PHOTOMETRIC STEREO AND APPEARANCE CAPTURE

ZHOU ZHENG LONG (B.Sc., SJTU)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2014

ii

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Mr Mar. 2014 Zhenglong Zhou

17.Mar.2014

ii

Acknowledgements

First of all, I would like to thank my PhD supervisor Professor Ping Tan. This thesis would not have been possible without the help, support and patience of Prof.Tan, not to mention his advice and unsurpassed knowledge of photometric stereo and computer vision. I would also like to acknowledge the financial, academic and technical support of the Department of Electrical and Computer Engineering and National University of Singapore. They gave me the chance to pursue my career in computer vision and graphics. I would also like to thank the committee for your effort in reviewing my thesis.

I would like to thank my parents and sister who support me at all times. My mere expression of thanks does not suffice.

Wu Zhe helped me a lot in my research. I am very grateful to him for his help in this thesis.

Last, but by no means least, I thank all lab staff and my friends for their support and encouragement throughout the past five years.

Contents

Co	onten	ts			iv
Su	ımma	ry			viii
Li	st of l	Figures			X
1	Intr	oductio	n		1
	1.1	Photor	netric Ster	eo	1
	1.2	Main o	challenges	in photometric stereo	8
		1.2.1	Auto-cal	ibration	8
			1.2.1.1	Light source calibration	10
			1.2.1.2	Light source calibration with perspective effect	13
		1.2.2	Non-Lan	nbertian Material	14
	1.3	Applic	cation in ap	ppearance capture	16
		1.3.1	Design g	oals	18
	1.4	3D rec	constructio	n methods	19
		1.4.1	Multi-vie	ew stereo	19
		1.4.2	Active ra	ngefinding	20
	1.5	Object	ives		22

CONTENTS

2	Pho	tometri	c stereo		23
	2.1	Basic	radiometry	7	23
		2.1.1	Basic co	ncepts in radiometry	24
	2.2	Surfac	e reflection	n and BRDF	27
		2.2.1	Lambert	ian reflection	29
		2.2.2	Microfac	cet models	30
			2.2.2.1	Oren-Nayar diffuse reflection	31
			2.2.2.2	Torrance-Sparrow model	32
		2.2.3	Measure	d BRDFs	33
			2.2.3.1	Isotropic reflectance model	34
	2.3	Basics	of photon	netric stereo	35
		2.3.1	Lambert	ian photometric stereo	35
		2.3.2	Lambert	ian photometric stereo: factorization approach	37
3	Auto	o-calibr	ation		39
	3.1	Ring-l	ight photo	metric stereo	39
		3.1.1	Ring-Lig	ght photometric stereo	40
			3.1.1.1	Uncalibrated photometric stereo	41
			3.1.1.2	Constraints from a ring-light	42
			3.1.1.3	Ring-light ambiguities	45
		3.1.2	A compl	ete stratified reconstruction	46
			3.1.2.1	Lights with equal interval	47
			3.1.2.2	Lights with equal intensity	47
			3.1.2.3	Two corresponding normals in two views	49
		3.1.3	Experim	ents	51

CONTENTS

			3.1.3.1	A prototype device	54
		3.1.4	Conclusi	on	56
	3.2	Near-l	ight Photo	metric Stereo	57
		3.2.1	Related	works	58
			3.2.1.1	Background	59
		3.2.2	Ambigui	ty in uncalibrated near-light photometric stereo	60
			3.2.2.1	Patch-based factorization	60
			3.2.2.2	Correlations of the ambiguities among patches	61
			3.2.2.3	Intrinsic shape-lighting ambiguities	62
		3.2.3	Disambi	guation methods	63
			3.2.3.1	Solution with one patch calibrated	63
			3.2.3.2	Solution with two patches calibrated	65
		3.2.4	Calibrate	ed perspective photometric stereo	65
			3.2.4.1	Light fall-off depth cue	66
			3.2.4.2	Depth consistency at neighboring pixels	67
			3.2.4.3	Graphical model for depth and normal recovery	68
		3.2.5	Experim	ents	69
		3.2.6	Conclusi	on and discussion	73
4	Non	-Lambe	ertian nho	tometric stereo	74
-	1 1	Leo de	eth contou	r actimation	75
	4.1	Iso-de	pui contou		15
	4.2	Experi	ment		81
		4.2.1	Errors in	iso-depth contours	84
		4.2.2	Number	of images at each viewpoint	84
	4.3	Conclu	usion		86

CONTENTS

5	App	earance capture by multi-view photometric stereo	88		
	5.1	Related work	89		
	5.2	System pipeline	91		
	5.3	Shape reconstruction: multi-view depth propagation	93		
	5.4	Reflectance capture	96		
	5.5	Experiment	98		
		5.5.1 Handheld system	99		
		5.5.2 Ring-light system	100		
		5.5.3 Comparison with existing methods	100		
	5.6	Re-rendering	103		
		5.6.1 Runtime efficiency	106		
	5.7	Discussion	106		
6	Con	clusions	108		
Re	feren	ces	110		
Ap	Appendix A: Proof of Proposition 1				
Ap	Appendix B: Determine <i>t</i> , <i>s</i> from <i>F</i>				
Ap	Appendix C: Constants in Equation 3.8–3.10				

Summary

In this thesis, we study photometric stereo and combine it with multi-view stereo to efficiently capture objects with complex geometry and materials. Photometric stereo recovers surface shape from images taken under different lighting conditions. Auto-calibration photometric stereo methods recover surface shape and lighting directions at the same time. In this thesis, we propose two novel auto-calibration methods. One method exploits special ring-light configurations to resolve general Lambertian surfaces which cannot be handled by previous methods. Another method is based on patch factorization which takes advantage of near-object light sources to resolve ambiguities.

For non-Lambertian surface, previous photometric stereo methods often assume parametric or known reflectance model, which limits their wide application. In this thesis, we propose a new multi-view photometric stereo technique that can work for general isotropic materials. Starting from a single viewpoint, we use a set of photometric images to identify 'iso-depth contours'. We collect these contours from multiple viewpoints and combine it with multi-view stereo to obtain a precise reconstruction of the complete 3D shape. The spatially varying isotropic bidirectional reflectance distribution function (SVBRDF) is captured by simultaneously inferring a set of basis BRDFs and their mixing weights at each surface point. We validate our approach with a wide range of objects of different materials. We show that with a single digital camera and a moving light source, both the details of the geometry and the reflectance can be faithfully obtained. We also show that a simplicity/quality trade-off can be achieved by using additional hardware such as a ring-light device.

List of Figures

1.1	Photometric stereo pipeline. (a) input image (b) estimated surface ori-	
	entations (c) integrated 3D surface	2
1.2	Lambertian reflectance model. Lights are reflected equally in all direc-	
	tions. The distribution of reflected energies according to the Lambert's	
	model in 2D and 3D space ((a) and (b)). (c) appearance of a Lamber-	
	tian diffuse sphere.	2
1.3	Complex behaviours when light interacts with physical world	3
1.4	Camera and lighting model of photometric stereo (a) Orthographic	
	projection (b) Correspondences between image and surface	3
1.5	Estimating normal from multiple light sources	4
1.6	Multiview photometric stereo Hernández et al. [2008] (a) data acquisi-	
	tion setup (b) one of the input images (c) recovered shape \ldots .	6
1.7	Shape and Spatially-Varying BRDFs From Photometric Stereo	6
1.8	Normals acquired with photometric stereo can improve 3D scanned	
	shape (a). The resulted shape (b) has lower noise and much real details.	7

1.9	The microgeometry capture system consists of an elastometric sen-	
	sor and a high-magnification camera (a). THe retrographics sensor re-	
	places the BRDF of the subject with its own (b), allowing microscopic	
	geometry to be accurately captured (c)	8
1.10	The microgeometry capture system consists of an elastometric sensor	
	and a high-magnification camera (a). The retrographics sensor replaces	
	the BRDF of the subject with its own (b), allowing microscopic geom-	
	etry to be accurately captured (c)	8
1.11	Capturing lighting directions by inserting a metal sphere	9
1.12	Top row shows a laser scanned human head rendered as a Lambertian	
	surface with constant albedo. The subsequent three rows show images	
	of heads whose shapes have been transformed by different generalized	
	bas-relief transformations. The profile views of the face in the third	
	column reveal the nature the individual transformations and the direc-	
	tion of the light source. From the frontal views (first column), the true	
	3-d structure of the objects cannot be determined	11
1.13	Challenging data for uncalibrated photometric stereo. (a) is too simple	
	and (b) is too complicate for most of existing methods	12
1.14	Example based photometric stereo matches test object's reflectance	
	measurements to the reference object to find the corresponding sur-	
	face normal (a). Objects with arbitrary complex reflectance model (b)	
	can be correctly recovered (c).	16

1.15	On the visible hemisphere of Gauss sphere, two normals form an <i>isotropic</i>	
	pair if they lie at the intersections of two circles centered at source di-	
	rection \mathbf{s} and view direction \mathbf{v} . If the BRDF is isotropic the observed	
	intensity at these points will be equal.	17
2.1	Lambert's law. As illumination is over a larger area at smaller inci-	
	dent angles, irradiance (E) arriving at a surface varies according to the	
	cosine of the angle of incidence of illumination.	25
2.2	Solid angle. By projecting c onto the unit sphere, the area of its projec-	
	tion is the solid angle s subtended by the object c in three dimensions.	26
2.3	Radiance L is defined as flux per unit solid angle $d\omega$ per unit projected	
	area $\mathrm{d}A^{\perp}$	27
2.4	BRDF model	29
2.5	Reflection from a surface can be generally categorized by the distri-	
	bution of reflected light: (a) diffuse, (b) glossy specular, (c) perfect	
	specular, and (d) retro-reflective distributions.	30
2.6	Color checker is made of special material which is very close to Lam-	
	bertian reflectance.	31
2.7	The BRDF of a target surface is measured by moving a light source	
	and a camera.	34
2.8	Sphere with same material can ease BRDF capturing	35

3.1	Ring-light photometric stereo. (a) Lighting directions lie on a view	
	centered cone. The term ω denotes the cone opening angle. (b) In	
	the projective plane, these lights lie on a ring centered at origin (i.e.	
	viewing direction). (c) When there is a linear ambiguity, these lights lie	
	on a general planar conic. Our algorithm resolves this linear ambiguity	
	by mapping lights back to their canonic positions	43
3.2	Geometric explanations of the components of the 'ring-light ambigu-	
	ity'. The first row shows the transformations induced to lighting di-	
	rections in the projective plane. The second row illustrates the corre-	
	sponding transformations to a 3D shape	44
3.3	The cost as a function of the hypothesized ϕ_1 . This functions has a	
	clear global minimum because ϕ_1, ϕ_2, s_1, s_2 are uniquely determined	
	in principle	49
3.4	The first and the second row are the angular errors of reconstructed	
	normal directions by using equal lighting interval and equal lighting in-	
	tensity constraint respectively. Typically, equal lighting intensity con-	
	straint generates more accurate results	51
3.5	Results for the challenging data in Figure 1. On the left are results up	
	to the ring-light ambiguity. In the middle is our reconstructed surface	
	normals. For a validation, we calibrate all incident lighting directions	
	with a metal sphere and use calibrated photometric stereo to compute	
	a ground truth in the right. Our result is very consistent to the ground	
	truth	52

3.0	6 Additional results. From left to right, they are one of the input images,	
	reconstruction up to the ring-light ambiguity and the final Euclidian	
	reconstruction. Some of the artifacts are due to non-Lambertian effects	
	like shadow and highlight.	53
3.7	7 Averaged angular error in the recovered normal directions as a function	
	of the cone opening angle and the angle between two viewpoints. In	
	most of time, the reconstruction error is smaller than 5 degrees	54
3.8	8 Shown in (a) is a prototype device. 20 LED bulbs lie on a circle with ra-	
	dius of 150 millimeters centered at the viewing direction. Our method	
	allows us to consider the weak perspective effects of the lighting which	
	is critical for a handheld photometric stereo setup operating at rela-	
	tively small distance. This weak perspective effects is illustrated in	
	(b). To ensure the opening angle of the cone is larger than 15 degrees,	
	the distance between the camera and captured objects should be within	
	1.2 meters	55
3.9	9 Perspective and light fall-off effects in near-light photometric stereo .	59
3.	10 The light sources positions $\mathbf{p}_1, \mathbf{p}_2$ can be obtained by intersecting the	
	corresponding lighting directions at two patches $\mathbf{o}^{(1)}, \mathbf{o}^{(2)}$. The depth	
	of $\mathbf{o}^{(2)}$ is decided by minimizing the distance between corresponding	
	light rays	64
3.	11 Reconstruction error w.r.t. the hypothesized depth d . Camera is located	
	at $d = 0$, and light sources are located at around $d = 110$. The left and	
	right are plots from synthetic and real data respectively.	67

69

- 3.12 Experiments on synthetic data. (a) shows one of the input image. (b) is the ground truth normal as a reference. (c) is the normal map computed by patch-based factorization. Each patch is subject to a different (and unknown) linear transform. (d) shows the result from the solution with one patch calibrated (known $\mathbf{A}^{(1)}, \mathbf{o}^{(1)}$). (e) shows the result from the solution with two patches calibrated (known $\mathbf{A}^{(1)}, \mathbf{o}^{(1)}$ and $\mathbf{A}^{(2)}, \mathbf{o}^{(2)}$).
- 3.13 Angular errors in the recovered normal map. From left to right, they are the errors of conventional directional lighting photometric stereo, errors in our solution with one patch calibrated and error in our solution with two patches calibrated. The angular error (in degrees) are visualized according to the color bar on the right.
 71
- 4.1 Illustration of an isotropic pair. Symmetric light source vectors are obtained by reflecting s about the plane spanned by the surface normal n and viewing direction v. The observed lighting intensities of s' and s are the same.
 4.2 (a) An input image with cast shadow; (b) Fourier series fitted image; 77

4.3	(a) The symmetry axis of intensity profiles tells the azimuth angle of	
	a pixel's normal direction; (b) cast shadows can break this symmetry;	
	(c) the intensity profile of most of isotropic BRDFs in Mcmillan et al.	
	[2003] can be well represented by a 2-order Fourier series	78
4.4	We compute a Delaunay triangulation of the original lighting direc-	
	tions (red dots) in the projective plane. The desired observations (blue	
	dots) on a view-centered circle are generated by linear interpolation	
	within these triangles. Left: the circle radius d is the mean distance	
	between the red dots and the viewpoint v. Right: the circle radius d is	
	set as $(d_i + d_o)/2$. Here, d_i (or d_o) is the largest (or smallest) distance	
	between \mathbf{v} and the red dots on the inner (or outer) conic	79
4.5	The hardware setup. Left: our device consists of a video camera and	
	two circles of LED lights. Right: we need to calibrate one parameter	
	θ_0 to determine lighting positions	82
4.6	Top view of the calibration setup. We capture a diffuse board at several	
	known positions (black lines) to calibrate camera vignetting and light	
	intensity. Some additional boards (blue lines) are used to calibrate the	
	angle θ_0	83
4.7	Average depth error of all iso-depth contours of the 'Cat' example.	
	This error increases linearly with contour length	85
4.8	Mean shape error of the 'Buddha' example. This error does not change	
	significantly with different number of LEDs	85
4.9	Relative RMSE BRDF error of a green paint. This error generally	
	decreases with more LEDs.	86

5.1	System pipeline. We recover iso-depth contours from photometric	
	stereo images and recover a sparse 3D point cloud by structure-from-	
	motion. In the figure showing iso-depth contours, the gray intensity	
	encodes the estimated azimuth angles, and the colored curves are iso-	
	depth contours. We then propagate the depths of these 3D points along	
	the iso-depth contours to recover the complete 3D shape. Once the	
	shape is fixed, we estimate the spatially varying BRDF from the origi-	
	nal input images.	92
5.2	We propagate the depth of \mathbf{x} to the iso-depth contour segment C_i that	
	passes through its projection in the i -th view. This propagation gener-	
	ates new 3D points, e.g. $\mathbf{y}_1, \mathbf{y}_2$, whose depths in other images can also	
	be propagated along their corresponding iso-depth contours C_{j1}, C_{j2} .	93
5.3	(a) initially reconstructed 3D points; (b) 3D points obtained by depth	
	propagation; (c) initial shape after Poisson surface reconstruction; (d)	
	final result.	96
5.4	Definition of θ_h , θ_d and ϕ	97
5.5	The normalized BRDF mixture weights are visualized in the different	
	color channels. The corresponding basis BRDFs are used to render a	
	sphere on the right.	98
5.6	Results from the handheld system. (a) one of the input images, (b) the	
	recovered shape rendered with uniform diffuse shading. (c) a render-	
	ing with the recovered reflectance model from the same viewpoint and	
	lighting condition as the image in (a). (d) the color-coded shape error	
	(in millimeters) compared to laser-scanned 'ground truth'	99

5.7	Results from the ring-light system. From left to right, these figures are	
	arranged in the same way as Figure 5.6	101
5.8	Results according to Alldrin et al. [2008]. (a) the color coded normal	
	map estimated. (b) the shape computed from the estimated normal	
	according to Wu and Tang [2006]. (c) a rendering under novel illumi-	
	nation. (d) the color coded shape error (in millimeters)	102
5.9	Results according to Hernández et al. [2008]	102
5.10	Results from handheld system. Left: reference real photographs. Right:	
	rendering under novel viewpoint and lighting condition	104
5.11	Results from the ring-light system. Left: reference real photographs.	
	Right: rendering under novel viewpoint and lighting condition	104
5.12	Rendering under environment lighting	105

Chapter 1

Introduction

1.1 Photometric Stereo

This dissertation studies the technique in computer vision called *photometric stereo*.

Photometric stereo was originally introduced by Woodham [1980] to determine surface orientation from multiple images. The idea is that from a smooth shaded object (Figure 1.1 (a)), we can usually perceive its shape from just the shading variation. However, a single image often does not provide enough information to determine the underlying shape. In order to overcome this limitation, photometric stereo holds the viewing direction constant, while varying the direction of incident illumination between successive images. Since the imaging geometry is not changed during capturing, the correspondence between image points is known. At each image point, these corresponding pixel intensities changes as a function of the angle between the local surface normal and incident illumination. As a result the recorded radiance values can determine the surface orientation. Figure 1.1 (b) and (c) show the estimated surface orientations from photometric stereo and the integrated surface from these orientations.



Figure 1.1: Photometric stereo pipeline. (a) input image (b) estimated surface orientations (c) integrated 3D surface



Figure 1.2: Lambertian reflectance model. Lights are reflected equally in all directions. The distribution of reflected energies according to the Lambert's model in 2D and 3D space ((a) and (b)). (c) appearance of a Lambertian diffuse sphere.

Lambertian photometric stereo is one of the most fundamental photometric stereo algorithms. There are three assumptions for Lambertian photometric stereo:

Lambertian reflectance model. A *reflectance model* describes how a surface interacts with light. When light strikes a surface, it can be absorbed, reflected, scattered, or travel along the surface and leave at some other points. Many of these effects are shown in Figure 1.3. The *Lambertian reflectance model* is named after Johann Heinrich Lambert, who introduced the concept of perfect diffusion in this 1760 book *Photometria*. The Lambertian model reflects the incoming light equally to all directions in the upper hemisphere above the surface (Figure 1.2). Thus the apparent brightness of such a surface to an observer is the same regardless of the observer's angle of view.



Figure 1.3: Complex behaviours when light interacts with physical world.



Figure 1.4: Camera and lighting model of photometric stereo (a) Orthographic projection (b) Correspondences between image and surface.

Orthographic camera model. The camera is assumed to be orthogonal, as illustrated in Figure 1.4. If the size of the object in view is small compared to the viewing distance, then the perspective projection can be approximated as an orthographic projection where all the projection lines are orthogonal to the projection plane. This projection ensures every pixel in the image have the same viewing direction to simplify the following analysis. Consider a camera that performs an orthographic projection. It is convenient to choose a coordinate system such that the viewing direction is aligned with the z-axis as shown in Figure 1.4 (b). It is also convenient to assume appropriate scaling of the image plane such that the object point Z(x, y) maps onto image point



Figure 1.5: Estimating normal from multiple light sources

 (x_0, y_0) where $x_0 = x$, $y_0 = y$. As a result, the viewing direction **v** is the same for all surface points.

Directional illumination. For a fixed light source, the incident lighting direction s is given by the positions of the surface point and the light, as shown in Figure 1.4 (b). We call the lighting condition *directional* if the incident lighting directions at all surface points are constant. For example, sunlight is generally considered as directional. In practice, if the light source is sufficiently far away from the object, we can approximate the lighting condition as directional illumination.

Under these three assumptions, given three or more light sources that are not colinear (Figure 1.5) and with known directions and intensities, Woodham [1980] showed the surface normal and albedo can be resolved at each pixel by solving a least square equation as explained in Section 2.3.1. If the lighting directions and intensities are unknown, the problem is called *uncalibrated phtometric stereo*. *Auto-calibration* methods solve uncalibrated photometric stereo problem by resolving both surface normals and lighting conditions at the same time. Section 2.3.2 discusses basic auto-calibration method (Hayakawa [1994]) that factorizes the intensity matrix *I* to compute surface normals and lighting conditions from normal matrix N and lighting matrix L.

$$I = \begin{pmatrix} i_{11} & \dots & i_{1F} \\ \vdots & \dots & \vdots \\ i_{Pl} & \dots & i_{PF} \end{pmatrix} = \begin{pmatrix} n_{1x} & n_{1y} & n_{1z} \\ \vdots & \vdots & \vdots \\ n_{Px} & n_{Py} & n_{Pz} \end{pmatrix} \times \begin{pmatrix} l_{1x} & \dots & l_{Fx} \\ l_{1y} & \dots & l_{Fy} \\ l_{1z} & \dots & l_{Fz} \end{pmatrix} = N \times L$$
(1.1)

Recent works of photometric stereo

There are many methods have been developed to relax the three previous assumptions. Recent uncalibrated photometric stereo methods such as Hernández et al. [2008] can obtain complete, detailed reconstructions of textureless shiny objects. This multiview photometric stereo algorithm is also based on the Lambertian reflectance assumption but can tolerate isolated specular highlights. It starts from the silhouettes to recover camera motion and construct the object's visual hull. This visual hull is then used to resolve the lighting directions. In contrast to previous Lambertian photometric stereo methods, Hernández et al. [2008] is not limited to a single viewpoint but produces accurate reconstructions in full 3D. Figure 1.6 (c) shows the recovered shape, the fine details on the body of the buddha are faithfully captured. Figure 1.6 (a) shows the acquisition setup. This kind of setup is quite commonly used in photometric stereo methods. The object is rotated on a turntable in front of a camera and a point light source. A sequence of images is captured, while the light source changes position between consecutive frames.

Besides Lambertian photometric stereo, photometric stereo can also be applied to objects with non-Lambertian materials. Goldman et al. [2005] proposed to recover the surfaces with spatially-varying BRDFs, including surfaces with both varying diffuse



Figure 1.6: Multiview photometric stereo Hernández et al. [2008] (a) data acquisition setup (b) one of the input images (c) recovered shape



Figure 1.7: Shape and Spatially-Varying BRDFs From Photometric Stereo

and specular properties. It is based on the observation that most objects are composed of a small number of fundamental materials. This approach recovers not only the shape but also material BRDFs and weight maps, producing compelling results for a wide range of objects. Figure 1.7 (b) and (f) shows recovered normal and reflectance.

Since photometric stereo is so good at recovering surface details. It can also be used to improve data acquired by other methods. For shape recovered by 3D scanners, the geometry can often be quite noisy as shown in Figure 1.8 (a). Nehab et al. [2005] present an algorithm that combines the 3D scanned shape and normals from photometric stereo and produces a new surface that approximates both. It treats high- and low-frequency components separately as stereo triangulation and photometric stereo have different error-vs.-frequency characteristics. Figure 1.8 (a) shows the optimized surface, which has much lower noise compared with 3D scanned one.

Besides 3D scanners, photometric stereo can also be used as a 2.5D 'scanner' for



Figure 1.8: Normals acquired with photometric stereo can improve 3D scanned shape (a). The resulted shape (b) has lower noise and much real details.

surface texture and shape. Johnson et al. [2011] developed a device consisting of a slab of clear elastomer covered with a reflective skin. When an object presses on the skin, the skin distorts to take on the shape of the object's surface. Thus, the sensor non-destructively changes the reflectance function of the object's surface. A camera records an image of this deformation using illumination from red, green and blue light sources at three different positions from behind(through the elastomer slab). Figure 1.9 (a) and (b) shows the sensor and captured image of the microscopic geometry(in this case, human skin). The photometric stereo presented here follows data-driven approach and build a data base with calibration objects. To reconstruct a region, the algorithm finds the closest matching region in the set of observed data and uses its shape. Figure 1.9 (c) shows the reconstruction of human skin.

Another application of photometric stereo is material classification. Gu and Liu [2012] uses coded illumination to directly measure discriminative features for material classification. Optimal illumination patterns are learned from training samples, after projecting to which, the spectral reflectance of different materials are maximally separated. Figure 1.10 shows the LED-based multi-spectral dome for producing discriminative illumination.



Figure 1.9: The microgeometry capture system consists of an elastometric sensor and a high-magnification camera (a). THe retrographics sensor replaces the BRDF of the subject with its own (b), allowing microscopic geometry to be accurately captured (c)



Figure 1.10: The microgeometry capture system consists of an elastometric sensor and a high-magnification camera (a). The retrographics sensor replaces the BRDF of the subject with its own (b), allowing microscopic geometry to be accurately captured (c)

1.2 Main challenges in photometric stereo

1.2.1 Auto-calibration

As photometric stereo uses lighting information to estimate normal information, most methods require lighting conditions as priori. However, capturing lighting conditions is tedious. One approach is to insert several calibration objects into the scene to determine the lighting direction and intensity. For example, in Figure 1.11 a mirror sphere is inserted into the scene during data capturing. Since the geometry of the calibration sphere is known, the lighting direction can be computed from location of the high-



Figure 1.11: Capturing lighting directions by inserting a metal sphere.

light. Similarly, a diffuse sphere can also be inserted to record the lighting intensity. Although this approach is quite flexible, it has several drawbacks. First, the calibration spheres should be segmented before hand to determine their position and size in the scene. The segmentation can be rather difficult because the sphere is highly reflective and mirrors the surrounding environments. Secondly, calibration objects will project shadow and inter-reflection onto the target object and this will cause errors in photometric stereo algorithms. Finally, due to the limited resolution of the image, the position of the highlight might not be precise. As a result, the estimation objects, another approach is to build a hardware that consists of multiple light sources and calibrate them only once as shown in Figure 1.10. This approach not only requires special hardware whose cost might be quite high, but also is confined itself to laboratory usage.

As we have already seen in Section 2.3.2, Hayakawa [1994] firstly introduced an auto-calibration method for Lambertian photometric stereo.

Hayakawa [1994] showed that surface normals can be recovered up to a general linear transformation if lighting directions are unknown. If one can identify six lights with equal intensity, or six normals with equal albedo, this general linear ambiguity can be reduced to a 3D rotation ambiguity. For example, assume we know six lights \hat{l}_i have equal intensity. Then $\hat{l}_i^{\top}A^{\top}A\hat{l}_i = 1$. Let $A^{\top}A = B$. $\hat{l}_i^{\top}B\hat{l}_i = 1$ is a linear problem. Once *B* is determined, we take the the SVD of *B*. The SVD of a symmetric

matrix is itself symmetric: $B = WDW^{\top}$, where D is diagonal and W is orthonormal. If we let $A = D^{1/2}W^{\top}$, the rest of the ambiguity must be a 3D rotation.

Most of the works in uncalibrated photometric stereo follow the seminal work by Belhumeur et al. [1999] that proved the linear ambiguity can be reduced to a generalized bas-reflief (GBR) ambiguity G by surface integrability:

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix}$$
(1.2)

This GBR ambiguity is fundamental: given any number of images taken from a fixed view-point, neither a computer vision algorithm nor biological process can distinguish two objects that differ by a GBR transformation. Prior knowledge about surface shape, surface albedo, light source direction, or light source intensity must be employed to resolve this ambiguity. Figure 1.12 is an illustration of the GBR ambiguity.

Since then, many works have been proposed to study and resolve this ambiguity. Drbohlav and Chaniler [2005]; Drbohlav and Šára [2002] showed spike-specular reflectance can resolve the GBR ambiguity. Tan and Zickler [2009]; Tan et al. [2007] further proved any homogenous isotropic reflectance can resolve it. The GBR ambiguity can also be resolved by inter-reflections Chandraker et al. [2005] and minimizing the entropy of surface albedos Alldrin et al. [2007].

1.2.1.1 Light source calibration

Almost all of these methods mentioned above rely on various assumptions about scene properties such as integrable surface, non-Lambertian materials, inter-reflections, six



Figure 1.12: Top row shows a laser scanned human head rendered as a Lambertian surface with constant albedo. The subsequent three rows show images of heads whose shapes have been transformed by different generalized bas-relief transformations. The profile views of the face in the third column reveal the nature the individual transformations and the direction of the light source. From the frontal views (first column), the true 3-d structure of the objects cannot be determined.

normals of equal albedo or small albedo entropy. ¹ Hence, these methods can work for certain types of scenes that meet their assumptions, but cannot handle other types. For example, the gift box shown in Figure 1.13 (a) contains a few discrete planes with only three different normal directions and no significant non-Lambertian reflection. Notice that a plane does not provide integrable constraint as a linearly transformed plane is

¹An exception is Hayakawa's work Hayakawa [1994] that used six lights with equal intensity to partially solve the problem.



Figure 1.13: Challenging data for uncalibrated photometric stereo. (a) is too simple and (b) is too complicate for most of existing methods.

also integrable. Hence, all these previous methods will fail on this simple example. Figure 1.13 (b) is another challenging data which contains many depth discontinuities. Methods based on integrability must first identify these discontinuities which is a non-trivial task. Indeed previous uncalibrated photometric stereo algorithms are mainly evaluated with a single segmented smooth curved object. Little work has been proposed to handle challenging data like those shown in Figure 1.13.

In Section 3.1, we propose to study uncalibrated photometric stereo by exploiting constraints in lighting configurations such that our method can be applied to more general data. We consider the case where a scene is illuminated by directional lights located on a view centered cone (a). We show that with at least five lights on such a cone, surface normal directions of a Lambertian scene can be recovered up to two kinds of rotations, and a scaling compounded with a mirror ambiguities. These ambiguities can be resolved if additional constraints are available, such as three lights of equal interval, five lights of equal intensity, surface integrability, non-Lambertian reflectance or corresponding normals from multiple viewpoints. To handle more general data, we choose to combine constraints derived from lighting configurations to achieve an

Euclidian reconstruction. We evaluate two possible approaches and choose the more robust one according to experiments. Our only requirement of the scene is that two corresponding normals can be identified from two views, a constraint which can be easily satisfied for most inputs. We use synthetic and real data to evaluate our algorithm and build a prototype device to demonstrate potential applications.

1.2.1.2 Light source calibration with perspective effect

The vast majority of previous works assume directional lighting and orthographic cameras so that every pixel in the image shares a common directional lighting. However, input images are often captured under nearby point light sources, and this makes the lighting directions vary over the scene points. Such a 'perspective' effect of nearby point light sources causes significant systematic errors in the normal directions under the assumption of a directional light. In practice, it is often difficult to assume a directional lighting due to a distant light source especially when dealing with large scenes or a limited working space.

In Section 3.2, we study photometric stereo under point light sources with intensity fall-off and perspective cameras. We always assume the camera is calibrated and study the photometric stereo problem under both known (calibrated) and unknown (uncalibrated) lighting positions. We begin by showing an inherent shape-light ambiguity that exists in the near-light photometric stereo when the light source positions are unknown. Under unknown *directional* lightings, it is well understood that a surface can only be recovered up to a linear ambiguity Hayakawa [1994]. We propose a patch-based factorization method for the *near-light* configuration, which surprisingly leads to a linear shape-light ambiguity that is similar to the directional lighting case. In other words, nearby point light sources introduce little additional shape ambiguities. In addition,

we propose two methods to resolve this shape-light ambiguity in the near-light setting.

Though the patch-based factorization is useful in analyzing the shape ambiguities, it is difficult to select an appropriate patch size. Hence, we further study the problem under calibrated lighting and propose a graph model based algorithm to estimate the depth and normal direction simultaneously at each pixel. Experiments indicate our method can significantly improve the shape reconstruction compared with methods based on directional lighting.

1.2.2 Non-Lambertian Material

Photometric stereo is all about estimating surface normal from observations under different lighting conditions, so if we assume different surface reflectance model, the surface normal will be different. In previous sections, we have assumed Lambertian reflectance model: brightness of such a surface to an observer is the same regardless of the observer's angle of view. This Lambertian assumption greatly simplifies the photometric stereo algorithms. However, in real world, the reflectance model of objects is barely Lambertian. For example, plastic toys, metal cans and wooden furnitures etc. are all non-Lambertian. These kind of objects can only be correctly reconstructed when their reflectance properties are precisely modelled. Early work such as Horn et al. [1978]; Woodham [1980] made strong assumptions on the reflectance models, typically requiring either explicit knowledge of the BRDF or simple parametric models. Thus these methods are limited to specific scenes.

Subsequent works Barsky and Petrou [2003]; Coleman Jr and Jain [1982]; Ikeuchi [1981]; Nayar et al. [1990] are based on the observation that the reflectance of many materials is well approximated by the sum of specular and a diffuse lobe. They often

assume a Lambertian diffuse lobe, while not imposing a parametric form on the specular lobe. Mallick et al. [2005] assumes the color of the specular lobe differs from the color of the diffuse lobe, so the separation of the specular and diffuse components is allowed.

Modern approaches such as Hertzmann and Seitz [2003, 2005] place reference objects of the same material as the test object and these reference objects provide the measurement of the BRDFs in the scene. By matching the test object's captured reflectance to the reference objects, as shown in Figure 1.14 (c), the corresponding normal can be computed from the reference object with known geometry. This approach can be applied for arbitrary BRDFs, but requires reference objects of the same material as the test object. For spatially varying BRDFs (SVBRDF), the algorithm treats the BRDF at each point on the test object as a linear combination of the basis BRDFs defined by a set of reference objects.

Another approach is built upon reflectance symmetry. These methods exploit symmetries exhibit by many BRDFs to avoid complicated and also imprecise parametric models.

Tan et al. [2007] resolves the GBR ambiguity if the non-Lambertian reflectance function is isotropic and spatially invariant. The key observation is that each point on a curved surface under directional illumination is a member of a family of points that are in isotropic or reciprocal configurations as illustrated in Figure 1.15. The GBR can be resolved in closed form by identifying members of these families in two or more images. This 'isotropic pairs' will be discussed further in Chapter 4. Alldrin et al. [2008] proposed a iterative optimization method based on these isotropic pairs to recover both surface normal and bi-variant BRDFs. Another reflectance symmetry based method Holroyd et al. [2008] works on anisotropic materials but requires up to thousand input


Figure 1.14: Example based photometric stereo matches test object's reflectance measurements to the reference object to find the corresponding surface normal (a). Objects with arbitrary complex reflectance model (b) can be correctly recovered (c).

images at each viewpoint. Tan et al. [2011] and Chandraker et al. Chandraker et al. [2011] both recovered iso-contours of depth and gradient magnitude for isotropic surfaces.

Our method also handles non-Lambertian materials. Based on Tan et al. [2011], we use reflectance symmetry to handle materials with general isotropic reflectance models, which consist of a very broad range of materials.

1.3 Application in appearance capture

Appearance capturing has always been a major goal of computer vision. Among many methods developed over decades, recovering 3D shape and associated material from 2D photographic images is an efficient, cost effective way to produce accurate 3D scans of objects. These image-based methods can compute a representation of the



Figure 1.15: On the visible hemisphere of Gauss sphere, two normals form an *isotropic pair* if they lie at the intersections of two circles centered at source direction s and view direction v. If the BRDF is isotropic the observed intensity at these points will be equal.

3D shape from two to thousands of images taken of the same object under different lighting conditions and allow rendering of the captured object from arbitrary view-points and lighting. However, due to the complexity of real world objects, developing a method that can simultaneously handle a wide range of materials and complex shape is a challenging problem as well as capturing illumination conditions.

In this section, we will first explain three design goals of our system and show how current reconstruction approaches reach their limits in terms of quality and performance when dealing with complex and multi-material scenes. In this context, both calibration simplicity and algorithm generalness are absolute necessities. We will see that massive 3D reconstruction of geometry and reflectance is impractical if the capturing hardware is too sophisticated or the reconstruction method is not powerful enough to handle general scenes. In the end, we will explain how we propose to rely on reflectance symmetry to handle the capturing problem.

1.3.1 Design goals

Simultaneously capturing of geometry and reflectance. Geometry and reflectance capturing has a wide application in industry and research. Debevec et al. [2000] introduced the Light Stage system to efficiently capture how an actor's face appears when lit from every possible lighting direction. Later the improved technique is employed in worldwide hit movies such as Avatar. Including Debevec et al. [2000], many of appearance capturing methods Debevec et al. [2000]; Goesele et al. [2004]; Lensch et al. [2003]; Sato et al. [1997]; Weyrich et al. [2006] require two process steps. The geometry of the object is estimated prior to the reflectometry and data are often captured with two separated sensors. As a result, accurate registration of the images and the 3D model for reflectometry is required for correct reflectance estimation. This registration remains difficult in practice and the quality of reflectance measurements near misalignment is inherently downgraded.

To overcome this limitation, we designed the system to use a single camera capturing images for both geometry and reflectometry. The registration of two types of sensors is completely avoided.

Capable of handling complex shapes and a wide range of materials. Our method is essentially based on photometric stereo. However, our methods differ from previous photometric stereo in the following aspects. Early photometric stereo algorithms jointly estimate local surface orientation (surface normal) with assume parametric reflectance models Georghiades [2003]; Goldman et al. [2005] and they can only handle materials of specific analytic reflectance model. Recent methods Hernández et al. [2008] can produce very accurate reconstructions based on robust Lambertian photometric stereo, but as the object reflectance deviates from Lambertian model, the errors

in the resulting geometry become hard to characterize. Finally, there are few attempts made to generate full 360 degree models from 2.5D height fields captured with these techniques Hernández et al. [2008].

Compared with previous methods, our method does not assume parametric reflectance models and can handle materials of general isotropic reflectance. By exploiting symmetric properties of isotropic reflectance models, our method first estimates normal azimuth angles of each view and then propagate them between multiple-views to reconstruct the complete shape. Even if the reflectance model changes dramatically on the surface, high frequency details can be faithfully recovered. However, our method is not designed for objects with extremely low albedo, or mirror-like materials.

1.4 3D reconstruction methods

Passive methods such as *stereo* and *multi-view stereo* constructs objects base on images taken under static environment light while active methods like *photometric streo* and *active rangefinding* use additional lighting sources to change the lighting condition in the images. We will see both passive and active methods rely on either simplification of materials and shapes, or special equipment and complicate capturing procedures, thus they have different kinds of limitations.

1.4.1 Multi-view stereo

Stereo matching takes of two or more images and recovers a 3D model by finding corresponding pixels in the images and computes 3D depths from 2D positions. For each of the input images, a sparse or dense depth map is assigned. It is a fundamental problem and it is still one of the most active research areas. Over the last few years,

a number of high-quality algorithms have been developed, and the state of the art is improving rapidly.

The greatest challenge to most stereo algorithms comes from establishing correspondence between points in different views. Successful correspondence depends on the reflectance of objects in the scene as well as on the lighting condition, which in real environments can be quite complex. As a result, various assumptions have been made on the photometric properties of the scene. The most common assumption is that the object is *Lambertian*: the energy radiated from a surface point does not depend on the outgoing direction. Under this assumption, the correspondence can be easily established by comparing the irradiance of individual images. This works well when the object is composed mostly of matte surfaces with few specular highlights. Although many works (Jin et al. [2003], Goesele et al. [2007]) have been done to address non-Lambertian reflection, they still cannot produce high-accuracy reconstruction for objects with high-frequency reflectance change.

1.4.2 Active rangefinding

Compared with multi-view stereo methods, *active rangefinding* methods greatly improve the performance of reconstruction systems by actively lighting a scene using highly reliable sensors. A variety of these rangefinding techniques have been used from the earliest days of machine vision Besl [1988]; Curless [1999]; Hebert [2000].

One of the most popular active illumination sensors is *time-of-flight* laser rangefinder. The laser rangefinder finds the distance of a surface by timing the round-trip time of a pulse of light. So the accuracy of a time-of-flight laser scanner depends on the precision of the measurement of the time. Since the laser rangefinder can only detects the distance of one point in its viewing direction, the scanner must change its viewing direction to scan different points. Many systems uses rotating mirrors to change the view direction as mirrors are much lighter than rangefinder and can thus be rotated faster and with greater accuracy.

Although time-of-flight range finders are capable of operating over very long distances, the accuracy of time-of-flight scanner is relatively low due to the high speed of light and difficult of timing the round-trip time.

Triangulation range finder is another type of active illumination laser scanner. One laser dot is emitted and it appears in two different cameras. As discussed in Section 1.4.1, the depth of the dot can be determined by stereo algorithm. Like stereo algorithm, triangulation range finders have a limited range of some meters, but their accuracy is relatively high, which are the opposite of time-of-flight scanners. The accuracy of triangulation range finders is on the order of tens of micrometers.

Structured light 3D scanners project a pattern of light on the object and infer the shape of the object from the deformation of the pattern. The pattern is usually projected by an LCD projector and a camera is mounted along with the projector but it is slightly offset a bit. These methods make use of features with similar appearance over time to find the correspondences in different views.

The advantage of structured light scanners is speed and precision. Compared with laser scanners which scan one point at a time, structured light scanners scan multiple points or the entire field of view at once. This greatly reduces the scanning time and distortion from motion.

Based on a precise 3D reconstruction using active ranger finder, parametric reflectance functions can be fitted at each surface point according to the image observations, as in Lensch et al. [2003]; Sato et al. [1997]. However, these methods require precise registration between images and 3D shapes. Since different sensors are used for shape and reflectance capture, this registration is difficult and often causes artefacts in misaligned regions.

1.5 Objectives

This thesis aims at making a combination of multi-view stereo and photometric stereo to recover shape and reflectance simultaneously with simple hardware requirement. While shape and reflectance capturing has been intensively used in the past, we aim to introduce a reflectance symmetry based approach to greatly simplify data capturing, improve reconstruction accuracy and support a wide range of materials.

We show that with a single camera and a hand-held light source, it is possible to reconstruct complex scene with isotropic SVBRDF. There are three key differences between our method and previous methods. First, we reconstruct a complete 3D shape rather than a single-view normal map. Second, we combine multi-view geometry and photometric cues to avoid fragile iterative optimization of shape and reflectance. Third, our method works with general tri-variant isotropic BRDFs while Alldrin et al. [2008] assumed bi-variant BRDFs to simplify the optimization.

Chapter 2

Photometric stereo

2.1 Basic radiometry

Radiometry provides a set of concepts and mathematical tools to describe light propagation and reflection. The derivation of the photometric stereo algorithms is based on radiometry and will be used throughout the rest of the thesis. In this thesis, we also assume the transfer of radiant energy operates at the *geometric optics* level, where the objects interacting with light are much larger than its wavelength, so effects like dispersion and interference will not be discussed. This leads to a few basic assumptions about the behavior of light throughout the thesis.

Linearity. The combined effect of two inputs to an optical system is always equal to the sum of the effects of each of the inputs individually.

Energy conservation. When light scatters from a surface, the scattering events can never produce more energy than they started with.

No polarization. The polarization of the electromagnetic field is ignored, so the only relevant property of light is its distribution by wavelength.

No fluorescence or phosphorescence. The behavior of light at one wavelength is completely independent of light's behavior at other wavelengths or times.

2.1.1 Basic concepts in radiometry

There are four radiometric concepts that are central to photometric stereo: flux, irradiance/radiant exitance, intensity, and radiance. All of these quantities are generally wavelength dependent, and we will not explicitly state this dependence in the remainder of this chapter.

Flux. *Radiant flux* is the total amount of energy passing through a surface or region of space per unit time. The units are joules/second(J/s).

Irradiance and radiant exitance. Irradiance (E) is the area density of flux arriving at a surface, and radiant exitance (M) is the area density of flux leaving a surface. The units are W/m^2 . For point light source, the amount of energy received falls off with the squared distance from the light.

$$E = \frac{\Phi}{4\pi r^2} \tag{2.1}$$

Lambert's law revisited. We have already seen Lambert's law in Section 2.3.1, which says the amount of light arriving at a surface is proportional to the cosine of the angle between the light direction and the surface normal. Here we will explain the details. Consider a light source with area A and flux Φ that is illuminating a surface. If the lighting direction is parallel to the surface normal (as on the left side of the figure), then the area on the surface receiving light A_1 is equal to A. So for points in A_1 , the irradiance is

$$E_1 = \frac{\Phi}{A}.\tag{2.2}$$



Figure 2.1: Lambert's law. As illumination is over a larger area at smaller incident angles, irradiance (E) arriving at a surface varies according to the cosine of the angle of incidence of illumination.

When the light is at an angle to the surface, the area on the surface reciving light is larger. The area receiving flux, A_2 , is roughly $A/\cos\theta$ and the irradiance of any point in A_2 is

$$E_2 = \frac{\Phi \cos \theta}{A}.$$
 (2.3)

The irradiance at a point is actually defined as

$$E = \frac{\mathrm{d}\Phi}{\mathrm{d}A},\tag{2.4}$$

where the differential flux from the light is computed over the differential area receiving flux.

Solid angle and intensity Solid angle is the extension of two-dimensional angles in a plane to an angle on a sphere. The total area s is the solid angle subtended by the object on a 3D unit sphere (Figure 2.2). The units of solid angles are *steradians* and the entire sphere subtends a solid angle of 4π . If we use the symbol ω to indicate the vectors which are described by set of points on the unit sphere centered at a point p,



Figure 2.2: Solid angle. By projecting c onto the unit sphere, the area of its projection is the solid angle s subtended by the object c in three dimensions.

the intensity is defined as flux density per solid angle:

$$E = \frac{\mathrm{d}\Phi}{\mathrm{d}\omega}.\tag{2.5}$$

Note that intensity is used to describe the distribution of light and is only meaningful for point light sources.

Radiance The final radiometric quantity is *radiance*, *L*. Radiance is defined as the flux density per unit area, per uni solid angle. In terms of flux, it is

$$L = \frac{\mathrm{d}\Phi}{\mathrm{d}\omega \mathrm{A}^{\perp}},\tag{2.6}$$

where A^{\perp} is the projected area of A on a hypothetical surface perpendicular to ω (Figure 2.3).

Of all of these radiometric quantities, radiance is the most fundamental one. If



Figure 2.3: Radiance L is defined as flux per unit solid angle d ω per unit projected area dA^{\perp}

radiance is given, then all of other values can be computed in terms of integrals of radiance over areas and directions. Another property is that radiance remains constant along rays through empty space.

2.2 Surface reflection and BRDF

When light strikes on a surface, the surface scatters the light, reflecting some of it back into the environment. The reflectance of a large class of objects is well modelled by the *bidirectional reflectance distribution function* (BRDF). The BRDF is a fourdimensional function $f_r(\omega_i, \omega_r)$ that defines how light is reflected at an opaque surface. Here ω_i is the negative incoming light direction and ω_r is the outgoing direction. Both ω_i and ω_r are defined with respect to the surface normal n as shown in Figure 2.4. The return value of the function is the ratio of reflected radiance exiting along ω_r to the irradiance incident $E(\omega_i)$ on the surface from direction ω_i . If the direction ω_i is considered as a differential cone of directions, the differential irradiance is

$$dE(\omega_i) = L_i(\omega_i)\cos\theta_i d\omega_i \tag{2.7}$$

The constant of proportionality defines the surface's BRDF for the particular pair of directions ω_i and ω_o :

$$f_r(\omega_o, \omega_i) = \frac{\mathrm{d}L_o(\omega_o)}{\mathrm{d}E(\omega_i)} = \frac{\mathrm{d}L_o(\omega_o)}{L_i(\omega_i)\cos\theta_i\mathrm{d}\omega_i}.$$
(2.8)

Physically based BRDFs obey the following two laws:

1. *Reciprocity:* For every pair of directions ω_i and ω_o , $f_r(\omega_i, \omega_o) = f_r(\omega_o, \omega_i)$.

2. *Energy conservation:* The total energy of light reflected is less than or equal to the energy of incident light.

$$\int_{H^2(\mathbf{n})} f_r(\omega_o, \omega') \cos \theta' \mathrm{d}\omega' \le 1.$$
(2.9)

Now for a surface point p, if the incident radiance \hat{L}_i is given along direction ω_i , the differential outgoing radiance will have the following form:

$$dL_o(\omega_o) = f_r(\omega_o, \omega_i) \hat{L}_i(\omega_i) \cos \theta_i d\omega_i.$$
(2.10)

To get the total radiance along direction ω_o , we can integrate this equation over the upper hemisphere $H^2(\mathbf{n})$ due to the incident illumination from all directions:

$$L_o(\omega_o) = \int_{H^2(\mathbf{n})} f_r(\omega_o, \omega_i) \hat{L}_i(\omega_i) \cos \theta_i \mathrm{d}\omega_i.$$
(2.11)

This reflection equation describes how an incident distribution of light at a point is



Figure 2.4: BRDF model

transformed into an outgoing distribution, based on the reflection properties of the surface.

Surface reflection can generally be split into four broad categories: *diffuse*, *glossy specular*, *perfect specular*, and *retro-reflective* as shown in Figure 2.5. The BRDF usually exhibits a mixture of these four types of reflection. The diffuse surfaces scatter incident light in all directions equally. However, a perfect diffuse surface does not exist in the real world. Dull chalkboards, matte paint and color calibration board (Figure 2.6) are examples of diffuse surfaces. Glossy surfaces scatter light preferentially along the outgoing direction. Examples include plastic or metal that shows blurry reflections of other objects. Perfect specular surfaces like mirror scatter light primarily back along the incident direction. The moon is a good example of retro-reflective surface.

2.2.1 Lambertian reflection

We have discussed the Lambertian reflection model in Section 1.1. Lambertian reflection models a perfect diffuse surface that scatters incident light equally in all directions. Although in the real world this reflection model is rarely seen, it is a good approxima-



Figure 2.5: Reflection from a surface can be generally categorized by the distribution of reflected light: (a) diffuse, (b) glossy specular, (c) perfect specular, and (d) retro-reflective distributions.

tion to surfaces such as matte paint. The BRDF of Lambertian reflection is

$$f_r(\omega_o, \omega_i) = \frac{R}{\pi},\tag{2.12}$$

where R gives the fraction of incident light scattered.

2.2.2 Microfacet models

Many surface reflection models are based on the ideas that rough surfaces can be considered as a collection of small *microfacts*. For these models, a mall patch on the surface is essentially a heightfield, where the distribution of facets is described statistically. When light strikes a small area, a large number of these small facets are illuminated, and their aggregate behavior determines the scattering.

For microfacet models, there are two main components: the distribution of facets and a BRDF that describes how light scatters from individual microfacets. Typically, perfect mirror reflection is assumed for the microfacets, although some models such as



Figure 2.6: Color checker is made of special material which is very close to Lambertian reflectance.

Oren-Nayar model (described in the next section) assume Lambertian BRDF.

Microfacet models consider local lighting effects between neighbor facets. Microfacets may be occluded by neighbor facet, may lie in the shadow of a neighboring microfacet, or reflect more light compared with direct lighting of low-level microfacet BRDF due to interreflection. A common simplication of microfacets structure is that all the microfacets make up symmetric V-shaped grooves.

2.2.2.1 Oren-Nayar diffuse reflection

Oren and Nayar observed rough surfaces generally appear brighter as the illumination direction approaches the viewing direction compared with perfect Lambertian reflection. They assumed that each individual microfacet exhibited perfect Lambertian reflection but describe rough surfaces as a collection of symmetric V-shaped grooves. So the derived BRDF models the aggregate reflection of the collection of grooves:

$$f_r(\omega_i, \omega_o) = \frac{\rho}{\pi} (A + B \max(0, \cos(\phi_i - \phi_o)) \sin \alpha \tan \beta)$$
(2.13)

where

$$A = 1 - \frac{\sigma^2}{2(\sigma^2, +0.33)}$$

$$B = \frac{0.45\sigma^2}{\sigma^2 + 0.09}$$

$$\alpha = \max(\theta_i, \theta_o)$$

$$\beta = \min(\theta_i, \theta_o).$$
(2.14)

2.2.2.2 Torrance-Sparrow model

Torrance and Sparrow [1967] modelled the surface as collections of perfectly smooth mirrored microfacets. The BRDF basically consists of three terms: distribution function $D(\omega_h)$ that gives the probability that a microfacet has orientation ω_h , geometric attenuation term $G(\omega_o, \omega_i)$ that describes the fraction of microfacets that are masked or shadowed, given directions ω_i and ω_o , and the Fresnel term $F_r(\omega_o)$.

The distribution function follows a Gaussian distribution:

$$D(\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\theta^2}{\sigma^2}},$$
(2.15)

where σ is called roughness. Surface is more shiny when σ is smaller.

By assuming that the microfacets are arranged along infinitely long V-shaped grooves, the geometric attenuation term takes into account masking and occlusion between microfacets.

$$sG(\omega_o, \omega_i) = \min\left(1, \min\left(\frac{2(\mathbf{n} \cdot \omega_h)(\mathbf{n} \cdot \omega_o)}{\omega_o \cdot \omega_h}, \frac{2(\mathbf{n} \cdot \omega_h)(\mathbf{n} \cdot \omega_i)}{\omega_o \cdot \omega_h}\right)\right).$$
(2.16)

The final Fresnel Term describe the behaviour of reflection when light moving

between media of differing refractive indices.

$$F(\theta) = F(0) + (1 - F(0))(1 - \cos\theta)^5$$
(2.17)

2.2.3 Measured BRDFs

Previous parametric reflection models are often not flexible enough to model the full complexity of scattering characteristics of real surfaces. Another effective approach is to use measured data about the reflection properties of real surfaces for realistic materials. This measured data can be used to set the parameter values for a parameterized BRDF like the Torrance-Sparrow model. If the the approximation introduced too much error, the accurate measured reflection data can be directly used for rendering and can faithfully re-create the surface's appearance.

Traditional BRDF capturing systems consist of moving camera and light source. As the camera and the light source move in the upper hemisphere of the target material, the incident lighting direction ω_i and outgoing lighting direction ω_o can be changed in the BRDF $f_r(\omega_o, \omega_i)$ and images are taken as samples (Figure 2.7). Although this approach can measure the BRDF accurately, it needs careful control of illumination and environment.

A simplified capturing system uses a spherically homogeneous sample of the material (Figure 2.8). Since a sphere contains normals of all directions, a single image contains many BRDF samples.



Figure 2.7: The BRDF of a target surface is measured by moving a light source and a camera.

2.2.3.1 Isotropic reflectance model

Given a particular category of reflection, the reflectance distribution function may be *isotropic* or *anisotropic*. Most objects are isotropic: if we choose a point on the surface and rotate it around its normal axis at the point, the amount of light reflected doesn't change. In contrast, anisotropic materials reflect different amounts of light as you rotate them in this way. Examples of anisotropic surfaces include brushed metal, phonographic records, and compact disks.

Now consider a table of measured isotropic BRDF samples, where each sample records the BRDF value for a given pair of directions defined by $(\theta_i, \phi_i, \theta_o, \phi_o)$. This is a natural form for measured data to be stored in. One shortcoming of this representation is that the isotropy of the BRDF isn't reflected in the representation. For isotropic BRDFs, rotation about the normal direction leaves the value unchanged. So

$$f_r((\theta_i, \phi_i), (\theta_o, \phi_o)) = f_r((\theta_i, \phi_i + \delta), (\theta_o, \phi_o + \delta))$$
(2.18)



Figure 2.8: Sphere with same material can ease BRDF capturing.

for all δ values. Therefore, a better approach would be to store the samples indexed by the two θ angles and the difference between ϕ directions, $(\theta_i, \theta_o, \Delta \phi)$, where $\Delta \phi = \phi_i - \phi_o$, thus reflecting this structure in the data. This gives a mapping from four directions to a three-tuple:

$$f_r(\theta_i, \phi_i, \theta_o, \phi_o) \to f_r(\theta_i, \theta_o, \Delta \phi).$$
 (2.19)

This representation of isotropic BRDF will be used in later chapters.

2.3 Basics of photometric stereo

2.3.1 Lambertian photometric stereo

Recall that Lambertian photometric stereo has three assumptions. First, it only handles the surface of a *Lambertian reflectance model* that scatters incoming light in all directions. The function value $f_r(\omega_o, \omega_i)$ of Lambertian model is a constant number (Equation (2.2.1)) which represents the albedo of a surface point (Section 2.2.1). The *orthographic camera model* assumes that the outgoing lighting direction ω_o relative to the camera does not change across surface points. Finally, *directional illumination* approximates light sources at infinity, so the lighting direction and intensity are the same at every surface point.

Under the previous three assumptions, the reflection equation at each image point can be formulated as

$$L_{o}(\omega_{o}) = \int_{H^{2}(\mathbf{n})} f_{r}(\omega_{o}, \omega_{i}) L_{i}(\omega_{i}) \cos \theta_{i} d\omega_{i}$$

$$= \rho n s_{i} = I_{i}.$$
 (2.20)

Here for each light source *i*, I_i is the observed pixel intensity. s_i is the lighting direction multiplied by the lighting intensity, which is a 3 × 1 vector and the same at all surface points. Since the lighting only comes from one direction, by using the property of the Dirac delta distribution ($\delta(x)$)

$$\int f(x)\delta(x-x_0)\mathrm{d}x = f(x_0), \qquad (2.21)$$

the integration over the upper hemisphere is simplified to $\rho n s_i$. *n* is the surface normal which is a 3 × 1 normalized vector and ρ represents the albedo of that surface point and may vary across the whole surface. ns_i obeys *Lambert's law*, i.e. the apparent brightness of a Lambertian surface is proportional to the cosine of the angle between the surface normal and the direction of the incident light.

Given three (or more) light sources that are not co-linear as shown in Figure 1.5,

we can write each equation in the matrix form (Equation (2.3.1))

$$I = S\rho n, \tag{2.22}$$

where

$$I = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}, \quad S = \begin{pmatrix} s_1^\top \\ s_2^\top \\ s_3^\top \end{pmatrix}.$$
(2.23)

If the lighting directions and intensities are pre-calibrated, we can solve for the surface normal and albedo by multiply S^{-1} to I. Thus $\rho = \|S^{-1}I\|$, $n = \frac{1}{\rho}S^{-1}I$

2.3.2 Lambertian photometric stereo: factorization approach

In the previous section, we have seen that per-pixel normal and albedo can be estimated by solving a least squares equation under known lighting directions and intensity. However, the lighting information is not always available, e.g, when estimating geometry of buildings under different amounts of sunlight from internet images. This photometric stereo without *priori* knowledge of lighting conditions is called *uncalibrated* photometric stereo problem, and its solution is called *autocalibration* of photometric stereo.

To solve the uncalibrated Lambertian photometric stereo problem, Hayakawa [1994] firstly introduced a factorization based approach. This method first rewrites the equa-

tions Equation (2.3.1) in into a matrix form.

$$I = \begin{pmatrix} i_{11} & \dots & i_{1F} \\ \vdots & \dots & \vdots \\ i_{Pl} & \dots & i_{PF} \end{pmatrix} = \begin{pmatrix} n_{1x} & n_{1y} & n_{1z} \\ \vdots & \vdots & \vdots \\ n_{Px} & n_{Py} & n_{Pz} \end{pmatrix} \times \begin{pmatrix} l_{1x} & \dots & l_{Fx} \\ l_{1y} & \dots & l_{Fy} \\ l_{1z} & \dots & l_{Fz} \end{pmatrix} = N \times L$$
(2.24)

Here, (n_{px}, n_{py}, n_{pz}) indicates the scaled surface normal (unit surface normal multiplied with albedo) at the *p*-th pixel, and (l_{fx}, l_{fy}, l_{fz}) is the scaled lighting direction (unit lighting direction multiplied with its intensity) for the *f*-th image. Only *I* is known and both *N* and *L* are unknown. The rank of both matrix *N*, and *L* are three.

By applying singular value decomposition (SVD), the matrix *I* can be decomposed as:

$$I = UDV^{\top} = (UD^{1/2})(D^{1/2}V^{\top}) = \hat{N}\hat{L}.$$
(2.25)

 \hat{N}, \hat{L} could differ from their true values by an arbitrary 3×3 invertible matrix A since $\hat{N}\hat{L} = \hat{N}A^{-1}A\hat{L}$.

Finally, the autocalibration of photometric stereo amounts to resolve the ambiguity matrix A. Once A is estimated, the true surface normals and lighting directions can be computed as $N = \hat{N}A^{-1}$, $L = A\hat{L}$.

Chapter 3

Auto-calibration

3.1 Ring-light photometric stereo

In the previous chapters, we have seen how auto-calibration photometric stereo algorithm Hayakawa [1994] recovers the surface normals and lighting directions up to a general linear transformation and Belhumeur et al. [1999] proved the linear ambiguity can be reduced to a generalized bas-relief (BGR) ambiguity by surface integrability. Since then, many works have been proposed to study and resolve this ambiguity. Drbohlav and Chantler Drbohlav and Chaniler [2005]; Drbohlav and Šára [2002] showed spike-specular reflectance can resolve the GBR ambiguity. Tan et al. Tan and Zickler [2009]; Tan et al. [2007] further proved any homogenous isotropic reflectance can resolve it. The GBR ambiguity can also be resolved by inter-reflections Chandraker et al. [2005] and minimizing the entropy of surface albedos Alldrin et al. [2007]. All these methods share a common limitation that depth discontinuities must be identified before integrability can be applied to obtain a reconstruction up to the GBR ambiguity. However, this identification of depth discontinuities is nontrivial in practice. Typically, a mask image is provided to separate the object from its background and the whole object surface is assumed to be integrable. This approach cannot handle complicated scenes like the one in Figure 1.13 (b). Furthermore, a piecewise planar scene like Figure 1.13 (a), though consists of integrable patches, does not provide any constraint to reduce the general linear ambiguity to the GBR ambiguity because a plane is always integrable after any linear transformation. Hence, these algorithms often require a pre-segmented and curved smooth surface.

Different from these previous works, in this chapter, we exploit partial information in the lighting conditions to resolve the shape ambiguity. Our method makes little assumption about the scene property. Hence, our method can be applied to more general data which cannot be handled by previous methods. Similar illumination configuration has been used by Alldrin and Kriegman Alldrin et al. [2008]; Alldrin and Kriegman [2007b] where lighting directions are calibrated beforehand, however, Alldrin and Kriegman [2007b] recovers only partial surface geometry and Alldrin et al. [2008] requires more than 100 input images. In comparison, our method requires only five images and our lighting directions are unknown.

Our method is also related to those works that combine photometric stereo and structure-from-motion Higo et al. [2009]; Joshi and Kriegman [2007]; Lim et al. [2005]. These methods assume the surface is differentiable and are difficult to be applied to complicated shapes like Figure 1.13 (b).

3.1.1 Ring-Light photometric stereo

Uncalibrated photometric stereo algorithms typically do not assume any prior knowledge about lighting conditions. In this section, we show that if the illumination is partially known, i.e. directional lights lying on a view centered cone, the problem can be significantly simplified. We first briefly review the shape ambiguity in uncalibrated photometric stereo. Then we show that lights lying on a view centered cone significantly reduce the ambiguity. Finally, we describe several ways to resolve the remaining ambiguities.

3.1.1.1 Uncalibrated photometric stereo

We first briefly review the factorization based formulation of uncalibrated photometric stereo. Suppose F images are captured for a Lambertian surface under a variant directional lighting and each image contains P pixels. Ignoring shadows, inter-reflections and non-Lambertian effects, we can formulate the image intensity matrix I as:

$$I = \begin{pmatrix} i_{11} & \dots & i_{1F} \\ \vdots & \dots & \vdots \\ i_{Pl} & \dots & i_{PF} \end{pmatrix} = \begin{pmatrix} n_{1x} & n_{1y} & n_{1z} \\ \vdots & \vdots & \vdots \\ n_{Px} & n_{Py} & n_{Pz} \end{pmatrix} \times \begin{pmatrix} l_{1x} & \dots & l_{Fx} \\ l_{1y} & \dots & l_{Fy} \\ l_{1z} & \dots & l_{Fz} \end{pmatrix} = N \times L$$
(3.1)

Here, (n_{px}, n_{py}, n_{pz}) indicates the scaled surface normal (unit surface normal multiplied with albedo) at the *p*-th pixel, and (l_{fx}, l_{fy}, l_{fz}) is the scaled lighting direction (unit lighting direction multiplied with its intensity) for the *f*-th image. In uncalibrated photometric stereo, only *I* is known and both *N* and *L* are unknown. The rank of both matrix *N*, and *L* are three. Applying singular value decomposition (SVD), the matrix *I* can be decomposed as:

$$I = UDV^{\top} = (UD^{1/2})(D^{1/2}V^{\top}) = \hat{N}\hat{L}.$$
(3.2)

 \hat{N},\hat{L} could differ from their true values by an arbitrary 3×3 invertible matrix A

since $\hat{N}\hat{L} = \hat{N}A^{-1}A\hat{L}$. The autocalibration of photometric stereo amounts to resolve the ambiguity matrix A. Once A is estimated, the true surface normals and lighting directions can be computed as $N = \hat{N}A^{-1}$, $L = A\hat{L}$.

3.1.1.2 Constraints from a ring-light

Suppose the lights are distributed on a cone centered at the viewing direction as shown in Figure 3.1 (a). We follow the work Tan and Zickler [2009] to analyze the problem in the projective plane where a lighting direction (l_x, l_y, l_z) is considered as a point $(l_x/l_z, l_y/l_z)$. We choose a world coordinate system such that the viewing direction is (0, 0, 1) and corresponds to the origin in the projective plane. In the projective plane, the true lighting directions should lie on a circle centered at origin as shown in Figure 3.1 (b). This circle can be denoted by a diagonal matrix $C = diag(s^2, s^2, -1)$ and $C = S^{\top}C_uS$. Here, $C_u = diag(1, 1, -1)$ is the unit circle and S = diag(s, s, 1)is a uniform scaling matrix. The SVD based reconstruction Equation (3.2) recovers lighting and normal directions up to an arbitrary invertible linear transformation A. The estimated lights form a general conic $\hat{C} = A^{\top}CA$ in the projective plane as shown in Figure 3.1 (c). Hence, we can resolve the ambiguity A by mapping \hat{C} back to C. In this subsection, we first reduce the ambiguity by mapping \hat{C} to the unit circle C_u . The remaining ambiguities are resolved in Section 3.1.1.3.

It is well known Coxeter [1989] that a conic can be computed from five points on it. Hence, we first use five estimated lighting directions to fit the conic \hat{C} which is a 3×3 symmetric matrix. We can apply SVD again to compute a linear transformation *B* that maps \hat{C} to C_u , i.e.

$$\hat{C} = UDU^{\top} = (UD_1^{1/2})C_u(D_1^{1/2}U^{\top}) = B^{\top}C_uB.$$



Figure 3.1: Ring-light photometric stereo. (a) Lighting directions lie on a view centered cone. The term ω denotes the cone opening angle. (b) In the projective plane, these lights lie on a ring centered at origin (i.e. viewing direction). (c) When there is a linear ambiguity, these lights lie on a general planar conic. Our algorithm resolves this linear ambiguity by mapping lights back to their canonic positions.

Here, $D_1^{1/2}C_uD_1^{1/2} = D$. Then the lighting and surface normal directions can be updated accordingly by $\tilde{L} = B\hat{L}$, $\tilde{N} = \hat{N}B^{-1}$. Now, the general linear ambiguity is reduced and the estimated lights \tilde{L} are on a view centered ring in the projective plane. But two kinds of ambiguities remain. First, the scaling matrix S between C and C_u is still unknown. Second, B can only be estimated up to a circle invariant transformation P that maps C_u to C_u . In other words, there could be an ambiguity matrix P such that $B^{\top}C_uB = B^{\top}P^{\top}C_uPB$. The following proposition specifies the structure of P.

Proposition 1: If a 3×3 linear transformation P maps the unit circle C_u to itself, i.e. $P^{\top}C_uP = C_u$, then P can be decomposed as $P = M^n R_{\phi} H_t R_{\theta}$, n = 1 or 2. Here, Mis a mirror transformation about y axis, R_{ϕ} , R_{θ} are rotations in the plane (centered at origin), and H_t is a hyperbolic rotation, i.e.

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_{\theta} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} H_{t} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cosh t & \sinh t \\ 0 & \sinh t & \cosh t \end{pmatrix}.$$
(3.3)



Figure 3.2: Geometric explanations of the components of the 'ring-light ambiguity'. The first row shows the transformations induced to lighting directions in the projective plane. The second row illustrates the corresponding transformations to a 3D shape.

Please refer to the appendix for a proof of this proposition. By this proposition, P is a compounded ambiguity that includes ordinary and hyperbolic rotations and a mirror transformation.

In the next section, we will discuss these ambiguities in more detail and propose methods to resolve them. Here we summarize these ambiguities by the following equation. The general conic \hat{C} can be decomposed as:

$$\hat{C} = B^{\top}C_{u}B = B^{\top}P^{\top}C_{u}PB$$

$$= B^{\top}P^{\top}S^{-\top}CS^{-1}PB = A^{\top}CA$$
(3.4)

Here, B is known, P and S are unknown transformations. Once P, S are determined, we can resolve the general linear ambiguity A. In the following, we refer to the compounded ambiguity $S^{-1}P$ as the *ring-light ambiguity*. It is also called the *ring-light transformation* depending on the context. The auto-calibration of ring-light photometric stereo amounts to estimate this compound transformation to upgrade the reconstruction \tilde{L}, \tilde{N} to Euclidian as: $L = S^{-1}P\tilde{L}, N = \tilde{N}P^{-1}S$.

3.1.1.3 Ring-light ambiguities

We first briefly study each component of the *ring-light ambiguity* and later propose methods to solve it. Figure 3.2 summarizes these components and their geometric implications. The ambiguity S is a scaling in the projective plane which corresponds to the classic bas-relief ambiguity. M flips the normal and lighting directions vertically. It corresponds to the convex-concave ambiguity along the vertical direction. R_{θ} rotates the lighting and normal directions around the origin. It preserves all origin centered circles and could map a continuous shape to a discontinuous one. H_t is a hyperbolic rotation that preserves the unit circle. The relative positions of points on the unit circle are changed after a hyperbolic rotation as shown in Figure 3.2. It could also map continuous shapes to discontinuous ones.

In the following, we show various priors that resolve these ambiguities. We first discuss some widely used priors and later introduce three novel priors.

Integrability: Surface integrability is a widely used scene prior to resolve the ambiguity in uncalibrated photometric stereo. If the scene is known to be integrable, the linear ambiguity A can be reduced to a GBR ambiguity Belhumeur et al. [1999]. The intersection of the GBR transformation group with the ring-light transformation contains only the classic bas-relief transformation. Hence, if applicable, integrability resolves all the other components except the scaling S.

Points with Equal Albedo: Hayakawa Hayakawa [1994] showed that six general normals with the same albedo can reduce the linear ambiguity to a 3D rotation. The intersection of the 3D rotation group with the ring-light transformation contains only the planar rotation R_{ϕ} . Hence, this prior reduces the ring-light ambiguity to a planar rotation. Lights with Equal Intensity Hayakawa's method Hayakawa [1994] can also be applied to six general lights with equal intensity. However, since our lights lie on a view centered cone, constraints derived this way are degenerated. Both S and a 3D rotation cannot be resolved (explained in the next section). Hence, it reduces the ring-light ambiguity to a planar rotation R_{ϕ} compounded with a scaling S.

Lights with Equal Interval If lights are uniformly distributed over the view centered cone, all lighting directions are determined up to a planar rotation (about the cone axis) and a scaling (corresponding to the unknown cone opening angle). Hence this constraint can reduce the linear ambiguity to a planar rotation R_{ϕ} compounded with a scaling *S*.

Multiple Viewpoint: Suppose a surface is observed from two different viewpoints with known relative motion and some corresponding points can be identified among these views. If the surface normals of both views are reconstructed up to some ambiguity, these corresponding points give constraints to resolve these ambiguities. In next section, we show that two corresponding normals from two views can resolve a planar rotation R_{ϕ} and a scaling S in both views.

Clockwise/Counter-Clockwise Lighting: M causes a vertical flipping of the estimated lighting and normal directions. If the lights on the ring are turned on one by one in clockwise or counter-clockwise, M reverses this order. Hence, M can be resolved if the order of lighting is known beforehand.

3.1.2 A complete stratified reconstruction

We choose to combine some of the priors discussed in previous section to achieve a Euclidian reconstruction. We favor priors on lighting configurations to handle more

general scene. We propose two methods to reduce the linear ambiguity to a planar rotation compounded with a scaling. In the next, we employ constraints from two corresponding normals from two views to obtain a Euclidian reconstruction. For this stratified reconstruction, we require observations from at least two viewpoints and five lights of equal interval or equal intensity distributed clockwise (or counterclockwise) on a view centered cone for each viewpoint.

3.1.2.1 Lights with equal interval

Suppose we know the order of lights (clockwise or counterclockwise). All lighting directions are determined up to the unknown cone opening angle and a planar rotation. We can assume arbitrary values of these two parameters to get pseudo lighting directions \check{L} up to a scaling S (corresponding to the cone opening angle) and a planar rotation R_{ϕ} (corresponding to the rotation about the cone axis). We can recover normal directions up to the same ambiguity according to $\check{N} = I\check{L}^{-1}$. However, as we will see in experiments, this approach generates larger errors. Hence, we derive a more sophisticated approach in the following.

3.1.2.2 Lights with equal intensity

We first apply the ring-light constraint described in Section 3.1.1.2 to reconstruct normal directions up to a *ring-light ambiguity*. Then we apply the equal lighting intensity constraint to reduce remaining ambiguities to a mirror transformation M, a planar rotation R_{θ} compounded with a scaling S. Afterwards, we use the known lighting order (clockwise in our experiments) to resolve M.

After applying the ring-light constraint, the estimated lighting direction **l** lies on the unit circle in the projective plane and is related to the true lighting direction **l** by $\mathbf{l} = S^{-1} P \tilde{\mathbf{l}}$. Suppose 5 lights are known to have equal intensity, we obtain

$$k_{1} = \mathbf{l}_{i}^{\top} \mathbf{l}_{i} = \tilde{\mathbf{l}}_{i}^{\top} P^{\top} S^{-\top} S^{-1} P \tilde{\mathbf{l}}_{i}$$
$$= \tilde{\mathbf{l}}_{i}^{\top} R_{\theta}^{\top} H_{t}^{\top} S^{-2} H_{t} R_{\theta} \tilde{\mathbf{l}}_{i} \qquad i = 1, 2, \dots 5.$$
(3.5)

It is easy to verify that M and R_{ϕ} are both eliminated from the equation. Here, k_1 is an unknown constant indicating the lighting intensity and i is an index of the lights. Let $F = R_{\theta}^{\top} H_t^{\top} S^{-2} H_t R_{\theta}$. Then Equation (3.5) is a linear equations about F, i.e. $\tilde{l}_i^{\top} F \tilde{l}_i = k_1$.

Hayakawa Hayakawa [1994] used six such equations from different lighting directions to solve F. However, in our problem there are at most five independent linear equations because of the special configuration of lights. More specifically, $\mathbf{\tilde{l}}' \doteq H_t R_{\theta} \mathbf{\tilde{l}}$ must lie on the unit circle on the projective plane, because $\mathbf{\tilde{l}}$ lie on the unit circle which is invariant under H_t and R_{θ} . Hence, no matter what S = diag(s, s, 1) is the expression, $\mathbf{\tilde{l}}_i^{\top} F \mathbf{\tilde{l}}_i = \mathbf{\tilde{l}}_i'^{\top} S^{-2} \mathbf{\tilde{l}}_i'$ is always a constant. In other words, S cannot be recovered from Equation (3.5) if these lights all lie on a view centered cone. To provide an experimental validation, we uniformly sample 360 lights on the unit circle. The six singular values of all these 360 equations are 17.72, 6.70, 6.70, 0.89, 0.63, 0.00. This suggests one degree of freedom of F cannot be determined.

Hence, we can only solve the 1D null space of F as $k_1F_1 + k_2F_2$. Here, F_1, F_2 satisfy $\tilde{\mathbf{l}}_i^{\top}F_1\tilde{\mathbf{l}}_i = 1$ and $\tilde{\mathbf{l}}_i^{\top}F_2\tilde{\mathbf{l}}_i = 0$ respectively, k_1 is the unknown but fixed constant and k_2 can vary to generate the whole 1D null space. We substitute $F = k_1F_1 + k_2F_2$ into $F = R_{\theta}^{\top}H_t^{\top}S^2H_tR_{\theta}$. We solve s, t, θ, k_1 for any given k_2 according to the formulas provided in Appendix B. It can be verified that the solutions of t and θ are independent of k_2 , while k_1 and s vary according to k_2 . Hence, we obtain a unique



Figure 3.3: The cost as a function of the hypothesized ϕ_1 . This functions has a clear global minimum because ϕ_1, ϕ_2, s_1, s_2 are uniquely determined in principle.

solution of H_t and R_{θ} but cannot determine S, and the original *ring-light ambiguity* is reduced to M, S and R_{ϕ} . The result of this subsection is summarized into the following proposition.

Proposition 2: If five lights with equal intensity can be identified, the *ring-light ambiguity* can be reduced to a mirror transformation, a planar rotation compounded with a scaling.

3.1.2.3 Two corresponding normals in two views

We further exploit the constraints from multiple views. Suppose \mathbf{n}_1 and \mathbf{n}_2 are two corresponding normals in different views. They are defined in their local camera coordinate system and are related by the relative rotation between the two cameras, i.e. $\mathbf{n}_1 = T\mathbf{n}_2$. The relative rotation T can be computed separately, for example, by structure-from-motion. Suppose $\tilde{\mathbf{n}}_1, \tilde{\mathbf{n}}_2$ are the estimated normals which are subject to a planar rotation R_{ϕ} and scaling S. We have the following equations:

$$\mathbf{n}_1 \simeq S_1 R_{-\phi_1} \tilde{\mathbf{n}}_1 \qquad \mathbf{n}_2 \simeq S_2 R_{-\phi_2} \tilde{\mathbf{n}}_2 \qquad \mathbf{n}_1 = T \mathbf{n}_2. \tag{3.6}$$

Here, \simeq means equal up to a scale. Hence,

$$\tilde{\mathbf{n}}_1 \simeq R_{\phi 1} S_1^{-1} T S_2 R_{-\phi 2} \tilde{\mathbf{n}}_2. \tag{3.7}$$

Let $E = R_{\phi 1}S_1^{-1}TS_2R_{-\phi 2}$. We get $\tilde{\mathbf{n}}_1 \simeq E\tilde{\mathbf{n}}_2$. This equation provides two independent constraints. Hence, the four ambiguities $S_1, S_2, R_{\phi 1}, R_{\phi 2}$ can be resolved from two corresponding normals in two views.

Equation (3.7) can be written as $\tilde{\mathbf{n}}_1 \times E\tilde{\mathbf{n}}_2 = 0$, where \times is the vector cross product. This vector equation expands to the following three equations:

$$s_2 \mathcal{A}^{(1)}(\phi_1, \phi_2) + \mathcal{B}^{(1)}(\phi_1) + s_1 s_2 \mathcal{C}^{(1)}(\phi_2) + s_1 \mathcal{D}^{(1)} = 0$$
(3.8)

$$s_2 \mathcal{A}^{(2)}(\phi_1, \phi_2) + \mathcal{B}^{(2)}(\phi_1) + s_1 s_2 \mathcal{C}^{(2)}(\phi_2) + s_1 \mathcal{D}^{(2)} = 0$$
(3.9)

$$s_2 \mathcal{A}^{(3)}(\phi_1, \phi_2) + \mathcal{B}^{(3)}(\phi_1) = 0.$$
 (3.10)

Here, $\mathcal{D}^{(i)}$ are constants and $\mathcal{A}^{(i)}$, $\mathcal{B}^{(i)}$ and $\mathcal{C}^{(i)}$ are polynomials of trigonometrical functions of ϕ_1, ϕ_2 .

$$\mathcal{A}^{(i)}(\phi_1, \phi_2) = a_1^{(i)} \cos\phi_1 \cos\phi_2 + a_2^{(i)} \sin\phi_1 \cos\phi_2 + a_3^{(i)} \cos\phi_1 \sin\phi_2 + a_4^{(i)} \sin\phi_1 \sin\phi_2$$
$$\mathcal{B}^{(i)}(\phi_1) = b_1^{(i)} \cos\phi_1 + b_2^{(i)} \sin\phi_1 \qquad \mathcal{C}^{(i)}(\phi_2) = c_1^{(i)} \cos\phi_2 + c_2^{(i)} \sin\phi_2$$

Here, $a_j^{(i)}, b_j^{(i)}$ and $c_j^{(i)}$ are all constants. These constants are provided in Appendix C.

Given two pairs of corresponding normals, it is nontrivial to derive an analytic solution for s_1, s_2, ϕ_1 and ϕ_2 . We apply a 1D search for ϕ_1 . For each hypothesized value of ϕ_1, ϕ_2 and s_2 can be easily solved from Equation (3.10) of both pairs. Then Equation (3.8) and Equation (3.9) from both pairs give totally 4 results for s_1 . We



Figure 3.4: The first and the second row are the angular errors of reconstructed normal directions by using equal lighting interval and equal lighting intensity constraint respectively. Typically, equal lighting intensity constraint generates more accurate results.

use the consistency of these four values to choose the optimal ϕ_1 and its associated ϕ_2, s_2, s_1 . In principle these four parameters are uniquely determined, so this 1D search has a global minimum and is robust as indicated in Figure 3.3. The result of this subsection is summarized in the following proposition.

Proposition 3: Given partial reconstructions of surface normals up to a planar rotation and a scaling from two views, if two pairs of corresponding normals can be identified, the reconstructions in both views can be upgraded to Euclidian.

3.1.3 Experiments

We first evaluate the two approaches to reduce the general linear ambiguity to a planar rotation and a scaling. We use a metal sphere to record lighting directions and apply calibrated photometric stereo to compute ground truth normal directions as a reference. Then we try both methods on a number of examples each with the same set of input images. The angular error in estimated normal directions of both methods are shown in


Figure 3.5: Results for the challenging data in Figure 1. On the left are results up to the ring-light ambiguity. In the middle is our reconstructed surface normals. For a validation, we calibrate all incident lighting directions with a metal sphere and use calibrated photometric stereo to compute a ground truth in the right. Our result is very consistent to the ground truth.

Figure 3.4. The average angular errors for the three tested scenes are 5.8, 16.4 and 7.5 degrees respectively when the equal lighting interval constraint is used. These errors are reduced to 3.0, 6.0 and 4.4 degrees if the equal lighting intensity is used. It is clear the later generates more accurate results. Hence, in the following, we will focus on the approach based on equal lighting intensity. The flower example has larger error in both methods due to its strong shadowing and inter-reflection.

We apply our method to the challenging data shown in Figure 1.13. As explained earlier, these two examples cannot be handled by previous methods because they are either too simple (too few normals and planar surfaces) or too complicated (too many depth discontinuities). Our method first recover a normal map up to the *ring-light ambiguity* as shown in the left column of Figure 3.5. This result is then upgraded to



Figure 3.6: Additional results. From left to right, they are one of the input images, reconstruction up to the *ring-light ambiguity* and the final Euclidian reconstruction. Some of the artifacts are due to non-Lambertian effects like shadow and highlight.

Euclidian as shown in the middle. The right is a validation computed by calibrated photometric stereo. The difference between our results and the ground truth is small. Some artifacts of the recovered normals on the box surface is due to the inaccuracy in radiometric calibration. Please notice that normals in the background (a black cloth on table to reduce inter-reflection) are very noisy which increase the average angular error by 0.5-1 degrees in general. In practice, to handle shadows and highlights, we use simple intensity thresholding to exclude points with non-Lambertian effects. Our method is applied to Lambertian pixels to calibrate lighting directions.

An additional result is shown in Figure 3.6. From left to right, we show one of the input image, the reconstruction up to the *ring-light ambiguity* and the final Euclidian reconstruction. Some of the artifacts are due to non-Lambertian effects like shadow and highlight which are not modeled in our method.

Next, we use a synthetic scene containing two spheres to evaluate our system under various conditions. Images are synthesized at 840×560 resolution. The images



Figure 3.7: Averaged angular error in the recovered normal directions as a function of the cone opening angle and the angle between two viewpoints. In most of time, the reconstruction error is smaller than 5 degrees.

are contaminated by Gaussian noise with zero mean and standard deviation 0.01 (pixel values are within [0,1]). We synthesize the scene from two viewpoints and at each viewpoint 10 lighting directions are generated on a view centered cone. Zero mean Gaussian noise with standard deviation of 0.5 pixel is added to the true corresponding pixel positions. We evaluate the reconstruction accuracy with respect to different values of the cone opening angle ω and the angle between the two viewing directions. The average angular error in reconstructed normal directions is shown in Figure 3.7. In most of the cases, the reconstruction is quite good with average error smaller than 5 degrees.

3.1.3.1 A prototype device

We manufacture a prototype device for ring-light photometric stereo according to our evaluation on synthetic data. The device is shown in Figure 3.8 (a) and consists 20 LED bulbs that are synchronized with a video camera to capture photometric stereo image sequences. The radius of the plate is 150 millimeters. To facilitate matching, the angle



Figure 3.8: Shown in (a) is a prototype device. 20 LED bulbs lie on a circle with radius of 150 millimeters centered at the viewing direction. Our method allows us to consider the weak perspective effects of the lighting which is critical for a handheld photometric stereo setup operating at relatively small distance. This weak perspective effects is illustrated in (b). To ensure the opening angle of the cone is larger than 15 degrees, the distance between the camera and captured objects should be within 1.2 meters.

between two viewing directions should be less than 20 degrees. Hence, according to Figure 3.7, the operation distance of the device should be less than 1.7 meters (cone opening angle larger than 10 degrees) to ensure reconstruction accuracy. This device is similar to those handheld photometric stereo setups proposed in Higo et al. [2009]; Joshi and Kriegman [2007]; Lim et al. [2005]. The advantage of our method is that our algorithm allows us to consider the weak perspective effects of lighting as illustrated in Figure 3.8 (b). We consider the lighting directions depend on the operation distance, e.g. $\omega_1 \neq \omega_2$. This effect is important for modeling accuracy especially when the operation distance is relatively small. A consequence is that this device cannot be pre-calibrated, because the incident lighting directions changes when the operation distance changes. For example, we pre-calibrate lighting directions for an operation distance of about 0.6 meters and apply it to the operation distance of about 0.8 meters. This incorrect pre-calibration causes average angular error on the box scene as large as 8.5 degrees. Almost three times larger than the 3.0 degrees error when our method is applied.

3.1.4 Conclusion

We have presented a stratified method for ring-light photometric stereo. We have shown that five lights on a view centered cone reduce the general linear ambiguity to two rotations, one mirror reflection compounded with a scaling. If these lights have equal intensity or equal interval, this compound *ring-light ambiguity* can be reduced to a planar rotation plus a scaling. If two corresponding normals from two viewpoints can be identified, Euclidian reconstruction can be obtained. Different from previous works on uncalibrated photometric stereo, we minimize the restriction on scene properties. Hence, our method can be applied to the most general data. We also built a prototype device to demonstrate our method. In Chapter 4 and Chapter 5, we also use this ring-light device to simplify data capturing.

3.2 Near-light Photometric Stereo

The photometric stereo methods we have studied so far assume directional lighting and orthographic cameras so that every pixel in the image shares a common directional lighting. However, real world cameras are of perspective projection model and sometimes they are very close to the object due to space limit. Furthermore, input images are often captured under nearby point light sources and this makes the lighting directions vary over the scene points. While these 'perspective' effects of camera model and nearby point light sources cause significant systematic errors in the normal directions under the assumption of a directional light, these effects can actually help us to recover the shape in another way. In this section, we study photometric stereo under point light sources with intensity fall-off and perspective cameras. We always assume the camera is calibrated and study the photometric stereo problem under both known (calibrated) and unknown (uncalibrated) lighting positions. We begin by showing an inherent shape-light ambiguity that exists in the near-light photometric stereo when the light source positions are unknown. Under unknown *directional* lighting, it is well understood that a surface can only be recovered up to a linear ambiguity Hayakawa [1994]. We propose a patch-based factorization method for the *near-light* configuration, which surprisingly leads to a linear shape-light ambiguity that is similar to the directional lighting case. In other words, nearby point light sources introduce little additional shape ambiguities. In addition, we propose two methods to resolve this shape-light ambiguity in the near-light setting.

3.2.1 Related works

Shape-from-shading (SfS) Horn [1989] studies a similar problem as photometric stereo where only a single input image is available. It was observed on synthetic data Prados and Faugeras [2003]; Tankus et al. [2003] that shape reconstruction can be significantly improved when a perspective camera model (rather than the conventional orthographic camera model) was used. This was further verified by experiments with real data in Tankus et al. [2004]. Later, it was shown Prados and Faugeras [2005] that, by considering the intensity fall-off effects of point light source, shape ambiguity in SfS can be resolved. These works motivate us to study the problem of photometric stereo under point light sources and perspective camera projection.

There have been several approaches that use nearby light sources. It has been pointed out in early works by Iwahori *et al.* Iwahori *et al.* [1992, 1990], Kim and Burger Kim and Burger [1991], and Clark Clark [1992] that the light fall-off effect due to nearby light sources carries information about the scene depth. Unlike traditional photometric stereo with distant lighting, the near-light photometric stereo problem becomes non-linear even with a Lambertian assumption. Clark later extended his method to use nearby distributed illuminants Clark [2006]. The advantage of the light fall-off effect for estimating depth is also used in the context of shape-from-shading Kao and Fuh [1995]; Okatani and Deguchi [1997]; Prados and Faugeras [2005]; Samaras and Metaxas [1999]; Wu et al. [2010].

More recently, Koppal and Narasimhan Koppal and Narasimhan [2007] showed that depth ordering can be achieved even with uncalibrated near-field lightings. Liao *et al.* Liao *et al.* [2007] developed a method that uses a sliding projector along the *z*-direction to measure the depth using the light fall-off effect. By observing the bright-



Figure 3.9: Perspective and light fall-off effects in near-light photometric stereo

ness changes of the scene across different projector locations, their method recovers depth from intensity variations that obey inverse square law.

3.2.1.1 Background

Near-light photometric stereo problem explicitly accounts for the intensity fall-off and perspective effects of the light. Scene irradiance \mathbf{l} is a function of the distance from the light position \mathbf{p} to the scene point \mathbf{o} described as

$$\mathbf{l}(\mathbf{p}, \mathbf{o}) = E \frac{\mathbf{o} - \mathbf{p}}{||\mathbf{o} - \mathbf{p}||^3},$$
(3.11)

where E is a scaling factor. The denominator is the product of a normalization term and the light fall-off term, which obeys the inverse-square law. Unlike traditional farlight setting, the near-light setting makes both the intensity and direction of the scene irradiance **l** vary with the scene position as illustrated in Figure 3.9.

3.2.2 Ambiguity in uncalibrated near-light photometric stereo

Photometric stereo algorithms are referred to as calibrated or uncalibrated when the lighting conditions are known beforehand or not. In this section, we assume the surface is Lambertian, the camera is calibrated, and all light sources have the same intensity (unit intensity) but unknown positions. We propose a patch-based factorization method, which has a global linear ambiguity in the recovered surface normal directions and the lighting positions. This result is surprising given that similar ambiguity exists for the uncalibrated photometric stereo under directional lighting. In other words, the 'perspective effect' introduces little additional shape ambiguities.

3.2.2.1 Patch-based factorization

If P Lambertian points are observed under K different directional lighting conditions, we can formulate the photometric stereo problem as the following matrix factorization:

$$\mathbf{I}_{P\times K} = \mathbf{N}_{P\times 3} \mathbf{L}_{3\times K}.$$

Here, each row of the matrix I is the radiance of a point under different lighting directions, each row of N is a normal direction, and each column of L is a lighting direction. There is a clear shape ambiguity as we can insert any invertible matrix A into this factorization like

$$\mathbf{I} = \mathbf{N}\mathbf{L} = \mathbf{N}\mathbf{A}^{-1}\mathbf{A}\mathbf{L} = \mathbf{N}\mathbf{\hat{L}}.$$
 (3.12)

Here, \hat{N} and \hat{L} are the ambiguous reconstructions of normal and lighting directions respectively.

When the scene is illuminated by point light sources, each point on the surface

has a different lighting direction and intensity, which makes the above factorization formulation invalid. However, we can divide the image into small patches such that the lighting conditions (directions and intensities) are consistent within a patch. We can then apply the factorization patch by patch¹. Suppose we reconstruct a lighting direction matrix $\hat{\mathbf{L}}^{(i)}$ at the *i*-th patch up to an ambiguity matrix $\mathbf{A}^{(i)}$. In the following subsection, we will show that these ambiguities matrices $\mathbf{A}^{(i)}$ are correlated and many of them can be resolved.

3.2.2.2 Correlations of the ambiguities among patches

Let the 3D coordinates of the K point light sources be $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_K$. Suppose the centroid of the *i*-th patch is at the position $\mathbf{o}^{(i)}$. The lighting direction (from the surface to the light source) and intensity at this patch is

$$\mathbf{l}_{k}^{(i)} = \frac{\mathbf{p}_{k} - \mathbf{o}^{(i)}}{||\mathbf{p}_{k} - \mathbf{o}^{(i)}||^{3}}, \qquad 1 \le k \le K,$$

as described in Equation (3.11). The direction and length of $\mathbf{l}_{k}^{(i)}$ indicate the lighting direction and intensity, respectively.

The estimated lighting conditions are related to the true configurations by a linear ambiguity matrix $\mathbf{A}^{(i)}$. Hence,

$$\hat{\mathbf{L}}^{(i)} = \left(\hat{\mathbf{l}}_{1}^{(i)}, \hat{\mathbf{l}}_{2}^{(i)}, \cdots, \hat{\mathbf{l}}_{K}^{(i)}\right) = \mathbf{A}^{(i)} \left(\frac{\mathbf{p}_{1} - \mathbf{o}^{(i)}}{||\mathbf{p}_{1} - \mathbf{o}^{(i)}||^{3}}, \frac{\mathbf{p}_{2} - \mathbf{o}^{(i)}}{||\mathbf{p}_{2} - \mathbf{o}^{(i)}||^{3}}, \cdots, \frac{\mathbf{p}_{K} - \mathbf{o}^{(i)}}{||\mathbf{p}_{K} - \mathbf{o}^{(i)}||^{3}}\right).$$
(3.13)

¹Note that the perspective effect of the camera does not matter here, because the scene radiance to all directions are the same for Lambertian surfaces. However it does affect the integration from normal to surfaces.

For simplicity, we consider the following bi-linear equations:

$$\hat{\mathbf{l}}_{1}^{(i)} \simeq \mathbf{A}^{(i)}(\mathbf{p}_{1} - \mathbf{o}^{(i)}) \Leftrightarrow \hat{\mathbf{l}}_{1}^{(i)} \times \mathbf{A}^{(i)}(\mathbf{p}_{1} - \mathbf{o}^{(i)}) = 0, \qquad (3.14)$$

$$\hat{\mathbf{l}}_{2}^{(i)} \simeq \mathbf{A}^{(i)}(\mathbf{p}_{2} - \mathbf{o}^{(i)}) \Leftrightarrow \hat{\mathbf{l}}_{2}^{(i)} \times \mathbf{A}^{(i)}(\mathbf{p}_{2} - \mathbf{o}^{(i)}) = 0, \qquad (3.15)$$

. . .

$$\hat{\mathbf{l}}_{K}^{(i)} \simeq \mathbf{A}^{(i)}(\mathbf{p}_{K} - \mathbf{o}^{(i)}) \Leftrightarrow \hat{\mathbf{l}}_{K}^{(i)} \times \mathbf{A}^{(i)}(\mathbf{p}_{K} - \mathbf{o}^{(i)}) = 0.$$
(3.16)

Here \simeq means equal up to a scale and \times is the cross product of two vectors. Note that the intensity fall-off effect in the original Equation (3.13) is not modeled in these bilinear equations.

We can do a simple counting for the number of unknowns and the number of equations. For each patch, we obtain K vector equations which correspond to 2K independent scalar equations. If we divide the image into N patches, we will obtain 2NKequations. We have 3K unknowns for the lighting positions $\mathbf{p}_1, \dots, \mathbf{p}_K, 3N$ unknowns for the patch centroids $\mathbf{o}^{(1)}, \dots, \mathbf{o}^{(N)}$ and another 9N unknowns for $\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}$. It seems that we should have enough equations (2NK) to solve all these unknowns (3K + 12N). However, these equations are not all independent. In the following, we first show there is an intrinsic linear ambiguity in these bilinear equations. In the next, we show that if $\mathbf{A}^{(1)}$ and $\mathbf{o}^{(1)}$ are known, then all the other unknowns can be solved.

3.2.2.3 Intrinsic shape-lighting ambiguities

There is a global intrinsic linear ambiguity in the Eqs. (3.14)-(3.16). For any 3×3 invertible matrix **T**, and any 3×1 vector **d**, we have

$$\hat{\mathbf{l}}_1^{(i)} \simeq \mathbf{A}^{(i)}(\mathbf{p}_k - \mathbf{o}^{(i)}) = \mathbf{A}^{(i)}\mathbf{T}^{-1}\mathbf{T}(\mathbf{p}_k + \mathbf{d} - \mathbf{d} - \mathbf{o}^{(i)}).$$

In other words, Eqs. (3.14)-(3.16) are not enough to uniquely determine all the $\mathbf{A}^{(i)}$. The best we can get are $\hat{\mathbf{A}}^{(i)}$, $\hat{\mathbf{p}}_k$ and $\hat{\mathbf{o}}^{(i)}$ which are related to their true values $\mathbf{A}^{(i)}$, \mathbf{p}_k and $\mathbf{o}^{(i)}$ by the following equations,

$$\hat{\mathbf{A}}^{(i)} = \mathbf{A}^{(i)}\mathbf{T}^{-1},$$

 $\hat{\mathbf{p}}_k = \mathbf{T}(\mathbf{p}_k + \mathbf{d}),$
 $\hat{\mathbf{o}}^{(i)} = \mathbf{T}(\mathbf{o}^{(i)} + \mathbf{d})$

3.2.3 Disambiguation methods

3.2.3.1 Solution with one patch calibrated

The last subsection shows there is at least a linear ambiguity in these bilinear equations. Here we show there is no more additional ambiguities. Suppose $\mathbf{A}^{(1)}$ and $\mathbf{o}^{(1)}$ are calibrated, $\mathbf{p}_1, \dots, \mathbf{p}_K$ can be determined from Equation (3.13). Basically, we can solve \mathbf{p}_k from

$$\frac{\mathbf{p}_k - \mathbf{o}^{(1)}}{||\mathbf{p}_k - \mathbf{o}^{(1)}||^3} = (\mathbf{A}^{(1)})^{-1} \hat{\mathbf{l}}_k^{(1)}; \qquad 1 \le k \le K.$$
(3.17)

Given $\mathbf{p}_1, \dots, \mathbf{p}_K$, we can then solve for $\mathbf{A}^{(i)}$ and $\mathbf{o}^{(i)}$ by minimizing the following function:

$$\sum_{k=1}^{K} ||\hat{\mathbf{l}}_{k}^{(i)} \times \mathbf{A}^{(i)}(\mathbf{p}_{k} - \mathbf{o}^{(i)})||^{2}.$$
(3.18)

We minimize this bilinear function iteratively. We first initialize $\mathbf{o}^{(i)}$ to have the same depth as $\mathbf{o}^{(1)}$, and fix it to solve $\mathbf{A}^{(i)}$. We then fix $\mathbf{A}^{(i)}$ to solve $\mathbf{o}^{(i)}$. At each iteration, we only need to solve a linear equation which makes the whole process converges quickly. Note, as the camera is calibrated, there is actually only one unknown in $\mathbf{o}^{(i)}$.

We can set the origin of the coordinate system at $\mathbf{o}^{(1)}$. Hence, this subsection



Figure 3.10: The light sources positions \mathbf{p}_1 , \mathbf{p}_2 can be obtained by intersecting the corresponding lighting directions at two patches $\mathbf{o}^{(1)}$, $\mathbf{o}^{(2)}$. The depth of $\mathbf{o}^{(2)}$ is decided by minimizing the distance between corresponding light rays.

shows that shape can be determined up to a linear ambiguity $\mathbf{A}^{(1)}$ with these bilinear Eqs. (3.14)-(3.16). Note that, it does not exclude the possibility of uniquely determining the shape from the original nonlinear Equation (3.13). A similar linear ambiguity (*i.e.*., Equation (3.12)) is reported in Hayakawa [1994] for photometric stereo under directional lighting. Please note that the ambiguity here has 9 degrees of freedom. Under directional lighting, one only needs to know the ambiguity matrix \mathbf{A} up to a scaling in Equation (3.12) to get the true normal directions. On the other hand, we need to know the exact $\mathbf{A}^{(1)}$ to obtain $\mathbf{p}_1, \dots, \mathbf{p}_K$. This is because our solution for the lighting positions in Equation (3.17) requires the lighting intensities to be calibrated for determining the lighting positions according to the intensity fall offs.

3.2.3.2 Solution with two patches calibrated

Since it is difficult to calibrate the absolute lighting intensity, we also propose a more practical solution here. We might instead calibrate the lighting directions at two patches, *e.g.*, obtain both $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ up to a scaling. This can be done by inserting a metal sphere at these patches to record their lighting directions. We might also apply any existing autocalibration method to recover them. We can then compute the lighting positions $\mathbf{p}_1, \dots, \mathbf{p}_K$ by intersecting the light rays centered at $\mathbf{o}^{(1)}$ and $\mathbf{o}^{(2)}$ as shown in Figure 3.10. Essentially, the lighting positions are solved by minimizing

$$||\hat{\mathbf{l}}_{k}^{(1)} \times \mathbf{A}^{(1)}(\mathbf{p}_{k} - \mathbf{o}^{(1)})||^{2} + ||\hat{\mathbf{l}}_{k}^{(2)} \times \mathbf{A}^{(2)}(\mathbf{p}_{k} - \mathbf{o}^{(2)})||^{2}.$$
 (3.19)

We can still set $\mathbf{o}^{(1)}$ at origin and use a 1D search for the depth of $\mathbf{o}^{(2)}$ to minimize Equation (3.19). Once the lighting positions are decided, we can apply the same iterative method as in the previous subsection to obtain the complete shape.

3.2.4 Calibrated perspective photometric stereo

The previous section uses a patch-based factorization to analyze the intrinsic shape ambiguity in the uncalibrated case, and derives two different methods to calibrate the lighting positions. Nevertheless, patch-based factorization has its limitations. If the patch size is too large, the directional lighting assumption within each patch is less valid. However, when the patch size is too small, many of the patches do not have enough shape changes for factorization (all normals within a patch must span a rank 3 linear space). In the following, we consider the problem under calibrated lighting positions and propose a more sophisticated method for shape reconstruction.

3.2.4.1 Light fall-off depth cue

Since the camera is calibrated, the 3D position \mathbf{x} of a pixel can be determined from its depth *d*. We can represent this position as $\mathbf{x}(d)$ and calculate the lighting direction and intensity at this pixel as

$$\mathbf{l}_k(d) = \frac{\mathbf{p}_k - \mathbf{x}(d)}{||\mathbf{p}_k - \mathbf{x}(d)||^3}, \qquad 1 \le k \le K.$$

As the camera is calibrated, the viewing direction \mathbf{v} is known at each pixel. Now, at each pixel, we obtain a viewing direction \mathbf{v} and a set of lighting directions $\mathbf{l}_k(d)$ for a hypothesized depth d. Suppose the surface reflectance is described by a reflectance model as $I = f(\mathbf{l}, \mathbf{v}, \mathbf{n})$, where I is the scene radiance, $\mathbf{l}, \mathbf{v}, \mathbf{n}$ are the lighting, viewing and normal directions respectively. We can fit a surface normal direction $\mathbf{n}(d)$ as follows,

$$\mathbf{n}(d) = \arg\min_{\mathbf{n}} C_d(\mathbf{n}) = \sum_k ||f(\mathbf{l}_k(d), \mathbf{v}, \mathbf{n}) - I_k||^2.$$
(3.20)

Here I_k is the recorded scene radiance under the k-th lighting direction, and $\mathbf{n}(d)$ is the optimal surface normal direction for the hypothesized depth d.

Note that the fitting residual, *i.e..*, the minimum of $C_d(\cdot)$, measures the validity of the hypothesized depth d. When d is close to the true depth, we should be able to find a normal **n** to reduce $C_d(\cdot)$ to zero. This gives us a strong cue to infer the depth at each pixel. We validate this depth cue on synthetic and real data in Figure 3.11. On the left, we show a synthetic image of a Lambertian surface. The fitting residual is plotted as a function of the hypothesized depth d for several pixels. The ground truth depth is also marked in this plot. It is clear that the fitting residual has a global minimum at the true depth value. On the right, we plot this fitting residual for some points on a real surface.



Figure 3.11: Reconstruction error w.r.t. the hypothesized depth d. Camera is located at d = 0, and light sources are located at around d = 110. The left and right are plots from synthetic and real data respectively.

Though the ground truth depth is unknown, the curve has a clear global minimum. Here we assume the real surface can be modeled by the Lambert's model, though we might use any parametric reflectance model such as the Cook-Torrance model Cook and Torrance [1982] or Oren-Nayar model Oren and Nayar [1994].

3.2.4.2 Depth consistency at neighboring pixels

The novel depth cue enables us to estimate both depth and normal direction at each pixel independently. However, the depth and normal directions of neighboring pixels are highly correlated. Here, we exploit this relationship to facilitate the shape reconstruction. For two neighboring pixels i, j, suppose their depths and associated 3D positions are d_i, d_j and $\mathbf{x}_i, \mathbf{x}_j$ respectively. We can define a tangent vector on the object surface as $\mathbf{x}_i - \mathbf{x}_j$. This tangent vector should be perpendicular to the surface normals at i, j. Hence, the estimated normals $\mathbf{n}_i, \mathbf{n}_j$ at these two pixels must satisfy the following

constraints:

$$\mathbf{n}_i^\top (\mathbf{x}_i - \mathbf{x}_j) = 0,$$

$$\mathbf{n}_j^\top (\mathbf{x}_i - \mathbf{x}_j) = 0.$$

3.2.4.3 Graphical model for depth and normal recovery

We can formulate the depth cue and the depth consistency at neighboring pixels into a graph optimization problem. In this graph model, each vertex represents a pixel where we need to infer a depth d. At each pixel, there is an associated normal $\mathbf{n}(d)$ for every depth value. Every pixel is connected to its four neighbors in the graph. Hence, we can define an objective function of the following form:

$$\sum_{i} D(i, d_i) + \sum_{(i,j) \in \mathbb{N}} S(i, j, d_i, d_j).$$
(3.21)

Here, \mathcal{N} is the set of all neighboring pixels. $D(i, d_i)$ is the data cost that measures the feasibility of assigning the depth d_i to the pixel *i*. We take the normal direction fitting residual as this cost. In other words, $D(i, d_i)$ is the minimum value of $C_{d_i}(\mathbf{n})$ at the pixel *i*. $S(i, j, d_i, d_j)$ is the smooth cost to enforce the depth consistency at neighboring pixels. We simply set it as

$$S(i, j, d_i, d_j) = (\mathbf{n}_i + \mathbf{n}_j)^{\top} (\mathbf{x}_i - \mathbf{x}_j).$$

 \mathbf{n}_i , \mathbf{n}_j are the normal direction at i, j associated with the depth d_i, d_j . This objective function Equation (3.21) can be optimized by the tree-reweighted belief propagation Kolmogorov [2006]; Szeliski et al. [2008]; Wainwright et al. [2005].



Figure 3.12: Experiments on synthetic data. (a) shows one of the input image. (b) is the ground truth normal as a reference. (c) is the normal map computed by patch-based factorization. Each patch is subject to a different (and unknown) linear transform. (d) shows the result from the solution with one patch calibrated (known $\mathbf{A}^{(1)}, \mathbf{o}^{(1)}$). (e) shows the result from the solution with two patches calibrated (known $\mathbf{A}^{(1)}, \mathbf{o}^{(1)}$) and $\mathbf{A}^{(2)}, \mathbf{o}^{(2)}$).

3.2.5 Experiments

We first evaluated our method for uncalibrated photometric stereo. For the method with one patch calibrated, we only validated it on synthetic data as it is difficult to calibrate the lighting intensity. Our results on two synthetic examples are provided in Figure 3.12. One of the input images and the ground truth normal map are provided in Figure 3.12 (a) and (b) respectively. Here, the x, y, z components of a normal direction is linearly scaled to the range [0, 1] for visualization through the RGB channels. (c) is the results of the patch-based factorization, where each patch is up to a different ambiguity matrix $\mathbf{A}^{(i)}$. (d) shows the normals corrected by our estimated $\mathbf{A}^{(i)}$ according to the known $\mathbf{A}^{(1)}, \mathbf{a}^{(2)}$ (up to a scaling) and $\mathbf{o}^{(1)}, \mathbf{o}^{(2)}$ in two patches. Our reconstructed normals visually resemble the ground truth.

For a numeric validation, we also provided a map of angular error for these esti-

mated normal directions in Figure 3.13. The error maps corresponding to our solutions with one patch and two patches calibrated are provided in Figure 3.13 (b) and (c) respectively. The average angular errors were 3.7, 6.4 degrees and 3.9, 8.8 degrees for the 'bunny' and 'dragon' data respectively. Part of this error came from the patch-based approach (note the patch seams in Figure 3.12 (d) and (e)). The results might be improved by making patches overlap. However, overlapping patches generally will not enhance the correlation significantly, because there are often not enough normal samples in the overlapped seam to solve the remaining $\mathbf{A}^{(i)}$ from $\mathbf{A}^{(1)}$. We also compared our method with conventional photometric stereo methods with directional lighting assumption. We chose the lighting directions at the object center as the lighting directions of the whole picture in applying conventional methods. The corresponding error maps are shown in Figure 3.13 (a). The average angular error here were 18.1 and 18.3 degrees for the 'bunny' and 'dragon' data respectively, which were significantly higher than that in our methods.

We also evaluated our method with real data in Figure 3.14. One of the input image is shown in Figure 3.14 (a). (b) is the results of the patch-based factorization, where each patch is up to a different ambiguity matrix $\mathbf{A}^{(i)}$. We applied our disambiguation method with two patches calibrated to generate the results in (c). Once the lighting positions are known, the graph-based method was applied to generate the results shown in (d). This graph-based method does not suffer from the patch seams and produces the best result.



Figure 3.13: Angular errors in the recovered normal map. From left to right, they are the errors of conventional directional lighting photometric stereo, errors in our solution with one patch calibrated and error in our solution with two patches calibrated. The angular error (in degrees) are visualized according to the color bar on the right.



Figure 3.14: Experiments on real data. (a) shows one of the input image. (b) is the normal map generated by patch-based factorization. Each patch is subject to a different (and unknown) linear transform. (c) shows the result from the solution with two patches calibrated (known $\mathbf{A}^{(1)}, \mathbf{o}^{(1)}$ and $\mathbf{A}^{(2)}, \mathbf{o}^{(2)}$). (d) is the result from our graph model.

3.2.6 Conclusion and discussion

We study photometric stereo with near point light sources where each surface point has different lighting direction and intensity. We design a bilinear patch-based factorization method which has a similar linear shape ambiguity as the conventional directional lighting photometric stereo. For practical applications, we extend this method to work with calibration within two patches. To overcome the difficulty in selecting an optimal patch size, we further propose a graph-based method to solve the problem when the lighting positions are calibrated. Our methods are tested on both synthetic and real data and produce good results.

In experiments, we also find some limitations of our methods. The iterative minimization of Equation (3.18) is sensitive to the estimation of lighting positions $\mathbf{p}_1, \dots, \mathbf{p}_K$. The graph-based solution for the calibrated case generate good normal directions. However, the resulted depth looks poor unless the distance between the light and object is kept very small to ensure significant lighting intensity fall-off effects.

Chapter 4

Non-Lambertian photometric stereo

In the previous chapter, we have studied auto-calibration photometric stereo with Lambertian reflectance model. In this section, we will further study reflectance symmetries and non-Lambertian photometric stereo. More specifically, we exploit reflectance symmetries to work on objects with spatially varying *isotropic* BRDF. Isotropic BRDF is very general and sometimes too complex to be approximated by the sum of a specular and a diffuse lobe. As a result, methods like Barsky and Petrou [2003]; Coleman Jr and Jain [1982]; Ikeuchi [1981]; Nayar et al. [1990] fail on general isotropic BRDFs. In this chapter, we design a robust algorithm to identify *iso-depth contours* from photometric stereo images with known lighting conditions. Although these iso-depth contours contain only partial of the information required to recover the surface, they can be combined with multi-view stereo to recover full geometry which will be studied in next chapter.



Figure 4.1: Illustration of an isotropic pair. Symmetric light source vectors are obtained by reflecting s about the plane spanned by the surface normal n and viewing direction v. The observed lighting intensities of s' and s are the same.

4.1 Iso-depth contour estimation

Alldrin and Kriegman Alldrin and Kriegman [2007a] observed that isotropy allows almost trivial estimation of iso-depth contours in the absence of global illumination effects such as shadows and inter-reflections. We propose an algorithm that is more robust in real data than the naïve approach described in Alldrin and Kriegman [2007a]. Specifically, we relax the assumption about lighting and propose a method to enhance robustness to global illumination effects.

Isotropic pair Two light source directions s and s' form an *isotropic pair* if they satisfy $\mathbf{n}^{\top}\mathbf{s} = \mathbf{n}^{\top}\mathbf{s}'$ and $\mathbf{v}^{\top}\mathbf{s} = \mathbf{v}^{\top}\mathbf{s}'$ where n is the normal of a surface patch and n is the viewing direction (Figure 4.1).

The main consequences of isotropic pairs can be summarized as follows.

Fact 1. For any isotropic pair of light sources defined relative to the surface normal and viewing direction, the value of an isotropic BRDF is identical.

Fact 2. For a given surface normal n and viewing direction v, the emitted radiance E(s) from an isotropic material is symmetric about the span of n and v (barring non-

local illumination effects). Consider an isotropic pair of light sources s and s'. From Equation (2.2.3.1), it can be known E(s) = E(s') since $\theta_o = \theta'_o$, $\theta_i = \theta'_i$ and $\Delta \phi = \Delta \phi'$. Since an isotropic pair can be form from any light source by reflecting it about the plane spanned by n and v it follows that the emitted radiance E(s) is symmetric about the plane spanned by n and v.

Fact 3. For a given surface normal n and and viewing direction v, consider the emitted radiance function E(s) resulting from isotropic BRDF ρ . Then in practice, E(s) is only symmetric about the span of n and v.

So based on the three facts above, we could recover the symmetry plane spanned by n and v at each point on the surface by dectecting symmetry in the emitted radiance function E(s) as measured over directional lighting that moves on a view-centered circle. In the camera local coordinate system, where the z-axis is aligned with the viewing direction, this plane gives the azimuth angle ϕ_g of the surface normal, which is the angle between the x-axis and the normal's projection in the xy-plane. We can recover ϕ_g at each pixel as shown in Alldrin and Kriegman [2007a] by minimizing the following objective function,

$$F(\phi_g) = \sum_{i=0}^{N} \min\left\{\eta; \frac{E(\phi_i)}{E(r(\phi_i, \phi_g))} + \frac{E(r(\phi_i, \phi_g))}{E(\phi_i)}\right\},$$
(4.1)

where η is a threshold to account for outliers and $r(\phi_i, \phi_g)$ is a function mapping angle ϕ_i to its reflected position about angle ϕ_g . In our experiments, we use a threshold of $\eta = 2.1$. Figure 4.3 (a) shows the observed pixel intensities under 36 different lighting directions on a view-centered circle. The vertical axis of the chart indicates pixel intensities, while the horizontal axis is the range of azimuth angles. The red symmetry axis of these observations provides a good estimation of the azimuth angle.



Figure 4.2: (a) An input image with cast shadow; (b) Fourier series fitted image;

Once azimuth angles are computed, at each pixel, we can recover an iso-depth contour by tracing along the directions perpendicular to the xy-plane projection of the surface normal there. For easier reference, we refer this direction of a projected surface normal as the azimuth direction in the rest of the thesis.

Figure 4.2 (a) shows one input image of the Buddha dataset. The cast shadow of the belly of the Buddha is removed using our fitting method. The artifacts on the belly of the Buddha are caused by normals facing the camera. In this situation, the emitted radiance curve should be flat and will be affected by noise. However, these artifacts will not increase reconstruction error since the depth in this area is not changing.

Light Source Interpolation In practice, it is more convenient to capture images with a handheld bulb, i.e. a point light source that does not lie precisely on a view-



Figure 4.3: (a) The symmetry axis of intensity profiles tells the azimuth angle of a pixel's normal direction; (b) cast shadows can break this symmetry; (c) the intensity profile of most of isotropic BRDFs in Mcmillan et al. [2003] can be well represented by a 2-order Fourier series.

centered circle. So we compute spatially variant lighting directions at each pixel, and interpolate the desired observations from recorded pixel intensities.

We take the average depth of an object (computed from the reconstructed sparse 3D points in Section 5.3) to estimate an approximate 3D position of each pixel. We also calibrate the 3D positions of the light source (see the experiments section). The lighting directions at each pixel are then computed according to the 3D positions of that pixel and the light sources.

To allow flexible data capture, we interpolate observations under lighting directions lying on a view-centered circle, and compute the azimuth angle from these interpolated observations. We study this interpolation problem in the projective plane where a unit 3D direction (x, y, z) is represented by a 2D point (x/z, y/z). As shown in the left of Figure 4.4, the original lighting directions at a pixel are represented by the red points.



Figure 4.4: We compute a Delaunay triangulation of the original lighting directions (red dots) in the projective plane. The desired observations (blue dots) on a view-centered circle are generated by linear interpolation within these triangles. Left: the circle radius d is the mean distance between the red dots and the viewpoint \mathbf{v} . Right: the circle radius d is set as $(d_i + d_o)/2$. Here, d_i (or d_o) is the largest (or smallest) distance between \mathbf{v} and the red dots on the inner (or outer) conic.

We compute a Delaunay triangulation of these points in the projective plane. The desired observations – those blue dots – on a view-centered circle are generated by linear interpolation within these triangles. The radius d of the blue circle is computed as the mean distance between the red dots and the viewpoint **v**.

Global Illumination Effects To improve accuracy, we need to identify cast shadows, which break the symmetry of pixel intensities. Figure 4.3 (b) shows an example pixel with cast shadow. (This pixel is marked in red in the input image of the 'Buddha' example in Figure 5.7.) The original intensity profile marked by red '×' is asymmetric. Though we might use an intensity threshold to detect shadows, it is hard to identify penumbra this way. Two samples in the penumbra are marked with red ' \otimes ' in Figure 4.3 (b). As shown in Figure 4.3 (b), the azimuth angle estimated by the naïve method in Alldrin and Kriegman [2007a] is far from the ground truth at this point. Points in the penumbra also cause problems in the reflectance estimation in Section 5.4. So we identify them as 'outliers' by fitting a parametric model to the observed intensity profiles. Consider a Lambertian point with surface normal $\mathbf{n} = (n_x, n_y, n_z)$ and albedo ρ . Its intensity should be $\rho r n_x \cos \theta + \rho r n_y \sin \theta - \rho z n_z$ when the lighting direction is $(r \cos \theta, r \sin \theta, -z)$. This motivates us to fit a truncated Fourier series

$$A_0 + \sum_k A_k \cos k\theta + \sum_k B_k \sin k\theta$$

to an intensity profile. We evaluate the fitting error on synthetic data generated according to the MERL BRDF database Mcmillan et al. [2003]. For each BRDF in the database, we uniformly sample ninety normals along a longitude on the visible upper hemisphere, and render them under a light moving on a view-centered circle. Figure 4.3 (c) plots the normalized RMSE (root-mean-square error) of all materials with different orders of Fourier series. For most materials, an intensity profile can be well represented by a second order (i.e. $1 \le k \le 2$) Fourier series with normalized RMSE less than 5%. So we always apply RANSAC to fit a second order Fourier series to each observed intensity profile, and estimate the azimuth angle according to the symmetry of the fitted curve. As shown by the green vertical line in Figure 4.3 (b), our estimated azimuth angle is closer to the ground truth. In fact, this fitting also makes our method less sensitive to specular inter-reflections, which are outliers above the fitted curve.

Tracing Contours Once an azimuth angle is computed at each pixel, we proceed to generate iso-depth contours. Starting from every pixel, we iteratively trace along the two directions perpendicular to the azimuth direction with a step of 0.1 pixel. Specifically, suppose the estimated azimuth angle is θ at a pixel **x**. We trace along the two 2D directions $\mathbf{d}_{+} = (\cos(\theta + \pi/2), \sin(\theta + \pi/2))$ and $\mathbf{d}_{-} = (\cos(\theta - \pi/2), \sin(\theta - \pi/2))$ to $\mathbf{x}_{+} = \mathbf{x} + 0.1\mathbf{d}_{+}$ and $\mathbf{x}_{-} = \mathbf{x} + 0.1\mathbf{d}_{-}$. We then replace \mathbf{d}_{+} and \mathbf{d}_{-} according to the azimuth angles of \mathbf{x}_+ and \mathbf{x}_- respectively and continue to trace. We stop tracing when the maximum number of iterations is reached (500 in our experiments). Pixels on one traced curve should have the same distance to the image plane. To avoid tracing across discontinuous surface points, we use the method described in the 'NPR camera' Raskar et al. [2004] to identify discontinuities. Further, we define a confidence measure for these traced contours as the inverse of the maximum curvature along them. Intuitively, smoother contours with relatively small curvature are more reliable.

4.2 Experiment

We evaluated our algorithm on real data with two hardware setups. Both setups used a PointGrey Grasshopper camera, which captures linear images at 1200×900 resolution. The first setup used a handheld bulb as light source to ensure data capture flexibility. The second one used blinking LED lights synchronized with the video camera to speedup capture.

Handheld System Consisting of just a video camera and a handheld light source, this system is compact and portable. At each viewpoint, we moved a handheld bulb to capture a short video clip (about two minutes), and then uniformly sampled about 100 images with different lighting directions. The light source positions and intensities were recorded with calibration spheres.

Ring-Light System To facilitate data capture, we built a simple device shown in Figure 4.5. 72 LEDs were uniformly distributed on two concentric circles of diameter 400 and 600 millimeters respectively. A video camera was mounted at the center of these circles, facing the direction perpendicular to the board ¹. The camera was syn-

¹The camera was mounted manually. It might not exactly sit on the circle center. Its direction



Figure 4.5: The hardware setup. Left: our device consists of a video camera and two circles of LED lights. Right: we need to calibrate one parameter θ_0 to determine lighting positions.

chronized with the LED lights such that at each video frame, there was only one light turned on.

We pre-calibrated the intensities and positions of these LEDs. Since they are uniformly distributed and the circle radiuses are known, we only need to calibrate one parameter θ_0 to determine their positions. Here, θ_0 is the reference angle of the first LED light as shown in Figure 4.5. To calibrate the rotation angle θ_0 , we first capture the diffuse board at some slanted positions (indicated as blue lines in Figure 4.6) and compute the azimuth angle of the normal direction of the board. The computed angle should be $\theta_0 + \alpha$, where α is the true azimuth angle. α is known, since the 3D position of the board is known from its calibration pattern. Hence, we can obtain θ_0 by subtracting α from the estimated values.

Real-world light sources like LEDs are not ideal point-light sources. Their inten-

might also be slightly off. We ignored these two factors as they introduce little errors according to our experiments.



Figure 4.6: Top view of the calibration setup. We capture a diffuse board at several known positions (black lines) to calibrate camera vignetting and light intensity. Some additional boards (blue lines) are used to calibrate the angle θ_0 .

sities toward different directions are different, and thus the intensity distribution of each LED needs calibration. We calibrated intensity distribution by capturing a diffuse board roughly parallel to the image plane at multiple depths as shown in Figure 4.6. The black lines indicate the positions of this board viewed from top-down. A check board calibration pattern is printed at the four corners of the board, such that its 3D position can be computed with the method described in Zhang [2000]. Assume the board is Lambertian with unit albedo. At each point, the observed pixel intensity should be $I = \mathbf{n}^{\top} \mathbf{l} V$. Here \mathbf{l}, \mathbf{n} are the local lighting and normal directions, and V is the light intensity. Hence, we can capture V at each point on the board as $I/\mathbf{n}^{\top} \mathbf{l}$. We linearly interpolate these captured values to obtain the result in a continuous 3D volume. During our experiments, we always divide an observed pixel intensity by the lighting intensity at its 3D position.

Since we considered LEDs to be point light sources. Hence, at a general surface point, the local lighting directions will form two conics in the projective plane as illustrated on the right of Figure 4.4. When computing azimuth angles, we performed a Delaunay triangulation based interpolation as introduced in Section 4.1. We chose a circle with diameter $d = (d_i + d_o)/2$ to interpolate the required observations. Here, d_i (or d_o) is the largest (or smallest) distance between **v** and the original lighting directions – the red dots – in the inner (or outer) conic.

4.2.1 Errors in iso-depth contours

We evaluated the accuracy of our obtained iso-depth contours with the 'Cat' example under both capture setups. Figure 4.7 shows the average depth error as a function of the contour length. The depth error was measured as the difference between the maximum and minimum 'ground truth' depths along an iso-depth contour. The red and green curves are the result from the handheld and ring-light systems respectively. The handheld system had larger error, since we only calibrated the overall light source intensity there, while we calibrated the light intensity distribution for the ring-light system. The error of the ring-light system will become larger as shown by the cyan curve, if we only calibrate the overall intensity of each LED. Typically, the depth error increased linearly with the length of iso-depth contours. Though shorter contours had less depth fluctuation, longer ones required less iterations of depth propagation to cover the whole surface, and hence, accumulated less error. In our experiments, we always used iso-depth contours of 50-pixel length, which often had about 0.6 and 0.3 millimeter depth changes for the handheld and ring-light system respectively.

4.2.2 Number of images at each viewpoint

We also evaluate the accuracy of the captured shape and BRDF (produced by our algorithm described in the next chapter) with different number of input images from each viewpoint. We only perform this evaluation with the ring-light system for simplicity in



Figure 4.7: Average depth error of all iso-depth contours of the 'Cat' example. This error increases linearly with contour length.



Figure 4.8: Mean shape error of the 'Buddha' example. This error does not change significantly with different number of LEDs.

data capture.

We first evaluated the shape accuracy on the 'Buddha' example shown in Figure 7 of the main paper. Figure 4.8 shows the mean shape reconstruction error (in millimeters) as a function of the number of LEDs in each viewpoint. We always chose equal number of uniformly distributed lights on both the outer and inner circles. Since our Fourier series fitting requires at least 5 LEDs from each viewpoint, we begin the plot from 10 lights (5 on each circle). We found the mean shape error did not change significantly for different number of LEDs.

We also evaluated the reflectance accuracy with different number of LEDs. We





Figure 4.9: Relative RMSE BRDF error of a green paint. This error generally decreases with more LEDs.

first measured the BRDF of a green paint after applying it to a sphere of known shape, and capturing the BRDF from images with calibrated directional lighting. We took this measurement as 'ground truth' and compared our result with it. This experiment is evaluated with another painted figurine shown on the right of Figure 4.9. The left of Figure 4.9 shows the relative root mean square error (RMSE) of our result with different number of LEDs from each viewpoint. We computed the RMSE by

$$E_{rms} = \left(\sum_{\theta_h, \theta_d, \phi} \frac{\left(f(\theta_h, \theta_d, \phi) - \hat{f}(\theta_h, \theta_d, \phi)\right)^2}{f(\theta_h, \theta_d, \phi)^2}\right)^{\frac{1}{2}}$$

Here, $f(\cdot)$, $\hat{f}(\cdot)$ are the 'ground truth' and recovered BRDFs respectively. This error converges to about 8.5% when about 30 LEDs are used. So in our experiments, we always used 30 LEDs for the ring-light system.

4.3 Conclusion

We have presented a robust algorithm to identify iso-depth contours from images under ring-light conditions. Global illumination effects such as cast shadows and specular inter-reflections are treated as outliers. Lighting conditions are pre-calibrated and we have shown that 30 images are sufficient to produce accurate iso-depth contours. In the next chapter, we will combine these iso-depth contours and multi-view stereo to recover complete object shape and spatially variant BRDF.
Chapter 5

Appearance capture by multi-view photometric stereo

Appearance capture methods recover both 3D shape and surface reflectance of objects, allowing photorealistic rendering of the captured objects from arbitrary viewpoints and lighting conditions. This capture is an important and challenging problem with many applications such as graphics and reverse engineering. Typically, appearance capture is performed with sophisticated hardware setups such as the light stage of Ghosh et al. [2009] and the coaxial lights of Holroyd et al. Holroyd et al. [2010]. Though these methods achieve highly accurate results, the data capture setup is expensive and complicated. We design a method with simple setup so that it can be used more widely. As shown in the previous chapter, our simplest setup only contains a digital camera and a handheld moving light source. Compared with Holroyd et al. [2010], our method achieves lower but still useful accuracy (0.3 millimeters vs. 50 microns). This lightweight solution provides a practical step towards enabling casual users the ability to perform appearance capture.

The appearance of opaque objects is well represented by a bi-directional reflectance distribution function (BRDF). Most previous methods on simultaneous shape and BRDF capture, e.g. Goldman et al. [2005]; Hernández et al. [2008], assumed specific parametric BRDF models. Their performance degrades when the real objects have different reflectance from the assumed model.

In the previous chapter, we exploit reflectance symmetries to work on objects with general spatially varying isotropic BRDF to identity 'iso-depth contours', i.e. pixels with the same distance to the image plane, from photometric stereo images. In this chapter, we collect iso-depth contours from multiple viewpoints to reconstruct the complete 3D shape. Specifically, we first apply structure-from-motion Hartley and Zisserman [2003] to reconstruct a sparse set of 3D points. We then propagate the depths of these 3D points along iso-depth contours. Each propagation generates additional 3D points, whose depths can be further propagated. A surprisingly small number of 3D points (about two hundred) can be propagated to reconstruct the complete 3D shape (about two hundred points). Once the shape is fixed, we use the same set of input images to infer the spatially varying reflectance. We assume the BRDF at each surface point is a linear combination of a few basis isotropic BRDFs which are represented by 3D discrete tables to handle general material. The basis BRDFs and mixing weights at each point are iteratively estimated by the ACLS method Lawrence et al. [2006].

5.1 Related work

Image-based modeling. These methods reconstruct a 3D shape and a 'texture map' to model objects from images. Furukawa and Ponce [2010]; Lhuillier and Quan [2005]

are two recent representative methods. Texture color at each surface point is decided according to its image projections. However, a texture map is often insufficient to represent general non-Lambertian materials.

Shape scanning and reflectance fitting. To obtain precise 3D shape, laser scanners and structured-light patterns were used in Levoy et al. [2000]; Rusinkiewicz et al. [2002]; Zhang et al. [2004]. Based on a precise 3D reconstruction, parametric reflectance functions can be fitted at each surface point according to the image observations, as in Lensch et al. [2003]; Sato et al. [1997]. These methods require precise registration between images and 3D shapes. Since different sensors are used for shape and reflectance capture, this registration is difficult and often causes artifacts in misaligned regions. Some methods Aliaga and Xu [2008]; Nehab et al. [2005] combine reflectance recovered from photometric stereo and shape recovered from structured-light, where registration is relatively simple. However, they need to capture images under both structured-light and varying directional light at *each* viewpoint, which is tedious and requires a more complicated setup than ours.

Photometric appearance capture. Our method belongs to photometric approaches that capture both shape and reflectance from the same set of images. Most of previous methods, e.g. Goldman et al. [2005]; Hernández et al. [2008]; Lim et al. [2005], assumed specific parametric BRDF models such as Lambert's or Ward's model Ward [1992]. The performance of these methods degrades when the real objects have different reflectance from the assumed model.

Some other methods employed a sophisticated hardware setup to achieve high quality results. Ma et al. [2007] and Ghosh et al. [2009] used a light stage where the intensity of each LED on the stage was precisely controlled. Holroyd et al. Holroyd et al. [2010] required specialized coaxial lights. This requirement of expensive and complicated hardware limits their wide application. Recently, a few algorithms Alldrin et al. [2008]; Holroyd et al. [2008] were proposed for appearance capture by exploiting various reflectance symmetries that are valid for a broader class of objects. However, Holroyd et al. [2008] required up to a thousand input images at *each* viewpoint and Alldrin et al. [2008] relied on fragile optimization. Tan et al. Tan et al. [2011] and Chandraker et al. Chandraker et al. [2011] both recovered iso-contours of depth and gradient magnitude for isotropic surfaces. Additional user interactions or boundary conditions are required to recover the 3D shape.

The work closest to our method is Alldrin et al. [2008]. Both methods are built upon reflectance symmetry embedded in 'isotropic pairs' introduced in Tan et al. [2007]. There are three key differences between our method and Alldrin et al. [2008]. First, we reconstruct a complete 3D shape rather than a single-view normal map. Second, we combine multi-view geometry and photometric cues to avoid fragile iterative optimization of shape and reflectance. Third, our method works with general tri-variant isotropic BRDFs while Alldrin et al. [2008] assumed bi-variant BRDFs to simplify the optimization.

BRDF acquisition. Our work is also related to BRDF acquisition methods such as Dong et al. [2010]; Ren et al. [2011]. These methods are only applicable to near-flat surfaces where the surface normals are known beforehand. Our method can be considered as a generalization of these methods to non-planar surfaces.

5.2 System pipeline

We provide an block diagram of our system in Figure 5.1. We capture images from multiple viewpoints. At each viewpoint, we capture photometric stereo images with



Figure 5.1: System pipeline. We recover iso-depth contours from photometric stereo images and recover a sparse 3D point cloud by structure-from-motion. In the figure showing iso-depth contours, the gray intensity encodes the estimated azimuth angles, and the colored curves are iso-depth contours. We then propagate the depths of these 3D points along the iso-depth contours to recover the complete 3D shape. Once the shape is fixed, we estimate the spatially varying BRDF from the original input images.

a moving light source, which can be simply a handheld bulb. We design a robust algorithm to identify iso-depth contours from these images (Chapter 4). Further, we apply structure-from-motion Hartley and Zisserman [2003] to images from different viewpoints to reconstruct a sparse set of 3D points. We then derive a complete 3D shape by propagating the depths of these points along the dense iso-depth contours. This initial shape is further refined according to the method described in Nehab et al. [2005]. Once the shape is fixed, we estimate a set of basis isotropic BRDFs and their mixing weights at each surface point by the ACLS method Lawrence et al. [2006] to model the surface reflectance.



Figure 5.2: We propagate the depth of **x** to the iso-depth contour segment C_i that passes through its projection in the *i*-th view. This propagation generates new 3D points, e.g. $\mathbf{y}_1, \mathbf{y}_2$, whose depths in other images can also be propagated along their corresponding iso-depth contours C_{j1}, C_{j2} .

5.3 Shape reconstruction: multi-view depth propagation

A standard structure-from-motion algorithm such as Lhuillier and Quan [2005]; Snavely et al. [2006] can reconstruct a set of sparse 3D points on the object. We capture experiment objects on a turntable with a checkboard pattern to ensure sufficient feature matching for textureless examples. Since structure-from-motion algorithms could be affected by moving highlights, we compute a median image at each viewpoint by taking the median intensity of each pixel and use these images for feature matching. Reconstructed 3D points are combined with the traced iso-depth contours to recover the complete 3D shape.

Depth Propagation As illustrated in Figure 5.2, given a reconstructed 3D point \mathbf{x} , we project it to all images where it is visible. Suppose an iso-depth contour C_i goes through its projection in the *i*-th image. We perform a depth propagation to assign the depth of \mathbf{x} to all pixels on C_i . (If the depth of a pixel on C_i is already known, we keep it

unchanged.) This propagation generates new 3D points, whose depths in other images can also be propagated. We begin with a sparse set of 3D points P reconstructed by structure-from-motion. Depth propagation with P in all images generates a large set of 3D points P'. We then replace P by P' and apply depth propagation iteratively. We keep iterating until P' is empty.

Direct application of the algorithm described above will generate poor results. There are a few important issues which must be addressed for robust 3D reconstruction.

Point Sorting We sort all points in P according to the confidence of their associated iso-depth contours. Note that if a point is visible in K different views, it is repeated K times in P and each repetition is associated with an iso-depth contour in one view. At each iteration, we only select half of the points in P of high confidence for depth propagation. We then remove those selected points, and insert P' into the sorted set P for the next iteration.

Visibility Check We should not propagate the depth of a 3D point in an image where it is invisible. However, the visibility information is missing for 3D points generated by propagation. So we apply a consistency check when propagating the depth of a 3D point \mathbf{x} to a contour C. We check pixels on C one by one, starting from the projection of \mathbf{x} to the two ends of C. If a pixel p fails the check, we truncate C at p, and only assign the depth of \mathbf{x} to pixels on the truncated contour. If the updated contour is too short (less than 5 pixels in our implementation), we do not propagate.

To evaluate consistency at a pixel p, we assign it the depth of \mathbf{x} to determine its 3D position. We then use the surface normal of \mathbf{x} to select L (L = 7 in our implementation) most front parallel views where \mathbf{x} is visible. We assume p is visible in all these L images and check the consistency of the azimuth angles at its projections. The

azimuth angles at corresponding pixels in two different views uniquely decide a 3D normal direction ¹. If different combinations of these L views all lead to consistent 3D normals (the angle between any two normals is within T degrees), we consider p as consistent. Otherwise, we discard the view that is most different from the mean view angle and check consistency with the remaining L - 1 views iteratively. We consider p consistent, if it is consistent over at least 3 views. Otherwise, it is inconsistent. For each consistent 3D point, we set its normal as the mean of all consistent normals. In our implementation, we begin with T = 3, and relax it by 1.3 times whenever P' is empty until T > 15.

We note the number of consistent views for each 3D point when inserting it to the set P'. Points are first sorted by the number of consistent views in descending order. Those with the same number of consistent views are sorted by the confidence of contours.

Shape Optimization After depth propagation, we have a set of 3D points, each with a normal direction estimated. We apply the Poisson surface reconstruction Kazhdan et al. [2006] to these points to obtain a triangulated surface. This surface is further optimized according to Nehab et al. [2005] by fusing the 3D point positions and their normal directions.

Figure 5.3 shows the reconstructed shape at different stages. Shown on the left are 3D points obtained from multi-view stereo. In the middle are the 3D points (with normal directions) obtained by depth propagation. On the right is the result after Poisson surface reconstruction. The final optimized shape is at Figure 5.3 (d). Note the face becomes clearly smoother after optimization.

¹An azimuth angle in one view (with the camera center) decides a plane where the normal must lie in. Intersecting two such planes determines the 3D normal direction.



Figure 5.3: (a) initially reconstructed 3D points; (b) 3D points obtained by depth propagation; (c) initial shape after Poisson surface reconstruction; (d) final result.

5.4 Reflectance capture

We assume the surface reflectance can be represented by a linear combination of several (K=2) basis isotropic BRDFs. Once the 3D shape is reconstructed, we follow Lawrence et al. [2006] to estimate the basis BRDFs and their mixing weights at each point on the surface. We consider the general tri-variant isotropic BRDF, which is a function of θ_h , θ_d , ϕ as shown in Figure 5.4. We discretize θ_h , θ_d and ϕ into 90, 2 and 5 bins respectively all in the interval $[0, \pi/2]$. Please refer to Romeiro and Zickler [2010] for a justification of choosing this interval. Hence, a BRDF is represented as a 900 × 1 vector by concatenating its values at these bins.

We build an $N \times M$ observation matrix **V**, and factorize it into a matrix of mixing weights **W** and a matrix of basis BRDFs **H** as,

$$\mathbf{V}_{N\times M} = \mathbf{W}_{N\times K}\mathbf{H}_{K\times M}.$$

M = 900 is the dimension of a BRDF. N is the number of 3D points. Each row



Figure 5.4: Definition of θ_h , θ_d and ϕ .

of V represents the observed BRDF of a surface point. In constructing the matrix V, we avoid pixels observed from slanted viewing directions (the angle between viewing direction and surface normal is larger than 40 degrees in our implementation), where a small shape reconstruction error can cause a big change in their projected image positions. V contains missing elements because of incomplete observation. We apply the Alternating Constrained Least Squares (ACLS) algorithm Lawrence et al. [2006] to iteratively compute the rows of W and columns of H.

To further improve reflectance capture accuracy, we first compute \mathbf{H} from a subset of precisely reconstructed 3D points, whose reconstructed normals from different combinations of azimuth angles are consistent within 1.5 degrees. We then fix \mathbf{H} and compute \mathbf{W} at all surface points.

Figure 5.5 visualizes the BRDF mixture weights and the basis BRDFs. The red and green channels are the normalized mixture weight of the first and second basis BRDFs. Each basis BRDF is applied to render a sphere under front lighting and viewing directions. Most of our examples consist of a shiny and a less shiny basis BRDFs. This can be seen clearly from the 'Cup' and 'Frog' examples.



Figure 5.5: The normalized BRDF mixture weights are visualized in the different color channels. The corresponding basis BRDFs are used to render a sphere on the right.

5.5 Experiment

We use the same two hardware setups described in Section 4.2: a handheld moving light bulb and a ring-light device. We captured images viewpoint by viewpoint. This process can be speeded up by an automatic turntable. But we used an LP player to simplify the setup. After capturing images at one viewpoint, we manually rotated the LP player to capture the next viewpoint.

In our experiments, the 3D points obtained from the structure-from-motion algorithm were often noisy. We only kept points with reprojection error less than 0.5 pixels. Typically, about 200 initial points were obtained for each example. Our system can also easily incorporate manual intervention in the form of matched feature points to handle textureless regions. To provide a 'ground truth' validation, all experimental objects were scanned using a Rexcan III industrial scanner, which is accurate to 10 microns.



Figure 5.6: Results from the handheld system. (a) one of the input images, (b) the recovered shape rendered with uniform diffuse shading. (c) a rendering with the recovered reflectance model from the same viewpoint and lighting condition as the image in (a). (d) the color-coded shape error (in millimeters) compared to laser-scanned 'ground truth'.

Our results were registered with the scanned shapes using the iterative closest point (ICP) algorithm Besl and McKay [1992].

5.5.1 Handheld system

An example is provided in Figure 5.6. Figure 5.6 (a) shows a sample input image. This example was captured from 10 viewpoints, which allow us to reconstruct part of its surface. To better visualize the recovered shape, we render it with uniform diffuse shading in (b). Most of the geometry details are successfully captured. (c) is a rendering according to the captured reflectance from the same viewpoint and lighting condition as the input image in (a). To provide a quantitative evaluation on shape capture, we visualize the shape reconstruction error (measured in millimeters) in (d). The larger errors at the surface boundary are due to insufficient and slanted observations. Overall, the median (or mean) shape errors are 0.62 and 0.53 (or 0.96 and 0.79) millimeters for the 'Cat' and 'Teapot2'. Here, the object diameters are 100 and 250 millimeters

respectively.

5.5.2 Ring-light system

An example, an polished wooden 'Buddha', is provided in the first row of Figure 5.7. This example has focused and strong highlight. The object diameter is 120 millimeters. We captured it from 41 different viewpoints. This example contains many discontinuities at clothes folds and large concavities at the shoulder. These shape details were faithfully captured, as shown in the rendering in (b) and (c). The median (or mean) shape error was 0.36 (or 0.57) millimeters in this example. Most of the large shape errors appeared at concave carvings with strong inter-reflection.

Another four examples, 'Cup', 'Frog', 'Cat' and 'Teapot' are included in Figure 5.7. Their diameters are 120, 90, 140 and 120 millimeters respectively. Our examples cover a wide range of different material. The rusted metal 'Cup' has quickly change reflectance over its surface. The painted 'Frog' also has significant spatial BRDF changes. We captured 30, 34, 35 and 30 viewpoints for the 'Cup', 'Frog', 'Cat' and 'Teapot' examples respectively. Our method consistently performed well on all of them. Their median (or mean) shape reconstruction error was 0.29, 0.25, 0.24 and 0.24 (or 0.5, 0.47, 0.53 and 0.66) millimeters respectively. The 'Teapot' example had relatively larger error at one side, mainly due to the imprecise structure-from-motion reconstruction caused by erroneous feature matching.

5.5.3 Comparison with existing methods

We compared our results with those obtained from Alldrin et al. [2008]. and Hernández et al. [2008]. We used the same code as the authors. The iterative shape and reflectance



Figure 5.7: Results from the ring-light system. From left to right, these figures are arranged in the same way as Figure 5.6.



Figure 5.8: Results according to Alldrin et al. [2008]. (a) the color coded normal map estimated. (b) the shape computed from the estimated normal according to Wu and Tang [2006]. (c) a rendering under novel illumination. (d) the color coded shape error (in millimeters).



Figure 5.9: Results according to Hernández et al. [2008].

optimization in Alldrin et al. [2008] is complicated and slow. It took over 40 hours to compute the results of *one* viewpoint with 72 input images at resolution of 200×350 . Figure 5.8 shows the results from Alldrin et al. [2008]. (a) is a color coded normal map where the x, y, z components of a normal direction are linearly encoded in the RGB channels, e.g. $(x + 1)/2 \rightarrow R$. Shown in (b) is a surface computed from the recovered normal map according to Wu and Tang [2006]. (c) is a rendering from novel lighting direction according to the estimated normal and reflectance. We can see clear artifacts in all these images. (d) is the color coded shape error (in millimeters). Notice the error range is from 0 to 5. The median (and mean) shape error is 2.38 (and 2.85) millimeters. The median (and mean) angular error of normal directions is 13.1 (and 17.6) degrees. Figure 5.9 shows the results from Hernández et al. [2008] which is designed for Lambertian surfaces, where most of the shape details are smoothed out.

5.6 Re-rendering

To provide an intuitive evaluation of our results, we rendered all examples under novel lighting and viewpoint and compared them with captured photographs in Figure 5.10 and Figure 5.11. Note these images were not used in our shape and reflectance capture system. The first two rows are results from the handheld system, while the others come from the ring-light system. From top to bottom, the median (or mean) intensity differences for each example were 4.8 6.9 8.3, 5.0, 2.6, 5.7 and 6.0 (or 7.3 12.2 12.5, 6.8, 4.0, 10.8 and 9.5) intensity levels with pixel values between [0,255] respectively. Figure 5.12 further shows rendering of the captured object under novel environment lighting.



Figure 5.10: Results from handheld system. Left: reference real photographs. Right: rendering under novel viewpoint and lighting condition.



Figure 5.11: Results from the ring-light system. Left: reference real photographs. Right: rendering under novel viewpoint and lighting condition.



Figure 5.12: Rendering under environment lighting.

5.6.1 Runtime efficiency

Our implementation was not optimized for speed. We did all experiments on a computer with 24GB RAM and a 8-core 3.0GHz CPU. At each viewpoint, our matlab code computed azimuth angles in 1 minute, and traced iso-depth contours in 1.5 minutes. Depth propagation took 16 minutes (for 40 viewpoints), and the final shape optimization took 1 minute. It took about 15 minutes to compute the basis BRDFs from 5,000 samples with ACLS. Our output mesh typically had about two million points with average spatial distance 0.095 millimeters. It took another 45 minutes to compute their BRDF mixing weights. Much of the involved process including azimuth angle computation, iso-depth contour tracing, and BRDF mixing weight computation can be easily parallelized.

5.7 Discussion

We propose a method to capture both shape and reflectance of real objects with spatially variant isotropic reflectance. Our method requires a simple hardware setup and is able to capture 3D shapes accurate to 0.3 millimeters and reflectance with 9% relative RMSE error.

Our method has a few limitations. First, our method cannot model anisotropic material. It also cannot handle translucent objects and mirror surfaces. Second, al-though our method is robust to cast shadows and strong specular inter-reflections with Fourier series fitting, it suffers from diffuse inter-reflections. To resolve this problem, we could replace LEDs by projectors and apply the method in Nayar et al. [2006] to separate inter-reflection. However, it would significantly complicate the hardware setup. Alternatively, we might iteratively estimate the shape and inter-reflection. Last,

our ring-light capture setup contains only two circles of LEDs. Hence, we only capture the BRDF of a point with two different θ_d values. (Note that θ_d is the angle between viewing and lighting directions as shown in Figure 5.4.) Hence, during reflectance capturing, we can only discretize θ_d to two levels, and cannot capture Fresnel effects faithfully. Note this limitation does not apply to the handheld setup. We could increase the number of circles of LED lights, or fit parametric Fresnel terms Schlick [1994] to solve this problem.

Chapter 6

Conclusions

In this thesis, we have proposed several solutions in order to solve difficult photometric stereo problems such as auto-calibration and non-Lambertian surfaces.

For uncalibrated photometric stereo, we have proposed a ring-light configuration to handle general Lambertian scenes. We have shown that five lights on a view centered cone can reduce the general linear ambiguity to two rotations, one mirror reflection compounded with a scaling. If these lights have equal intensity or equal interval, this compound *ring-light ambiguity* can be reduced to a planar rotation plus a scaling. If two corresponding normals from two viewpoints can be identified, Euclidian reconstruction can be obtained.

In real-world, light source is often so close to the target object that it violates directional lighting assumption used in many photometric stereo methods. We also proposed an auto-calibration method which exploits near point light sources and designed a bilinear patch-based factorization method which has a similar linear shape ambiguity as the conventional directional lighting photometric stereo. Our methods are tested on both synthetic and real data and produce good results. For non-Lambertian surfaces, we proposed a method to capture both shape and reflectance of real objects with spatially variant isotropic reflectance. We have demonstrated that reflectance symmetry are very powerful and robust in surface estimation. Combined with existing multi-view stereo algorithm, the reconstruction of full object is accurate to 0.3 millimeters. By assuming the BRDF at each surface point is a linear combination of a few basis isotropic BRDFs, the basis BRDFs and mixing weights can be recovered with 9% relative RMSE error. The captured shape and material can be easily used in many applications such as photo-realistic rendering.

Future work The main challenge of our appearance capturing algorithm is the mix of photometric stereo and multi-view stereo. Although reflectance symmetries recovers accurate surface azimuth angles robustly, to recover the full surface normal it still lacks information which can only be obtained from multi-view stereo. As a result, *correspondence error* is introduced into photometric stereo which previously does not have. Our method does not explicitly address this problem. We believe future algorithms can take point correspondences into consideration and optimize them together with reflectance symmetry.

References

- Daniel G Aliaga and Yi Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
 90
- Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with nonparametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008. xviii, 15, 22, 40, 91, 100, 102, 103
- Neil G Alldrin and D Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Computer Vision*, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007a. 75, 76, 79
- Neil G Alldrin and David J Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Computer Vision,* 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007b. 40

Neil G Alldrin, Satya P Mallick, and David J Kriegman. Resolving the generalized

bas-relief ambiguity by entropy minimization. In *CVPR*, volume 1, page 5. Citeseer, 2007. 10, 39

- Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1239–1252, 2003. 14, 74
- Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999. 10, 39, 45
- Paul J Besl. Active, optical range imaging sensors. *Machine vision and applications*, 1(2):127–152, 1988. 20
- Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. pages 586–606, 1992. 99
- Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. A theory of differential photometric stereo for unknown isotropic brdfs. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2505–2512. IEEE, 2011.
 16, 91
- Manmohan Krishna 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. 10, 39
- James J Clark. Active photometric stereo. In Computer Vision and Pattern Recognition,

1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on, pages 29–34. IEEE, 1992. 58

- James J Clark. Photometric stereo with nearby planar distributed illuminants. In *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*, pages 16–16. IEEE, 2006. 58
- E Coleman Jr 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. 14, 74
- Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24, 1982. 67
- Harold Scott Macdonald Coxeter. Introduction to Geometry. Wiley Classics Library. Wiley, 1989. ISBN 9780471504580. URL http://books.google.com.sg/books?id=N8i1QgAACAAJ. 42
- Brian Curless. From range scans to 3d models. *ACM SIGGRAPH Computer Graphics*, 33(4):38–41, 1999. 20
- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the* 27th annual conference on Computer graphics and interactive techniques, pages 145–156. ACM Press/Addison-Wesley Publishing Co., 2000. 18
- Yue Dong, Jiaping Wang, Xin Tong, John Snyder, Yanxiang Lan, Moshe Ben-Ezra, and Baining Guo. Manifold bootstrapping for svbrdf capture. ACM Transactions on Graphics (TOG), 29(4):98, 2010. 91

- 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. 10, 39
- Ondřej Drbohlav and Radim Šára. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Computer Vision, ECCV 2002*, pages 46–60. Springer, 2002. 10, 39
- Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(8): 1362–1376, 2010. 89
- 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. 18
- Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Proceedings of the Twentieth Eurographics conference on Rendering*, pages 1161–1170. Eurographics Association, 2009. 88, 90
- Michael Goesele, Hendrik Lensch, Jochen Lang, Christian Fuchs, and Hans-Peter Seidel. Disco: acquisition of translucent objects. In ACM Transactions on Graphics (TOG), volume 23, pages 835–844. ACM, 2004. 18
- Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M Seitz.
 Multi-view stereo for community photo collections. In *Computer Vision*, 2007. *ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 20

- DB Goldman, B Curless, A Hertzmann, and SM Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 341–348. IEEE, 2005. 5, 18, 89, 90
- Jinwei Gu and Chao Liu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 797–804. IEEE, 2012. 7
- Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 89, 92
- Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11(11):3079–3089, 1994. 4, 9, 11, 13, 37, 39, 45, 46, 48, 57, 64
- Martial Hebert. Active and passive range sensing for robotics. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 102–110. IEEE, 2000. 20
- Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3): 548–554, 2008. x, xviii, 5, 6, 18, 19, 89, 90, 100, 102, 103
- Aaron Hertzmann and Steven M Seitz. Shape and materials by example: A photometric stereo approach. In 2013 IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 533–533. IEEE Computer Society, 2003. 15

Aaron Hertzmann and Steven M Seitz. Example-based photometric stereo: Shape re-

construction with general, varying brdfs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1254–1264, 2005. 15

- Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1234–1241. IEEE, 2009. 40, 55
- Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler. A photometric approach for estimating normals and tangents. In *ACM Transactions on Graphics* (*TOG*), volume 27, page 133. ACM, 2008. 15, 91
- Michael Holroyd, Jason Lawrence, and Todd Zickler. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. *ACM Transactions on Graphics (TOG)*, 29(4):99, 2010. 88, 90
- Berthold K. P. Horn. Obtaining shape from shading information. pages 123–171, 1989. 58
- Berthold KP Horn, Robert J Woodham, and M Silverwilliam. Determining shape and reflectance using multiple images. 1978. 14
- Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):661–669, 1981. 14, 74
- Y Iwahori, RJ Woodham, and N Ishii. Shape from shading with a nearby moving point light source. In *Proceedings of the 2nd International Conference on Automation, Robotics and Computer Vision, Singapore*, 1992. 58

- Yuji Iwahori, Hidezumi Sugie, and Naohiro Ishii. Reconstructing shape from shading images under point light source illumination. In *Pattern Recognition, 1990. Proceedings., 10th International Conference on*, volume 1, pages 83–87. IEEE, 1990.
 58
- Hailin Jin, Stefano Soatto, and Anthony J Yezzi. Multi-view stereo beyond lambert. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–171. IEEE, 2003. 20
- 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. 7
- Neel Joshi and David J Kriegman. Shape from varying illumination and viewpoint. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–7. IEEE, 2007. 40, 55
- Sheng-Liang Kao and Chiou-Shann Fuh. Shape from shading using near point light sources. In *Image Analysis Applications and Computer Graphics*, pages 487–488. Springer, 1995. 58
- Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics Symposium on Geometry processing*, 2006. 95
- Byungil Kim and Peter Burger. Depth and shape from shading using the photometric stereo method. volume 54, pages 416–427. Elsevier, 1991. 58

- Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28 (10):1568–1583, 2006. 68
- Sanjeev J Koppal and Srinivasa G Narasimhan. Novel depth cues from uncalibrated near-field lighting. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 58
- Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. 25(3):735–745, 2006. 89, 92, 96, 97
- Hendrik Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. ACM Transactions on Graphics (TOG), 22(2):234–257, 2003. 18, 21, 90
- Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the* 27th annual conference on Computer graphics and interactive techniques, pages 131–144. ACM Press/Addison-Wesley Publishing Co., 2000. 90
- Maxime Lhuillier and Long Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):418–433, 2005. 89, 93
- Miao Liao, Liang Wang, Ruigang Yang, and Minglun Gong. Light fall-off stereo. In

Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1–8. IEEE, 2007. 58

- Jongwoo Lim, Jeffrey Ho, Ming-Hsuan Yang, and David Kriegman. Passive photometric stereo from motion. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1635–1642. IEEE, 2005. 40, 55, 90
- Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proceedings of the 18th Eurographics conference* on Rendering Techniques, pages 183–194. Eurographics Association, 2007. 90
- Satya P Mallick, Todd E Zickler, David Kriegman, and Peter N Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 619–626. Ieee, 2005. 15
- Leonard Mcmillan, Arthur C Smith, Wojciech Matusik, and Wojciech Matusik. A data-driven reflectance model. 2003. xvi, 78, 80
- Shree K Nayar, Katsushi Ikeuchi, and Takeo Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *Robotics and Automation, IEEE Transactions on*, 6(4):418–431, 1990. 14, 74
- Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. 25(3):935–944, 2006. 106
- Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently

combining positions and normals for precise 3d geometry. 24(3):536–543, 2005. 6, 90, 92, 95

- Takayuki Okatani and Koichiro Deguchi. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer vision and image understanding*, 66(2):119–131, 1997. 58
- Michael Oren and Shree K Nayar. Generalization of lambert's reflectance model. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 239–246. ACM, 1994. 67
- Emmanuel Prados and Olivier Faugeras. "perspective shape from shading" and viscosity solutions. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 826–831. IEEE, 2003. 58
- Emmanuel Prados and Olivier Faugeras. Shape from shading: a well-posed problem?
 In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 870–877. IEEE, 2005. 58
- Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. Nonphotorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. 23(3):679–688, 2004. 81
- Peiran Ren, Jiaping Wang, John Snyder, Xin Tong, and Baining Guo. Pocket reflectometry. 30(4):45, 2011. 91
- F Romeiro and T Zickler. Inferring reflectance under real-world illumination. 2010. 96

- Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. Real-time 3d model acquisition. 21(3):438–446, 2002. 90
- Dimitris Samaras and Dimitris Metaxas. Coupled lighting direction and shape estimation from single images. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 868–874. IEEE, 1999. 58
- Yoichi Sato, Mark D Wheeler, and Katsushi Ikeuchi. Object shape and reflectance modeling from observation. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 379–387. ACM Press/Addison-Wesley Publishing Co., 1997. 18, 21, 90
- Christophe Schlick. An inexpensive brdf model for physically-based rendering. 13(3): 233–246, 1994. 107
- Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. *ACM transactions on graphics (TOG)*, 25(3):835–846, 2006. 93
- Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothnessbased priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30 (6):1068–1080, 2008. 68
- Ping Tan and Todd Zickler. A projective framework for radiometric image analysis. In *Computer Vision and Pattern Recognition*, 2009. CVPR 2009. IEEE Conference on, pages 2977–2984. IEEE, 2009. 10, 39, 42
- Ping Tan, Satya P Mallick, Long Quan, David Kriegman, and Todd Zickler. Isotropy,

reciprocity and the generalized bas-relief ambiguity. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 10, 15, 39, 91

- Ping Tan, Long Quan, and Todd Zickler. The geometry of reflectance symmetries. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(12):2506– 2520, 2011. 16, 91
- Ariel Tankus, Nir Sochen, and Yehezkel Yeshurun. A new perspective [on] shapefrom-shading. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pages 862–869. IEEE, 2003. 58
- Ariel Tankus, Nir Sochen, and Yehezkel Yeshurun. Perspective shape-from-shading by fast marching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–43. IEEE Computer Society; 1999, 2004.
 58
- Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *JOSA*, 57(9):1105–1112, 1967. 32
- Martin J Wainwright, Tommi S Jaakkola, and Alan S Willsky. Map estimation via agreement on trees: message-passing and linear programming. *Information Theory, IEEE Transactions on*, 51(11):3697–3717, 2005. 68
- Gregory J Ward. Measuring and modeling anisotropic reflection. volume 26, pages 265–272. ACM, 1992. 90
- Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, et al.

Analysis of human faces using a measurement-based skin reflectance model. 25(3): 1013–1024, 2006. 18

- Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980. 1, 4, 14
- Chenyu Wu, Srinivasa G Narasimhan, and Branislav Jaramaz. A multi-image shapefrom-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision*, 86(2-3):211–228, 2010. 58
- Tai-Pang Wu and Chi-Keung Tang. Visible surface reconstruction from normals with discontinuity consideration. In *Computer Vision and Pattern Recognition*, 2006 *IEEE Computer Society Conference on*, volume 2, pages 1793–1800. IEEE, 2006. xviii, 102, 103
- Li Zhang, Noah Snavely, Brian Curless, and Steven M. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23:548–558, 2004. 90
- Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis* and Machine Intelligence, IEEE Transactions on, 22(11):1330–1334, 2000. 83

Appendix A: Proof of Proposition 1

Proposition 1: If a 3×3 linear transformation P maps the unit circle C_u to itself, i.e. $P^{\top}C_uP = C_u$, then P can be decomposed as $P = M^n R_{\phi} H_t R_{\theta}$, n = 1 or 2.

Proof: Our proof is based on the following two lemmas:

Lemma 1: If a conic C is mapped to another conic C' by a projective transformation P, then P maps the interior/exterior of C to the interior/exterior of C'.

Lemma 2: Suppose A and A' are two points on two different conics C and C'. B, B' lies inside of C, C' respectively. Then there are precisely two projective transformations which map C to C', A to A', and B to B'.

In the following, for a general linear transformation P that maps C_u to C_u , we assume the pre-images of (1, 0, 1) and (0, 0, 1) are A and B respectively. We explicitly derive two transformations $P_1, P_2, P_1 \neq P_2$, with the form $M^n R_{\phi} H_t R_{\theta}$ that maps A, B to (1, 0, 1) and (0, 0, 1) respectively. Then according to Lemma 2, we know Proposition 1 is true.

According to the Lemma 1, B is a point within C_u . So we can denote B as $(rcos\theta, rsin\theta, 1)$, where 0 < r < 1. It is easy to verify that $H_t R_{\pi/2-\theta}$ maps the point B to the origin. Here, t is uniquely decided by r = -sinh(t)/cosh(t). It is also easy to verify that $H_t R_{\pi/2-\theta}$ maps A to another point A' on the circle. We can denote A' as $(cos\phi, sin\phi, 1)$. Then a rotation $R_{-\phi}$ will maps A' to the point
(1, 0, 1) and keep the origin invariant. As a result, we get the following transformation $P_1 = R_{-\phi}H_tR_{\pi/2-\theta}$, that maps B to (0, 0, 1) and A to (1, 0, 1). Note that, we can define $P_2 = MR_{-\phi}H_tR_{\pi/2-\theta}$. P_2 should also maps B to origin and A to (1, 0, 1). Further, $P_1 \neq P_2$. Hence, according to Lemma 2, they are the only two transformations that map A, B to (1, 0, 1) and (0, 0, 1) respectively.

Appendix B: Determine t, s from F

 $\begin{aligned} \theta \text{ can be directly computed from } F, \quad \theta &= \arctan\left(-F_{13}/F_{23}\right) \\ k_1 \text{can be solved from equation } (a^2 - b^2 - c^2)k_1^2 - (a + 3c)k_1 - 2 = 0 \\ \text{where } a &= \frac{1}{2}\left(F_{11} + F_{22}\right) + \frac{3}{2}F_{33} \quad b = \frac{1}{2}\left(F_{11} + F_{22} - F_{33}\right) \quad c = \frac{2F_{23}}{\cos\theta} = -\frac{2F_{13}}{\sin\theta} \\ s^{-2} &= \frac{1}{2}\left(k_1(F_{11} + F_{22} - F_{33}) + 1\right) \\ t &= \frac{1}{2}arcsinh\left(\frac{2k_1F_{23}}{\cos\theta(s^{-2}+1)}\right) = \frac{1}{2}arccosh\left(\frac{k_1(F_{11} + F_{22} + F_{33}) - s^{-2}}{s^{-2}+1}\right) \end{aligned}$

Appendix C: Constants in Equation 3.8–3.10

$$\begin{split} T &= \{t_{ij}\}_{3\times 3} \\ a_{1}^{(1)} &= -t_{21}n_{21}n_{13} - t_{22}n_{22}n_{13} \quad a_{1}^{(2)} = +t_{11}n_{21}n_{13} + t_{12}n_{22}n_{13} \\ a_{2}^{(1)} &= +t_{11}n_{21}n_{13} + t_{12}n_{22}n_{13} \quad a_{2}^{(2)} = +t_{21}n_{21}n_{13} + t_{22}n_{22}n_{13} \\ a_{3}^{(1)} &= -t_{22}n_{21}n_{13} + t_{21}n_{22}n_{13} \quad a_{3}^{(2)} = +t_{12}n_{21}n_{13} - t_{11}n_{22}n_{13} \\ a_{4}^{(1)} &= +t_{12}n_{21}n_{13} - t_{11}n_{22}n_{13} \quad a_{4}^{(2)} = +t_{22}n_{21}n_{13} - t_{21}n_{22}n_{13} \\ a_{4}^{(1)} &= +t_{12}n_{21}n_{11} + t_{22}n_{22}n_{11} - t_{11}n_{21}n_{12} - t_{12}n_{22}n_{12} \\ a_{1}^{(3)} &= +t_{21}n_{21}n_{11} + t_{22}n_{22}n_{11} - t_{21}n_{21}n_{12} - t_{22}n_{22}n_{12} \\ a_{3}^{(3)} &= +t_{22}n_{21}n_{11} - t_{21}n_{22}n_{11} - t_{22}n_{21}n_{12} + t_{11}n_{22}n_{12} \\ a_{3}^{(3)} &= +t_{22}n_{21}n_{11} - t_{21}n_{22}n_{11} - t_{22}n_{21}n_{12} + t_{21}n_{22}n_{12} \\ a_{4}^{(3)} &= -t_{12}n_{21}n_{11} + t_{11}n_{22}n_{11} - t_{22}n_{21}n_{12} + t_{21}n_{22}n_{12} \\ b_{1}^{(1)} &= -t_{23}n_{23}n_{13} \quad b_{2}^{(1)} &= +t_{13}n_{23}n_{13} \quad b_{1}^{(2)} &= +t_{23}n_{23}n_{13} \\ b_{1}^{(3)} &= +t_{23}n_{21}n_{11} - t_{13}n_{23}n_{12} \quad b_{2}^{(3)} &= -t_{13}n_{23}n_{11} - t_{23}n_{23}n_{13} \\ c_{1}^{(1)} &= +t_{31}n_{21}n_{12} + t_{32}n_{22}n_{12} \quad c_{1}^{(2)} &= -t_{31}n_{21}n_{11} - t_{32}n_{22}n_{11} \\ c_{2}^{(1)} &= +t_{32}n_{21}n_{12} - t_{31}n_{22}n_{12} \quad c_{2}^{(2)} &= -t_{32}n_{21}n_{11} + t_{31}n_{22}n_{11} \\ c_{1}^{(3)} &= c_{2}^{(3)} &= \mathcal{D}^{(3)} &= \mathcal{D}^{(1)} &= +t_{33}n_{23}n_{12} \quad \mathcal{D}^{(2)} &= -t_{33}n_{23}n_{11} \end{split}$$