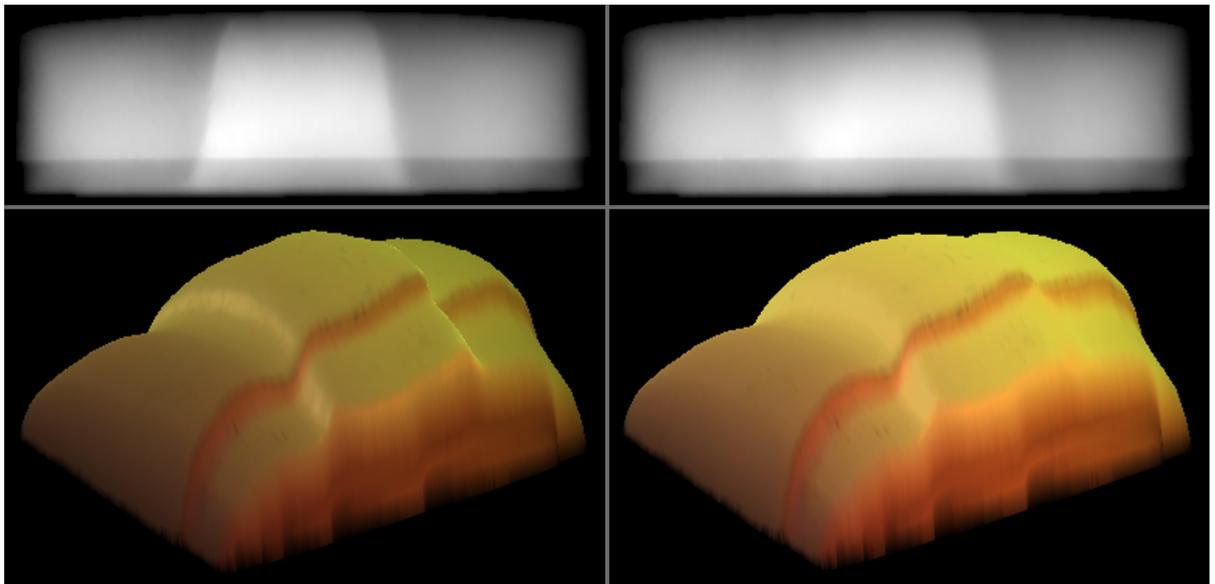


FIGURE 6.10: Top: Height maps. Bottom: Textured meshes. Left: Without pre-processing. Right: With pre-processing.

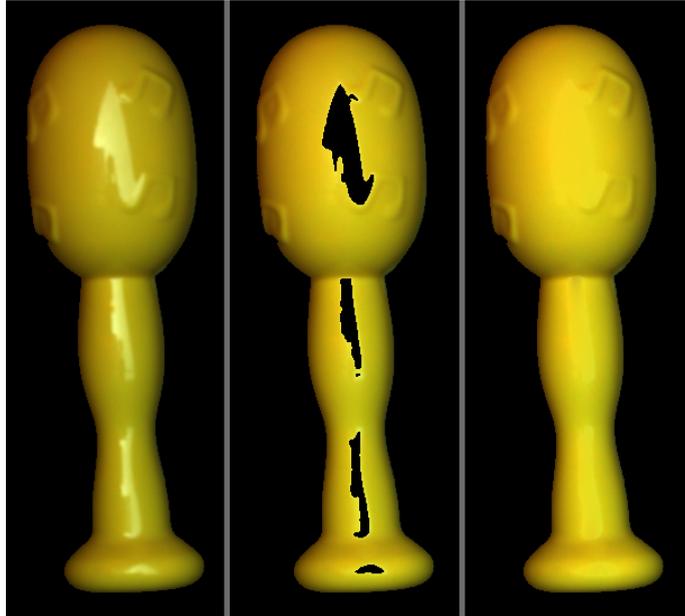


FIGURE 6.11: Top: Original image. Middle: Specular pixels detected and removed. Bottom: Inpainted image.

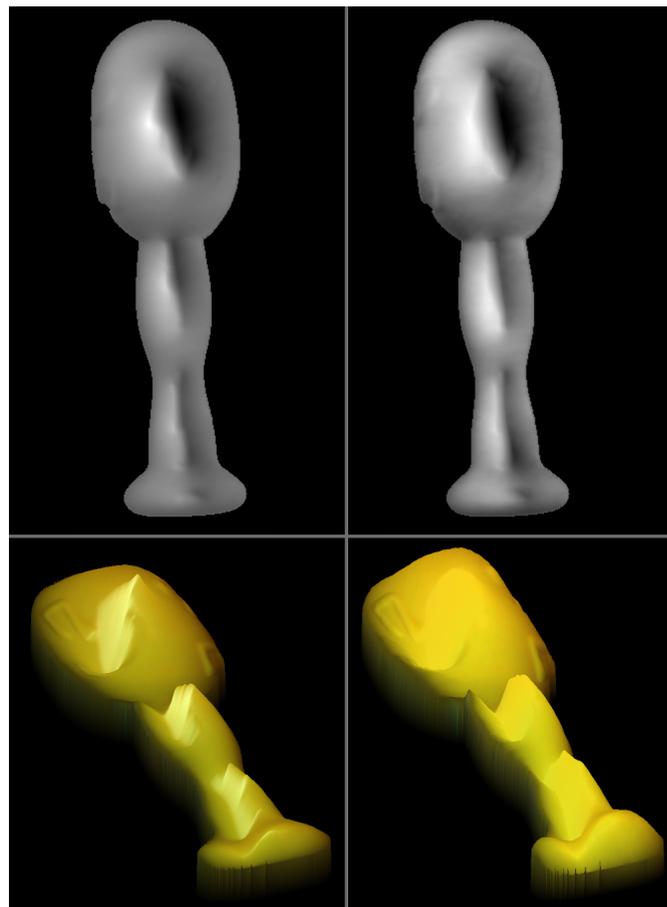


FIGURE 6.12: Top: Height maps. Bottom: Textured meshes. Left: Without pre-processing. Right: With pre-processing.

Although the inpainted pixels no longer appear specular, they still exhibit changes in colour which do not conform to the corresponding surface normals. Additionally, the flat colouring exhibited by inpainted areas does not conform to the smoothly varying colour of the truly Lambertian pixels. In terms of the dichromatic reflectance model - we have not successfully recovered the body component of the pixels. Thus the recovered shapes are still subject to error.

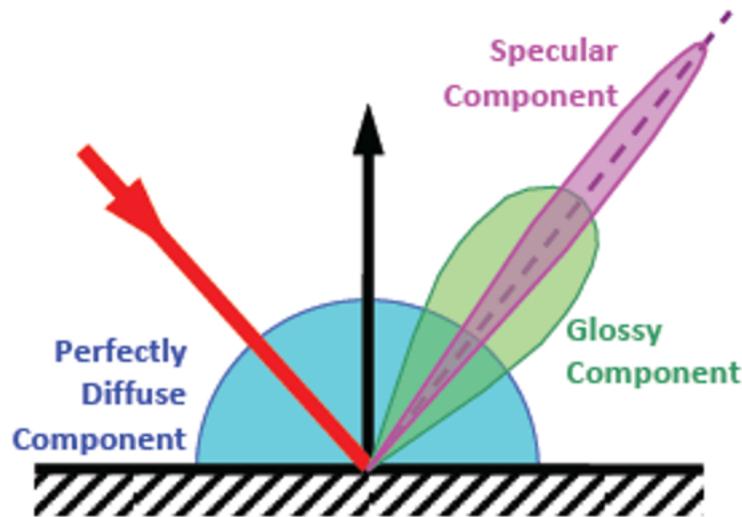


FIGURE 6.13: A specular lobe consists of the specular and glossy components above. The wider the lobe, the wider the range of surface normals which will produce pixels with specular components (Image taken from [Virial International](#)).

Indeed, inpainting in general is better suited to filling small gaps in images. In our case the condition of a surface's "specular lobe" is particularly important. The term specular lobe refers to the range of directions over which a surface will exhibit specularities (Figure 6.13). Surfaces with a narrow specular lobe will tend to produce less specular pixels when imaged. To demonstrate this, Figures 6.14 and 6.15 show the same process for a synthesised image which exhibits a single, relatively small specular highlight.

The relatively wide specular lobes of the surfaces in Figures 6.12 and 6.9 produce large specularities. It is difficult for inpainting to recover cognizant results for

such areas and there is significant potential for object details to be lost (e.g. the musical notes in 6.11).

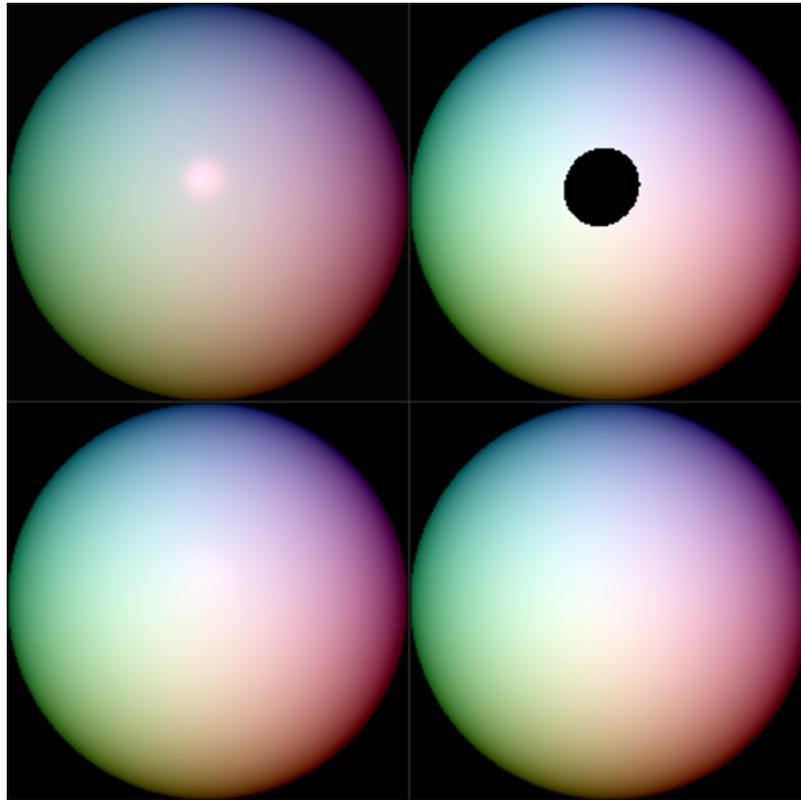


FIGURE 6.14: Top left: Synthesised image with a specular highlight. Top right: Highlight is removed. Bottom left: Inpainted image. Bottom right: Synthesised image without specular highlight.

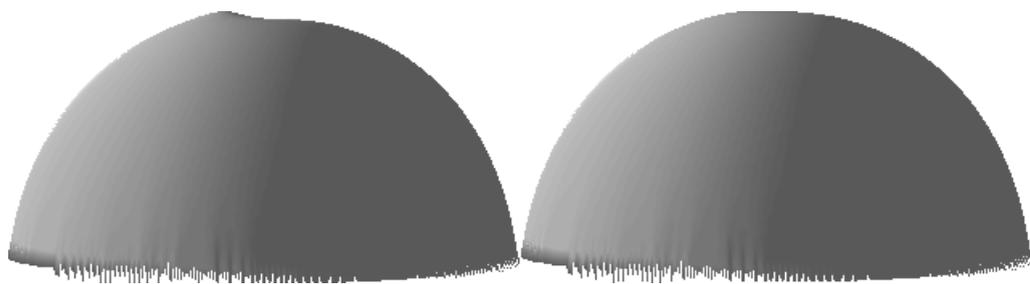


FIGURE 6.15: Recovered shapes. Left: With specular highlight. Right: After inpainting.

6.2.1.2 Inherent removal

In our second approach towards pre-processing images containing specularities, we adapt a method which performs inherent removal of specular highlights. That is,

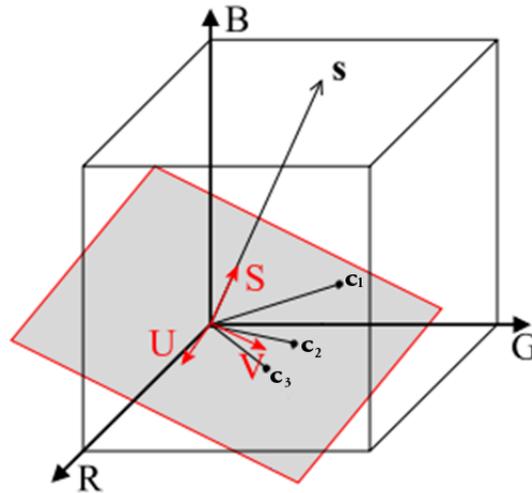


FIGURE 6.16: RGB colour space is rotated such that one of the axes aligns with the light colour \mathbf{s} (Image adapted from [22]).

there is no explicit detection step involved.

Mallick et al. [22] propose the removal of highlights by manipulation of image colour space. They observed that for the majority of surfaces any highlights will have the same spectral properties as the light source causing the effect. If we know the colour of the light source \mathbf{s} , we can find a rotation matrix R that aligns one of the RGB colour axes with \mathbf{s} . This new data-dependent space is referred to as SUV. The pixel values \mathbf{c}_{RGB} in the original image are then also rotated by R (Figure 6.16),

$$\mathbf{c}_{SUV} = R\mathbf{c}_{RGB}. \quad (6.7)$$

The result, is that any specular highlights caused by the light source should exist solely in the new S channel. The UV channels should contain only diffuse shading information. Thus a single rotated RGB image transformed to SUV will leave two useful channels. In terms of the dichromatic model, we are removing the interface whilst retaining the body. Though of course, this is assuming that there is only a single light source which is causing any highlights. In reality this will not

necessarily always be the case for our system (e.g. in Figure 6.9 the highlights are caused by the walls of the box).

To find the colour of the light source \mathbf{s} , we once again use the mirrored sphere for calibration. We take the highest intensity pixel on the mirrored sphere to be the light colour. In Figure 6.17 it can be seen that this method works to some degree, though too much diffuse information is lost (we fail to retain the body component).

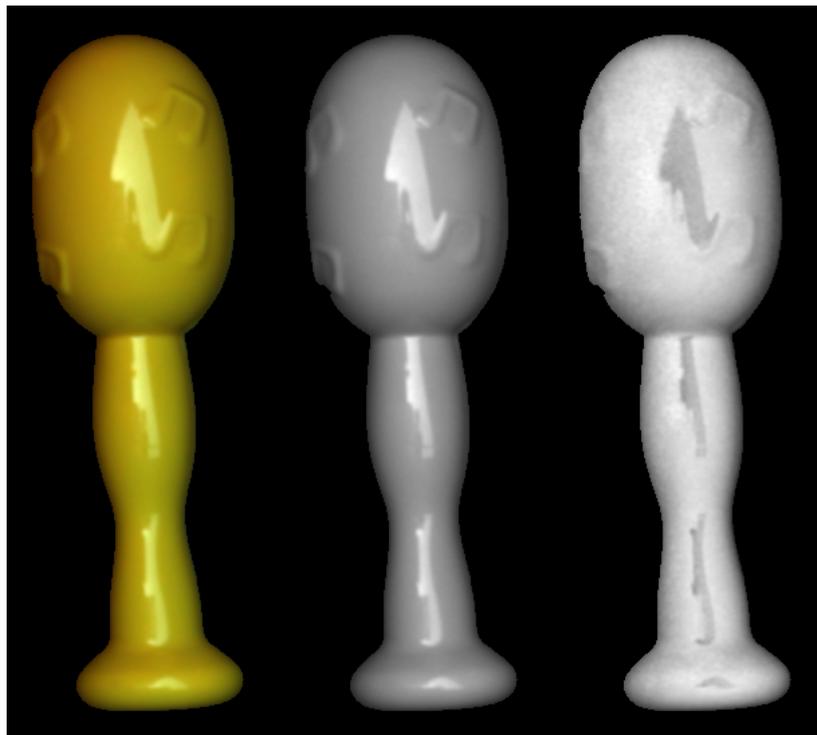


FIGURE 6.17: Left: RGB image. Middle: Specular image (S). Right: Diffuse image ($\|U + V\|$).

The two remaining channels are then processed using a reduced dimensionality version of our shape recovery method from Chapter 5. We calculate a 2x2 matrix transform between surface gradients and UV pixel values, instead of the usual 3x3 transform between surface normals and RGB values. This reduction in data dimensionality inevitably causes poor shape estimation (Figure 6.18).

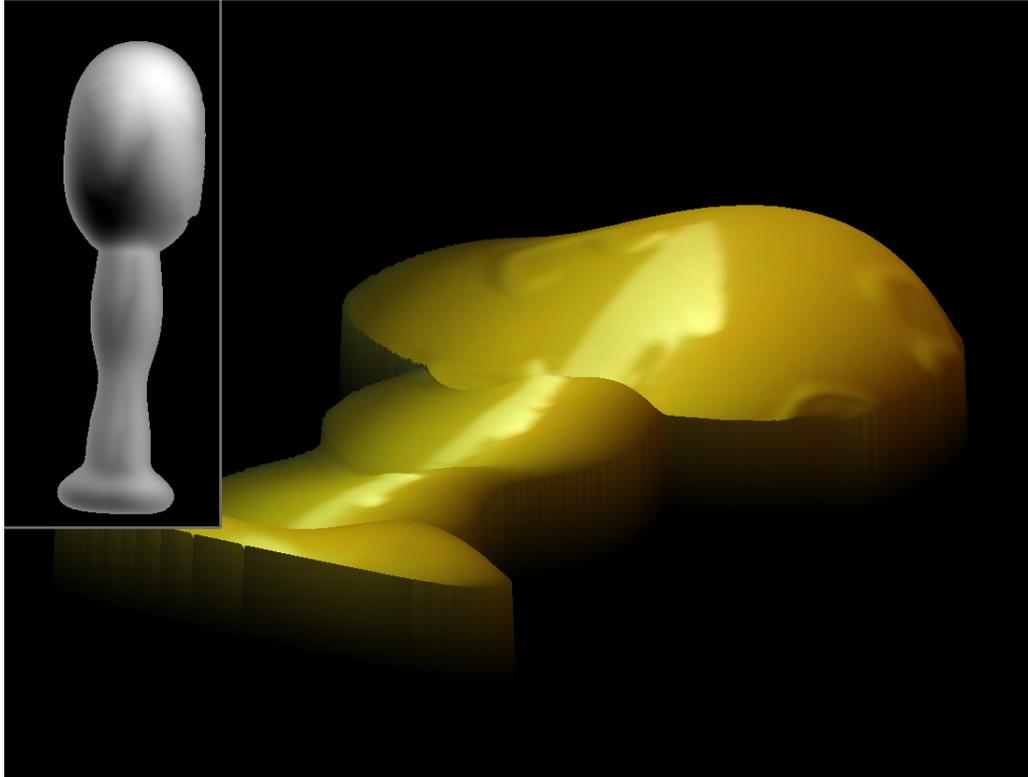


FIGURE 6.18: Shape recovered from the UV channels using a two-dimensional modification of our shape recovery process.

Our existing mutual illumination photometric stereo method (Chapter 5) requires three image channels to accurately recover shape. Accordingly, here we experiment with a modification to the work of Mallick et al. [22]. Still using the box environment, we capture a second image with a lens filter applied giving a total of six image channels (which reduces to five after rotation - assuming a single causal light source). Thus there should be enough information remaining to recover the shape. In theory this seems a sensible approach, however we experienced very limited success with the method (in fact, limited to just the one object in Figures 6.19 and 6.20)!

In practice our adapted method gives poor results. Even when the light colour is known (rather than estimated), the S channel is always larger than it needs to be, effectively taking some amount of useful diffuse information with it. Still, a method that attempts to preserve diffuse shading information offers an attractive prospect for future work in handling specular highlights in single-image photometric stereo.



FIGURE 6.19: Images of a shiny bowl. Left: No filter. Right: Using an orange lens filter.

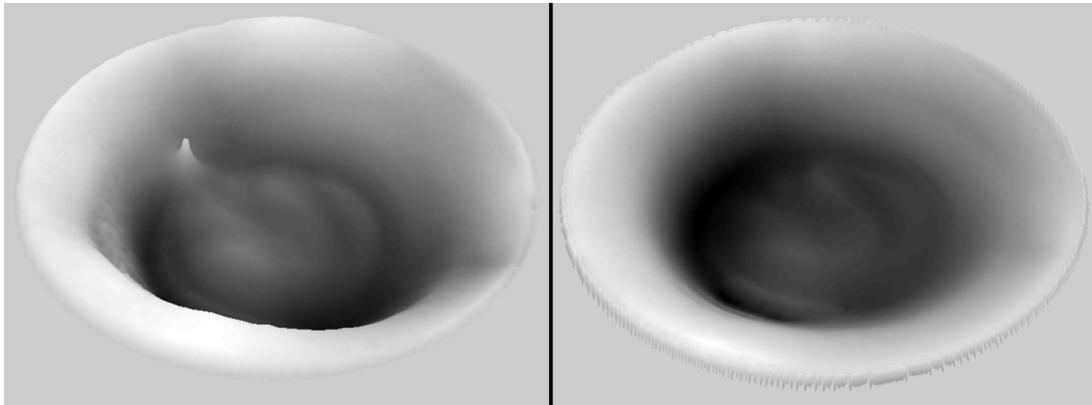


FIGURE 6.20: Recovered shape of a shiny bowl. Left: Recovery with highlight present. Right: Recovery after highlight removal.

6.2.2 Conclusion

Handling of specular highlights is a difficult problem in shape recovery. Although many solutions have been proposed (Chapter 2), to our knowledge there is no prior work which addresses the issue specifically for single-image colour photometric stereo systems such as ours. We believe that the two most promising approaches are either a more sophisticated inherent removal method or incorporation of the Phong reflectance model.

The simple detection and removal method implemented in this Chapter relies too heavily on inpainting. Any object with a significant specular component requires

a high proportion of image pixels to be removed and thus a lot of estimation is needed. Inpainting methods are designed largely with aesthetic performance in mind, rather than the restoration of data for input to shape recovery.

Inherent removal of highlights presents a more promising route for future work. Although we had very limited success with it, methods such as those of Mallick et al. are designed to preserve diffuse shading information and thus potentially well suited to single-image photometric stereo.

Another potential approach would be to adopt the Phong reflectance model. This is essentially the commonly used Lambertian model with additional terms to incorporate specular reflectance. Some shape from shading works [80, 81] have successfully recovered the shape of specular surfaces in this manner. However shape from shading uses only a single light source, so solving for the specular parameters is a simpler problem (we have upwards of four potential sources to cause highlights - the walls and any direct lights in the exterior environment).

Chapter 7

Conclusion

7.1 Thesis contributions

In this thesis we have presented the following contributions:

We have shown that given sufficiently varied colourful mutual illumination in an environment it is possible to recover the 3D shape of objects from a single image (provided that objects have uniform Lambertian reflectance, unsaturated surface colour and are of appropriate size given the environment). Despite the relaxed conditions, our mutual illumination photometric stereo method estimates shape with almost equivalent accuracy to the established Shape from Color method [48, 87, 88]; which we implemented to provide a benchmark comparison. Thus we have added to existing evidence (e.g. [7, 8]) that mutual illumination can provide hugely important information in image captures and should certainly not be ignored (as it is throughout most of computer vision).

We have provided details of an engineering process which can be used to measure the potential of lighting environments for shape recovery. In our work we used this process to create a capture environment which is specifically engineered to produce the required level of mutual illumination. The requirement for a colourful box is

somewhat restrictive, but it has served as an excellent proof of concept and given its portability we have made an important contribution in moving photometric stereo towards real-world conditions. Using our method the shape of a suitable object can potentially be recovered anywhere that one can carry a box.

We proposed novel sources for ground truth data in shape recovery. Simple paper-craft objects can be constructed quickly and are extremely accessible, they offer cheap and fast sources of objects with known ground truth. 3D-printed objects are highly customisable - any desired shape can be designed and printed from a wide range of materials. They are less accessible (given the requirement of a 3D printer) but can meet specific demands and have inherent ground truth data.

We presented an adaptation of our own system which removes the need for a calibration step with only a minimal impact on accuracy (a more sophisticated method would likely remove the loss).

We investigated the problem of single image shape recovery for objects with specular surface reflectance. Although our experiments had limited success they still provide a valuable contribution in directing future works. We have demonstrated that two possible approaches are not viable and as a result have proposed two further possible approaches which we believe to have potential at the end of Chapter 6.

Lesser contributions of this thesis which some may find particularly helpful (they certainly would have been helpful for the author at least!) include a review of literature encompassing many different subjects, an easy to follow explanation of surface normal integration and the necessary details for implementing spherical harmonic rendering.

7.2 Future work

We have proposed some future work in the previous section and throughout this thesis, however mutual illumination photometric stereo suffers from other problems which we are yet to address.

Issues in our work include the assumption of distant light sources, which is shown in the maths by not accounting for any falloff of light over distance - the enclosed box environment violates this as the box walls are definitely not infinitely distant from contained objects! Results in Chapter 5 showed that spatial location of objects within the box environment seems to influence accuracy. Also we fail to account for self-cast mutual illumination and shadows, thus it is possible for complex objects to result in pixel values which do not agree with our expected relationship between pixels and surface normals. We believe that adopting an inverse radiosity solution could potentially address these issues. There has already been some limited work in this area concerned with lighting estimation [111, 112] as well as shape recovery ([113]). However the work of Liu et al. [113] only looks at recovery for very simple planar objects with a final step which fits planes to the estimated surfaces - thus recovery of any useful level of detail is not possible. Given a known environment (either using our box or the scene understanding methods mentioned in the previous section) one would essentially have a radiosity model with only the object as an unknown quantity to be solved for. It is perhaps an ambitious approach, but modelling the transfer of energy in this way has the potential to inherently fix many problems in photometry-based shape recovery.

Our approach towards removal of a calibration step could be improved using existing work. We propose incorporation of 3D scene understanding for the purpose of extracting information about the capture environment visible in an image. For example, Gupta et. al [23] are able to estimate the spatial layout of rooms (Figure 7.1). The geometry and colour of surfaces would then be known and we could

estimate light sources as in our current attempt at removal of calibration. Adaptation of such work could allow us to remove the box requirement entirely (instead using the mutual illumination of real scenes to recover shape) as well as giving this approach to calibration removal a wider applicability (at present it is only useful for capture environments of known shape like ours.)

With concern to the issue of surfaces with specular components, recent work [114] has used a library of bidirectional reflectance distribution functions (BRDFs) to aid recovery of such objects from multiple images. This is a concept which we also played with briefly (without success), but perhaps poses a promising approach; though it will be more difficult for our case of a single colour image.



FIGURE 7.1: Left: Input image. Right: The room is segmented according to 3D scene understanding [23].

As currently presented, mutual illumination photometric stereo falls short of being useful for direct implementation in the examples given in Chapter 1 Section 1.2. However with further work we believe both applications are possible. The security corridor problem could be solved by incorporating the work of [10] (wherein shape from colour is used as part of a shape recovery system which processes video capture). The other example given of a possible application was for mutual illumination photometric stereo to be used in remote situations where power sources are not available. Our method as it stands could be of some limited use for this purpose, however the range of objects which it could be applied to is limited -

i.e. Lambertian, suitably coloured and small enough to fit inside the box (which must also be transported to said remote location). Addressing issues with specular surfaces would increase the range of suitable objects and the aforementioned radiosity model combined with scene understanding approach could remove the need for box transportation.

7.3 Final remarks

In time recovery of 3D shape will likely become an easier problem as specialist hardware becomes increasingly sophisticated (e.g. development of higher resolution depth sensors), but for now it remains a difficult task.

In this thesis we have presented a new approach to shape recovery which requires only a single image, moves photometric stereo closer to real-world conditions and demonstrates the importance of mutual illumination in images.

Ultimately our work makes an important contribution to the field of photometric shape recovery and offers several interesting options for future research.

Bibliography

- [1] Daniel Vlasic, Pieter Peers, Ilya Baran, Paul Debevec, Jovan Popović, Szymon Rusinkiewicz, and Wojciech Matusik. Dynamic shape capture using multi-view photometric stereo. In *ACM Transactions on Graphics (TOG)*, volume 28, page 174. ACM, 2009.
- [2] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [3] Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1067–1073. IEEE, 1997.
- [4] Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International journal of computer vision*, 38(3):199–218, 2000.
- [5] George Vogiatzis and Carlos Hernández. Self-calibrated, multi-spectral photometric stereo for 3d face capture. *International Journal of Computer Vision*, 97(1):91–103, 2012.
- [6] Katsushi Ikeuchi and Berthold KP Horn. Numerical shape from shading and occluding boundaries. *Artificial intelligence*, 17(1-3):141–184, 1981.
- [7] Alan L Gilchrist. *The perception of surface blacks and whites*. WH Freeman, 1979.

-
- [8] Alexa I Ruppertsberg and Marina Bloj. Reflecting on a room of one reflectance. *Journal of Vision*, 7(13):12–12, 2007.
- [9] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2553–2560, 2011.
- [10] Gabriel J Brostow, Carlos Hernández, George Vogiatzis, Björn Stenger, and Roberto Cipolla. Video normals from colored lights. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2104–2114, 2011.
- [11] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International journal of computer vision*, 35(1):33–44, 1999.
- [12] Athinodoros S Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 816–823. Ieee, 2003.
- [13] Manmohan Krishna Chandraker, Chandraker Fredrik Kahl, and David J Kriegman. Reflections on the generalized bas-relief ambiguity. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 788–795. IEEE, 2005.
- [14] Jae Hyun Won, Man Hee Lee, and In Kyu Park. Active 3d shape acquisition using smartphones. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 29–34. IEEE, 2012.
- [15] Feng Lu, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, and Yoichi Sato. From intensity profile to surface normal: photometric stereo for unknown light sources and isotropic reflectances. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(10):1999–2012, 2015.

-
- [16] Grant Schindler. Photometric stereo via computer screen lighting for real-time surface reconstruction. In *Proc. International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), CD-ROM*, 2008.
- [17] Ajmal Mian. Illumination invariant recognition and 3d reconstruction of faces using desktop optics. *Optics express*, 19(8):7491–7506, 2011.
- [18] Micah K Johnson, Forrester Cole, Alvin Raj, and Edward H Adelson. Microgeometry capture using an elastomeric sensor. In *ACM Transactions on Graphics (TOG)*, volume 30, page 46. ACM, 2011.
- [19] Todd Zickler, Satya P Mallick, David J Kriegman, and Peter N Belhumeur. Color subspaces as photometric invariants. *International Journal of Computer Vision*, 79(1):13–30, 2008.
- [20] Y He, Nitin Khanna, Carol J Boushey, and Edward J Delp. Specular highlight removal for image-based dietary assessment. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 424–428. IEEE, 2012.
- [21] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. *Proceedings of SIGGRAPH 2001*, pages 497–500, 2001.
- [22] Satya P Mallick, Todd E Zickler, David J Kriegman, and Peter N Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 619–626. Ieee, 2005.
- [23] Abhinav Gupta, Martial Hebert, Takeo Kanade, and David M Blei. Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces. In *Advances in neural information processing systems*, pages 1288–1296, 2010.

-
- [24] Peter Kovési. Shapelets correlated with surface normals produce surfaces. *Tenth IEEE International Conference on Computer Vision (ICCV)*, 2:994–1001, 2005.
- [25] Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.
- [26] Tal Simchony, Rama Chellappa, and M Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):435–446, 1990.
- [27] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.
- [28] Berthold KP Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, Massachusetts Institute of Technology, 1970.
- [29] Edwin H Land and John J McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971.
- [30] David Forsyth and Andrew Zisserman. Mutual illumination. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*., pages 466–473. IEEE, 1989.
- [31] Shree K Nayar, Katsushi Ikeuchi, and Takeo Kanade. Shape from inter-reflections. *International Journal of Computer Vision*, 6(3):173–195, 1991.
- [32] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011.
- [33] Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004.

-
- [34] JL Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1): 1–17, 1982.
- [35] Henry Harlyn Baker. Depth from edge and intensity based stereo. Technical report, DTIC Document, 1982.
- [36] Stan Birchfield and Carlo Tomasi. Depth discontinuities by pixel-to-pixel stereo. In *Computer Vision, 1998. Sixth International Conference on*, pages 1073–1080. IEEE, 1998.
- [37] Myron Z Brown, Darius Burschka, and Gregory D Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.
- [38] Nalpantidis Lazaros, Georgios Christou Sirakoulis, and Antonios Gasteratos. Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, 2(4):435–462, 2008.
- [39] Philip Pritchett and Andrew Zisserman. Wide baseline stereo matching. In *Computer Vision, 1998. Sixth International Conference on*, pages 754–760. IEEE, 1998.
- [40] Adrien Treuille, Aaron Hertzmann, and Steven M Seitz. Example-based stereo with general brdfs. In *European Conference on Computer Vision*, pages 457–469. Springer, 2004.
- [41] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528. IEEE, 2006.
- [42] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Patrice Horaud. *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media, 2012.

- [43] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (tof) cameras: a survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.
- [44] Valentina Campanelli, Stephen M Howell, and Maury L Hull. Accuracy evaluation of a lower-cost and four higher-cost laser scanners. *Journal of biomechanics*, 49(1):127–131, 2016.
- [45] Ivo Ihrke, Kiriakos N Kutulakos, Hendrik Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. In *Computer Graphics Forum*, volume 29, pages 2400–2426. Wiley Online Library, 2010.
- [46] Sebastian Schuon, Christian Theobalt, James Davis, and Sebastian Thrun. High-quality scanning using time-of-flight depth superresolution. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–7. IEEE, 2008.
- [47] Graham D. Finlayson and Christopher W. Powell. Shape in a box. *Submitted for review to British Machine Vision Conference*, 2014.
- [48] Mark S Drew. Shape from color. Technical report, Simon Fraser University, 1992.
- [49] Berthold KP Horn and Michael J Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986.
- [50] Berthold KP Horn and Michael J Brooks. *Shape from shading*. MIT press, 1989.
- [51] Ruo Zhang, P-S Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.
- [52] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Transactions on Graphics (TOG)*, 22(3):313–318, 2003.

-
- [53] Amit Agrawal, Ramesh Raskar, and Rama Chellappa. What is the range of surface reconstructions from a gradient field? In *European Conference on Computer Vision*, pages 578–591. Springer, 2006.
- [54] Zak Murez, Tali Treibitz, Ravi Ramamoorthi, and David J Kriegman. Photometric stereo in a scattering medium. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1880–1891, 2017.
- [55] Jingtang Liao, Bert Buchholz, Jean-Marc Thiery, Pablo Bauszat, and Elmar Eisemann. Indoor scene reconstruction using near-light photometric stereo. *IEEE Transactions on Image Processing*, 26(3):1089–1101, 2017.
- [56] Carlos Hernández, George Vogiatzis, Gabriel J Brostow, Björn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. *IEEE 11th International Conference on Computer Vision (ICCV)*, pages 1–8, 2007.
- [57] Paulo F. U. Gotardo, Tomas Simon, Yaser Sheikh, and Iain Matthews. Photogeometric scene flow for high-detail dynamic 3d reconstruction. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [58] Alan Gilchrist and Alan Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13(1):5–19, 1984.
- [59] Gunter Wyszecki and Walter Stanley Stiles. *Color science*, volume 8. Wiley New York, 1982.
- [60] David Forsyth and Andrew Zisserman. Shape from shading in the light of mutual illumination. *Image and vision computing*, 8(1):42–49, 1990.
- [61] Bas Boom, Sergio Orts-Escolano, Xi Ning, Steven McDonagh, Peter Sandilands, and Robert B Fisher. Point light source estimation based on scenes recorded by a rgb-d camera. In *BMVC*, 2013.

-
- [62] Andrew D Spence. Calibrated and uncalibrated photometric stereo for surface texture acquisition. 2005.
- [63] Richard F Murray. The statistics of shape, reflectance, and lighting in real-world scenes.
- [64] Ondrej Drbohlav and M Chaniler. Can two specular pixels calibrate photometric stereo? In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1850–1857. IEEE, 2005.
- [65] Paolo Favaro and Thoma Papadimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 821–828. IEEE, 2012.
- [66] Christopher Powell and Graham D Finlayson. Mutual illumination photometric stereo without calibration. *Color and Imaging Conference*, 2015(1): 212–218, 2015.
- [67] James J Clark. Photometric stereo using lcd displays. *Image and Vision Computing*, 28(4):704–714, 2010.
- [68] Lina Bi, Zhan Song, and Linmin Xie. A novel lcd based photometric stereo method. In *Information Science and Technology (ICIST), 2014 4th IEEE International Conference on*, pages 611–614. IEEE, 2014.
- [69] Thoma Papadimitri and Paolo Favaro. Uncalibrated near-light photometric stereo. In *BMVC*, 2014.
- [70] Micah K Johnson and Edward H Adelson. Retrographic sensing for the measurement of surface texture and shape. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1070–1077. IEEE, 2009.

- [71] Klett, Eberhard Witwe, Detleffsen, Christoph Peter, et al. *IH Lambert... Photometria sive de mensura et gradibus luminis, colorum et umbrae*. sumptibus viduae Eberhardi Klett, 1760.
- [72] Johann Heinrich Lambert and DL DiLaura. Photometry, or on the measure and gradations of light, colors and shade. *New York: Illuminating Engineering Society of North America*, 2001.
- [73] E North Coleman and Ramesh Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer graphics and image processing*, 18(4):309–328, 1982.
- [74] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.
- [75] Svetlana Barsky and Maria Petrou. Design issues for a colour photometric stereo system. *Journal of Mathematical Imaging and Vision*, 24(1):143–162, 2006.
- [76] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Shadows in three-source photometric stereo. *European Conference on Computer Vision (ECCV)*, pages 290–303, 2008.
- [77] Hossein Ragheb and Edwin R Hancock. A probabilistic framework for specular shape-from-shading. *Pattern Recognition*, 36(2):407–427, 2003.
- [78] Gudrun J Klinker, Steven A Shafer, and Takeo Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1):7–38, 1990.
- [79] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

- [80] Oliver Vogel, Michael Breuß, and Joachim Weickert. Perspective shape from shading with non-lambertian reflectance. *Pattern Recognition*, pages 517–526, 2008.
- [81] Guohui Wang, Wei Su, and Fen Gao. A specular shape from shading by fast marching method. *Procedia Engineering*, 24:192–196, 2011.
- [82] Ali Madooei and Mark S Drew. Detecting specular highlights in dermatological images. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 4357–4360. IEEE, 2015.
- [83] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000.
- [84] Nirali Pandya and Bhailal Limbasiya. A survey on image inpainting techniques. *International Journal of Current Engineering and Technology*, 3(5):1828–1831, 2013.
- [85] Robby T Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):178–193, 2005.
- [86] Qingxiong Yang, Shengnan Wang, and Narendra Ahuja. Real-time specular highlight removal using bilateral filtering. In *European Conference on Computer Vision*, pages 87–100. Springer, 2010.
- [87] Mark S Drew. Photometric stereo without multiple images. *Electronic Imaging '97*, pages 369–380, 1997.
- [88] Mark S Drew and Michael H Brill. Color from shape from color: A simple formalism with known light sources. *Journal of the Optical Society of America (JOSA)*, 17(8):1371–1381, 2000.

-
- [89] Michael F Cohen and John R Wallace. *Radiosity and realistic image synthesis*. Elsevier, 2012.
- [90] Christoph Reinhart and Pierre-Felix Breton. Experimental validation of autodesk® 3ds max® design 2009 and daysim 3.0. *Leukos*, 6(1):7–35, 2009.
- [91] Francesc Moreno-Noguer and Josep M Porta. Probabilistic simultaneous pose and non-rigid shape recovery. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1289–1296. IEEE, 2011.
- [92] Paul Francis Jacobs. *Rapid prototyping & manufacturing: fundamentals of stereolithography*. Society of Manufacturing Engineers, 1992.
- [93] Chee Kai Chua, Kah Fai Leong, and Chu Sing Lim. *Rapid prototyping: principles and applications*. World Scientific, 2010.
- [94] GENES Miller and C Robert Hoffman. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Advanced Computer Graphics Animation seminar notes*, volume 190, 1984.
- [95] MJ Kilgard. Perfect reflections and specular lighting effects with cube environment mapping. *Technical Brief, nVidia Corp*, 1999.
- [96] Tien-Tsin Wong, Liang Wan, Chi-Sing Leung, and Ping-Man Lam. Real-time environment mapping with equal solid-angle spherical quad-map. *Shader X4: Lighting & Rendering*, W. Engel, ed., Charles River Media, 2006.
- [97] Ned Greene. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications*, 6(11):21–29, 1986.
- [98] Robin Green. Spherical harmonic lighting: The gritty details. *Archives of the Game Developers Conference*, 2003.
- [99] Volker Schönefeld. Spherical harmonics. Technical report, RWTH Aachen University, 2005.

-
- [100] Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2):218–233, 2003.
- [101] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *Proceedings of SIGGRAPH 2007*, pages 369–378, 2007.
- [102] Kristian Kirk and Hans Jørgen Andersen. Noise characterization of weighting schemes for combination of multiple exposures. *British Machine Vision Conference (BMVC)*, pages 1129–1138, 2006.
- [103] Berthold KP Horn. Understanding image intensities. *Artificial Intelligence*, 8(2):201–231, 1977.
- [104] Tiangong Wei and Reinhard Klette. Height from gradient using surface curvature and area constraints. In *ICVGIP*. Citeseer, 2002.
- [105] Tsai Ping-Sing and Mubarak Shah. Shape from shading using linear approximation. *Image and Vision computing*, 12(8):487–498, 1994.
- [106] Vishal M Patel and Rama Chellappa. Approximation methods for the recovery of shapes and images from gradients. *Applied and Numerical Harmonic Analysis*, pages 377–398, 2013.
- [107] Ariel Tankus and Nahum Kiryati. Photometric stereo under perspective projection. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 611–616. IEEE, 2005.
- [108] Thoma Papadhimetri and Paolo Favaro. A new perspective on uncalibrated photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1474–1481, 2013.

-
- [109] Michael F Cohen and Donald P Greenberg. The hemi-cube: A radiosity solution for complex environments. In *ACM SIGGRAPH Computer Graphics*, volume 19, pages 31–40. ACM, 1985.
- [110] Selim Esedoglu and Jianhong Shen. Digital inpainting based on the mumford–shah–euler image model. *European Journal of Applied Mathematics*, 13(04):353–370, 2002.
- [111] Magali Contensin. Inverse lighting problem in radiosity. *Inverse Problems in Engineering*, 10(2):131–152, 2002.
- [112] Yizhou Yu, Paul Debevec, Jitendra Malik, and Tim Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. *Proceedings of SIGGRAPH 1999*, pages 215–224, 1999.
- [113] Siying Liu, Tian-Tsong Ng, and Yasuyuki Matsushita. Shape from second-bounce of light transport. In *European Conference on Computer Vision*, pages 280–293. Springer, 2010.
- [114] Feng Lu, Xiaowu Chen, Imari Sato, and Yoichi Sato. Symps: Brdf symmetry guided photometric stereo for shape and light source estimation. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):221–234, 2018.