PHOTOMETRIC STEREO CONSIDERING HIGHLIGHTS AND SHADOWS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SONER BÜYÜKATALAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENG. DEPT., METU

SEPTEMBER 2011

Approval of the thesis:

PHOTOMETRIC STEREO CONSIDERING HIGHLIGHTS AND SHADOWS

submitted by **SONER BÜYÜKATALAY** in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronics Eng. Dept., METU Department, Middle East Technical University by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Eng. Dept., METU** ________________

Prof. Dr. Uğur Halıcı
Supervisor, **Electrical and Electronics Eng. Dept., METU** ________________

Assist. Prof. Dr. Özlem Birgül
Co-supervisor, **TÜBİTAK UZAY** ________________

Examining Committee Members:

Prof. Dr. Aydın Alatan
Electrical and Electronics Engineering Dept., METU ________________

Prof. Dr. Uğur Halıcı
Electrical and Electronics Engineering Dept., METU ________________

Assist. Prof. Dr. Alptekin Temizel
Informatics Institute Dept., METU ________________

Assist. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Engineering Dept., METU ________________

Assist. Prof. Dr. Hasan Şakir Bilge
Computer Engineering Dept., Gazi University ________________

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: SONER BÜYÜKATALAY

Signature :

ABSTRACT

PHOTOMETRIC STEREO CONSIDERING HIGHLIGHTS AND SHADOWS

Büyükatalay, Soner
Ph.D., Department of Electrical and Electronics Eng. Dept., METU
Supervisor : Prof. Dr. Uğur Halıcı
Co-Supervisor : Assist. Prof. Dr. Özlem Birgül

September 2011, 130 pages

Three dimensional (3D) shape reconstruction that aims to reconstruct 3D surface of objects using acquired images, is one of the main problems in computer vision. There are many applications of 3D shape reconstruction, from satellite imaging to material sciences, considering a continent on earth or microscopic surface properties of a material. One of these applications is the automated firearm identification that is an old, yet an unsolved problem in forensic science. Firearm evidence matching algorithms rely on the fact that a firearm creates characteristic marks on surfaces of the bullets and the cartridge cases. These marks should be digitized unaffected from different surface material properties of evidences. Accuracy of 3D shape is one of the most important parameters affecting the overall identification performance. A very high resolution, accurate 3D data have to be reconstructed in the order of minutes. Photometric stereo (PS) method is capable of reconstructing high resolution surfaces in a fast manner. But, the metallic material and the surface topology of the firearm evidences generate highlights and shadows on their images that does not comply with the assumptions of conventional PS. In the scope of this work, it is intended to design an accurate, fast and robust 3D shape reconstruction scheme using PS considering highlights and shadows. These new PS procedures to be developed here should not be limited only to the ballistic evidences but they also could be used for a wider range of objects reflection properties and texture. For this purpose, masked PS methods which are quite fast when compared to other approaches, were classified and implemented. Simple additional masking methods are also proposed. A novel weighted PS method, using weighted least square estimation, is presented to eliminate false edges created by the masks. Concurrently, the calibration processes and the illumination configuration were improved. The disturbances due to close light sources were removed by image calibrations. From experimental tests to simulate the light positioning problem, it is concluded that the double zenith illumination configuration have better performance than the optimal single zenith illumination configuration, when the highlights and the shadows are considered. Double zenith illumination configuration results were further improved by the weighted normal PS with a small additional computation cost. All the implemented methods were tested firstly on the controlled environment using synthetic images. Later the same tests were conducted on real objects with varying characteristics as well as the firearm evidences.

Keywords: photometric stereo, optimal illumination, highlights, shadows

ÖZ

PARLAMA VE GÖLGELERİ GÖZÖNÜNE ALAN FOTOMETRİK STEREO

Büyükatalay, Soner
Doktora, Elektrik-Elektronik Müh. Bölümü
Tez Yöneticisi : Prof. Dr. Uğur Halıcı
Ortak Tez Yöneticisi : Yard. Doç. Dr. Özlem Birgül

Eylül 2011, 130 sayfa

Objelerin görüntülerinden üç boyutlu (3B) şeklini geri çatma, bilgisayarlı görünün temel problemlerinden biridir. 3B geri çatmanın, uydu görüntülemesinden malzeme bilimine, dünya üzerindeki bir kıtadan, mikroskobik malzeme özelliklerine kadar pek çok alanda değişik uygulamaları bulunmaktadır. Bu uygulamalardan bir tanesi de, eski ama daha çözülmemiş bir adli bilim problemi olan otomatik silah tanıma sistemidir. Silah delili eşleştirme algoritmaları, silahların kovan ve mermi çekirdekleri üzerinde karakteristik izler bıraktığı gerçeğine dayanmaktadır. Bu izler yüzey malzeme özelliklerinden etkilenmeden sayısallaştırılmalıdır. 3B şeklin doğruluğu, bütün tanıma başarımını en çok etkileyen değişkenlerden biridir. Çok yüksek çözünürlüklü, doğru 3B veri dakikalar içinde oluşturulmalıdır. Fotometrik stereo (FS) yöntemi hızlı bir şekilde yüksek çözünürlüklü geri çatım yapabilmektedir. Fakat ateşli silah delillerinin metalik malzemelerinin ve yüzey topolojilerinin ürettiği gölge ve parlamalar, genel FS varsayımları ile uyuşmamaktadır. Bu doktora tezi kapsamında amacım, gölge ve parlamaları göz önüne alan FS kullanarak doğru, hızlı ve güvenilir bir 3B geri çatma akışı tasarlanması amaçlanmıştır. Bu yeni FS yönteminin, sadece balistik deliller ile sınırlandırılmayıp, değişik yansıma ve doku özelliklerine sahip geniş bir objeler sınıfı içinde kullanılabilecek bir geri çatım yöntemi olması amaçlanmaktadır. Bu amaçla, literatürde

bulunan, hızlı çalışan maskeli FS yöntemleri sınıflandırılmış ve gerçekleştirilmiştir. Basit ek maskeleme yöntemleri de önerilmiştir. Maskelerin oluşturduğu hatalı kenarları elemek için, ağrılıklı en küçük kareler kestirimi kullanan, yeni bir ağırlıklı FS yöntemi sunulmuştur. Aynı zamanda, kalibrasyon süreçleri ve aydınlatma konfigürasyonları geliştirilmiştir. Yakın ışıklardan kaynaklanan bozulmalar görüntü kalibrasyonları ile temizlenmiştir. Işık kaynağı konumlandırma problemi için yapılan testlerden, çift zenit açılı aydınlatma seçiminin gölge ve parlamaların olduğu durumlarda, en iyi tek zenit açılı seçimden daha iyi bir başarımı olduğu sonucu elde edilmiştir. Çift zenit aydınlatma seçiminin sonucu, az bir ek hesaplama maliyeti ile daha da iyileştirilmiştir. Bütün yöntemler önce, sentetik görüntüler kullanan kontrollü koşullarda test edilmiştir. Sonra aynı testler, değişen karakteristikteki gerçek objeler ve ateşli silah delilleri üzerinde uygulanmıştır.

Anahtar Kelimeler: fotometrik stereo, en iyi aydınlatma, parlamalar, gölgeler

To Elçin and Zeynep

ACKNOWLEDGMENTS

I am greatly indebted to my supervisor Prof. Dr. Uğur Halıcı and to my co-supervisor Assist. Prof. Dr. Özlem Birgül for their invaluable guidance and encouragement throughout my graduate studies. Their contributions in every stage of this research are gratefully acknowledged.

I am grateful to my thesis committee members Prof. Dr. Aydın Alatan and Assist. Prof. Dr. Alptekin Temizel for their valuable criticism and suggestions.

I would like to express my heartfelt gratitude to Erman Acar and Ahmet Oğuz Öztürk for their contributions to this work during the endless days and nights on calibrating and testing the hardware.

Special thanks to BALİSTİKA2010 hardware team, especially Mustafa Sancay Kırık, Erol Güven, Emre Ünses and Himmet Ateş for their efforts on the image acquisition hardware used in this work.

I would like to thank to all the people working on BALİSTİKA2010 project, especially to Assist. Prof. Dr. Özlem Birgül, Dr. Şükrü Alphan Es, Dr. Uğur Murat Leloğlu and Dr. Ufuk Sakarya for their invaluable support and trust.

I wish to thank to my dear friends Hakan Kocakulak, Kerem Önal, Çağlar Bilir, Yeşim Tantekin, Murat Hüdaverdi, Serkan Girgin, Argün Katırcı, Serhan Poçan, Burçak Poçan, Ozan Keysan, Asuman Keysan, Varlık Kılıç and Utku Kütük for their morale support at all stages of the thesis work.

Finally, I wish to express my deep gratitude to my family for their support and patience.

This work is partially supported by TÜBİTAK 107G194 Research Grant.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

Source Zenith (Z_{source}) Light source zenith angle

Surface Zenith ($Z_{surface}$) Surface normal zenith angle

Image Count (n_i) Number of images used in PS

Diffuse (Di) Images rendered with Lambert reflection model

Highlight (Hi) Images rendered with TS reflection model

Shadow (Sh) Cast shadows created with shadow mapping

Single Zenith ($1Z$) A configuration of light sources, sources are placed on a single circle around camera with equal zenith angles

Double Zenith ($2Z$) A configuration of light sources, sources are placed on two circles around camera with equal zenith angles

Photometric Stereo (PS) Ordinary linear PS using all images

Weighted Normal PS (WPS) Uses Double Zenith. PS is executed on each subset of lights sources, two resultant normals are weighted to calculate final normals

CHAPTER 1

INTRODUCTION

1.1 3D Shape Reconstruction

3D shape reconstruction is one of the main problems in computer vision. These methods aim to reconstruct 3D surface of the object using acquired images. Since methods focused on this problem have similar names like “shape from shading ”or “shape from stereo ”, general name “shape from X ”is used to define all of them.

The one of the early shape from X method was shape from shading (SFS), which was formulated firstly by Horn in 1970 [5]. This method reconstructs the surface from only a single image [6], which is ill conditioned without restricting with some constraints [7]. Especially, “The depth values of some singular points must be known”constraint limits the feasibility of the method.

Binocular stereo (BS) is an other shape from X method that is also used by humans. In BS, two or more images viewing the same scene are used to calculate the depth of a surface point. The main problem in BS is finding conjugate pairs in two images, which is called matching. In case of a calibrated BS system, epipolar geometry can be utilized to reduce 2 dimensional matching problem to 1 dimensional line search problem. However matching problem still has to be solved.

There are two main approaches to matching problem, area-based (intensity matching) approach and feature-based approach. Area-based approach tries to match pixel gray levels of two images. It results dense depth map since matching is executed for each pixel of the images. But this approach is very sensitive to noise. Feature-based approach searches for features (usually edges or corners) to match in two images. This approach is robust if features are visible in both images. But, only depth values of features can be calculated directly, depth values of other points must be interpolated to reconstruct the whole surface. Resolution of depth values in the feature-based approach depends on the surface texture quality, which is a problem for smooth textured surfaces. This resolution problem can be eliminated by using a structured light source [8] [9].

Depth from defocus (DFD) is a fairly new shape from X method. It has similar properties with triangulation methods like BS and motion [10]. Basic idea is the amount of blur in the image is directly related to the camera settings and the depth values of surface [11]. The sensitivity of the DFD depends on the camera's aperture diameter and practically it is not easy to select desired aperture diameter of the camera [10].

One of the well known shape from X methods is photometric stereo that was firstly proposed by Woodham in 1978 [12]. In photometric stereo (PS) method, still images of a surface are captured under variable, known light sources. Intensity values of image pixels can be related to the surface normals and the reflection properties of the surface. This relation can be formulated by reflection models and different models can be selected depending on surface reflection properties. Surface normal and albedo values are calculated using the selected reflection model and images under different lighting conditions.

The main advantages of PS are very high resolution output and faster computation compared to other 3D reconstruction methods. On the other hand, PS has some shortcomings when shadows and specular reflections occur. In the literature, there are two main approaches to solve these problems. The first approach utilizes non-linear reflection equations and resulting constraint optimization problem is to be solved. These non-linear PS solutions have good results but work much slower than linear PS. In the second approach, pixels with shadows or highlights (i.e. specular reflection regions) are masked out, and linear PS is applied to the rest of pixels. This approach increases the accuracy of the results and still performs fast.

1.2 Automated Firearm Identification

Automated firearm identification is an unsolved important problem in forensic science. Ballistic experts use stereo microscopes capable of viewing two firearm evidences side by side to compare bullets or cartridge cases. This setup can compare only two bullets or cartridge cases at a time. With the increasing number of evidences, identification problem grows geometrically making it nearly impossible to solve without an automated identification system.

Identification of firearms depends on striated and impressed marks on the metallic surfaces of bullets and cartridge cases [13] [14]. In most cases, the useful information for evidence comparison is the geometry information, independent of texture and reflection [13] [14]. Early versions of automated identification systems use 2D images of cartridge cases [15]. These systems have poor performance since the marks on the metallic surface of evidences are very sensitive to the type and direction of light sources [13]. 3D shape is a better choice to identify the geometrical marks independently from illumination.

1.3 Motivation and Objective of The Thesis

Beside the automated firearm identification, the 3D shape reconstruction has many applications in industry, from satellite imaging to material sciences that may consider a continent on earth to microscopic surface properties of a material. A very hot topic in computer vision and pattern recognitions is three dimensional face recognition, that works on the reconstructed 3D facial data. On the other hand, 3D media consumer products that developed recently, increases demand on 3D data.

This thesis work is about the 3D reconstruction of any surface topology from the images. However the problems occurring on firearm evidences were concentrated most. The images of the evidence metallic surfaces mainly suffer from highlights and shadows. The 3D reconstruction problem has to be solved considering these highlights and shadows.

The two main requirements of 3D reconstruction for automated fire identification are;

- The most discriminative 3D shape should be generated.
- The method should generate 3D shape in the order of minutes.

The discriminative 3D shape is very important, since thousands of similar evidences will be identified. The 3D generated shape should have high pixel resolution to be discriminative. Previous systems were using 1024x1024 pixel images for 2D investigation. Higher resolution is needed with the growing database of evidences. This requirement limits the 3D reconstruction methods to dense 3D shape generation methods.

The feasibility of the identification system depends on the speed performance of acquisition and comparison system. The acquisition hardware and software should reconstruct the surface of the evidence in the order of minutes. This performance is required to handle daily work load of the evidence acquisition system.

1.4 Contributions of The Thesis

Firstly, masked PS methods that removes highlights and shadows were implemented. Masked PS methods depends on the fact that objects with highlights and shadows in some regions, still have diffuse (Lambert) reflection on other parts of the surface [16]. The locations of highlights and shadows change with lighting conditions. With proper selection of light sources, one may acquire a set of images so that for each pixel, a subset with at least three images satisfying linear reflection model (Lambert) can be chosen. Pixel values at three images is necessary in order to find out three unknowns, one is related to albedo and other two are related to surface normal at that point.

In this study, the implemented masks are listed as follows;

- **Image Masks**
  - Threshold Mask (Th)
  - Non-Lambert Quadruple Mask (NL) [2]
- **Normal Masks**
  - Self Shadow Mask (SS)
  - Cast Shadow Mask (CS)
  - Highlight Mask (Hi)
  - Reflection Mask (Re) [2]
  - Shadow Mask (Sh) [4]
- **Subset Masks**

- Coleman and Jain Mask (CJ) [17]
- Extended Non-Lambert Quadruple Mask (xNL) [3] [18]
- RANSAC (RA) [19]

The masked methods listed above are combined within a unified framework that we name as unified PS. This framework provides a platform for fair comparison among various methods. This is achieved by using the same codes for PS calculation except the masking procedures. Secondly, using unified PS framework, masks can be fused in different ways using logical operations easily. Hence, different combinations can be selected and evaluated based on the characteristics of each mask and requirements of a specific application to find the most robust PS method.

Threshold mask, self shadow mask, cast shadow mask and highlight mask are fast, simple and easy to implement masking methods that are not found in the literature. Also extended Non-Lambert Quadruple Mask method for six images is generalized to any number of images in this study. Other masking methods are taken from previous works and implemented as they were defined.

Secondly, weighted masked PS was introduced. The previously mentioned masks were used as true/false flags that presents an intensity value will be used or not used in PS. This hard decision created false edges at the boundaries of the mask, where the used intensity sets are changed. This new weighted mask method utilizes residual errors of PS to create weights to masks. Also smoothing masks further reduced the false edges.

Thirdly, optimal illumination configuration for PS is investigated. Most of the previous works on optimal illumination configuration did not include any highlights or shadows while calculating optimal lighting configuration. They commonly proposed that the optimal illumination configuration was placing lights on a single circle around the camera. This study investigated this configuration with realistic highlights and shadows in the synthetic images. The simulations with highlights and shadows showed that it is not possible to find a single optimal configuration for all smooth or rough surfaces. Hence, we propose to place light sources on two circles around the camera with two zenith angles.

Lastly, usage of the double zenith light sources also produced new opportunities to reduce highlight and shadow errors. Instead of using all light sources at once, each light sources set with equal zenith angles is used exclusively in PS. The resultant normal vectors are combined with weights that were calculated based on the estimated errors. This novel weighted normal PS approach and general weight functions are presented in this work.

The methods explained above are also tested on the real images. The calibration of the hardware highly affects the performance of the reconstructed 3D shapes, thus, calibration procedures were also studied in the scope of this thesis.

Since the relative positioning of the sample with respect to the camera, determines the amount of highlights and shadows on the images, a priori adjustment of orientation has a potential to improve 3D generation performance. To explore this potential a plane correction algorithm was implemented and integrated.

As a summary, the contributions of this thesis can be grouped into three categories;

The first group includes the PS configuration improvements on the illumination configuration and calibration processes. The optimal illumination configuration without highlights and shadows were replaced with the double zenith configuration. Double zenith configuration reduced the errors due to highlights and shadows. Also surface illumination due to near light sources was modeled and this effect is normalized with an additional calibration procedure.

Secondly, mask PS methods in the literature are classified and implemented in the unified framework. This framework is capable of applying any combination of the previously implemented masks. Some fast working, simple and easy to implement masking methods such as threshold mask, self shadow mask, cast shadow mask and highlight mask were added to the unified framework. These masks are novel methods to the best of our knowledge.

Thirdly, new weighted approaches to PS were implemented. Firstly weighted mask PS that employs weighted least square estimation was proposed. The previously implemented masks can be weighted in this implementation. Also, weighted normal PS is proposed on double zenith illumination configuration.

1.5 Outline of Thesis

This thesis has been divided into eight chapters including this introduction chapter. Second chapter describes the previous works related to this thesis. The third chapter defines the unified PS framework, masks and weighted mask PS. Chapter four proposes the novel illumination configuration and weighted normal PS method theoretically. Next three chapters explain the tests done with the implemented methods on synthetic images, real objects and cartridge cases, respectively. Finally, in the last chapter results are summarized and discussed.

CHAPTER 2

BACKGROUND

2.1 Introduction

In this chapter, firstly, general definitions used throughout the thesis are explained. The image formation theory with the reflection models and shadows is explained. PS and masked PS methods in the literature are placed in the next section. Previous works on the optimal illumination configuration for PS and resultant normal evaluation schemes like uncertainty are defined. Lastly the error definitions used throughout the thesis are presented.

2.2 Imaging Geometry and General Definitions

The general geometry of the imaging system used in definitions through out this thesis is as follows; the world coordinate system is defined in as camera-centered coordinate system; that is the origin is the center of the field of view, the camera axis is placed on the $+Z$ axis, and the up direction of the image is parallel with $+Y$ axis. In this definition, the X and Y axis of the imaging plane and field of view are parallel. It is generally assumed that the object is placed faraway from the camera's imaging plane so that the orthogonal projection assumption can be safely used.

The general notation vectors used in this work are presented in Fig.2.1. The camera direction vector $\mathbf{v}$, is the optical axis of the camera and with orthogonal projection assumption it is given as $[x\ y\ z]^T = [0\ 0\ 1]^T$ all over the field of view. The unit surface normal vector is represented with $\mathbf{n}$. The light source direction is an unit vector defined from surface point to center of the point light source and is shown with $\mathbf{s}_i$. The subscript i indicates the index of the light source, the photometric stereo (PS) demands multiple light source. In the PS calculations, the light direction $\mathbf{s}_i$ is assumed to be constant, but this assumption is corrected on the real images with the image calibration processes which will be explained in Chapter 6.

The reflection direction vector $\mathbf{r}_i$ is the direction where the specular reflection occurs [20]. This is defined as the symmetric vector of light source direction with respect to surface normal and is formulated as Eq. 2.1. Since, it is a function of surface normal $\mathbf{n}$ and the light source direction $\mathbf{s}_i$, it lies in the plane defined by them.

$$\mathbf{r}_i = 2(\mathbf{s}_i^T\mathbf{n})\mathbf{n} - \mathbf{s}_i \tag{2.1}$$

The surface normals resulting a specular reflection from the light source to the camera direction can be checked directly with specular normal direction defined in Eq. 2.2.

$$\mathbf{n}_S = \frac{\mathbf{s} + \mathbf{v}}{|\mathbf{s} + \mathbf{v}|} \tag{2.2}$$

where |.| denotes the norm of the vector.

All vectors are defined with their zenith and polar angles in this work. The zenith angle of a vector is the positive angle between itself and $+Z$ axis. The polar angle is measured positively counterclockwise from $+Z$ axis.

2.3 Reflectance Map

Reflectance map is the relation of the image intensity with the surface normal [16]. It is a function of the light source radiance and the surface bidirectional reflectance distribution function (BRDF). Source radiance is simply the light power emitted to the surface of interest.

Source radiance function ($L_S(\mathbf{x})$) of a single distant point light source is given in Eq. 2.3. The solid angle delta function δ_ω, is defined as in Eq. 2.4. Here E_o is the irradiance of the light source and $\mathbf{s}$ is the light source direction.

$$L_S(\mathbf{x}) = E_o\delta_\omega(\mathbf{x} - \mathbf{s}) \tag{2.3}$$

$$\int_\omega h(\mathbf{x})\delta_\omega(\mathbf{x} - \mathbf{x}_o)d\omega = h(\mathbf{x}_o) \tag{2.4}$$

BRDF can be defined shortly as the ratio of the reflected light power from the surface to the incident light power with respect to each income and outgoing light directions. BRDF

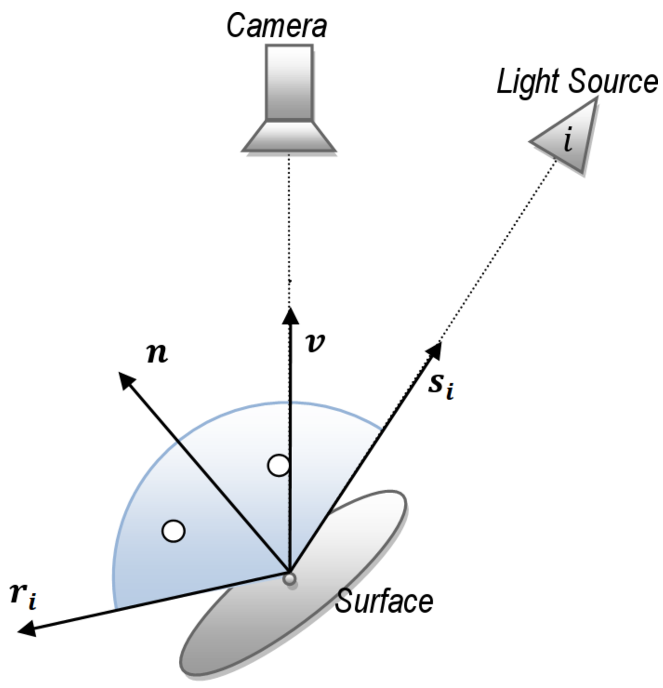


Figure 2.1: Vectors used in reflection models

is a function of the incident and the reflection directions. These directions are defined with respect to the surface normal. BRDF is the ratio of the radiance of the reflected light at the given reflection direction to the irradiance of incident light. For the surface patch, with the surface normal $\mathbf{n}$, isotropic BRDF can be expressed with $f_r(\mathbf{s}, \mathbf{n}, \mathbf{v})$ [16].

The surface radiance of reflected light along $\mathbf{v}$ is $L_v()$;

$$L_v(\mathbf{s}, \mathbf{n}, \mathbf{v}) = \int_\omega f_r(\mathbf{x}, \mathbf{n}, \mathbf{v}) L_s(\mathbf{x}) max(0, \mathbf{x}^T \mathbf{n}) d\omega \tag{2.5}$$

On the other hand, it is proved that the surface reflection is proportional to image irradiance [21]. This can be written as $i = kL_v()$ where I is the intensity value and k is the constant of proportionality. If I_{max} is the maximum intensity value of the image, the reflectance map $R(\mathbf{s}, \mathbf{n}, \mathbf{v})$ is defined as;

$$R(\mathbf{s}, \mathbf{n}, \mathbf{v}) = k/I_{max} \int_\omega f_r(\mathbf{x}, \mathbf{n}, \mathbf{v}) L_s(\mathbf{x}) max(0, \mathbf{x}^T \mathbf{n}) d\omega \tag{2.6}$$

In Eq. 2.6, the reflectance map of the source is related to the BRDF and the orientation of the surface. The BRDF will be defined later to substitute in this equation. A perfectly smooth planar surface reflects only in specular manner. All incident radiation from a point light source

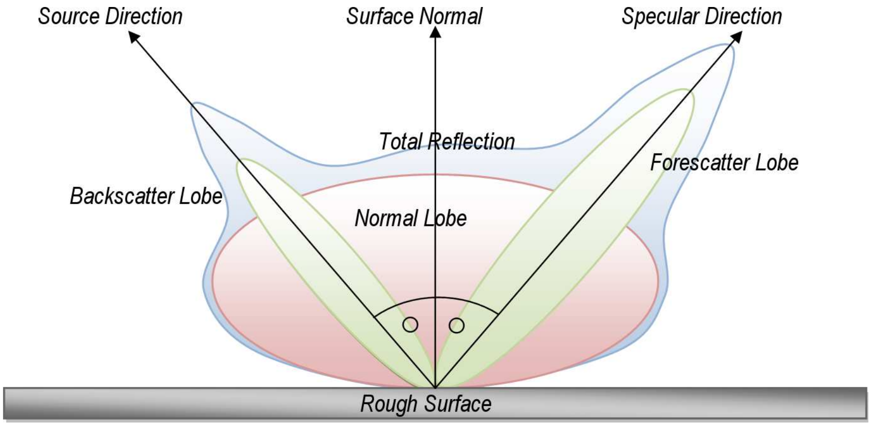


Figure 2.2: Lobes of total reflection

is reflected along this direction resulting a specular BRDF as in Eq. 2.7

$$f_v(\mathbf{s}, \mathbf{n}, \mathbf{v}) = \frac{\delta_\omega(\mathbf{n} - \mathbf{n}_s)}{2\mathbf{s}^T \mathbf{n}} \tag{2.7}$$

On the other side, a perfectly rough planar surface reflects only in diffuse manner. All incident radiation is from a point light source is reflected uniformly along all directions with constant BRDF as in Eq. 2.8.

$$f_v(\mathbf{s}, \mathbf{n}, \mathbf{v}) = 1/\pi \tag{2.8}$$

The uniform BRDF value is calculated by the fact that all the incident light is reflected by the surface.

Most solids have a combination of diffuse and specular reflection properties which is sum of three lobes, forescatter lobe, backscatter lobe and normal lobe (see Fig.2.2). The forescatter lobe represents the specular reflection of the surface. Normal lobe is diffuse component of the reflection map. Backscatter lobe is spread around source direction, most materials have very little backscatter. Only some paints have strong backscatter lobe [16].

The forescatter lobe is spread around the specular direction and spread function purely depends on the surface properties. It can be explained simply by modeling rough surface as a collection of infinitely small perfect specular reflectors called facets. Each facet inclined randomly around the surface normal. Facet models also assume that surface is isotropic that is uniform in all orientations. The incident light is reflected from each facets specularly. Since the facets inclination have a mean value around the surface normal, forescatter lobe have a monotonically decreasing distribution around the reflection direction.

Normal lobe is spread around the surface normal. The bulk scattering of surface is assumed to be the origin of the normal lobe. In the bulk scattering model, the surface is composed of an optically uniform material with facets in it. The scattering in this model is analyzed using radiative transfer function and can be approximated as Lambertian.

The BRDF of a real life surface can be modeled with the summation of specular, normal, forescatter and backscatter components, which is written as;

$$f_v(\mathbf{s},\mathbf{n},\mathbf{v}) = \mu_{spec} f_{spec}(\mathbf{s},\mathbf{n},\mathbf{v}) + \mu_{fsc} f_{fsc}(\mathbf{s},\mathbf{n},\mathbf{v}) + \mu_{norm} f_{norm}(\mathbf{s},\mathbf{n},\mathbf{v}) + \mu_{bsc} f_{bsc}(\mathbf{s},\mathbf{n},\mathbf{v}) \quad (2.9)$$

where μ's are the weights of each component in the BRDF.

If this generalized BRDF definition in Eq. 2.9, is placed in the reflectance map Eq. 2.6, the generalized reflectance map equation is created; however this generalization is beyond the scope of this work, and some assumptions will be made to simplify the general problem. Firstly, the specular reflection component of the general BRDF results an unbounded reflectance map and saturated intensity value. So $\mu_{spec} = 0$ is assumed throughout this thesis. Also rare backscatter is assumed to be $\mu_{bsc} = 0$, in this context. The simplified BRDF Eq. 2.10 and corresponding derived simplified reflectance map from Eq. 2.6 are formulated as in Eq. 2.11;

$$f_r(\mathbf{s},\mathbf{n},\mathbf{v}) = \mu_{fsc} f_{fsc}(\mathbf{s},\mathbf{n},\mathbf{v}) + \mu_{norm} f_{norm}(\mathbf{s},\mathbf{n},\mathbf{v}) \quad (2.10)$$

$$R(\mathbf{s},\mathbf{n},\mathbf{v}) = \rho_{spec}\Phi(\mathbf{n}_s^T\mathbf{n}) + \rho_{diff}(\mathbf{s}^T\mathbf{n}) \quad (2.11)$$

The constant terms ρ_{spec} and ρ_{diff} in reflectance map Eq. 2.11 are called the specular forescatter (specular in short) and diffuse albedo values of the surface respectively.

Maximum value of a reflectance map can be 1, which corresponds to the case all the light power from the source is reflected to camera. However, measured intensity values of the camera are not normalized to 1 but have a gain value depending on hardware and acquisition configuration of the camera. Hence the intensity values are normalized before any further calculation.

The term 'reflection model' is used for a mathematical function that generates intensity values with given surface normals and light source directions and some additional parameters. These additional parameters simulate the reflection properties of the surface material.

2.3.1 Lambert model

This simplest reflection model that only implements the diffuse reflection of the surface is Lambert reflection model and is formulated in Eq. 2.12. The constant diffuse BRDF is used for this model.

$$I_i = \rho_{diff}(\mathbf{s}_i^T \mathbf{n}) \quad : \quad (\mathbf{s}_i^T \mathbf{n}) > 0 \tag{2.12}$$

Lambert model is very simple and holds for many type of materials if the camera direction is away from the forescatter lobe [2]. If the camera direction is close to the specular direction, Phong or Torrance-Sparrow model may be used, that specular reflection is taken into account.

2.3.2 Phong model

Even though Phong model is physically not correct, since it is very fast to implement it on the hardware shader and includes the specular reflection component, it is widely used in computer graphics.

$$I_i = \rho_{spec}(\mathbf{r}_i^T \mathbf{v})^m + \rho_{diff}(\mathbf{s}_i^T \mathbf{n}) \quad : \quad (\mathbf{s}_i^T \mathbf{n}) > 0 \tag{2.13}$$

The forescatter lobe is modeled by a power of cosine function of angle between reflection direction $\mathbf{r}_i$ and camera direction $\mathbf{v}$. The ρ_{spec} and ρ_{diff} are specular and diffuse albedo values respectively in Eq. 2.13.

2.3.3 Torrance Sparrow Reflection Model

For visible light, Torrance Sparrow model is a good approximation of surfaces that can be modeled with facets [16].

$$I_i = \rho_{spec} e^{-m^2[\arccos(\mathbf{n}_h^T \mathbf{n})]^2} + \rho_{diff}(\mathbf{s}^T \mathbf{n}) \tag{2.14}$$

$$\mathbf{n}_h = \frac{\mathbf{c} + \mathbf{s}}{|\mathbf{c} + \mathbf{s}|} \tag{2.15}$$

Here the intensity value i is composed of specular and diffuse terms. The specular term depends on the angle between the normal reflection $\mathbf{n}_h$ and normal $\mathbf{n}$ direction. The normal reflection direction is bisector of the camera $\mathbf{c}$ and light source $\mathbf{s}$ directions where highlight will occur with full power if $\mathbf{n} = \mathbf{n}_h$.

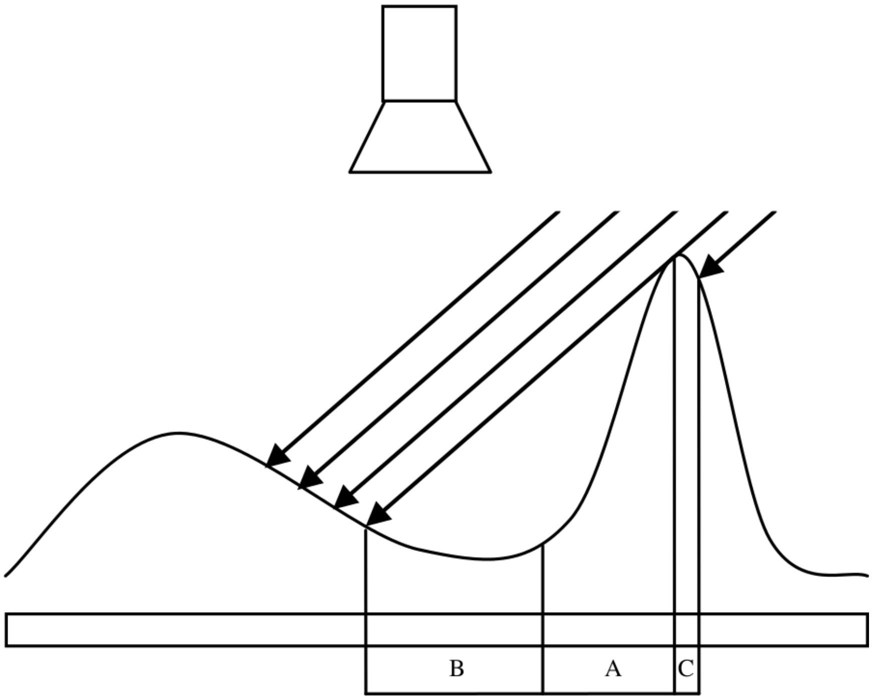


Figure 2.3: Cross section of the surface with self shadows (A), cast shadows (B) and highlights (C).

2.3.4 Shadows

Shadows can be classified in two groups; self shadow and cast shadow [2] as shown in Fig. 2.3. A self shadowed pixel is shaded by itself that can be formulated $(\mathbf{s}_i^T\mathbf{n}) < 0$. Geometrically, the angle between the surface normal vector and the light source direction is more than $90°$. With a single point light source, these pixels have zero intensity value.

Beside self shadows, the cast shadows may exist on images due to some other parts of the surface that occludes the light source. Although $(\mathbf{s}_i^T\mathbf{n}) > 0$ condition holds, zero intensity value will appear due to lack of illumination at these pixel coordinates.

Theoretically, since both of these shadows results in zero intensity value, they may be recognized from intensity values. In real images, however, shaded regions may not have perfect zero intensity values due to secondary illumination. In such cases, although self shadows may be extracted from surface normals, cast shadows can not be detected locally.

2.4 Photometric Stereo

In photometric stereo (PS), still images of a surface are captured under different illumination configurations [12]. Intensity values of image pixels can be related to surface normals and the reflection properties of the surface. Various reflection models were proposed to formulate the surfaces reflection properties. The most effortless model is Lambert reflection model since it provides a linear solution to the PS problem.

2.4.1 Linear Photometric Stereo

Linear PS method uses simple, linear Lambert reflection model. In this model, intensity of a pixel, illuminated by single point light source, is defined as follows;

$$i = \rho_d \mathbf{s}^T \mathbf{n} \tag{2.16}$$
$$\mathbf{s} = \mu \mathbf{d} \quad \mathbf{s}^T \mathbf{n} > 0, \quad |\mathbf{d}| = 1, \quad |\mathbf{n}| = 1$$

In Eq. 2.16, i is the intensity value, ρ_d is albedo, $\mathbf{n}$ is the unit surface normal vector, $\mathbf{s}$ is the illumination vector which is the multiplication of unit light source direction vector, $\mathbf{d}$, and illumination strength, μ. Intensity value is directly proportional to the cosine of the angle between the surface normal and the light source direction with a factor, albedo (ρ_d) times the illumination strength (μ).

Assuming that the illumination vector is known, Eq. 2.16 has three unknowns; one is the albedo and two comes from the surface normal. If three intensity values of the same surface point are captured using independent light sources, the linear equation system can be solved as:

$$\mathbf{I} = \rho_d \mathbf{S} \mathbf{n}$$
$$\rho_d = |\mathbf{S}^{-1}\mathbf{I}| \quad \mathbf{n} = \frac{\mathbf{S}^{-1}\mathbf{I}}{\rho_d} \tag{2.17}$$
$$\mathbf{I} = \begin{bmatrix} i_1 & i_2 & i_3 \end{bmatrix}^T, \mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \mathbf{s}_3 \end{bmatrix}^T$$

In Eq. 2.17, $\mathbf{I}$ is intensity vector composed of the intensity values, i_m from m^{th} image, illuminated by $\mathbf{s}_m$. Similarly, $\mathbf{s}_m$, illumination vectors for each light sources are concatenated to form the illumination matrix, $\mathbf{S}$. If more than 3 images are given, the equation system can be solved using the $\mathbf{S}^{\dagger}$ (pseudo inverse of $\mathbf{S}$) calculated as in Eq. 2.18.

$$\mathbf{S}^{\dagger} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T \tag{2.18}$$

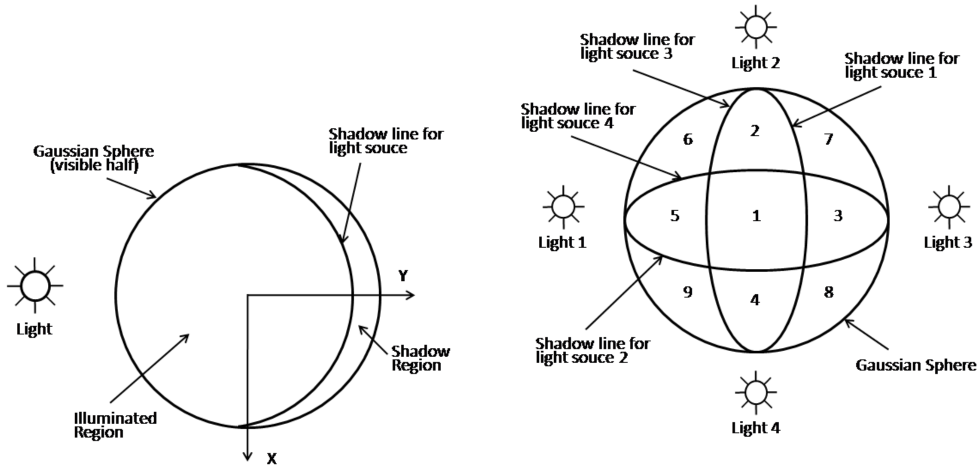
2.4.2 Optimal Illumination For PS

In 2003, Spence and Chantler, published the first paper on the optimal illumination for three image PS [22]. Later in 2006, they extended their work with intense sensitivity analysis [23]. All the derivations were made with the assumption that the surface is Lambertian (no highlights) and there is no shadow. Additive, zero mean Gaussian noise on images, representing CCD noise, was considered as the source of all resultant normal error in their work. For three images, light sources were placed with equal polar angles of 120^o optimally. Zenith angles of the light sources (Z_{source}) were the same and were chosen depending on the surface roughness (55^o for the rough surfaces and 90^o for the smooth ones).

Later, Drbohlav and Chantler further extended the previous works with more than three images PS [24]. Again only Gaussian CCD noise was modeled. For more than three images ($n_i > 3$) optimal zenith angle was found to be 54.74^o and light sources were equally spaced with $(360/n_i)^o$ tilt angles. Besides the circular light placement with constant zenith angle Z_{source}, they also investigated the case where one of the light sources was at 0^o zenith angle position (coincide with the camera) and others are on the circle. With the additional 0^o zenith angle light source, the optimal Z_{source} changed, but when n_i goes to infinity, it converges to the same optimal value given above.

Concurrently, Barsky and Petrou published the general design issues paper for color PS, which is an extension of their previous paper [2], [25]. The CCD noise, illumination estimation errors, shadows and highlights were investigated. Similar to Sakane and Sato illumination configuration was designed to minimize singular value decomposition (SVD) condition number [26]. They suggested a circle of light sources with constant zenith angles in the range 30^o to 45^o. This is less than the previously calculated optimal, since shadows were considered.

Lastly in 2007, Sun et al, showed that the orthogonal light source is the best for 3 image case with no highlights and no shadows exist in the images [3]. Also the uncertainty of the results depends on surface albedo in their work.



(a) (b)

Figure 2.4: Segmentation of different regions defined in [1]

2.4.3 Masked PS methods

The main idea of these methods is that nearly all non-Lambertian materials behave close to Lambertian, if the surface orientation is far from specular region [16]. So it is very attractive to solve PS problem linearly, by eliminating non-Lambert pixels and using Lambert pixels.

In 1982, Coleman and Jain implemented the first filtered PS method using four images [17]. This method can filter one specular pixel among four. It assumes that there is no shadow in images. With some restrictions on four light source positions, this condition can be satisfied easily. Their method is based on the assumption that surface albedo is constant for any light combination regarding there is no shadows and surface is Lambertian. Three image PS is calculated four times and four albedo values and normal vectors are found. If the variance of albedo values is less than a threshold, algorithm decides that there is no highlight and the average of all normals is the resultant normal vector. On the other hand, if there is a highlight pixel then the solution with the smallest albedo is used.

Later, Solomon and Ikeuchi built up a new filtered PS method using four images [1]. This method can handle two shadow pixels among four. In this two-image PS case, albedo is assumed to be constant and known. Although the algorithm states what to do when there is a shadow or highlight, it does not inform about how to sense them.

Algorithm applies a different solution for each region defined as with the shadows as the fig-

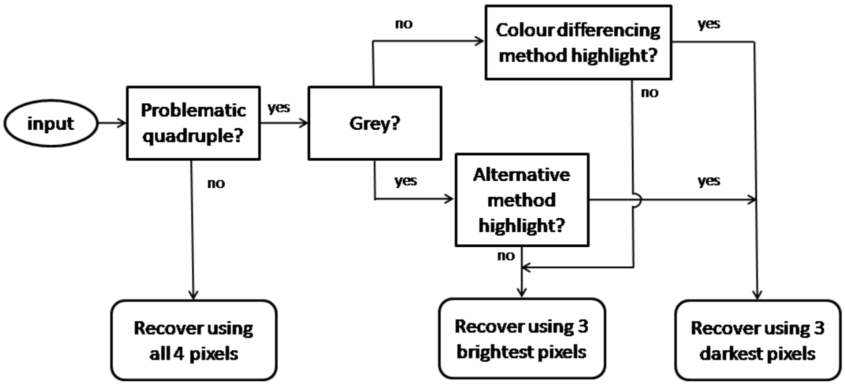


Figure 2.5: General flow chart of method defined in [2]

ure. Generally when there is no shadow, for a region Coleman, Jain [17] specular detection method is used with adaptive threshold depends on camera noise. When the region is shadowed in only one of the images, region lighted by the opposite source has highlight. So, two remaining lights with constant albedo are used. When there are two shadowed pixels, algorithm evaluates normals with them, hoping there is no highlight at the rest of the images.

In 2003, Barsky and Petrou, built new filters for PS method [2] defined by Coleman and Jain [17]. Beside highlights also shadows can be sensed and filtered with these methods. Method is capable of sensing one erroneous pixel among four. Two erroneous pixels may cause filters to make false decisions that results a worse case than using all four pixels together.

They defined three filters, non-Lambert quadruple detection filter, color differencing highlight filter, alternative highlight filter. Method works as follows;

Non-Lambert quadruple detection filter utilizes linear dependence of source directions which indicates that any four vectors in three-dimensional world are linearly dependent.

$$a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + a_3\mathbf{s}_3 + a_4\mathbf{s}_4 = \mathbf{0} \tag{2.19}$$

If we multiply both sides with local albedo and surface normal;

$$a_1\rho_{diff}(\mathbf{s}_1\mathbf{n}) + a_2\rho_{diff}(\mathbf{s}_2\mathbf{n}) + a_3\rho_{diff}(\mathbf{s}_3\mathbf{n}) + a_4\rho_{diff}(\mathbf{s}_4\mathbf{n}) = \mathbf{0} \tag{2.20}$$

This is equivalent to;

$$
\begin{aligned}
a_1 I_1 + a_2 I_2 + a_3 I_3 + a_4 I_4 &= \mathbf{0} \\
\mathbf{aI} = \mathbf{0} \quad \mathbf{a} &= [a_1 a_2 a_3 a_4]
\end{aligned}
\tag{2.21}
$$

Vector defines **a** hyper plane in four dimensional space and can be computed directly from source directions. If all four intensities are perfectly Lambert, above equation must hold. For near-Lambert intensities, ”Lambertian error” is defined as $(\mathbf{aI})^2$. If Lambert error is less than a threshold, all four intensities are near-Lambert.

Color differencing highlight filter employs different color characteristics of specular and diffuse reflection. While diffuse reflection has chromaticity of body color, the specular reflection has the chromaticity of illuminant.

If both illuminant and body have close colors, color differencing does not work. In these cases, alternative highlight filter is used. Alternative method is intersection of two threshold filters. If brightness of a pixel is high and normal vector is close to bisector of light source and camera vector, it is marked as specular and other three pixels are used in PS. Error definition in this paper is different from the majority of the PS literature. The count of erroneously constructed pixels (over a threshold value) instead of mean normal error of all pixels is used. Also a low (30°, less than optimal) zenith angle light sources are used to reduce the risk of multiple errors for pixel.

Later, Chandraker et al, published a four or more image PS method [27]. They implemented a shadow detection filter (term light source visibility is used in paper). Shadow graphs are similar to shadow maps (also called z-buffer in computer graphics), which keep the information of visibility of each pixel from camera or from light sources. Shadow graphs were employed as constraints in integration of normals to create surface height data.

At the same time, Sun et al, proposed a hierarchical filtering strategy to eliminate shadows and highlights, from 6 images [3]. They claimed that for any convex object, at minimum 6 light sources is needed to solve for entire visible surface. 6 light sources are placed on a circle with 45° zenith angle. In this illumination configuration, at most one specular and two shadows can occur for each pixel on a convex surface.

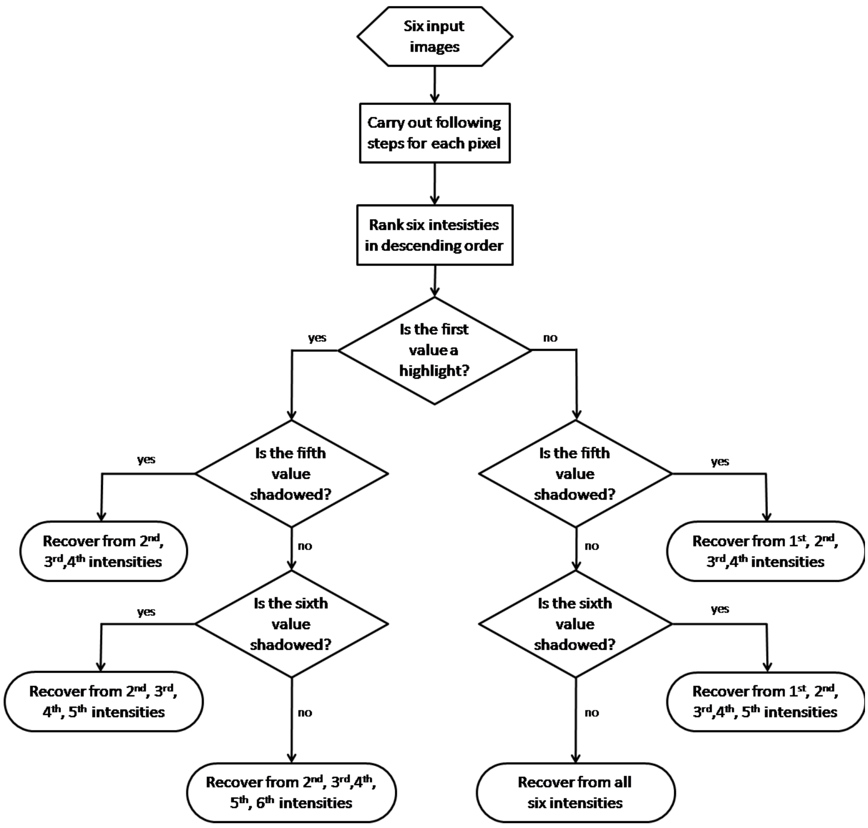


Figure 2.6: General flow chart of decision making, defined in [3]

The flow chart of the algorithm is seen in Fig 2.6. Sun et al assume that highlight occurs at the brightest pixel and shadows occur at the darkest pixels. With these assumptions when the pixels are sorted with their intensity values, 2nd, 3rd and 4th pixels are guaranteed to be Lambertian. Brightest one pixel and darkest two pixels are tested one-by-one with Non-Lambert quadruple detection filter [2].

Later, Argyriou and Petrou, enhanced their previous method with shadow mapping (z-buffer) for four images [4]. The new method includes a recursive loop that gradually finds shadows and highlights, corrects normals and heights. It starts with eliminating non-Lambert pixels with Non-Lambert quadruple detection filter. These pixels are unreliable, so they are interpolated with neighbor pixels. The recursive section searches the non-Lambert quadruples for shadows with shadow mapping filter. At the rest of the unreliable pixels, brightest pixels in the quadruples are classified as highlight. Shadows and highlights are filtered and PS is calculated again.

Here shadow mapping filter works for both cast and self shadows. To find shadowed pixels, firstly height map is generated from normal map. Z-Buffer of height map, which contains closest distance value to view point, is rendered from each light source view point. Simple distance comparison is used to recognize shadowed pixels.

In 2009, Miyazaki et al, developed a median PS method [28] using more than four images. This is again a recursive process that terminates when the changes in calculated normals are less than a convergence threshold. For each pixel, surface normal is estimated from median candidate of all normals which each one is generated with different 3 light combinations. This is said to be less sensitive to outliers compared to mean normal. Also neighbor surface normals are added to median set. At the end, median normals are again weighted with neighbors, which is a smoothing process on normals. This is done for while all normals converge.

Lastly, in 2010, Argyriou et al, extended their previous works for four images to any number of images [29]. They proposed a recursive algorithm removing highlights and shadows by using least squares error similar to intensity error in this work.

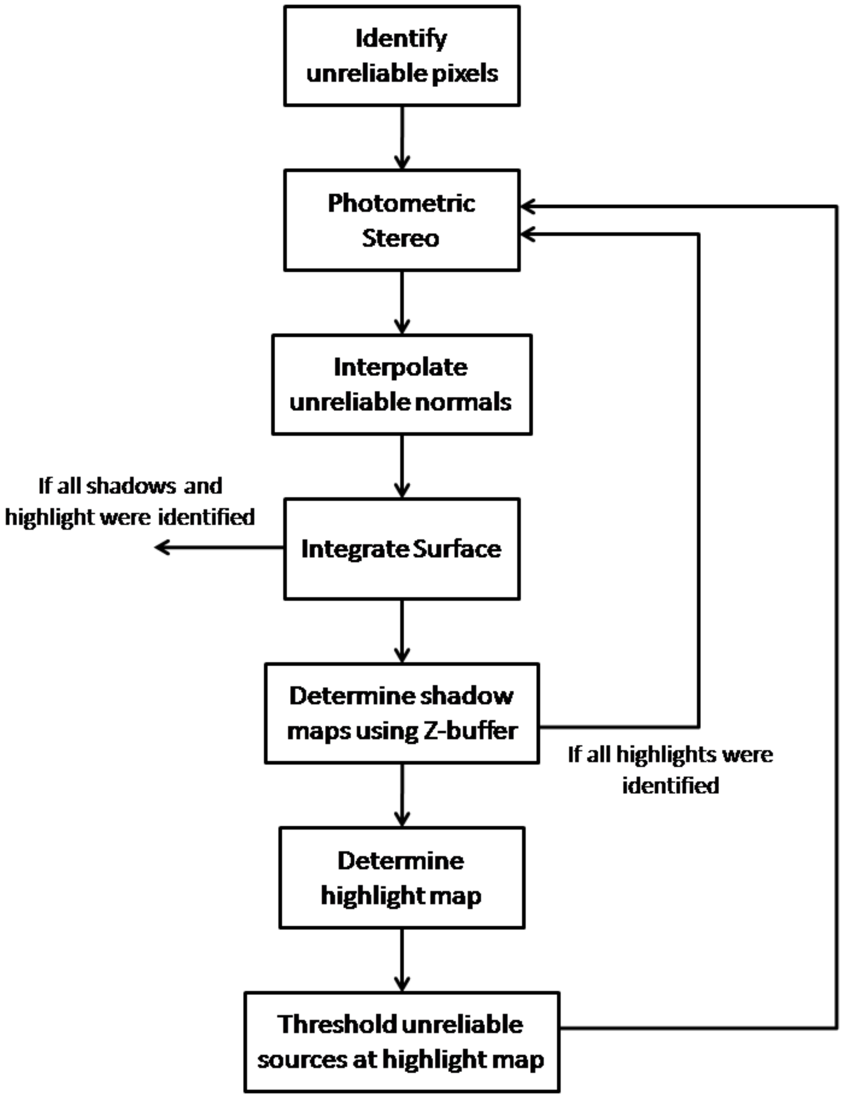


Figure 2.7: Recursive flow chart of method, defined in [4]

2.4.4 Non-Linear PS

In this part, relatively complex methods are presented. These methods do not only reconstruct normals and heights, but are also interested in complex surface reflectance properties. They employ long optimization solutions which run for hours. All these works are generally aim to solve only highlight problems. Shadows are discarded most of the time.

Hertzmann and Seitz became the first who used a reference sphere having the same BRDF with target object [30]. The base of method depends on orientation-consistency cue, which is two points with the same surface orientation reflect the same light toward the viewer. So, both reference and target are imaged at the same time. Approximate nearest neighbor search structure finds the corresponding points between reference and target. All the shadows are ignored. 14 images are used at total and reported to run about 5 hours of computation.

Later, Goldman et al, advanced the research [31] with similar constraints. They have calculated 3D and BRDF at the same time [32] in 2010. The non-linear optimization methods run on 12 images about 5-10 hours.

There are other methods that do not use parametric reflectance model [33] [34] [35]. They employ non-parametric BRDF measurements from many images counting up to thousands.

2.5 Uncertainty of Normals in PS

Woodham [12] who first defined the photometric stereo method, also placed the basic restriction for location of the three light sources. The illumination matrix has to be inverted; so it must be non-singular. This means that light sources must not be placed on a line. Besides the condition of non-singularity, he stated that orthogonal three light sources combination will be optimal for Lambert reflection model.

Later, Sakane and Sato [26] implemented an active PS that optimizes camera and light positions. In their work, reliability or accuracy of the solution is evaluated for any number of light sources by singular value analysis of the illumination matrix. Lambert reflectance PS for N images can be formulated as follows;

$$\mathbf{I} = \rho_{diff}(\mathbf{S}^T\mathbf{n}) \quad \mathbf{I} = [i_1 \ldots i_N]^T \quad \mathbf{S} = [\mathbf{s}_1 \ldots \mathbf{s}_N]^T \tag{2.22}$$

Table 2.1: Condition values of different light source configurations

Light source count	**Zenith angle**	**Condition value**
4	5	16.1645
4	15	5.2779
4	30	2.4495
4	45	1.4142
4	60	1.2247
8	5	16.1645
8	15	5.2779
8	30	2.4495
8	45	1.4142
8	60	1.2247

In equation (2.22), **I** is intensity vector, **S** is illumination matrix, **n** is surface normal and ρ_{diff} is local diffuse albedo of material. When singular value decomposition is applied to **S**, condition value can be calculated as ratio of maximum to minimum singular values.

$$\mathbf{S} = \mathbf{U}\Sigma\mathbf{V}^T \quad \Sigma = \mathbf{diag}(\sigma_1\sigma_2\sigma_3) \quad \mathbf{cond}(\mathbf{S}) = \frac{\sigma_{max}}{\sigma_{min}} \tag{2.23}$$

If $\Delta\mathbf{n}$ and $\Delta\mathbf{I}$ are defined as errors in normal and intensity vectors, following inequality relates these errors with condition value.

$$\frac{\Delta\mathbf{n}}{\mathbf{n}} <= \mathbf{cond}(\mathbf{S}) * \frac{\Delta\mathbf{I}}{\mathbf{I}} \tag{2.24}$$

Therefore, if the condition value is smaller, PS solution will estimate more reliable normal vectors with the errors in the intensities.

As an example, condition value of 4 and 8 light sources are presented in the Table 2.1. The number of light sources does not change the condition value. Also distant light sources, with larger zenith angles, have better condition value.

Woodham, in 1994 [36], published a real time PS that uses lookup table relating 2 or 3 intensity values to surface normals. Calibration sphere with the same target material was used to create the lookup table of desired material for the given illumination configuration. As a result surface gradient and distance measure to surface was outputted. With the usage of lookup table, he eliminated the Lambert reflectance assumption, which solves the highlight problems. Shadow problems were not solved but a local confidence estimate was generated.

Lookup table is defined from 3D intensity space to 2D gradient space and 1D confidence estimate. Lookup table is created in three steps. Firstly measured intensity values and surface gradients are stored in the table. In the second step, gaps in the 2D gradient space are filled by interpolation for both intensities and gradients. All the resultant table entries are direct hit, so has a 0 confidence estimate value. In the third step, 3D intensity space is filled with closest gradient values and distance to closest measured data is local confidence estimate.

Results of the algorithm are promising, and the algorithm is practically easy to implement.

In 2003, Spence and Chantler [22], first published a paper optimal illumination for three image PS. Later in 2006 [23], they extended their work with intense sensitivity analysis. All the derivations are made with assumption that surface is Lambertian and there is no shadow. Only Gaussian noise on images is the source of resultant normal errors. For three images, optimally light sources placed with equal tilt angles of 120 degrees. Zenith angle (term slant angle used in the paper) depends on surface roughness and 55 degrees rough surfaces and 90 degrees for smooth ones.

In their sensitivity analysis, three light positions that minimize the sum of the variances of each normal direction components. The ratio of σ_{n_x} variance of x component of scaled normal vector (product of albedo and unit normal vector), to σ_i variance of Gaussian noise on images is formulated as follows;

$$\frac{\sigma_{n_x}}{\sigma_i} = \sqrt{(\frac{\partial n_x}{\partial i_1})^2 + (\frac{\partial n_x}{\partial i_2})^2 + (\frac{\partial n_x}{\partial i_3})^2} \tag{2.25}$$

Here, partial derivatives of scaled normal with respect to intensity values are called sensitivity expressions. There are 9 of them for three light source PS.

Followings are the figure of merits for rough and smooth surfaces; σ_{n_z}/σ_i term is ignored for smooth surfaces.

$$M_r = \frac{\sigma_{n_x}}{\sigma_i} + \frac{\sigma_{n_y}}{\sigma_i} + \frac{\sigma_{n_z}}{\sigma_i} \quad M_s = \frac{\sigma_{n_x}}{\sigma_i} + \frac{\sigma_{n_y}}{\sigma_i} \tag{2.26}$$

Sensitivity is found to be inversely proportional to sine of zenith angle for x and y normal components and inversely proportional to cosine of zenith angle for z normal component. When z is ignored for smooth surfaces, sensitivity is inversely proportional to sine of zenith angle.

Also signal to relight error ratio (**SER**) is defined as the error definition for real image tests.

$$\mathbf{SER} = 10\log\left(\frac{\mathbf{var(I)}}{\mathbf{var(I - I}_{relight})}\right) \tag{2.27}$$

Variances of pixel intensities are experimentally calculated from 10 images of same light source. This error definition includes only CCD noise error in images.

Later, Drbohlav and Chantler [24] further extended their previous works with more than three images PS. Again only Gaussian CCD noise is modeled. This time, the uncertainty of normals, mean value of squared distance of scaled normals to reference, are used as a figure of merit. For three images same results are found. For more than three images (image count, N ¿ 3) optimal slant angle (zenith angle) is 54.74 and light sources are equally spaced with 360/N tilt angles. Minimum uncertainty will be $9\sigma^2/N$ for N images PS, with zero mean σ^2 variance Gaussian noise.

Beside the circular light placement pattern with constant slant angle, they also investigated the case that one of the light sources is at 0 slant angle position (coincide with the camera) and others are on circular pattern. Optimal slant angle of circular lights changed, but when n goes to infinity, slant angle converges to same optimal value.

Concurrently, Barsky and Petrou [25] published a general design issues paper for color PS that is an extension of their previous paper [2]. Synthetic images are used to create fine control errors. CCD noise, illumination estimation errors, shadows and highlights were investigated. Similar to Sakane and Sato [26] illumination configuration is designed to minimize SVD condition number. They suggested a circular pattern of light sources with constant slant angles in range 30° to 45°. This is less than previously calculated optimal, because increasing shadows with increasing slant angles are considered.

Lastly in 2007, Sun et al, [3] presented a good proof showing that the orthogonal light source is the best for 3 image case. The Lambertian reflectance and no shadows assumptions were used again. They insisted that uncertainty also depends on surface albedo. This was not that obvious in the previous works related to PS, since they all use scaled normal vectors as uncertainty value. Here dark colored surfaces will have errors compared to lighter colored ones.

In their work, uncertainty of normals is defined as covariance matrix of normals in the presence of noise in images. Objective function for optimality is defined as the trace of the covariance matrix of normals.

2.6 Error Definitions

The primary output of PS is the surface normal vectors, so the resultant performance was measured by considering error on normal vectors in this study. The NE is the angular deviation of the normal vectors of PS ($\mathbf{n}_r$), from the originals ($\mathbf{n}_o$) as in Eq. 2.28. However, for real image tests, the original normal vectors are not known. In that case, the intensity error (IE) definition, that is the average of the residual errors in all images, are used as in Eq. 2.29. This definition is extended with weights to evaluate the weighted PS method as in Eq. 2.30.

$$NE = \frac{\sum_{Surface} \angle(\mathbf{n}_r, \mathbf{n}_o)}{\sum_{Surface}} \tag{2.28}$$

$$IE = \frac{\sum_{Images} |\mathbf{I} - \mathbf{ISS}^{\dagger}|}{\sum_{Images}} \tag{2.29}$$

$$IE = w_1 IE_1 + w_2 IE_2 \tag{2.30}$$

Intensity error is defined on image intensities, and without solving PS, it can be calculated from images directly (term relight error is used for intensity error in [22]). Since only images are used, reference 3D data is not needed to calculate it. Intensity error definition used here is residual error of least square estimation problem.

(a)

(b)

Figure 2.8: Normal error of PS (a) and square of intensity error calculated (b).

$$\mathbf{I} = \rho_{diff}(\mathbf{S}^T\mathbf{n}) \quad \mathbf{I} = [i_1 \ldots i_N]^T \quad \mathbf{S} = [\mathbf{s}_1 \ldots \mathbf{s}_N]^T \tag{2.31}$$
$$\rho'_{diff} = |\mathbf{S}^P\mathbf{I}| \quad \mathbf{n}' = \frac{\mathbf{S}^P\mathbf{I}}{\rho'_{diff}} \quad \mathbf{S}^P = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T$$
$$\mathbf{R} = \mathbf{I} - \mathbf{S}^T(\rho'_{diff}\mathbf{n}')$$
$$\mathbf{R} = (\mathbf{1} - \mathbf{S}^T(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T)\mathbf{I}$$

Linear system of Lambertian reflectance equations in more than three images PS is solved with least square estimation optimally. ρ'_{diff} and $\mathbf{n}'$ are optimal estimations of the albedo and the surface normal respectively. The intensity error is defined as $|\mathbf{R}|$, where $|.|$ is column vector norm operator that calculate norm of residual errors of all pixels, resulting a scalar non-negative error value for each pixel coordinate. L_1, L_2, L_∞ vector norm definitions can be used depending on needs.

Square of intensity error is good approximation for the trace of covariance of normal vectors. This would be theoretically true if errors in images are independent, zero mean Gaussian functions. In that case, covariance of normal vectors must be diagonal matrix according to least square estimation theory.

Fig. 2.8, presents similarity of normal error and intensity error up to a scale. The plotted normal error image is the norm of vectorial difference of original and calculated normal.

Notice that normal errors on sharp edges of Mozart does not exists in intensity error.

$$QF = \frac{TP}{TP + FP + FN} \tag{2.32}$$

Besides the normal error and intensity error, the performance of the mask generation can be measured by comparing the calculated masks with the initially known, synthetic shadows and highlights. For this purpose, the quality factor (QF) given in Eq. 2.32 is used. Here, the number of true positives (TP) is defined as the pixel count of matching original and final masks. The number of false negative (FN) is the pixel count of unrecognized shadows and highlights and the false positive count (FP) is the pixel count of masked regions that are actually not shadows and highlights. The QF is 1 if all the pixels of mask match with the desired mask. With each false decision, the QF drops down to 0.

CHAPTER 3

UNIFIED PHOTOMETRIC STEREO

3.1 Introduction

The masked methods listed above are combined within a unified framework that we name as unified PS. Firstly, this framework provides a platform for fair comparison among various methods. This is achieved by using the same codes for PS calculation except the masking procedures. Secondly, using unified PS framework, masks can be fused in different ways using logical operations easily. Hence, different combinations can be selected an evaluated based on the characteristics of each mask and requirements of a specific application to find the most robust PS method.

3.2 Flow Diagram for Unified PS

The most generic flow diagram of unified PS is presented in Fig. 3.1. The first step in the flow is to calculate setup parameters and calibrate images. These calibrated images are then fed into subset selection algorithm that forms combinations of images. Each of these combinations will be the input of image masking stage. In the next step, PS and normal masks are calculated iteratively. Finally, results of previously selected subsets are fused to create resultant normals.

1. *Calculate setup parameters*: These are the parameters such as pixel size of camera, light source directions and powers. Parameters related to light sources will be used in PS solution whereas pixel size is required for height map calculation.

2. *Calibrate Images*: Image calibration step corrects non-linear radiometric response of

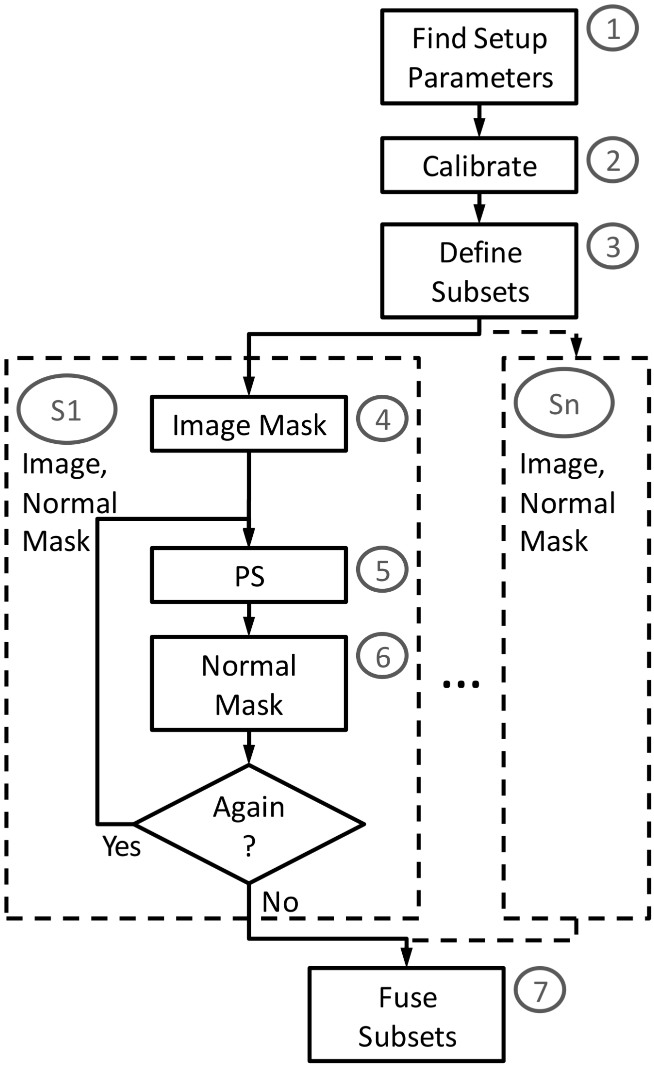


Figure 3.1: Flow chart of unified framework PS. Image masks work in Step 4. Normal masks are applied in Step 6 inside the iterative loop. Subset masks are utilized in both Step 3 and Step 7. Step 4,5,6 are executed for each subset of images.

camera using response curve.

3. *Define Subset Images*: Some masking approaches suggests to solve the overdetermined PS problem by dividing images into subsets. These subsets will be merged in later steps after masking stages. Any combination with at least three images among calibrated images is valid as a subset.

4. *Calculate Image Masks*: In this step, image pixels are masked with respect to their intensity values. Masked pixels are not used in PS calculation.

5. *Photometric Stereo (PS)*: PS is calculated as explained in the previous section. Illumination matrix **S** and intensity vector **I** are formed with respect to calculated mask at that point. The mask must leave at least three related illumination vectors in **S** and intensity values in **I** for each point.

6. *Calculate Normal Masks*: At this step, the mask is generated from normal vectors calculated in the previous PS stage. Height values of the surface may or may not be required depending on the selected normal mask type. Once normal mask is calculated, Step 5 and 6 can be iterated to improve mask quality, thus the normals.

7. *Fuse Subset Results*: PS results of subset images are merged in this step to form final normals. The fusion (in Step 7) and define subset (in Step 3) strategy depends on the subset mask utilized.

Some steps of unified PS can be omitted if they are not used. For example, Coleman et al. [17] uses only subset masks, so Step 4 and 6 are omitted. On the other hand, Argyriou at al. [4] method does not use subsets, thus Step 3 and 7 are omitted, but both Step 4 and 6 are utilized.

The unified PS framework is capable of employing more than one masks at once. These masks are logically AND, so that if a pixel is masked with any of the utilized masks, it will be excluded from the PS calculation. In other words, only the consensus set of all masks are used. The AND operation is also required since some of the masks are focused on resolving different types of errors, like highlight errors and shadow errors.

There are some problematic cases in PS solutions. Firstly, if surface albedo is close to zero at a pixel, all intensity values from images will be approximately zero for that pixel. Since no valid intensity data is acquired for those coordinates, the normal vector **n** can not be calculated. Perpendicular normal vector $(0, 0, 1)$ is used in such cases.

Although the normals can be improved with the masks, their usage may bring exceptional cases that must be carefully handled. In order to have a solution, at least three pixels are required. If less than three pixels remain after masking, three pixels with least error must be used anyway. The masked pixels with least residual errors are employed to complete pixel count to three. In these cases, the residual errors of pixels are calculated as in Eq. 3.1

$$\begin{aligned} \mathbf{R} &= |\mathbf{I} - \mathbf{L}| \\ \mathbf{L} &= \mathbf{IS}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T \end{aligned} \tag{3.1}$$

where **R** is residual error vector, and **L** is Lambert intensity vector.

3.3 Masks In Unified Masked PS

Three classes of masks are defined in unified PS. These are image masks, normal masks and subset masks.

3.3.1 Image Masks

Two image mask algorithms are implemented in Step 4. These algorithms utilize only intensity values of the images and they are not iterative.

3.3.1.1 Threshold Mask (Th)

This mask estimates both highlights and shadows with respect to their intensity values. Since the lightest and the darkest pixels of the images are probably the specular and shadowed regions, they are masked by threshold values as in Eq. 3.2.

$$F_{Th} = \begin{cases} 0 & i < k_{low}(i_{max} - i_{min}) \\ 0 & i > k_{up}(i_{max} - i_{min}) \\ 1 & otherwise \end{cases} \tag{3.2}$$

In Eq. 3.2, threshold values are defined for each image separately, with respect to dynamic ranges of the images, $(i_{max} - i_{min})$.

3.3.1.2 Non-Lambert Quadruple Mask (NL)

Non-Lambert quadruple mask detects if there is either a highlight or a shadow among four [2]. Additional information is needed to indicate the specific erroneous pixel. This mask uses the linear dependence of source directions which states that any four vectors in three-dimensional world are linearly dependent as in Eq. 3.3.

$$a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + a_3\mathbf{s}_3 + a_4\mathbf{s}_4 = \mathbf{0} \tag{3.3}$$

If both sides of Eq. 3.3 are transposed and multiplied with local diffuse albedo, ρ_d and surface normal, $\mathbf{n}$;

$$a_1\rho_d\mathbf{s}_1^T\mathbf{n} + a_2\rho_d\mathbf{s}_2^T\mathbf{n} + a_3\rho_d\mathbf{s}_3^T\mathbf{n} + a_4\rho_d\mathbf{s}_4^T\mathbf{n} = 0 \tag{3.4}$$

This is equivalent to;

$$\begin{aligned} a_1i_1 + a_2i_2 + a_3i_3 + a_4i_4 &= 0 \\ \mathbf{a} &= [\begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \end{array}] \\ \mathbf{I} &= [\begin{array}{cccc} i_1 & i_2 & i_3 & i_4 \end{array}]^T \\ \mathbf{aI} &= 0 \end{aligned} \tag{3.5}$$

In Eq. 3.5, the vector $\mathbf{a}$ can be computed directly from source directions. If all four intensities perfectly satisfy Lambert reflection model, above equation must hold. For near-Lambert cases, "Lambertian error" is defined as $(\mathbf{aI})^2$. Non-Lambert quadruples can be detected by comparison of the Lambertian error with a threshold value, as in Eq. 3.6.

$$F_{NL} = \begin{cases} 0 & (\mathbf{aI})^2 > t_{NL} \\ 1 & otherwise \end{cases} \tag{3.6}$$

3.3.2 Normal Masks

Normal masks use both image intensity values and calculated normals. Since normals depend on masks and masks depend on normals, these masks must be calculated iteratively.

3.3.2.1 Self Shadow Mask (SS)

This mask works on self shadowed pixels. Self shadows occurs when surface normal vector makes an angle more than 90^o with respect to the source direction. This condition can be checked using the dot product of surface normal and source direction vectors as in Eq. 3.7.

$$F_{SS} = \begin{cases} 0 & \mathbf{s}^T\mathbf{n} > 0 \\ 1 & otherwise \end{cases} \tag{3.7}$$

3.3.2.2 Reflection Mask (Re)

Similar to the self shadow mask, the reflection mask uses surface normal and source direction vectors to find highlights. If the reflection direction of the light from the surface is close to the camera direction, specular reflection is highly probable [2]. This principle is checked by using a threshold value , t_{Re}, for the angle between surface normal vector and bisector vector of the source direction and the camera direction as in Eq. 3.8.

$$F_{Re} = \begin{cases} 0 & \mathbf{n}^T \dfrac{\mathbf{s}+\mathbf{c}}{2} < t_{Re} \\ 1 & otherwise \end{cases} \tag{3.8}$$

where $\mathbf{c}$ is camera direction and $(\mathbf{s} + \mathbf{c})/2$ term is the bisector of the source and the camera directions.

3.3.2.3 Cast Shadow Mask (CS)

Cast shadowed pixels are in shade by a near peak that occludes the light source. Unlike self shadows, cast shadows are not caused by the local pixel data, hence, only local clue of a cast shadow is its intensity value. The actual intensity value, I is compared with the expected Lambert intensity value and if it is less than expected, the pixel is masked as a cast shadow as in Eq. 3.9.

$$F_{CS} = \begin{cases} 0 & i - \rho_d \mathbf{s}^T\mathbf{n} < k_{CS}(i_{max} - i_{min}) \\ 1 & otherwise \end{cases} \tag{3.9}$$

where, the cast shadow threshold (t_{CS}) is chosen to be a fraction of the image dynamic range.

3.3.2.4 Highlight Mask (Hi)

Similar to the cast shadow mask, highlight mask uses the difference of actual intensity and the expected Lambert intensity values. If the actual intensity value, I is greater than the expected value, it is masked as highlight as in Eq. 3.10.

$$F_{Hi} = \begin{cases} 0 & i - \rho_d \mathbf{s}^T \mathbf{n} > k_{Hi}(i_{max} - i_{min}) \\ 1 & otherwise \end{cases} \tag{3.10}$$

Highlight mask can be used in conjunction with reflection mask to increase its accuracy.

3.3.2.5 Shadow Mask (Sh)

Shadow mask works on both cast and self shadows by using height values of the surface that are calculated by integrating surface normal vectors [4], [37]. The shadow map, i.e. the visibility map of a coordinate from the source position is used as explained in [38]. The camera is positioned at the each light source and the Z-Buffer is rendered. Z-Buffer contains distances between the camera and non-occluded points of the surface corresponding to pixels in the image. Later distance of the surface to the source is compared with the Z-Buffer for each pixel coordinate. If the distance is larger than the Z-Buffer value, this means that there is a blocking region between this coordinate and the light source, so this pixel is shadowed.

3.3.3 Subset Masks

Subset masks are applied in both Step 3 and 7. In Step 3, subsets of images are formed. Each subset is solved with PS exclusively. The results of all subsets are fused in the Step 7 with various methods.

3.3.3.1 Coleman and Jain Mask (CJ)

This mask detects and masks out a single highlighted pixel among four pixels. In Step 3, four triplets of 4 images are formed. These four triplets are fed to PS separately without any masks. In Step 7, the four resultant normals are fused with respect to their albedo values. If the standard deviation of albedo values are less than a threshold, there is no highlight in pixels, hence resultant normals are averaged. Otherwise, the normal with the smallest albedo is selected as the result.

3.3.3.2 Extended Non-Lambert Quadruple Mask (xNL)

Barsky et al. [2] and Argyriou et al. [4] used Non-Lambert quadruple mask (NL) with exactly four images defined in image masks. Non-Lambert quadruple Mask can be extended to more than four images, eg. Sun et al. [3] used this mask with exactly six images. The six intensity values from different images are sorted from darkest to brightest. For their illumination configuration and a convex surface, at most, a single highlight with brightest intensity, and two shadows at the darkest two intensities can occur, leaving at least three correct pixels for PS. Possible erroneous pixels are eliminated with checking the Non-Lambert quadruple masks, calculated from the sorted quadruples. Later, Argyriou et al. updated Non-Lambert quadruple mask to arbitrary number of illuminants [29] with single highlight constraint.

This six-image approach is generalized to any number of images in this study with some additions similar to our previous work [18]. Assume that, *k* images are acquired to be used in PS. For each point, brightest pixels are possible highlights and darkest ones are possible shadows. Similar to Sun et al. intensities are sorted from darkest to brightest. For each consecutive quadruples in the sorted intensities "Lambertian error" is calculated. Thus, for *k* sorted intensities, *k-3* Lambertian errors are calculated. These Lambertian errors are compared to a threshold to find quadruples with errors. The threshold operation produces a sequence of binary pattern of errors. Normally, this pattern should contain consecutive true values for Lambertian reflection pixels. The first false values indicate shadows whereas the last ones indicate highlights. As an example, for $k = 8$, the binary pattern $\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix}$ indicates that darkest one pixel is shadow and brightest two pixels are highlights. With extended non-Lambert quadruple method, up to *k-4* erroneous pixels can be detected for *k* images.

In some cases, due to improper selection of threshold value or image noise resultant binary pattern may not have consecutive true values. In this case binary pattern is corrected by checking brightest pixels with reflection mask. For example, in calculated binary pattern $\begin{bmatrix} 0 & 1 & 1 & 0 & 1 \end{bmatrix}$, either the last true value or second and third true values are miscalculated. In this case, brightest four pixels are tested with reflection mask. If any of them are highlight, binary pattern is corrected to $\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix}$, else $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

3.3.3.3 RANSAC (RA)

The robust estimation method random sample consensus (RANSAC) [39], was first used in PS by Mukaigawa et al. [19]. The application of RANSAC approach to PS problem can be classified as subset mask in unified PS.

The idea behind the RANSAC method is simply as follows [40];

1. Randomly select minimal set of data from all data set. (For PS, select 3 intensity values.)
2. Calculate residual errors of all data set.
3. Find consensus set that their residual errors are within the threshold error.
4. **If** the size of consensus set is greater than a threshold or maximum trial count achieved, estimate model with all consensus set.
5. **Else** select new subset and repeat above.

For PS, each pixel should be estimated with RANSAC individually. This individual solution results long computation times with large images. Instead of running RANSAC pixel by pixel, the faster scheme is to apply RANSAC to whole images together. In that case, random images subset is selected. The consensus set with the maximum size is saved for each pixel that is every pixels may have different consensus set at the end. The algorithm stops when all the planned random subsets are used.

It is often not necessary to select all possible subsets and calculate their consensus set. Instead the number of subsets is chosen sufficiently high to ensure with a probability, p, that at least one of the random subsets is free from outliers. Assume that ϵ is the probability of pixel to be outlier, highlight or shadow. Then at least N subsets must be used to satisfy p;

$$N = log(1 - p)/log(1 - (1 - \epsilon)^s) \tag{3.11}$$

Usually, $p = 0.99$ is used and for PS $s = 3$. For the worst case assume that half of the pixels are outliers that is $\epsilon = 0.5$. With these assumptions number of subsets is $N = 35$. If 7 images are available to PS, all the possible triplets ($C_7^3 = 35$) can be selected, else if more than 7 images are available, randomly selected 35 subsets will be adequate to achieve $p = 0.99$.

3.4 Weighted Least Square Estimation in PS

The robust estimation method, weighted least square estimation (WLSE) is used in the surface integration [9] (2004) and in shape from shading [41] (2006). The PS problem that conventionally uses least square estimation (LSE) with more than three images, was solved to WLSE firstly in this thesis.

LSE proposes pseudo inverse solution for optimal results with the following assumptions [42];

- Mean value of the errors are zero.
- Errors must be uncorrelated (independent).
- Variance of each error must be equal.

Also the distribution of the error function must be Gaussian. But for other distribution functions with same assumptions LSE still works.

In the linear PS that utilizes the Lambert model, the images contain generally three class of errors, image noise, highlights and shadows. Actually the shadows and the highlights are not errors, but since they are not considered in Lambert model they distort the results like outliers.

The listed assumptions hold for the thermal noise and the discretization errors. But for the highlights and the shadows, these assumptions does not hold. Shadows and highlights in all images are the results of the same surface topology. So they are all correlated with the same surface and indirectly with each other. For example, the images illuminated with near light sources will both have similar shadowed regions.

The zero mean assumption is also not valid for the highlights and the shadows. While the highlight errors increase intensity values and the shadows decrease. The amount of highlight error is not necessarily equal to the shadow error. Similarly, the variance of the image errors need not be equal, since there is no rule for the surface geometry, hence for the shadows and highlights. This algorithm proposes a weighted least square estimation (WLSE) to eliminate equal variance assumption [43].

Assume that all three errors can be represented with an additional error term $\mathbf{e}$ to Lambert model as in Eq. 3.12;

$$\mathbf{I} = \rho_{diff}(\mathbf{S}^T\mathbf{n}) + \mathbf{e} \tag{3.12}$$

Assume that the covariance matrix $\mathbf{W}$ of error $\mathbf{e}$ is known. Then, pseudo inverse solution can be generalized without the listed assumptions with given $\mathbf{W}$, which is WLSE.

$$\begin{gathered} \rho_{diff} = |\mathbf{S}^W\mathbf{I}| \quad \mathbf{n} = \frac{\mathbf{S}^W\mathbf{I}}{\rho_{diff}} \\ \mathbf{S}^W = (\mathbf{S}^T\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{W}^{-1} \end{gathered} \tag{3.13}$$

But since ρ_{diff} and $\mathbf{n}$ are unknown, the additional error term $\mathbf{e}$ and it's covariance matrix $\mathbf{W}$ are also unknown too. Even if $\mathbf{W}$ is known, it is a square matrix with image count rank, and the inverse of $\mathbf{W}$ must be calculated for each pixel. The computation cost of the inversion operation for each pixel is very high. Instead, if the uncorrelated errors assumption is reused, $\mathbf{W}$ square matrix is reduced to diagonal matrix composed of pixel variance values. These variance values are the weights for each pixel value, and represent the reliability of that pixel.

The reliability term for each pixel can be use in two ways. First, an iterative PS can be implemented with iteratively recalculating the normals, the errors and the variances. This method is called feasible weighted least square estimation [43]. Secondly, previously calculated masks can be weighted. Here, The aim is not to improve the resultant numerical error but generate smooth transitions at the boundaries of the masks. As a result, fake edges created by mask boundaries are removed.

3.4.1 Feasible Weighted Least Square Estimation (FWLSE) PS

Iterative variance PS is an application of feasible weighted least square estimation. Steps of calculation are;

1. Calculate the normals with LSE.

2. Render Lambert images from normals.

$$\mathbf{L} = \rho_{diff}(\mathbf{S}^T\mathbf{n}) \tag{3.14}$$

3. Define variance of a pixel as the square of difference of Lambert and real intensity (σ_i^2)

for image i. To guarantee non-singular $\mathbf{W}$, define lower limit for the variance σ^2_{min};

$$\sigma_i = \begin{cases} I_i - L_i & I_i - L_i > \sigma_{min} \\ \sigma_{min} & \text{else} \end{cases} \tag{3.15}$$

4. Form diagonal variance matrix. For k images;

$$\mathbf{W} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix} \tag{3.16}$$

5. Calculate the inverse of variance matrix. Inverse of a diagonal matrix is simply inverse of each diagonal element.

$$\mathbf{W}^{-1} = \begin{bmatrix} \sigma_1^{-2} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^{-2} \end{bmatrix} \tag{3.17}$$

6. Apply WLSE.

$$\begin{gathered} \rho_{diff} = |\mathbf{S}^W\mathbf{I}| \quad \mathbf{n} = \frac{\mathbf{S}^W\mathbf{I}}{\rho_{diff}} \\ \mathbf{S}^W = (\mathbf{S}^T\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{W}^{-1} \end{gathered} \tag{3.18}$$

7. If the difference between variance values calculated in two consecutive steps are below a threshold, stop the iteration, else go to 2.

In other words, FWLSE PS, weights the intensities with respect to their deviations from Lambert render intensities. If the real image is close to Lambert assumption, it is weighted more than others.

3.4.2 Weighted Masks

In unified PS, the masks are applied as boolean flags to use each image in the PS solution. The mask is composed of a bit pattern of total image count size. Each bit indicates that the pixel value will be utilized in PS or not. For each bit pattern $\mathbf{S}$ is recalculated with the used pixels and light sources. This masking implementation causes problems when the bit patterns of two neighbor pixels are different. Different mask patterns means different image subsets will be used. Different subsets possibly generate very different normals that will be discontinuous. These discontinuous normals caused the main problem of masked PS, where sharp false edges occurred on the boundaries of the masks.

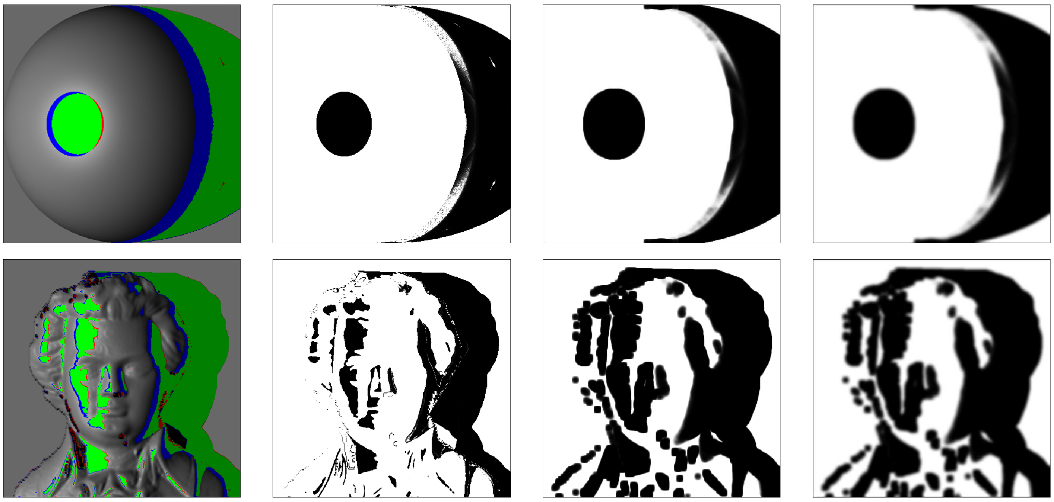
This problem can be solved if the masks are used not as 1 or 0, rather as weights reducing at the boundaries of the masks. WLSE that introduces weights in the solution, was used to solve the this problem. This approach is novel in the literature. The flow of the algorithm is similar to FWLSE PS. The main difference is that **W** is calculated once, so there is no loop in algorithm. The variations are written bold as follows;

1. Calculate the normals with LSE **with binary masks**.
2. Render Lambert images from normals.
3. Define variance of a pixel as the square of difference of Lambert and real intensity (σ_i^2) for image i **only for masked regions**. **The unmasked pixels will have minimum variance** σ_{min}^2.

$$\sigma_i = \begin{cases} |I_i - L_i| & mask_i = 1, \quad |I_i - L_i| \geqslant \sigma_{min} \\ \sigma_{min} & mask_i = 0, \quad |I_i - L_i| \geqslant \sigma_{min} \\ \sigma_{min} & |I_i - L_i| < \sigma_{min} \end{cases} \tag{3.19}$$

4. **Form variance images from variance value of each pixel. Dilate variance images.**
5. **Form inverse variance images from inverse variance value of each pixel. Filter inverse variance images with a smoothing Gaussian kernel.**
6. Form diagonal variance matrix. For k images;
7. Calculate the inverse of variance matrix. Inverse of a diagonal matrix is simply inverse of each diagonal element.
8. Apply WLSE.

The key operations in weighting masks are dilation and smoothing operations done in step 4 and 5. The dilation operation in step 4, enlarges the erroneous regions and makes sure that



(a) (b) (c) (d)

Figure 3.2: The calculated masks (a) for sphere and mozart objects are shown with green and blue colors. The weighted masks without any dilation or smoothing (step 4 and 5 are omitted) (b), dilated masks (step 5 is omitted) (c), and dilated, smoothed masks (d) are shown.

near boundary errors are also eliminated. Most of the false edges caused at the cast shadow boundaries can be removed with this dilation operation. The smoothing filter at step 5 that applied on weights (inverse variances), smooths the noisy results caused by binary masks.

The sizes of the dilation and Gaussian kernel should be equal to overlap them at the smoothed transition regions of the mask boundaries. The size was selected by trial and error with the synthetic images. The minimum standard deviation that clears the false edges is desired. Hence, the dilation operation causes the correct pixels near the false ones to be treaded as erroneous ones.

In Fig. 3.2, a test of weighted PS is plotted for sphere and mozart objects. The Th mask is used as shown in Fig. 3.2(a). The weighted masks without any operations (step 4 and 5 are omitted) is shown at Fig. 3.2(b). Other plots represents the effect of Gaussian filters step by step.

The detailed results of the methods explained in this Chapter will be presented on the synthetic images in Chapter 5, on the real images in Chapter 6 and on the cartridge cases in Chapter 7.

CHAPTER 4

DOUBLE ZENITH LIGHT SOURCES

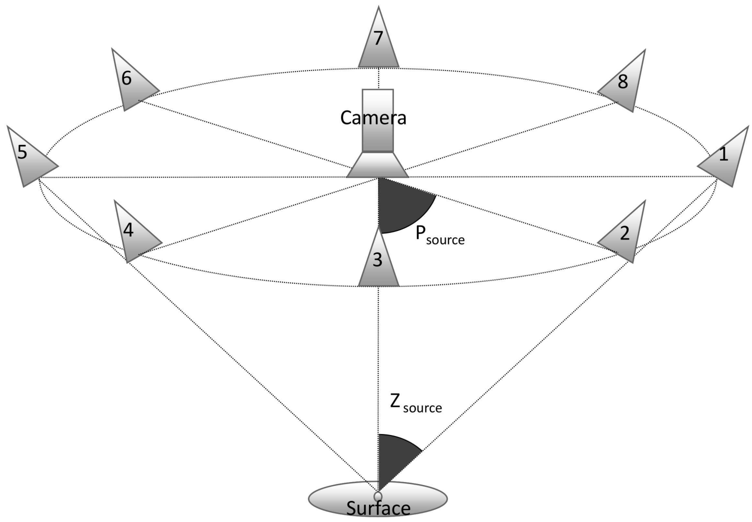
4.1 Introduction

The performance of the light source configurations depends on the errors in the images. In this study, three sources of error, i.e. the image noise, the highlights and the shadows were investigated. The image noise can be modeled with a zero mean, independent Gaussian added on images. Independent, zero mean Gaussian noise is a well known type of error in linear system theory and its performance analysis can be made theoretically. However, highlights and shadows are not zero mean errors. Also they are not independent among all images, since they all occur on the same surface topology. Since theoretical approach will be very complex for performance analysis of *PS* with highlights and shadows, simulations were made to compare different light source configurations. These simulations are executed on a control test configuration that defines light source placements, test surfaces, surface reflection parameters, image rendering properties and error definitions.

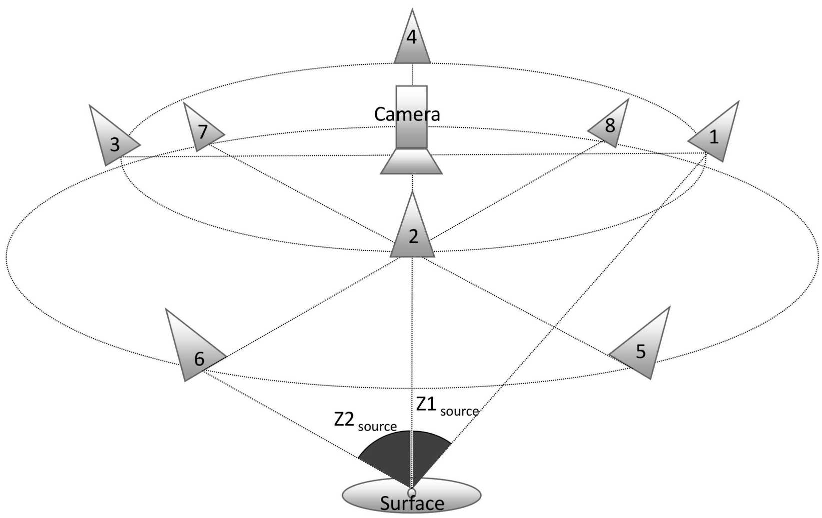
4.2 Light Source Configurations

In single zenith light sources tests, light sources were placed on a circular ring around the camera having the same zenith angle and equal polar distances as in Fig. 4.1a. This type of illumination was found to be optimal in previous works [23] [44] [25] [3]. A second novel illumination configuration is also used in the tests, that the lights are placed on two circles as in Fig. 4.1b.

Three sombreros with 1750 (sombrero1), 3500 (sombrero2) and 7000 (sombrero3) peak val-



(a)



(b)

Figure 4.1: 8 light sources placed around camera with (a) single zenith $Z_{source} = 45^o$ and (b) double zenith $Z1_{source} = 30^o$, $Z2_{source} = 60^o$ illumination configurations.

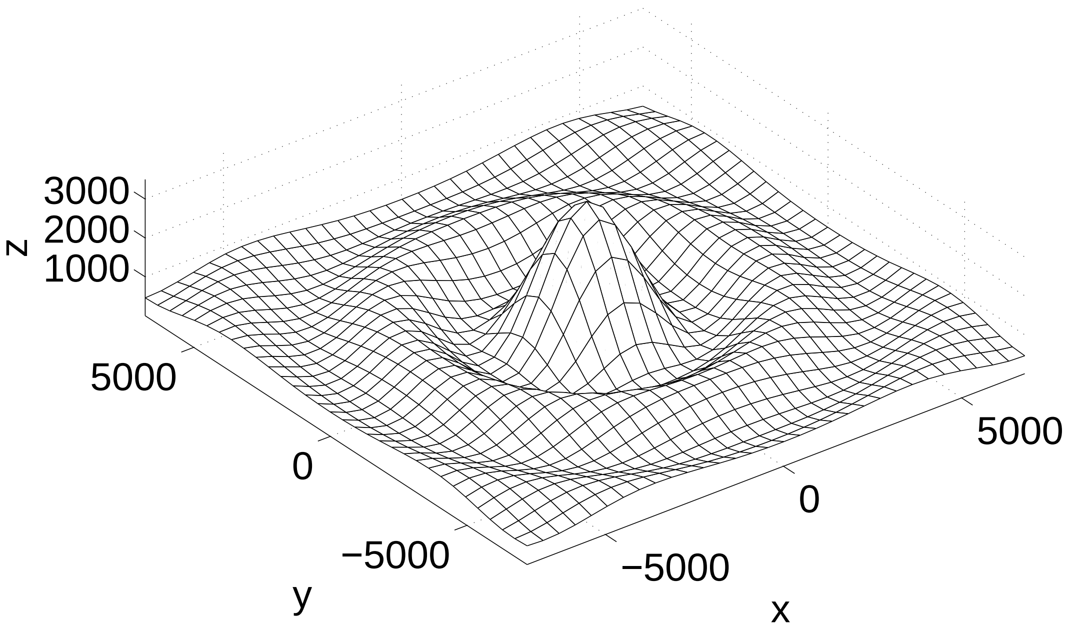


Figure 4.2: Sombrero2 ranging $[-7200, 7200]$ with the peak 3500.

ues were used to simulate both smooth and rough surfaces. The one with the middle peak value is shown in Fig. 4.2.

In the tests, the synthetic images with highlights were rendered with Torrance-Sparrow reflection model given at Eq. 2.14 [16]. Uniform specular albedo ($\rho_s = 1.0$) and uniform diffuse albedo ($\rho_d = 1.0$) were used. The uniformity condition, however, was not a precondition for the solutions. Shininess parameter m is chosen to be 2.0, which was found to be the worst case in our previous study [45].

The CCD noise (also known as the thermal noise) of a conventional camera was represented with a zero mean Gaussian noise added to the intensity values. The standard deviation (noise power) of Gaussian noise was chosen to be 5% of the image dynamic ranges.

4.2.1 Single Zenith Light Sources Configuration ($1Z$)

The first test was conducted to observe the effect of image count n_i. A typical example presented in Fig. 4.3 plotting resultant normal error (NE) versus image count (n_i) for both diffuse (Di) images and images with highlight and shadow (Hi-Sh). In both cases, decrease

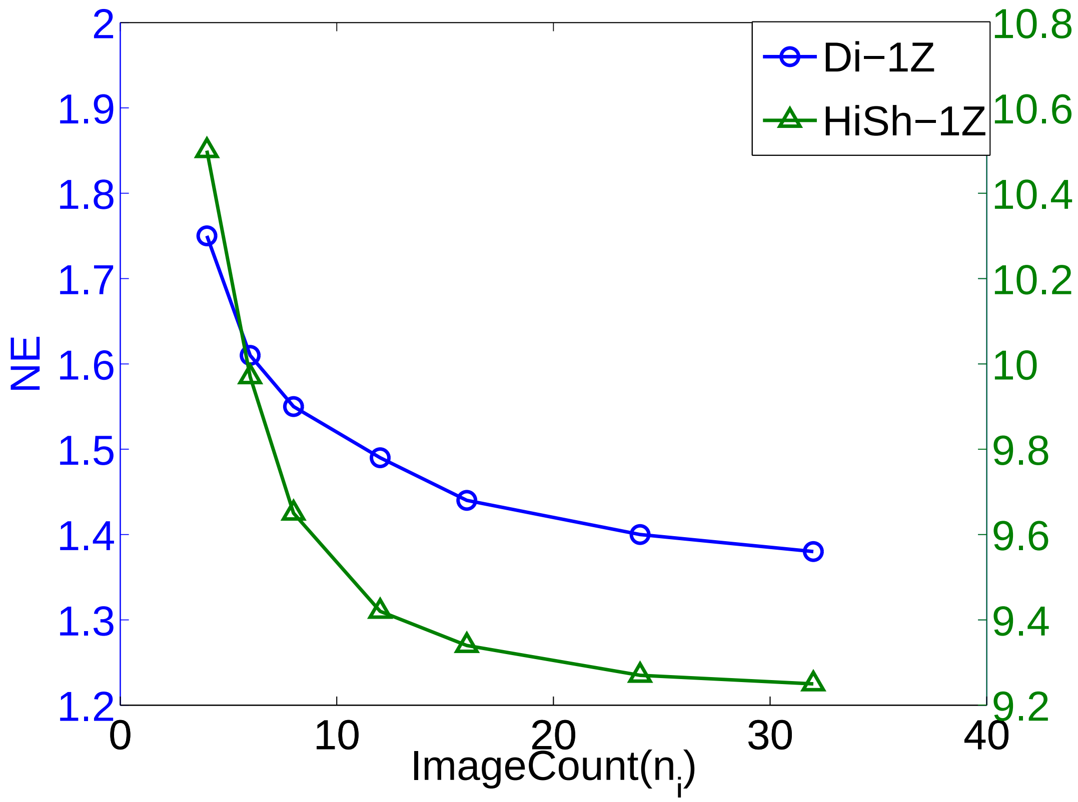


Figure 4.3: Normal Error (NE) vs image count (n_i) for diffuse (Di) and highlight shadow(Hi-Sh) images. In both cases error drop saturates with increasing image count (n_i).

in NE was saturated with increasing n_i. Here increasing n_i from 16 to 32 decreased NE from 1.44^o to 1.38^o, the difference being only 4% for diffuse (Di) images. For highlight shadow images (Hi-Sh), same change in n_i, decreased NE from 9.34^o to 9.25^o, having 1% difference.

Secondly, the simulations were executed on the optimal light source configurations from the previous works. The synthetic diffuse images without highlights and shadows were used. Gaussian noise with 5% of dynamic range of images were added on these images. The change of NE with respect to source zenith angle (Z_{source}) is plotted in Fig. 4.4 for 4, 6, 8, 12 and 16 images. Similar to the previous works, with all image counts and all sombreros, $Z_{source} = 45^o$ is found to be optimum.

Later, the optimal Z_{source} is tested with highlights and shadows. The amount of inserted highlight and shadow errors depends on the surface topology and Z_{source} as seen in Fig. 4.5. First row of images were generated from sombrero1 (smooth surface), the second from sombrero2 and the third row were from sombrero3 (rough surface). Images were rendered with

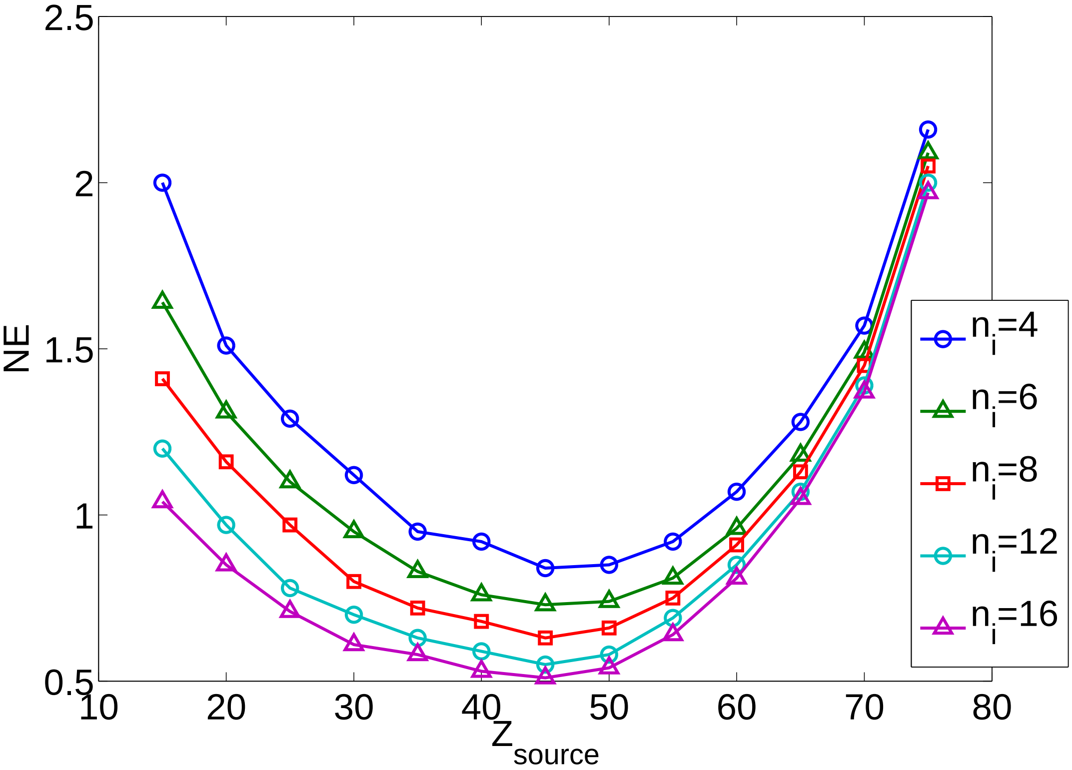
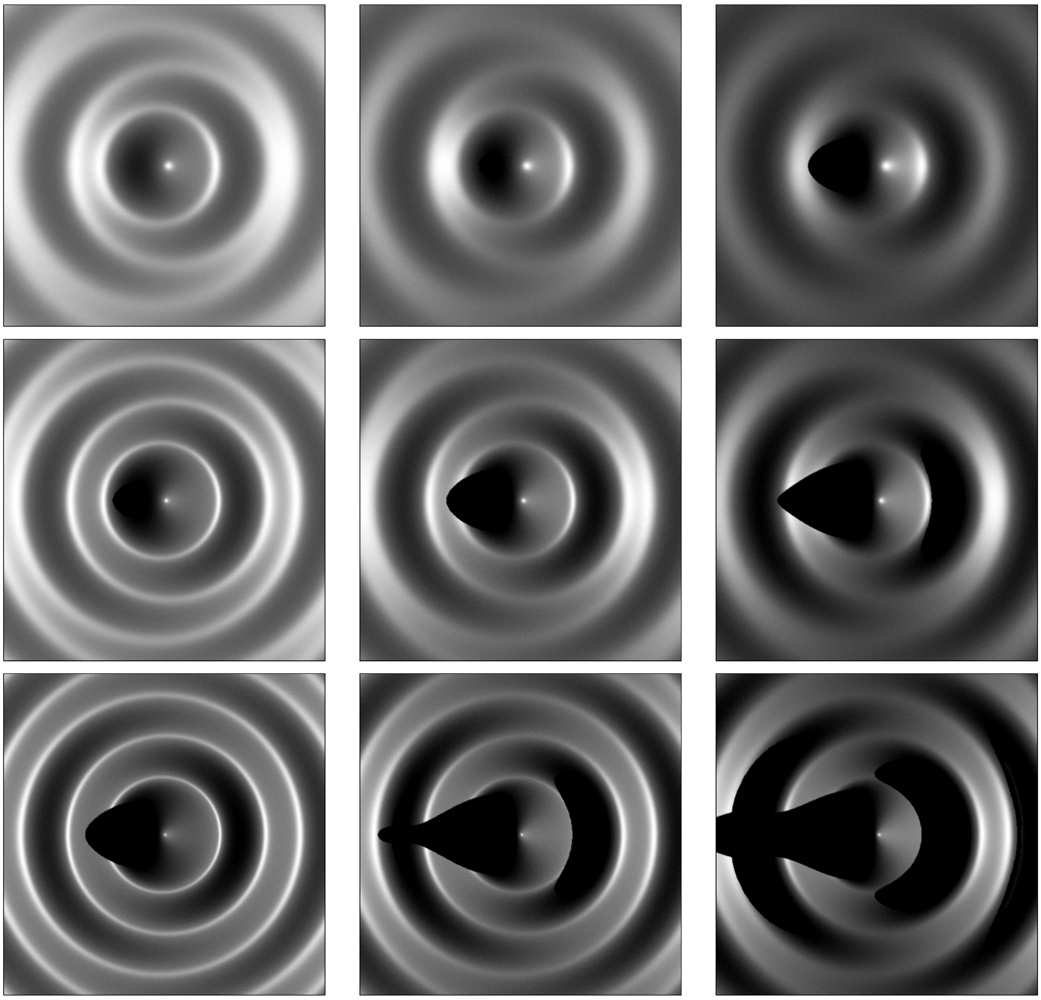


Figure 4.4: Normal Error (NE) vs light sources zenith angle (Z_{source}) for 4, 6, 8, 12 and 16 images. Images were rendered without highlight and shadow. 5% gaussian noise is added to images.

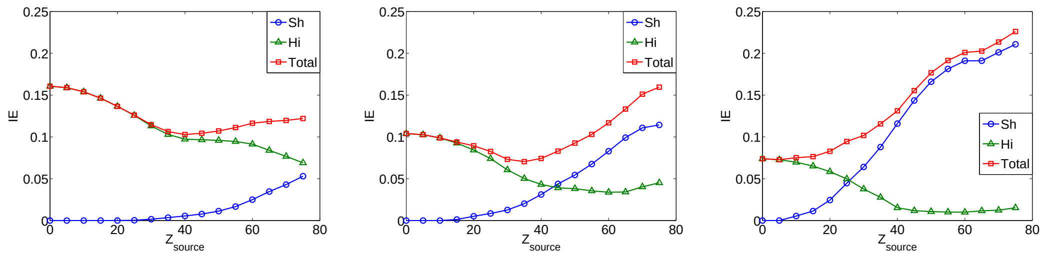


(a) (b) (c)

Figure 4.5: Sombreros, (from top to bottom) sombrero1, sombrero2 and sombrero3 with (a) $Z_{source} = 30^o$, (b) $Z_{source} = 45^o$ and (c) $Z_{source} = 60^o$.

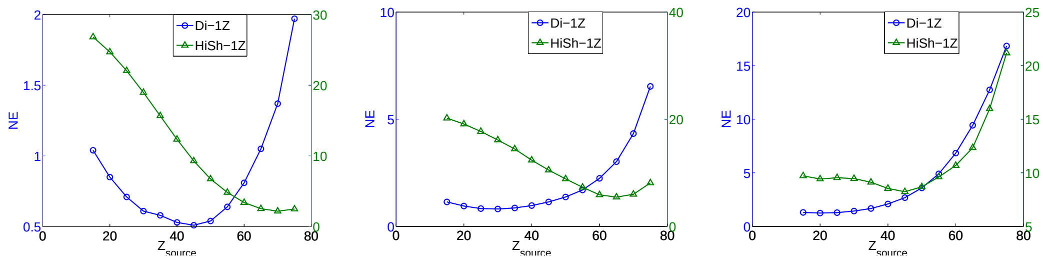
$Z_{source} = 30^o$ for first column, $Z_{source} = 45^o$ for second and $Z_{source} = 60^o$ for last column of images. While for sombrero1, the highlights were the major problem with $Z_{source} = 30^o$ (Fig. 4.5 upper left), for sombrero3, shadows occurred widely with $Z_{source} = 60^o$ (Fig. 4.5 lower right).

Fig. 4.6 shows the change of generated intensity error (IE) from highlights, shadows for each sombrero with changing Z_{source}. The total IE of the images changes with both surface topology and Z_{source}. The NE versus Z_{source} is plotted in Fig. 4.7 for each sombrero. At each plot the results of diffuse (Di) and highlight shadow (Hi-Sh) images are shown. At each case optimal value of Z_{source} changed differently. For smooth sombrero1, (see Fig. 4.7a) the



(a) (b) (c)

Figure 4.6: Shadow, highlight and total errors for (a) sombrero1, (b) sombrero2 and (c) sombrero3.



(a) (b) (c)

Figure 4.7: Normal error vs light sources zenith angle with / without shadows and highlights for (a) sombrero1, (b) sombrero2 and (c) sombrero3.

optimal Z_{source} was changed from 45^o to 70^o due to strong highlight errors with low Z_{source}. On the other hand, for rough sombrero3, (see Fig. 4.7c) the optimal Z_{source} was still increased but settled at 45^o because of large shadow errors at high Z_{source}.

As a result, there is no single optimal value for every surface if highlights and shadows occur at the input images. With some estimations of the general surface topology, like smooth and rough [22], some weak rule of thumb may be used.

4.2.2 Double Zenith Light Sources Configuration (2Z)

The previous section focused on a single zenith light sources configuration around the camera. The main optimization parameter was the zenith angle of the light sources. This type of illumination configuration was proved as the optimal configuration while considering only image noise. On the other hand, when highlights and shadows presents in the images, optimal Z_{source} differs with surface topology. In this section, light source configuration with two zenith angles (see Fig. 4.1b) will be investigated.

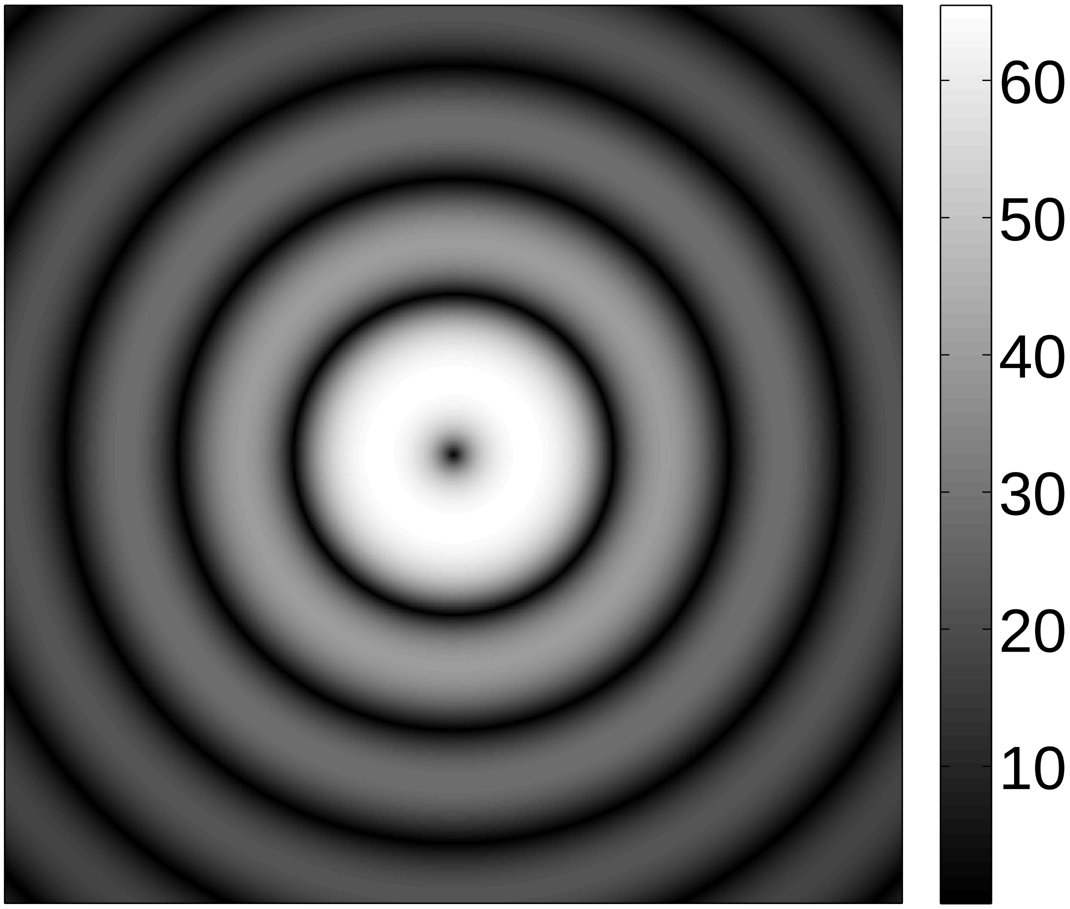


Figure 4.8: Surface zenith angles of sombrero2

When using double zenith light sources configuration, two source zenith angles ($Z1_{source}$ and $Z2_{source}$) should be decided. For this purpose, the relation between the surface topology and NE should be considered in depth. Since surface topology is unknown, and a general solution good for all type of surfaces is desired, surface topology is summarized with surface zenith angles ($Z_{surface}$) that is the angle between surface normal at a point and camera direction. An example to $Z_{surface}$ for sombrero2 is presented in Fig. 4.8.

PS is executed with single zenith light sources to analyze relation between surface zenith ($Z_{surface}$) and NE. Different Z_{source} values and surfaces are used. As a result, NE versus $Z_{surface}$, i.e. distribution of mean NE over $Z_{surface}$, is plotted for each solution. In Fig. 4.9 NE versus $Z_{surface}$ of sombrero2 with $Z_{source} = 30^o$, 45^o and 60^o is shown. For $Z_{source} = 30^o$, NE is concentrated at low $Z_{surface}$, that are caused by highlights. For $Z_{source} = 60^o$, NE gradually

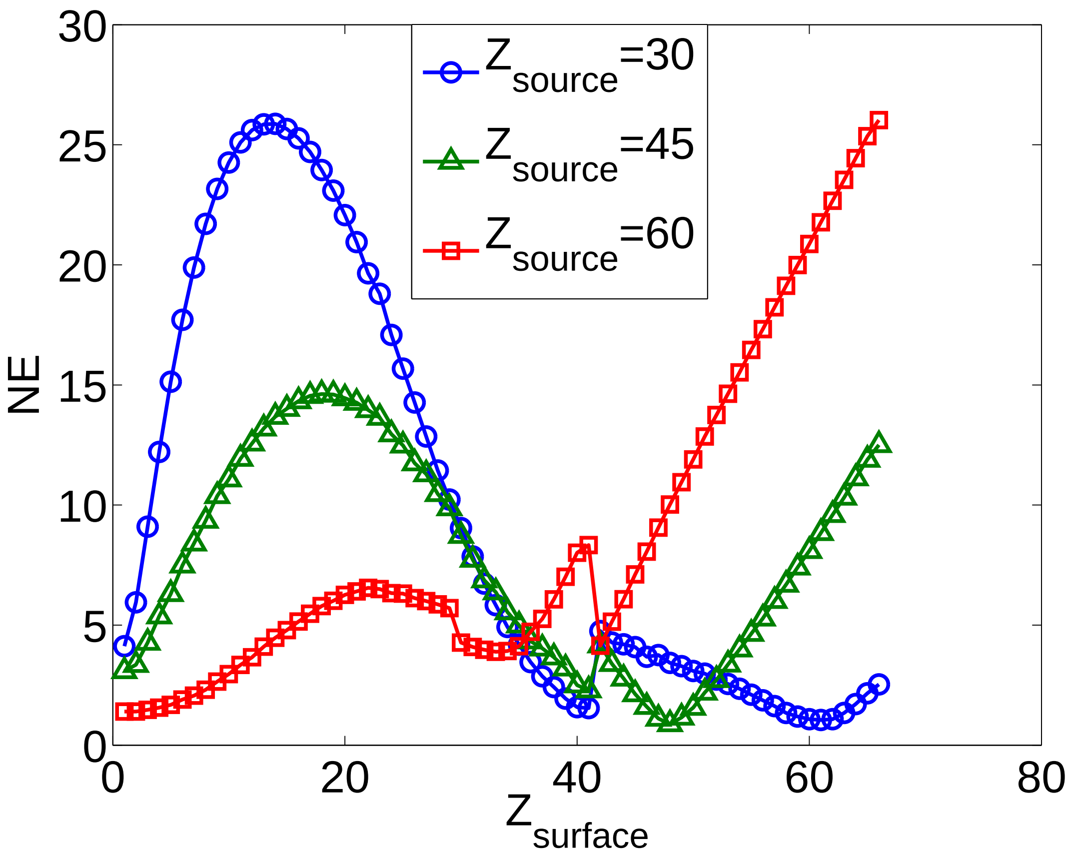


Figure 4.9: Normal Error VS $Z_{surface}$ for sombrero2 with 30^o, 45^o and 60^o light source zenith angles. (Maximum $Z_{surface}$ is 66^o for sombrero2)

increases with increasing $Z_{surface}$. The discontinuities around $Z_{source} = 40^o$ are caused by the cast shadows that may occur at anywhere independent of surface zenith angle. The peak at the center of sombrero causes these cast shadows at the same $Z_{source} = 40^o$ because of the circular symmetry of the shape.

This NE distribution is very similar to the intensity error (IE) distribution caused by highlights and shadows in Fig. 4.10. The plotted intensity error can be formulated for TS as in Eq. 4.1. The first line is the highlight term in TS model and the second line is the self shadow error, which is not modeled in linear PS. The intensity errors caused by the cast shadows are omitted in this equation.

$$IE_{TS} = \begin{cases} \rho_s e^{-m^2[\arccos(\mathbf{n}_h^T \mathbf{n})]^2} & \mathbf{s}^T\mathbf{n} \geq 0 \\ -\rho_d(\mathbf{s}^T\mathbf{n}) & \mathbf{s}^T\mathbf{n} < 0 \end{cases} \tag{4.1}$$

The IE around $Z_{surface} = Z_{source}/2$ caused by highlights where $Z_{source}/2$ is the reflection direction where surface directly reflects light to camera. The standard deviation of IE (STD_h)

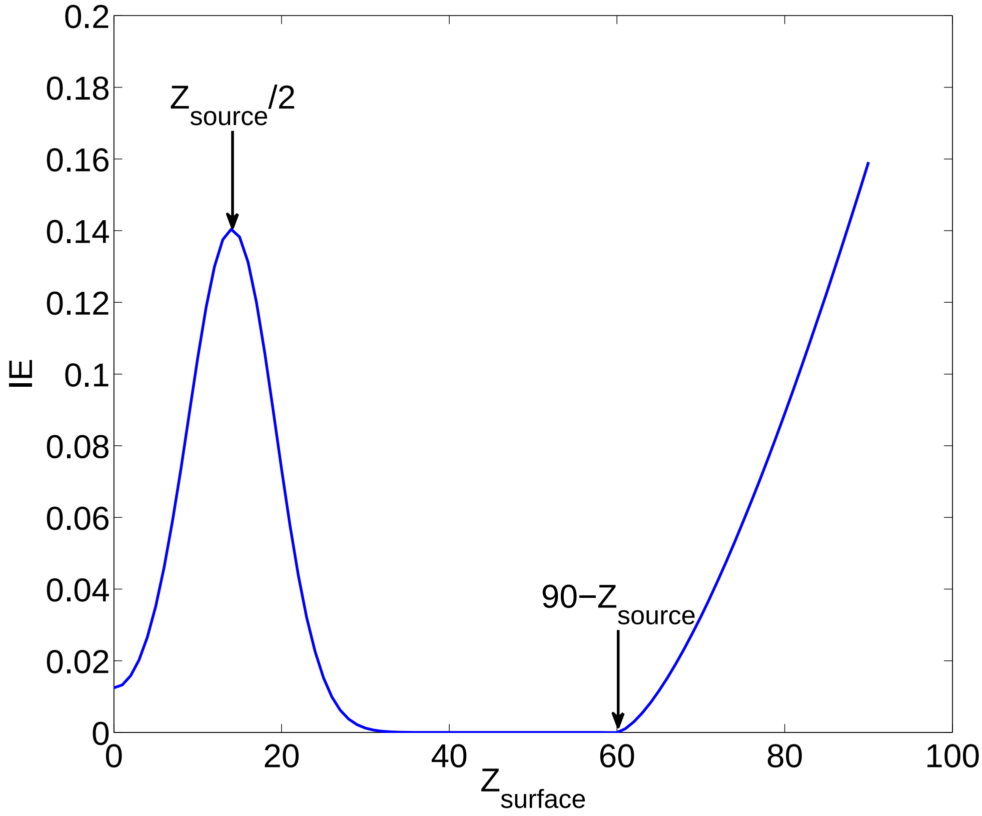


Figure 4.10: IE distribution versus surface zenith angle ($Z_{surface}$)

around $Z_{source}/2$ depends on the shininess (range that highlight occurs) of the surface. After $Z_{surface} = 90 - Z_{source}$ shadows start to occur causing increasing IE. More pixels will be shaded with increasing $Z_{surface}$.

With this IE (NE) characteristics in the hand, a rule of thumb can be suggested to select two light source zenith angles. A general purpose PS should have a constant error all over the source that a flat NE versus $Z_{surface}$ is desired. So two Z_{source} should be selected in such a way that two high NE peaks, i.e. caused by shadows and highlights, should not coincide. Also the average of two zenith angles should be the optimal light source zenith for single circle sources $Z_{sourceOp}$. The Eq. 4.2 presents the general rule of thumbs for the general purpose PS.

$$Z1_{source}/2 + STD_h \approx 90 - Z2_{source} \tag{4.2}$$

$$(Z1_{source} + Z2_{source})/2 \approx Z_{sourceOp}$$

If $STD_h = 15^o$ and $Z_{sourceOp} = 45^o$ then, $Z1_{source} = 30^o$ and $Z2_{source} = 60^o$.

4.3 Weighted Multi Zenith PS

With double zenith light sources ($2Z$), NE was reduced if highlights and shadows were present in the images. But the erroneous images were still used in PS calculation. In this section, the effect of erroneous images were tried to be reduced. Since a weighted summation of normal vectors is used for this purpose, this method is called weighted PS (WPS).

When the single zenith sources ($1Z$) were used in PS, the normal error is concentrated at very low or very high surface zenith regions. This is the main clue for the weighted PS. In the double zenith sources configuration, light sources can be grouped with respect to their Z_{source}. The two subsets of light sources were solved exclusively resulting two normal vectors for each pixel. Later, the two resultant normal vectors were fused with weights that are calculated from the estimated errors of each normal vector. These weights can be rounded to 0 and 1 for binary weighting that is selecting normal vector with small error.

The following flow presents the mathematical details of weighted PS;

1. Calculate normal vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ from illumination matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. $\mathbf{S}_1$ is composed of $Z1_{source} = 30^o$ light sources and $\mathbf{S}_2$ is composed of $Z2_{source} = 60^o$ light sources.

$$\rho_{1diff} = |\mathbf{S}_1^{\dagger}\mathbf{I}_1| \quad \mathbf{n}_1 = \frac{\mathbf{S}_1^{\dagger}\mathbf{I}_1}{\rho_{1diff}} \tag{4.3}$$

$$\rho_{2diff} = |\mathbf{S}_2^{\dagger}\mathbf{I}_2| \quad \mathbf{n}_2 = \frac{\mathbf{S}_2^{\dagger}\mathbf{I}_2}{\rho_{2diff}} \tag{4.4}$$

2. With normals from (4.3) and (4.4) equation, $\mathbf{n}_1$ and $\mathbf{n}_2$ calculate Lambert images.

$$\mathbf{I}_1 = \rho_{1diff}(\mathbf{S}_1^T\mathbf{n}_1) \tag{4.5}$$

$$\mathbf{I}_2 = \rho_{2diff}(\mathbf{S}_2^T\mathbf{n}_2) \tag{4.6}$$

3. Define residual errors (e_1 and e_2) as the norm of difference of Lambert and real images.

$$e_1 = |\mathbf{I}_1 - \mathbf{L}_1| \tag{4.7}$$

$$e_2 = |\mathbf{I}_2 - \mathbf{L}_2| \tag{4.8}$$

4. Merge two normal vectors by weights (w_1 and w_2), calculated from residual errors.

$$\rho_{diff}\mathbf{n} = w_1 * (\rho_{1diff}\mathbf{n}_1) + w_2 * (\rho_{2diff}\mathbf{n}_2) \tag{4.9}$$

Weight definitions have very important rule in the resultant performance. The main rule in weights definition is that their summation must be unity as in Eq. 4.10. So each weight is scaled with normalization factor, k as defined in Eq. 4.11.

$$w_1 + w_2 = 1 \tag{4.10}$$

$$k = \frac{1}{w_1 + w_2} \tag{4.11}$$

Weight definitions can be converted to binary as in Eq. 4.12 to simply select the normal with less error.

$$(w_1, w_2) = \begin{cases} (0, 1) & w_1 < w_2 \\ (1, 0) & else \end{cases} \tag{4.12}$$

In this work, we tested four different weight definitions.

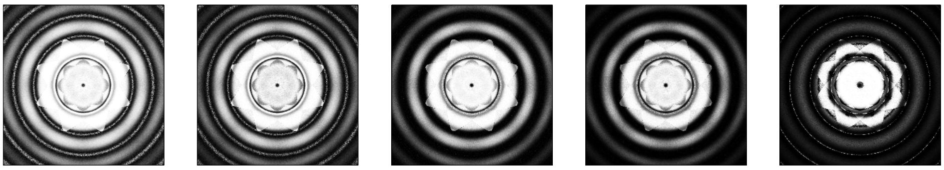
- *Weight Definition 1 (e)* : Weights are inversely proportional to uncertainty of normals that is square of the residual error (e).

$$w_1 = ke_2^2 \tag{4.13}$$
$$w_2 = ke_1^2$$

- *Weight Definition 2 (eZ_{source})* : For smooth surfaces, the sensitivity of normals are inversely proportional to sine of Z_{source} [23]. Using this heuristic, weights are updated with constant multiplier of $\sin(Z_{source})$. Normals generated with larger Z_{source} are stressed more in the resultant normals.

$$w_1 = k\sin(Z1_{source})e_2^2 \tag{4.14}$$
$$w_2 = k\sin(Z2_{source})e_1^2$$

- *Weight Definition 3 ($eZ_{surface}$)* : Another heuristic can be generated from the highlights and shadows of images. At the regions where the $Z_{surface}$ is limited, (smooth regions)



(a) (b) (c) (d) (e)

Figure 4.11: (a) e, (b) eZ_{source}, (c) $eZ_{surface}$, (d) $eZ_{source}Z_{surface}$, (e) optimal weights for minimum NE of 30 degree solution on TS rendered sphere.

the major risk is highlights. To avoid highlights, large Z_{source} should be employed. Visa versa, for rough regions with large $Z_{surface}$, most probably, shadows will distort the results, little Z_{source} should be weighted more. This is implemented with a sigmoid function on the $Z_{surface}$.

$$\begin{aligned} w_1 &= k\mathbf{sig}(Z1_{surface})e_2^2 \\ w_2 &= k\mathbf{sig}(Z2_{surface})e_1^2 \\ \mathbf{sig}(x) &= 1/(1+e^{-10(x-0.5)}) \end{aligned} \tag{4.15}$$

- *Weight Definition 4 ($eZ_{source}Z_{surface}$)* : The last weighting definition is simply the combination of W2 and W3, using both Z_{source} and $Z_{surface}$ multipliers.

$$\begin{aligned} w_1 &= k\sin(Z1_{source})\mathbf{sig}(Z1_{surface})e_2^2 \\ w_2 &= k\sin(Z1_{source})\mathbf{sig}(Z2_{surface})e_1^2 \end{aligned} \tag{4.16}$$

Fig. 4.11 presents the weight image of $Z1_{source}$ for Hi-Sh rendered sombrero2 for each of the weight definition given above. The weight image of $Z2_{source}$ is not presented since it is $w_2 = 1 - w_1$. Fig. 4.11e displays the weights calculated not from the estimated error but exact NE, that is the theoretical limit of WPS. The closest weight image to optimal weights is found to be fourth weight definition ($eZ_{source}Z_{surface}$).

Some sample normal vectors are plotted on the surface of sombrero2 in Fig. 4.12. $z1$, $z2$ (red) are normal vectors calculated with $Z1_{source}$ and $Z2_{source}$ light sources. w (blue) is the weighted normal vector. r (green) is the reference normal vector. Since w is linear combination of $z1$ and $z2$, it is always going to be in the arc between $z1$ and $z2$. In general, r does neither have to be between $z1$ and $z2$, nor in the same plane.

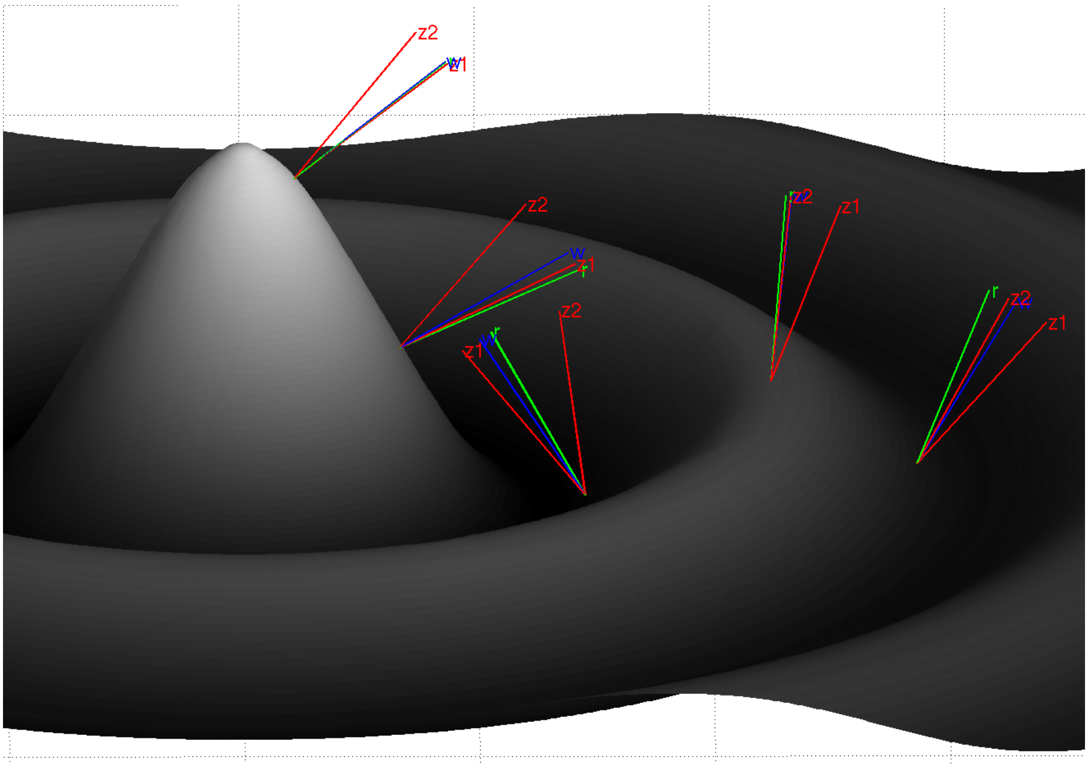


Figure 4.12: Sample normal vectors plotted on surface. $z1$, $z2$ (red) are normal vectors calculated with $Z1_{source}$ and $Z2_{source}$ light sources. r (green) is the reference normal vector. w (blue) is the weighted normal vector.

The detailed results of the methods explained in this Chapter will be presented on the synthetic images in Chapter 5, on the real images in Chapter 6 and on the cartridge cases in Chapter 7.

CHAPTER 5

SYNTHETIC TESTS

5.1 Introduction

The synthetic image results are presented in this chapter. Firstly, the synthetic image rendering methods and used configurations are given. All thresholds of the masks are fine tunned in the next section. Since the original ground truths are available with synthetic images, the results of masking methods are evaluated with detailed numerical analysis under various conditions. Also visual results, representing improvements of weighted mask PS are displayed in the next section. Lastly, both numerical and visual results of the double zenith and the weighted normal PS are given.

5.2 Synthetic Image Generation (Rendering)

Synthetic image generation process (rendering) is completely renewed to create realistic images for the tests. Instead of using output images of commercial 3D rendering tools, rendering process is fully implemented in this thesis work. The main aim of renderer implementation is to use it in PS solution as well. Secondary aim is to have complete control on generated test images.

Following is the list of new features of the rendering process.

- External camera parameters like camera coordinate frame, (position, look at direction, up vector) can be defined.
- Internal camera parameters like pixel size, vertical-horizontal pixel count, focal length

can be defined.

- Cast and self shadows can be generated.
- Finite distance point light source can be used.
- Inverse square lighting power can be used.
- Perspective projection can be used.

Some of these features are implemented for future use in alternative 3D modeling process. The most important feature for PS is shadow generation. Finite distance lighting, inverse square lighting, and perspective projection are all implemented but have not been utilized yet.

The main steps of the rendering are as follows;

1. Generate the intensity value
   (a) Calculate the intensity value for each surface coordinate using selected reflectance model.
   (b) If perspective projection is desired, use finite distance point light source.
   (c) If inverse square lighting power is desired, multiply the original source power with inverse square of the distance between the given pixel and the light source
2. Generate shadows
   (a) Generate the shadow map.
   (b) Modify intensity values using the shadow map.
3. Project objects in 3D to 2D images
   (a) Translate the object so that the look-at position is origin.
   (b) Rotate the object so that the camera direction is at z axis and camera-up vector is at y axis.
   (c) If perspective projection is desired, apply perspective deformation to the object.
   (d) Use Z-Buffer for back face culling (i.e. occlusion; z-buffer will be explained later in the text).

Above steps are explained in detail in the following part:

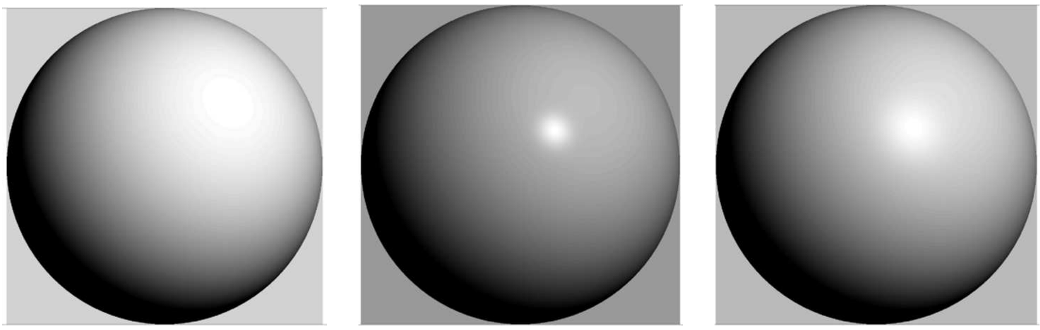


Figure 5.1: (a) Lambert, (b) Phong and (c) Torrence-Sparrow reflection models on sphere surface.

5.2.1 Generate the Intensity Value

Three images generated using three different types of reflection models for a constant albedo semi sphere are shown in Fig. 5.1 . All images are projected orthogonally, illuminated with light sources at infinity and there are no shadows. Phong model configuration has sharp specular region. Torrance Sparrow models are more realistic with softer and wider specular regions.

In Fig. 5.2, different Lambert illumination calculation methods are illustrated on a flat surface at x-y plane. Fig. 5.2.1 shows the simplest case of illumination, point light source at infinity. Since the light source is at infinity, the source direction is constant all over the surface creating a uniform intensity. Fig. 5.2.2 presents finite distance light source effect, i.e. the light source direction is calculated for each surface coordinate. For a planar surface, source direction deviates from surface normal as the distance between surface coordinate and source position increases, resulting in decreased intensity. Light source power can be modified with respect to inverse square law, to generate a more realistic point light source illumination as in Fig. 5.2.3.

5.2.2 Generate Shadows

Cast and self shadows are created with an algorithm called shadow mapping [46]. The main idea of shadow mapping is presented in Fig. 5.3. For example a sombrero object will be rendered with shadows in the default configuration. Fig. 5.3(a) shows the image without

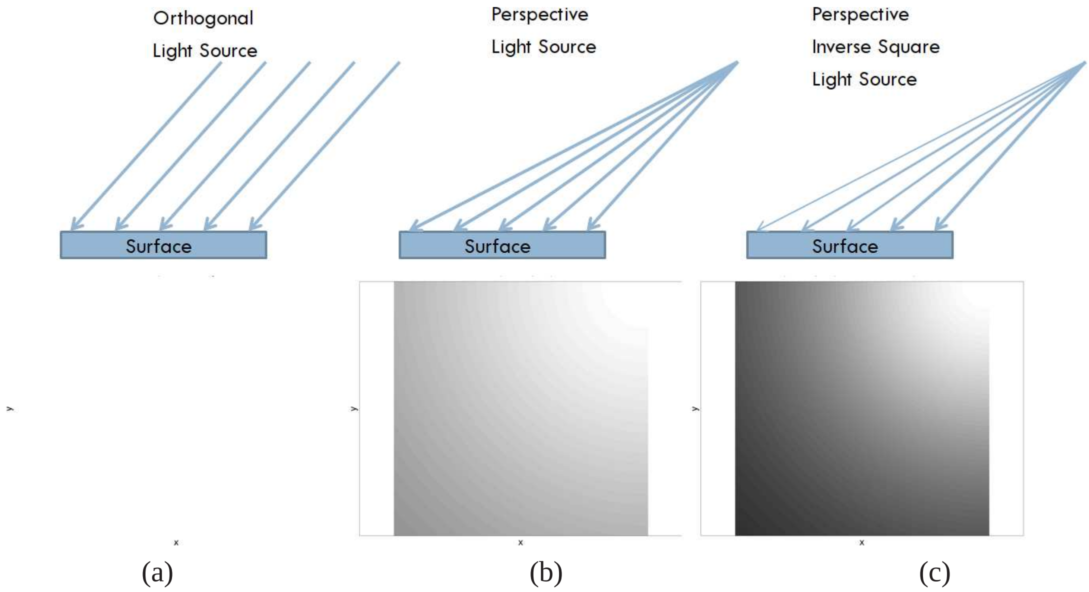


Figure 5.2: (a) Orthogonal, (b) Perspective and (c) Perspective and inverse square illumination calculations.

shadows. To generate shadows, a temporary image is rendered with the camera at the light source as seen in Fig. 5.3(b). Since the camera and light source coincide, no shadows can occur in the temporary image. Inverse of this is also true, i.e. invisible coordinates are all shadowed. The matrix that contains the visibility information for each coordinate is called the shadow map. To generate shadowed image in Fig. 5.3(c), image in Fig. 5.3a is filtered with shadow map i.e. shadowed areas are replaced with shadow intensity, 0.

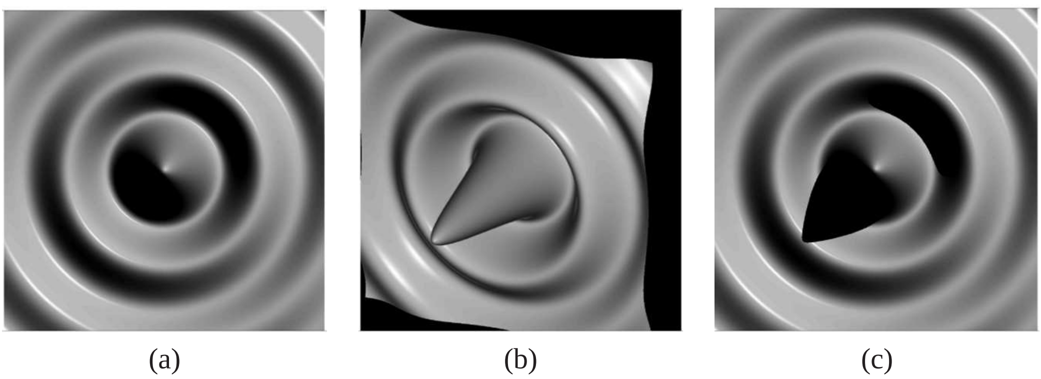


Figure 5.3: (a) No shadow sombrero, (b) temporary image viewed from light source and (c) sombrero with shadows.

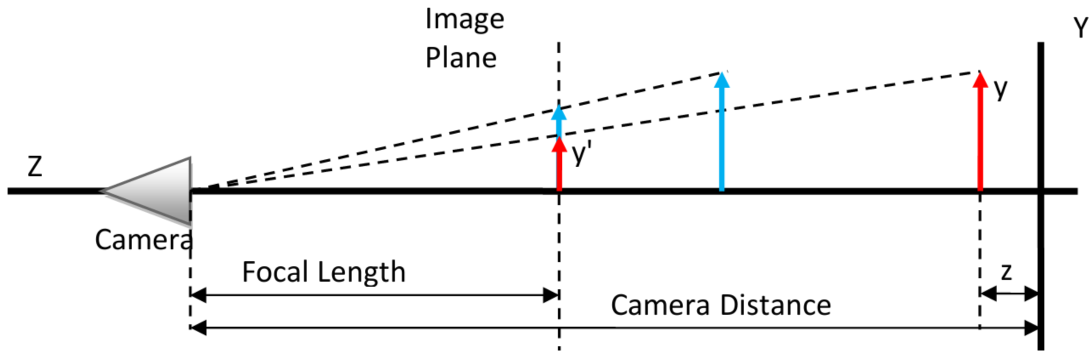


Figure 5.4: Perspective projection deformation.

5.2.2.1 Z-Buffer

Visibility of a coordinate can be tested with a method called Z-Buffer [38]. This method is also employed for back face culling, i.e. removing occluded coordinates from images. The idea behind the Z-Buffer is to use an extra 2 dimensional buffer to keep the closest depth values for each pixel. When a new coordinate is to be rendered, its depth value is checked with the one in the Z-Buffer. If the new depth value is less than the old one, Z-Buffer is updated with the new one. After updating Z-Buffer, in shadow mapping, old coordinate is marked as shadowed. In back face culling, new coordinates' intensity values overwrite old ones.

5.2.2.2 Projections

A pin hole camera model creates an image of the real world to the 2D image plane with the perspective projection. The ratio of real object size and its image depends on the distance of the object to the camera center. If the variation of distances of the objects to camera are limited the perspective projection can be further simplified to orthogonal projection that image size to object size ratio is constant.

Before perspective projection, object is translated and rotated to the camera coordinate frame that the camera is at $+Z$ axis, looking at origin and up vector is parallel to $+Y$ axis as in Fig. 5.4.

The effects of the orthogonal and perspective projections are presented in Fig. 5.5 for both shape and shadows. The Fig. 5.5(a) is the result of the orthogonal projection and Fig. 5.5(b) is the render image with the perspective projection. While the camera is placed at a finite distance in perspective image, for the orthogonal image, the camera is at the infinity. In Fig. 5.5(c) represents the orthogonal projection of the light source that the light source is at infinity. The Fig. 5.5(d) is the render image of the same object with the perspective projection.

In the tests the perspective projection of the shape is discarded, since the variation of the surface Z values are very small, compared the focal length. However this assumption is not valid for light sources, whose distances are comparable with the surface Z variation.

5.3 Test Configuration

The comparison among the defined masks were conducted under a controlled test configuration. The test configuration defines light source placements, test surfaces, surface reflection parameters, image rendering properties and error definitions.

Light sources were placed on a single circular ring around the camera with equal zenith angle and equal polar distances as in Fig. 4.1(a). This type of illumination was found to be optimal for noisy images [44].

Different surface topologies may create different problems that has to be solved using masks. For a fair comparison among the masks, five sample surfaces were selected for tests as shown in Fig. 5.6.

The amount of highlight and shadow errors depend jointly on the shape of the surface and light source placement. If the surface slope is low, highlight problems are dominant, since they occur on wide regions. Oppositely, shadows occur more on high slope surfaces. On the other hand, the increasing light source zenith angle increases shadows and decreases highlights. The variations of resultant shadows (marked with green) and highlights (marked with blue) were presented on the sombrero and sphere objects in Fig. 5.7. Large amount of highlight was occurred on the low slope sombrero with 30° light source zenith angle as seen at upper left image. Large shadowed area was created with high slope sphere object and 60° light source zenith angle.

The amount of highlight and shadow errors are presented on sombrero and sphere with

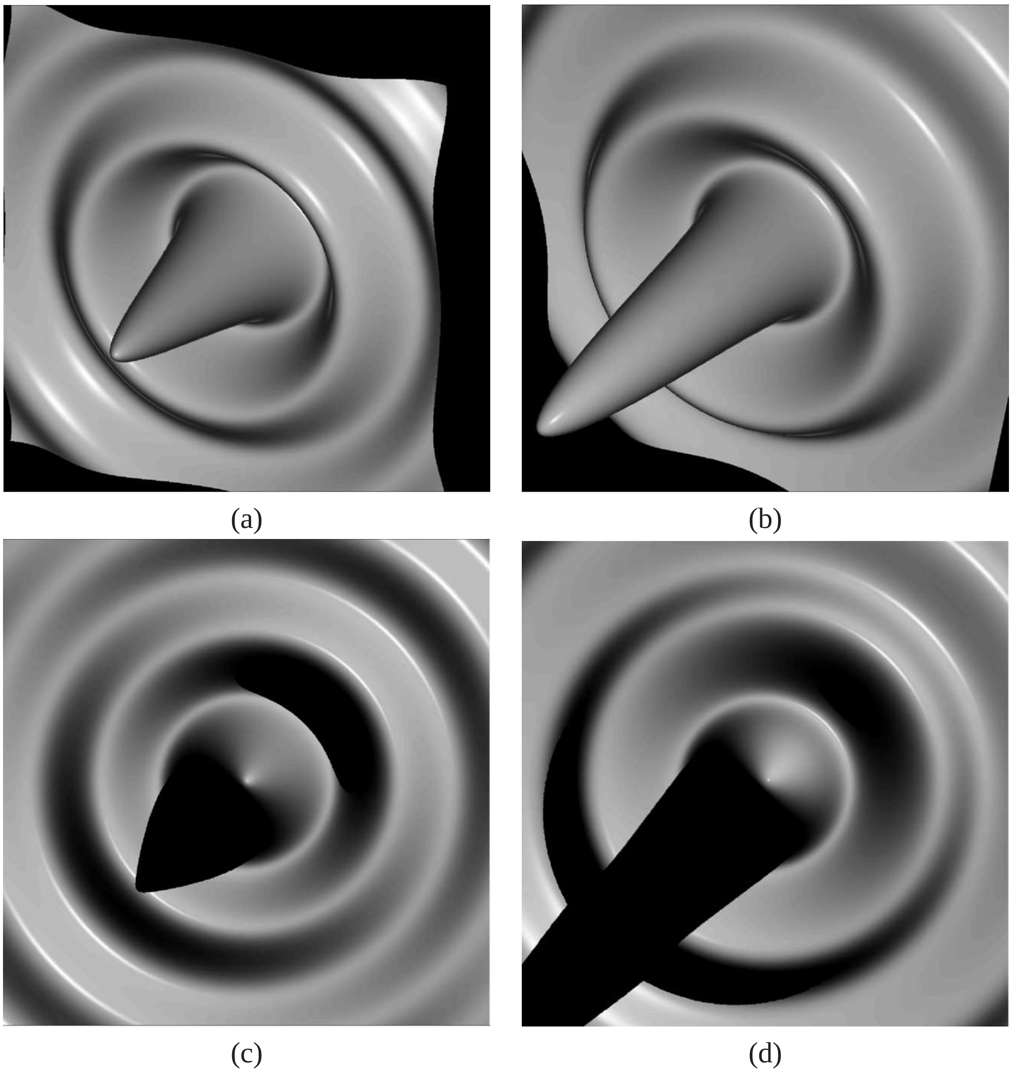


Figure 5.5: (a) Orthogonal, and (b) perspective projection of the sombrero viewed from the light source. The top view of the same sombrero with (c) orthogonal and (d) perspective projections.

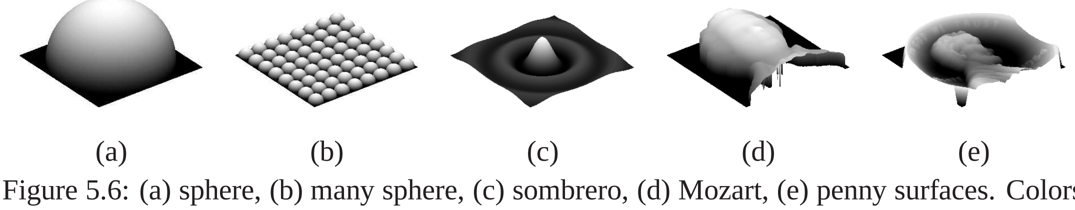


Figure 5.6: (a) sphere, (b) many sphere, (c) sombrero, (d) Mozart, (e) penny surfaces. Colors indicates the depth of surface.

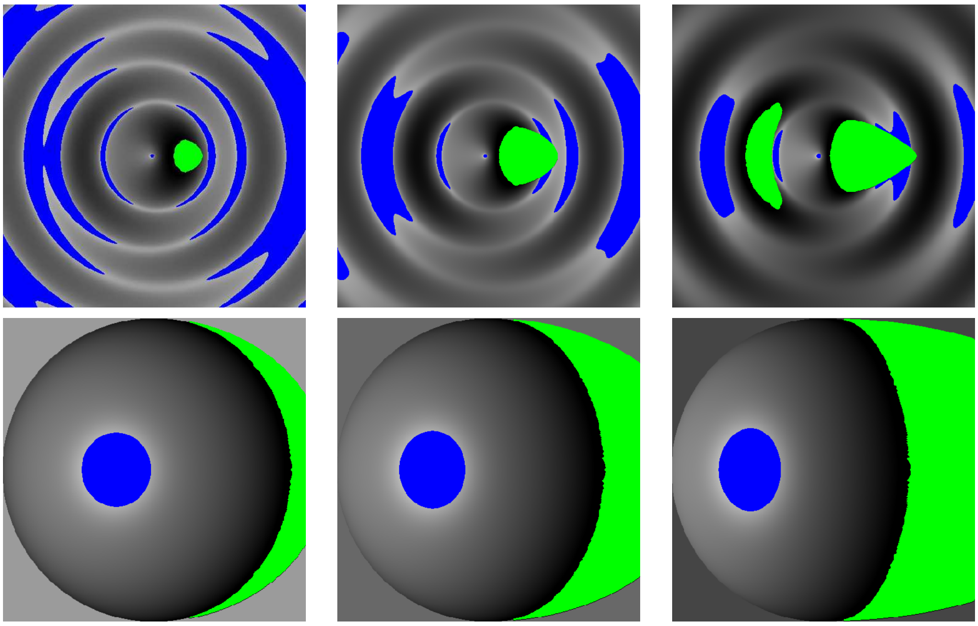
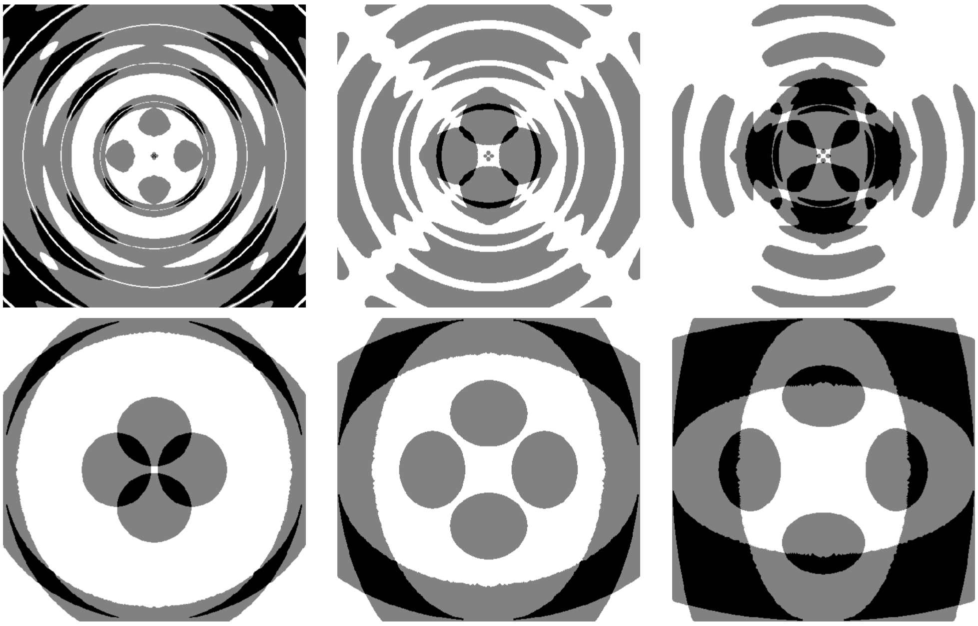


Figure 5.7: Specular (blue) and shadowed (green) regions on sombrero (top) and sphere (bottom) objects with (a) 30°, (b) 45°, (c) 60° light source zenith angles.

$30^\circ, 45^\circ, 60^\circ$ light source zenith angles in Fig. 5.8. In this figure, the sum of diffuse pixels, where Lambert reflection equation holds, from four light sources placed around the camera with 90° polar angles is plotted. The black regions indicate only two pixels are diffuse, the grays indicate three and whites indicate all four images are free of highlights and shadows. Better results should be expected for the sphere with 30° zenith angle light sources, since more diffuse pixels are present. On the other hand, for sombrero with 45° zenith angle produces least amount of highlights and shadows.

In the tests, the synthetic images were rendered with Torrance-Sparrow reflection model given at Eq. 2.14 [16]. Here the intensity value I is composed of specular and diffuse terms. The specular term depends on the angle between the light reflection direction $\mathbf{r}$ and the camera direction $\mathbf{c}$. In tests, uniform specular albedo ($\rho_s = 1.0$) and uniform diffuse albedo ($\rho_d = 1.0$) were used. The uniformity condition, however, was not a precondition for the solutions. Shininess parameter m is chosen to be 2.0, which was found to be the worst case in our previous study [45].

In synthetic image rendering, the orthogonal projection was used for both the image and shadow map projection processes. The depth of the test surfaces were less than 1% of the



(a) (b) (c)

Figure 5.8: The total count of diffuse pixels, where Lambert reflection equation holds on sombrero (top) and sphere (bottom) objects with (a) 30°, (b) 45°, (c) 60° light source zenith angles. Black, gray and white pixel values indicate that 2, 3 and 4 pixels among 4 have diffuse intensity values.

camera and light working distance, thus, orthogonal projection can be accepted as a good approximation.

The thermal noise of conventional camera was represented with a zero mean Gaussian noise added to the intensity values [47]. The standard deviation (noise power) of Gaussian noise was varied from 5% to 20% of the image dynamic ranges. The rendered images of the Mozart with no noise and 20% Gaussian noise are shown in Fig. 5.9. Both scenes are lightened with a single light source from the left with 45° zenith angle.

Image generation parameters are fixed throughout the test and can be found in Table 5.1. Perspective projection and inverse square law are not used.

5.4 Fine Tuning Thresholds

The quality of the mask depend on the threshold values. For a fair comparison among the masks, each individual threshold value should be tuned to minimize the resultant normal

(a) (b)

Figure 5.9: Mozart images with 45° zenith angle light source from left with (a) no noise and (b) additive Gaussian noise σ = 20% dynamic range.

Table 5.1: Image generation parameters

Parameters	**Symbol**	**Value**
Image Size		[512 512]
Pixel Intensity	I	
Light Source Position	$d * (distancetoLookAt)$	[1 1 2]
Light Source Power	μ	1.0
Illumination Vector	$s = \mu d$	
Reflection Direction	$r = 2(s^T n)n - s^T$	
Camera Position	$v * (distancetoLookAt)$	[0 0 2.41]
Camera Look at Position		[0 0 0]
Camera Up Direction		[0 1 0]
Focal Length		1.0
Reflection Model	Torrance Sparrow	
Diffuse Albedo Factor	ρ_{diff}	1.0
Diffuse Albedo Type		Constant
Shininess	m	2.0
Specular Albedo Factor	ρ_{spec}	1.0
Specular Albedo Type		Constant

Table 5.2: Threshold values for 4, 6, 8, 12 images

Threshold Name	**Image Count**			
	4	**6**	**8**	**12**
k_{low}	0.02	0.06	0.06	0.08
k_{up}	0.35	0.35	0.35	0.35
k_{Hi}	0.03	0.04	0.05	0.06
k_{CS}	-0.1	-0.1	-0.1	-0.1
t_{NL}	0.02	0.04	0.02	0.01
t_{Re}	$min(z, (90 - z)/2, p)$			

error. In the tests, the best value found for threshold values mainly depends the noise and number of the images.

The effect of the image noise is presented in Fig. 5.10 at the left side plots. As the noise level increases, threshold values also increase for most cases. An exception is the reflection threshold, such that its value seems to be less susceptible to the noise change.

In general, the amount of the image noise is not known. Hence average of the no noise, the 10% and the 20% Gaussian noise cases are used to fine tune the thresholds.

In Fig. 5.10 at the right side, the change of quality factor with respect to the threshold values for 4, 6, 8, 12 images are shown. Noisy images with 10% and 20% were used and averaged in these plots. In general, the threshold values increase from 4 to 12 images. This is expected, since when more images are available for PS, masking out more pixels are tolerable.

For the reflection mask, the threshold angle should be selected such that none of the reflection cones coincide with any other cones or shadowed regions [2]. This fact is concluded with the formula 5.1, given below, derived from surface geometry.

$$t_{Re} = min(z, (90 - z)/2, p) \tag{5.1}$$

Here z is the zenith angle, and p is the polar angle of all light sources. The formula was verified with synthetic image tests as seen in Fig. 5.11.

The selected threshold values of the masking methods are listed in Table 5.2.

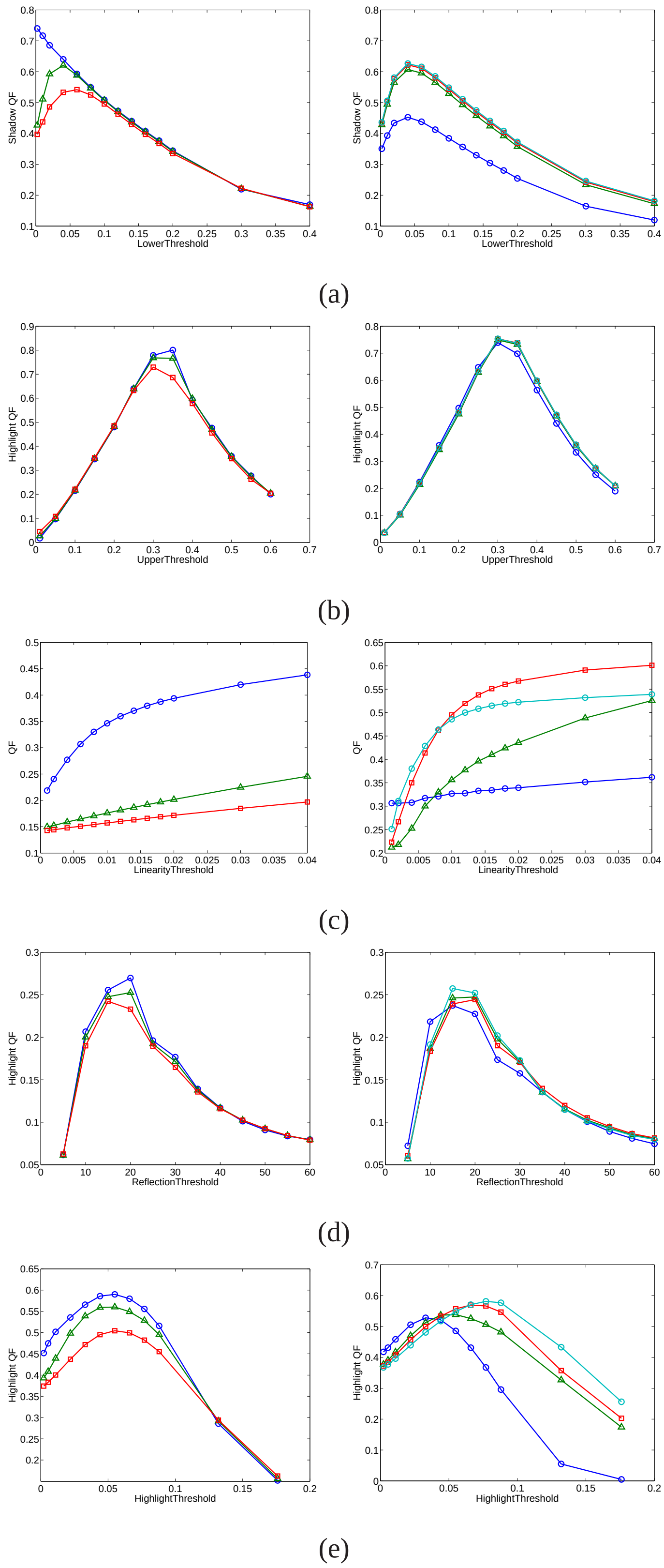


Figure 5.10: Change of quality factor with respect to (a) intensity lower k_{low}, (b) intensity upper k_{up}, (c) linearity t_{NL}, (d) reflection t_{Re} and (e) highlight k_{Hi} thresholds. At the left side, different noise levels, no noise (blue), the 10% (green) and the 20% (red) and at the right side, different image counts, 4 (blue), 6 (green), 8 (red), 12 (light blue), are plotted.

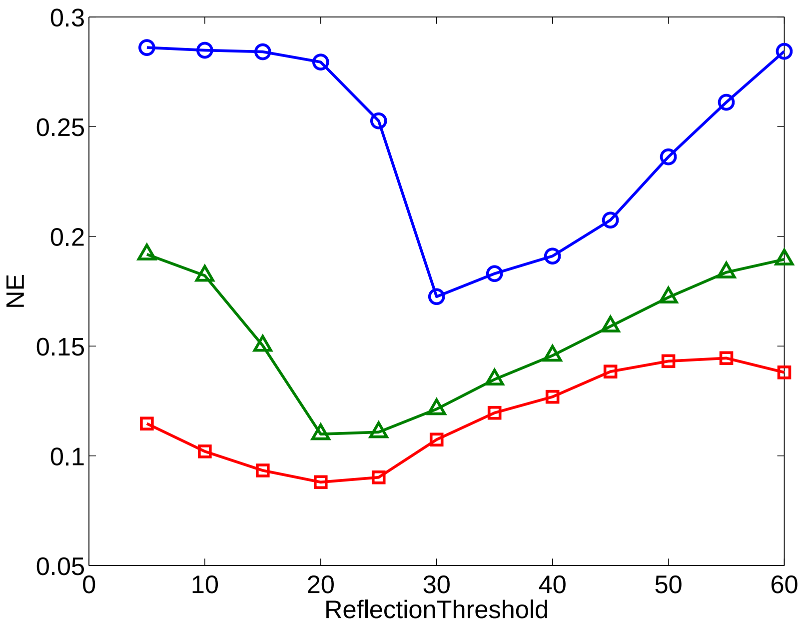


Figure 5.11: Change of normal error with reflection threshold for 30° (blue), 45° (green) and 60° (red) light source zenith angles.

5.5 Unified PS Results

The unified PS framework is capable of using any combination of masks in PS. In this study, only a subset of all possible combinations are discussed to refine the amount of presented data. The refined set includes single mask performances, masking algorithms in previous works and combinations that masks both shadows and highlights at the same time.

The previous works that are implemented for comparison, are Coleman et al. [17], Barsky et al. [2], Sun et al. [3], Argyriou et al. [4] and Mukaigawa et al. [19]. Coleman's method is based on Coleman and Jain mask (CJ), that is defined in their work. Barsky's method uses Non-Lambert quadruple mask and reflection mask (NL-Re) for exactly four images. Sun et al. uses the extended Non-Lambert quadruple mask (xNL) for six images and claims that it can be generalized for any number of images. Argyriou's method combines Non-Lambert quadruple and shadow masks (NL-Sh). Mukaigawa et al. employs random sample consensus (RA) on intensity values to filter outliers such as highlights and shadows.

In Table 5.3, the normal errors, elapsed times and quality factors of highlights, shadows and total are presented. These values are the average of all objects, all light combinations and

Table 5.3: The overall results of all mask methods, for all light sources and 4 light sources only.

Masks	**Normal Error** (°)	**Elapsed Time (sec)**	**Highlight QF**	**Shadow QF**	**Total QF**
		All Light Sources			
None	**14.78**	**0.11**	**N/A**	**N/A**	**N/A**
xNL	**11.37**	**16.94**	**0.28**	**0.23**	**0.25**
Th	**8.63**	**1.67**	**0.74**	**0.56**	**0.64**
Re	13.39	4.30	0.33	N/A	0.22
SS	14.54	3.37	N/A	0.35	0.21
Sh	12.74	12.23	N/A	0.54	0.32
Hi Re	12.80	5.60	0.48	N/A	0.28
Hi Re SS Sh	9.89	16.91	0.46	0.60	0.53
Hi Re Sh	10.53	16.31	0.47	0.54	0.51
CS	13.55	5.34	N/A	0.14	0.11
xNL Sh	**11.36**	**28.80**	**0.23**	**0.25**	**0.23**
xNL Re Sh	11.34	30.38	0.23	0.25	0.23
xNL Hi Sh	11.34	32.66	0.26	0.25	0.25
xNL Hi Re Sh	11.31	32.50	0.28	0.25	0.26
Th Re Sh	8.63	17.77	0.38	0.56	0.44
Th Hi Sh	9.12	22.83	0.27	0.55	0.39
Th Hi Re Sh	**8.28**	**20.12**	**0.55**	**0.56**	**0.56**
Th Hi Re SS Sh	8.29	20.26	0.55	0.56	0.56
RA	**13.23**	**5.00**	**0.29**	**0.22**	**0.27**
		4 Light Sources			
None	**15.18**	**0.08**	**N/A**	**N/A**	**N/A**
NL Re	**12.27**	**1.01**	**0.49**	**0.16**	**0.27**
Th	12.29	0.56	0.66	0.43	0.54
NL Re Sh	12.39	7.67	0.41	0.17	0.26
CJ	**14.14**	**0.48**	**N/A**	**N/A**	**N/A**
RA	**17.72**	**0.68**	**0.14**	**0.08**	**0.12**

all noise levels. The overall normal error without any masks (None) is 14.78°. Utilizing the extended Non-Lambert quadruple mask proposed by Sun (xNL), reduced the normal error to 11.37°. More improvement is accomplished with the addition of the shadow mask to Non-Lambert quadruple mask (xNL-Sh) as proposed by Argyriou et al.

The threshold masks (Th) produced impressing enhancements on the normal errors. The threshold mask reduced normal error to 8.63° by its own. Although the minimum normal error (8.28°) is achieved with the combination of threshold mask, highlight mask, reflection mask and shadow mask, (Th-Hi-Re-Sh) the improvement with respect to the threshold mask only case (Th) is less than a degree with the cost of eightfold computation time for the cases presented here.

At the bottom of the Table 5.3, results with only four light sources are presented to compare Barsky's (NL-Re) and Coleman's (CJ) methods. For four images, Coleman and Jain (CJ) method did not perform as good as other masking methods and Barsky's method (NL-Re) performed slightly better among other.

The overall results of (RA) method was 13.23°, which was a slight improvement compared to the others. Also, as seen at the bottom of the Table 5.3, four light sources results were even worse than None results. These results indicates that image count used in PS have a great influence on (RA) performance. With the increasing number of images, (RA) method may have better results.

The maximum time elapsed for the execution of these was $32 secs$. The best performing configuration (Th-Hi-Re-Sh) calculated results in $20 secs$. All of the tests were conducted on a regular PC (Intel Core 2 Quad CPU) and Matlab. The native implementations (e.g. C++) of these algorithms were expected to work faster than Matlab scripts.

(None), (NL), (Th), (NL-Sh) , (Th-Hi-Re-Sh) and (RA) masking combinations were selected for in depth analysis. The average normal error versus the noise plot for the selected masks is presented in Fig. 5.12. Obviously, with more noise on the images, normal errors increased. For masked PS methods, the normal errors increased faster indicating that masked PS methods are more sensitive to noise than None mask PS. Also Th mask was found to be less sensitive to noise than NL mask. Oppositely, RA mask is found to be very sensitive to noise such that the resultant error became worse than None mask PS when added Gaussian noise standard

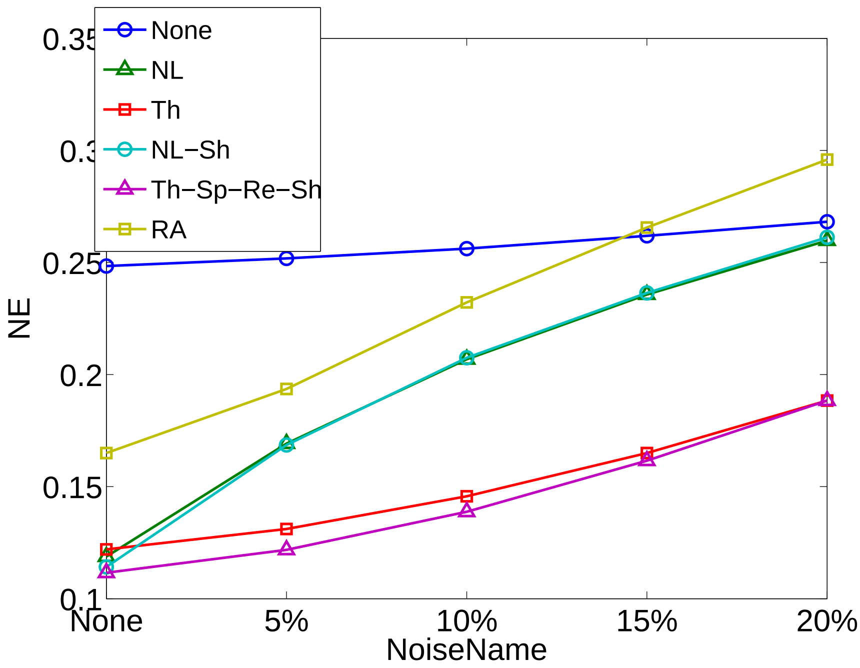


Figure 5.12: Normal error versus noise for the selected masks.

deviation is more than %15 of image dynamic ranges.

In Fig. 5.13, the average normal error versus the light count (image count) plot for the selected masks is presented. For the no mask PS (None), the normal error slightly decreased with increasing number of images for PS from 4 to 12 images. Utilization of the masks, however, decreased normal errors more. In addition, the relative improvement in the normal error in masked PS with respect to no mask PS increases with increasing images count. Hence, with more images in hand, masks were able to recognize shadows and highlights better. The most dramatic improvement is performed by (RA) mask with the increasing image count. With more images, consensus set of RANSAC became robust to the image noises. These results denoted that at least 6 images should be used with RA mask and with noisy images the image count should be further increased.

The Non-Lambert quadruple mask (NL) with 12 images performed worse than with for 12 images as seen in Fig. 5.13. This is because of low threshold value has to used for 12 images. Noise performance of the NL mask with various image counts are plotted in Fig. 5.14. Without any noise (blue), NL was able to mask with 12 images, but with 5% (green) and 10% (red) noise, it performed worse than 8 images result.

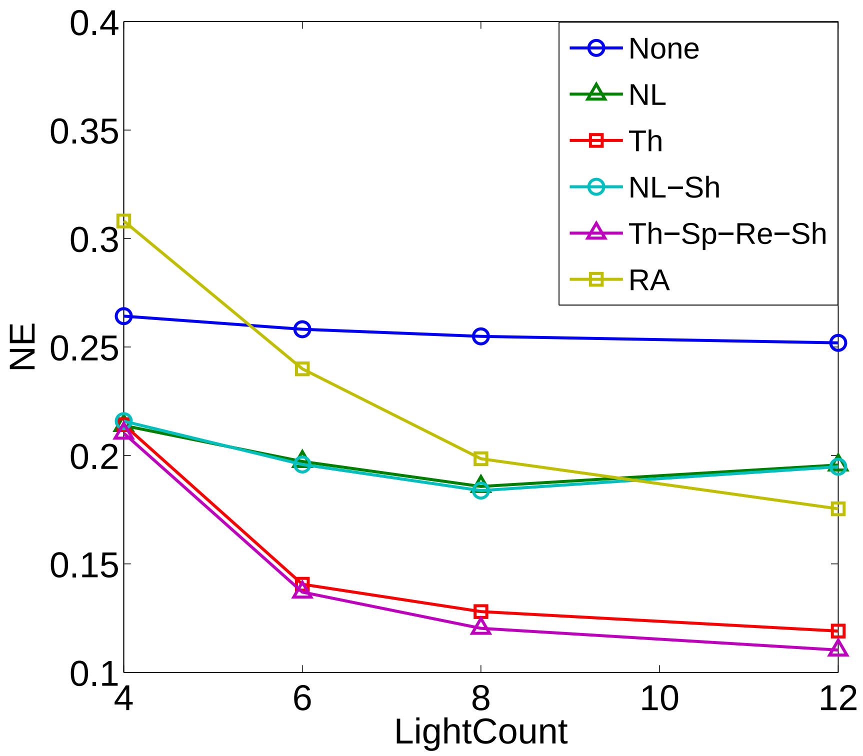


Figure 5.13: Normal error versus light count for the selected masks.

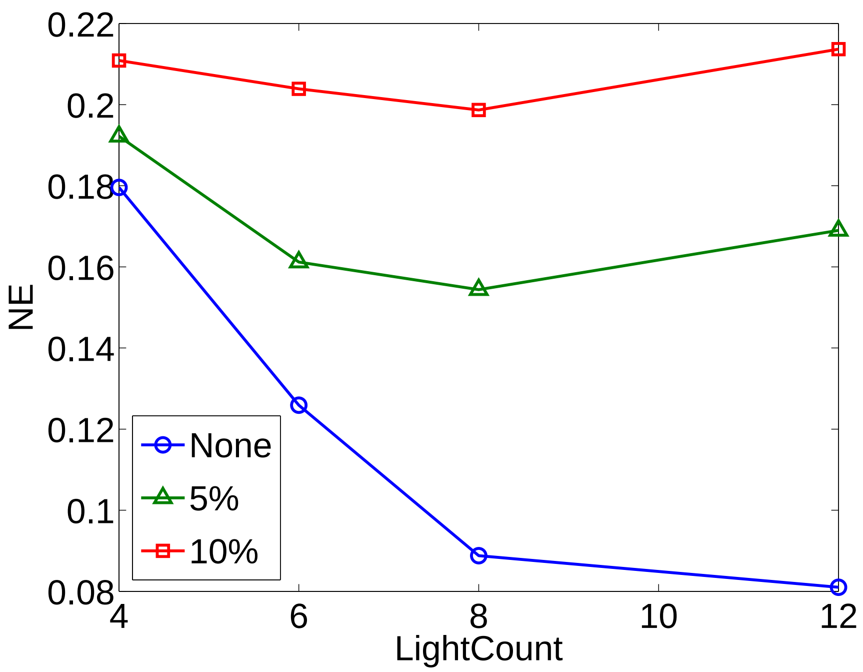
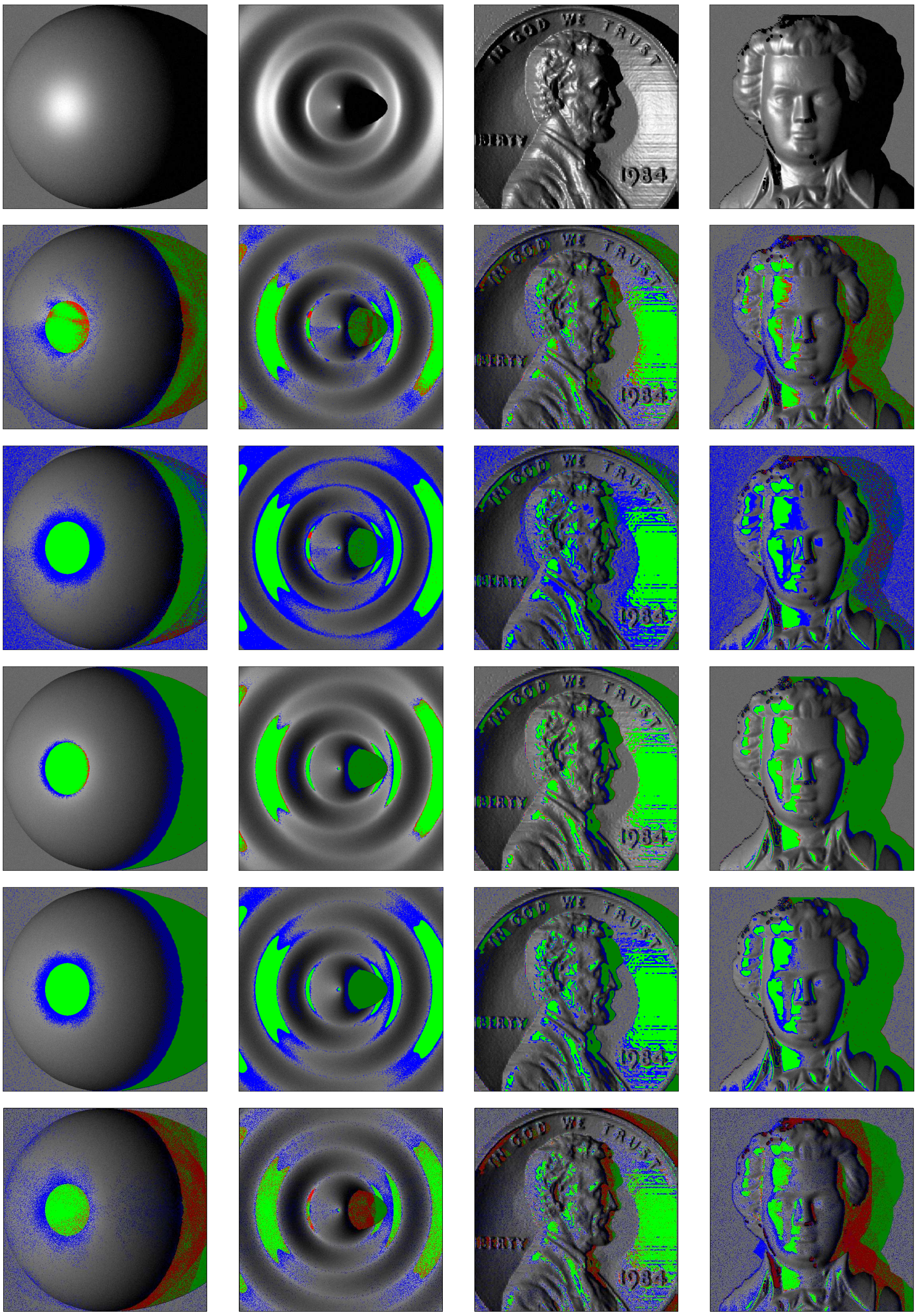


Figure 5.14: Normal error versus light count for the NL mask with different noise powers.



(1) (2) (3) (4) (5) (6)

(a) (b) (c) (d)

Figure 5.15: (1) Sample images with 10% Gaussian noise, illuminated from left with 45° zenith angle. Calculated masks are plotted on images with green for TP, with blue for FP and with red for FN for (2) NL, (3) NL-Sh, (4) Th, (5) Th-Hi-Re-Sh, and (6) RA.

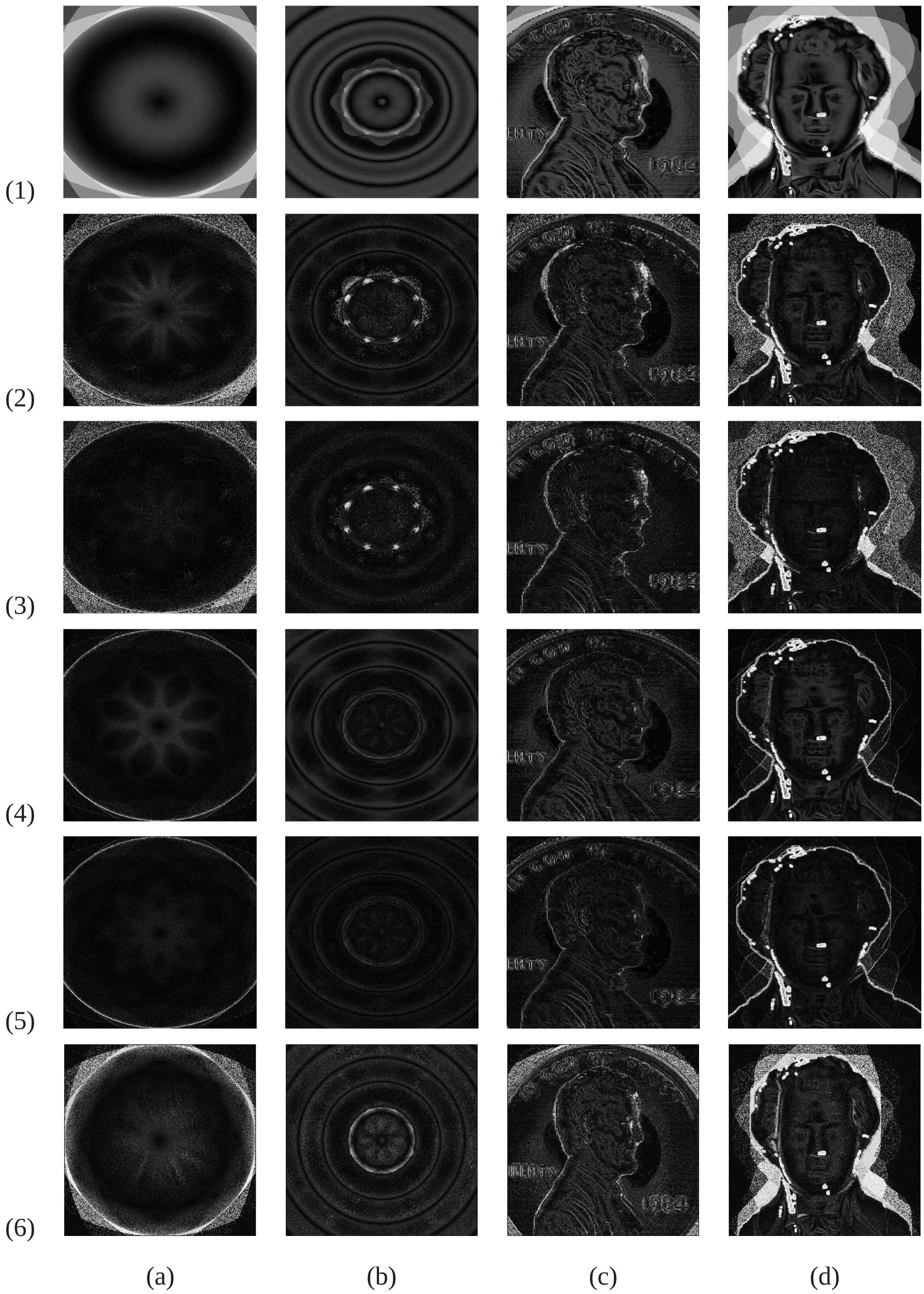


Figure 5.16: Resultant normal errors from 10% Gaussian noise, illuminated with 45° zenith angle lights for (1) None, (2) NL, (3) NL-Sh, (4) Th, (5) Th-Hi-Re-Sh, and (6) RA.

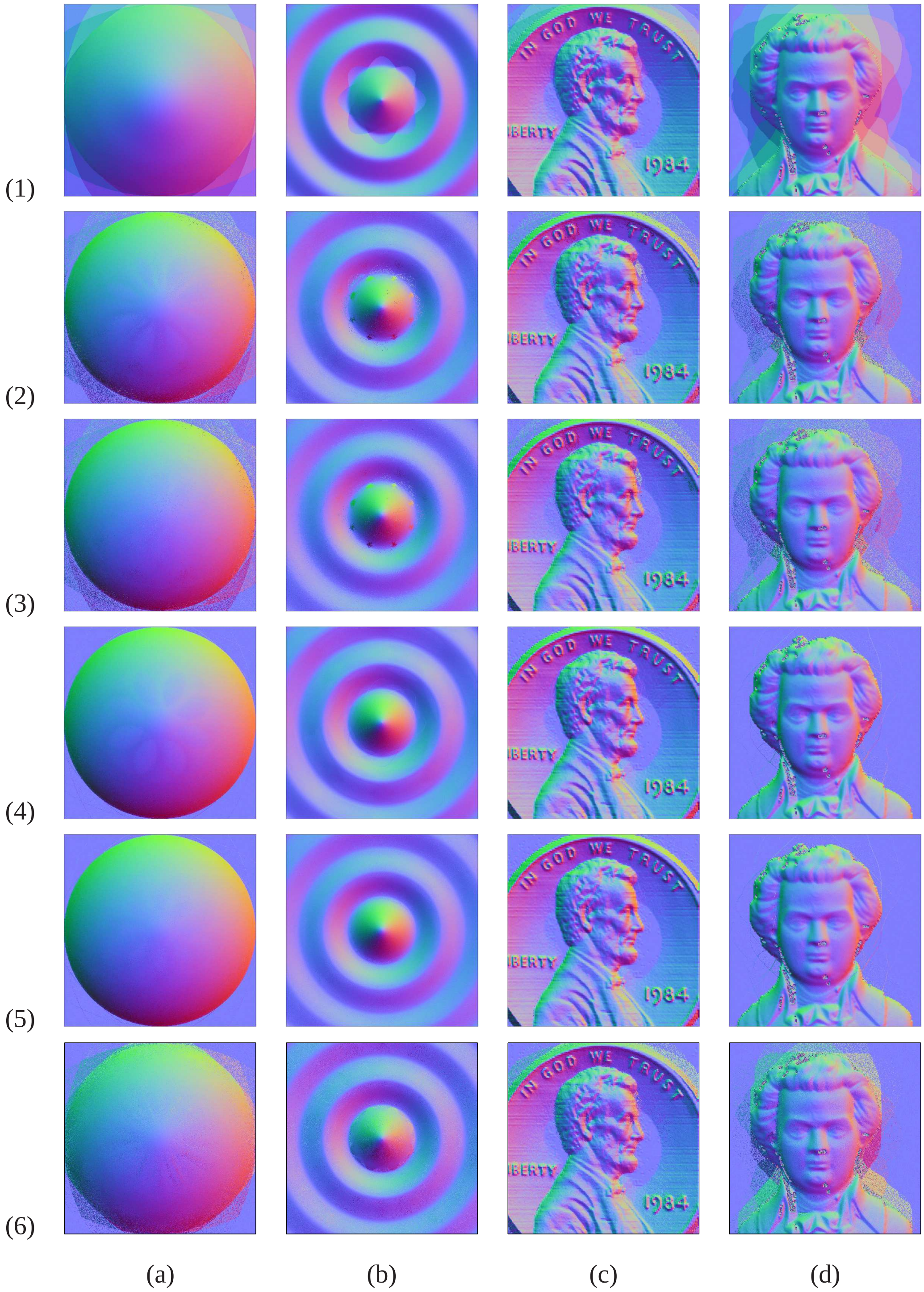


Figure 5.17: Calculated normals from 10% Gaussian noise, illuminated with 45° zenith angle lights. Normals are encoded in RGB for (1) None, (2) NL, (3) NL-Sh, (4) Th, (5) Th-Hi-Re-Sh, and (6) RA.

A sample test configuration that uses 8 images illuminated with 45° zenith angle light sources with 10% Gaussian noise, is selected for the visual presentation of calculated masks. The masks found are shown for four test surfaces (a) semisphere, (b) sombrero, (c) penny and (d) mozart in Fig. 5.15. Calculated masks are colored with green for the true positives (TP), with blue for the false positives (FP) and with red for false negatives (FN). The first row 5.15 (1) is row images where no mask is applied. The extended Non-Lambert quadruple mask in Fig. 5.15 (2) have some shortcomings in detecting shadows. In Fig. 5.15 (3) shadow FN are reduced with the additional shadow mask. In Fig. 5.15 (4), the threshold mask detected nearly all the erroneous pixels with some FP. Since 8 images were used, false alerts did not ruined the results. Blue false alerts spread all over images for the extended Non-Lambert quadruple mask indicating the high noise sensitivity of the mask. For the threshold mask (Th), noisy pixels, recognized as FP, are created only around TP. In Fig. 5.15 (5), with addition of Hi, Re and Sh masks to Th mask, the masked regions enlarged. The major part of the error have been removed by Th and additional masks handled the small errors at the boundaries of the Th masks. In Fig. 5.15 (6) that represents only RA masks, the red FN regions were at self shadowed pixels. The RA mask missed the self shadows when more than one intensity values were shaded. The main cause is the residual errors of self shadows were lesser compared to cast shadows and highlights. Two small outliers, the self shadows, were included in the consensus set of RANSAC.

In Fig. 5.16, norm of calculated and original normal vectors are plotted. The white regions indicate large normal errors, mainly due to the cast shadows and highlights as seen in Fig. 5.16 (1). Cast shadow errors were accumulated around the objects while highlight errors were on the objects. Both of the highlight and shadow errors reduced with masks. The noisy image performance of each masks can be seen in these images as salt and paper effect on the normal errors. The NL and RA masks at Fig. 5.16 (2, 3, 6) have these salt and paper effect heavily. These masks are more sensitive to noise than others. Also, at the boundaries of the masks, the normal errors changed rapidly resulting a flower like pattern in Fig. 5.16 (4a). These mask patterns were also transfered resultant normals in Fig. 5.17 (4a).

Lastly, in Fig. 5.17, calculated normals vectors are plotted with RGB values $[r, g, b] = [(n_x + 1)/2, (n_y + 1)/2, n_z]$. In Fig. 5.17 (1), without any masks, errors due to the cast shadows appeared around the Mozart and the sphere. These errors reduced with usage of the masks. However, due to the noisy images used in the tests, the calculated extended Non-Lambert

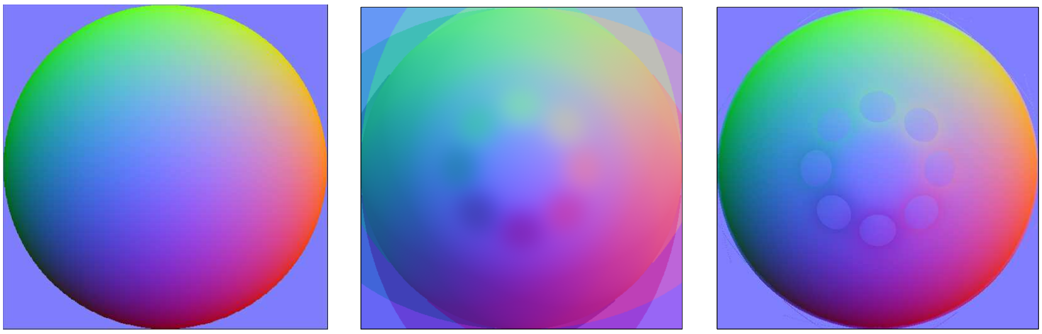


Figure 5.18: Normals of semisphere (a) and normals created by None PS (b) and Th PS (c) with eight 45° zenith light sources.

quadruple mask is very noisy, resulting noisy normals. The threshold mask performed better under same conditions as seen in Fig. 5.17 (4) and (5).

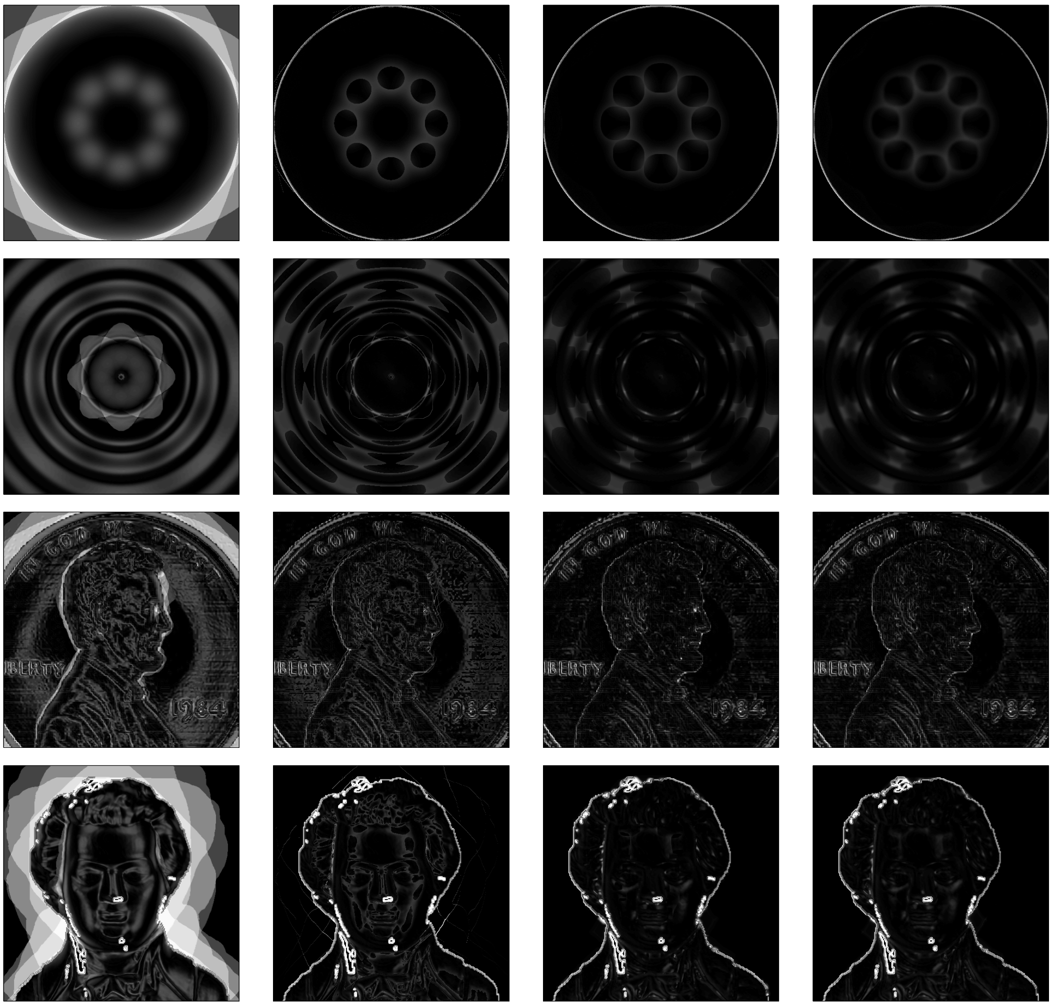
5.6 Weighted PS Results

The false edges generated at the boundaries of the masks was noted in the previous section. This problem can be seen in Fig. 5.18 on semisphere with no noise. The original normals seen in Fig. 5.18(a) is clearly better reconstructed with Th mask PS as in seen in Fig. 5.18(c). However at the mask boundaries where utilized images set were changed, the normal vectors are changed rapidly and caused false edges.

In this section weighted PS method that works on clearing this false edges, is explained. The flow of the method is explained in "Weighted PS" section of Chapter 3. The weighted PS method can be applied to any mask explain in this work. Since the used mask is not very important, simple Th mask is used as an example. No noise added to the images to see the false boundaries clearly. Eight light sources with $Z_{source} = 45°$ are used as PS input images.

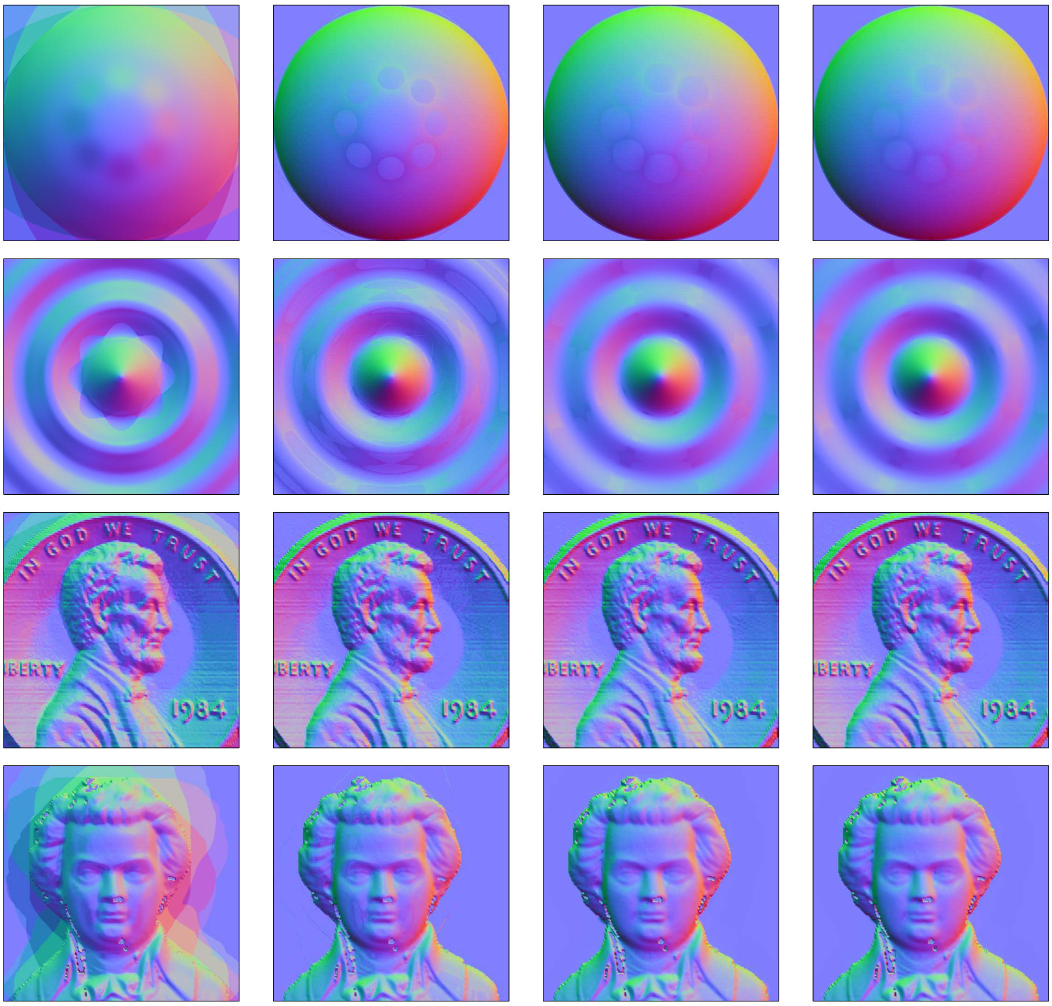
The size of the dilation and smoothing filters depends on the error margin of the used mask. Here dilation kernel is selected as a circular disk with 7 pixel radius. The smoothing filter is a Gaussian with 3 standard deviation. Both kernels have 15x15 size.

The normal error images that are the norm of difference vector between original and calculated normal vectors are presented in Fig. 5.19. The no mask results suffered from highlights and



(a) (b) (c) (d)

Figure 5.19: None mask (a), Th mask (c), dilated Th mask (b) and dilated smoothed Th mask (c) normal errors for semisphere, sombrero, penny and mozart.



(a) (b) (c) (d)

Figure 5.20: None mask (a), Th mask (c), dilated Th mask (b) and dilated smoothed Th mask (c) normals for semisphere, sombrero, penny and mozart.

shadows as seen in Fig. 5.19(a) with no mask. Th mask removed major errors as seen in Fig. 5.19(b) but sharp false edges were created. Fig. Fig. 5.19(c) displays the dilated Th mask results. The dilation operation enlarged the masked regions and clear the mask misses around the threshold. The mozart normal error image have false edges at the cast shadow boundaries, and were cleared at bottom image of Fig. 5.19(c). The left most column, Fig. 5.19(d), represents dilated and smoothed mask results that the sharp false edges at highlights were smoothed. The color coded resutant normals of the same test were presented in Fig. 5.20.

The effect of weighted PS is obvious when render images of the generated normals are created

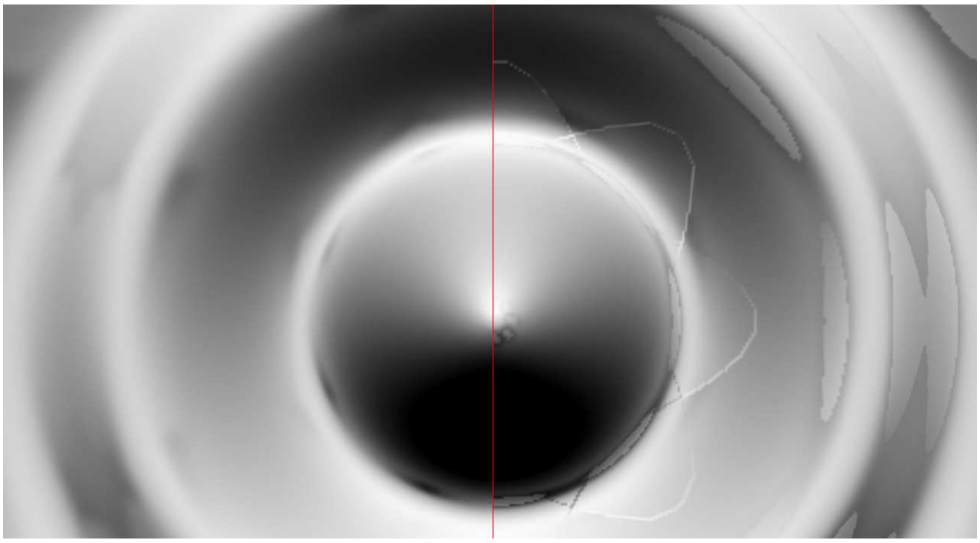


Figure 5.21: Weighted Th mask PS (left) and Th mask PS (right)

as in Fig. 5.21. The false edges of Th mask on smooth sombrero surface can be seen right side. The left image is rendered from the normals of weighted Th PS that have exactly same mask but weighted, dilated and smoothed.

5.7 Multiple Zenith Results

The Table 5.4 represents the mean NE of three sombreros with various illumination configurations. The illumination configuration is shown with a formula ($Z_{source} x n_i + \cdots + Z_{source} x n_i$). For example, the formula $30° x4$ + $45° x4$ + $60° x4$ represents that four light sources were placed on a circular ring with $Z_{source} = 30°$, $Z_{source} = 45°$ and $Z_{source} = 60°$ and totally 12 light sources were used. Each ring of light sources placed with an polar angle offset value to have equal polar distances between two consecutive sources.

The first part of the Table 5.4, includes samples of double zenith illumination configurations with total 16 lights. The average value of the Z_{source} was changed from 35° to 65° and empirically configuration $35° x8 + 65° x8$ with 50° average value have the minimum NE. The rule of thumb given in Eq. 4.2 proposed the configuration $30° x8$ + $60° x8$, which is the second best solution.

Similarly, at second part of the Table 5.4, the difference of two Z_{source} have been experimented.

Again previously proposed $30^\circ x8 + 60^\circ x8$ illumination configuration have the second best NE performance. These results verified the rule of thumb in Eq. 4.2 is a good prediction to place the double zenith light sources.

The last part of the Table 5.4 shows the results of the multiple zenith illumination configurations. All the illumination configurations have 12 light sources. These light sources were located on double, triple and quadruple circles around the camera. The triple circles illumination combination have slight performance improvement with respect to double. Others are not better than double illumination configuration.

The $Z1_{source} = 30^\circ, Z2_{source} = 60^\circ$ double zenith light sources configuration ($2Z$) is compared with $Z_{source} = 45^\circ$ single zenith light sources configuration ($2Z$) for all three sombreros. Both diffuse (Di) and highlight shadow (Hi-Sh) images were used. Resultant NE versus image count n_i is plotted in Fig. 5.22. For diffuse images (bottom two plots), NE with $1Z$ was always less than $2Z$. This result was expected since $1Z$ is the optimal solution without highlights and shadows. But when Hi-Sh images were used, $2Z$ performs better than $1Z$ if $n_i > 4$. With more images used, the difference between $1Z$ and $2Z$ increases in favor of $2Z$. On this figure, WPS is weighted PS which will be explained in the next section.

In the first two rows of Table 5.5, numerical results of this experiment are shown. By using 8 lights at $Z1_{source} = 30^\circ$ and 8 lights at $Z2_{source} = 60^\circ$ instead of using 16 lights at $Z_{source} = 45^\circ$ the resultant NE is reduced 12.9%.

Table 5.5 presents the mean NE of all three sombreros with various image configurations and PS solution methods. In this table, also the improvement percentages with respect to $1Z$ PS are given. First thing to notice is NE was improved with double zenith ($2Z$), weighted PS methods. The maximum improvement is achieved with $eZ_{source}Z_{surface}$ weight, that is for each image count (n_i), angular NE is improved more than 30%. For 16 images, improvement is 39.4%, that is much better than 12.9% $2Z$ PS. Secondly, the angular NE reduces with increasing n_i from left to right. For $1Z$, (first row) the percent improvement is about 4% from 8 images to 32 images. On the other hand, $2Z$, the improvement is more than 10%, which indicates that if many images are available for PS, double zenith lights configuration is a better choice than squeezing them to a single circle.

The normal errors of the three sombreros (s1, s2, s3) with binary weighted and weighted

Table 5.4: NE for various Illumination Configurations

Illumination Configuration	NE
Double Zenith Average	
$20° x8 + 50° x8$	5.96
$25° x8 + 55° x8$	5.83
$30° x8 + 60° x8$	**5.51**
$35° x8 + 65° x8$	**5.33**
$40° x8 + 70° x8$	5.57
$45° x8 + 75° x8$	6.21
$50° x8 + 80° x8$	7.53
Double Zenith Difference	
$5° x8 + 85° x8$	13.79
$10° x8 + 80° x8$	9.37
$15° x8 + 75° x8$	6.08
$20° x8 + 70° x8$	5.16
$25° x8 + 65° x8$	**4.85**
$30° x8 + 60° x8$	**5.01**
$35° x8 + 55° x8$	5.50
$40° x8 + 50° x8$	5.86
Multi Zenith	
$30° x6 + 60° x6$	**5.73**
$30° x4 + 45° x4 + 60° x4$	**5.72**
$15° x4 + 45° x4 + 75° x4$	6.35
$15° x3 + 35° x3 + 55° x3 + 75° x3$	6.27
$30° x3 + 40° x3 + 50° x3 + 60° x3$	5.80

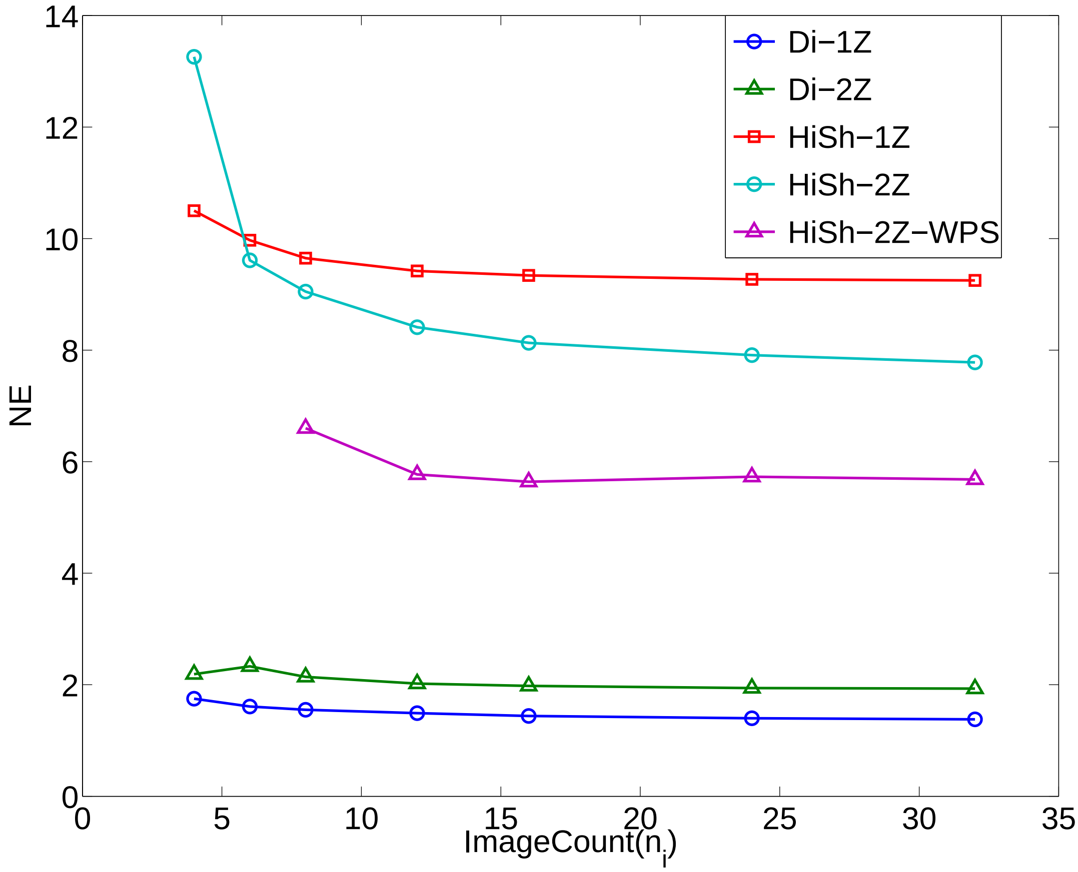


Figure 5.22: Normal Error versus image counts for single and double zenith angle light sources, Lambert and TS images

Table 5.5: NE for various image configurations and PS solution methods

		8	**12**	**16**	**24**	**32**
Source	Weight			**NE**		
$1Z$	$None$	9.65	9.42	9.34	9.27	9.25
$2Z$	$None$	9.05	8.41	8.13	7.91	7.78
$2Z$	e	8.44	7.84	7.89	8.05	8.08
$2Z$	eZ_{source}	7.92	7.14	7.11	7.25	7.26
$2Z$	$eZ_{surface}$	7.00	6.25	6.17	6.26	6.22
$2Z$	$eZ_{source}Z_{surface}$	6.60	5.77	5.64	5.73	5.68
Source	Weight	**% Improvement NE w.r.t.** $1Z$				
$2Z$	$None$	6.2	10.7	12.9	14.7	15.9
$2Z$	e	12.6	16.7	15.5	13.2	12.6
$2Z$	eZ_{source}	17.9	24.2	23.8	21.8	21.5
$2Z$	$eZ_{surface}$	11.6	12.4	13.3	13.6	14.3
$2Z$	$eZ_{source}Z_{surface}$	31.6	38.7	39.6	38.2	38.6

Table 5.6: Angular NE for binary and weighted PS results for three sombreros

	s1	**s2**	**s3**	**Mean**
Weigthed	4.08	6.12	6.74	**5.65**
Binary	3.39	5.19	7.00	**5.19**

normal vectors are shown in Table 5.6. Binary weights that simply selects the less error normal, performed better than the weighted summation of normal vectors if only normal error is considered. But when all normals are plotted as in Fig. 5.23, binary selection of two data set caused the false edges. On the other hand, weights provides a smooth transition from one data set to the other one.

In Fig. 5.24, NE images of sombrero2 are displayed for in depth investigation. In Fig. 5.24a 8 images with $Z_{source} = 30°$ and in Fig. 5.24b 8 images with $Z_{source} = 60°$ were utilized in PS. At each case NE was concentrated at different regions due to highlights and shadows. In Fig. 5.24c all 16 images were used in total PS and now NE is reduced and uniform. In Fig. 5.24d the NE image resulted from weighted PS with $eZ_{source}Z_{surface}$ weights is shown. This

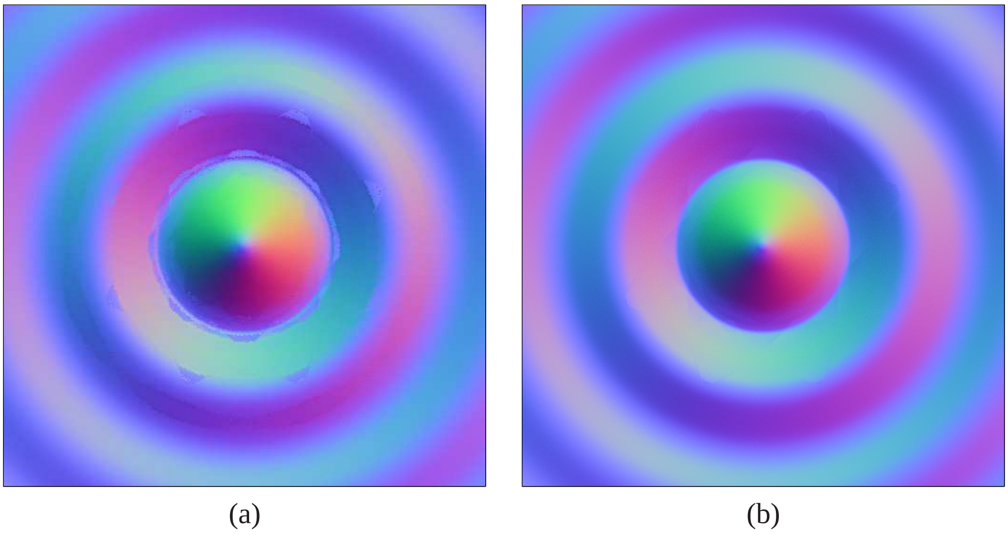


Figure 5.23: (a) False edges were occurred in binary normal vectors. (b) With weighted summation of normals, false edges on weight transitions were smoothed.

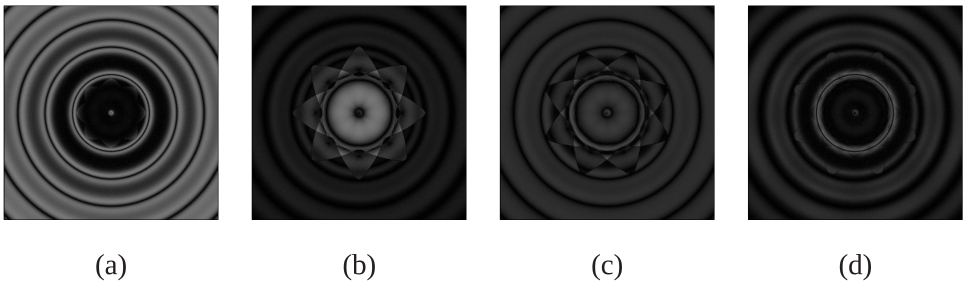
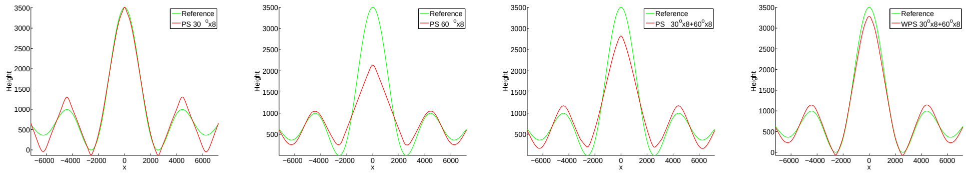


Figure 5.24: NE errors of sombrero2 for (a) $PS, Z_{source} = 30°, n_i = 8$, (b) $PS, Z_{source} = 60°, n_i = 8$, (c) $PS, Z_{source} = 30°, 60°, n_i = 16$, (d) $WPS, Z_{source} = 30°, 60°, n_i = 16$. White pixels have IE of 20% of image dynamic range.

image is nearly union of dark regions (less NE) of Fig. 5.24a and Fig. 5.24b that indicates the weights can effectively select the correct solution.

Fig. 5.25 displays the same result set on sombrero2 cross section. The cross section of original sombrero2 is plotted in **green** while the calculated surface that is the surface integral of PS normal vectors [37] is plotted with **red**. The cross section of the weighted PS solution fitted the best to cross section of original height map.

In this work, triple zenith illumination configuration and triple WPS was also tested. The triple zenith angles was chosen to be $Z1_{source} = 30°$, $Z2_{source} = 45°$ and $Z3_{source} = 60°$. The triple zenith configuration is compared with equal image count double zenith configuration.



(a) (b) (c) (d)

Figure 5.25: Cross sections of sombrero for (a) $PS\, Z_{source} = 30°, n_i = 8$, (b) $PS\, Z_{source} = 60°, n_i = 8$, (c) $PS\, Z_{source} = 30°, 60°, n_i = 16$, (d) $WPS\, Z_{source} = 30°, 60°, n_i = 16$,

The double zenith configuration resulted better than the triple zenith illumination. Also when the triple zenith configuration was employed in WPS, the weights of $Z2_{source} = 45°$ were mostly very small indicating that they were nearly never used.

CHAPTER 6

REAL IMAGE TESTS

6.1 Introduction

Before the real image tests, calibrations of the image acquisition hardware must be completed. The precise measurements of the physical placements of the hardware components are explained. Also image disturbances due to hardware and assumptions of PS method are corrected in this process. Later, visual results of all methods are presented on metallic surfaces with highlights and shadows.

6.2 Hardware Components

Image acquisition hardware used in real image tests can be seen in Fig. 6.1. The hardware used in the these tests was developed by BALISTIKA2010 System hardware group in TÜBİTAK UZAY. All parts of hardware are assembled in a black opaque box to avoid secondary illumination. Camera is placed vertically at the center of lights. 16 light sources were placed around camera pointing to field of view of camera. Detailed light source placement can be seen in Fig. 6.2. Cartridge case and bullet holders were placed on a computer controlled motorized linear X, Y, Z stages and rotational yaw, pitch, stages. Bullet could be rotated around itself with an extra motor. Also camera focus and zoom could be controlled with motors via serial port computer interface.

High power light emitting diodes (LED) were used as light sources. These warm white LEDs sink $350mA$ rated current from computer controlled current source. They were packed with small plastic lens, collimating light to $10°$ circular cone. Film diffusers were glued in front of

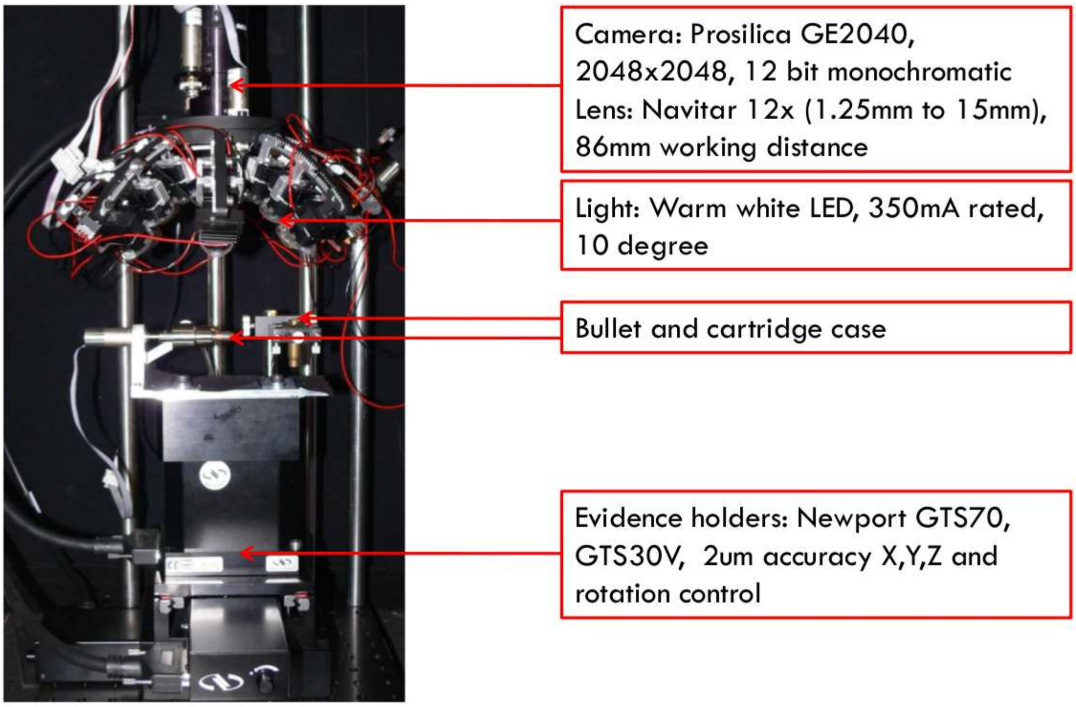


Figure 6.1: Hardware used for image acquisition

LEDs to have homogeneous illumination distribution in the field of view.

Camera used in the setup was Prosilica GE2040 that has a 2048x2048, 12 bit monochromatic CCD sensor. It can deliver 15 frames per second via gigabit Ethernet adapter. Camera configuration has done before image acquisition once, and the same configuration was used during all tests.

Navitar 12x lens was attached to the camera with appropriate adapters to have variable field of view from 1.25x1.25mm to 15x15mm and 86mm working distance. This lens has par-focal zoom system that focus is not distorted by zooming operations. The lens is controlled with two motors, one for the focus and one for the zoom. The pixel size at the image plane can be changed with the zoom motor from 7.32μm at the minimum magnification to 0.61μm for the maximum magnification setting.

In the real image tests, images were acquired with largest field of view and the minimum pixel resolution. The spatial resolution of the images were 7.32μm.

Evidence holders are placed on Newport GTS70 and Newport GTS30V motorized stages. These stages are capable of positioning the evidences with 2μm accuracy.

Figure 6.2: (a) Imaging camera at middle and 16 LEDs placed around it

6.3 Light Source Direction Calibration

Although a special hardware was designed and produced for the light positioning, the exact positions of the light sources were calculated with a metallic bearing ball test [48]. A specular metallic bearing ball with 9*mm* diameter was placed in the field of view and images were taken with each light source individually. Fig. 6.3(a) displays a sample image of the bearing ball illuminated with the light source 7. The center of the highlighted region was calculated from this image. The vector from the center of ball to the highlight was the reflection direction and the bisector vector of the camera direction and the light source direction. The camera direction was defined as [0, 0, 1] (z axis) and the light source direction was calculated from reflection direction. Light source to ball distance was measured physically and the average distance was found to be 86*mm*.

6.4 Light Source Power Calibration

The PS calculation employs the illumination powers of the light sources. In the Balistika 2010 hardware, all the light sources are same model LEDs emitting equal light sources up

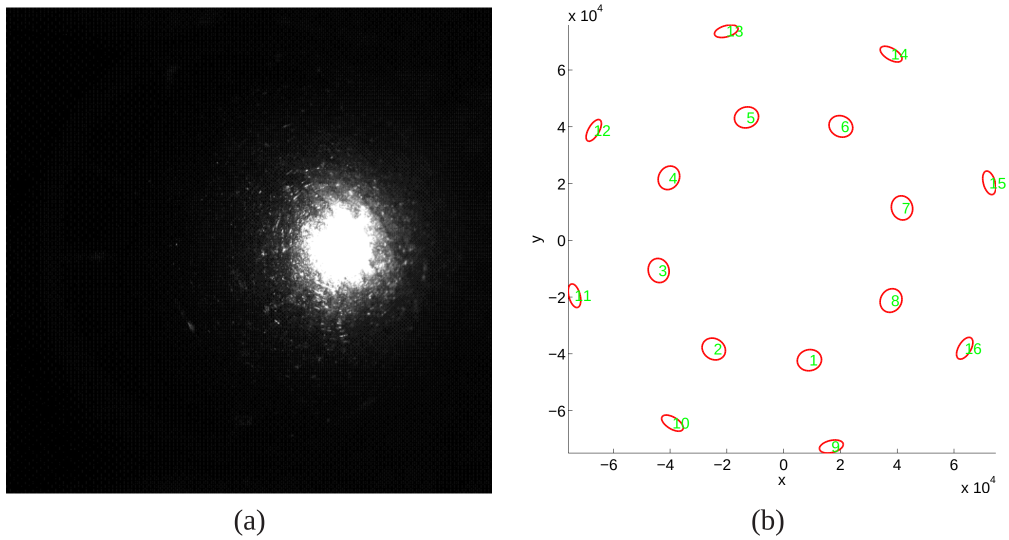


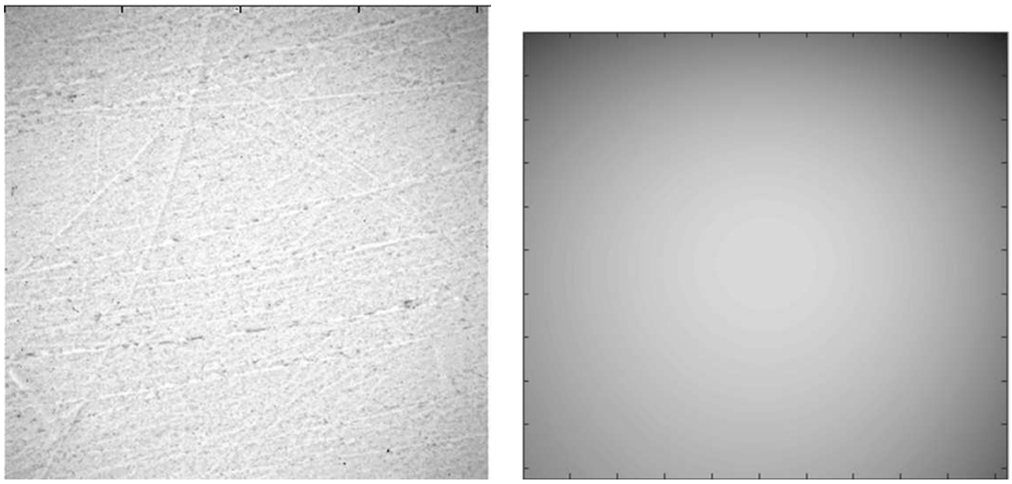
Figure 6.3: Bearing ball image illuminated from the light source 7 and the calculated light source positions looking from z, y and x axis respectively.

to 1% tolerance under same usage. However the irradiance on the target surface may not be equal due to targeting differences of the collimators and diffusers in front of the each LEDs. Consecutively, unbalanced illumination may occur on the target surface. The light source powers were calibrated with a planar white target. The test images of a planar white target, for each light source were acquired with same camera configuration. The mean values of image intensities are directly proportional to the irradiance on the surface. Light source power can be calculated by scaling irradiance value with solid angle of each light source that is $\sin(Z_{source})$.

If the target surface is placed horizontally (mean of the all surface normals are [001]), and nearly symmetric for each light sources, mean values of target images can be approximated to illumination powers. This approximation should be used carefully. If the surface is not placed horizontally, the resultant normals will be deformed. The cartridge cases are good example for this type of surfaces.

6.5 Radiometric Calibration

PS method estimates the surface normals from the irradiance changes of the surface with changing light source positions. Instead of the irradiance, the image intensities values are em-



(a) (b)

Figure 6.4: White calibration image and light distribution function.

ployed with linear camera response assumption. Secondly, uniform illumination assumption over the whole field of view is used to simplify calculations. These assumptions are checked and corrected with the radiometric calibration on images.

Radiometric calibration is composed of following calibration processes;

6.5.1 White image calibration

The Light sources used in image acquisition hardware do not have a uniform illumination distribution over the field of view [13]. This calibration process measures intensity values on a planar reference, white diffuse surface called Spectralon. Intensity values of the Spectralon were fitted to 2 dimensional polynomial function. This light distribution function representing illumination power at each pixel was used as a scale factor for each pixel. The images were scaled with the inverse of this function to correct illumination variations. Fig. 6.4 shows a sample of white calibration image and fitted light distribution function.

6.5.2 Gray image calibration

Radiometric response curve is the relation between the scene radiance and image intensity. In many computer vision systems, it is assumed that the image intensity of a point directly

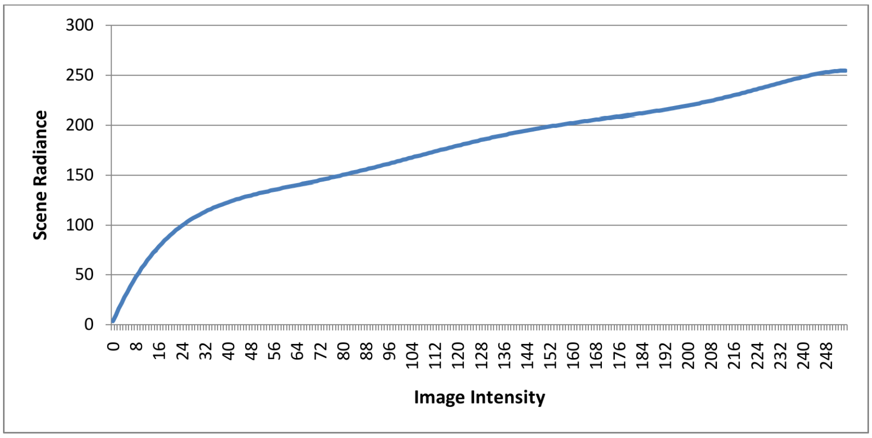


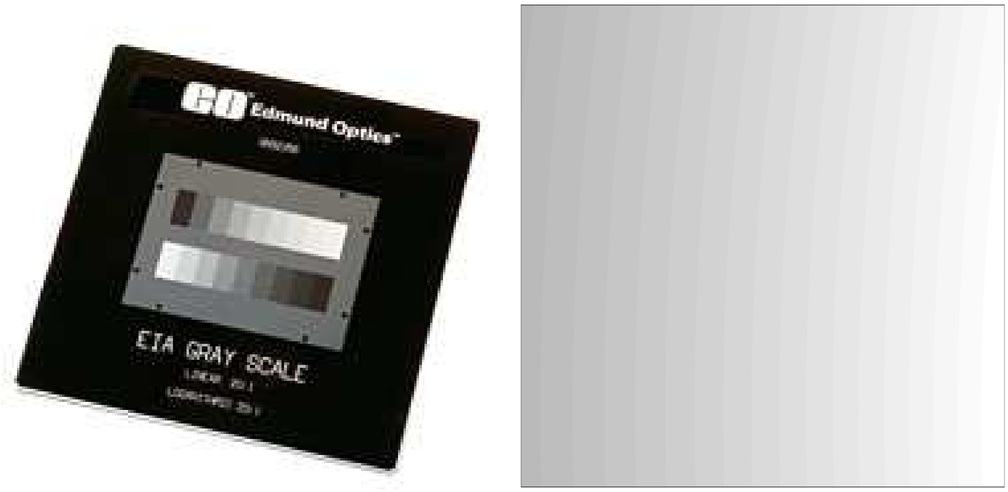
Figure 6.5: Radiometric response curve of the camera.

reflects the scene radiance of the point. However, this assumption does not hold in most cases. Camera producers use nonlinear camera response functions in order to compress the dynamic range of the scene to 8 bits [49]. PS algorithm uses scene radiance to reconstruct the scene in 3D. In order to calculate the scene radiance, all the images must be corrected by radiometric response function of the camera.

Camera response function was estimated from a gray scale pattern having 9 linear gray radiance values (see Fig. 6.6a). Mean intensity values of each these 9 gray regions is fitted to a 6^{th} order polynomial function resulting the graph in Fig 6.5.It is obvious that the relation between the scene radiance and image intensity is not linear.

6.5.3 Dark image calibration

Due to the thermal noise in CCD some pixels of the images may have relatively very high intensity values, even if all the lights are off. When these few erroneous pixels (around 200 in 4 million) are used in PS, sometimes unacceptable peaks occurred on the 3D surface. Dark image calibration simply detects these erroneous pixels using intensity threshold operation, and averages them with surrounding pixels.



(a) (b)

Figure 6.6: White calibration image and synthetic calibration image.

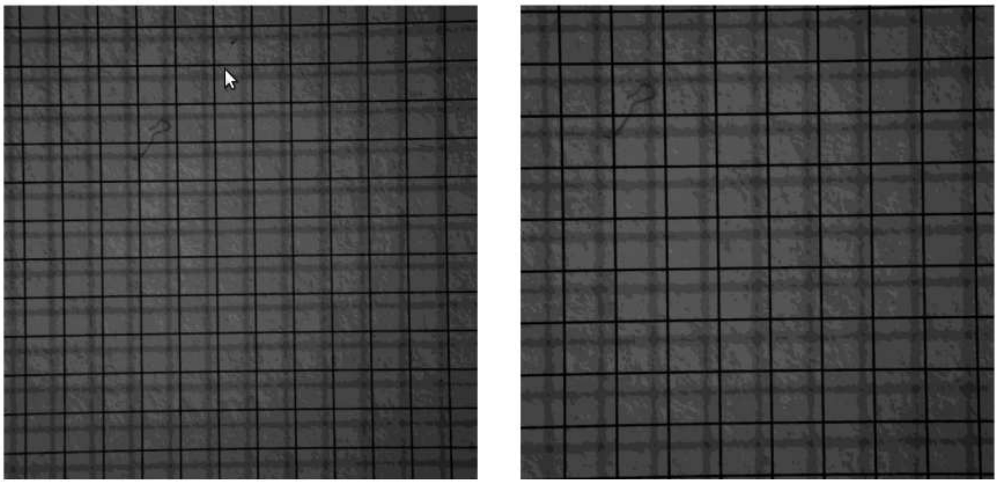
6.5.4 Synthetic image calibration

Since light sources were not placed at infinity, their illumination distribution will not be uniform over the field of view. There are two reasons of this, the illumination power reduces with light source distance, and secondly the angle of light source changes over the surface. After hardware assembly, planar diffuse white surface is rendered with measured light source and camera positions. Resultant synthetic images are used to scale images same as white image calibration. This calibration solves most of the convexity error on 3D surfaces. Fig. 6.6b shows a sample of the synthetic calibration image on the right.

6.6 Geometric Calibration

The camera lens combination used in the image acquisition has a very low lens distortion which can be seen at Fig. 6.7. So, it is not necessary to model and correct it.

In order to use absolute lengths in the generated 3D surfaces, absolute length corresponding one pixel must be calculated. For each zoom level, this calculation is done on the grid pattern having $1x1mm$ cells. Left image of Fig. 6.7 has pixel size of $6.00\mu m$ and right image has $4.64\mu m$ pixel size.



(a) (b)

Figure 6.7: Geometric calibration image for 12.3mmx12.3mm (a) and 9.5mmx9.5mm (b).

6.7 Unified PS Results

The synthetic image results presented in the previous chapter were also verified with real images. Three metallic test surfaces, a coin, a relief of two horses and viking statue, were selected to present both highlights and shadows. 2048x2048 8 bit monochromatic images of these test surfaces were acquired with Prosillica GE2040 camera and Navitar 12x zoom lens. Eight warm white power LEDs were placed around the camera with 62° zenith angles and 45° polar angles. In Fig. 6.8 (1), one of the eight images illuminated from lower right corner are shown. All eight images like these samples were filled with large shadows and highlights, especially for two horses and viking statue. From Fig. 6.8 (2) to Fig. 6.8 (6) calculated masks were plotted on the same images for xNL, xNL-Hi-Re-Sh, Th, Th-Hi-Re-Sh and RA. While the green pixels represents the shadows, the blue ones stands for highlights. The threshold values calculated in synthetic image test were utilized. The xNL masks were not very successful due to noisy image characteristic and reduced threshold value for eight images. On the other hand simple Th mask performed better in all three cases. The additional Hi and Re masks also increased highlight detection performance as in Fig. 6.8 (5). The RA mask, presented in Fig. 6.8 (6), was able to find highlights (blue) but clearly missed the shadows. The same problem was also occurred with the synthetic images. The main cause of this failure was the residual error magnitude difference of highlights and shadows. While the shadows with small errors were included in the consensus set, the highlights with large errors were excluded.

In Fig. 6.9, encoded normals were plotted for (1) None, (2) xNL, (3) xNL-Hi-Re-Sh, (4) Th, (5) Th-Hi-Re-Sh and (6) RA masks. Without any masks, as seen in the first row, smooth normals were reconstructed. As seen in Fig. 6.9 (4) and (5) variation of normals increased with the usage of masks, resulting a more contrast normal image. This improvement came with some side effects. With the usage of the masks, the calculated normals have noisy results at the boundaries of the masks. Especially when the masks are discontinuous and noisy, the resultant normals are also noisy as in 6.9 (2) xNL, 6.9(3) xNL-Hi-Re-Sh and 6.9(6) RA masks. Weighted PS, in the next sections will be focused on the false edges and noisy look of normals.

In Fig. 6.10, calculated normals and albedos were rendered with the same illumination in Fig. 6.8. Similar to encoded normals, the render images of Th mask results have contrast but noisy results.

In Fig. 6.11, render images of normals and hight values were plotted. The view angle is tilted with 45° to show effect of masks on calculated hight values. The shadows around the nose of the viking statue were found with Th and Th-Hi-Re-Sh as seen in bottom Fig. 6.8 (4) and (5). Hence the height of the nose was calculated higher which is closer to the correct height.

6.8 Weighted PS Results

Weighted PS was tested real images with Th mask which was also used with the synthetic images. Eight images with $Z_{source} = 62°$ were used in weighted PS. The resultant normals with and without weighting is shown in Fig. 6.12. The left, Fig. 6.12(a), not weighted normals have false edges at the helmet and the neck of the viking statue where shadows and highlights were masked out. Normals at Fig. 6.12(b), which were calculated with weighted PS, the most of the false edges were smoothed.

The smoothing effect of weighted PS is more obvious in the render images, presented at Fig. 6.13. Both images were rendered with diffuse surface reflection and illuminated from top. The false edges at the helmet, at the left eye and under the beard were smoothed.

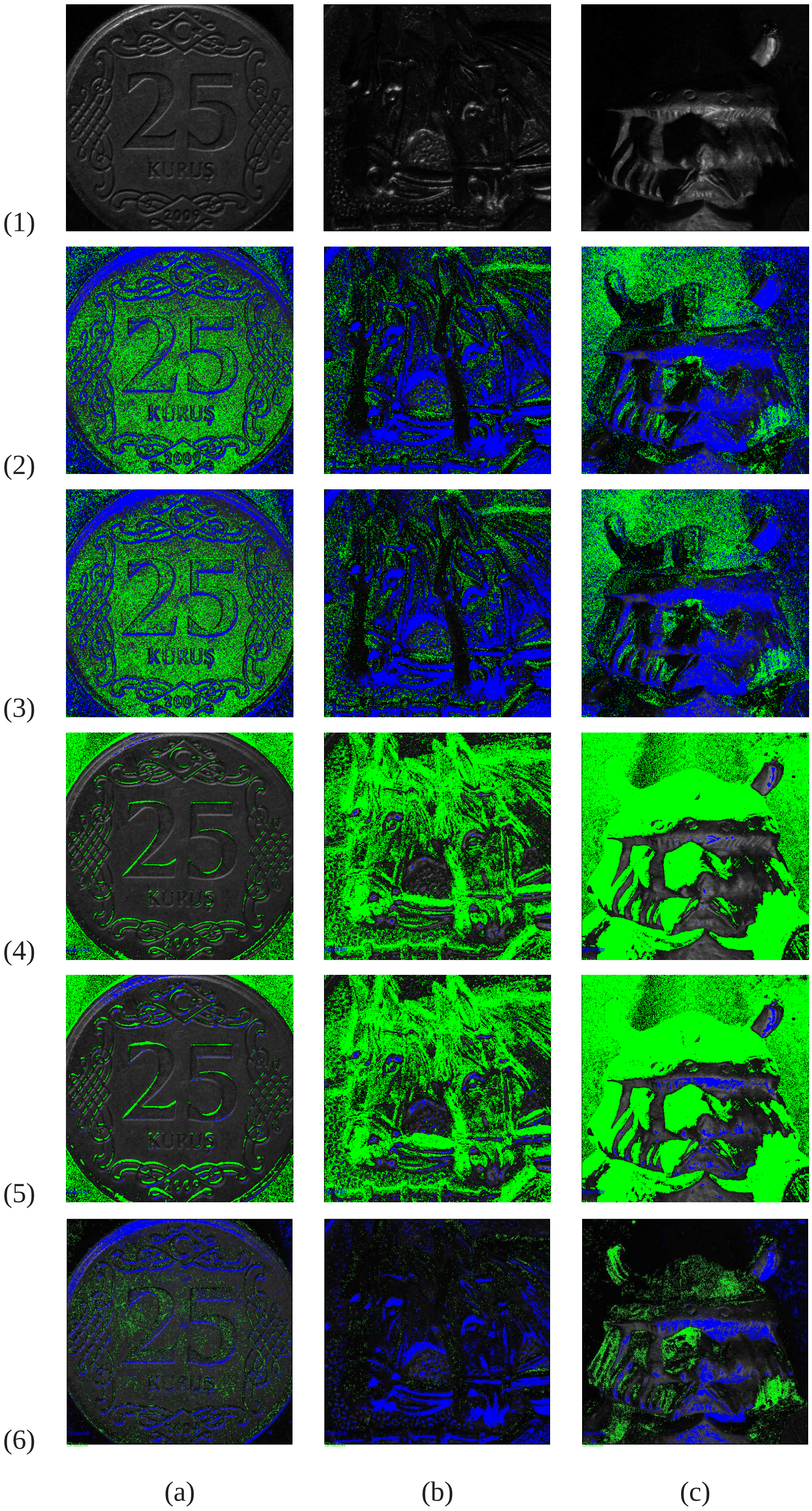


Figure 6.8: (1) Real images with illuminated from right bottom corner with 60° zenith angle. Calculated masks are plotted on images with green for shadows, with blue for highlights for (2) xNL (3) xNL-Hi-Re-Sh (4) Th, (5) Th-Hi-Re-Sh and (6) RA.

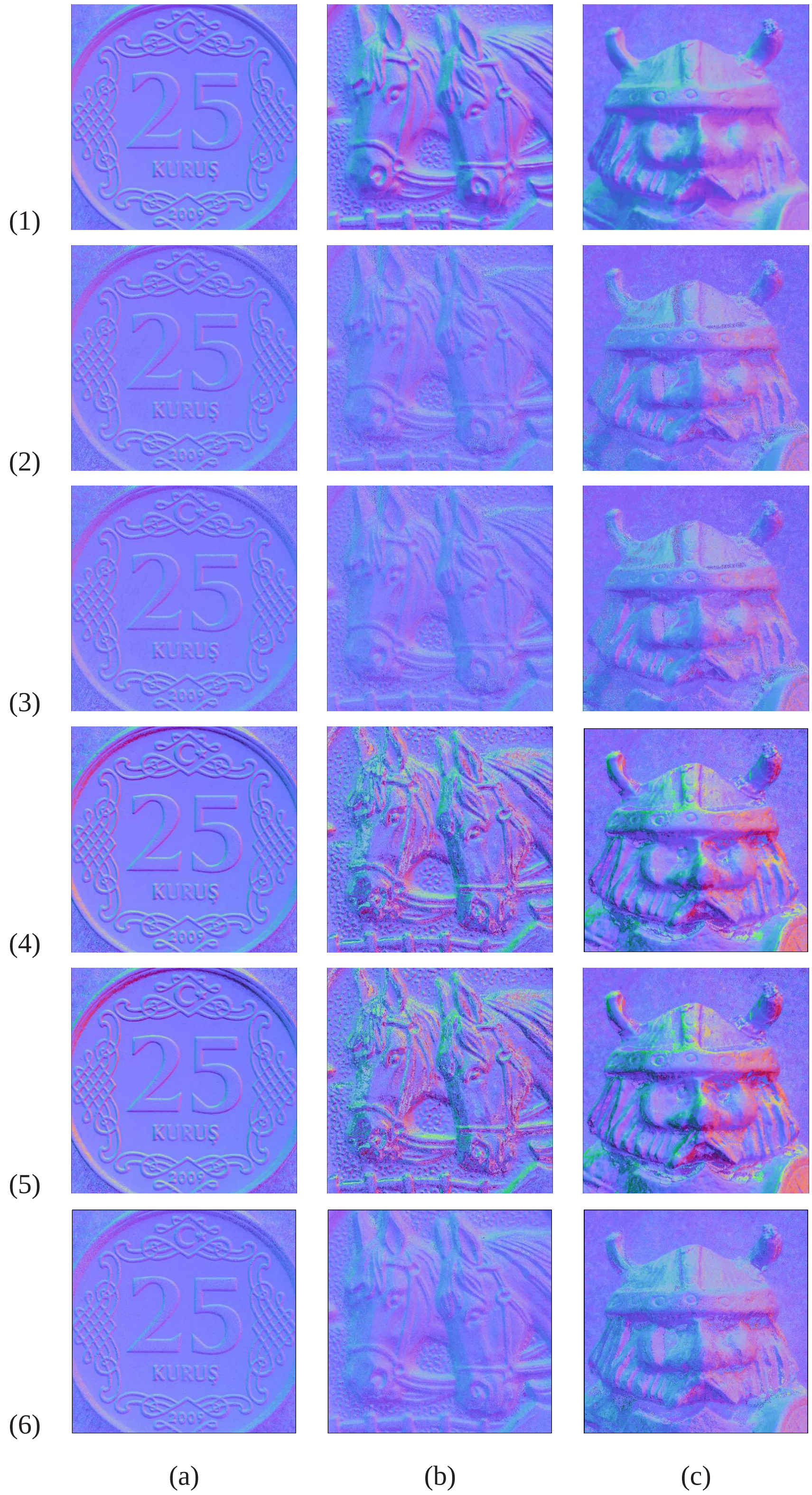
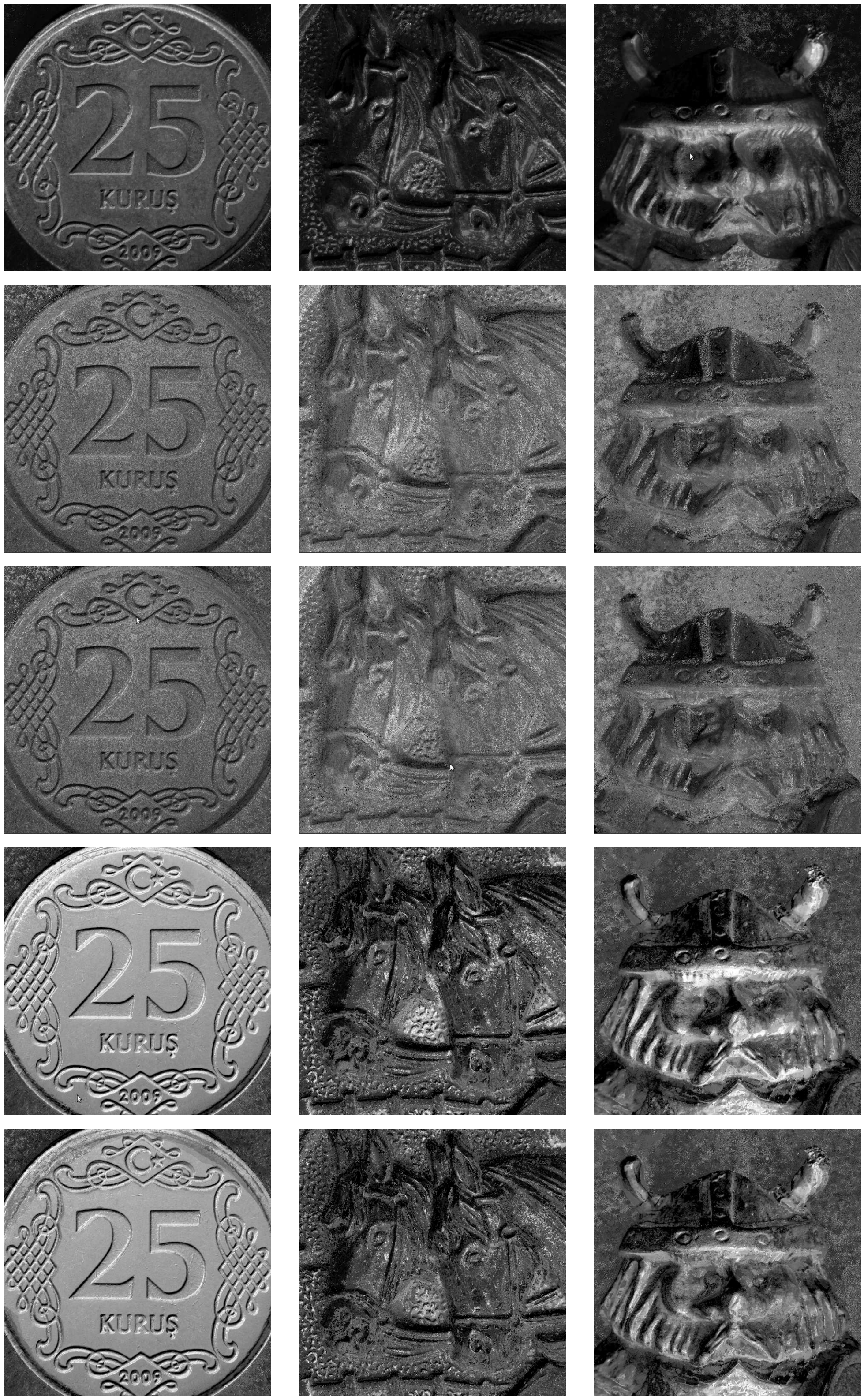


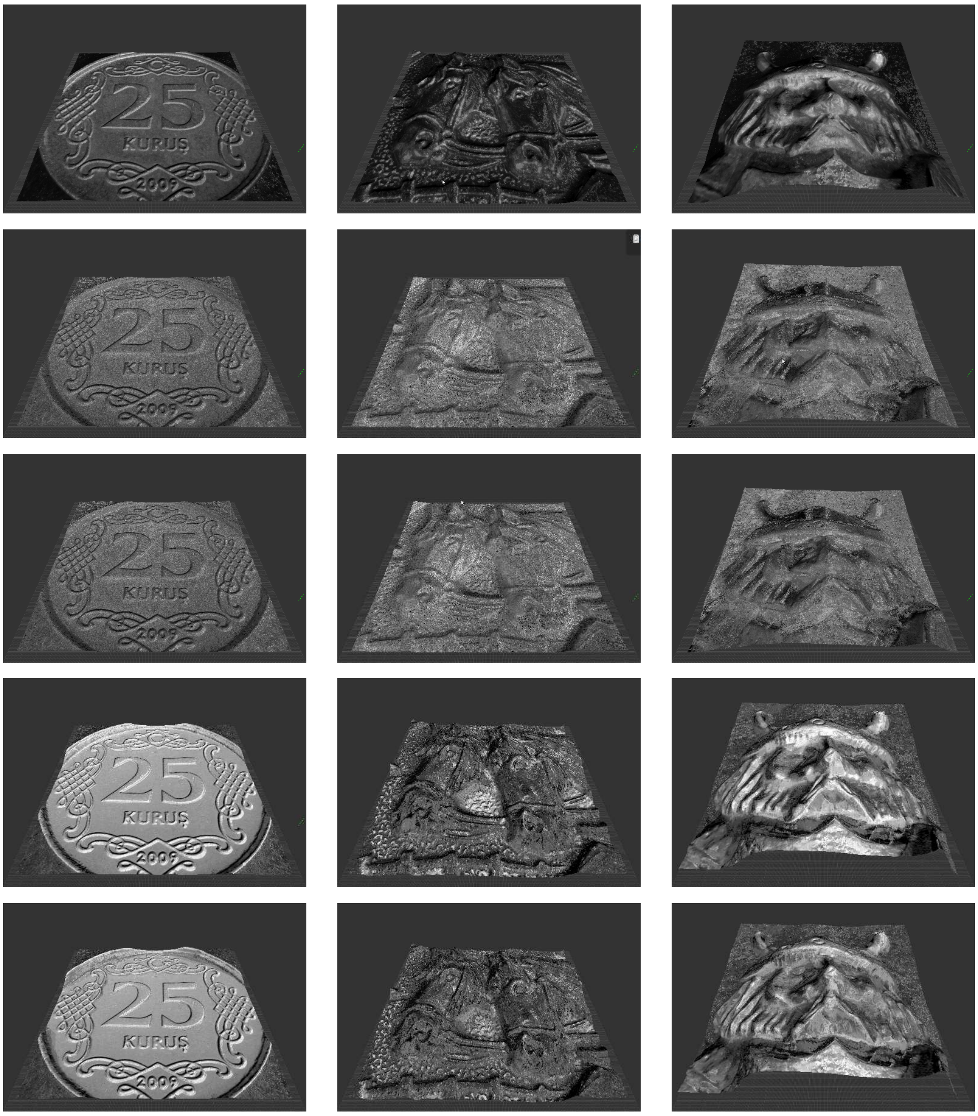
Figure 6.9: (a) Calculated normals for (1) None (2) xNL (3) xNL-Hi-Re-Sh (4) Th and (5) Th-Hi-Re-Sh and (6) RA.



(1) (2) (3) (4) (5)

(a) (b) (c)

Figure 6.10: (Render images for (1) None (2) xNL (3) xNL-Hi-Re-Sh (4) Th and (5) Th-Hi-Re-Sh.



(1) (2) (3) (4) (5)

(a) (b) (c)

Figure 6.11: Render images for (1) None (2) xNL (3) xNL-Hi-Re-Sh (4) Th and (5) Th-Hi-Re-Sh.

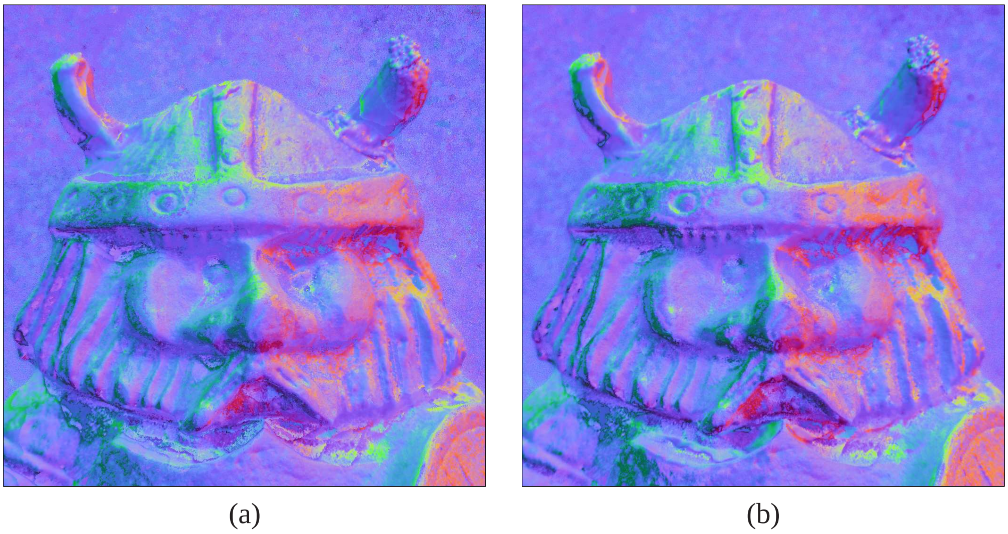


Figure 6.12: Normals of Th mask PS (a) and weighted Th mask PS.

Weighted PS smooths the masks not the generated surface normals. The false edges of the normals can also be smoothed with simple Gaussian filtering operation on the resultant normals. But this smoothing operation will also smooth the details of the reconstructed normals, which is not desired. As seen in Fig. 6.12, the weighted PS preserved the surface details and smoothed only the false edges and noisy look caused by the masks.

6.9 Multi Zenith Results

The synthetic image results presented up to now were also verified with the real images. Five test surfaces, bearing ball, a coin, a relief of two horses, plastic toy face and metal toy face were selected to present both highlights and shadows. 2048x2048 8 bit monochromatic images of these test surfaces were acquired with Prosillica GE2040 camera and Navitar 12x zoom lens. 16 warm white power LEDs with diffusers were placed around the camera with $Z1_{source} = 31°, Z2_{source} = 62°$ and polar distances being 45° as shown in Fig. 6.2.

5 test objects were used in the real image tests as seen in Fig. 6.14. Plastic and metallic toy faces were good examples for diffuse and specular surfaces. Coin, horses, bearing ball and metal toy face were selected to have changing amount shadow and highlights on their surfaces.

Figure 6.13: Render images of normals of Th mask PS (a) and weighted Th mask PS.

Figure 6.14: The real objects used in tests. Left to right, plastic toy face, metal toy face, coin 25 kurus, bearing ball and horses.

Fig. 6.15 displays the test images of the metal toy face after the calibration process that linearized CCD gain characteristics. The images at left most two columns are taken with $Z1_{source} = 31°$, and the ones at right most two columns are taken with $Z2_{source} = 62°$ light sources. Wide highlighted regions can be seen at the left ones. The right ones suffer from shadows, especially around the nose of the metal toy face.

The weights used in WPS are presented in Fig. 6.16. The sum of two weight images adds up to 1.0 all over the image. The white intensity values indicate heavy weights that it's normal vector will influence more to the final normal vector. The normals of the slanted regions, for example, around the nose, have less error when calculated with $Z1_{source} = 31°$ lights, hence $Z2_{source} = 62°$ lights generated many shadows at these regions. The weights around the nose were calculated as desired, i.e. reduced NE, as in Fig. 6.16.

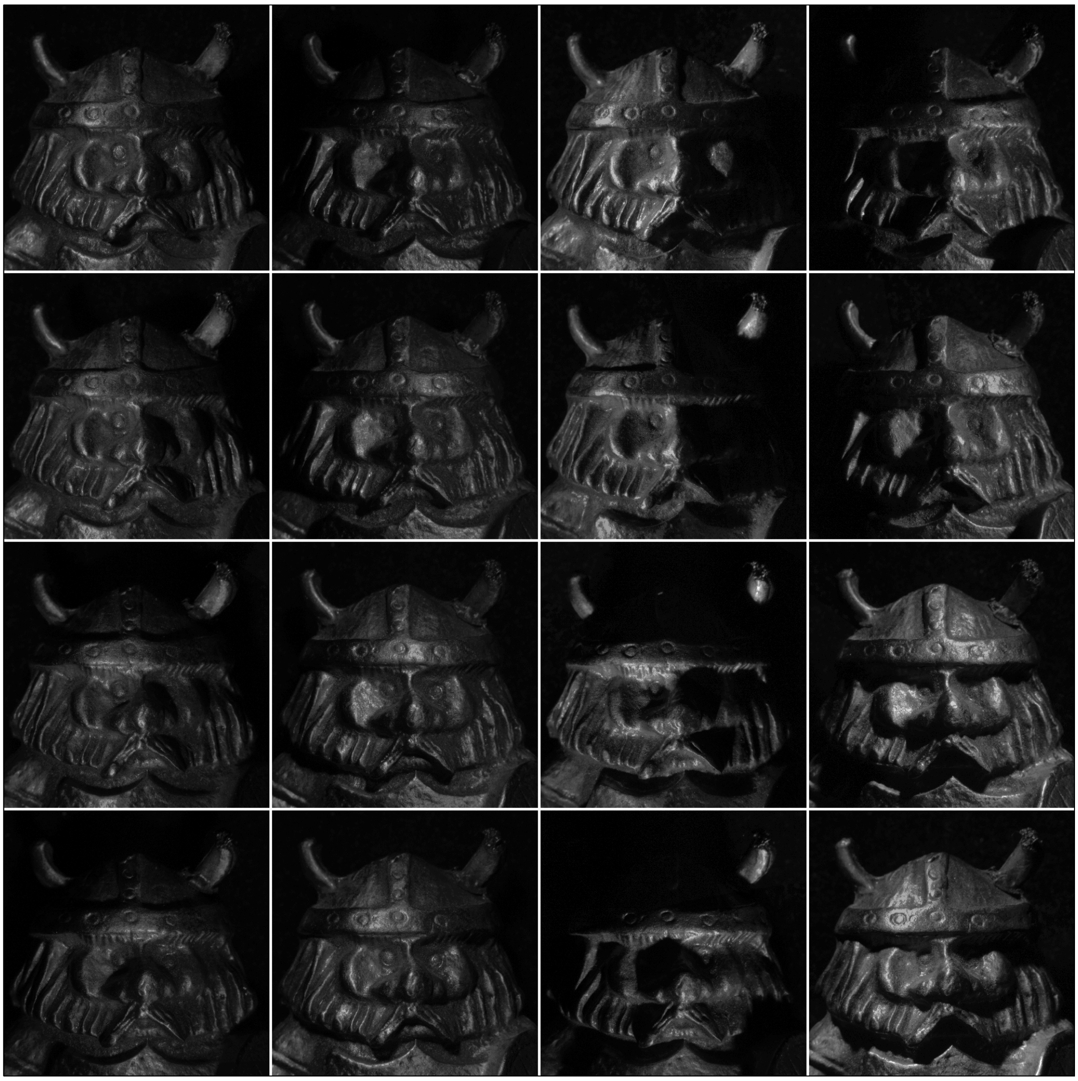
In Fig. 6.17, the intensity error images of PS and WPS solution are shown. The IE of the slanted regions decreased with the proposed WPS.

Fig. 6.18 displays the color coded normal vectors of PS and WPS solutions. While the normals calculated with PS (see Fig. 6.18a) have a limited range, the normals of WPS method (see Fig. 6.18b) ranged larger.

Lastly, real and render images of the metal toy face is presented in Fig. 6.19 from different angles. Left most images (see Fig. 6.19a) were acquired with an ordinary digital camera. Fig. 6.19b and 6.19c are render images of PS and WPS. As seen the Fig. 6.19b bottom image, the nose of the metal toy face could not be reconstructed with PS. It was much smoother than the real image. But this error reduced in Fig. 6.19c indicating that WPS worked better in the presence of shadows.

Table 6.1 lists the percent intensity errors (IE) of all test surfaces with 8 and 16 images configuration. In both cases, IE is reduced significantly with the WPS method, compared to the ordinary PS method. The minimum IE was achieved with only residual error (e) weights. This is because of the similarity of the definition of IE and e. Other weights disturbed the results with additional informations other than e. But from the synthetic image results, it was shown that $eZ_{source}Z_{surface}$ weights have the better NE performance.

Table 6.2 presents the average elapsed time of the PS and WPS with 8 and 16 lights configurations. The computation cost of WPS is about 3 times of the ordinary PS. The computational



$Z1_{source} = 31°$ $Z2_{source} = 62°$

Figure 6.15: 16 calibrated images of metal toy face used in *PS* tests.

(a) (b)

Figure 6.16: Weights of metal toy face (a) $Z1_{source} = 31°$ and (b) $Z2_{source} = 62°$.

(a) (b)

Figure 6.17: Intensity error images of metal toy face with (a) *PS* method and (b) *WPS* method. White pixels have IE of 20% of image dynamic range.

(a) (b)

Figure 6.18: Normal vectors of the metal toy face with (a) PS method and (b) WPS method.

(a) (b) (c)

Figure 6.19: (a) real (b) PS render and (c) WPS render images of the metal toy face from front (top) and from right (bottom). The nose was calculated as smooth peak in PS. But this error is reduced in WPS.

Table 6.1: Intensity Error (% of *None*)

	bearing ball	**coin 25 kurus**	**horses**	**metal toy face**	**plastic toy face**	**Overall**
			Average of IE light_30_4_60_4			
None	100.0	100.0	100.0	100.0	100.0	100.0
e	49.8	55.5	38.8	61.1	54.4	50.5
eZ_{source}	49.5	55.7	38.6	65.1	58.9	51.9
$eZ_{surface}$	53.3	52.9	41.3	68.3	72.9	56.0
$eZ_{source}Z_{surface}$	52.2	52.3	40.1	70.1	76.7	56.3
			Average of IE light_30_8_60_8			
None	100.0	100.0	100.0	100.0	100.0	100.0
e	64.7	92.0	62.0	70.6	59.9	68.4
eZ_{source}	63.6	90.2	60.5	77.6	65.9	70.0
$eZ_{surface}$	68.5	82.7	62.3	81.1	87.9	75.0
$eZ_{source}Z_{surface}$	67.2	80.5	60.5	86.2	94.0	76.1

cost increase is mainly caused by the weight calculation. This cost is still much lower than non-linear *PS* methods.

Table 6.2: Average elapsed time (Normalized w.r.t $2Z$).

Weight	**light_30_4_60_4**	**light_30_8_60_8**
None	1.00	1.00
e	2.89	2.65
eZ_{source}	2.91	2.62
$eZ_{surface}$	3.06	2.68
$eZ_{source}Z_{surface}$	3.06	2.69

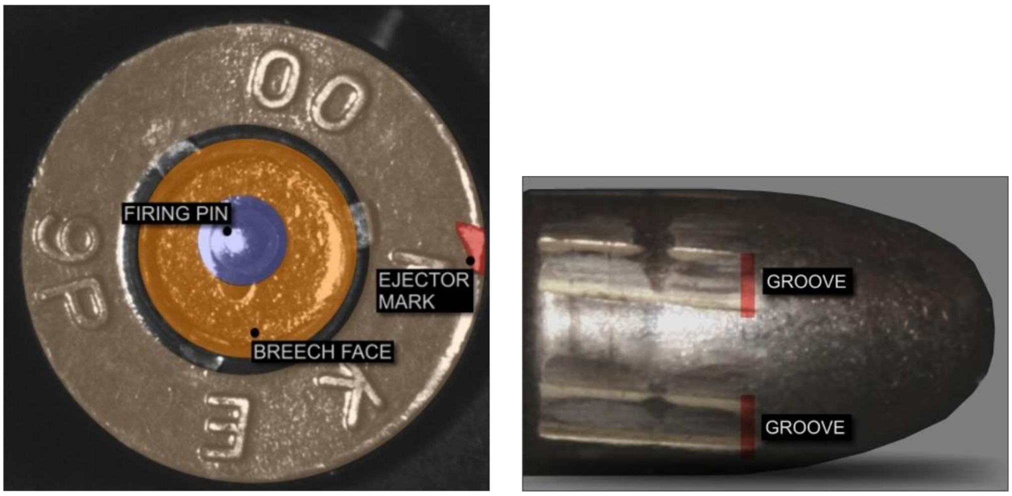
CHAPTER 7

FIREARM EVIDENCES

Firearm identification is the matching problem of the firearm evidences from the striation and impressed marks left by the firearm [13]. Traditionally, the matching operation is executed with a microscope by the trained experts. Each couple of evidences are placed under the microscope side by side and are investigated by eye. The side by side comparison operation have to be done for each new evidence with every one in the database outnumbering thousands. Automated firearm identification systems helps experts to overcome this cumbersome task of comparisons.

The firearm evidences are mainly the cartridge cases and the bullets as seen in Fig 7.1. The cartridge case is full of gun powder and the bullet is attached in front of it. The mechanism of the gun, the firing pin, hits on the soft metal part of the cartridge case, the breech face to ignite the gun powder inside the case. The powder burns very fast creating very high heat and pressure. The pressure forces the bullet to exit from the barrel with a speed of $300m/sec$ for an ordinary firearm. In modern firearms, inside the barrel, there are helical grooves that spins the bullet around its motion direction. Later, the cartridge case is ejected from gun automatically. These firing process deforms the metallic surface of the cartridge case and the bullet. The deformation is firearm specific, so that surface topology of the firearm is impressed and striated on the evidences.

The cartridge cases mostly have impression marks on its bottom. The firearm identification system, acquires the images of the bottom of the cartridge case, shown in Fig. 7.1(a). There are three important regions at the bottom of the case that are investigated by experts separately. The ejector mark is created by the firearm ejector pin that pushes the cartridge case out of the gun after firing. This mark usually has a small area and is difficult to match but it indicates



(a) (b)

Figure 7.1: The regions of a cartridge case (a) and a bullet (b).

Figure 7.2: Test spend cartridge cases with various caliber type and material

the orientation of the cartridge case inside the gun. The breech face is the soft part of the case where gun's firing pin strikes on it and create firing pin mark. These marks on the deformable thin metal contains majority of the characteristic marks of the gun.

The bullets are usually composed of two parts the heavy soft core that is made of lead and the harder brass shell. This type of bullet is called "full metal jacket". While bullet travels inside the barrel, the helical grooves of the barrel spin the bullet and create strained marks on the bullet. Even after hitting to the target, these groove marks can be used to identify the firing gun. Usually there are four to six groove marks on a bullet as seen in Fig. 7.1(b).

The cartridge cases used in the tests are presented in Fig. 7.2. The cartridge cases with different metals having a variety of colors and reflection properties are selected. Also different

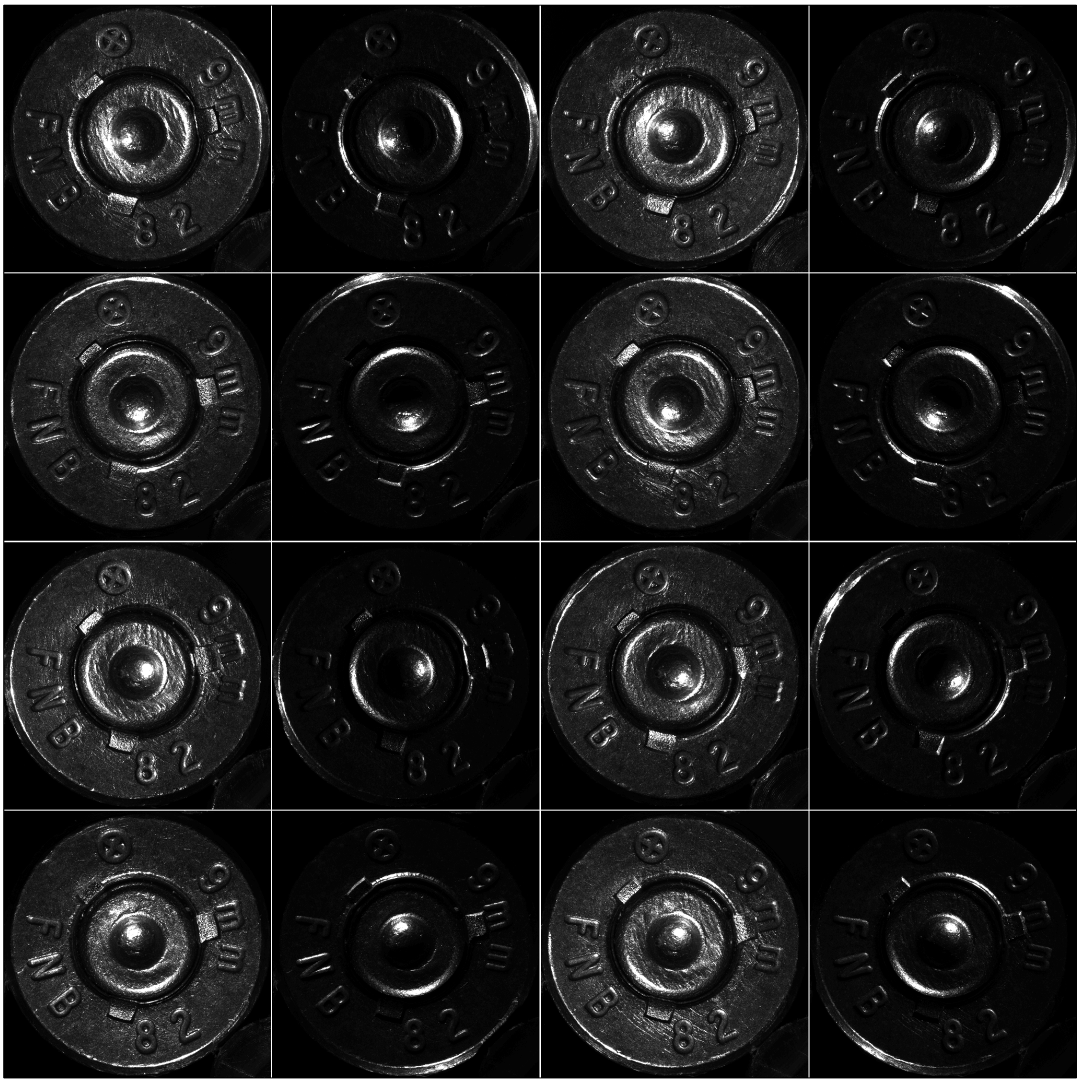


Figure 7.3: 16 raw images of cartridge case 3.

caliber types used by guns and rifles exist in the test set. The 3^{rd} and 5^{th} from the left are "sister cartridge cases" that have been fired from the same gun. The results of these two cases will be presented in this chapter.

The 16 raw images acquired by the BALISTIKA2010 System are presented in Fig. 7.3. The images have $2048x2048$ pixels with 8 bit gray scale intensity value. The field of view is $10mmx10mm$. The spatial resolution of the acquired images were $4.9\mu m$. The camera's exposure was set to the same value for all images. This common exposure value was selected considering the mean value of all intensity values to be around 75. The dark ones are illuminated by $Z_{source} = 62°$ light sources and bright ones are by $Z_{source} = 31°$. Highlights and

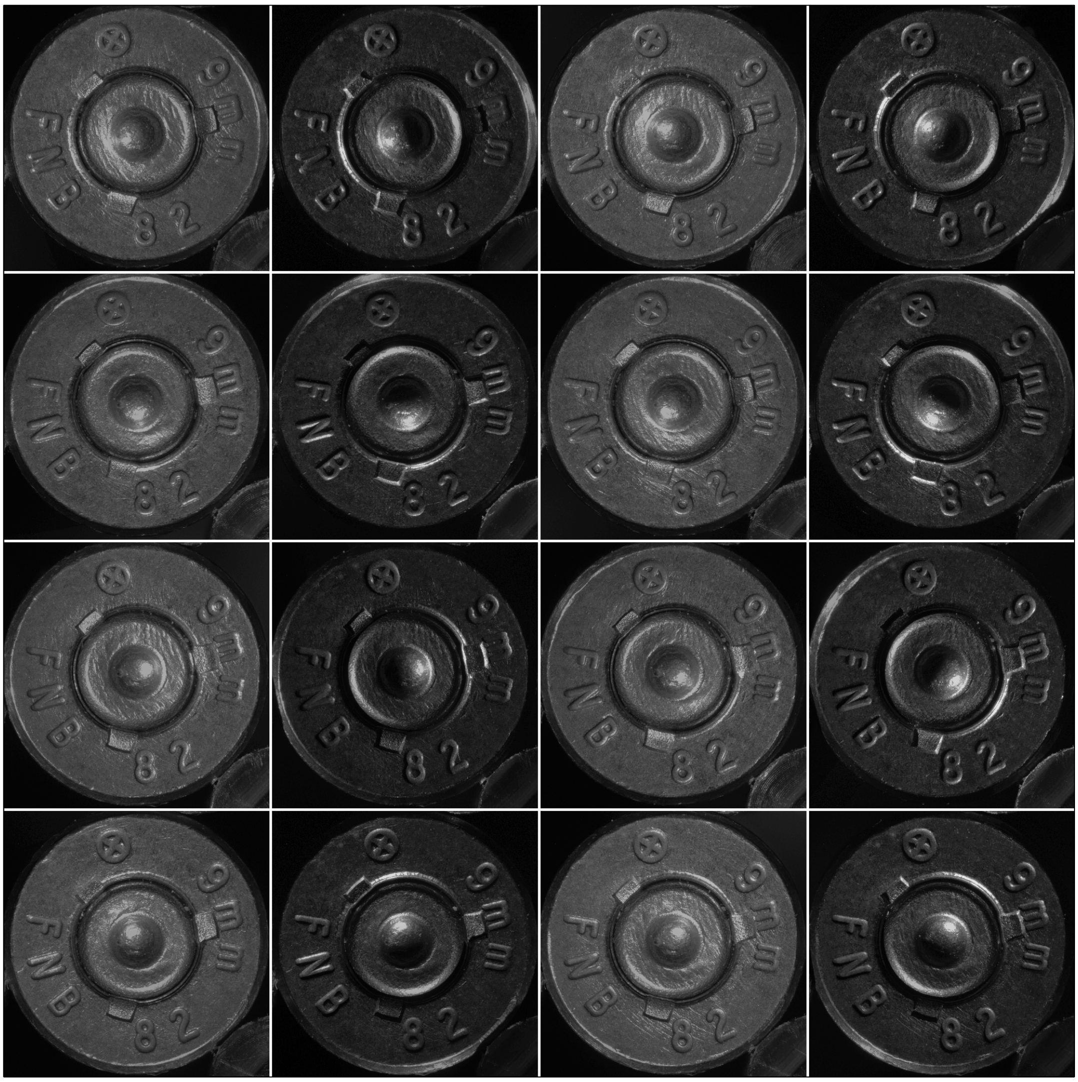


Figure 7.4: The calibrated images of cartridge case 3.

shadows were present in this ordinary cartridge case.

Fig. 7.4 displays the resultant images of the calibration process. The 8 bit images are converted to floating point images before calibration. The calibrated images that will be used in PS, are also in floating point intensity images.

The masks that yielded best results in the synthetic images were applied to the cartridge cases as seen in Fig. 7.5. The green pixels are shadows and the blues are the highlights. Although the outside of the cartridge case was marked as shadows, these regions are discarded in the next steps. The images with $Z_{source} = 31°$ mainly have highlights indicated with blue. On

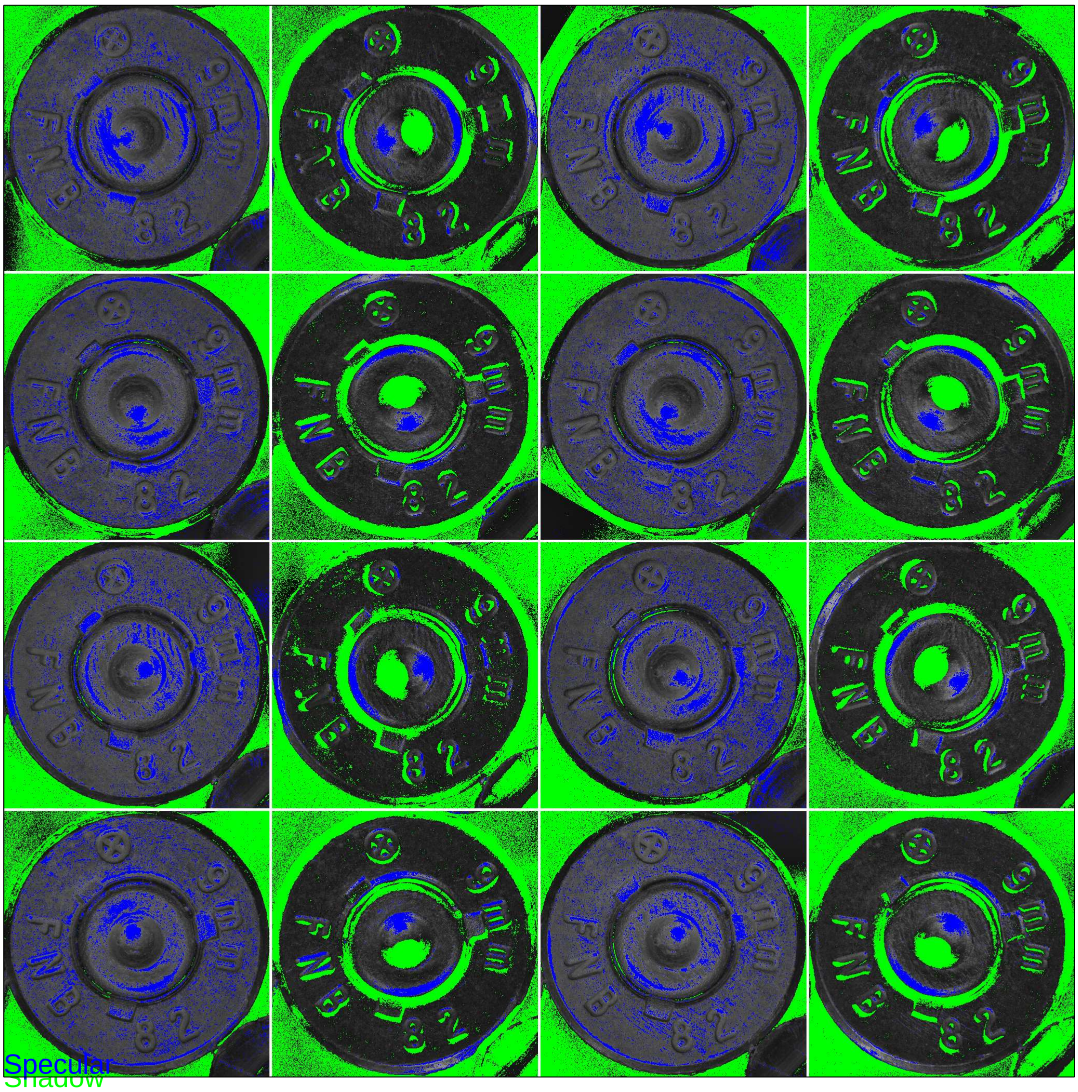
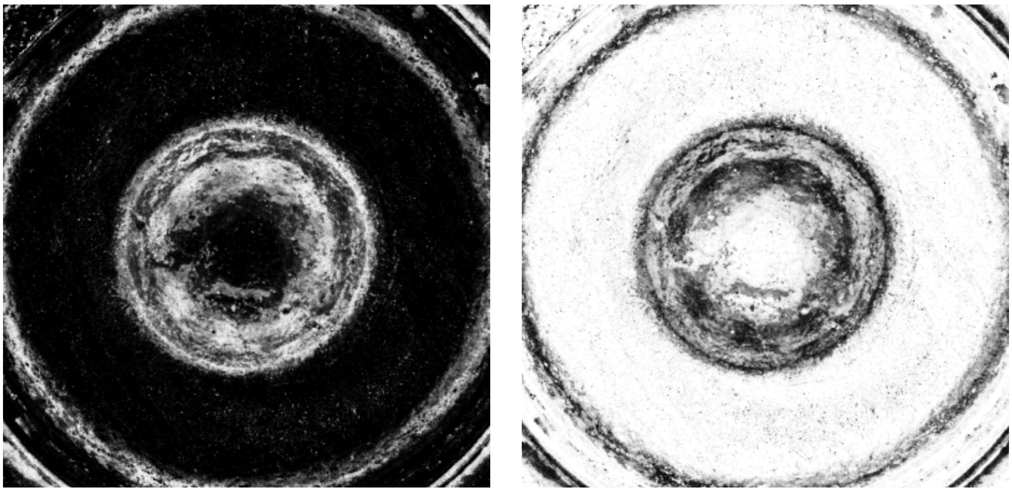


Figure 7.5: The Th-Hi-Re-Sh masks calculated on the cartridge case 3.



(a) (b)

Figure 7.6: $Z_{source} = 31°$ (a) and $Z_{source} = 62°$ (b) weights of the firing pin of the cartridge case 3.

the other hand, the images with $Z_{source} = 62°$ have both highlights and shadows on the case especially on the critical firing pin mark. Considering the real height map of the cartridge case, the highlights and shadows masks are calculated successfully. Note that many pixels from the concave firing pin was masked out, leaving few pixels to be used in PS. These masks may be used both as a binary mask or can be weighted to obtain smooth mask edges.

The weighted normal PS was also tested on these cartridge cases. Fig. 7.6 presents the calculated weights for each normals with (a) $Z_{source} = 31°$ and (b) $Z_{source} = 62°$ where white pixels indicate high weight values. The most of the resultant normals were calculated with images illuminated by $Z_{source} = 62°$ light sources. On the other hand, when the shadows occurred at the walls of the firing pin, $Z_{source} = 31°$, the light sources have greater influence on the resultant normals. These masks yielded satisfactory results from the synthetic image results.

The render images of two reconstructed sister cartridge cases, cartridge case 3 (left) and cartridge case 5 (right), are shown in Fig. 7.7. The top images are tilted 3D view of the cases, and bottom images are breech face close view. The orientation of the cases were matched as much as possible so that the characteristic surface deformations can be seen in the place in the images. The horizontal short line close to the middle of the firing pin center and diagonal long lines on the breech face are characteristic marks of this gun as commented by ballistic

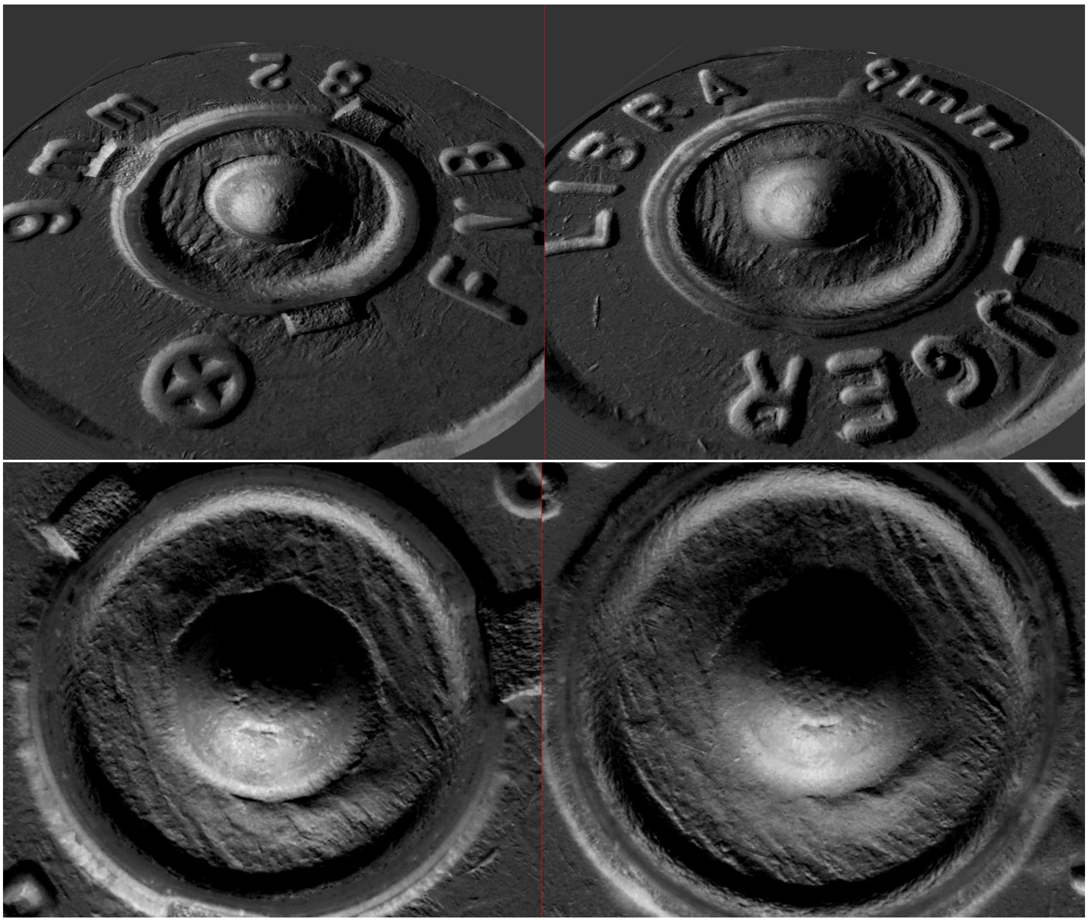
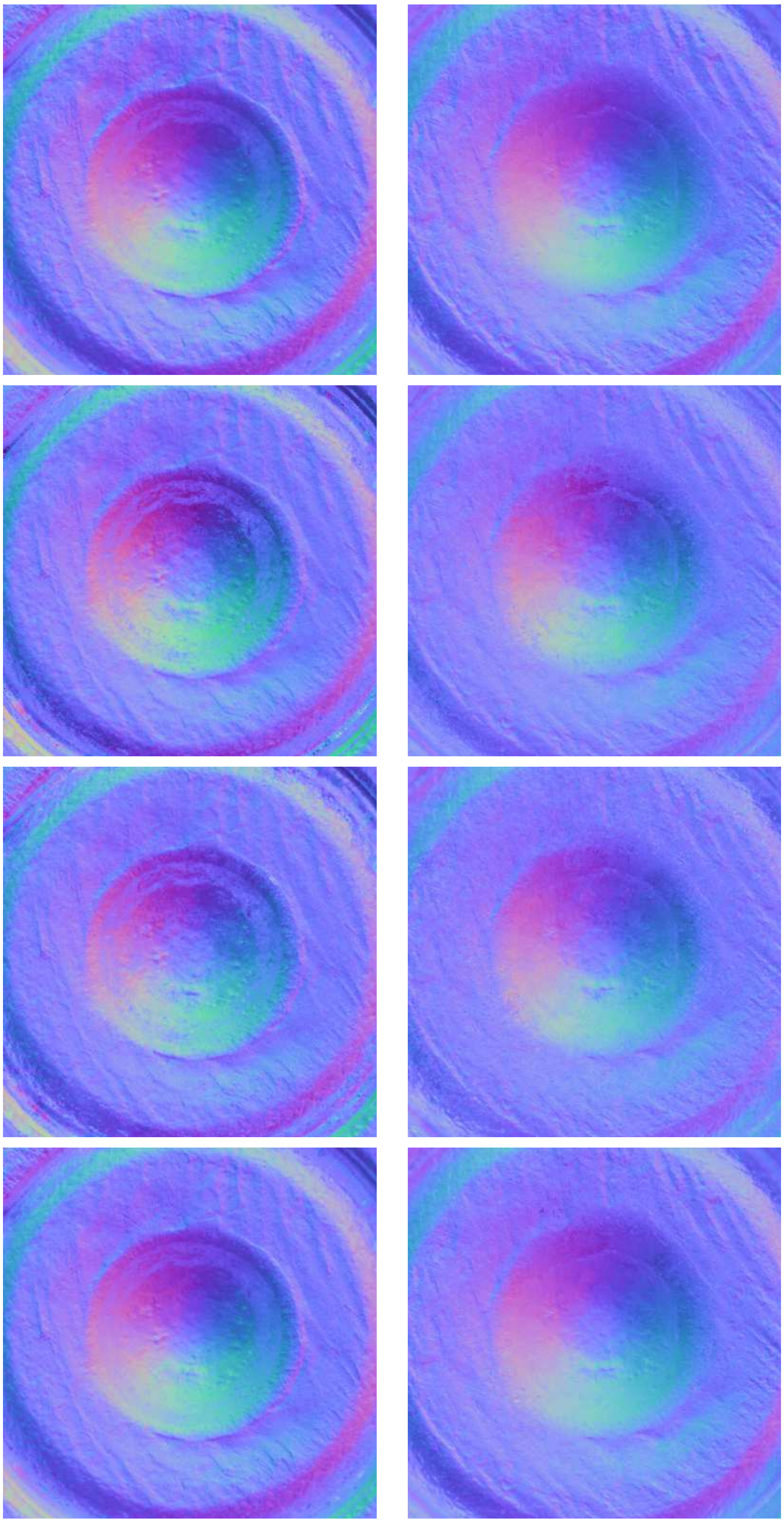


Figure 7.7: Render images of cartridge case 3 (left) and cartridge case 5 (right)

experts. They also suggested that the number of characteristic marks are very good that these two can be easily identified as sisters.

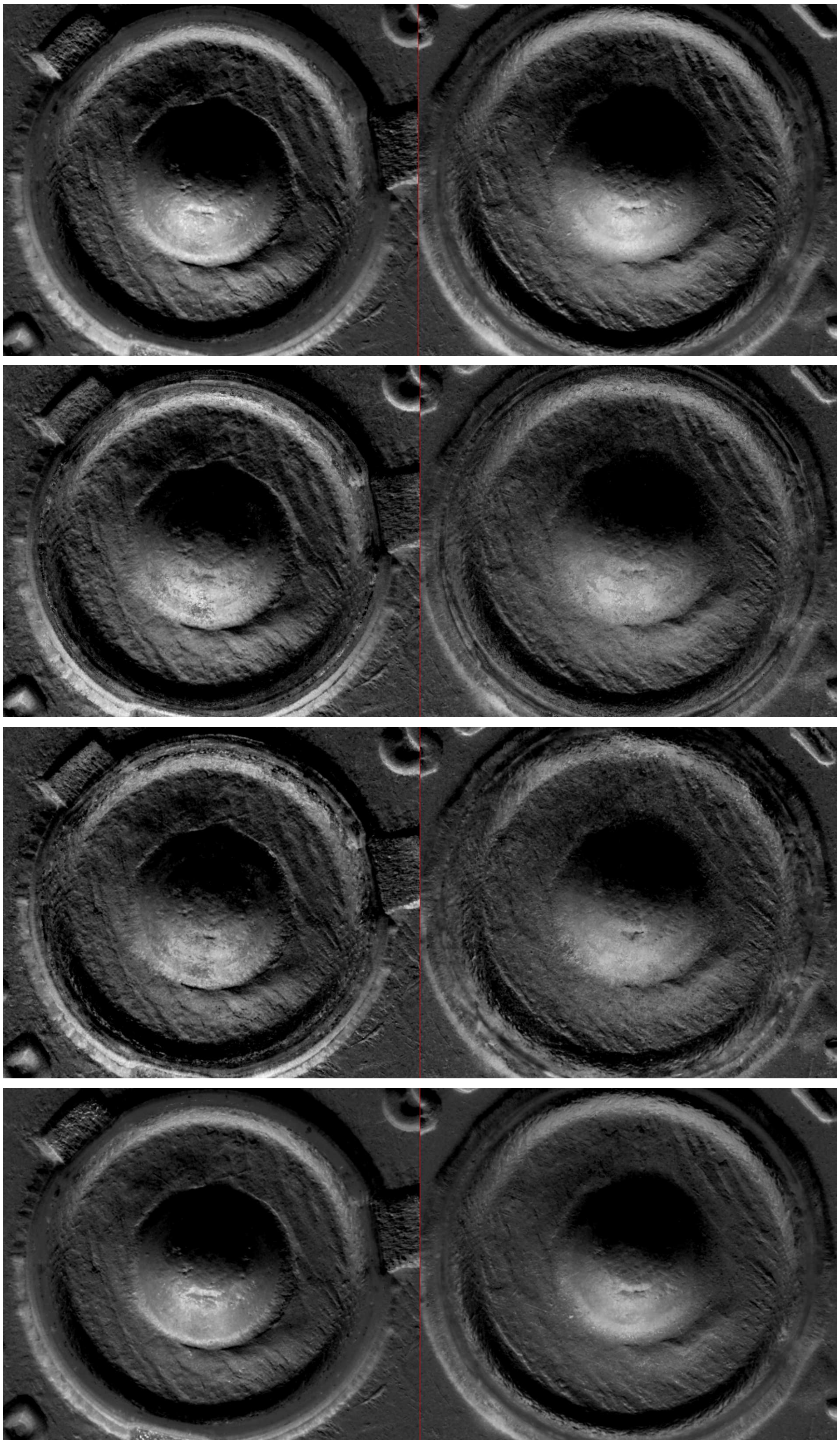
The resultant normals of cartridge case 3 (a) and case 5 (b), generated using different methods are shown in Fig. 7.8. The sisters are placed side by side to present the matched characteristic marks for each PS method. The first row of normals are generated without any masks. The second row, Fig. 7.8(2), presents the binary Th-Hi-Re-Sh masks results where more details at the walls of the firing pin are visible. Some of these are mask false edges. In Fig. 7.8(3), the false edges are removed when weighted PS compared to Fig. 7.8(2). The bottom row, the results weighted normal PS, were the smoothest normals among all.

Lastly, the render images are displayed in Fig. 7.9. These images are captured from the screens of the BALISTIKA2010 Identification System. For each row, two sisters were identified by the ballistic experts easily.



(1)
(2)
(3)
(4)
(a) (b)

Figure 7.8: Normals of the cartridge case 3 (a) and 5 (b) for (1) None, (2) Th-Hi-Re-Sh (3) weighted Th-Hi-Re-Sh (4) and weighted normal PS.



(1)

(2)

(3)

(4)

Figure 7.9: Render images of the cartridge case 3 (a) and 5 (b) for (1) None, (2) Th-Hi-Re-Sh (3) weighted Th-Hi-Re-Sh (4) and weighted normal PS.

CHAPTER 8

CONCLUSION

8.1 Summary

Following tasks were completed in this thesis;

- Masked PS methods in the literature and simple new masking methods like threshold mask were classified and implemented on a unified framework.
- The weighted PS method, using weighted least square estimation (WLSE), is suggested and developed to eliminate false edges created by the masks.
- The calibration processes were investigated. The disturbances due to close light sources were removed by image calibrations.
- Different illumination configurations including double zenith illumination configuration, were suggested and tested.
- Double zenith illumination configuration results were further improved by the weighted normal PS.
- The synthetic image generation codes were implemented. Controlled tests were conducted to fine tune the masks.
- The methods were tested on real objects with varying characteristics as well as the firearm evidences.

8.2 Discussion

Firstly, this thesis proposed the unified PS framework and weighted PS, which is capable of utilization of any implemented mask combinations. The unified PS has the key role to make a fair comparison among masks. The unified PS was justified by achieving similar enhancements with the previous works in literature. Novel weighted PS is integrated to the unified PS framework to weight masks and remove side effects of masking operations.

In addition to some masks proposed earlier in the literature, new simple masks were also proposed. Also, widely used NL method is further extended from 6 images to any number of images. The masking threshold values were optimized with the synthetic images for various image counts. With the test configuration stated in this thesis, the normal error was reduced from 14.78^o to 8.63^o (44% improvement) with simple fast working Th mask. This simple yet powerful masking method was combined with highlight, reflection and shadow masks to achieve the best improvement.

Without masks, increasing image count slightly reduces the normal error because new images bring more highlights and shadows and without masks the improvement is limited. One of the most significant findings to emerge from this thesis is that with simple masks, normal errors can be further decreased with a reasonable computation cost.

This thesis showed that the normal errors increased faster with the masked PS methods than no mask PS, indicating that masked PS is more sensitive to noise. Among all the masks implemented in this thesis, NL mask was the most noise sensitive method with increasing image count. This is caused by the small threshold value used for many images. The threshold value have to be reduced for increasing image count to have a discriminating mask with the closer light source illumination directions.

The experiments on the synthetic image test were also tested on objects with metallic surfaces. The real image results supported earlier conclusions derived from synthetic cases. In real object cases threshold (TH) and threshold highlight reflection and shadow (Th-Hi-Re-Sh) mask combination produced the most detailed normals and height values.

The masked PS methods have a side effect that disturbs the results. The boundary patterns of the masks were transfered to the normals. This false edge disturbance was caused by the change in the input image combinations at the neighboring pixels. If the calculated masks do not have well defined boundaries, this results in noisy normals.

The false edges at the masked boundaries were handled with the novel weighted PS method, using weighted least square estimation. Instead of binary masks, continuous weights were used in this method. The weight of each pixel was inversely proportional to the square of the intensity error. Later, the mask boundaries were dilated and smoothed on the weight images. The false edges and noisy normals caused by discontinuous masks were also smoothed. This method only filters out the false edges and noisy normals. The rest of the details at the normals were not affected.

Secondly, a better illumination configuration considering not only the image noise, but also highlights and shadows was investigated. Realistic highlights and shadows were rendered on synthetic surfaces. It is experimentally shown that there is no single optimal illumination configuration for every surface with highlights and shadows. However, placing the light sources on two circles around the camera with different zenith angles reduced the resultant normal error. Also with the double zenith sources configuration, NE was more evenly distributed over the surface which may be a desired feature. For a *PS* setup to reconstruct a variety of shapes with highlights and shadows, results suggests to place the light sources on two circles with 30^o and 60^o zenith angles if more than 8 of sources are available.

With the double zenith illumination configuration, weighted normal PS that uses double zenith light sources, was proposed. In this method the light sources are grouped with their zenith angles and two subsets are solved exclusively. The two resultant normal vectors for each pixel was weighted with respect to their estimated errors. Different weighting methods were considered. The weighted normal PS improved normal error more than 30% compared to single zenith light sources and ordinary PS. The cost of using weighted normal PS is the overhead of weight computation and weighting operation. With the same volume of input data, the elapsed time of weighted normal PS is about 3 times of the PS on the average. This cost is still much lower than non-linear PS methods using iterative error minimizations.

The results of the firearm evidences were also presented in this thesis. The breech face of the cartridge case is a planar region with shallow characteristics marks on it. These planar parts usually do not create highlights or shadows. Most of the problems occurred at the concave firing pin mark, which is a very important region including most of the characteristics marks of the cartridge. The mask combination with best synthetic image performance was used. The masks and the weights were calculated properly. The 3D shapes were investigated by ballistic examiners and concluded that all the proposed methods create easy to identify results. Generally the ordinary PS was enough to reconstruct a discriminative normals, since surface topology is simple to solve for PS. On the other hand, some marks at the firing pin walls can be better reconstructed by masks.

Both of the synthetic and the real images were have 8 bit intensity values. These integer intensity values were converted to 32 bit floating point values normalized to have maximum value one. The resultant normal vectors were also have 32 bit floating point values. The calculated normal maps have the same pixel resolution with the input images, $2048x2048$. For the test objects $7.32\mu m$ spatial resolution normal and height maps were calculated. The cartridge cases and bullet grooves were placed at the scene to have the best resolution. For cartridge cases, spatial resolution is around $4.9\mu m$, and for bullet grooves spatial resolution can be reduced down to $1.0\mu m$.

For the 16 synthetic images with $512x512$ pixel resolution, the maximum time elapsed for the execution of the masked PS algorithm was $32secs$. The best performing masked PS calculated results in $20secs$. All of the tests were conducted on a regular PC and Matlab. The native implementations (e.g. C++) of these algorithms were expected to work faster than Matlab scripts.

8.3 Future Works

The concave shape of the firing pin causes another important problem, which is secondary reflection. The illuminated region reflects light inside the firing pin creating secondary reflections. The secondary illumination is not solved in this thesis and remained as future work. A mask working similar to the ray tracing algorithm in the computer graphics may be implemented. The ray tracing can be executed quickly up to two or three reflections. The back reflections to the camera may be identified as secondary reflections. Since the removal of these back reflecting pixels will change the normals and surface topology, the ray tracing should be executed again. Hence, secondary reflection mask will be an iterative normal mask in unified PS.

Another improvement can be achieved by the implementation of multi view PS, that is the fusion of binocular stereo and PS. This method would be very useful when creating 3D shape of complex topologies like bullets. Detailed PS results can be fused to form a single real 3D shape of the bullet.

REFERENCES

[1] F. Solomon and K. Ikeuchi, "Extracting the shape and roughness of specular lobe objects using four light photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 449–454, Apr. 1996.

[2] S. Barsky and M. 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*, vol. 25, no. 10, pp. 1239–1252, 2003.

[3] J. Sun, M. Smith, L. Smith, S. Midha, and J. Bamber, "Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities," *Image and Vision Computing*, vol. 25, no. 7, pp. 1050–1057, 2007.

[4] V. Argyriou and M. Petrou, "Recursive photometric stereo when multiple shadows and highlights are present," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, no. 1, pp. 1–6, Jun. 2008.

[5] B. K. Horn, "Shape From Shading," Ph.D. dissertation, 1970.

[6] R. Zhang, P. Tsai, J. Cryer, and M. Shah, "Analysis of shape from shading techniques," in *1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94.* IEEE Comput. Soc. Press, 1994, pp. 377–384.

[7] R. Klette, R. Kozera, and K. Schluns, "Shape from shading and photometric stereo methods," *Handbook of computer vision and applications*, vol. 2, no. May, pp. 532–590, 1999.

[8] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3D geometry," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, p. 543, 2005.

[9] I. Horovitz and N. Kiryati, "Depth from gradient fields and control points: bias correction in photometric stereo," *Image and Vision Computing*, vol. 22, no. 9, pp. 681–694, Aug. 2004.

[10] Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: how different really are they?" *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, vol. 39, no. 2, pp. 1784–1786, 2000.

[11] M. Asif and a.S. Malik, "3D shape recovery from image defocus using wavelet analysis," *IEEE International Conference on Image Processing 2005*, vol. 2, no. 5, pp. I–1025, 2005.

[12] R. J. Woodham, "Photometric stereo," Ph.D. dissertation, MIT, 1978.

[13] U. Sakarya, U. M. Leloglu, and E. Tunali, “Three-dimensional surface reconstruction for cartridge cases using photometric stereo.” pp. 209–17, Mar. 2008.

[14] M. S. Bonfanti and J. de Kinder, “The influence of manufacturing processes on the identification of bullets and cartridge cases–a review of the literature.” *Science & justice : journal of the Forensic Science Society*, vol. 39, no. 1, pp. 3–10, 1999.

[15] C. Smith and J. Cross, “Optical imaging techniques for ballistics specimens to identify firearms,” *Proceedings The Institute of Electrical and Electronics Engineers. 29th Annual 1995 International Carnahan Conference on Security Technology*, pp. 275–289, 1995.

[16] H. Tagare and R. DeFigueiredo, “A theory of photometric stereo for a class of diffusenon-Lambertian surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, pp. 133–152, 1991.

[17] E. Coleman and R. Jain, “Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry,” *Computer Graphics and Image Processing*, vol. 18, no. 4, pp. 309–328, Apr. 1982.

[18] S. Buyukatalay, O. Birgul, and U. Halici, “Surface Reconstruction from Multiple Images Filtering Non-Lambert Regions,” in *ICIP*, no. 1. Ieee, Nov. 2009, pp. 533–536.

[19] Y. Mukaigawa, Y. Ishii, and T. Shakunaga, “Analysis of photometric factors based on photometric linearization.” *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 24, no. 10, pp. 3326–34, Oct. 2007.

[20] F. Kobayashi, T. Fukuda, K. Shimojima, and T, “Shape reconstruction system integrating stereo vision and shape from shading with evolutionary programming,” *Journal of Electronic*, 1998.

[21] B. K. Horn and R. W. Sjoberg, “Calculating the reflectance map.” *Applied optics*, vol. 18, no. 11, pp. 1770–9, Jun. 1979.

[22] A. Spence and M. Chantler, “Optimal illumination for three-image photometric stereo acquisition of texture,” in *Texture 2003: Proceedings of the 3rd international workshop on texture analysis and synthesis*, vol. pages, 2003, pp. 89–94.

[23] A. Spence and M. Chantler, “Optimal illumination for three-image photometric stereo using sensitivity analysis,” *IEE Proceedings-Vision, Image and Signal Processing*, vol. 153, no. 2, pp. 149–159, 2006.

[24] O. Drbohlav and M. Chantler, “Can two specular pixels calibrate photometric stereo?” in *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005*, vol. 2, 2005.

[25] S. Barsky and M. Petrou, “Design issues for a colour photometric stereo system,” *Journal of Mathematical Imaging and Vision*, vol. 24, no. 1, pp. 143–162, 2006.

[26] S. Sakane and T. Sato, “Automatic planning of light source and camera placement for an active photometric stereo system,” *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pp. 1080–1087, 1991.

[27] M. Chandraker, S. Agarwal, and D. Kriegman, "ShadowCuts: Photometric Stereo with Shadows," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Jun. 2007.

[28] D. Miyazaki, K. Hara, and K. Ikeuchi, "Median Photometric Stereo as Applied to the Segonko Tumulus and Museum Objects," *International Journal of Computer Vision*, vol. 86, no. 2-3, pp. 229–242, Jun. 2009.

[29] V. Argyriou, M. Petrou, and S. Barsky, "Photometric stereo with an arbitrary number of illuminants," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 887–900, Aug. 2010.

[30] a. Hertzmann and S. Seitz, "Shape and materials by example: a photometric stereo approach," *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, pp. I–533–I–540, 2003.

[31] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz, "Shape and spatially-varying brdfs from photometric stereo," in *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005*, vol. 1, 2005.

[32] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and spatially-varying BRDFs from photometric stereo." *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1060–71, Jun. 2010.

[33] T.-P. Wu and C.-K. Tang, "Photometric stereo via expectation maximization." *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 546–60, Mar. 2010.

[34] M. Holroyd, J. Lawrence, G. Humphreys, and T. Zickler, "A photometric approach for estimating normals and tangents," *ACM Transactions on Graphics*, vol. 27, no. 5, p. 1, Dec. 2008.

[35] N. Alldrin, T. Zickler, and D. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, vol. pp, 2008, pp. 1–8.

[36] R. J. Woodham, "Gradient and curvature from the photometric-stereo method, including local confidence estimation," *Journal of the Optical Society of America A*, vol. 11, no. 11, p. 3050, Nov. 1994.

[37] T. Simchony and R. Chellappa, "Direct analytical methods for solving Poisson equations in computer vision problems," *Pattern Analysis and*, vol. 12, no. 5, pp. 435–446, May 1990.

[38] M. Segal, C. Korobkin, R. van Widenfelt, J. Foran, and P. Haeberli, "Fast shadows and lighting effects using texture mapping," *ACM SIGGRAPH Computer Graphics*, vol. 26, no. 2, pp. 249–252, Jul. 1992.

[39] M. Fischler, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, 1981.

[40] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2000.

[41] W. Smith and E. Hancock, “Estimating Cast Shadows using SFS and Class-based Surface Completion,” *18th International Conference on Pattern Recognition (ICPR’06)*, pp. 86–90, 2006.

[42] C. Lawson and R. Hanson, *Solving least squares problems*, ser. Classics in applied mathematics. SIAM, 1995.

[43] Å. Björck, *Numerical methods for least squares problems*, ser. Miscellaneous Bks. SIAM, 1996.

[44] O. Drbohlav and M. Chantler, “On Optimal Light Configurations in Photometric Stereo,” *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, vol. II, pp. 1707–1712, 2005.

[45] S. Buyukatalay, O. Birgul, and U. Halici, “Effects of Light Sources Selection and Surface Properties on Photometric Stereo Error,” in *SIU*, no. 2, 2010.

[46] C. Everitt, A. Rege, and C. Cebenoyan, “Hardware shadow mapping,” *White paper, nVIDIA*, vol. 2, 2001.

[47] A. O. Ozturk, O. Birgul, E. Esen, and U. Sakarya, “Effects of Noises in Images used for Height Map Reconstruction of Cartridge Cases using Photometric Stereo,” in *SIU*, 2009.

[48] M. Powell and S. Sarkar, “A simple strategy for calibrating the geometry of light sources,” *IEEE Transactions on Pattern*, vol. 23, no. 9, p. 320, 2001.

[49] S. J. Kim and M. Pollefeys, “Robust radiometric calibration and vignetting correction.” *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 4, pp. 562–76, Apr. 2008.

VITA

Soner Büyükatalay was born in Erzurum, Turkey in 1979. He received her B.Sc. and M.Sc degrees from Middle East Technical University (METU), Department of Electrical and Electronics Engineering in 2001 and 2004 respectively. He is working towards Ph. D. degree in the same department since 2005.

He joined TÜBİTAK, UZAY in 2001 and has been working as a researcher since then. His areas of interest are Computer Vision and Pattern Recognition.

PUBLICATIONS DURING Ph.D. STUDY

International Peer-Reviewed Journal Papers

- **Büyükatalay S.**, Birgül Ö., Halıcı U., "Evaluation of Photometric Stereo using Different Masks with Variable Number of Images in a Unified Framework," Summited at March 2011, under review
- **Büyükatalay S.**, Birgül Ö., Halıcı U., "New Illumination Configuration and Solution Method for Photometric Stereo with Shadows and Highlights," Summited at August 2011, under review

International Conference Papers

- **Büyükatalay S.**, Birgül Ö., Halıcı U., "Surface Reconstruction from Multiple Images Filtering Non-Lambert Region," IEEE International Conference on Image Processing, Cairo, Egypt, 2009

National Conference Papers

- **Büyükatalay S.**, Birgül Ö., Halıcı U., "Effects of Light Sources Selection and Surface Properties on Photometric Stereo Error," IEEE 18th Signal Processing and Communications Applications Conference, Diyarbakır, Turkey, 2010.
- Öztürk A. O., **Büyükatalay S.**, Birgül Ö., "Generating Height Map of Cartridge Case by using Single Camera and Single Structured Light Source," IEEE 18th Signal Processing and Communications Applications Conference, Diyarbakır, Turkey, 2010.

- **Büyükatalay S.**, Birgül Ö., Halıcı U., “Iterative Photometric Stereo with Shadow and Specular Region Detection for 3D Reconstruction,” IEEE 17th Signal Processing and Communications Applications Conference, Antalya, Turkey, 2009.