

# **Face Shape and Reflectance Acquisition using a Multispectral Light Stage**

Abhishek Dutta

Submitted for the degree of Master of Science (M.Sc.)

Department of Computer Science

**THE UNIVERSITY of York**

October 2010

## **Abstract**

In this thesis, we discuss the design and calibration (geometric and radiometric) of a novel shape and reflectance acquisition device called the “Multispectral Light Stage”. This device can capture highly detailed facial geometry (down to the level of skin pores detail) and Multispectral reflectance map which can be used to estimate biophysical skin parameters such as the distribution of pigmentation and blood beneath the surface of the skin.

We extend the analysis of the original spherical gradient photometric stereo method to study the effects of deformed diffuse lobes on the quality of recovered surface normals. Based on our modified radiance equations, we develop a minimal image set method to recover high quality photometric normals using only four, instead of six, spherical gradient images. Using the same radiance equations, we explore a Quadratic Programming (QP) based algorithm for correction of surface normals obtained using spherical gradient photometric stereo.

Based on the proposed minimal image sets method, we present a performance capture sequence that significantly reduces the data capture requirement and post-processing computational cost of existing photometric stereo based performance geometry capture methods.

Furthermore, we explore the use of images captured in our Light Stage to generate stimuli images for a psychology experiment exploring the neural representation of 3D shape and texture of a human face.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contributions . . . . .	3
<b>2</b>	<b>Related Work</b>	<b>5</b>
2.1	Shape Acquisition . . . . .	5
2.2	Reflectance Acquisition . . . . .	9
2.3	Alignment . . . . .	17
2.4	Real Time Performance Capture of Human Face . . . . .	17
2.5	Stimuli for Psychology Experiments . . . . .	19
<b>3</b>	<b>Design and Calibration of the Multispectral Light Stage</b>	<b>20</b>
3.1	Creating the Spherical Illumination Environment . . . . .	21
3.1.1	Selection of the Light Source . . . . .	23
3.1.2	Estimation of Light Source's 3D Position . . . . .	24
3.1.3	Light Source Intensity and Camera Shutter Controller . . . . .	28
3.2	Diffuse and Specular Reflectance Separation . . . . .	32
3.2.1	Light Source Polariser Orientation . . . . .	33
3.2.2	Simultaneous Capture of Cross Polarized Images Using a Beam Splitter	34
3.2.3	Registration of Cross Polarized Images . . . . .	36
3.2.4	Results of Diffuse and Specular Separation . . . . .	37
3.3	Extending the Basic Light Stage Design for Multispectral Capture . . . . .	39
<b>4</b>	<b>Multispectral Light Stage Data Processing</b>	<b>43</b>
4.1	Spherical Gradient Photometric Stereo using Diffuse Images . . . . .	44

4.1.1	Radiance Equation for Gradient Illumination . . . . .	45
4.1.2	Radiance Equation for Constant Illumination . . . . .	46
4.1.3	Surface Normal Estimation . . . . .	47
4.2	Spherical Gradient Photometric Stereo using Specular Only Images . . . . .	47
4.2.1	Radiance Equation for Gradient Illumination . . . . .	48
4.2.2	Radiance Equation for Constant Illumination . . . . .	50
4.2.3	Surface Normal Estimation . . . . .	50
4.3	Analysis of the Normalizing Constant Value - $N_d$ and $N_s$ . . . . .	51
4.4	Quadratic Programming based Normal Correction . . . . .	53
4.4.1	Modified Radiance Equations for Gradient Illumination . . . . .	55
4.4.2	Modified Radiance Equation for Constant Illumination . . . . .	57
4.4.3	Quality of Surface Normal Estimated Using Original Spherical Gradient Photometric Stereo Method . . . . .	58
4.4.4	Modified Radiance Equations for Complement Gradient Illumination . .	59
4.4.5	Correcting Recovered Surface Normals Using Quadratic Programming .	61
4.4.6	Discussion . . . . .	64
4.5	Minimal Image Sets for Robust Spherical Gradient Photometric Stereo . . . . .	66
4.5.1	Results . . . . .	69
4.5.2	Discussion . . . . .	70
4.6	Registration of Spherical Illumination Images . . . . .	73
4.6.1	Result from alignment of gradient images . . . . .	75
<b>5</b>	<b>Applications of the Light Stage</b>	<b>77</b>
5.1	Real Time Performance Capture . . . . .	77
5.1.1	Original Performance Geometry Capture Method . . . . .	79
5.2	Performance Capture Sequence based on Minimal Image Sets . . . . .	81
5.2.1	Performance Geometry Capture using the New Image Capture Sequence	85
5.2.2	Results . . . . .	88
5.2.3	Discussion . . . . .	90
5.3	Stimuli Image Dataset for Psychology Experiment . . . . .	91
5.3.1	Stimuli Image Dataset . . . . .	92

5.3.2	Results and Discussion . . . . .	95
<b>6</b>	<b>Conclusion</b>	<b>97</b>
6.1	Future Work . . . . .	98

# List of Figures

3.1	Centroid of ideal specular and diffuse reflectance lobe . . . . .	21
3.2	Images of an apple captured under spherical gradient and constant illumination	22
3.3	Our Light Stage . . . . .	22
3.4	Light source intensity for constant and gradient illumination environment . . .	23
3.5	Emission spectrum of VIO (Vio/3.6W/741) LED measured using our CCD spectrometer . . . . .	24
3.6	Centroid $h(x, y)$ of specular highlight depicted in a full illumination mirror ball image (numbers correspond to light source unique identifier) . . . . .	25
3.7	Estimation of light source 3D position using position of its specularity in a mirror ball . . . . .	27
3.8	Position of light sources depicted as black spots on a unit sphere . . . . .	28
3.9	Led intensity and camera trigger control diagram . . . . .	29
3.10	Inter-packet delay parameter introduces controlled amount of delay between ethernet packets generated by a camera . . . . .	31
3.11	Images captured from four test illumination environments . . . . .	32
3.12	Cross polarization . . . . .	33
3.13	Cross polarized images of a hardened chrome steel ball bearing (mirror ball). (left) Specular and diffuse $I_0$ and (right) diffuse only $I_1$ . . . . .	34
3.14	Cross polarized images of a snooker ball. (left) Specular and diffuse $I_0$ and (right) diffuse only $I_1$ . . . . .	35
3.15	Polarizing cube beam splitter and two camera setup for simultaneous capture of cross polarized images ( $I_0$ and $I_1$ ) . . . . .	35

3.16 Manually selected correspondence points in CAM0 (left) and CAM1 (right) image for normalized DLT algorithm . . . . .	37
3.17 Result of diffuse and specular reflectance separation . . . . .	38
3.18 (left and centre) Optical filter wheel attached to existing beam splitter setup for Multispectral acquisition. (right) Optical filters centred at different wavelength of the visible range in a filter wheel . . . . .	39
3.19 Transmission spectrum of single-bandpass optical filters used for the Multispectral light stage . . . . .	41
3.20 Multispectral diffuse and specular image set for a face region . . . . .	42
 4.1 Global ( $X, Y, Z$ ) and local coordinate ( $u, v, n$ ) frame for diffuse images . . . . .	44
4.2 Global ( $X, Y, Z$ ) and local coordinate ( $u, v, n$ ) frame for specular images . . . . .	48
4.3 Centroid (depicted by small white circle) of diffuse and specular reflectance lobe	51
4.4 Distribution of diffuse $N_d$ (bottom left) and specular $N_s$ (bottom right) normalizing constant value for a region (depicted with white rectangle) in the diffuse (top left) and specular (top right) normal map of a white cement plaster statue.	52
4.5 Deformed diffuse (left) and specular (right) lobes due to inter-reflection, ambient occlusion and coarse approximation of spherical illumination . . . . .	54
4.6 Ambient occlusion in concave surfaces . . . . .	55
4.7 Shadows clearly visible in the constant $C$ and gradient ( $X, Y, Z$ ) images (top row) of a white cement plaster statue. The normal map (bottom leftmost - normal components mapped to R,G,B) and the X,Y,Z normal components depicted as grayscale image (bottom right three) do not show the effect of shadows. . . . .	59
4.8 Complement coordinate frames in a Light Stage . . . . .	60
4.9 Result of QP based normal correction applied to surface normals of a white cylinder when initial solution is (left) $\mathbf{x}_0 = (0, 0, 0, 0, 0, 0, n_x^{\text{Ma}}, n_y^{\text{Ma}}, n_z^{\text{Ma}})$ and (right) $\mathbf{x}_0 = (0, 0, 0, 0, 0, 0, n_x^{\text{Wil}}, n_y^{\text{Wil}}, n_z^{\text{Wil}})$ . . . . .	63
4.10 (left) 1 pixel wide verticle region and (right) 1 pixel wide horizontal region in gradient images of the face region of a statue selected for analysis of QP based normal correction. (center) Side view photograph of the statue's face region. . .	63

4.11 $n_z$ component of surface normals in 1 pixel wide selected vertical region (Fig. 4.10 - left) obtained after applying QP based normal correction with initial estimate of surface normals from (top) Ma et al. method i.e $n_0 = n^{[Ma2007]}$ and (bottom) Wilson et al. method i.e. $n_0 = n^{[Wilson]}$ . . . . .	65
4.12 $n_z$ component of surface normals in 1 pixel wide selected horizontal region (Fig. 4.10 - right) obtained after applying QP based normal correction with initial estimate of surface normals from (top) Ma et al. method i.e $n_0 = n^{[Ma2007]}$ and (bottom) Wilson et al. method i.e. $n_0 = n^{[Wilson]}$ . . . . .	66
4.13 z-component ( $n_z$ ) of estimated surface normals of a cylinder . . . . .	69
4.14 Photometric normals of a statue (static) computed using our minimal image set method (top) and that computed using Ma et al. [26] (middle left) and Wilson et al. [37] (middle right). All the three complement base pairs — $(X, \bar{X})$ , $(Y, \bar{Y})$ and $(Z, \bar{Z})$ — possible in our minimal image set method was used to generate similar photometric normals. (Bottom) Distribution of angular difference between normal maps computed using our minimal image set method and that computed using [26] and [37]. . . . .	71
4.15 Photometric normals of a face (non-static) computed using our minimal image set method (top) and that computed using Ma et al. [26] (middle left) and Wilson et al. [37] (middle right). All the three complement base pairs — $(X, \bar{X})$ , $(Y, \bar{Y})$ and $(Z, \bar{Z})$ — possible in our minimal image set method was used to generate similar photometric normals. (Bottom) Distribution of angular difference between normal maps computed using our minimal image set method and that computed using [26] and [37]. . . . .	72
4.16 Alignment of spherical gradient images used Joint Photometric Alignment[37]. For illustration purpose, all the intensity values were scaled by 2 except the warped $r_{\bar{x}}(v)$ which was scaled by 3 because it falls on the dark side of spherical gradient illumination. . . . .	75
4.17 Complement constrain residual for 100 iterations of the joint photometric alignment technique applied to a $298 \times 182$ spherical X gradient image . . . . .	76

5.1	Image capture sequence used by Wilson et al. [37] (top). The proposed image capture sequence (bottom) based on our minimal image sets method. . . . .	78
5.2	Development of the modified capture sequence . . . . .	82
5.3	Development of the modified capture sequence . . . . .	85
5.4	Photometric normals obtained from proposed image capture sequence based on minimal image sets . . . . .	89
5.5	Image capture requirement analysis for performance capture using Wilson et al. [37] and our 4 image method . . . . .	91
5.6	(left) normal map obtained using spherical gradient photometric stereo, and corresponding (right) <i>shape only</i> stimuli image . . . . .	93
5.7	(left) <i>texture-only</i> stimuli image, and (right) combined <i>shape and texture</i> stimuli image . . . . .	94

# List of Tables

3.1	Manual and automatic measurement of $d$ and $S_c$ . . . . .	27
3.2	Single bandpass filters used for the Multispectral light stage . . . . .	40
5.1	Image capture requirements for performance capture using Wilson et al. and our 4 image method . . . . .	90

# Acknowledgments

First, I would like to thank Dr. William Smith, my supervisor, for always keeping his office doors open. This not only meant that he was accessible all the time but also ensured that I could discuss my problems as I encountered them while exploring my research terrain. Due to his friendly and cheerful nature, I never hesitated to discuss any of my ideas. He deserves to be “Supervisor of the Year”. Thanks also to him for patiently standing in the Light Stage during long data capture sessions.

Thanks to my sister, Dr. Smita Dutta, for sending delicious food to me every week. My brother-in-law, Dr. Abhishek Kabra, tirelessly shuttled between York and Driffield to make sure that I was never left alone during weekends and holidays. I never felt like staying thousands of miles away from home.

Thanks also to Manoj mama(*maithili* word for maternal uncle) for giving me a laptop at the age of 14 and introducing me to Computers. To be honest, I did not even deserve a pocket calculator from him as I had destroyed all his electronics kits and tools while exploring the wonderful world of hobby electronics. I am also truly grateful to Pawan Kumar Pandey (Computer Science teacher at DPS, Dharan) for being the architect of my C/C++ programming skills.

Finally, I would also like to acknowledge the EURECA project (a project funded by the Erasmus Mundus External Cooperation Window (EMECW) of the European Commission) for funding my MSc research.

*On your shoulders I stand,  
For a vantage point, very few have.  
Narrow minded thoughts fear to tread in my conscience,  
Thanks to your teachings that are always on guard.  
Wisdom made sense after you explained Russel's "Knowledge and Wisdom",  
And so did Tagore's "Where the mind is without fear (Gitanjali)".*

*- to Baba and my parents*

# **Declaration**

I declare that all the work in this thesis is solely my own, except where attributed and cited to another author. Some of the material in this thesis has been previously published by the author: please refer to Section 1.1 for details.

# Chapter 1

## Introduction

The 3 dimensional shape and skin texture (i.e. 2 dimensional skin reflectance) of a human face determines its visual appearance. Face shape and reflectance acquisition devices aim to capture these two information from real human faces, either separately or in a combined form. Extensive research has been pursued in the past three decades to develop devices that are capable of acquiring shape and reflectance information accurately and conveniently. Photorealistic renderings of human faces can be created using the highly detailed shape and texture information acquired from these capture devices. As a result, digital actors, who have natural looking faces, can be extensively used in movies and animation. Additionally, access to accurate shape and reflectance information is the key to developing a new generation of face recognition algorithms that can maintain accuracy even under arbitrary pose and illumination variation.

In this thesis, we discuss the design and calibration (geometric and radiometric) of a novel shape and reflectance acquisition device that is able to capture highly detailed facial geometry (down to the level of skin pores detail) and a Multispectral reflectance map. The Multispectral reflectance map can be used to estimate biophysical skin parameters such as the distribution of pigmentation and blood beneath the surface of the skin. This device, called the *Multispectral Light Stage*, is an extension of the Light Stages developed at UC Berkeley and University of South California (USC ICT) [1]. We use a beam splitter based capture device to simultaneously acquire parallel and cross polarised images. This ensures that both the acquired images are in perfect registration and hence results in very accurate diffuse and specular reflectance

separation. Previous Light Stages [1] relied on a servo motor to flip the plane of polarisation of a polarising filter. As a result, the capture time increases. This increase in capture time does not affect the shape and reflectance recovery of a static object. However, for a non-static objects like a human face, it is extremely difficult to remain in the same position and maintain a facial expression for the capture duration. Hence, increased capture time compromises the quality of recovered surface geometry and reflectance information because the captured images are no longer perfectly aligned.

Capturing the facial geometry of human actors during a dynamic performance is the first step towards creating digital actors that can produce realistic and natural facial expressions. Such digital actors are in high demand for movies and animation. The human face is capable of producing a large number of facial expressions with small non-rigid motion of facial muscles. Hence, to reproduce such expressions in a digital actor, it is essential that, in addition to capturing overall facial expression, the motion of fine scale skin features (like wrinkles, pores, scars, etc) are also captured. Such fine details are the key ingredient to reproducing natural facial expressions.

Photometric stereo based methods can capture all the fine scale details of a dynamic performance. However, they require expensive high speed photography equipment and are data intensive. In this thesis, we have proposed a novel real time performance capture sequence by exploiting the fact that high quality photometric normals can be recovered using just 4 images. This new capture sequence not only reduces the data capture requirements for realtime performance geometry capture, but reduces the need for expensive high speed photography equipment for capture of highly detailed performance geometry.

Understanding the way the human brain represents and processes visual information is the key to creating machine vision algorithms that can match the capabilities of the human visual cortex. Unfortunately, non-invasive reverse engineering methods are the only practical tools available for such study. One of the popular choices for such reverse engineering approach is to study the brain activity of a human observer, usually in a controlled environment, when they are exposed to various types of visual stimuli. Brain activity during such experiments is mostly monitored using functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG). In addition to the capability of monitoring devices, the effectiveness

of these experiments also depend on the ability to control various aspects of the visual stimulus. For example: the neural representation of 3D shape and 2D skin reflectance function (i.e. texture) can be effectively studied if we can create stimuli image that only contain the 3D shape or the 2D skin reflectance information. The photographs that we capture using a standard camera contain a mix of these two information sources. Availability of such stimulus data is the key to unravelling the psychology and neuropsychology of face perception.

In this thesis, we have also explored the use of our Light Stage for generating such a stimulus dataset. This dataset was used by Jones et al. [22] in their study of the neural representation of 3D shape and 2D skin reflectance of human faces. To our knowledge, this is the first research to use Light Stage data (the high quality photometric normals and image of a face under spherical illumination) for generating a psychological stimulus dataset. We envisage that this will lead to further exploration of the use of a Light Stage in psychology experiments.

## 1.1 Contributions

The major contributions of this thesis are:

**Design and Calibration of Multispectral Light Stage** We have described the design and calibration (geometric and radiometric) of an extended version of the original Light Stage. Our Light Stage design consists of a “beam splitter” based setup that allows simultaneous capture of parallel and cross polarised images. Furthermore, our capture device uses a filter wheel containing narrowband optical filters, to separately record reflectance in different bands of the visible spectrum.

**Minimal Image Sets for Robust Spherical Gradient Photometric Stereo** We extend the analysis of original spherical gradient photometric stereo developed by Ma et al. [26] to consider the effect of deformed diffuse lobes on the quality of recovered surface normals. Based on our modified radiance equations, we explore a Quadratic Programming (QP) approach to correction of surface normals recovered using existing spherical gradient photometric stereo methods. Using the same set of equations, we propose that a minimal set of 4 images can recover surface normals of the quality provided by the

existing 6 image method. This minimal image set method has also been described in the following publication:

- Abhishek Dutta and William A. P. Smith. Minimal image sets for robust spherical gradient photometric stereo. In ACM SIGGRAPH ASIA 2010 Sketches, SA 10, pages 22:122:2, New York, NY, USA, 2010. ACM.

It is important to realise that our analysis is based on the following simplifying assumption described in section 4.4.1 and 4.4.4: overall deformation in the diffuse reflectance lobe for gradient and complement gradient illumination environment can be quantified using a single scalar parameter  $\delta_{\{x,y,z\}}$  and  $\delta_{\{x,y,z\}} + \delta_{\{\bar{x},\bar{y},\bar{z}\}}$  respectively.

**Novel Capture Sequence for Real Time Performance Capture** Based on our “Minimal Image Sets” analysis and building on the work of Wilson et al. [37], we propose a new image capture sequence for facial performance geometry capture during dynamic performance. In addition to reducing the data capture requirement of Wilson et al. performance capture framework, the proposed capture sequence also reduces its computational cost by requiring alignment of only one, instead of three, pair of gradient and complement gradient images.

**Stimulus Image Dataset for Psychology Experiment** We explore, for the first time, the use of images captured in a Light Stage to generate stimulus images for a psychology experiment. For a given face, we generate three stimulus images: the first contains only the 3D shape information, the second contains only 2D skin reflectance (texture) information and the third contains both shape and texture information. This image dataset has been used for studying the neural representation of 3D shape and texture of a human face.

# Chapter 2

## Related Work

In this chapter, we will discuss the previous work done in shape and reflectance acquisition and the method used to align captured data affected by subject motion during the capture process. In the later part of this thesis, we discuss how the data captured in our Light Stage can be used to produce better stimulus images for psychology experiments. Moreover, we also discuss a new performance capture strategy to reduce the data capture requirement of existing methods used to capture facial geometry during dynamic performance. Therefore, in this section, we also review the previous work done in these two areas.

### 2.1 Shape Acquisition

Two of the most popular methods for 3D shape acquisition are: Depth from triangulation and Photometric Stereo. As a result, two types of representation exist for conveying the 3D shape information: 3D meshes (vertices and their connectivity) and normal map (surface normal at each image point).

In the depth from triangulation method, a surface point is viewed from two (or more) viewpoints using a calibrated camera and the corresponding image points are recorded. In the ideal case (i.e. with no imaging noise), the rays through these two (or more) corresponding images points will intersect at a point in 3D space. This 3D location represents the original surface point represented by the corresponding image points observed using calibrated camera. There exist several methods to determine the corresponding image points in multiple views of a 3D object recorded using a calibrated camera. Nehab [30] developed the “spacetime stereo”

framework to classify all the existing depth from triangulation methods. This classification was based on the domain (spatial or temporal) in which the corresponding image points are located. The class of methods which determine corresponding image points by the analysis of similar pixels in the image plane are classified as “spatial domain” methods. On the other hand, methods which determine image point correspondance by analysis of pixel intensity variation over time are classified as “temporal domain” methods. This classification not only provides a unified view of all the existing depth from triangulation methods, but also provided valuable insight for development of two new methods in [30] that exploited both “spatial” and “temporal” domain constraints of corresponding image points.

Woodham [40] proposed the photometric stereo method to determine the surface geometry of each image point using diffuse images captured by varying the direction of incident illumination while keeping the view direction constant. The basis for this technique is the observation that each pixel intensity of a Lambertian surface image illuminated by a point source results in a linear photometric equation. If the direction of point source is known, then this system of linear equations can be inverted to recover unknown diffuse albedo and surface orientation using at least 3 images. The unit surface normal constraint allows separation of these two quantities from a system of 3 linear photometric equations. This early version of photometric stereo method developed in [40] did not consider the effect of specular highlight, shadow and inter-reflection in the captured images.

It is convenient to invert a linear system resulting from photometric equations of a Lambertian reflection. However, no such linear system exists for non-Lambertian surface reflection. Hence, several previous research in photometric stereo has focused on developing methods to detect image points affected by specular highlight and shadow. Coleman and Jain [7] proposed the use of 4 point light sources, instead of just 3, to detect and exclude pixels affected by specular highlights and shadow. Three surface normals corresponding to a single surface patch were available from these 4 images captured using point light sources. They predicted that a large amount of deviation in both direction and magnitude of these three surface normals would occur for a pixel affected by specular highlight. This allowed them to tag and remove the specular source. Their method was based on the assumption that only one of the four light sources can cause specular highlight at any given image point.

Following the 4 source strategy, Barsky and Petrou [2] used four spectrally distinct light sources to exploit the linearly independent photometric equations resulting from different color channels of a color image. They used spectral or directional cues to detect shadows and highlights in the input images. However, the choice of threshold parameter proved pivotal to the detection accuracy. [2] observed that with increase in imaging noise, a single threshold value cannot detect all the specular highlights and shadows present in the captured images.

A completely different approach to photometric stereo was pursued by Basri et al. [3]. They felt the need for photometric stereo technique to work under general illumination condition. They argued that it was not always possible to control illumination for large outdoor structures or have knowledge of light source direction and strength for photographs taken under everyday lighting condition. The fact that any image of a convex Lambertian object under complex illumination can be approximated as a linear combination of 4 (first order) or 9 (second order) harmonic images<sup>1</sup> forms the basis of their proposed photometric stereo algorithm for general illumination. Harmonic images can be expressed in terms of surface albedo and normal components and hence such decomposition allowed them to estimate these two quantities. They propose 9D (which requires at least 9 images) method and 4D (which requires at least 4 images) method of photometric stereo under general lighting condition. The 9D method produces slightly better results at the expense of higher computational cost of decomposing images captured under general lighting condition into 9 harmonic images. The authors illustrate the quality of surface geometry reconstruction by using more images (64, 32, 11, 10) than the required minimum. For example: the fine scale surface details of a volleyball was recovered by using 64 images of the ball lit by point light sources (strength and direction unknown). Hence, at the expense of large computational cost and comparatively larger number of images, they were able to estimate good quality surface geometry under general illumination conditions. It is important to realise that this method is not applicable to images containing specular highlight or shadows.

All the previous photometric stereo methods ([40], [7], [2], [3]) treated specular highlight in an image as a undesirable effect which restricted the application domain of photometric stereo. Extensive research has been done to develop methods for tagging and removal of

---

<sup>1</sup>harmonic images represent the image of an object in low frequency lighting condition

specular highlights. However, Ma et al. [26] used specularity to their advantage and acquired specular normal maps containing fine surface details of a human face never recovered by previous methods. They have shown how high resolution shape and reflectance information can be measured using an extended version of photometric stereo called the spherical gradient photometric stereo. An object is placed at the centre of a “light stage” which uses polarised spherical gradient illumination arranged such that the plane of polarisation after reflection, from the object towards the camera, are all the same. This allows the setup to separate diffuse and specular reflectance components by acquiring parallel and cross polarised images. The key observation underpinning this approach is that the centroid of the diffuse or specular reflectance lobe coincides with the surface normal or reflection vector respectively. The insight of Ma et al. was to show how to estimate the reflectance centroids using spherical gradient illumination conditions. When integrated with an illumination gradient in X, Y, or Z direction, the corresponding components of the reflectance centroid, and hence surface normal, can be recovered. This extended version of photometric stereo was capable of recovering fine scale surface details that was unmatched by the existing photometric stereo methods in terms of quality and level of detail.

The quality of surface geometry recovered using Ma et al. [26] method is affected by the extent to which the following assumptions are satisfied: *a)* no shadowing of light sources , i.e. object is convex; *b)* no inter-reflections; *c)* Fresnel term<sup>2</sup>; and *d)* light sources closely approximate a continuous illumination environment. The last assumption can be addressed by maximising the number of light sources in the light stage: Ma et al. used 156 LEDs attached to vertices and edges of a twice subdivided icosahedron. This method also ignores light source attenuation effects, which is equivalent to assuming all the points on the object lie exactly at the centre of the light stage.

Wilson et al. [37] proposed using gradient and complement gradient images to reduce the effect of shadowing. Instead of using the “ratio” method of Ma et al. to compute the surface normal components, they used the difference of gradient and complement gradient images to estimate more accurate surface geometry. They argued, “since the pixels that

---

<sup>2</sup>The proportion of light transmitted into the surface and subsequently diffusely reflected varies with incidence angle according to Fresnel’s equations. The same effect will occur when the diffused light exits the surface again.

are dark under one gradient illumination condition are most likely well exposed under the complement gradient illumination condition” [37]. Recently, Dutta and Smith [12] have proved the validity of this claim by showing that the difference image method of Wilson et al. result in cancellation of symmetric deformation in diffuse lobes. This deformation cancellation property is not present in the method of Ma et al. because it involves estimation of surface normal components from the ratio images which preserves the term quantifying deformation in diffuse reflectance lobe. Dutta and Smith [12] have also shown that a minimal four image set can achieve the “improved robustness” quality of [37] while preserving the “reduced data capture” benefit of [25]. They used a “light stage” with only 41 LEDs (attached only to vertices of a twice subdivided icosahedron) to study the degradation in the quality of recovered surface geometry with increase in “light discretization” i.e. coarse approximation of continuous spherical gradient illumination. Their minimal image set method was able to recover high quality normal map using a spherical illumination created with only 41 LEDs.

## 2.2 Reflectance Acquisition

Reflectance models are an attempt to mathematically capture the interaction of light with a given material or class of materials. In Computer Graphics and Computer Vision, the reflectance properties of human skin have been investigated extensively in the past two decades. Reflectance models allow the creation of photo-realistic renderings of human faces in arbitrary pose and under complex illumination. It helps with development of natural looking cosmetics because reflectance models provide insight into the way light interacts with human skin [33]. Photo therapy (or Laser based treatment) of skin disease requires good understanding of the interaction between light and human skin. Skin reflectance models help improve the precision of such treatment methods by allowing designers to simulate the effect of light based skin treatment methods [19].

Reflectance models mostly rely on measured reflectance data for estimation of their model parameters. The practicability of reflectance models depend on the ease with which reflectance properties of real world objects can be acquired. Most reflectance models discuss the related capture device that can acquire reflectance measurements required for estimation of the model parameters. Often, new capture devices trigger the development of reflectance models that

can make full use of the available reflectance data. Hence, in addition to reviewing existing skin reflectance models, we will also discuss about the corresponding reflectance measurement device. In this section, we will discuss previous work done in the reflectance models related to human skin. These models can easily be modified to simulate light interaction in other types of materials like milk, marble, etc.

Marschner et al. [27] developed a reflectance capture device that, for the first time, measured the *in vivo* surface reflectance of human skin. A set of three machine readable targets were used for geometric calibration. This allowed automatic estimation of the relative position of the light source, sample material and the camera. For radiometric calibration, they used a calibrated reference source to determine the spectral characteristic of the camera. The radiometric calibration allowed them to relate the recorded pixel values with radiance reflected from the sample under study. A section of forehead was imaged under several incident illumination directions (capture time  $\sim 30\text{min}$  ). This region of the face was selected because it was relatively smooth, convex and involved least amount of deformation during long capture session. Using the machine readable targets, the geometric arrangement of sample, camera, light source and reference white target was automatically determined for each captured image. All these information was supplied to a “derenderer” which computed the BRDF value at each pixel positon by dividing the measured pixel radiance with the source irradiance. The scene geometry required by the “derenderer” was captured using a 3D range scanner. The authors produced renderings of human head using the measured BRDF of the skin sample. This rendering had a hard look and lacked the features of actual human skin because the proposed skin reflectance model only considered the surface reflectance component of the overall skin reflectance.

Overall skin reflectance from human skin can be decomposed into two components: surface reflectance (modelled using BRDF) and subsurface reflection (modelled using BSSRDF). In facial skin, the subsurface reflection component dominates the overall reflection [11][19]. Hence, a skin reflectance model involving only the surface reflectance component cannot achieve photorealistic rendering of human skin. Debevec et al. [9] developed a novel capture device called the “light stage” which can illuminate a face from a dense set of spherical positions while recording the appearance from multiple viewpoints. Using the images captured

in this device, they propose a method to recreate facial appearance under novel illumination and viewpoint. Their method exploits the fact that a given facial appearance under general lighting condition can be represented as linear combination of facial appearance under illumination by point light sources densely distributed over a sphere surface. In other words, if all the possible appearance of a human face lie in a  $\mathbf{N}$  dimensional space, then the face images captured under illumination by a dense sampling of incident illuminaton direction forms the basis of this vector space. To recreate facial appearance from novel viewpoint, they create a geometric model of the face using structured lighting. The facial appearance from original viewpoint is projected onto this geometric model and appearance from novel viewpoint is computed based on this projected appearance.

The facial appearance projected from the original viewpoint cannot reproduce the shifting and scaling in the measured reflectance function caused by change in viewpoint. Hence, the viewpoint specific changes to diffuse and specular reflectance components from a region in forehead — a region also selected by Marschner et al. [27] — is used to extrapolate the corresponding deviation in other regions of the face. Specular and diffuse reflectance components are separated using the difference of parallel and cross polarized images. Colorspace analysis is used to separate the diffuse and specular reflectance components in other parts of the face. These separated components undergo shifting and scaling according to novel viewpoint specific scaling and shift observed for a  $2 \times 5$  pixel in forehead region. The specular reflectance component is fitted to the microfacet based rough surface model of Torrance and Sparrow [32].

The authors aimed to produce realistic rendering of subsurface reflectance phenomena. Hence, the subsurface scattering data was not fitted to any skin reflectance model and instead was only used to determine viewpoint specific changes to subsurface reflectance component for a given illumination environment. Hence, the renderings produced using this method cannot reproduce correct subsurface scattering effect due to heterogeneous illumination environment. Also, it is a data driven technique and hence requires capture of a large number of images ( $64 \times 32 = 2048$  photographs) resulting in long capture procedure (1 min). Moreover, the data driven nature of this method prevents its use for editing or transfer of facial appearance characteristics among the captured subjects i.e. we are locked in the facial appearance space

spanned by the captured data.

Hanrahan and Krueger [17] developed a reflectance model which, for the first time, related the physical properties (like refractive index, thickness, absorption and scattering coefficients) of a layered material to the subsurface reflectance properties of that material. They presented a model — suitable for Computer Graphics — for reflection of light due to subsurface scattering in a layered material. This model treated a physical material as a layered homogeneous scattering medium. The authors suggested modeling heterogeneity in a material using random noise or a texture map. Reflection from outer surface of the material was modeled using the Torrance and Sparrow [32] microfacet model and subsurface scattering was modeled using the proposed reflectance model based on 1D linear transport theory. Rendering of a human face, whose 3D geometry was acquired using a medical MRI scanner, was generated using a two layered model which correspond to the epidermis and dermis layers of a human skin. The model parameters for each layer were chosen manually to generate renderings that were close in appearance to real human skin. The authors did not present a method to capture reflectance data required for estimation of the model parameters. Hence, although the model was anatomically motivated and produced acceptable face renderings, the layer parameters used for generating these renderings were not based on measurements from actual human skin.

Jensen et al. [21] took a different approach to modeling subsurface reflectance by introducing the dipole model based on “diffusion theory” to the graphics community. The diffusion theory existed in the optics community prior to this but had never been used for subsurface skin reflectance modeling. They modeled subsurface scattering of light using diffusion approximation which is based on the observation that light distribution in a highly scattering media is isotropic. The authors acknowledge inspiration for this model from the use of diffusion theory used in describing the scattering of laser light in human tissue in medical physics research. Unlike [17], they described a capture device setup to capture reflectance data of real world objects which can be used to estimate all the model parameters. This device focused a beam of white light on the sample material and recorded a High Dynamic Range (HDR) image corresponding to radiance fall off from the point of incident beam i.e. the radially symmetric diffusion profile. Using this setup, they measured the diffusion profile  $R_d(r)$  of a

wide variety of real world objects like milk, human skin, marble, etc. The model parameters, absorption  $\sigma_a$  and reduced scattering coefficients  $\sigma'_s$ , were estimated from these measured diffusion profiles. Hence, for the first time the measured subsurface scattering characteristics of real world material was plugged into a reflectance model.

The capture device proposed by Jensen et al. is not suitable for facial skin as focused beam of white light may harm the skin during the capture process. Hence, they measured the diffusion profile of skin in the arm region for their experiment and extrapolated the subsurface reflectance properties to skin in other body parts. This reflectance model assumed the scattering medium to be semi-infinite i.e. only one side of the medium had well defined boundary. In other words, this model assumed that every component of the incident light will eventually be reflected back. Hence, the model failed to account for incident light that gets transmitted into the material. Also, this model can only be used with highly scattering media because the diffusion approximation, used by this model, is only applicable to highly scattering medium i.e.  $\sigma'_s \gg \sigma_a$ .

Weyrich et al. [36] overcame this practical limitation of the diffusion profile capture device by building a contact probe consisting of a linear array of optical fibres: one of them being the source fibre and the remaining are detectors. Using this contact probe, all the parameters of the dipole model can be robustly estimated. They modeled skin as a single layered homogeneous scattering medium. They added a spatially varying absorptive film of zero thickness — called the modulation texture layer — to simulate inhomogeneous scattering in human skin. Use of texture map (or random noise) was also suggested by [17] to simulate the effects due to heterogeneity in a scattering medium. The parameters of this layer were estimated from albedo map and the dipole model parameters  $\sigma'_s$ ,  $\sigma_a$  were obtained from the diffusion profile captured using their contact probe. They measured 3D face geometry, skin reflectance and subsurface scattering using custom built devices for 149 subjects of varying age, gender and race. This allowed them to study the variation of subsurface scattering parameters for a large population of skin types. Moreover, flexibility in their reflectance model allowed intuitive editing of facial appearance. For example: they presented the results of face renderings obtained by transfer of skin features like freckle and skin type (BRDF and albedo).

Weyrich et al. added a suction pump to the contact probe in order to maintain the con-

tact and position during the capture of diffusion profile. The total capture time of 88 sec. necessitated addition of the suction feature to the contact probe. It is known that physical pressure alters the normal blood flow mechanism in a human skin. Hence, the scattering and absorption coefficients obtained from the diffusion profile captured using such contact probe may be biased to some extent. Moreover, the design of contact probe limits its use to flat areas of a human face. Hence, they extrapolate the reflectance measurements from forehead, cheek and below the chin to other parts of a human face. Although, the addition of a modulation texture closely reproduces effects due to heterogeneity in skin, it makes the reflectance model anatomically implausible.

Ghosh et al. [15] have described a skin reflectance model which treats overall skin reflectance as the linear sum of four reflectance components: specular, single scattering, shallow scattering and deep scattering. These components are classified according to the depth of skin from which they get reflected. They are able to estimate all the model parameters from just 20 photographs of human face captured under spherical illumination (developed by [25]) and projected lighting condition. The spectral difference between these two sources is compensated by computing a colour transformation matrix which transforms both photographs to a common colorspace. A 24 ColorChecker square and 10 skin patches are imaged under these two illumination condition to compute this colour transformation matrix. First, the specular and single scattering components are separated from overall reflection by exploiting the fact that these components preserve the polarization of incident light. This is also true for single scattering reflectance because the probability of depolarization of light increases exponentially with each additional scattering event. Furthermore, these two components are separated from each other by exploiting another interesting difference between these two components: any non-specular reflectance component that preserves polarization of incident light is treated as single scattering term. The specular reflectance component is modeled using Torrance and Sparrow [32] microfacet model and Hanrahan and Krueger [17] first order single scattering BRDF is used to model the single scattering term.

Multiple scattering is composed of shallow and deep scattering reflectance. The diffuse only image obtained from polarization difference image of parallel and cross polarized images contain the multiple scattering reflectance component. Using the method of Nayar et al. [29],

they separate the multiple scattering reflectance component into direct and indirect reflectance components. The key observation underpinning this separation is that when the frequency of illumination pattern is in the order of thickness of epidermis, then the direct component relates to shallow scattering and indirect component corresponds to the deep scattering reflectance component.

Ghosh et al. modelled subsurface scattering (or multiple scattering) using reflectance in a two layered medium. Although, they do not explicitly model the epidermis and dermis layers in a human skin, they use the notion of deep and shallow scattering to roughly model the light interactions occurring in these two layers. Deep scattering, caused by the bottom layer, was modelled using the dipole model of Jensen et al. [21] which treats the scattering layer to be semi-infinite. The semi-infinite assumption is practical for the bottom layer but not for the top layer. Were the top layer semi-infinite, it would not have a transmission profile and hence the bottom layer would not receive any portion of the incident light. Hence, the shallow scattering, caused by the top layer, is modeled using the multipole diffuse model of Donner and Jensen [10]. The transmission profile of the top layer obtained using the multipole model becomes the incident profile for the bottom layer which is modeled using a dipole model.

Ghosh et al. model the overall skin reflectance as the linear sum of four reflectance components: specular, single scattering, shallow scattering and deep scattering. The purely additive nature of these reflectance components prohibits the modelling of phenomenon involving interaction between skin layers; for instance, the epidermal effects on dermal scattering.

Donner et al. [11] proposed the “physiologically most advanced skin reflectance model that is still practical for rendering”[35]. They described a two layered skin reflectance model which used, for the first time, spatially varying model parameters for each layer to account for heterogeneous light transport in human skin. Using diffuse images captured in 9 different bands of the visible region, they were able to model spectral dependence of subsurface scattering characteristics. The proposed two layered skin reflectance model has 6 spatially varying model parameters which relate to physiological skin parameters and are represented as 2D chromophore <sup>3</sup> map. The two layers in this model correspond to epidermis (top layer) and dermis (bottom) layers of a human skin. A thin absorbing layer was added between these

---

<sup>3</sup>skin constituent that selectively absorb some spectral bands of the incident light

two scattering layers which corresponds to pigmentation concentrated in a narrow region between epidermis and dermis of actual human skin.

They demonstrated the strength of anatomically motivated reflectance model by generating photo-realistic renderings of human hand from just the user painted 2D chromophore maps corresponding to the 6 model parameters. In addition to the user painted chromophore map method, they also devised a inverse rendering based approach to estimate the 6 model parameters from multispectral images of a flat skin sample. They developed a filterwheel based multispectral capture device to capture the multispectral reflectance map of a flat skin sample in the arm region. Scattering in each layer was modeled using the multipole diffusion model of [10].

The proposed inverse rendering approach is not scalable to the Multispectral images of the full face. Inverse rendering resembles a “brute force” approach in which estimation of model parameters from Multispectral images involves searching a 6D space for model parameter values that minimised the difference between rendered Multispectral images and the captured Multispectral photographs. This strategy is not applicable to estimation of model parameters from Multispectral images of full face because the complex geometry of human face makes the process of inverse rendering computationally intractable. Also, the capture process required the skin surface to be coated with Ultrasound gel. Ultrasound gel has same refractive index as a human skin and hence created a smooth surface over the skin sample under observation. This allowed the use of Fresnel transmission term for estimation of radiance transmitted into the skin. It is impractical to apply the Ultrasound gel to full face for acquisition of Multispectral images.

Ghosh et al. [15] and Donner et al. [11] have proposed the current state-of-the-art skin reflectance models. While the data driven model of Ghosh et al. can estimate model parameters of complete face in natural expressions using just 20 photographs captured in 5 sec, the anatomically plausible skin reflectance model of Donner et al. can produce realistic renderings of human hand with just a user painted 2D chromophore map representing the model parameters. On the other hand, the data acquisition procedure of Donner et al. is not scalable to complete human face whereas the reflectance model proposed by Ghosh et al. lacked biophysically meaningful parameters.

## 2.3 Alignment

Almost all shape and reflectance acquisition system has to deal with motion of non-static objects, like a human face, during the capture process. Marschner et al. [27] used a set of three machine readable targets for automatic estimation of the relative position of the light source, sample material and the camera during the 30 minute capture process. Debevec et al. [9] proposed using a head rest to reduce motion during the capture process which lasted for 1 minute. In reality, it would be extremely difficult to maintain facial expression and position for 1 minute in spite of a head rest. Weyrich et al. [36] used a contact probe with suction to maintain the position of the subsurface reflectance capture device during the 90 seconds of capture time. Ghosh et al. [15] have not mentioned how they corrected for subject motion during the capture of 20 photographs in 5 seconds. Donner et al. [11] use a filterwheel based multispectral capture device to capture 9 multispectral photographs of a skin sample in the arm region. As compared to human face, it is relatively easy to maintain position of arm during the capture process. They marked a rectangular region in the skin sample which allowed them apply rigid alignment methods.

Recently, Wilson et al. [37] have developed the Joint Photometric Alignment technique for the registration of gradient images captured in a Light Stage. Traditional optical flow based alignment techniques were not applicable to the alignment of gradient images as the “brightness constancy” assumption is violated in each of these images. Wilson et al. exploited the complement image constraint to devise an iterative algorithm for alignment of gradient images. Photometric normals computed from aligned gradient images can recover the fine surface details like wrinkles, scar, etc present in a human face.

## 2.4 Real Time Performance Capture of Human Face

Marker based facial motion capture is widely used for the capture of geometric deformations in human face during a dynamic performance. The 3D position and velocity of these reflective markers are used as cues to the 3D motion of body structure to which these markers are attached. A limitation of this method is that it requires placement of a very large number of markers on the target face in order to accurately model the 3D motion of each facial muscle.

In addition to inconvenience caused by these markers during facial performance, there exists a limit to which these markers can be attached to a face. This limitation prevents from acquiring fine scale geometric details of face muscles during a dynamic performance. Human observers have a mastery in detecting unnatural facial motion caused by sparse distribution of these markers. Hence, marker based motion capture techniques are not used for close up shots of the human faces.

Furukawa and Ponce [14] have developed a markerless 3D motion capture method for human faces. They track the nonrigid motion of vertices in the 3D mesh of the face obtained from multiview stereo technique. Their method is capable of dealing with unreliable texture information due to fast motion, self occlusion, etc. However, this method involves a data intensive capture process and is affected by specular highlights on a face.

Wilson et al. [37] have developed facial performance geometry capture method which is not data intensive and can capture highly detailed facial geometry without requiring expensive and complex setup of high speed photography. They capture a set of gradient and complement gradient spherical illumination images which flank the constant illumination image, also called the tracking frame. Using their Joint Photometric Alignment method, the gradient images are aligned to the tracking frame which allowed computation of photometric normal at each tracking frame. These photometric normals are warped to the intermediate gradient frames according to the flow fields computed by the alignment stage. This process is called “Temporal Upsampling” because it increases the effective performance capture frame rate by adding warped photometric normals at the temporal position of intermediate gradient frames. They use the spherical gradient photometric stereo technique to recovery very high resolution photometric normal at each tracking frame. Hence, their dynamic performance capture algorithm is able to capture the motion of very fine facial features like wrinkles, pores visible during skin deformation, etc. Moreover, the absence of makers on face allows capture of natural facial expressions.

The proposed Joint Photometric Alignment method is an iterative process requiring two optical flow computation in each iteration. Because this alignment procedure has to be applied to each of the gradient and complement gradient image pairs separately, the computational cost of performance capture using this method is very high. This does not affect the prac-

ticability of this performance capture method because the alignment is a post-processing operation which can be carried out offline.

## 2.5 Stimuli for Psychology Experiments

Research in Psychology and Neuropsychology of face perception has always relied on Computer Vision and Computer Graphics community for stimuli image dataset required for their experiments. Ability to control different aspects of facial appearance is the key to success of these experimental procedures designed to unravel the face representation and processing mechanisms of the visual cortex in human brain.

Caharel et al. [6] used a 3D Morphable Model to generate stimuli images for studying the time course (i.e. temporal sequence) for processing of 3D shape and 2D skin reflectance information of a human face. Their stimuli image contained face images in which texture and shape information of the test subjects were controlled. Although, the 3D Morphable Model produces facial rendering close to natural human faces, it does not include the high frequency skin texture detail. Lack of detailed skin texture, which is known to contribute to face perception, can bias the results of such psychology experiments.

Recently, we explored the application of Light Stage in generating stimuli image dataset for the psychology experiment conducted by Jones et al. [22]. This experiment investigated the neural representation of 3D shape and 2D skin reflectance information of a human face. Using the image data captured in our Light Stage, we were able to separate 3D shape and 2D skin reflectance information for a given face. Spherical illumination of a Light Stage ensured that the texture images were not affected by shadows. Also, these “texture only” face images included all the high frequency facial skin details like wrinkles, mole, frackles, etc.

## Chapter 3

# Design and Calibration of the Multispectral Light Stage

The Photometric stereo technique was developed by Woodham [40] to determine the surface geometry of each image point using images captured by varying the direction of incident illumination while keeping the view direction constant. Ma et al. [26] have proposed spherical gradient photometric stereo — an extended version of original photometric stereo — for acquisition of high resolution shape and reflectance information. A spherical illumination environment is pivotal to this state-of-the-art shape and reflectance acquisition technique because it requires images of an object captured under spherical gradient and constant illumination environment. In this chapter, we discuss the design and calibration of a device that can be used to create a spherical gradient illumination environment. We also propose an extended version of the required acquisition device setup to allow capture of multispectral images in a spherical illumination environment.

The mass centroid of an ideal diffuse reflectance lobe coincides with the surface normal  $\vec{n}$ . For the specular reflectance lobe, the centroid coincides with the reflection vector  $\vec{r}$  as shown in Fig. 3.1. The centroid  $(x_0, y_0, z_0)$  of a diffuse or specular reflectance function  $f(x, y, z)$  can be computed by integrating it with a linear gradient. Mathematically,

$$(x_0, y_0, z_0) = \frac{1}{\int_{-1}^1 f(x, y, z) dx} \left( \int_{-1}^1 x f(x, y, z) dx, \int_{-1}^1 y f(x, y, z) dy, \int_0^1 z f(x, y, z) dz \right) \quad (3.1)$$

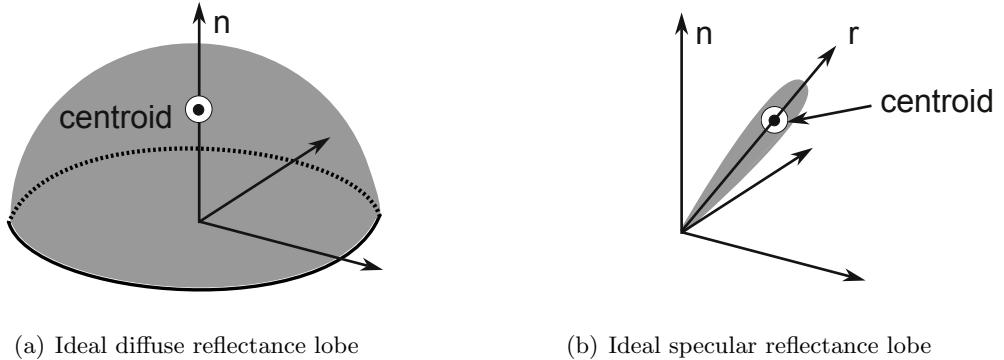


Fig. 3.1: Centroid of ideal specular and diffuse reflectance lobe

The insight of Ma et al. [26] was to show how to estimate the diffuse and specular reflectance centroids using spherical gradient illumination. They proposed the spherical gradient photometric stereo technique which suggests that when integrated with a linear illumination gradient in the X, Y or Z direction, the corresponding component of the reflectance centroid, and hence surface normal, can be recovered. The key observation underpinning this approach is evident when we look at the radiance equation for diffuse and specular reflection:

$$r = \int_{\Omega} P(\omega) R(\omega, n) d\omega, \quad (3.2)$$

where  $P(\omega)$  is the intensity of light incident from direction  $\omega$  and  $R(\omega, n)$  is the Lambertian or specular Bidirectional Reflectance Distribution Function (BRDF). According to (3.2), if we replace the illumination environment  $P(\omega)$  with a spherical gradient illumination in X, Y or Z, the radiance value recorded by an imaging device is related linearly to the centroid of Lambertian or specular BRDF  $R(\omega, n)$ . In the next section, we will discuss design and calibration of a “light stage” : a device proposed by Ma et al. [26] to create a spherical illumination environment.

### 3.1 Creating the Spherical Illumination Environment

Spherical illumination refers to an illumination environment in which every surface patch of an object receives illumination incident from every direction of its visible hemisphere. Fig. 3.2 shows the images of an apple illuminated by spherical gradient and constant illumination. An object placed at the centre of a sphere can be illuminated by spherical illumination by



Fig. 3.2: Images of an apple captured under spherical gradient and constant illumination

using light sources distributed evenly and finely over the surface of that sphere. Ma et al. [26] argued that as the position of edges and vertices of a twice subdivided icosahedron closely approximates the surface of a sphere, LEDs attached to these positions can create spherical illumination. They proposed a device called “light stage” (or led sphere) that consists of 156 LEDs attached to the edges and vertices of a twice sub-divided icosahedron. Fig. 3.3 shows an image of our light stage of diameter 1.58 meter consisting of 41 LEDs attached only to the vertices of twice sub-divided icosahedron. In 3.1.1, we discuss the reason behind using only 41 LED in our Light Stage.

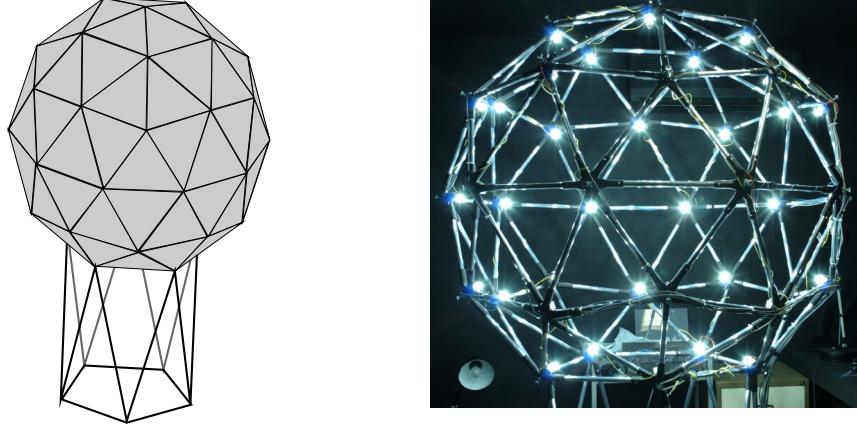


Fig. 3.3: Our Light Stage

Constant illumination is created by switching all the LEDs to their maximum brightness level as shown in Fig. 3.4(a). For the X, Y or Z gradient illumination environment, the intensity of each LED is proportional to the X, Y or Z coordinate of their 3D position respectively. If the 3D position coordinate of each LED is normalized i.e.:  $\|(x, y, z)\| = 1$  then Fig. 3.4(b) depicts the plot of LED intensity for gradient illumination environment. We can setup a gradient illumination environment by assigning each LED an intensity level that is proportional to their 3D position with respect to the center of the light stage. Hence, the knowledge of

light source 3D position is essential to setup a gradient illumination environment.

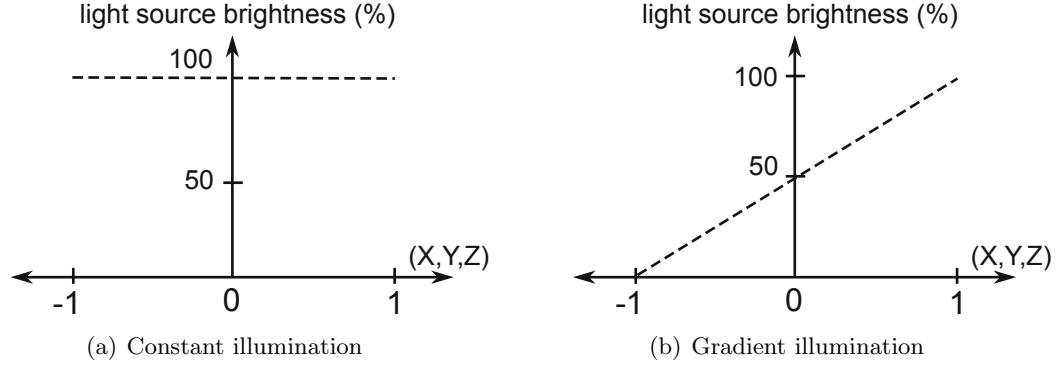


Fig. 3.4: Light source intensity for constant and gradient illumination environment

### 3.1.1 Selection of the Light Source

We have extended the basic light stage design of Ma et al. [26] to achieve the following additional functionalities

- simultaneous capture of cross polarized images using a polarizing beam splitter
- Multispectral capture capability using a set of narrow band optical filters

We have selected the VIO (Vio/3.6W/741) HighPower White LED (manufactured by General Electric Illumination) as the light source for our light stage because :

- the light reaching the camera sensor is attenuated by the light source polariser(< 50% transmission), optical filter(< 90% transmission) and the polarizing beam splitter(< 50% transmission). Hence the camera sensor receives only 22% of the total emitted light even if we image a perfect reflector. The VIO LED has the brightness of 196 lumens [24]. This level of brightness is adequate to image human skin when the attenuation factor of the capture device is 0.78.
- The 180° beam angle of these LED provide complete coverage of large objects like human face in a small light stage of diameter 1.58 meter.

The data obtained from our Multispectral light stage can be used to obtain parameters of skin reflectance models like Donner et al. [11]. It is known that there occurs peak absorption by human skin chromophores in the 400 – 450nm visible range band [11, Fig. 7]. This fact can

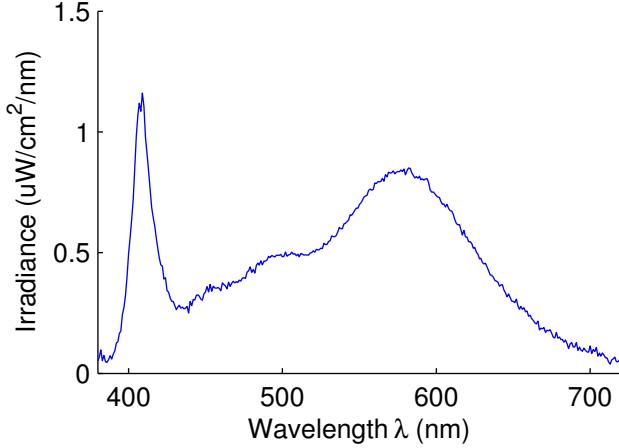


Fig. 3.5: Emission spectrum of VIO (Vio/3.6W/741) LED measured using our CCD spectrometer

be exploited to obtain accurate model parameters for parameteric skin reflectance models. Fig. 3.5 shows that apart from peak emission in 550 – 600nm, the VIO LED also has peak emission in the 380nm to 420nm range: a common feature of most LEDs. This behaviour is ideal for capturing the multispectral reflectance map of human skin.

Because of the high cost of VIO LED, we decided to attach these light sources only to the vertices (and not to the edges) of a twice subdivided icosahedron. Although, this design decision causes extreme “light discretization”, we have developed an algorithm in section 4.5 to exploit the complement gradient constraint in order to reduce the effects of inter-reflection, ambient occlusion and “light discretization”.

### 3.1.2 Estimation of Light Source’s 3D Position

From the discussion in 3.1, it is evident that the knowledge of 3D position of each light source is essential for the setup of spherical gradient illumination environment. These 3D positions should be represented with reference to the center of the light stage as the object placed at the center of the light stage needs to be illuminated by gradient illumination. The 3D positions of each light source can be estimated by manual measurement or by using Computer Aided Design (CAD) drawing of the light stage. However, this method does not provide accurate measurement of 3D position. Moreover, due to limitations related to manufacture of geodesic domes, there is always some asymmetry and deformation introduced during assembly of the light stage.

We estimate the 3D position of each light source in a viewer centred coordinate system by exploiting the relationship between light source position and the position of its specularity on a mirror ball placed at the centre of the light stage.

### Determining the Location of Specular Highlight

We place a 76.2mm hardened chrome steel ball bearing (mirror ball whose boundary<sup>1</sup> is shown in white in Fig. 3.6) at the center of the light stage and capture its photograph when each of the light sources are switched on individually. These images contains the specular highlight corresponding to each light source. The captured images are preprocessed with morphological erosion and dialation operations to remove any stray bright spots and make the specular highlight more symmetric. The centroid  $h(x, y)$  (as shown in Fig. 3.6 by red cross hair) of bright spot in each image forms the location of specular highlight caused by a given light source.

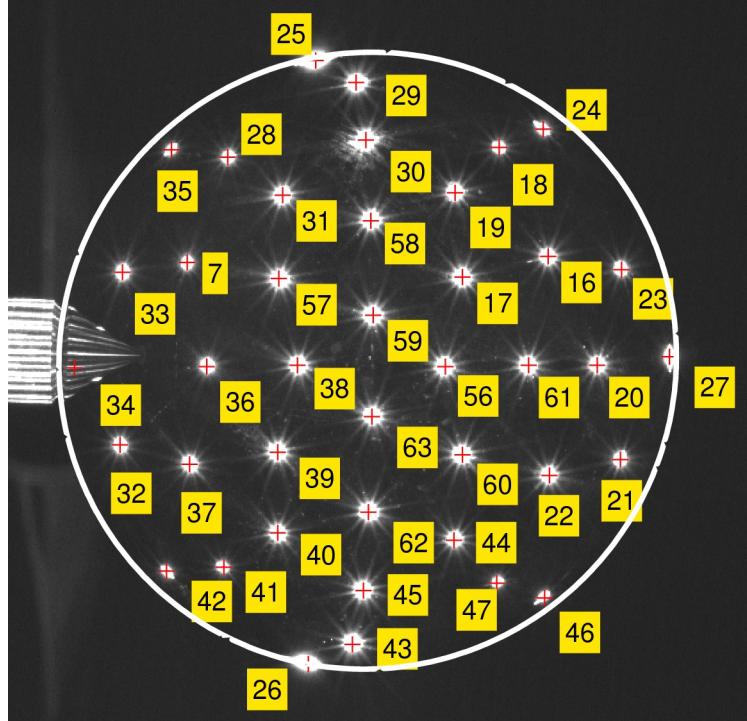


Fig. 3.6: Centroid  $h(x, y)$  of specular highlight depicted in a full illumination mirror ball image (numbers correspond to light source unique identifier)

<sup>1</sup>The edge of cylindrical rod supporting the mirror ball tapers to a smaller radius to provide threading for screws and this forms the contact point for mirror ball. Therefore, the white boundary fit on the far left of Fig. 3.6 is correct.

In Fig. 3.6, notice that there is no specular highlight for LED 34. The reason being that the specular highlight caused by this LED falls in the blind spot region of the sphere surface visible in the captured images. This blind spot is caused by the stand that supports the metallic sphere. These specular highlight locations can be determined by interpolation of symmetric position of neighbouring light sources in the light stage.

### Mirror Ball Sphere Centre

The 3D coordinate of centre of the mirror ball is essential for the computation of each light source's direction. We apply the method proposed by Wong et al. [39] to recover the sphere center. First, we manually select at least 6 points on the conic  $C$  formed by the boundary of the mirror ball in its image. Using the direct least square fitting method of Fitzgibbon et al. [13], we obtain the parameters  $(a, b, c, d, e, f)$  that define the conic  $C$  such that any point  $x$  lying on  $C$  satisfies the equation

$$\tilde{x}^T C \tilde{x} = 0 \quad \text{where, } C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix},$$

and,  $\tilde{x}$  is the homogeneous coordinate representation of  $x$ . The result of this fitting process is depicted by the white conic shown in Fig. 3.6.

The calibration matrix  $K$  was computed using the Matlab camera calibration toolbox [4]. To remove the effect of camera calibration matrix  $K$ , we normalize the image with  $K^{-1}$ . This normalization transforms conic  $C$  to a normalized conic  $\hat{C} = K^T C K$ . Using singular value decomposition, we diagonalize conic  $\hat{C}$  into

$$\hat{C} = M D M^T = M \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix} M^T. \quad (3.3)$$

Finally, the sphere centre can be computed using

$$S_c = M \begin{bmatrix} 0 & 0 & d \end{bmatrix}^T \quad \text{where } d = R \sqrt{\frac{a+b}{b}}. \quad (3.4)$$

Table 3.1: Manual and automatic measurement of  $d$  and  $S_c$

	Manual (mm)	Using Wong et al. [39] (mm)
$d$	890	898.35
$S_c$	(0, 0, 890)	(21.33, -16.56, 897.94)

Here,  $R$  is the radius of mirror ball and  $d$  is the distance between camera center and sphere center. Wong et al. [39] have also proved that the light direction estimated from an observed specular highlight in an image of a sphere will be independent of the radius used in recovering the location of the sphere center. It is important to recognise that this observation is valid only when the light sources are placed at infinity. In this experiment, the light sources are present close to the mirror ball and therefore requires accurate measurement of mirror ball radius to recover correct values of  $d$  and  $S_c$  using [39]. Very small deviation between manually measured values for  $d$  and  $S_c$  and that measured using [39] (as shown in Table 3.1) support the fact that manual measurement of mirror ball radius was quite accurate.

### Light Source Position Estimation

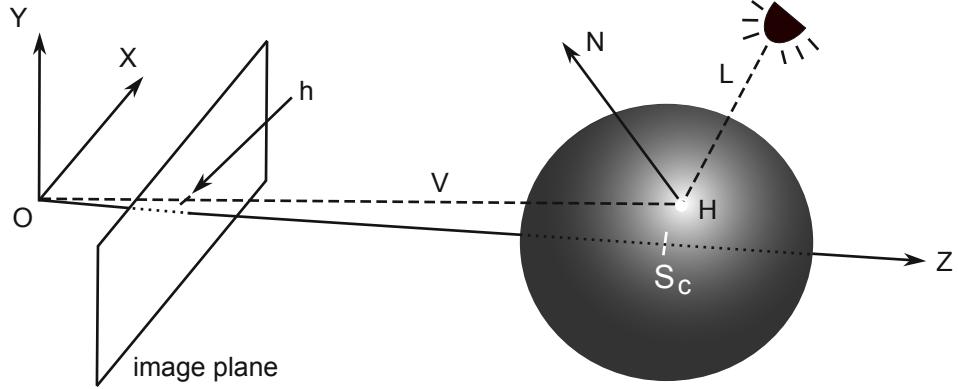


Fig. 3.7: Estimation of light source 3D position using position of its specularity in a mirror ball

In Fig. 3.7,  $H(X, Y, Z)$  and  $h(x, y)$  represent the location of specular highlight on the surface of sphere and the image plane respectively. The ray from the camera center  $O$  to the location of specular highlight  $H$  on the sphere surface forms the view vector  $V$ .  $L$  represents the light source direction and  $N$  is the surface normal at point  $H$ .

From the discussions in previous two sections (3.1.2 and 3.1.2), we have the values for sphere center  $S_c$  and image plane location of specular highlight  $h(x, y)$ . However, to estimate

the light source direction  $L$ , we need the value of one more quantity  $H$  : the location of specular highlight on the surface of mirror ball.

To determine the values of  $H$ , we first construct a ray  $\vec{l}$  originating at the camera center  $O$  through the pixel coordinate of the specularity in the image plane  $h(x, y)$ . If  $K$  is the camera calibration matrix and  $\tilde{h} = [x \ y \ 1]$  represents  $h(x, y)$  in homogeneous coordinates, then

$$\vec{l} = K^{-1}\tilde{h} + O.$$

The location  $H(X, Y, Z)$  of specular highlight on the mirror ball is the point of intersection of ray  $\vec{l}$  and a sphere centered at  $S_c$  with radius  $R$ . [31, p116] discusses the method to compute the point of intersection of a ray and a sphere.

With all these measurements to hand, we can now compute the light source direction vector  $L$  using

$$L = (2N \cdot V)N - V \quad (3.5)$$

where,  $V = \frac{H-O}{|H-O|}$  and  $N = \frac{H-S_c}{|H-S_c|}$ . The positions of light sources recovered using this method are depicted in Fig. 3.8

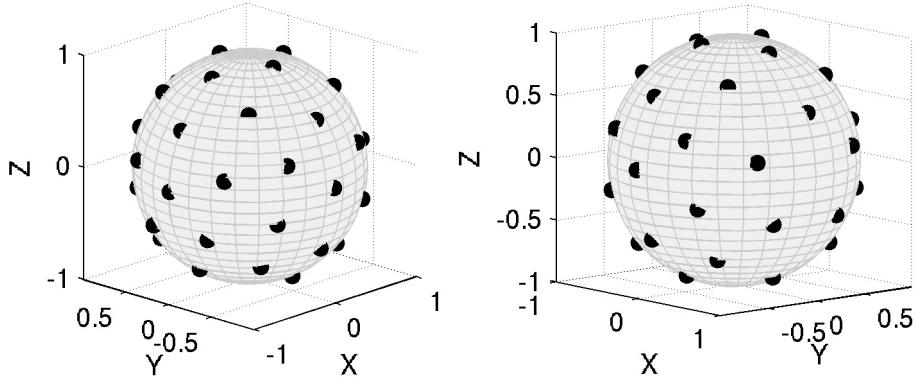


Fig. 3.8: Position of light sources depicted as black spots on a unit sphere

### 3.1.3 Light Source Intensity and Camera Shutter Controller

The LED controller used in this project was designed and built by Cooper et al. [8]. This controller is based on an MBED<sup>2</sup> board and PCA9685<sup>3</sup> I2C LED controller. The MBED

---

<sup>2</sup><http://www.mbed.org>

<sup>3</sup>[www.nxp.com/documents/data\\_sheet/PCA9685.pdf](http://www.nxp.com/documents/data_sheet/PCA9685.pdf)

(LPC1768) board acts as the control hardware for PCA9685 and camera shutter. PCA9685 is 16 channel I2C LED controller that uses 12 bit (4096 brightness levels) Pulse Width Modulation (PWM) to control LED brightness. The LED controller [8] uses four PCA9685 to provide control interface for 41 LEDs in our Light Stage. As PCA9685 is controlled using I2C bus, this design can be easily extended to provide control interface for even larger number of LEDs.

We have used the “Geodesic Light Dome” designed by Cooper et al. [8] to control all the light sources and the camera shutter in our Light Stage. AnMBED (LPC1768) board acts as the control hardware for the camera shutter and PCA9685 LED controller. PCA9685 is 16 channel I2C LED controller that uses 12 bit (4096 brightness levels) Pulse Width Modulation (PWM) to control LED brightness. We have used four PCA9685 to provide the control interface for 41 LEDs in our Light Stage. As PCA9685 is controlled using I2C bus, this design can be easily extended to provide control interface for even larger number of LEDs. TheMBED board provides a “C” like programming environment for the control of LED intensity and camera shutter.

We use two JAICM200GE machine vision camera along with a polarizing beam splitter to capture cross polarized images (refer to 3.2.2 for details). The connection diagram ofMBED, PCA9685, two cameras and a computer is shown in Fig. 3.9.

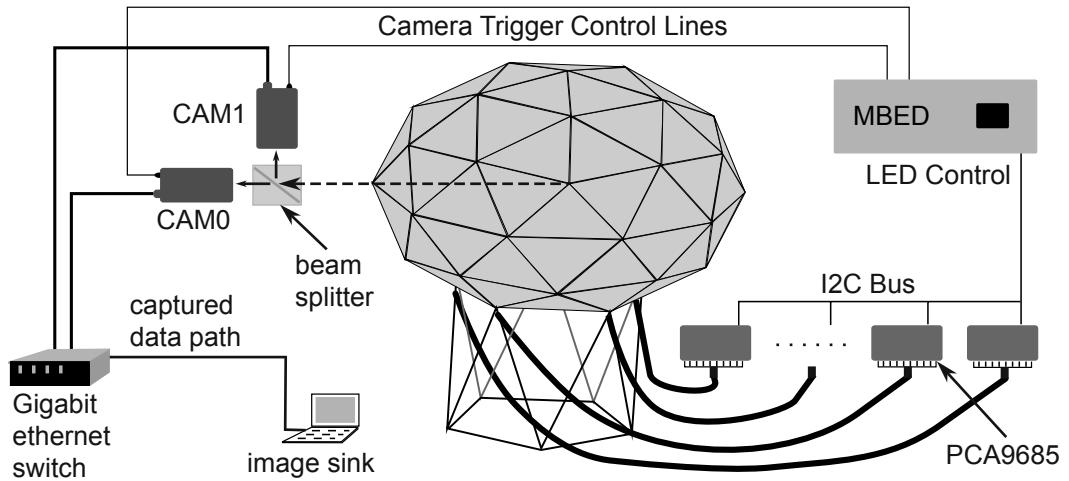


Fig. 3.9: Led intensity and camera trigger control diagram

## **LED Intensity Control**

The LED controller [8] uses MBED board and PCA9685 to provide a “C” like programming environment for LED intensity control. We program the MBED board such that it writes the LED channel identifier (explained in the next paragraph) and corresponding brightness level (0 to 4095) to its I2C bus pins which in turn is connected to the PCA9685. The PCA9685 chip provides a very simple interface (in the form of I2C commands) for LED brightness control and therefore allows us to avoid the intricacies of Pulse Width Modulation (PWM) based LED intensity control.

The LED identifier assigned to specular highlight shown in Fig. 3.6 represents the channel identifier of the corresponding LED. The 3D position of each LED (obtained using the method discussed in 3.1.2) is used to create an Intensity Lookup Table (ILT). This lookup table contains the intensity level (0 to 4095) of all the LED for X, Y and Z gradient illumination environment and we store the ILT in the flash memory of MBED. This allows us to setup X, Y or Z gradient illumination in just  $23161\mu s$ .

## **Camera Trigger Control**

The two JAICM200GE camera are connected to a computer (henceforth referred to as “image sink” because it receives all the captured images) via a Gigabit ethernet switch. All the camera functions, including the camera shutter, can be controlled via the GigE vision interface. However, to synchronise the illumination environment setup with the image capture, we use the MBED board (present in the LED controller [8]) to control the shutter of both camera. We built a cable to use the digital I/O lines available in the MBED board of the LED controller [8] for control of the two cameras via their General Purpose Input Output (GPIO) interface. The camera is configured to use Pulse Width Modulation (PWM) based shutter control in which the rising edge and falling edge represent shutter open and close events respectively. After configuring the camera shutter control mode to “Pulse Width Trigger Mode” (PWC), the “image sink” then waits for ethernet packets containing the image data. Now all the capture sequence is handled by the MBED board which is programmed to control the camera shutter using PWM.

MBED board performs the following two operations in a sequence to allow capture of

spherical gradient images in X, Y, Z and constant illumination environment:

- Setup the brightness of each LED according to the data in Intensity Lookup Table (ILT) corresponding to required illumination environment
- Send a pulse to GPIO pins of both camera such that the rising and falling edge indicate the shutter open and close events respectively.

This process is repeated to setup X, Y, Z or constant illumination environment.

The two cameras are triggered simultaneously and hence they start sending ethernet packets, containing the captured image, at the same time. The network switch has sufficient memory to avoid congestion when a single set of gradient images is captured. However, for real time performance capture, a huge amount of data is generated every second: two camera capturing  $1200 \times 1000$  image at 10 bit/pixel ( $\sim 2$  byte/pixel) generate  $2 \times 93$  grayscale images per second for 30 fps tracking frame rate. This causes congestion in the network switch resulting in loss of ethernet packets due to limited buffer memory of the camera and network switch.

The camera manufacturer recommends using the inter-packet delay feature available in the camera to avoid congestion in the network switch[20, p22, p26]. They provide a tool to compute optimal inter-packet delay in order to make best use of the available video bandwidth. The inter-packet delay parameter of a camera determines the time interval delay between two adjacent packets transmitted by the camera to the receiving computer. If this delay time is larger than the packet size of other camera, the “image sink” will receive ethernet stream in which the packets from both camera are interleaved as shown in Fig. 3.10. This allows for optimal use of available video bandwidth. For details on computing the inter-packet delay, refer to [20, p26].

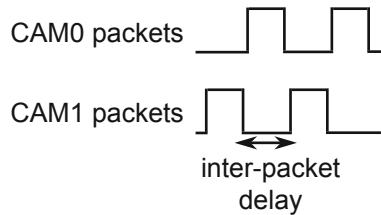


Fig. 3.10: Inter-packet delay parameter introduces controlled amount of delay between ethernet packets generated by a camera

## Verification of Camera and Illumination Synchronisation

The MBED code is executed sequentially. Hence, if we trigger the camera only after setting up intensity of all the LEDs in the Light Stage, the illumination and capture process would always be synchronised. However, to verify correct synchronisation of our setup, we created four test illumination environment. The test illumination environment have the same setup time as the original gradient illumination environment. Moreover, as these test illumination have simple patterns of light, it allows us to verify if there is any “illumination leak” from neighbouring illumination environments. Fig. 3.11 shows the mirror ball captured under the four test illumination environment. These images support our assumption of correct synchronization.



Fig. 3.11: Images captured from four test illumination environments

## 3.2 Diffuse and Specular Reflectance Separation

Reflection from a surface consists of diffuse and specular reflectance components. The specular component is caused by light reflected directly from the surface and hence is also called a surface phenomena. The diffuse component results from light rays penetrating the surface, undergoing multiple reflections and refractions, and re-emerging at the surface [28]. For linearly polarized incident light, specular reflection has polarization oriented perpendicular to the plane of incidence<sup>4</sup> and the diffuse component is essentially unpolarized [38, p84]. The fact that the specular component has the same polarization as the incident light is the basis of the “cross polarization” technique for separation of the diffuse and specular reflection components. The axis of polarization of the linear polariser is such that the plane of polarization of the incident light is orthogonal to the plane of incidence as shown in Fig. 3.12. When the axis of polarization of the analyzer<sup>5</sup> is aligned with the plane of incidence(as shown in Fig. 3.12),

<sup>4</sup>plane of incidence at a given surface point is defined as the plane containing view vector  $\vec{V}$  and the surface normal  $\vec{n}$  at that point

<sup>5</sup>the linear polarizer placed in front of the camera

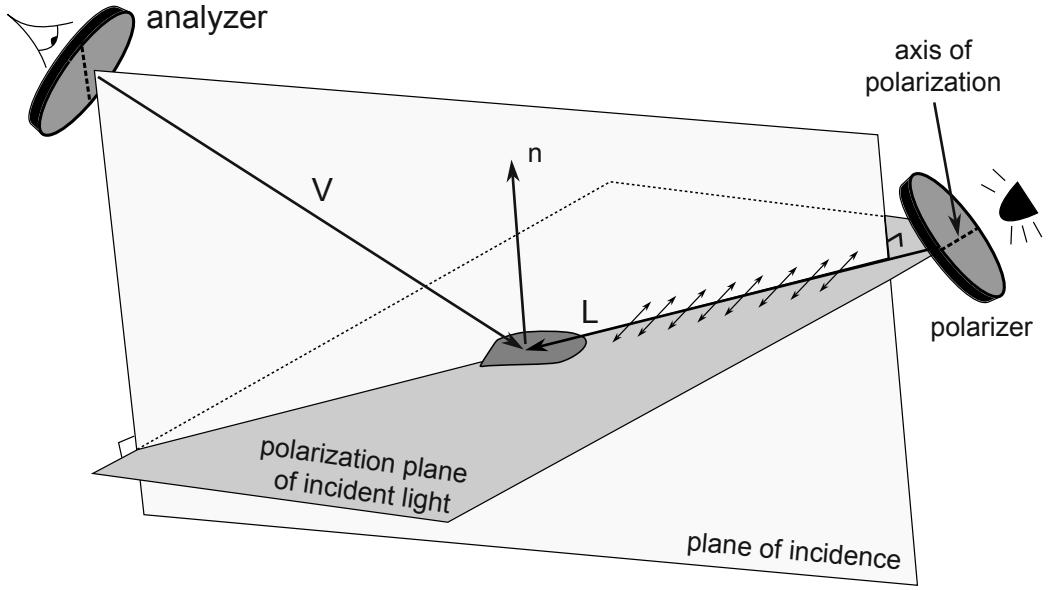


Fig. 3.12: Cross polarization

only the diffuse component of the reflected light can be observed ( $I_1$  - diffuse only image). This is because the specular component of the reflected light has polarization perpendicular to the plane of incidence. To record both the unpolarized diffuse and the polarized specular reflection ( $I_0$  - specular and diffuse image), the axis of polarization of the analyzer is oriented orthogonal to the plane of incidence. Hence, we can write the following expression for the two cross polarized images [25, p40 ]

$$I_0 = \frac{1}{2}I_D + I_S, \quad I_1 = \frac{1}{2}I_D.$$

Finally, the images containing only the diffuse and specular reflectance components can be recovered using :  $I_S = I_0 - I_1$  and  $I_D = 2I_1$ .

### 3.2.1 Light Source Polariser Orientation

Spherical gradient illumination requires all the light sources to be distributed uniformly on the surface of a sphere. Hence, we require a spherical field of linear polarization for all the light sources in which all of them have the same plane of polarization. The optimal orientation of each light source polariser is achieved when the “diffuse only” image  $I_1$  of the two cross polarized images contain no specular highlights. In addition to the numerical optimization method, Ma et al. [25] also describe a simpler method to obtain such optimal orientation by

manually tuning the orientation of each light source polarizer until all the specularity from a mirror ball gets cancelled in one of the cross polarized images. The “live view” feature<sup>6</sup> of our acquisition device allowed us to quickly find the optimal orientation of each polarizer.



Fig. 3.13: Cross polarized images of a hardened chrome steel ball bearing (mirror ball). (left) Specular and diffuse  $I_0$  and (right) diffuse only  $I_1$

We came across a peculiar behaviour while looking for optimal polariser orientation using a hardened chrome steel ball bearing (mirror ball) placed at the centre of the light stage. It was not possible to completely cancel specular highlight in the “diffuse only” image. This effect was more pronounced for the specular highlights corresponding to the light sources for which the angle of incidence was close to the  $90^\circ$  as shown in Fig. 3.13. Suspecting the way metallic surface interact with polarized light, we tried using a snooker ball (made of PVC - a dielectric). We were able to quickly find the optimal polarizer orientation using a snooker ball as shown in Fig. 3.14. Ghosh et al. [15] have also emphasised the use of a dielectric spherical reflector (i.e plastic ball) for the polariser orientation calibration. As we intend to only capture gradient images of dielectric materials (like human face, ceramic and plastic objects, etc), we did not further investigate the peculiar behaviour of metallic surfaces.

### 3.2.2 Simultaneous Capture of Cross Polarized Images Using a Beam Splitter

A square linear polariser mounted on a servo motor in front of the camera lens was used by [25, p46] as the analyzer. The servo motor rotated the filter rapidly to allow capture of cross polarized images. The polariser rotation time required by this mechanical setup

---

<sup>6</sup>real time view of both the cross polarized images

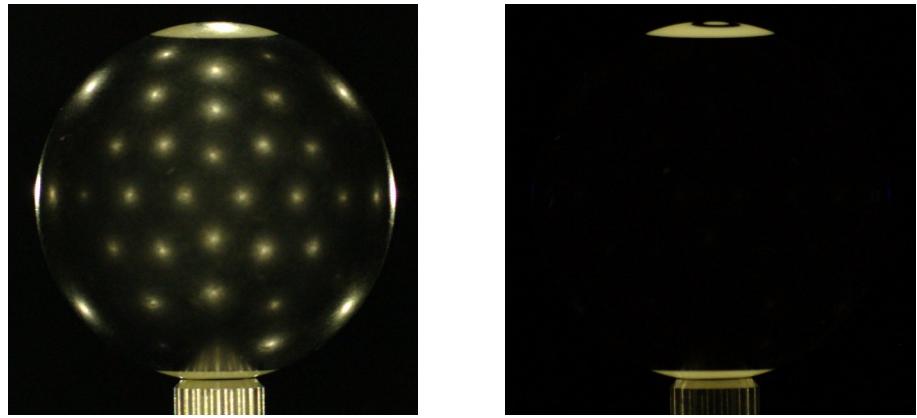


Fig. 3.14: Cross polarized images of a snooker ball. (left) Specular and diffuse  $I_0$  and (right) diffuse only  $I_1$

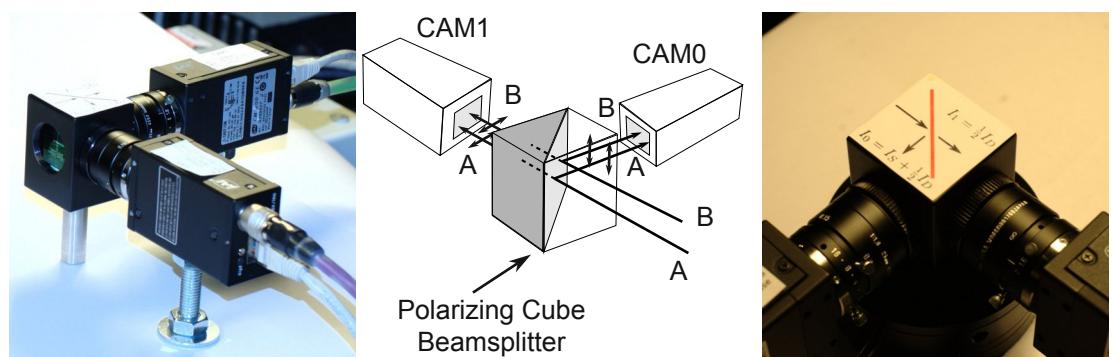


Fig. 3.15: Polarizing cube beam splitter and two camera setup for simultaneous capture of cross polarized images ( $I_0$  and  $I_1$ )

caused some delay in the capture of the two cross polarized images. In the case of static objects, this does not cause any problem. However, when cross polarized images of a human face are captured by such servo motor based setup, slight motion between the two images cannot be avoided. So the two cross polarized images are not in perfect registration and thus requires some alignment before the diffuse and specular only images can be computed from them. Moreover, the mechanical rotation setup achieved by a servo motor cannot ensure equal amount of rotation in every instance.

To avoid the problems introduced by a mechanical servo motor based system, we used the Techspec®Polarizing Cube Beam splitter<sup>7</sup> (25mm, Visible Range) to split the incoming light into *S* polarized<sup>8</sup> and *P* polarized<sup>9</sup> components. These two reflection components are recorded by two cameras attached to the two faces of a cube beam splitter as shown in Fig. 3.15. This setup ensures that both the cameras simultaneously capture the cross polarized images. It is interesting to note that one of the cameras needs to be rotated by 180° and the captured image be compensated for mirror reflection (using MATLAB `fliplr()`) to undo the inversion of image caused by splitting of incoming light along two orthogonal axes. In the leftmost image of Fig. 3.15, the rotation of one of the camera in our setup is evident from the flipped sequence of the ethernet and power cables.

### 3.2.3 Registration of Cross Polarized Images

The images captured by both the cameras are automatically registered if their principal axes intersect. However, such a setup is not possible to achieve as the connection adapter used to screw in the camera lens to the beam splitter mount introduces offset between the principal axes of both cameras. Hence, to align the two cross polarized images, we need to compute a 2D homography matrix  $H$  that transforms one of the cross polarized image in order to align it with the other image. Note that this alignment step is quite different from the registration in Ma et al. [25] setup required to compensate for the motion of the subject during the capture process. This alignment is performed to cancel the offset in the principal axis of the two camera so that the images captured by the two camera are in perfect registration.

---

<sup>7</sup><http://www.edmundoptics.com/onlinecatalog/displayproduct.cfm?productID=2986>

<sup>8</sup>In *s* polarization, the electric field vector is perpendicular to the plane of incidence.

<sup>9</sup>In *p* polarization the electric field vector is parallel to the plane of incidence



Fig. 3.16: Manually selected correspondence points in CAM0 (left) and CAM1 (right) image for normalized DLT algorithm

We obtain an initial estimate of the 2D homography matrix  $H$  using normalized Direct Linear Transform (DLT) [18, p109]. The automatic corner detection tool in [4] is used to detect 65 corner points in a checkerboard image captured using the two camera and beam splitter setup. These points are corrected manually, as shown in Fig. 3.16, for subpixel accuracy and then supplied to the DLT algorithm as ( $n = 65$ ) initial 2D to 2D point correspondances. Using the initial estimate of  $H$  from DLT, we determine the Maximum Likelihood Estimate (MLE) of  $\hat{H}$  that minimizes the Samson's error [18, p114]. This homography matrix is applied to all the images captured by CAM0 so that the transformed CAM0 images are in alignment with the CAM1 image.

Lens distortion is another effect that can lead to misalignment of images by the two cameras as described in [41]. However, the effect is likely to be very small and therefore we ignore the contribution of lens distortion.

### 3.2.4 Results of Diffuse and Specular Separation

The CAM0 gradient images ( $I_0$ ) can be transformed using 2D homography  $\hat{H}$  to obtain images  $\hat{I}_0$  that are aligned with the CAM1 gradient images ( $I_1$ ). Diffuse and specular only images can now easily be obtained using

$$I_S = \hat{I}_0 - I_1 \quad \text{and} \quad I_D = 2I_1.$$

The result of specular and diffuse separation for constant spherical illumination of faces and a static object is shown in Fig. 3.17.

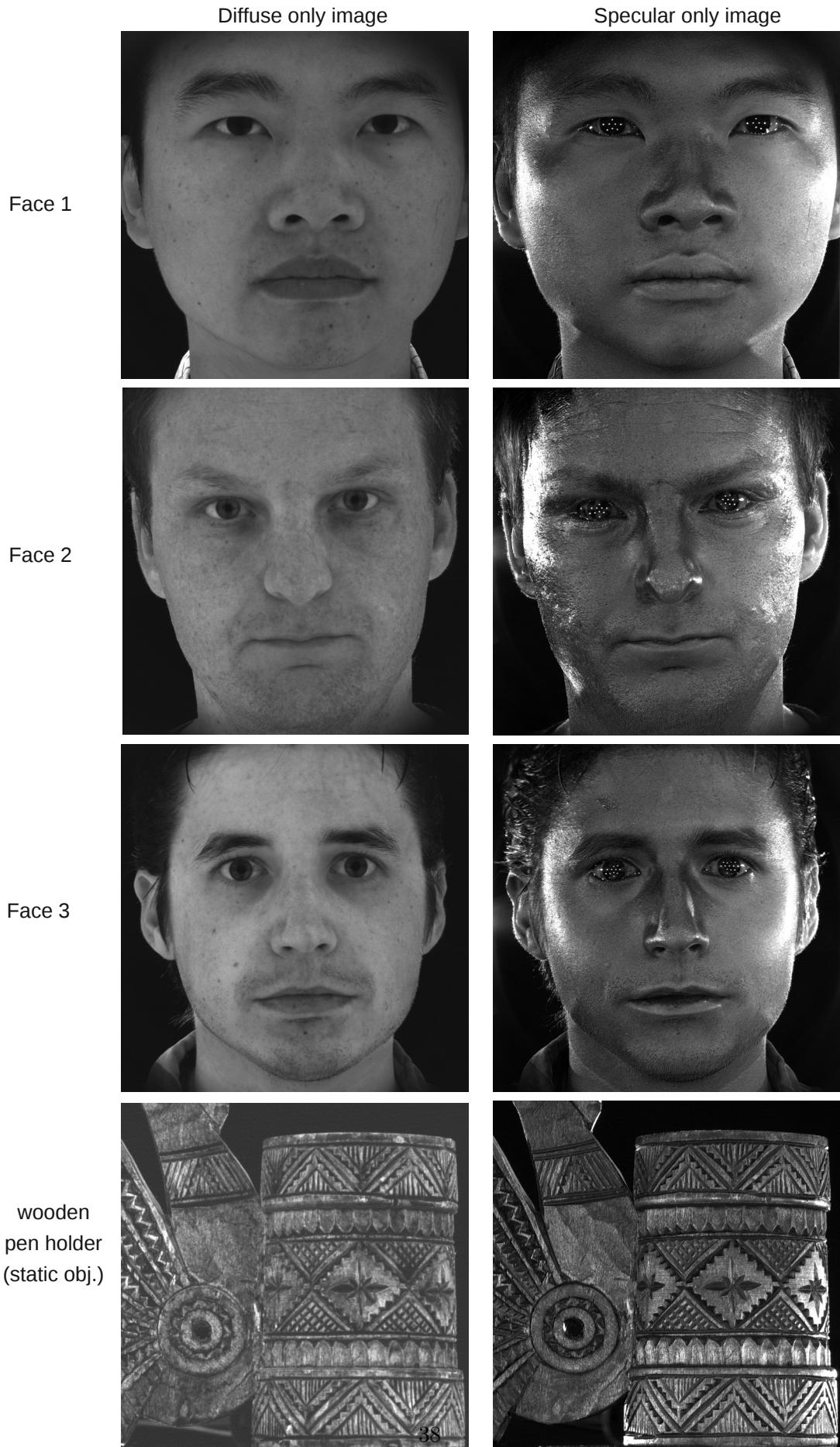


Fig. 3.17: Result of diffuse and specular reflectance separation

### 3.3 Extending the Basic Light Stage Design for Multispectral Capture

Most real world objects (human skin, fruits, etc..) are made up of multiple layers having different absorption and reflectance properties. These properties are very useful in Computer Graphics and Computer Vision research because they reveal the reflectance and absorption characteristics of underlying layers. The visible white light consists of radiation in the visible range (380nm to 720nm) and they have differential penetration depth in human skin with the red band (620 – 750 nm) going deepest. Hence, a multispectral image set — reflectance recorded at sparse set of bands in the visible range — contains the reflectance information from different layers of an object. For example: the Multispectral images of a fruit can reveal the properties of its inner layers. This information can be used to study the quality of a fruit [23] [34]. Similarly, parametric skin reflectance models like [11] rely on skin images captured at a set of narrow bands in the visible range. This information helps in creating a model of light interaction in different layers of the skin.

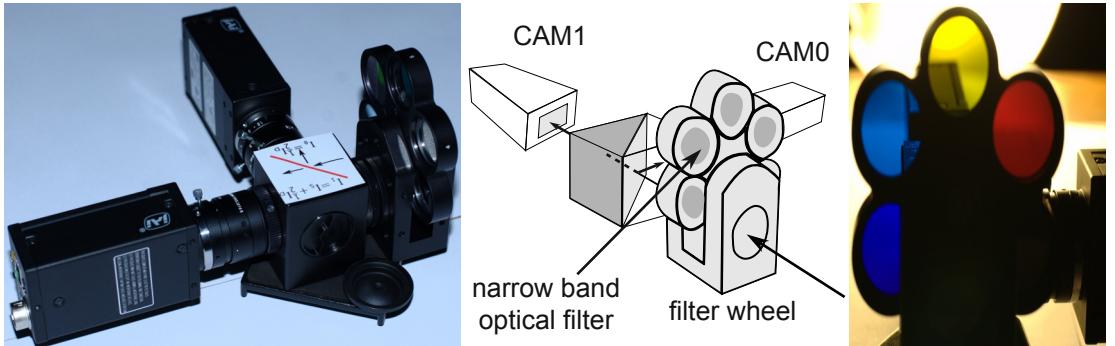


Fig. 3.18: (left and centre) Optical filter wheel attached to existing beam splitter setup for Multispectral acquisition. (right) Optical filters centred at different wavelength of the visible range in a filter wheel

We have made a modification in the light stage capture device proposed by Ma et al. [26] to allow capture of multispectral images. These images are captured in spherical illumination environment and are very useful for analysis of multiple layered objects because they do not contain shading information. We placed BrightLine single bandpass optical filters in front of our existing beam splitter setup (discussed in 3.2.2) to allow simultaneous capture of cross polarized multispectral images. Six filters are mounted to a filter wheel (as shown in Fig.

3.18) which snaps into preset position when the filter wheel is rotated. Manual rotation of the filter wheel increases the total capture time to  $\sim 12$  seconds. Stepper motor driven filter wheel<sup>10</sup> (access time of  $\sim 650$ ms) can be used to reduce the capture time to  $\sim 2$ sec. However, these electronic filter wheels are very expensive and hence we choose to use the manual filter wheel.

The specification and transmission spectrum of 6 single bandpass filters used in our Multispectral light stage is given in Table 3.2 and Fig. 3.19 respectively. These optical filters are polarization preserving: a property critical for simultaneous acquisition of cross polarized multispectral images using our existing beam splitter setup. The Semrock Brightline ® single bandpass filters<sup>11</sup> have the polarization preserving property with more than 90% transmission in the pass band. Such high transmission property is crucial for our setup because we lose more than 80% of the light source emission due to linear polarizer and the beam splitter. The center wavelength of these filters were chosen to sample the most significant points in the chromophore absorption curve obtained by [11, Fig.7]. Hence, this filter set targets the subsurface reflectance characteristics of human skin. The result of multispectral capture is shown in Fig. 3.20.

The diffuse Multispectral images clearly show the effect of absorption by in multiple skin layers. For example, the 655 nm diffuse image in Fig. 3.20 do not show freckle and moles which are visible in other bands of the multispectral diffuse image set. It is the subject of future work to use these multispectral images to recover parameters of a skin reflectance model like [11].

Table 3.2: Single bandpass filters used for the Multispectral light stage

Filter	Center Wavelength (nm)	Bandwidth (nm)	Average Transmission (%)
FF01-407/17-25	407	17	> 90
FF01-434/17-25	434	17	> 90
FF01-445/20-25	445	20	> 93
FF01-497/16-25	497	16	> 90
FF01-576/10-25	576	10	> 90
FF01-655/15-25	655	15	> 90

<sup>10</sup>[http://www.thorlabs.de/NewGroupPage9.cfm?ObjectGroup\\_ID=988](http://www.thorlabs.de/NewGroupPage9.cfm?ObjectGroup_ID=988)

<sup>11</sup><http://www.semrock.com/Catalog/Category.aspx?CategoryID=27>

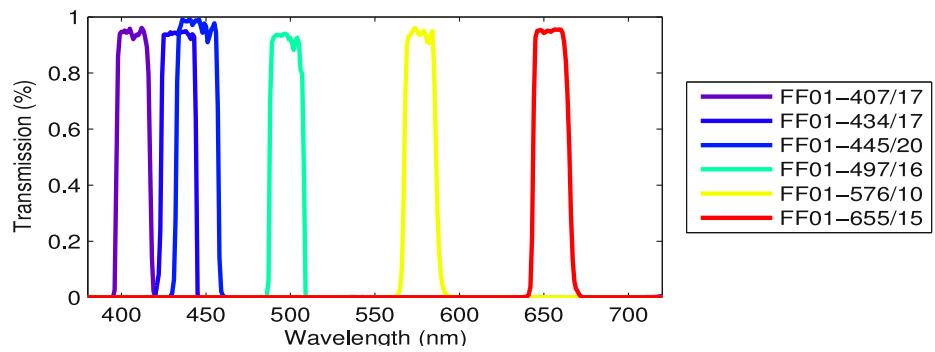


Fig. 3.19: Transmission spectrum of single-bandpass optical filters used for the Multispectral light stage

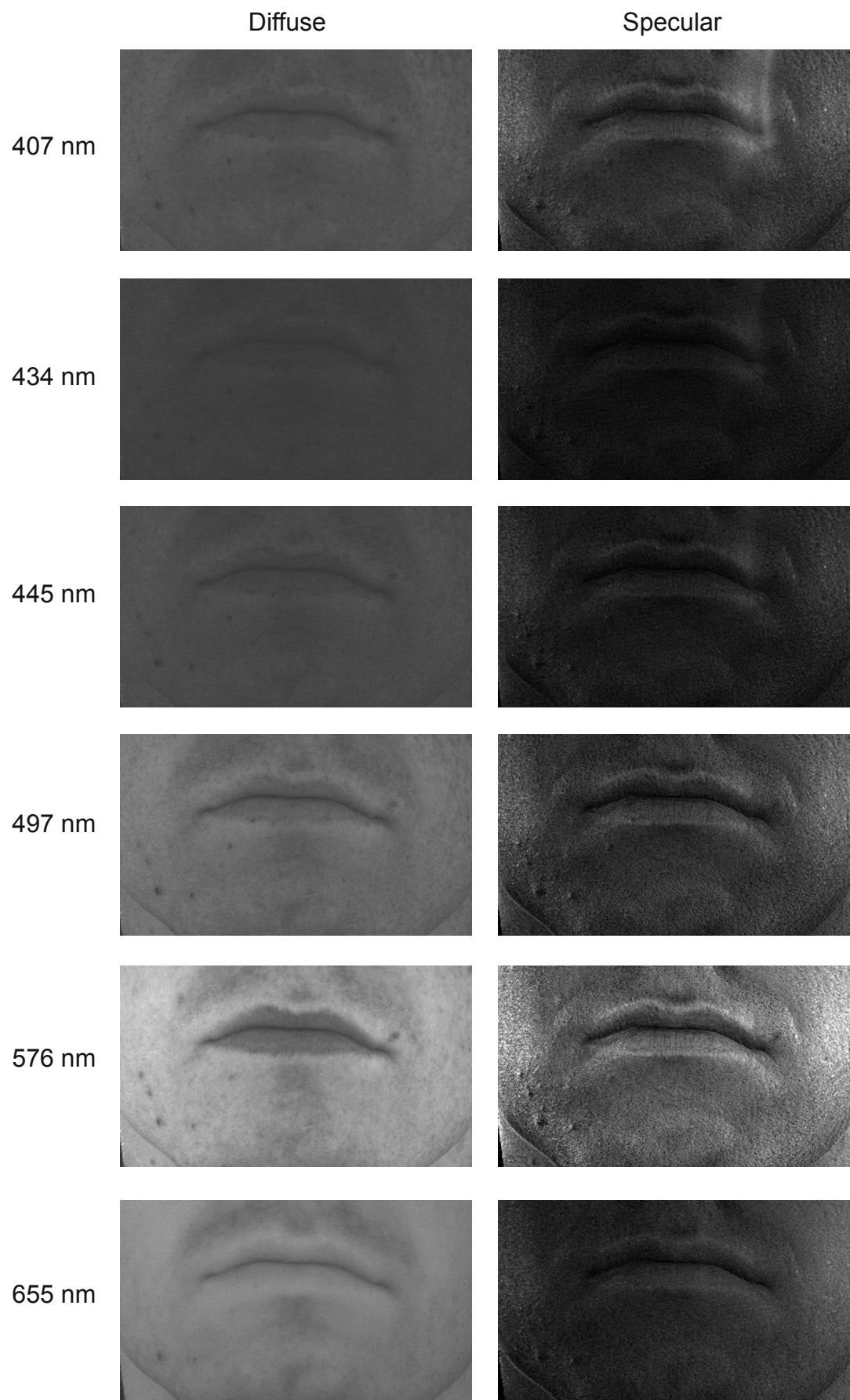


Fig. 3.20: Multispectral diffuse and specular image set for a face region

## Chapter 4

# Multispectral Light Stage Data Processing

Ma’s original Light Stage [26] allows capture of spherical gradient and constant illumination images in full visible spectrum. These images can be used to recover high resolution surface geometry using the spherical gradient photometric stereo technique. Addition of single bandpass optical filters to the existing capture device setup of the Light Stage allows capture of multispectral spherical gradient and constant illumination images at six narrow bands in the visible spectrum. These multispectral images capture the reflectance properties of multi-layered materials like human skin. Such multispectral reflectance maps can be used with parametric skin reflectance models like [11].

In this chapter, we first discuss the theoretical background of the spherical gradient photometric stereo method of Ma et al. [26]. This method assumes perfect registration (or alignment) of all the gradient images being used for computation of photometric normal. For a non-static object like a human face, it is not possible to remain still during capture of all the four ([26]) or six ([37]) gradient images. To correct for motion during the capture process, we discuss the “Joint Photometric Alignment” method proposed by Wilson et al. [37]. Using our modified radiance equations, we explore a Quadratic Programming (QP) based normal correction algorithm for surface geometry recovered using spherical gradient photometric stereo. Finally, based on our analysis of modified radiance equations, we propose a method to compute photometric normals using minimal four image set consisting of  $(X, Y, Z, \{\bar{X}, \bar{Y}, \bar{Z}\})$ .

We also show that the proposed method has the improved robustness property of [37] and reduced data capture requirement benefit of [26].

## 4.1 Spherical Gradient Photometric Stereo using Diffuse Images

For every image point (i.e. pixel) in a diffuse image, we define a local coordinate frame  $[\vec{u}, \vec{v}, \vec{n}]$  such that the  $\vec{n}$  axis aligns with the surface normal of the surface patch corresponding to that image point as shown in Fig. 4.1. We use primed symbols, i.e.  $\omega'$ , to represent vectors in a local coordinate frame. The axes of local coordinate frame  $[\vec{u}, \vec{v}, \vec{n}]$  can be defined in terms of

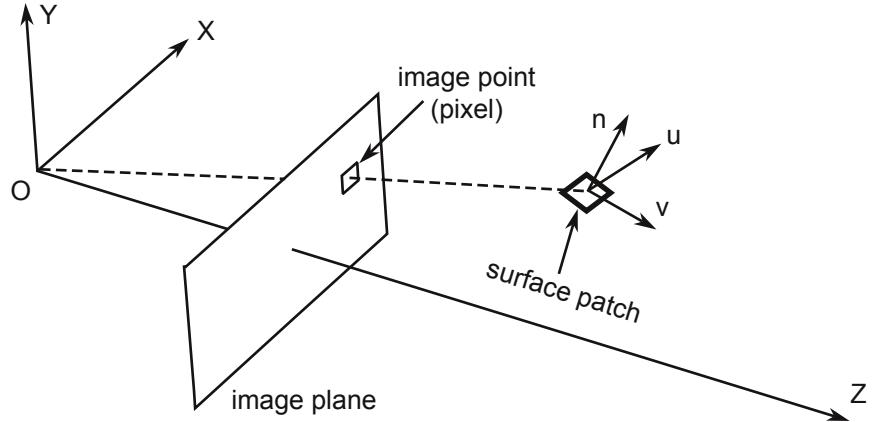


Fig. 4.1: Global  $(X, Y, Z)$  and local coordinate  $(u, v, n)$  frame for diffuse images

the global coordinate frame  $[O, X, Y, Z]$  as

$$\begin{aligned}\vec{u} &= (u_x i + u_y j + u_z k), \\ \vec{v} &= (v_x i + v_y j + v_z k), \\ \vec{n} &= (n_x i + n_y j + n_z k).\end{aligned}$$

Let us also define  $\omega' = (\omega'_u, \omega'_v, \omega'_n)$  as the spherical direction in local coordinates such that the corresponding global coordinates are given by

$$\omega = (\omega'_u u_x + \omega'_v v_x + \omega'_n n_x) i + (\omega'_u u_y + \omega'_v v_y + \omega'_n n_y) j + (\omega'_u u_z + \omega'_v v_z + \omega'_n n_z) k.$$

For any Lambertian surface, the value of radiance under spherical illumination is given by

$$r = \int_{\Omega} P(\omega) R(\omega, n) d\omega = \int_{\Omega} P(\omega') R(\omega', [0, 0, 1]) d\omega', \quad (4.1)$$

where  $P(\omega)$  and  $P(\omega')$  represent the intensity of light incoming from direction  $\omega$  (global coordinate) and  $\omega'$  (local coordinate) respectively and  $R(\omega, n)$  is the Lambertian Bidirectional Reflectance Distribution Function (BRDF). Recall that both  $\omega$  and  $\omega'$  represent the same physical direction but in different coordinate frames. In (4.1) we have substituted  $[0, 0, 1]$  vector for the normal because the  $n$  axis of local coordinate frame aligns with the surface normal.

In the case of X-gradient spherical illumination, the intensity of light incident from direction  $\omega' \in \Omega$  is proportional to the X-component of  $\omega \in \Omega$  (the corresponding incident direction represented in global coordinate).

$$P(\omega') = P_x(\omega) = (\omega'_u u_x + \omega'_v v_x + \omega'_n n_x) \in [-1, 1] \quad (4.2)$$

As it is not possible to emit light with negative value of intensity, we cannot realize an X-gradient illumination with  $P(\omega') \in [-1, 1]$ . Hence, we rescale as follows:

$$P(\omega') = \frac{P_x(\omega) + 1}{2} = \frac{(\omega'_u u_x + \omega'_v v_x + \omega'_n n_x) + 1}{2} \in [0, 1]. \quad (4.3)$$

#### 4.1.1 Radiance Equation for Gradient Illumination

Substituting (4.3) in (4.1), we can write the radiance equation for X-gradient illumination as:

$$r_x = \int_{\Omega} \left( \frac{\omega'_u u_x + \omega'_v v_x + \omega'_n n_x + 1}{2} \right) R(\omega', [0, 0, 1]) d\omega'. \quad (4.4)$$

Both Ma et al. [26] and Wilson et al. [37] assumed that the surface is convex and that the diffuse reflectance is symmetric about the surface normal. Hence, the integral over the

hemisphere along  $u_x$  and  $v_x$  axes becomes 0 and the gradient radiance simplifies to

$$\begin{aligned} r_x &= \frac{1}{2} \left\{ n_x \int_{\Omega} \omega'_n R(\omega', [0, 0, 1]) d\omega' + \int_{\Omega} R(\omega', [0, 0, 1]) d\omega' \right\}, \\ &= \frac{\pi \rho_D}{2} \left\{ n_x \int_0^1 \omega'_n \omega'_n d\omega' + \int_0^1 \omega'_n d\omega' \right\}, \\ r_x &= \frac{\pi \rho_D}{2} \left\{ \frac{1}{3} n_x + \frac{1}{2} \right\}, \end{aligned} \quad (4.5)$$

where  $\rho_D$  is the diffuse albedo. In a similar way, we can arrive at the following equations for diffuse radiance in Y and Z gradient illumination:

$$r_y = \frac{\pi \rho_D}{2} \left\{ \frac{1}{3} n_y + \frac{1}{2} \right\}, \quad (4.6)$$

$$r_z = \frac{\pi \rho_D}{2} \left\{ \frac{1}{3} n_z + \frac{1}{2} \right\}. \quad (4.7)$$

#### 4.1.2 Radiance Equation for Constant Illumination

For ideal constant spherical illumination, the intensity of light incident from all the possible spherical directions is a constant, i.e.

$$P(\omega') = 1 \quad \text{for all } \omega' \in \Omega.$$

Thus, the expression for radiance under constant illumination becomes:

$$\begin{aligned} r_c &= \int_{\Omega} R(\omega', [0, 0, 1]) d\omega' \\ &= \int_{-1}^1 (\pi \rho_D) \max(0, \omega' \cdot [0, 0, 1]) d\omega' \\ &= \pi \rho_D \int_0^1 \omega'_n d\omega', \\ r_c &= \frac{\pi \rho_D}{2}. \end{aligned} \quad (4.8)$$

It is evident from (4.8) that the constant illumination image is used to recover the diffuse albedo  $\rho_D$ .

### 4.1.3 Surface Normal Estimation

Ma et al. [26] used the ratio of gradient images to the constant illumination image to recover high resolution surface geometry of the surface visible in the gradient images. Hence, the ratio of (4.5),(4.6), (4.7) to (4.8) results in:

$$\begin{aligned} n_x &= \frac{1}{N_d} \left( \frac{r_x}{r_c} - \frac{1}{2} \right), \\ n_y &= \frac{1}{N_d} \left( \frac{r_y}{r_c} - \frac{1}{2} \right), \\ n_z &= \frac{1}{N_d} \left( \frac{r_z}{r_c} - \frac{1}{2} \right), \end{aligned}$$

where,  $N_d$  is a normalizing constant given by

$$N_d = \sqrt{\left( \frac{r_x}{r_c} - \frac{1}{2} \right)^2 + \left( \frac{r_y}{r_c} - \frac{1}{2} \right)^2 + \left( \frac{r_z}{r_c} - \frac{1}{2} \right)^2}. \quad (4.9)$$

## 4.2 Spherical Gradient Photometric Stereo using Specular Only Images

For the analysis of specular radiance, let us define global  $[X, Y, Z]$  and local coordinate  $[\vec{s}, \vec{t}, \vec{u}]$  frames as shown in Fig. 4.2.  $v_i$  represents the view vector and  $v_r$  is the reflected direction of view vector which is obtained by  $180^\circ$  rotation of  $v_i$  around the surface normal  $n$ . The local coordinate frame  $[\vec{s}, \vec{t}, \vec{u}]$  for every image point (i.e. pixel) in a specular image is defined such that  $u$  axis aligns with the reflected direction of view vector  $v_r$  and the orthogonal axes  $\vec{s}, \vec{t}$  are orthogonal to  $u$  axis. The axes of local coordinate frame  $[\vec{s}, \vec{t}, \vec{u}]$  can be defined in terms of the global coordinate frame  $[O, X, Y, Z]$  as:

$$\begin{aligned} \vec{s} &= (s_x i + s_y j + s_z k), \\ \vec{t} &= (t_x i + t_y j + t_z k), \\ \vec{u} &= (u_x i + u_y j + u_z k). \end{aligned}$$

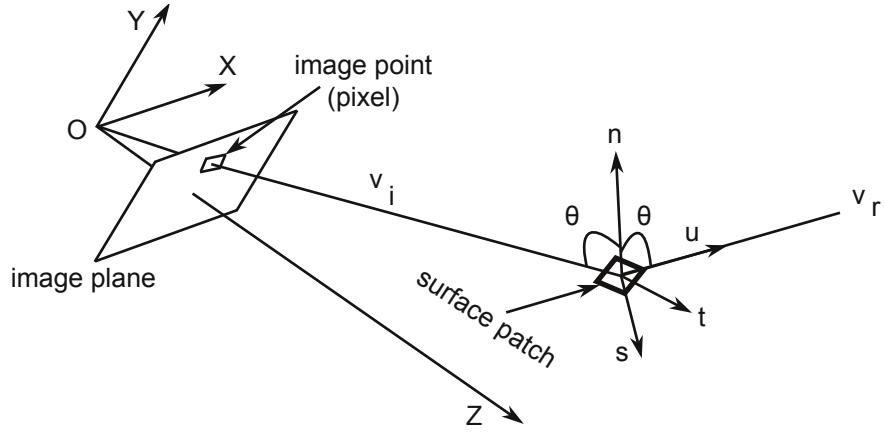


Fig. 4.2: Global ( $X, Y, Z$ ) and local coordinate ( $u, v, n$ ) frame for specular images

Let us also define  $\omega' = (\omega'_s, \omega'_t, \omega'_u)$  as the spherical direction in local coordinate frame such that the corresponding global coordinate frame direction is given by

$$\omega = (\omega'_s s_x + \omega'_t t_x + \omega'_u u_x)i + (\omega'_s s_y + \omega'_t t_y + \omega'_u u_y)j + (\omega'_s s_z + \omega'_t t_z + \omega'_u u_z)k.$$

The value of specular radiance under spherical illumination is given by

$$r = \int_{\Omega} P(\omega) R(\omega, v_i, n) d\omega = \int_{\Omega} P(\omega') R(\omega', v'_i, n') d\omega', \quad (4.10)$$

where,  $P(\omega)$  and  $P(\omega')$  represent the intensity of light incoming from direction  $\omega$  (global coordinate) and  $\omega'$  (local coordinate) respectively and  $R(\omega, v_i, n)$  is the specular Bidirectional Reflectance Distribution Function (BRDF). Recall that both  $\omega$  and  $\omega'$  represent the same physical direction but in different coordinate frames. The specular BRDF can be expressed as

$$R(\omega, v_i, n) = S(r, v_i, n)\Psi(\omega, n), \quad (4.11)$$

where,  $r = 2(n \cdot \omega)n - \omega$  is the perfect specular reflected direction,  $S$  is the specular reflectance lobe which is non-zero around a small solid angle around  $r$  and  $\Psi$  is the foreshortening factor.

#### 4.2.1 Radiance Equation for Gradient Illumination

In the case of X-gradient spherical illumination, the intensity of light incident from direction  $\omega' \in \Omega$  is proportional to the X-component of  $\omega \in \Omega$  (the corresponding incident direction

represented in global coordinate), i.e.

$$P(\omega') = P_x(\omega) = (\omega'_s s_x + \omega'_t t_x + \omega'_u u_x) \in [-1, 1]. \quad (4.12)$$

As it is not possible to emit light with negative value of intensity, we cannot realize an X-gradient illumination with  $P(\omega') \in [-1, 1]$ . Hence, we rescale as follows:

$$P(\omega') = \frac{P_x(\omega) + 1}{2} = \frac{(\omega'_s s_x + \omega'_t t_x + \omega'_u u_x) + 1}{2} \in [0, 1]. \quad (4.13)$$

Substituting (4.13) and (4.11) in (4.10), we can write the radiance equation for X-gradient illumination as:

$$r_x = \int_{\Omega} \left( \frac{\omega'_s s_x + \omega'_t t_x + \omega'_u u_x + 1}{2} \right) S(r', v'_i, n') \Psi(\omega', n') d\omega' \quad (4.14)$$

where the superscript ' is added to represent coordinates in local coordinate frame. Ma et al. [25, p27] assumed the foreshortening factor  $\Psi$  to be constant (say  $c_f$ ). This assumption is not valid for glossy reflections (i.e. the specular lobe  $S$  is non-zero around a large solid angle around  $r$ ) and surface patches that lie at grazing angle with respect to the viewer (i.e.  $v_i \sim 90^\circ$ ).

The ideal specular lobe  $S$  is symmetric along the  $u$  axis and hence the first two terms involving  $s_x$  and  $t_x$  in (4.14) become zero resulting in

$$r_x = \frac{1}{2} \left\{ u_x \int_{\Omega} \omega'_u S(r', v'_i, n') \Psi(\omega', n') d\omega' + \int_{\Omega} S(r', v'_i, n') \Psi(\omega', n') d\omega' \right\}. \quad (4.15)$$

In a similar way, we can arrive at the following equations for specular radiance in Y and Z gradient illumination:

$$r_y = \frac{1}{2} \left\{ u_y \int_{\Omega} \omega'_u S(r', v'_i, n') \Psi(\omega', n') d\omega' + \int_{\Omega} S(r', v'_i, n') \Psi(\omega', n') d\omega' \right\}, \quad (4.16)$$

$$r_z = \frac{1}{2} \left\{ u_z \int_{\Omega} \omega'_u S(r', v'_i, n') \Psi(\omega', n') d\omega' + \int_{\Omega} S(r', v'_i, n') \Psi(\omega', n') d\omega' \right\}. \quad (4.17)$$

### 4.2.2 Radiance Equation for Constant Illumination

For ideal constant spherical illumination, the intensity of light incident from all the possible spherical directions is a constant, i.e.

$$P(\omega') = 1 \quad \text{for all } \omega' \in \Omega.$$

Thus, the expression for specular radiance under constant illumination becomes

$$r_c = \int_{\Omega} S(r', v'_i, n') \Psi(\omega', n') d\omega'. \quad (4.18)$$

### 4.2.3 Surface Normal Estimation

It is evident from (4.15), (4.16), (4.17) and (4.18) that we can recover the reflected direction of the view vector  $v_r = u(u_x, u_y, u_z)$  by subtracting the constant illumination specular image from the X, Y and Z gradient illumination specular image followed by normalization. Mathematically,

$$\begin{aligned} u_x &= \frac{1}{N_s} (r_x - \frac{1}{2} r_c), \\ u_y &= \frac{1}{N_s} (r_y - \frac{1}{2} r_c), \\ u_z &= \frac{1}{N_s} (r_z - \frac{1}{2} r_c), \end{aligned} \quad (4.19)$$

where,  $N_s = \sqrt{(r_x - \frac{1}{2} r_c)^2 + (r_y - \frac{1}{2} r_c)^2 + (r_z - \frac{1}{2} r_c)^2}$  is a normalizing constant. The half way vector between view vector  $v_i = [0 \ 0 \ -1]^T$  and the reflected direction of view vector corresponds to the surface normal and is given by

$$\vec{n} = \frac{1}{\bar{N}} (v_r + v_i), \quad (4.20)$$

where,  $\bar{N}$  is a normalizing constant and  $v_r = u$ .

The specular normal map is able to capture fine surface details because unlike diffuse radiance — which is a subsurface phenomena — specular reflection is a surface phenomena. A fine structure due to white paint on the nose tip of a white cement statue is revealed in the specular normal map of Fig. 4.4 (top right) while the diffuse normal map (top left) does not

capture this fine detail. The constant foreshortening factor  $F$  assumption of [25, p27] breaks down at grazing angle as revealed by large amount of noise in the boundary of face and both sides of the nose bridge in the specular normal map of Fig. 4.4 (top right).

### 4.3 Analysis of the Normalizing Constant Value - $N_d$ and $N_s$



Fig. 4.3: Centroid (depicted by small white circle) of diffuse and specular reflectance lobe

The vector along the direction of diffuse and specular lobe centroids (as shown by white circle in Fig. 4.3) can be converted to a unit vector by normalization: operation in which a vector is divided by its magnitude (also called normalization constant or  $\ell^2$ -norm ) to obtain a unit vector in its direction. The expression for normalization of diffuse and specular centroid is given by:

$$N_d = \frac{1}{r_c^d} \sqrt{\left(r_x^d - \frac{1}{2} r_c^d\right)^2 + \left(r_y^d - \frac{1}{2} r_c^d\right)^2 + \left(r_z^d - \frac{1}{2} r_c^d\right)^2},$$

$$N_s = \sqrt{\left(r_x^s - \frac{1}{2} r_c^s\right)^2 + \left(r_y^s - \frac{1}{2} r_c^s\right)^2 + \left(r_z^s - \frac{1}{2} r_c^s\right)^2}.$$

Note that these expressions for  $N_d$  and  $N_s$  are same as (4.9) and (4.19) with only the superscript  $d$  and  $s$  added to depict the diffuse and specular radiance values. The unit vector along diffuse and specular centroid direction correspond to the surface normal and reflected direction of the view vector respectively.

The normalizing constant values  $N_d$  and  $N_s$  are proportional to the size of diffuse and specular reflectance lobes respectively. Hence, for a typical diffuse surface, we would expect  $N_s < N_d$  to hold true. The distribution of  $N_d$  and  $N_s$  for a white cement plaster statue (diffuse object) shown in Fig. 4.4 supports this hypothesis. The distribution of diffuse normalizing constant  $N_d$  value reveals another interesting fact. Most of the  $N_d$  values are clustered in (0.37, 0.41) region. We need to investigate deeper into the nature of ideal diffuse reflectance

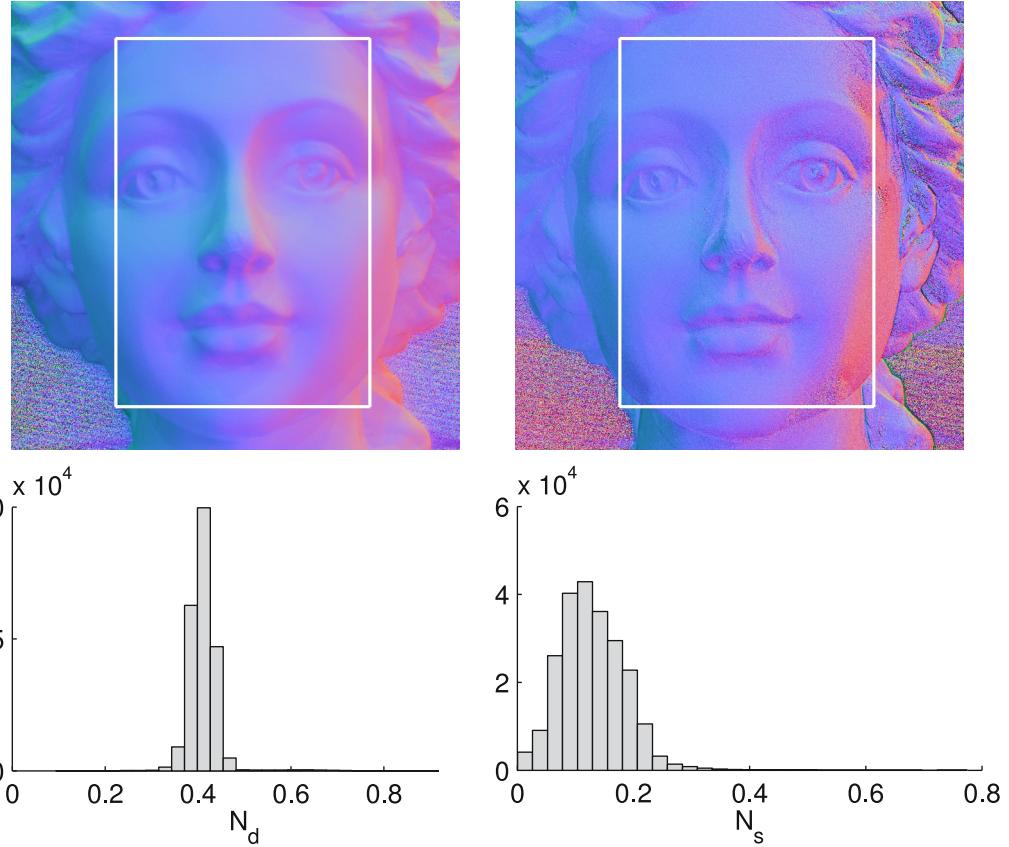


Fig. 4.4: Distribution of diffuse  $N_d$  (bottom left) and specular  $N_s$  (bottom right) normalizing constant value for a region (depicted with white rectangle) in the diffuse (top left) and specular (top right) normal map of a white cement plaster statue.

centroid to be able to explain this behaviour. Let us consider an ideal diffuse reflectance lobe symmetric along  $n$  axis and stretching  $k$  units along this axis of the local coordinate frame  $[u, v, n]$  as shown in Fig. 4.3. Such diffuse reflectance lobe can be defined by the solution for  $f(u, v, n) = 0$ , where the function  $f(u, v, n)$  is defined as:

$$f(u, v, n) = u^2 + v^2 + n^2 - k^2 \quad n \in [0, k].$$

As the diffuse reflectance function is symmetric about  $n$  axis, its centroid is given by

$$(u_0, v_0, n_0) = \left( 0, 0, \frac{\int n f(u, v, n) dn}{\int f(u, v, n) dn} \right) = \left( 0, 0, \frac{3(k^2 - 2k)}{4(k^2 - 3)} \right).$$

So, for a unit diffuse reflectance lobe (i.e.  $k = 1$ ), the centroid lies at  $(0, 0, 0.375)$  and the values of diffuse normalizing constant is  $N_d = 0.375$ . Hence, the distribution of  $N_d$  in Fig. 4.4 confirms that the white cement plaster surface has reflecting properties that are close to an ideal diffuse surface.

For a wide variety of real world surfaces, the diffuse and specular reflectance lobes get distorted due to inter-reflection, ambient occlusion and coarse approximation of spherical illumination due to light discretization. This causes the diffuse and specular lobe centroids to shift away from its ideal position on the surface normal and reflected direction of the view vector respectively. For these reasons, the normalizing constant values cannot be used to infer the nature of diffuse and specular reflectance lobes. For example, it is possible for a completely distorted diffuse lobe to acquire centroid value of a unit diffuse lobes (i.e.  $(0, 0, 0.375)$ ). Hence, although the value of normalizing constant is a good measure of the reflecting properties of a surface, it cannot be used to quantify the nature of distortion in the reflectance lobes. Analysis of the normalizing constant values provides a good insight into the basis of spherical gradient photometric stereo technique and its limitations.

## 4.4 Quadratic Programming based Normal Correction

Quality of surface geometry recovered using spherical gradient photometric stereo [26] is affected by the extent to which the following assumptions are satisfied:

1. no shadowing of light sources i.e. object is convex
2. no inter-reflection i.e. light incident on a surface patch is solely due to light source and not because of reflections from nearby surface patches
3. light sources closely approximate a continuous illumination environment i.e. effect of “*light discretization*”<sup>1</sup> is minimal



Fig. 4.5: Deformed diffuse (left) and specular (right) lobes due to inter-reflection, ambient occlusion and coarse approximation of spherical illumination

In this section, we will introduce new parameters to the original Ma et al. [26] radiance equations in order to quantify the extent of violation of these three assumptions. These modified radiance equations not only helps uncover the limitations of Ma et al. method, but also provide insight into possible modifications of this technique in order to improve the quality of recovered surface geometry. Using these modified equation, we show why the quality of normal estimated by Ma et al. method degrades with deformed diffuse lobe. We also propose a Quadratic Programming (QP) based normal correction technique to compensate for the effects of deformed diffuse lobes and hence improve the quality of recovered surface normals. Finally, based on analysis of our modified radiance equations, we propose a minimal image sets method for spherical gradient photometric stereo which has the improved robustness property of Wilson et al. [37] and reduced data capture requirement benifit of Ma et al. [26].

Here, we present an analysis of diffuse lobes deformation only because similar approach can be used to analyse the effects of deformed specular lobes.

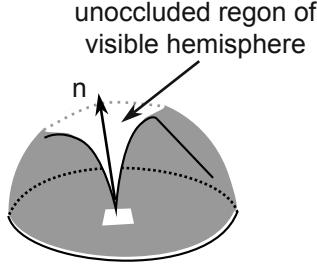


Fig. 4.6: Ambient occlusion in concave surfaces

#### 4.4.1 Modified Radiance Equations for Gradient Illumination

Ambient occlusion limits the portion of hemisphere visible to a surface patch as shown in Fig. 4.6. Hence, to quantify the effect of ambient occlusion at an arbitrary surface patch  $p$  in spherical illumination images, we introduce the following binary visibility function:

$$V_{p,\omega'} = \begin{cases} 1 & \text{if direction } \omega' \text{ is unoccluded,} \\ 0 & \text{otherwise.} \end{cases} \quad (4.21)$$

We can now rewrite the X gradient radiance equation of (4.4) as:

$$r_x = \int_{\Omega} V_{p,\omega'} \left( \frac{\omega'_u u_x + \omega'_v v_x + \omega'_n n_x + 1}{2} \right) R(\omega', [0, 0, 1]) d\omega'. \quad (4.22)$$

Inter-reflections and coarse approximation of spherical illumination deforms the diffuse reflectance lobe along the  $u_x$  and  $v_x$  axes. Hence, contribution of the integrals along  $u_x$  and  $v_s$  axes cannot be ignored in the case of deformed diffuse lobe. In other words, the diffuse reflectance lobe is no more symmetric along the  $n_x$  axes. Ma et al. [26] and Wilson et al. [37] assumed a diffuse reflectance lobe symmetric along the  $n_x$  axis and therefore they were able to ignore the contribution of these integrals in their analysis.

We do not ignore the effect of asymmetry in diffuse reflectance lobe. However, as it is not possible to evaluate the integrals along  $u_x$  and  $v_x$ , we quantify the extent of distortion in diffuse reflectance lobe using a single scalar  $\delta'_x$  (distortion coefficient). This parameter scales the diffuse albedo  $\pi\rho_D$  to quantify the contribution of integrals along  $u_x$  and  $v_x$  axes in (4.22). In other words, we make the simplifying assumption that overall deformation in

---

<sup>1</sup>a term used by Ma et al. [26] to refer to coarse approximation of spherical illumination caused by LEDs attached to discrete positions on a twice subdivided icosahedron. It is important to realise that the term “light discretisation” does not imply that intensity of light sources is discrete.

the diffuse reflectance lobe for a gradient illumination environment can be quantified using a single parameter  $\delta'_x$  (distortion coefficient).

Adding this parameter to (4.22) gives:

$$r_x = \frac{\pi\rho_D}{2} \left\{ \delta'_x + n_x \int_0^1 \omega'_n V_{p,\omega'} \omega'_n d\omega' + \int_0^1 V_{p,\omega'} \omega'_n d\omega' \right\} \quad (4.23)$$

where,

$$\delta'_x(\pi\rho_D) = u_x \int_{\Omega} \omega'_u V_{p,\omega'} R(\omega', [0, 0, 1]) d\omega' + v_x \int_{\Omega} \omega'_v V_{p,\omega'} R(\omega', [0, 0, 1]) d\omega'$$

To simplify the evaluation of (4.23), we first consider the ideal case value for the visibility function i.e. when complete hemisphere is visible. In this ideal case,  $V_{p,\omega'} = 1$  for all  $\omega' \in \Omega$ , (4.23) simplifies to:

$$r_x = \frac{\pi\rho_D}{2} \left\{ \delta'_x + n_x \int_0^1 \omega'_n \omega'_n d\omega' + \int_0^1 \omega'_n d\omega' \right\} = \frac{\pi\rho_D}{2} \left\{ \delta'_x + \frac{1}{3}n_x + \frac{1}{2} \right\}. \quad (4.24)$$

For real world objects, the ideal case value of visibility function is not valid i.e.  $\exists \omega' \in \Omega : V_{p,\omega'} \neq 1$ . This implies that the actual value of two integrals in (4.23) will be less than their ideal case values i.e.

$$\int_0^1 \omega'_n V_{p,\omega'} \omega'_n d\omega' < \frac{1}{3} \quad \text{and} \quad \int_0^1 V_{p,\omega'} \omega'_n d\omega' < \frac{1}{2}.$$

To quantify the overall effect of shadowing, we define the ambient occlusion term  $V_p \in [0, 1]$  such that:  $V_p = 1$  when complete hemisphere is visible and  $V_p = 0$  for completely occluded hemisphere. The intermediate values  $0 < V_p < 1$  apply to partial occlusion. Substituting this visibility parameter in (4.24), we obtain the following expression for radiance from real world surfaces under X gradient illumination:

$$r_x = \frac{\pi\rho_D V_p}{2} \left\{ \delta_x + \frac{1}{3}n_x + \frac{1}{2} \right\}, \quad (4.25)$$

where,  $\delta'_x = V_p \delta_x$ . In a similar way, we can obtain the expression for radiance in Y and Z

gradients illumination

$$r_y = \frac{\pi\rho_D V_p}{2} \left\{ \delta_y + \frac{1}{3}n_y + \frac{1}{2} \right\}, \quad (4.26)$$

$$r_z = \frac{\pi\rho_D V_p}{2} \left\{ \delta_z + \frac{1}{3}n_z + \frac{1}{2} \right\}. \quad (4.27)$$

#### 4.4.2 Modified Radiance Equation for Constant Illumination

For constant spherical illumination, Ma et al. [26] assumed the intensity of light incident from all possible spherical directions to be unity, i.e.

$$P(\omega') = 1 \quad \text{for } \omega' \in \Omega.$$

This is true for ideal case spherical illumination. However, this assumption ignores: a) light source attenuation effects, which is equivalent to assuming all points on the object lie exactly at the centre of the light stage; b) contribution of inter-reflection and shadowing which can increase or decrease the intensity of light incident from a particular spherical direction. In (4.4.1), we introduced the binary visibility function  $V_{p,\omega'}$  which models whether a spherical direction  $\omega'$  is visible at any surface patch  $p$ . For a surface patch, the intensity light incident from a direction  $\omega'$  is dependent on the binary visibility function defined for that surface patch. Therefore, we can now define  $P(\omega')$  as:

$$P(\omega') = \begin{cases} c_{p,\omega'} & \text{if } V_{p,\omega'} = 1 \\ 0 & \text{otherwise,} \end{cases} \quad (4.28)$$

where,  $c_{p,\omega'}$  models the angular deviation of intensity under constant illumination for a surface patch  $p$ . As it is not possible to evaluate radiance integral using this definition of  $P(\omega')$ , we make the simplifying assumption that the intensity of incident light is unity when a spherical incident direction is visible from a surface patch. In other words, we also use the unit incident

intensity assumption of Ma et al. but only for visible spherical directions. Mathematically,

$$P(\omega') = \begin{cases} 1 & \text{if } V_{p,\omega'} = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (4.29)$$

This simplifying assumption ignores the light source attenuation effects and contribution of inter-reflection and only includes the contribution of shadowing effects under constant spherical illumination. Using this simplifying assumption, the expression for radiance under constant spherical illumination becomes:

$$\begin{aligned} r_c &= \int_{\Omega} P(\omega') V_{p,\omega'} R(\omega', [0, 0, 1]) d\omega' \\ &= \int_{-1}^1 V_{p,\omega'} (\pi \rho_D) \max(0, \omega' \cdot [0, 0, 1]) d\omega' \\ &= \pi \rho_D \int_0^1 V_{p,\omega'} \omega'_n d\omega' = \frac{\pi \rho_D V_p}{2}. \end{aligned} \quad (4.30)$$

#### 4.4.3 Quality of Surface Normal Estimated Using Original Spherical Gradient Photometric Stereo Method

The expression for computing photometric normal using Ma et al. [26] method is:

$$\vec{n} = \frac{(\frac{r_x}{r_c} - \frac{1}{2}, \frac{r_y}{r_c} - \frac{1}{2}, \frac{r_z}{r_c} - \frac{1}{2})}{\|(\frac{r_x}{r_c} - \frac{1}{2}, \frac{r_y}{r_c} - \frac{1}{2}, \frac{r_z}{r_c} - \frac{1}{2})\|} \quad (4.31)$$

Now, we use our modified radiance equations to represent the surface normal computed using the above method:

$$N_{\{x,y,z\}} = \frac{r_{\{x,y,z\}}}{r_c} - \frac{1}{2} = \delta_{\{x,y,z\}} + \frac{1}{3} n_{\{x,y,z\}}, \quad (4.32)$$

where,  $N_{\{x,y,z\}}$  is the unnormalized surface normal vector. It is evident from above expression that although the occlusion term ( $V_p$ ) cancel in this “ratio method”, the diffuse lobe distortion term  $\delta_{\{x,y,z\}}$  does not cancel out. Therefore, we conclude that the quality of surface normals computed using Ma et al. method will degrade with deformation in diffuse lobe. It is important to understand that the cancellation of occlusion term ( $V_p$ ) results from the following simplifying assumption used while evaluating the radiance equation for constant illumination:

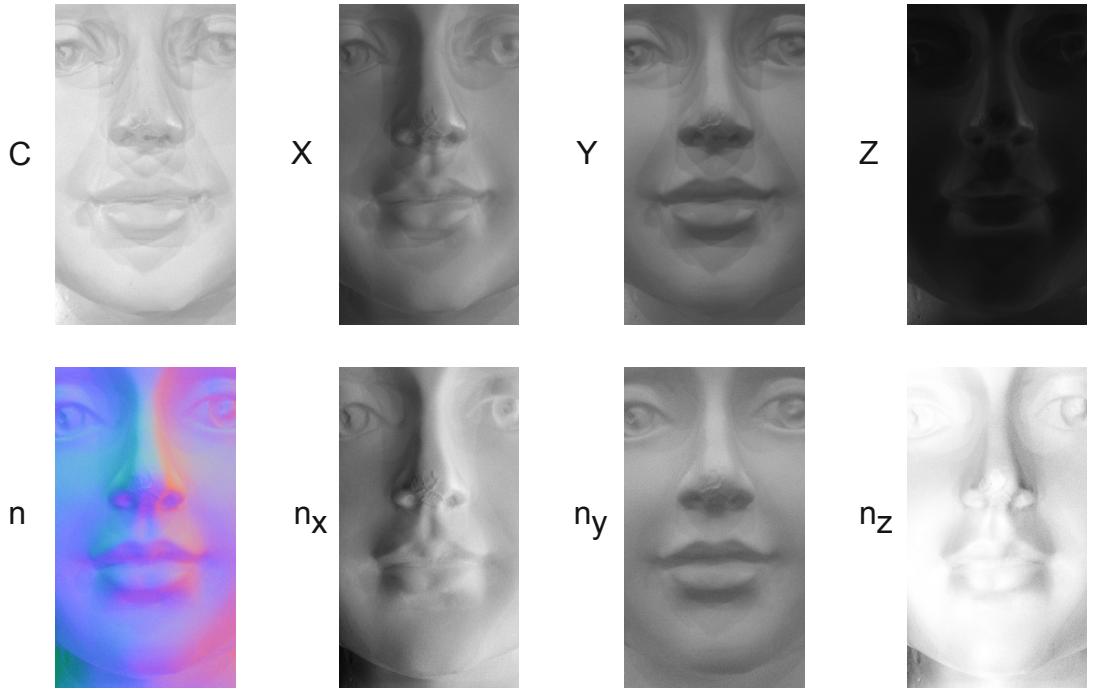


Fig. 4.7: Shadows clearly visible in the constant  $C$  and gradient ( $X, Y, Z$ ) images (top row) of a white cement plaster statue. The normal map (bottom leftmost - normal components mapped to R,G,B) and the  $X, Y, Z$  normal components depicted as grayscale image (bottom right three) do not show the effect of shadows.

the intensity of incident light is unity when a spherical incident direction is visible from a surface patch i.e.  $P(\omega') = 1$  if direction  $\omega'$  is unoccluded.

The modified radiance equations of (4.25, 4.26, 4.27 and 4.30) form an underdetermined system with 3 equations and 6 unknowns. In the next section, we explore the concept of complement image constraint in order to obtain additional constraints for this underdetermined system. This analysis will form the basis for our Quadratic Programming (QP) based normal correction.

#### 4.4.4 Modified Radiance Equations for Complement Gradient Illumination

Light Stage uses a reference coordinate frame  $[O, X, Y, Z]$  to setup gradient illumination. In addition to this gradient condition, complementary coordinate frame  $[O, \bar{X}, \bar{Y}, \bar{Z}]$  can also be used to setup complement gradient illumination environment. Here,  $O$  is the center of light stage and  $[\bar{X}, \bar{Y}, \bar{Z}]$  are the coordinate axes obtained by flipping  $[X, Y, Z]$  as shown in Fig. 4.8. Although the true surface normal  $n$  remains same in both coordinate frames, the distortion

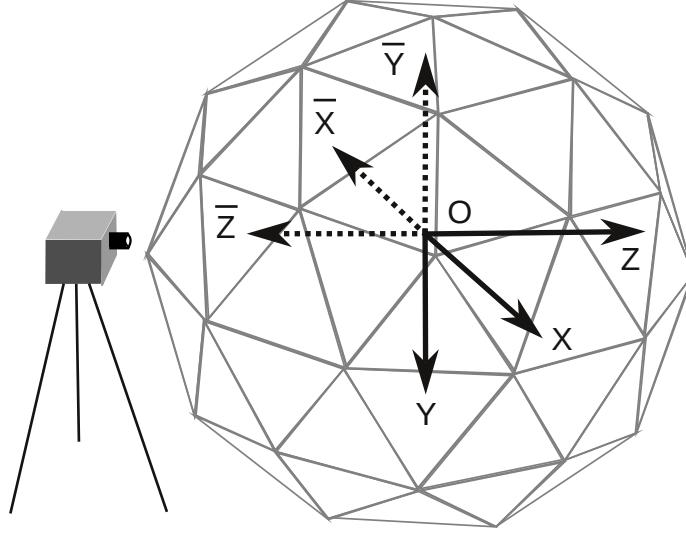


Fig. 4.8: Complement coordinate frames in a Light Stage

of diffuse reflectance lobe may not be identical. In other words, the distortion of diffuse lobe in complement gradient illumination contains some asymmetry with respect to the distortion observed in the gradient illumination. In the case of ideal spherical illumination and absence of ambient occlusion and inter-reflection, diffuse lobe distortion is symmetric in gradient and complement gradient illumination. To model the asymmetry present in complement gradient illumination, we represent the diffuse lobe distortion as the sum of symmetric distortion observed under gradient illumination ( $\delta_x$ ) and an asymmetric component ( $\delta_{\bar{x}}$ ). Therefore, we rewrite (4.25) in order to include asymmetry in diffuse lobe under the complement gradient illumination:

$$r_{\bar{x}} = \frac{\pi\rho_D V_p}{2} \left\{ \delta_x + \delta_{\bar{x}} + \frac{1}{3}n_{\bar{x}} + \frac{1}{2} \right\}, \quad (4.33)$$

where,  $\delta_{\bar{x}}$  is a scalar quantifying the amount of asymmetry (with respect to distortion in gradient illumination) in the distortion of diffuse lobe. Flipping the reference coordinate frame does not alter the true surface normal and therefore (4.33) can be rewritten as:

$$r_{\bar{x}} = \frac{\pi\rho_D V_p}{2} \left\{ \delta_x + \delta_{\bar{x}} - \frac{1}{3}n_x + \frac{1}{2} \right\}, \quad (4.34)$$

In a similar way, we can obtain the following expression for radiance under Y and Z complement gradient illumination:

$$r_{\bar{y}} = \frac{\pi\rho_D V_p}{2} \left\{ \delta_y + \delta_{\bar{y}} - \frac{1}{3}n_y + \frac{1}{2} \right\}, \quad (4.35)$$

$$r_{\bar{z}} = \frac{\pi\rho_D V_p}{2} \left\{ \delta_z + \delta_{\bar{z}} - \frac{1}{3}n_z + \frac{1}{2} \right\}. \quad (4.36)$$

The complement gradient images have also been used by Wilson et al. [37] to formulate an iterative algorithm (Joint Photometric Alignment) for estimation of optical flow of subject's motion during performance capture in a Light Stage.

#### 4.4.5 Correcting Recovered Surface Normals Using Quadratic Programming

In this section, we will explore a Quadratic Programming (QP) based correction of surface geometry recovered using spherical gradient photometric stereo method. From the analysis so far (section 4.4.1, 4.4.2 and 4.4.4), we have the following expressions for radiance under gradient and complement gradient spherical illumination:

$$r_{\{x,y,z\}} = \delta_{\{x,y,z\}} + \frac{1}{3}n_{\{x,y,z\}} + \frac{1}{2}, \quad (4.37)$$

$$r_{\{\bar{x},\bar{y},\bar{z}\}} = \delta_{\{x,y,z\}} + \delta_{\{\bar{x},\bar{y},\bar{z}\}} - \frac{1}{3}n_{\{x,y,z\}} + \frac{1}{2}. \quad (4.38)$$

From (4.37) and (4.38), we have 6 linear equations resulting in an underdetermined system in 9 unknowns  $\mathbf{x} = (\delta_x, \delta_y, \delta_z, \delta_{\bar{x}}, \delta_{\bar{y}}, \delta_{\bar{z}}, n_x, n_y, n_z)$  which can be expressed in matrix form as:

$$\mathbf{Ax} = \mathbf{b} \quad \text{where } \mathbf{A} \in \mathbf{R}^{6 \times 9}, \quad (4.39)$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & -1 & 0 & 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \frac{2}{3} \end{bmatrix}, \text{ and } b = \begin{bmatrix} \frac{r_x}{r_c} - \frac{1}{2} \\ \frac{r_y}{r_c} - \frac{1}{2} \\ \frac{r_z}{r_c} - \frac{1}{2} \\ \frac{r_x - r_{\bar{x}}}{r_c} \\ \frac{r_y - r_{\bar{y}}}{r_c} \\ \frac{r_z - r_{\bar{z}}}{r_c} \end{bmatrix}.$$

We apply a Quadratic Programming (QP) approach to perform correction to the surface normals computed using our minimal image sets method (or that computed using [37] or [26]). We regularize the problem such that QP computes new surface normals and estimates for distortion coefficients such that our linear system is satisfied and the new surface normals are closest to an initial solution. For example, if we take the diffuse centroid  $(n_x^{\text{Wil}}, n_y^{\text{Wil}}, n_z^{\text{Wil}})$  estimated by the method of Wilson et al. [37] and define  $\mathbf{x}_0 = (0, 0, 0, 0, 0, 0, n_x^{\text{Wil}}, n_y^{\text{Wil}}, n_z^{\text{Wil}})$ , then we can correct for deformed diffuse lobes by solving the following quadratic programming problem

$$\text{minimise } \|\mathbf{x} - \mathbf{x}_0\|^2 \quad \text{subject to } \mathbf{Ax} = \mathbf{b}. \quad (4.40)$$

### Results from QP based Surface Normal Correction

First, we analyse the results of QP based surface normal correction for a simple static object (white cylinder) because the captured gradient images are perfectly aligned and its ground truth data is known. Moreover, the simple convex surface of this object allows us to evaluate the performance of our QP based normal correction strategy.

When the initial solution to the QP based normal correction is the surface normal recovered using Ma et al. [26] method, the corrected normals tend to move towards the true surface normal as shown in Fig. 4.9 (left). The correction algorithm cannot recover true normals because we seek the corrections that are closest to the initial solution given by the Ma et al. [26] method. When the initial solution to the QP based normal correction is the surface normal recovered using Wilson et al. [37] method, the corrected normals tend to remain close to the initial solution as shown in Fig. 4.9 (right). This indicates that the surface normals recovered using Wilson et al. [37] are already close to the true surface normals.

Now, we analyse the results of QP based surface normal correction applied to the face

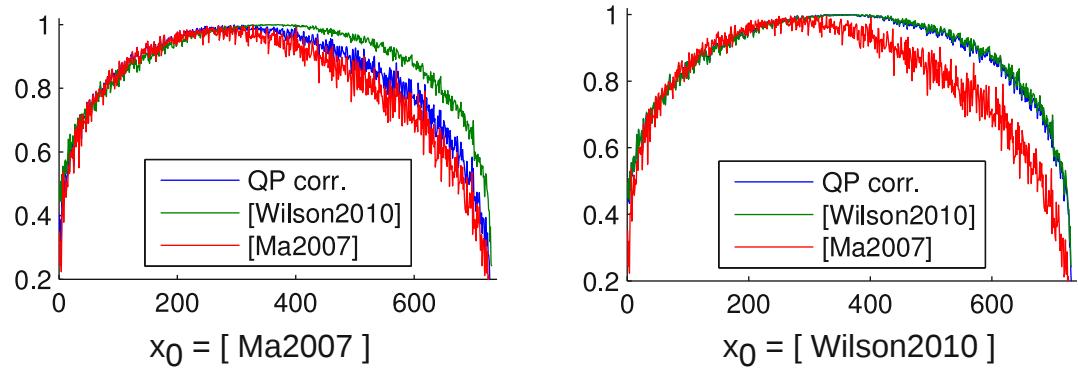


Fig. 4.9: Result of QP based normal correction applied to surface normals of a white cylinder when initial solution is (left)  $\mathbf{x}_0 = (0, 0, 0, 0, 0, 0, n_x^{\text{Ma}}, n_y^{\text{Ma}}, n_z^{\text{Ma}})$  and (right)  $\mathbf{x}_0 = (0, 0, 0, 0, 0, 0, n_x^{\text{Wil}}, n_y^{\text{Wil}}, n_z^{\text{Wil}})$ .



Fig. 4.10: (left) 1 pixel wide verticle region and (right) 1 pixel wide horizontal region in gradient images of the face region of a statue selected for analysis of QP based normal correction. (center) Side view photograph of the statue's face region.

region of a statue made of white cement: a material that exhibits property very close to an ideal diffuse surface. First, let us consider the surface normals in a 1 pixel wide vertical region (Fig. 4.10 - left) as shown in Fig. 4.11. When the normals computed by Ma et al. method is the initial solution, the corrected normals tend to move closer towards the surface normals computed using Wilson et al. method. On the other hand, when the normal computed by Wilson et al. method is the initial solution, the corrections computed by QP is insignificant. This suggests that the normals computed using Wilson et al. method is already very close to satisfying the constraints i.e.  $\mathbf{Ax} = \mathbf{b}$ . Moreover, Fig. 4.11 clearly shows that the corrected normals retain the noise characteristics of the initial solution, irrespective of the choice of initial solution. This behaviour could be attributed to the fact that we apply QP based correction to each image pixel independently and therefore the noise characteristics is propagated to corrected normals. We obtain similar results for the 1 pixel wide horizontal region (Fig. 4.10 - right) as shown in Fig. 4.12. From Fig. 4.12 (top), it is evident that QP based correction is significant in the region (250 to 450 pixel region) where the initial solution had large deviation from the surface normals computed by Wilson et al. method.

We used MATLAB 7.9 (R2009b) implementation of Quadratic Programming, *qprog()*, running on Slackware 13.1-2-12 on 3 GHz Intel®Core2 Duo CPU for testing this normal correction algorithm. It takes around 1.78 hours to perform normal correction on a photometric normal map of size  $1624 \times 1236$ .

#### 4.4.6 Discussion

QP based normal correction algorithm provides insignificant improvement in the recovered surface geometry. However, it will be evident in the next section that this analysis is pivotal to the development of minimal image sets method for robust spherical gradient photometric stereo. The modified radiance equations, which resulted in QP based correction algorithm, not only reveal the limitations of original Ma et al. [26] method but also provide an explanation for the improvement in quality of surface normals recovered by the complement gradient method of Wilson et al. [37]. Furthermore, in the next section, we use this analysis to show that our proposed minimal image sets method combines the advantage of the original method of Ma et al. (reduced data capture requirement) with that of Wilson et al. (improved

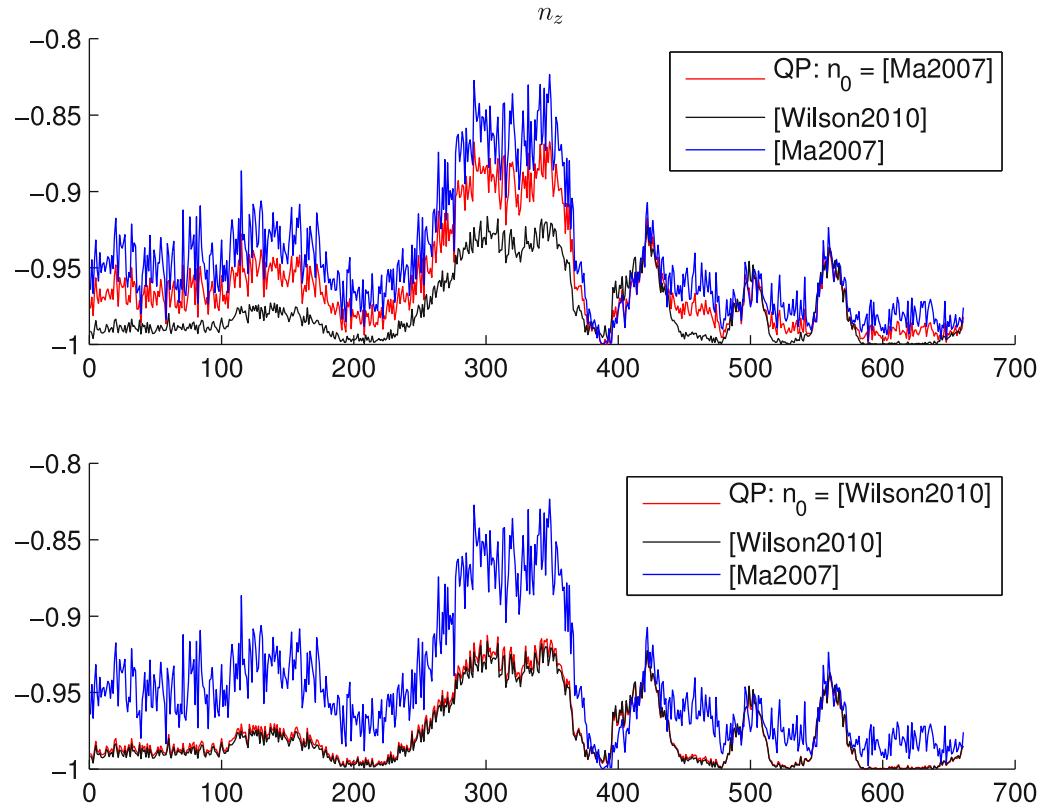


Fig. 4.11:  $n_z$  component of surface normals in 1 pixel wide selected vertical region (Fig. 4.10 - left) obtained after applying QP based normal correction with initial estimate of surface normals from (top) Ma et al. method i.e  $n_0 = n^{[Ma2007]}$  and (bottom) Wilson et al. method i.e.  $n_0 = n^{[Wilson]}$ .

robustness). Hence, although the QP based normal correction algorithm did not result in significant improvement over existing methods, it provided us with valuable insight into the limitations and strength of the spherical gradient photometric stereo technique.

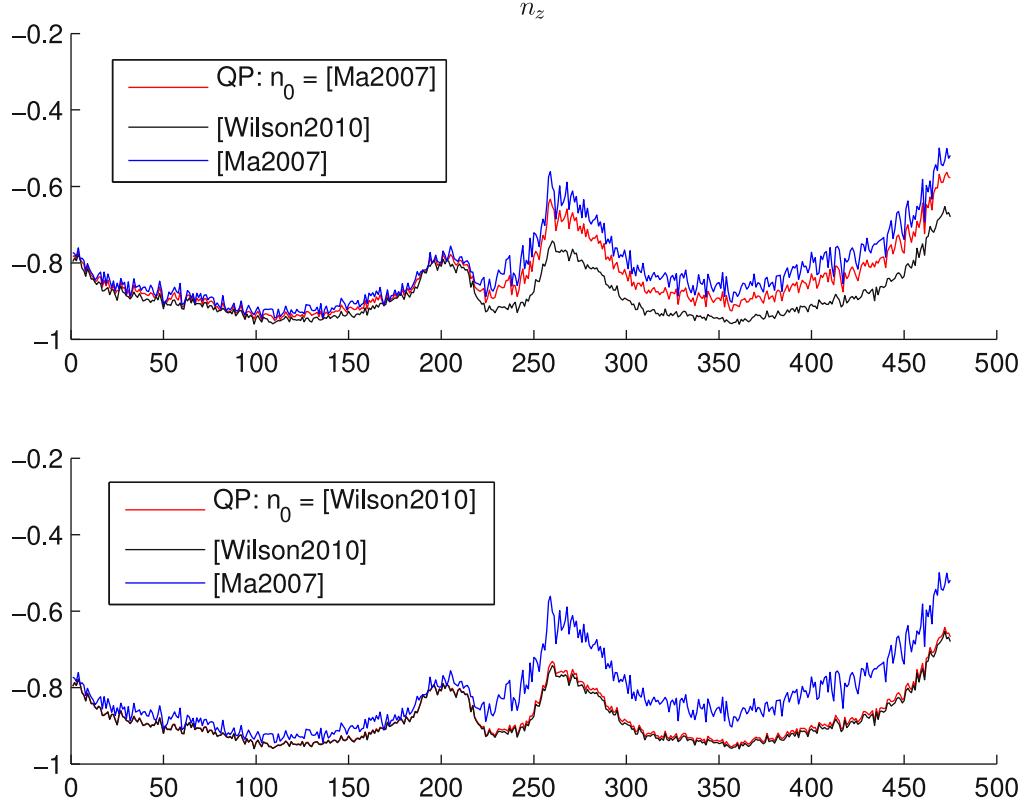


Fig. 4.12:  $n_z$  component of surface normals in 1 pixel wide selected horizontal region (Fig. 4.10 - right) obtained after applying QP based normal correction with initial estimate of surface normals from (top) Ma et al. method i.e  $n_0 = n^{[\text{Ma2007}]}$  and (bottom) Wilson et al. method i.e.  $n_0 = n^{[\text{Wilson}]}$ .

## 4.5 Minimal Image Sets for Robust Spherical Gradient Photometric Stereo

In section 4.4.3, we used our modified radiance equations to show that the quality of surface normals computed using Ma et al. [26] method will degrade with deformation in diffuse lobe. In this section, using the same modified radiance equations, we show how the method of Wilson et al. [37] uses a set of 6 gradient and complement gradient images to cancel out

the effects of deformed diffuse lobe. Finally, based on the analysis of spherical gradient photometric stereo using our modified radiance equations, we propose a minimal 4 image set method that combines the advantage of the original method of Ma et al. (reduced data capture requirement) with that of Wilson et al. (improved robustness).

Recently, Wilson et al. [37] proposed the use of complement gradient images, in addition to gradient images, to improve the quality of recovered surface normals. This method used the difference of gradient and complement gradient images to recover surface normals. Mathematically, the “difference method” of [37] is given by:

$$\vec{n} = \frac{[r_x - r_{\bar{x}}, r_y - r_{\bar{y}}, r_z - r_{\bar{z}}]^T}{\| [r_x - r_{\bar{x}}, r_y - r_{\bar{y}}, r_z - r_{\bar{z}}] \|}. \quad (4.41)$$

They claimed that this method improves the quality of the normal estimates over estimates from Ma et al. [26], since “the pixels that are dark under one gradient illumination condition are most likely well exposed under the complement gradient illumination condition” [37, p17:5]. Indeed, the validity of this claim is easily demonstrated by our modified radiance equations. Once again, we use our modified radiance equations to represent the surface normal computed using Wilson et al. method:

$$\begin{aligned} N_{\{x,y,z\}} &= r_{\{x,y,z\}} - r_{\{\bar{x},\bar{y},\bar{z}\}} \\ &= \frac{\pi\rho_D V_p}{2} \left\{ \delta_{\{\bar{x},\bar{y},\bar{z}\}} + \frac{2}{3} n_{\{x,y,z\}} \right\}. \end{aligned} \quad (4.42)$$

As the reader considers equation (4.42), it is critical to understand that we arrived at this expression using the modified radiance equations that are based on the following simplifying assumption described in section 4.4.1 and 4.4.4: overall deformation in the diffuse reflectance lobe for gradient and complement gradient illumination environment can be quantified using a single scalar parameter  $\delta_{\{x,y,z\}}$  and  $\delta_{\{x,y,z\}} + \delta_{\{\bar{x},\bar{y},\bar{z}\}}$  respectively.

The interesting observation in (4.42) is that the symmetric distortion of diffuse lobe cancel out and the only component contributing to error in the recovery of surface normal is the asymmetric distortion parameter  $\delta_{\{\bar{x},\bar{y},\bar{z}\}}$ . In other words, symmetric deformations in the reflection lobe are averaged out and therefore the surface normals recovered by Wilson et al. [37] method are less affected by deformation to the diffuse reflectance lobe caused by shadowing,

inter-reflection and coarse approximation of spherical illumination due to light discretization. Therefore, we conclude that the method of Wilson et al. [37] recovers surface geometry closer to the true surface geometry because its “difference method” involves cancellation of symmetric deformation in diffuse reflectance lobes. This “symmetric deformation” cancellation property is not present in the “ratio image” method proposed by Ma et al. [26]. Note that the remaining constant cancel out during the vector normalization step.

Using our modified radiance equations and building upon the “difference method” proposed by [37], we derive a minimal four image solution in which symmetric deformations of diffuse lobes still cancel. We exploit the following complement image constraint to arrive at the minimal four image solution:

$$r_x + r_{\bar{x}} = r_y + r_{\bar{y}} = r_z + r_{\bar{z}} = r_c. \quad (4.43)$$

Using this X complement image constraint, we rewrite (4.41) as:

$$\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \bar{\mathbf{x}})} = \frac{[r_x - r_{\bar{x}}, 2r_y - (r_x + r_{\bar{x}}), 2r_z - (r_x + r_{\bar{x}})]^T}{\|[r_x - r_{\bar{x}}, 2r_y - (r_x + r_{\bar{x}}), 2r_z - (r_x + r_{\bar{x}})]^T\|} \quad (4.44)$$

Similarly, Y and Z base complement pairs can also be used to obtain  $\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \bar{\mathbf{y}})}$  and  $\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \bar{\mathbf{z}})}$  as follows:

$$\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \bar{\mathbf{y}})} = \frac{[2r_x - (r_y + r_{\bar{y}}), r_y - r_{\bar{y}}, 2r_z - (r_y + r_{\bar{y}})]^T}{\|[2r_x - (r_y + r_{\bar{y}}), r_y - r_{\bar{y}}, 2r_z - (r_y + r_{\bar{y}})]^T\|} \quad (4.45)$$

$$\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \bar{\mathbf{z}})} = \frac{[2r_x - (r_z + r_{\bar{z}}), 2r_y - (r_z + r_{\bar{z}}), r_z - r_{\bar{z}}]^T}{\|[2r_x - (r_z + r_{\bar{z}}), 2r_y - (r_z + r_{\bar{z}}), r_z - r_{\bar{z}}]^T\|} \quad (4.46)$$

In a similar way, we can also derive expressions for complement minimal image sets:  $\mathbf{n}_{(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \mathbf{x})}$ ,  $\mathbf{n}_{(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \mathbf{y})}$  and  $\mathbf{n}_{(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \mathbf{z})}$ . Therefore, we have total six image sets in our minimal image sets formulation:  $\mathbf{n}_{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \{\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}\})}$  and  $\mathbf{n}_{(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})}$ .

In the next section, we show that there is very small angular deviation ( $\sim 3.9^\circ$ ) between the normals computed using Wilson et al. method and our method. This observation supports our claim that above substitution indeed preserves the “symmetric deformation cancellation property”.

The non-symmetric deformation  $\delta_{\{\bar{x}, \bar{y}, \bar{z}\}}$  do not cancel and still contribute to error in the recovered surface normals. To analyze the influence of non-symmetric deformation, we

computed surface normal of a cylinder as shown in Fig. 4.13 using Wilson et al. [37] and our method. If there were significant contribution of non-symmetric deformation, the normals computed using Wilson et al. method and our method would have deviated strongly from the ground truth (not shown in the plot as it aligns with surface recovered using Wilson et al. method). Therefore, we conclude that in practice the contribution of non-symmetric deformation is very small. Unavailability of ground truth data prevented us from verifying this claim for other more complex surfaces like a human face.

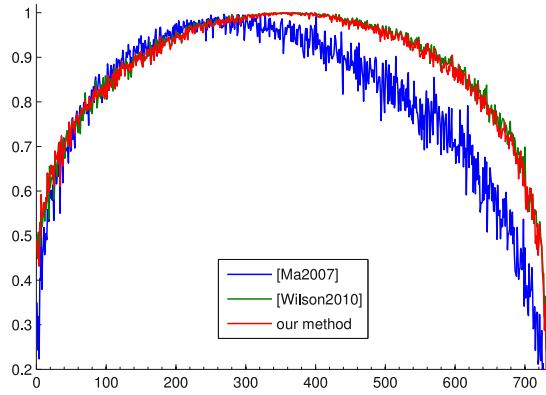


Fig. 4.13: z-component ( $n_z$ ) of estimated surface normals of a cylinder

#### 4.5.1 Results

From our analysis in previous section, we concluded that the method of Wilson et al. [37] recovers optimal surface geometry as it involves cancellation of symmetric deformation in diffuse reflectance lobes. Hence, we use the normal map recovered using [37] to assess the quality of normals recovered using our minimal image sets method and that obtained from Ma et al. [26]. First let us analyse the results for a static object (a statue). This object is made up of white cement plaster and hence its reflectance properties are very close to an ideal diffuse surface. Moreover, the static nature of this object ensures that the captured gradient images are perfectly aligned<sup>2</sup>.

In Fig. 4.14 (top row), we show the normal maps computed using the three possible minimal image sets :  $(X, Y, Z, \bar{X})$ ,  $(X, Y, Z, \bar{Y})$  and  $(X, Y, Z, \bar{Z})$ . The normal maps computed for same gradient images using Wilson et al. and Ma et al. is shown in the middle row of Fig.

---

<sup>2</sup>misalignment can be caused by motion of the subject during the capture process

4.14. The bottom row in this figure shows the distribution of angular error between normal map computed using our minimal image set method and that computed using [37] and [26]. It is evident from these histogram that the normal map estimated using our minimal image sets method (requiring just 4 images) is very close to that estimated by [37] (requiring 6 images). Also, the angular difference is not very large for the normal map estimate by [26]. Hence, analysis of normal map estimates of a static object suggests that there is not much significant difference in the normal maps computed using these three methods.

Now let us perform similar analysis for the normal map of a human face. Small motion between gradient images of non-static objects, like a human face, is unavoidable. This results in misalignment of gradient images and therefore causes surface normal deviations that cannot be modelled using our “diffuse reflectance lobe distortion” framework. Our Light Stage has significant “*light discretization*” as we use only 41 LED (74% less than the Light Stage of [26] and [37]). This contributes significantly to the deformation of diffuse lobes. The normal estimation technique of Ma et al. is unable to cope with distortion in the diffuse lobes. This causes the recovered surface normal to have large deviation from the true surface normal. From our analysis in 4.5, we know that if the deformation in diffuse lobes is symmetric in the complement images, the normal estimation technique of Wilson et al. results in cancellation of these deformations. Our minimal image sets formulation preserves this “deformation cancellation” property and hence there is very small angular difference ( $\sim 7.3^\circ$ ) with the normal map computed using Wilson et al. as shown in Fig. 4.15. The inability of Ma et al. method to cope with deformation in diffuse lobe is also evident from the distribution of angular deviation shown in Fig. 4.15 (bottom). It exhibits large angular deviation ( $> 42^\circ$ ) with our minimal image set normal map and that of Wilson et al. [37].

### 4.5.2 Discussion

The method of Ma et al. also used 4 images,  $r_x, r_y, r_z, r_c$ , but it does not use the additional information about deformation in reflectance lobes obtained from complement gradient images. On the other hand, Wilson et al. method requires 6 images,  $r_x, r_y, r_z, r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}$ , to compensate for deformation of reflectance lobes. Our method requires only 4 images,  $r_x, r_y, r_z, r_{\bar{x}}$ , because it exploits the information obtained from X and complement X gradient condition in the es-

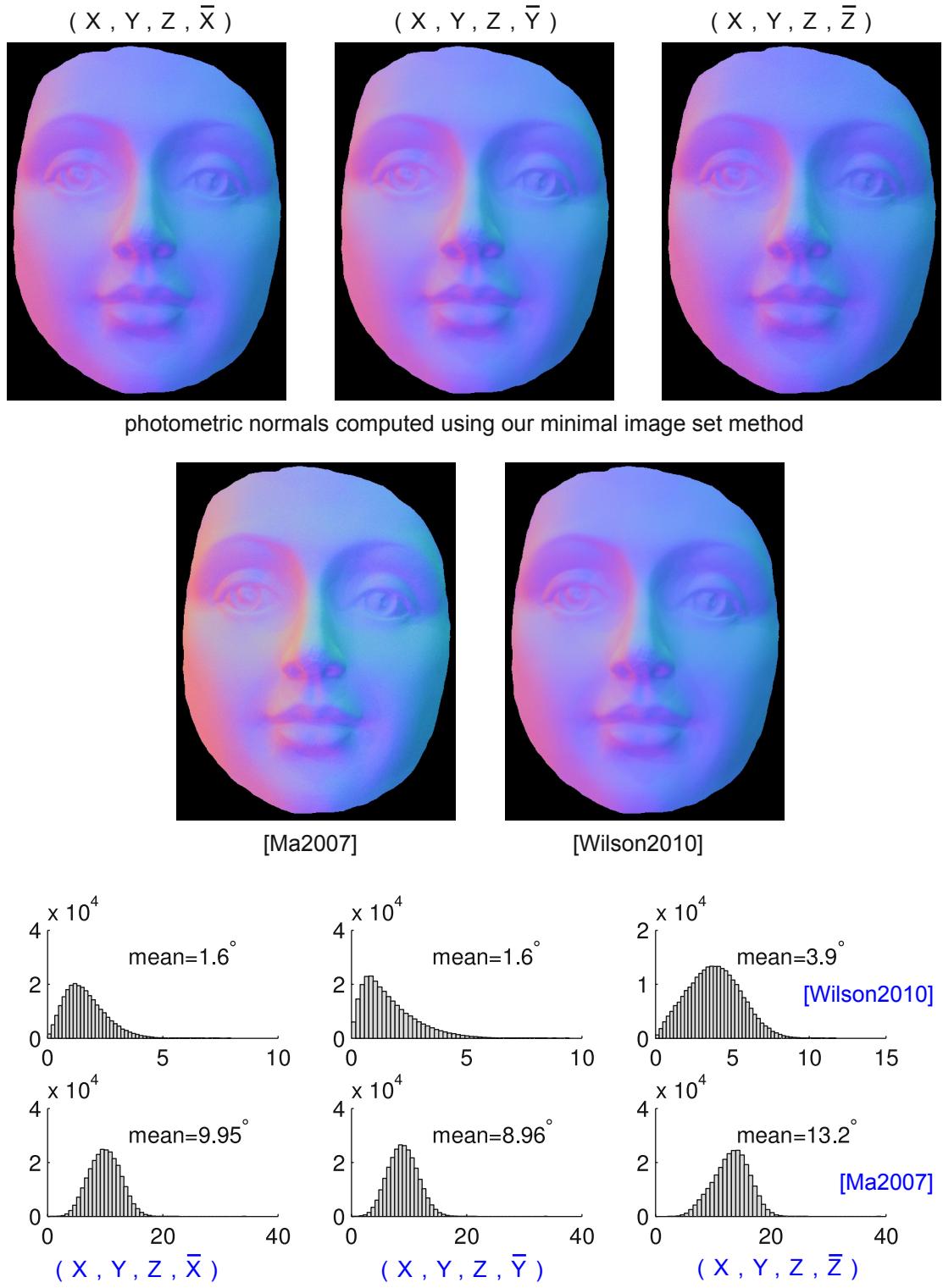
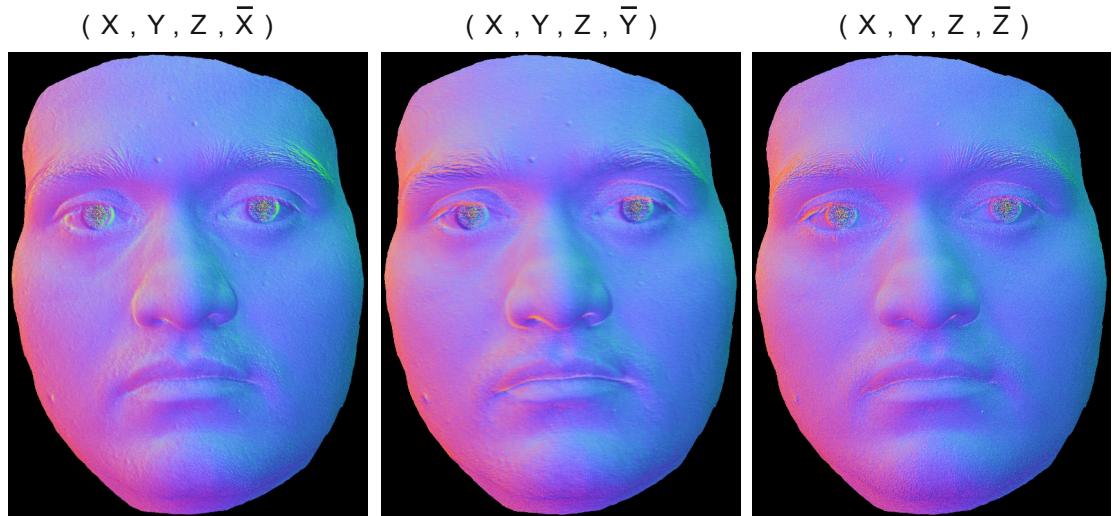


Fig. 4.14: Photometric normals of a statue (static) computed using our minimal image set method (top) and that computed using Ma et al. [26] (middle left) and Wilson et al. [37] (middle right). All the three complement base pairs —  $(X, \bar{X})$ ,  $(Y, \bar{Y})$  and  $(Z, \bar{Z})$  — possible in our minimal image set method was used to generate similar photometric normals. (Bottom) Distribution of angular difference between normal maps computed using our minimal image set method and that computed using [26] and [37].



photometric normals computed using our minimal image set method

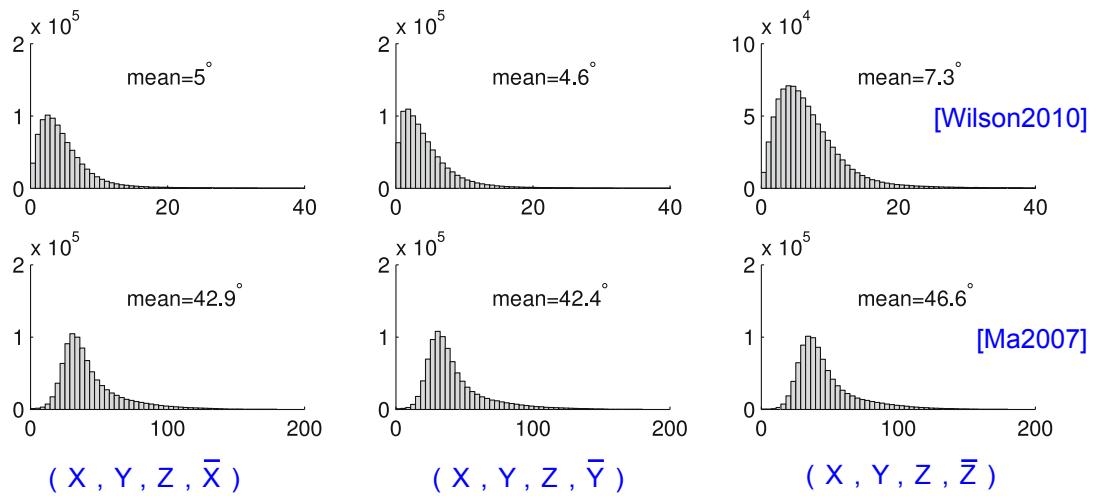
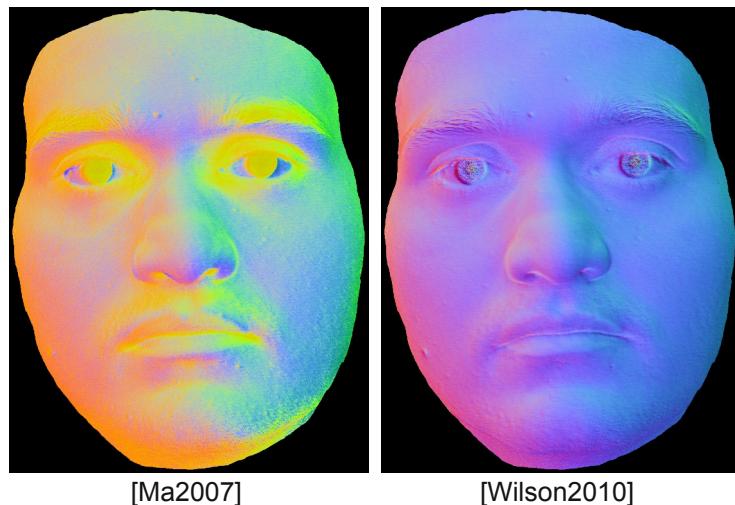


Fig. 4.15: Photometric normals of a face (non-static) computed using our minimal image set method (top) and that computed using Ma et al. [26] (middle left) and Wilson et al. [37] (middle right). All the three complement base pairs —  $(X, \bar{X})$ ,  $(Y, \bar{Y})$  and  $(Z, \bar{Z})$  — possible in our minimal image set method was used to generate similar photometric normals. (Bottom) Distribution of angular difference between normal maps computed using our minimal image set method and that computed using [26] and [37].

timation of Y and Z complements. In other words, this new method combines the advantage of the original method of Ma et al. (reduced data capture requirement) with that of Wilson et al. (improved robustness). This new formulation is able to reduce the data requirements which is extremely important if the spherical gradient photometric stereo is to be used for real time performance capture as discussed in 5.1.

It is important to understand that our analysis is based on the following simplifying assumption described in section 4.4.1 and 4.4.4: overall deformation in the diffuse reflectance lobe for a gradient and complement gradient illumination environment can be quantified using a single scalar parameter  $\delta_{\{x,y,z\}}$  and  $\delta_{\{x,y,z\}} + \delta_{\{\bar{x},\bar{y},\bar{z}\}}$  respectively.

## 4.6 Registration of Spherical Illumination Images

Spherical gradient photometric stereo technique requires capture of 4 spherical illumination images ( $X, Y, Z, C$ ) with the assumption that the imaged object remains at the same position during the capture process. In other words, a pixel position in all the gradient images should correspond to the same surface patch. However, for non-static objects like a human face, it is difficult to remain at same position during the capture of these 4 images. Even at high capture frame rate, apparent motion between 1<sup>st</sup> and 4<sup>th</sup> image is unavoidable which causes some inaccuracy in the photometric normals computed using misaligned gradient images. Hence, in order to recover accurate photometric normals, we must align these gradient images to the constant illumination image. This task is achieved by the Joint Photometric Alignment method proposed by Wilson et al. [37].

Traditional optical flow techniques have been successfully applied for alignment of images consisting of small motion of the imaged object. Such techniques estimate the apparent motion of object in a sequence of images by exploiting the brightness constancy assumption i.e. corresponding image points maintain their brightness level despite apparent motion. Mathematically, this assumption can be expressed as:

$$I(x, t) = I(x + u, t + 1)$$

where,  $I(x, t)$  is the image pixel value at a 2D spatial location  $\vec{x} = [x \ y]^T$  and time  $t$ . Optical

flow based alignment techniques estimate the 2D warp function  $u$  (flow field) to minimise

$$u \leftarrow \operatorname{argmin}_u \varepsilon(I(x + u, t + 1), I(x, t))$$

where,  $\varepsilon(\cdot)$  is the error function which quantifies the extent of misalignment between the source  $I(x, t)$  and target  $I(x + u, t + 1)$  images. The 4 illumination conditions ( $X, Y, Z, C$ ) in spherical gradient photometric stereo are deliberately designed to dramatically change the pixel brightness of each image point in order to reveal the corresponding surface geometry. This causes violation of the brightness constancy assumption in the 4 images and hence traditional optical flow based tecniques cannot be directly applied to align the gradient images.

The Joint Photometric Alignment technique of Wilson et al. [37] can align these gradient images at the expense of capturing additional 3 images called the complement gradient images<sup>3</sup> -  $(\bar{X}, \bar{Y}, \bar{Z})$ . They exploit the complement image constraint to align the gradient  $(X, Y, Z)$  and complement gradient images  $(\bar{X}, \bar{Y}, \bar{Z})$  to the constant illumination image  $C$  (also called tracking frame). Mathematically, the complement image constraint can be expressed as:

$$r_{\{x,y,z\}} + r_{\{\bar{x},\bar{y},\bar{z}\}} = r_c \quad (4.47)$$

where,  $r_{\{x,y,z\}}$ ,  $r_{\{\bar{x},\bar{y},\bar{z}\}}$  and  $r_c$  represent gradient, complement gradient and constant illumination image respectively. The Joint Photometric Alignment method is an iterative algorithm that estimates optimal 2D warp functions  $u$  and  $v$  (flow fields) for the gradient and complement gradient images such that extent of complement constraint violation is minimized. Bootstrapping both flows ( $u$  and  $v$ ) initialized to zero, the iterative algorithm proceeds to minimise the following error in each iteration:

$$\begin{aligned} u^{(i+1)} &\leftarrow \operatorname{argmin}_u \varepsilon \left( r_{\{x,y,z\}}(u), c - r_{\{\bar{x},\bar{y},\bar{z}\}}(v^{(i)}) \right) \\ v^{(i+1)} &\leftarrow \operatorname{argmin}_v \varepsilon \left( r_{\{\bar{x},\bar{y},\bar{z}\}}(v), c - r_{\{x,y,z\}}(u^{(i+1)}) \right) \end{aligned}$$

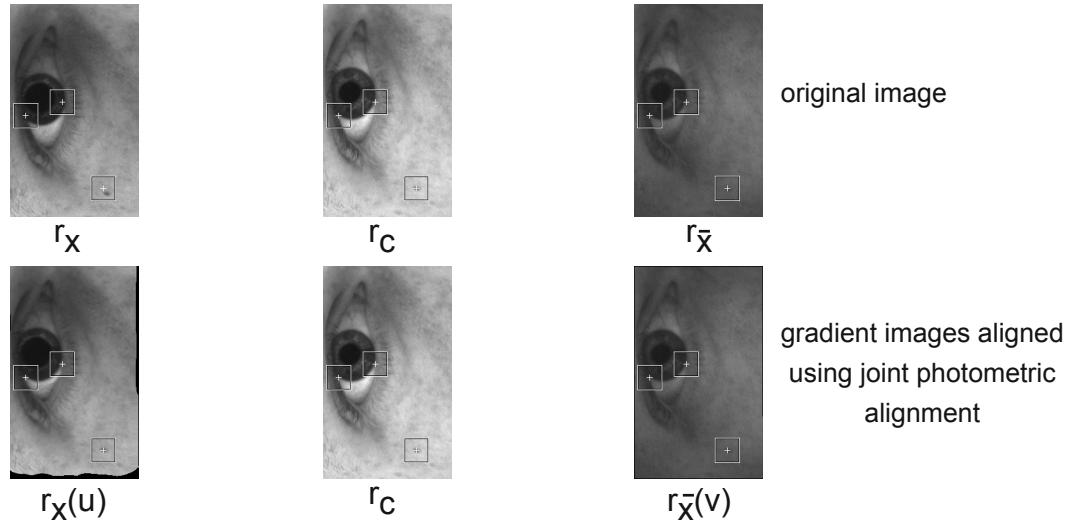


Fig. 4.16: Alignment of spherical gradient images used Joint Photometric Alignment[37]. For illustration purpose, all the intensity values were scaled by 2 except the warped  $r_{\bar{X}}(v)$  which was scaled by 3 because it falls on the dark side of spherical gradient illumination.

#### 4.6.1 Result from alignment of gradient images

To illustrate the result of joint photometric alignment, we marked 3 feature points (cross hair inside a bounding rectangle) in the gradient (top left), constant (top center) and complement gradient (top right) illumination images as shown in Fig. 4.16. We used the following image capture sequence:  $X, Z, Y, C, \bar{X}, \bar{Z}, \bar{Y}$ . As  $C$  and  $\bar{X}$  are consecutive frames in the capture sequence, the apparent motion of feature points (clearly visible due to bounding rectangle) is negligible and hence requires no warping. However,  $X$  and  $C$  are two frames apart in the capture sequence and hence there is significant displacement of the feature points. After the application of joint photometric alignment technique, the marked feature points get aligned in the warped  $X$  gradient image (bottom left) as shown in Fig. 4.16. Large value of flow field  $u$  for  $X$  gradient image is evident from dark regions in the boundary of the corresponding warped image.

As reported in [37], the iterative nature of this alignment technique requires considerable amount of time to arrive at acceptable level of alignment. For a  $298 \times 182$  grayscale image, it took 595.30 sec ( $\sim 10$  min.) to complete 10 iterations<sup>4</sup>. The plot of residual<sup>5</sup> at each iteration

<sup>3</sup>introduced in section 4.4.4

<sup>4</sup>each iteration involves two execution of Brox et al. [5] optical flow technique (C implementation provided by the authors) running in Slackware 13.1-2-12 on 3 GHz Intel®Core2 Duo CPU

<sup>5</sup>this residual,  $\sum |r_c - (r_x + r_{\bar{X}})|$ , quantifies the extent of complement constraint violation

is shown in Fig. 4.17. This plot depicts that at each iteration there is significant reduction in residual and hence the number of iterations should be large for the joint photometric alignment technique to converge at the optimal flow field value.

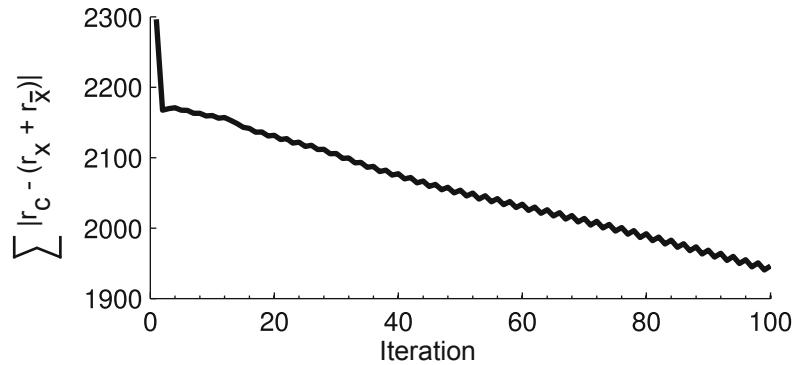


Fig. 4.17: Complement constrain residual for 100 iterations of the joint photometric alignment technique applied to a  $298 \times 182$  spherical X gradient image

# Chapter 5

## Applications of the Light Stage

A Light Stage provides a rich source of geometric and photometric information that is useful for many research avenues in, but not limited to, Computer Vision and Computer Graphics. In this chapter, we will discuss two such applications of the Light Stage which we have explored.

### 5.1 Real Time Performance Capture

The real time facial geometry of a dynamic performance can be captured using the spherical gradient photometric stereo based performance capture and photometric alignment method proposed by Wilson et al. [37]. We modify the capture sequence proposed by Wilson et al. based on our minimal image sets for robust spherical gradient photometric stereo (discussed in section 4.5). This modified performance capture sequence results in:

- Reduced data capture requirement for real time performance capture without compromising the quality of recovered photometric normals. In other words, we show that only 5 spherical illumination images, instead of 7, is sufficient for estimation of tracking frame photometric normal and corresponding warped normals.
- Lower post processing overhead because the modified capture sequence requires joint photometric alignment of only one pair, instead of three, of gradient and complement gradient images. In other words, the post processing time required for alignment of gradient images is significantly reduced because only one pair of gradient images, farthest from the tracking frame, require alignment.

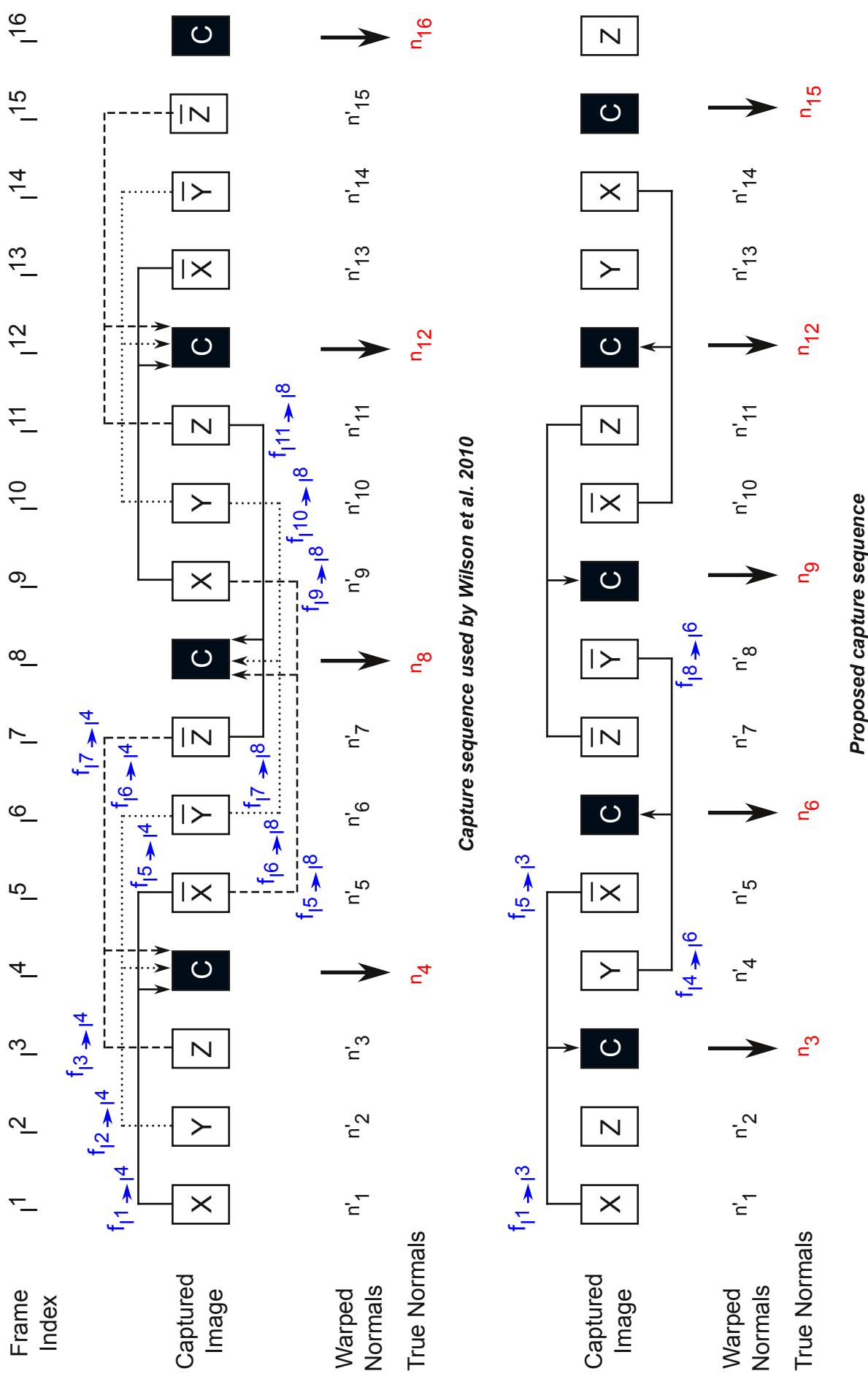


Fig. 5.1: Image capture sequence used by Wilson et al. [37] (top). The proposed image capture sequence (bottom) based on our minimal image sets method.

### 5.1.1 Original Performance Geometry Capture Method

In this section, we briefly describe the real time performance geometry capture method proposed by Wilson et al. [37]. The performance capture sequence developed by Wilson et al. is depicted in Fig. 5.1 (top). The frame index ( $I^1, I^2, I^3, \dots$ ) is used to indicate the temporal sequence of each illumination condition in the capture sequence. Wilson et al. developed the Joint Photometric Alignment method (discussed in section 4.6), for the alignment of gradient  $\{X, Y, Z\}$  and complement gradient  $\{\bar{X}, \bar{Y}, \bar{Z}\}$  images to the flanked constant illumination  $C$  image i.e. the tracking frame. For example: the Joint Photometric Alignment applied to frames ( $I^1, I^4, I^5$ ) results in two optical flow fields,  $f_{I^1 \rightarrow I^4}$  and  $f_{I^5 \rightarrow I^4}$ , that align the gradient  $X$  and complement gradient  $\bar{X}$  images to the tracking frame  $C$  respectively. In other words, warping  $X$  and  $\bar{X}$  according to  $f_{I^1 \rightarrow I^4}$  and  $f_{I^5 \rightarrow I^4}$  respectively, aligns both  $X$  and  $\bar{X}$  to the tracking frame  $C$ . We represent the warping operation of an image  $I$  by the flow field  $f$  as  $\mathbf{W}(I, f)$  where  $I \in \mathbb{R}^{N \times M}$ . Note that warping of a vector field  $N$ , involves warping of each normal vector component followed by renormalisation which can be represented as  $\mathbf{W}(N, f)$  where  $N \in \mathbb{R}^{N \times M \times 3}$ .

#### Tracking Frame Photometric Normals

The photometric normal computed at each tracking frame represents the true<sup>1</sup> normal. At each tracking frame, we have access to 3 pairs of aligned gradient and complement gradient images. These aligned gradient images are used for computation of photometric normal at each tracking frame. For example, in Fig. 5.1, photometric normal at the tracking frame  $I^4$  is given by:

$$n_4 = \frac{[X_w - \bar{X}_w, Y_w - \bar{Y}_w, Z_w - \bar{Z}_w]}{\| [X_w - \bar{X}_w, Y_w - \bar{Y}_w, Z_w - \bar{Z}_w] \|} \quad (5.1)$$

where,

$$\begin{aligned} X_w &= \mathbf{W}(X, f_{I^1 \rightarrow I^4}) \quad , \quad \bar{X}_w = \mathbf{W}(\bar{X}, f_{I^5 \rightarrow I^4}) \\ Y_w &= \mathbf{W}(Y, f_{I^2 \rightarrow I^4}) \quad , \quad \bar{Y}_w = \mathbf{W}(\bar{Y}, f_{I^6 \rightarrow I^4}) \\ Z_w &= \mathbf{W}(Z, f_{I^3 \rightarrow I^4}) \quad , \quad \bar{Z}_w = \mathbf{W}(\bar{Z}, f_{I^7 \rightarrow I^4}) \end{aligned}$$

---

<sup>1</sup>referring to non-warped normal and not the ground truth normal

In a similar way, all the tracking frame photometric normals can also be computed. Note that, we represent tracking frame photometric normals by the symbol  $n_{\{x\}}$  while the warped photometric normals are depicted as  $n'_{\{x\}}$ .

### Warped Normals at Intermediate Gradient Frames

With the flow field from each gradient and complement gradient to a common tracking frame at hand, Wilson et al. warped the tracking frame photometric normals to obtain normals corresponding to the temporal location of gradient and complement gradient images. Wilson et al. used the term “Temporal Up-sampling” to refer to this operation of estimating photometric normal at non-tracking frames.

The warped normal at frames  $I^1, I^2, I^3$  is given by:

$$n'_1 = \mathbf{W}(n_4, -f_{I^1 \rightarrow I^4}) , \quad n'_2 = \mathbf{W}(n_4, -f_{I^2 \rightarrow I^4}) , \quad n'_3 = \mathbf{W}(n_4, -f_{I^3 \rightarrow I^4})$$

Recall that we use the symbol  $n'_{\{x\}}$  to represent the warped photometric normals.

For each subsequent frames, every gradient frame is flanked by two tracking frames. Therefore, two flow fields exist for each gradient frame and hence, there are two versions of warped photometric normal corresponding to each gradient frame. For example, if we consider the gradient image  $\bar{X}$  at frame location  $I^5$ , we have the following two warped normals for this frame location:

$$n''_5 = \mathbf{W}(n_4, -f_{I^5 \rightarrow I^4}) , \quad n'''_5 = \mathbf{W}(n_8, -f_{I^5 \rightarrow I^8})$$

Wilson et al. used the weighted average (weighted according to the temporal distance) of these two warped normals as the photometric normal for intermediate gradient frames.

$$n'_5 = \frac{3n''_5 + n'''_5}{||3n''_5 + n'''_5||}$$

In a similar way, we can compute warped normals at all the remaining intermediate gradient frames.

## 5.2 Performance Capture Sequence based on Minimal Image Sets

We can modify the original performance capture sequence of Wilson et al. in order to exploit the minimal image set method of computing photometric normal. Based on the analysis presented in 4.5, we can say that a set of 3 gradient images ( $X, Y, Z$ ) and any 1 of the three complement gradient images ( $\bar{X}, \bar{Y}, \bar{Z}$ ) is sufficient to compute photometric normals. So we can mathematically represent our minimal 4 image set as  $[r_x, r_y, r_z, \{r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}\}]$ . For example, if our minimal 4 image set is  $[r_x, r_y, r_z, r_{\bar{x}}]$ , then we have the following expression for the corresponding photometric normal:

$$\vec{n} = \frac{[r_x - r_{\bar{x}}, 2r_y - (r_x + r_{\bar{x}}), 2r_z - (r_x + r_{\bar{x}})]^T}{\| [r_x - r_{\bar{x}}, 2r_y - (r_x + r_{\bar{x}}), 2r_z - (r_x + r_{\bar{x}})] \|}. \quad (5.2)$$

It is imperative to recall that minimal 4 image set and the expression for surface normal is valid only when the gradient and complement gradient image satisfy the complement image constraint. Mathematically, if  $[r_x, r_y, r_z, r_{\bar{x}}]$  is the 4 image set, then the following complement image constraint must hold true.

$$r_x + r_{\bar{x}} = r_c$$

where,  $r_c$  is the constant illumination image and  $r_x$  and  $r_{\bar{x}}$  form the base complement pair.

The dual of this proposition also exists. The dual minimal 4 image set can be represented as  $[r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}, \{r_x, r_y, r_z\}]$ . For example, if our minimal 4 image set is  $(r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}, r_x)$ , then the expression for photometric normal is given by:

$$\vec{n} = \frac{[r_x - r_{\bar{x}}, -2r_{\bar{y}} + (r_x + r_{\bar{x}}), -2r_{\bar{z}} + (r_x + r_{\bar{x}})]^T}{\| [r_x - r_{\bar{x}}, -2r_{\bar{y}} + (r_x + r_{\bar{x}}), -2r_{\bar{z}} + (r_x + r_{\bar{x}})]^T \|}. \quad (5.3)$$

We can exploit this flexibility in computation of photometric normal to develop a new image capture sequence compatible with the realtime facial geometry capture framework developed by Wilson et al. .

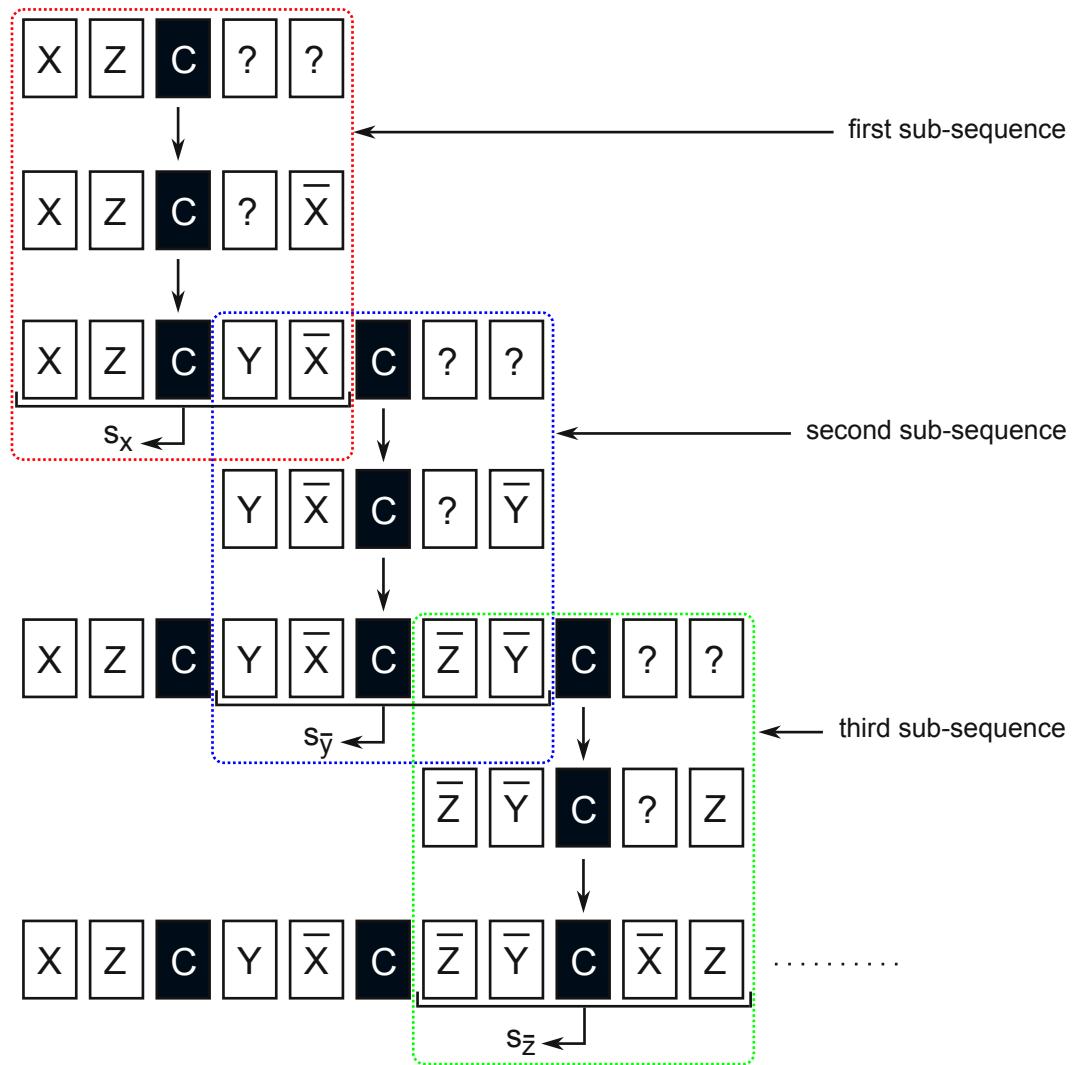


Fig. 5.2: Development of the modified capture sequence

## The New Image Capture Sequence

Using the flexibility provided by minimal image sets in computation of photometric normal, we develop an image capture sequence containing gradient and complement gradient images interleaved in such a way that either a minimal  $[r_x, r_y, r_z, \{r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}\}]$  or dual minimal image set  $[r_{\bar{x}}, r_{\bar{y}}, r_{\bar{z}}, \{r_x, r_y, r_z\}]$  always flanks the tracking frame. Such a sequence can be created with the help of following rules:

1. base complement pair (i.e.  $r_x$  and  $r_{\bar{x}}$  in  $[r_x, r_y, r_z, r_{\bar{x}}]$ ) should be placed farthest from the tracking frame (i.e. the constant illumination image)
2. exactly two gradient images should lie between any two tracking frames in a sequence
3. tracking frame should be flanked by two gradient images and these frames should have linear subject motion with respect to the tracking frame

The third rule is based on the assumption that, at high frame rate of capture, three consecutive frames do not have significant subject motion and hence the subject motion can be assumed to be linear in these frames. Even for exaggerated facial motion, this assumption is reasonable given that the gradient image capture frame rate is large (for example 60 fps). Based on this assumption, we align the gradient images adjacent to a tracking frame by half of the flow field from base gradient image pairs in that subsequence.

To illustrate the development of image capture sequence based on minimal image sets, let us consider an example in which we start the sequence with any two arbitrary gradient images and a constant illumination image(tracking frame)  $[X \rightarrow Z \rightarrow C \dots]$  as shown in Fig. 5.2. The constant illumination image  $C$  is preceded by two gradient images in accordance to Rule 2. According to Rule 1, the base complement pair must be placed farthest i.e. at the two ends of a sub-sequence. As the first position is occupied by  $X$ , the other base complement pair  $\bar{X}$  must appear at the other end as illustrated in the next stage of first sub-sequence shown in Fig. 5.2 (second row from top). We have now partial minimal 4 image set of  $[X, ?, Z, \bar{X}]$ . It is evident that the unknown image in the set can only be filled by  $Y$  gradient. Hence, the final minimal image set corresponding to this sub-sequence is  $[X, Y, Z, \bar{X}]$ . We name this sub-sequence as  $s_x$  because  $X$  is the base complement pair and gradient images (not the complement gradient images) form the first three members of the minimal set. At this

stage, we have the following subsequence:  $[X \rightarrow Z \rightarrow C \rightarrow Y \rightarrow \bar{X} \dots]$ . This subsequence is sufficient to compute photometric normals. Moreover, the warp from  $X$  to  $C$  and  $\bar{X}$  to  $C$  can be computed using the Joint Photometric Alignment.

Using the same three rule, we are now ready to further grow the first sub-sequence to second sub-sequence. According to Rule 2, we first place the tracking frame  $C$ . The other end of this sub-sequence will be filled by  $\bar{Y}$  according to Rule 1. The partial minimal image set is  $[\bar{X}, \bar{Y}, ?, Y]$ . From the dual of our image set, it is evident that the blank space will be occupied by  $\bar{Z}$ . We name this sub-sequence as  $s_{\bar{y}}$  because  $Y$  is the base complement pair and complement gradient images form the first three members of the minimal set. Now the capture sequence becomes:  $[X \rightarrow Z \rightarrow C \rightarrow Y \rightarrow \bar{X} \rightarrow C \rightarrow \bar{Z} \rightarrow \bar{Y} \dots]$ .

In a similar way of growing the sequence, we obtain the minimal image set for third sub-sequence as  $[?, \bar{Y}, \bar{Z}, ?, Z]$  and we name this sub-sequence as  $s_{\bar{z}}$ . At this stage, the combination of three sub-sequences has resulted in the unit sequence  $(s_x, s_{\bar{y}}, s_{\bar{z}})$  whose expanded form is given by:

$$[X \rightarrow Z \rightarrow C \rightarrow Y \rightarrow \bar{X} \rightarrow C \rightarrow \bar{Z} \rightarrow \bar{Y} \rightarrow C \rightarrow X \rightarrow Z \rightarrow C \dots]$$

The end of this capture unit sequence can be combined with the unit sequence generated similarly by  $(s_x, s_y, s_{\bar{z}})$  which in turn can be combined with  $(s_{\bar{x}}, s_y, s_z)$  and so on. There are total 6 possible combinations to form sub-sequences:  $(s_{\{x, \bar{x}\}}, s_{\{y, \bar{y}\}}, s_{\{z, \bar{z}\}})$ . Out of these, two unit sequences are not possible:  $(s_{\bar{x}}, s_{\bar{y}}, s_{\bar{z}})$  and  $(s_x, s_y, s_z)$ . Hence, it is only possible to have the following 4 unique sub-sequences:  $(s_x, s_{\bar{y}}, s_{\bar{z}})$ ,  $(s_x, s_y, s_{\bar{z}})$ ,  $(s_{\bar{x}}, s_y, s_z)$ ,  $(s_{\bar{x}}, s_{\bar{y}}, s_z)$ . Thus, the final capture sequence for real time performance capture is:

$$(s_x, s_{\bar{y}}, s_{\bar{z}}) \rightarrow (s_x, s_y, s_{\bar{z}}) \rightarrow (s_{\bar{x}}, s_y, s_z) \rightarrow (s_{\bar{x}}, s_{\bar{y}}, s_z) \rightarrow (s_x, s_{\bar{y}}, s_{\bar{z}}) \rightarrow \dots$$

as depicted in Fig. 5.1 (bottom).

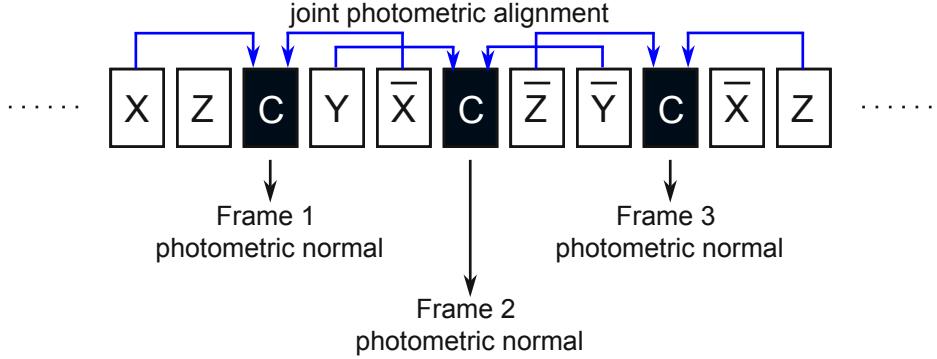


Fig. 5.3: Development of the modified capture sequence

### 5.2.1 Performance Geometry Capture using the New Image Capture Sequence

From the discussion in previous section, we now have a image capture sequence as shown in Fig. 5.3. The images captured in this sequence is sufficient to recover true photometric normals at each tracking frame (frame **C** in Fig. 5.3) using the equations from our minimal image sets analysis discussed in section 4.5.

In Fig. 5.3, the alignment of  $X$  and  $\bar{X}$  to tracking frame  $C$  can be achieved using the Joint Photometric Alignment method of Wilson et al. . However, the photometric normal at tracking frame  $C$  cannot be computed from the images of this capture sequence ( $[X \rightarrow Z \rightarrow C \rightarrow Y \rightarrow \bar{X}]$ ) because the two images ( $Z$  and  $Y$ ) flanking the tracking frame remain misaligned.

At high frame rate of image capture, three consecutive frames do not have significant non-linear subject motion and hence the subject motion can be assumed to be linear in these frames. Even for exaggerated facial motion, this assumption is reasonable given that the gradient image capture frame rate is large (for example 60 fps). Therefore, for 3 images captured in a sequence, the flow between 2<sup>nd</sup> and 3<sup>rd</sup> frames can be approximated by half of the flow between 1<sup>st</sup> and 3<sup>rd</sup> frames. For example: the optical flow field from  $Z$  to  $C$  in Fig. 5.3 can be approximated as half of the flow field from  $X$  to  $C$  i.e.  $f_{Z \rightarrow C} = \frac{f_{X \rightarrow C}}{2}$  and similarly,  $f_{Y \rightarrow C} = \frac{f_{\bar{X} \rightarrow C}}{2}$ . Hence, the optical flow field for base gradient image pair ( $X$  and  $\bar{X}$ ) obtained using Joint Photometric Alignment method of Wilson et al. can be used to approximate the flow field of intermediate gradient frames ( $Z$  and  $Y$ ) that flank the tracking frame ( $C$ ).

We will illustrate the procedure of alignment and computation of tracking frame and gradient frame photometric normals using the capture sequence given in Fig. 5.1 (bottom). The Joint Photometric Alignment applied to  $(I^1, I^3, I^5)$  results in two optical flow fields,  $f_{I^1 \rightarrow I^3}$  and  $f_{I^5 \rightarrow I^3}$ , that align the gradient  $X$  and complement gradient  $\bar{X}$  images to the flanked tracking frame  $C$  respectively. In other words, warping  $X$  and  $\bar{X}$  according to  $f_{I^1 \rightarrow I^4}$  and  $f_{I^5 \rightarrow I^4}$  respectively, aligns both  $X$  and  $\bar{X}$  to the tracking frame  $C$ .

### Tracking Frame Photometric Normals

At each tracking frame, we have access to 1 pair of aligned gradient and complement gradient images. Assuming that the two gradient images flanking the tracking frame ( $Z$  and  $Y$  in this sequence) have linear subject motion, photometric normal at the tracking frame  $I^3$  is given by:

$$n_3 = \frac{[X_w - \bar{X}_w, 2Y_w - (X_w + \bar{X}_w), 2Z_w - (X_w + \bar{X}_w)]}{\| [X_w - \bar{X}_w, 2Y_w - (X_w + \bar{X}_w), 2Z_w - (X_w + \bar{X}_w)] \|}, \quad (5.4)$$

where,

$$X_w = \mathbf{W}(X, f_{I^1 \rightarrow I^3}) \quad , \quad \bar{X}_w = \mathbf{W}(\bar{X}, f_{I^5 \rightarrow I^4}),$$

$$Y_w = \mathbf{W}\left(Y, \frac{f_{I^5 \rightarrow I^4}}{2}\right) \quad , \quad Z_w = \mathbf{W}\left(Z, \frac{f_{I^1 \rightarrow I^3}}{2}\right)$$

Recall that we have used our minimal image set method to compute photometric normal using just 4 spherical illumination images. Also, note that we have warped  $Y$  and  $Z$  according to the average flow of the gradient and complement gradient images with respect to the tracking frame.

Similarly, the normal at frame  $I^6$  is given by:

$$n_6 = \frac{[-2\bar{X}_w + (Y_w + \bar{Y}_w), Y_w - \bar{Y}_w, -2\bar{Z}_w + (Y_w + \bar{Y}_w)]}{\| [-2\bar{X}_w + (Y_w + \bar{Y}_w), Y_w - \bar{Y}_w, -2\bar{Z}_w + (Y_w + \bar{Y}_w)] \|}, \quad (5.5)$$

where,

$$Y_w = \mathbf{W}(Y, f_{I^4 \rightarrow I^6}) \quad , \quad \bar{Y}_w = \mathbf{W}(\bar{Y}, f_{I^8 \rightarrow I^6}).$$

$$\bar{X}_w = \mathbf{W}\left(X, \frac{f_{I^4 \rightarrow I^6}}{2}\right) \quad , \quad \bar{Z}_w = \mathbf{W}\left(Z, \frac{f_{I^8 \rightarrow I^6}}{2}\right).$$

Note the slight change in photometric normal formula caused by dual minimal image set  $(\bar{X}, \bar{Y}, \bar{Z}, Y)$ .

For frame  $I^9$ , the tracking frame photometric normal is given by:

$$n_9 = \frac{[-2\bar{X}_w + (Z_w + \bar{Z}_w), -2\bar{Y}_w + (Z_w + \bar{Z}_w), Z_w - \bar{Z}_w]}{\|[-2\bar{X}_w + (Z_w + \bar{Z}_w), -2\bar{Y}_w + (Z_w + \bar{Z}_w), Z_w - \bar{Z}_w]\|}, \quad (5.6)$$

where,

$$Z_w = \mathbf{W}(Z, f_{I^7 \rightarrow I^9}) \quad , \quad \bar{Z}_w = \mathbf{W}(\bar{Y}, f_{I^{11} \rightarrow I^9}).$$

$$\bar{X}_w = \mathbf{W}\left(X, \frac{f_{I^{11} \rightarrow I^9}}{2}\right) \quad , \quad \bar{Y}_w = \mathbf{W}\left(Y, \frac{f_{I^7 \rightarrow I^9}}{2}\right).$$

In a similar way, all the other tracking frame photometric normals can also be computed.

### Warped Normals at Intermediate Gradient Frames

With the flow field from gradient and complement gradient frames to a common tracking frame at hand, we can compute warped normals corresponding to temporal location of gradient and complement gradient images. The warped normal at frames  $I^1, I^5$  is given by:

$$n'_1 = \mathbf{W}(n_3, -f_{I^1 \rightarrow I^3}) \quad , \quad n'_5 = \mathbf{W}(n_3, -f_{I^5 \rightarrow I^3})$$

Assuming linear subject motion between frame  $I^1$  and  $I^3$ , the warped normal at frame  $I^2$  is given by:

$$n'_2 = \mathbf{W}\left(n_3, -\frac{f_{I^1 \rightarrow I^3}}{2}\right)$$

For each subsequent frames, each gradient frame is flanked by two tracking frame. Therefore, two flow fields exist for each gradient frame and hence, there are two version of warped photometric normal corresponding to each gradient frame. For example, if we consider the gradient image  $\bar{X}$  at frame location  $I^5$ , we have the following two warped normals for this frame location:

$$n_5'' = \mathbf{W}(n_3, -f_{I^5 \rightarrow I^3}) \quad , \quad n_5''' = \mathbf{W}\left(n_6, -\frac{f_{I^4 \rightarrow I^6}}{2}\right)$$

Note that, computation of the warped normal  $n_5'''$  is based on linear subject motion assumption. Based on the temporal distance of  $I^5$  with respect to the two flanking tracking frames  $I^3$  and  $I^6$ , the warped normal at frame  $I^5$  is given by:

$$n_5' = \frac{2n_5'' + n_5'''}{\|2n_5'' + n_5'''\|}$$

In a similar way, we can compute warped normals at all the remaining intermediate gradient frames.

### 5.2.2 Results

Before discussing the results, we describe our capture device setup and its limitations. We used a monochrome JAICM200GE GigE camera. We could not capture gradient images at 60 fps (capture rate of Wilson et al.) for the following two reasons:

- The computer that received the captured image packets via ethernet only supported a maximum ethernet packet size of 1428 bytes. Although, our camera is capable of using jumbo ethernet packet (9000 bytes) for a very high frame rate capture, we could not use this feature due to the limitation of our receiving network node.
- Our Light Stage uses only 41 LED. Hence, based on the sensitivity of our camera, we observed that a minimum exposure time of 50 ms is required to capture well exposed face images.

We observed that when quick exaggerated facial motion is performed, there occurs drastic change in the 1<sup>st</sup> and 5<sup>th</sup> frames of a subsequence block i.e.  $(X, Y, Z, C, \bar{X})$ . Hence, to address the capture rate limitation of our device, we asked our subject to change facial expression

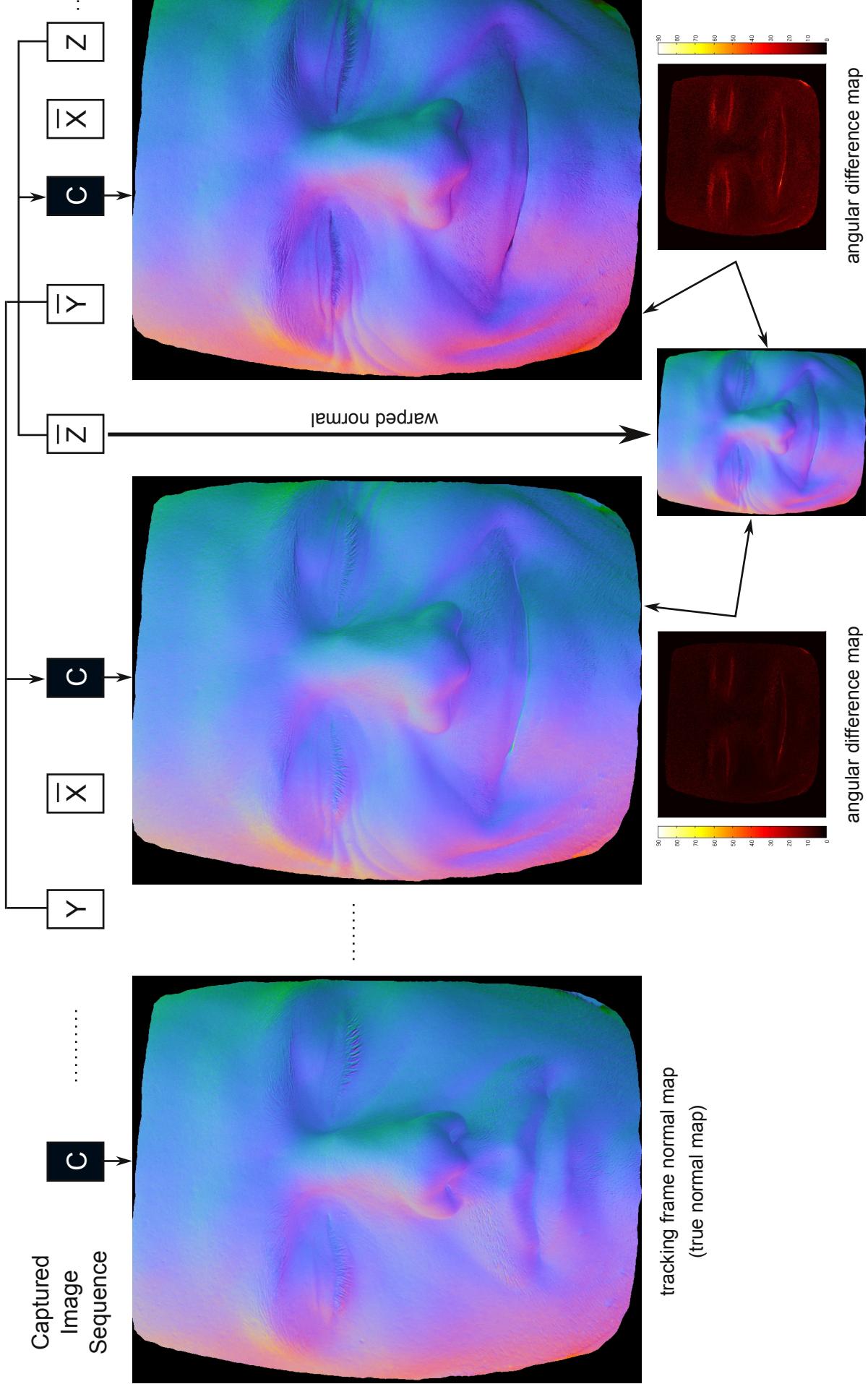


Fig. 5.4: Photometric normals obtained from proposed image capture sequence based on minimal image sets

Table 5.1: Image capture requirements for performance capture using Wilson et al. and our 4 image method

Tracking Frame Count ( $n$ )	Total Number of Images Captured	
	Wilson et al. [37]	our method
1	7	5
2	11	9
3	15	11
4	19	15
5	23	17
6	27	21
...	...	...
$n$	$4n + 3$	$\begin{cases} 6(\lfloor \frac{n}{2} \rfloor + 1) - 1 & \text{odd } n \\ 3n + 3 & \text{even } n \end{cases}$

slowly while we captured gradient frames at the rate of 20 fps. At higher frame rate, we believe that our proposed sequence can resolve the facial performance geometry more finely.

Fig. 5.4 shows the photometric normals computed using images from the modified image capture sequence based on minimal image sets. The tracking frame photometric normals accurately captures the facial geometry during facial motion. Warped photometric normal is computed as weighted average of the tracking frame normals. The weighting of tracking frame normal is performed according to the temporal distance of warped normal from these tracking frames. This weighting strategy is evident from the angular difference map shown in Fig. 5.4 (bottom). These angular difference maps also depict very small motion in the lips and eyes region.

### 5.2.3 Discussion

We have shown that minimal image sets can be exploited to form a capture sequence that can not only reduce the data capture requirement of a realtime performance geometry capture but also reduces the computational cost involved in alignment of the captured images. Table 5.1 shows the relationship between tracking frame capture rate and required number of images to be captured for Wilson et al. and our 4 image method. The impact of reduction in image capture requirement for real time performance capture is pronounced for higher frame rate as shown in Fig. 5.5.

Using our proposed real time performance capture sequence, we can compute true photometric normal map after every two image capture. This allows us to densely sample the

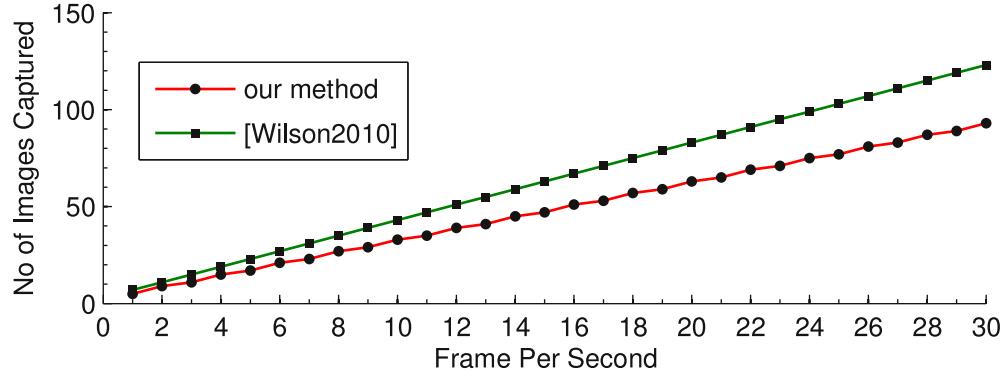


Fig. 5.5: Image capture requirement analysis for performance capture using Wilson et al. [37] and our 4 image method

complete dynamic performance even at lower frame rate.

### Limitations

The proposed capture sequence and method of photometric normal computation assumes high capture frame rate ( $\sim 60\text{fps}$ ) for gradient images. For lower frame rates ( $\sim 20\text{fps}$ ), the geometry of quick exaggerated facial motion cannot be correctly recovered because the alignment algorithm cannot handle non-linear changes in facial geometry. Moreover, the assumption that three consecutive frames do not have significant subject motion and that the subject motion can be assumed to be linear in these frames become invalid for such quick exaggerated facial motion.

### 5.3 Stimuli Image Dataset for Psychology Experiment

The overall appearance of a human face is due to its 3D shape and 2D skin reflectance (skin texture) property. Hence, these two parameters are believed to play a critical role in face processing and recognition carried out by the human brain. Knowledge of how these two sources of information are represented and processed in the neural level is the key to understanding the face recognition mechanism of the human brain.

The face adaptation paradigm is commonly used to study the representation and processing of these two information i.e. 3D shape and 2D skin reflectance information. Face adaptation refers to the decay in neuronal response of face processing regions in human brain

when a human observer is exposed repeatedly to same stimulus (e.g. face image). Original neuronal response can be recovered by altering some properties of the stimulus. Face adaptation paradigm is based on the assumption that the changes in stimuli that causes recovery of the neuronal response relate to the functional properties of cortical neurons [16].

Application of adaptation paradigm requires the ability to control specific properties of the stimuli. Caharel et al. [6] used 3D morphable model to control the 3D shape and 2D reflectance information of stimuli images. They examined the time course (i.e. temporal sequence) for the processing of 3D shape and 2D skin reflectance information using the Event Related Brain Potential - ERP<sup>2</sup> adaptation paradigm. They discovered that 3D shape information caused early sensitivity ( $\sim 160 - 250$ ms) to human faces. Furthermore, they also found that both 3D shape and 2D skin reflectance information (skin texture) contributed equally to ERP on the later time window ( $\sim 250 - 350$ ms).

We collaborated with Jones et al. [22] to study the neural representation of face's 3D shape and 2D skin reflectance information in face selective regions of the human brain. Using fMRI adaptation paradigm [16], Jones et al. analyzed the adaptation of face selective regions in the Fusiform Face Area (FFA), Occipital Face Area (OFA) and Superior Temporal Sulcus (STS). Participants were shown stimuli face images that contained :

1. 3D shape information (shape only)
2. 2D skin reflectance information (texture only)
3. both shape and texture information

### 5.3.1 Stimuli Image Dataset

Using our Light Stage, we created the stimuli images dataset required for this study aiming to investigate the neural representation of 3D shape and 2D skin reflectance in the visual cortex. The experiment was conducted by Jones et al. and a detailed account of the experimental procedure and discussion of the results is available in [22]. Here, we discuss the method that was used to create stimuli image dataset consisting of (a) texture only (b) shape only and (c) texture and shape images of human faces. The frontal view spherical gradient ( $X, Y, Z$ ) and

---

<sup>2</sup>Electroencephalography (EEG) recording during an epoch (time slot in which stimulus is shown) constitute ERP



Fig. 5.6: (left) normal map obtained using spherical gradient photometric stereo, and corresponding (right) *shape only* stimuli image

constant illumination  $C$  images of participants in neutral expression was captured and the corresponding *shape-only* and *texture-only* images were generated as follows:

### Computing the 3D shape only image

We can acquire highly detailed (down to the level of skin pore detail) photometric normal map (left - Fig. 5.6) of a human face using the spherical gradient photometric stereo technique discussed in Chapter 4. This normal map recovers facial geometry in the form of surface normal vector at each surface point covered by individual pixels of an imaging device. Using this normal map, we can generate a front lit Lambertian rendering as follows:

$$I_{\text{shape-only}} = n \cdot l_1 + n \cdot l_2$$

where,  $l_1$  and  $l_2$  are the two front lighting direction vectors (chosen manually to create realistic shape only image) and  $n$  is the facial normal map computed using the spherical gradient photometric stereo technique. The resulting *shape-only* rendered image is shown in Fig. 5.6

(right).

### Computing the texture only image

Skin texture is the result of light reflected after subsurface scattering. In other words, the portion of incident light reflected after entering the skin surface constitutes the characteristic skin colour. Using cross polarization, we separate the facial reflectance into the diffuse and specular components as described in section 3.2. The diffuse only image captured under constant full spherical illumination records the reflectance component responsible for skin texture as shown in Fig. 5.7 (left). We use this image as the *texture-only* stimuli.

### Combined shape and texture image

Combining the *shape-only* image with the *texture-only* image (i.e. diffuse albedo) results in a combined shape and texture image as shown in Fig. 5.7 (right). This stimuli image represents the facial images as captured by a real world camera.

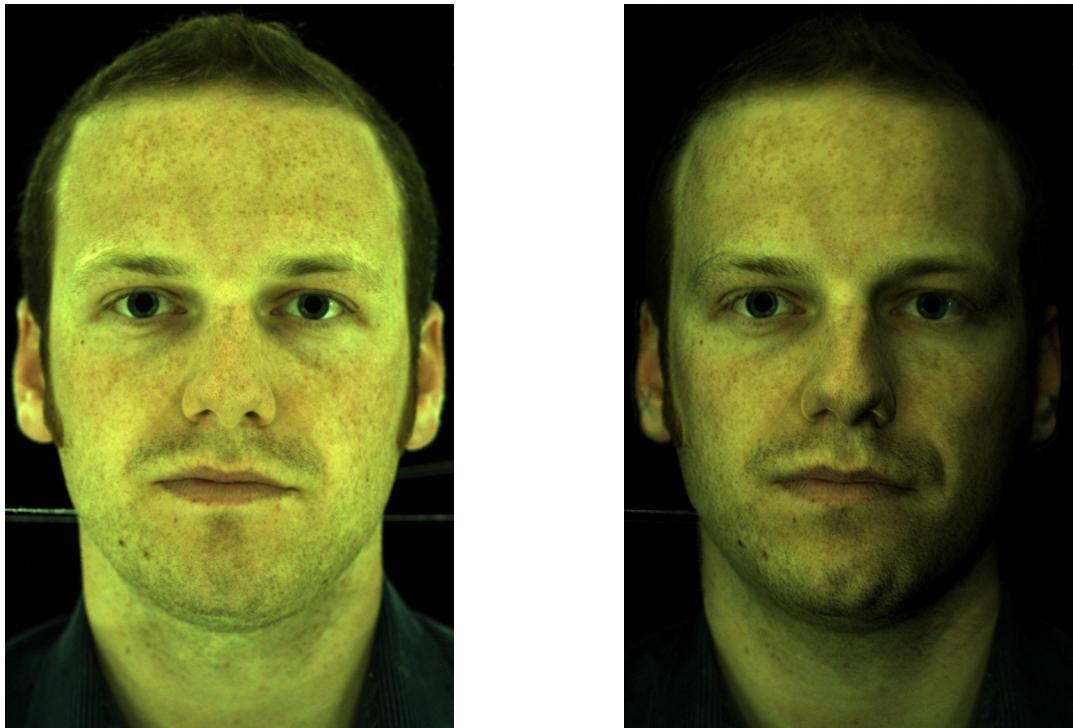


Fig. 5.7: (left) *texture-only* stimuli image, and (right) combined *shape and texture* stimuli image

### 5.3.2 Results and Discussion

A complete explanation of experimental procedure and results are available in Jones et al. [22]. Here, we present a brief summary of results discussed in [22]. Both FFA and OFA regions exhibited significant adaptation to all the image types in the stimuli image dataset. Moreover, there was no significant difference in the activation of the hemispheres. Based on the adaptation paradigm, Jones et al. concluded that the 3D shape and 2D skin reflectance information are represented equally in the face selective regions of the brain. Furthermore, they also found that there was no significant effect of familiarity in the activation of FFA region. This indicated that the FFA region is largely involved in general face processing task rather than in dealing with facial identity.

A Light Stage with 41 LEDs was used to generate the stimuli images for this experiment. It required capture of just 4 images (capture time  $\sim 1$  sec) and almost no post processing to compute the *shape-only* and *texture-only* stimuli images. Electronic control of each LED brightness and data capture in a dark room ensured that the level of illumination remained consistent across different face images.

The spherical gradient photometric stereo technique of Ma et al. [26] was used to compute the facial normal map which in turn allowed rendering of *shape-only* images. The quality of photometric normals computed using [26] is known to degrade with light discretization i.e. coarse approximation of spherical illumination (see 4.4 for details). Our light stage used only 41 LED: 74% less light sources as compared to 151 LED used by [26]. At the time of the stimuli dataset creation, we had not discovered our minimal image sets method discussed in section 4.5. We were also not aware of the normal map computation technique proposed by Wilson et al. [37] which required capture of 6 gradient images ( $X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}$ ). Hence, increasing the number of light source in our Light Stage was the only possible but expensive route to improve the quality of normal map computed using [26]. However, with the minimal image sets method (see section 4.5) in hand, we can now use the same 41 LED light stage to compute very accurate photometric normals without incurring the cost of capturing extra images as required by [37].

Concave regions of a human face (like corner of the eyes) do not receive full hemispherical illumination. In other words, non-convex regions of a face are affected by ambient occlusion.

Although, ambient occlusion helps add realism in 3D Computer Graphics, this effect is not desirable for *texture only* stimulus image because it adds shading information to the non-convex regions. Hence, the *texture only* stimulus images of a human face have some shading effect in the non-convex facial regions.

# Chapter 6

## Conclusion

In this thesis, we have presented a detailed analysis of design and calibration (geometric and radiometric) of a novel shape and reflectance acquisition device called the Multispectral Light Stage. Using the spherical gradient photometric stereo method, we capture highly detailed facial geometry (down to the level of skin pores detail). We used a beam splitter based capture device setup to simultaneously capture both the parallel and cross polarised reflectance components. Therefore, the image alignment procedure is not required to compute specular and diffuse images from the captured parallel and cross polarised images. To record Multispectral skin reflectance map, we added a set of narrow bandpass optical filters to our image capture device. These reflectance maps can be used to estimate biophysical skin parameters such as the distribution of pigmentation and blood beneath the surface of the skin.

We have extended the analysis of original spherical gradient photometric stereo method to consider the effect of diffuse lobes distortion on the quality of recovered surface geometry. Using our modified radiance equations, we show that the symmetric deformation in diffuse reflectance lobe under gradient and complement gradient illumination cancel when computing surface normal using Wilson et al. [37] 6 image method. In addition, we also show that the method of Ma et al. [25], which requires 4 images, is highly affected by deformed diffuse lobes. We propose a minimal image set method, requiring just 4 images, that combines the advantage of the original method of Ma et al. (reduced data capture requirement) with that of Wilson et al. (improved robustness). We show that our method maintains the quality of

Wilson et al. while requiring fewer gradient images. Using our modified radiance equations, we also explore a Quadratic Programming (QP) based normal correction algorithm for surface normals recovered using spherical gradient photometric stereo.

Based on our minimal image sets method, we have proposed a modification to the original performance geometry capture sequence of Wilson et al. [37]. Minimal image sets method provides the flexibility of computing accurate photometric normals from all the possible combinations in minimal image set  $(X, Y, Z, \{\bar{X}, \bar{Y}, \bar{Z}\})$  or the dual minimal image set  $(\bar{X}, \bar{Y}, \bar{Z}, \{X, Y, Z\})$ . We exploit this flexibility to create a performance capture sequence which contain gradient and complement gradient images interleaved in such a way that it always becomes possible to compute aligned photometric normals at the tracking frame (i.e. constant illumination image). This new capture sequence not only reduces the data capture requirement but also reduces the postprocessing computation cost of existing photometric stereo based performance geometry capture methods like [37].

We have also explored the use of Light Stage data for creating stimulus image dataset for a psychology experiment investigating the neural representation of 3D shape and 2D skin reflectance (texture) of a human face. For a given face, we generate three stimulus images: the first contains only the 3D shape information, the second contains only 2D skin reflectance (texture) information and the third contains both shape and texture information. This image dataset has been used by Jones et al. [22] for studying the neural representation of 3D shape and texture of a human face. The high quality photometric normal map obtained from spherical gradient images is used to create a front lit Lambertian rendering of that face. This *shape only* rendered image contains only the 3D shape information. The constant spherical illumination image represents the *texture only* because no shading cues are present due to spherical illumination.

## 6.1 Future Work

The present design of Multispectral Light Stage discussed in Chapter 3 can be improved in many ways. First, finer approximation of spherical gradient illumination can be achieved by increasing the number of light sources to 162. Present version of our Light Stage consists of only 41 LEDs attached to the vertices of a twice subdivided icosahedron. The light reaching

the camera sensor is attenuated by the light source polarizer(< 50% transmission), optical filter(< 90% transmission) and the polarizing beam splitter(< 50% transmission). Hence the camera sensor receives only  $\sim 22\%$  of the total emitted light even if we image a perfect reflector. As a result, capture of multispectral reflectance map requires longer exposure which increases the overall capture time. Adding LEDs to the edges of the twice subdivided icosahedron will not only result in finer approximation of spherical illumination but also ensure that more light is reflected off the object present at the center of the Light Stage.

The second improvement in the design of Multispectral Light Stage can be accomplished by using a stepper motor driven filter wheel. This would help reduce the capture time of Multispectral skin reflectance map. Adding electronic control to the filter wheel would also help automate the whole capture process. For geometric calibration, using a sphere instead of planar checker board would result in more accurate model of image formation.

Chapter 4 describes the minimal image sets method that not only reduces the data capture requirement of spherical gradient photometric stereo but also improves the quality of recovered surface geometry when diffuse lobes are distorted. Future work in this area can explore such correction mechanism for specular reflectance lobes as well. Moreover, we also investigated a Quadratic Programming (QP) based approach for correction of deformed diffuse lobe. However, as the resulting system was severely underconstrained (6 equations and 9 unknowns), our optimization based correction strategy did not result in significant improvement. Future research can also explore improved modeling of specular and diffuse reflectance lobe deformation and search for more constraints.

We used the Joint Photometric Alignment method proposed by Wilson et al. [37] to align gradient and complement gradient images to a common tracking frame. However, this alignment technique is not applicable to multispectral reflectance maps. Hence, future research in this area could investigate into alignment methods for multispectral reflectance maps. One interesting observation regarding this future work is that as the specular reflectance is a surface phenomena, the value of specular radiance should remain constant throughout all the multispectral images. This relationship between multispectral specular reflectance maps can be used to align the specular images and in turn, also the diffuse images.

Chapter 5 also invites future work. The alignment method used in the realtime perfor-

mance geometry capture cannot handle changes in facial geometry. Moreover, it is based on the assumption that the subject motion is linear. Future work in this area can investigate to overcome these limitations of the photometric alignment method. Additionally, capture of performance geometry does not allow tracking of facial feature points. This prevents transfer of facial performance geometry to 3D face models (obtained from range scanner) because no correspondance between image point and its corresponding 3D vertex can be established. Furthermore, the application of Light Stage in creating stimulus image dataset for different types of psychology experiment can also be explored.

Further application of Light Stage data can also explore multi view photometric stereo. This involves capturing facial geometry and reflectance map using two or more camera capturing different view of a face. This would allow reconstruction of high quality 3D geometry based on photometric normals from multiple views.

# Bibliography

- [1] *The Light Stages at UC Berkeley and USC ICT*. USC Institute for Creative Technologies, <http://gl.ict.usc.edu/LightStages/> , 2010. Accessed On : Oct. 22, 2010.
- [2] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1239–1252, 2003.
- [3] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72:239–257, 2007. 10.1007/s11263-006-8815-7.
- [4] J. Y. Bouguet. *Camera calibration toolbox for matlab*. [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc) , 2004. Accessed On : Sep. 12, 2010.
- [5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision (ECCV)*, volume 3024 of *LNCS*, pages 25–36, Prague, Czech Republic, May 2004. Springer.
- [6] Stphanie Caharel, Fang Jiang, Volker Blanz, and Bruno Rossion. Recognizing an individual face: 3d shape contributes earlier than 2d surface reflectance information. *NeuroImage*, 47(4):1809 – 1818, 2009.
- [7] E. North 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.

- [8] Pete Cooper and Marc Thomas. *Geodesic Light Dome*. Department of Computer Science, University of York, UK, <http://www-users.cs.york.ac.uk/~pcc/Circuits/dome> , Mar. 2010. Accessed On : Sep. 10, 2010.
- [9] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [10] Craig Donner and Henrik Wann Jensen. Light diffusion in multi-layered translucent materials. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 1032–1039, New York, NY, USA, 2005. ACM.
- [11] Craig Donner, Tim Weyrich, Eugene d’Eon, Ravi Ramamoorthi, and Szymon Rusinkiewicz. A layered, heterogeneous reflectance model for acquiring and rendering human skin. *ACM Trans. Graph.*, 27(5):1–12, 2008.
- [12] Abhishek Dutta and William A. P. Smith. Minimal image sets for robust spherical gradient photometric stereo. In *ACM SIGGRAPH ASIA 2010 Sketches*, SA ’10, pages 22:1–22:2, New York, NY, USA, 2010. ACM.
- [13] A. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5):476–480, may 1999.
- [14] Y. Furukawa and J. Ponce. Dense 3d motion capture for human faces. pages 1674 –1681, jun. 2009.
- [15] Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. Practical modeling and acquisition of layered facial reflectance. In *SIGGRAPH Asia ’08: ACM SIGGRAPH Asia 2008 papers*, pages 1–10, New York, NY, USA, 2008. ACM.
- [16] Kalanit Grill-Spector and Rafael Malach. fmr-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107(1-3):293 – 321, 2001.

- [17] Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 165–174, New York, NY, USA, 1993. ACM.
- [18] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [19] Takanori Igarashi, Ko Nishino, and Shree K. Nayar. The appearance of human skin: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(1):1–95, 2007.
- [20] Jai. *User Manual for CM0299GE - Digital Monochorome Progressive Scan GigE Vision Camera*. [http://www.jai.com/ProtectedDocuments/JP\\_datasheets/CM\\_CB-200GE\\_200907.pdf](http://www.jai.com/ProtectedDocuments/JP_datasheets/CM_CB-200GE_200907.pdf) , Sep. 2009. Accessed On : Sep. 10, 2010.
- [21] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518, New York, NY, USA, 2001. ACM.
- [22] N. Jones, S. Singh, and T.J. Andrews. How is the 3d shape and texture of familiar and unfamiliar faces represented in the brain?, 2010.
- [23] O. Kleynen, V. Leemans, and M.-F. Destain. Development of a multi-spectral vision system for the detection of defects on apples. *Journal of Food Engineering*, 69(1):41 – 49, 2005.
- [24] General Electric Lighting. *Vio - High Power White LED*. [http://www.gelighting.com/na/business\\_lighting/products/led/downloads/vio\\_data\\_sheet.pdf](http://www.gelighting.com/na/business_lighting/products/led/downloads/vio_data_sheet.pdf) , 2009. Accessed On : Sep. 08, 2010.
- [25] W. C. Ma. *A Framework for Capture and Synthesis of High Resolution Facial Geometry and Performance*. PhD thesis, National Taiwan University, 2008.
- [26] W. C. Ma, T. Hawkins, P. Peers, C. F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Eurographics Symposium on Rendering*, pages 183–194, 2007.

- [27] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-based brdf measurement including human skin. *10th Eurographics Workshop on Rendering*, pages 139–152, June 1999.
- [28] Shree K. Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163, 1997.
- [29] 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. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 935–944, New York, NY, USA, 2006. ACM.
- [30] Diego Nehab. *Advances in three-dimensional shape acquisition*. PhD thesis, Princeton, NJ, USA, 2007.
- [31] Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2010.
- [32] K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. pages 32–41, 1992.
- [33] Norimichi Tsumura, Nobutoshi Ojima, Kayoko Sato, Mitsuhiro Shiraishi, Hideto Shimizu, Hirohide Nabeshima, Syuuichi Akazaki, Kimihiko Hori, and Yoichi Miyake. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph.*, 22:770–779, July 2003.
- [34] Devrim Unay. *Multispectral image processing and pattern recognition techniques for quality inspection of apple fruits*. PhD thesis, Belgium, 2005.
- [35] Tim Weyrich, Jason Lawrence, Hendrik P. A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. Principles of appearance acquisition and representation. *Found. Trends. Comput. Graph. Vis.*, 4(2):75–191, 2009.
- [36] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross.

- Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, 2006.
- [37] Cyrus A. Wilson, Abhijeet Ghosh, Pieter Peers, Jen-Yuan Chiang, Jay Busch, and Paul Debevec. Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graph.*, 29(2):1–11, 2010.
- [38] Lawrence B. Wolff. Polarization vision: a new sensory approach to image understanding. *Image and Vision Computing*, 15(2):81–93, 1997.
- [39] Kwan-Yee K. Wong, Dirk Schnieders, and Shuda Li. Recovering light directions and camera poses from a single sphere. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 631–642, Berlin, Heidelberg, 2008. Springer-Verlag.
- [40] J. Robert Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [41] Zhengyou Zhang. On the epipolar geometry between two images with lens distortion. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 407 –411 vol.1, aug 1996.