博士論文

Photometric Stereo Using Constrained Regression

（制約付き回帰に基づく照度差ステレオ）

氏名　　池畑　諭

Photometric Stereo Using Constrained Regression

by

Satoshi Ikehata

Submitted to the Information and Communication Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Information Science and Technology

at the

UNIVERSITY OF TOKYO

December 2013

Author . . . .
Information and Communication Engineering
Dec. 13, 2013

Certified by . . . .
Kiyoharu Aizawa
Professor
Thesis Supervisor

Accepted by . . . .
...
...

Photometric Stereo Using Constrained Regression

by

Satoshi Ikehata

Submitted to the Information and Communication Engineering
on Dec. 13, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Information Science and Technology

Abstract

Photometric stereo is an inverse problem of recovering surface normals of a scene from appearance variations under different lightings. When the reflectance of a scene obeys a simple Lambertian assumption, the problem is well-constrained and there is a very simple algorithm capable of recovering the surface normals and albedos of the scene. However, the image formation process of the real world scene involves more complex interactions between scene shape, reflectance and illumination, making the problem more difficult.

Nearly all current approaches to this problem can be viewed as performing the standard regression analysis which finds the relationship between actual observation and unknown model parameters (*e.g.*, surface normals and unknown parameters of the image formation model) by minimizing the residual between the observation and estimated model output. From this standpoint, the statistical complexity of the photometric stereo problem can be discussed with regard to the behavior of outliers deviated from the underlying model assumption, and its complexity *i.e.*, the number and non-linearity of parameters in the model. For instance, the Lambertian diffuse reflectance model has only 3-DOF assuming the lightings are known a priori, and therefore less complex and easy to be solved inversely. However, more stable outlier rejection techniques are required when the behavior of non-Lambertian effects becomes more complex. On the other hand, some of non-Lambertian effects can be considered as inliers with the complex reflectance model and the derivative image formation model, however the increased complexity may make the inference difficult or even impossible. Thus, there is a constant struggle between model complexity and tractability.

This dissertation argues that new penalties and constraints applied to a simple (therefore mostly linear) image formation model well formulate the photometric stereo problem of the non-Lambertian scene. The dissertation begins by deriving optimal solutions to the sparse regression formulation of the photometric stereo problem, where a sparsity penalty is applied to outliers of Lambertian diffuse reflectance model to robustly neglect various non-Lambertian effects such as shadows, specular highlights, and sensor noises.

Then, we develop a branch to handle the non-Lambertian diffuse reflections with sparse non-diffusive outliers, where the inlier model can no longer be represented by a simple Lambertian reflectance model. Additionally, we present that the assumption on BRDF that a diffuse component is represented by a monotonic function of the surface normal and lighting dot-product, reasonably resolves the non-linearity of BRDF in the regression, which dramatically reduces the complexity of the problem. This function is constructed using a piecewise linear approximation, which is modeled as latent variables embedded within a hierarchical Bayesian model such that we may accurately compute the unknown surface normals while simultaneously separating diffuse from non-diffuse components.

Finally, we challenge to recover the surface normals of general isotropic materials without assuming any parametric BRDF nor special light configurations. A simple bivariate reflectance model is derived from a non-linear but versatile sum-of-lobes representations of the isotropic BRDF. Then, we prove that the non-linearity of the model is also canceled by assuming the monotonicity of the BRDF which is commonly observed in most physically plausible BRDF. Approximating the regression function by a smooth, bivariate Bernstein polynomials, the unknown normal vector is separated from the unknown inverse reflectance function, and then we may accurately compute the unknown surface normals by solving a simple and efficient quadratic programming problem.

Thesis Supervisor: Kiyoharu Aizawa
Title: Professor

Acknowledgments

This dissertation is the product of the constant endeavor of my supervisor, co-working researchers, colleagues, friends and family. We first thank all of them for supporting, discussing, and collaborating with me during three years of my PhD experience.

Prof. Kiyoharu Aizawa provided me excellent supervision and guidance. The atmosphere of the research environment created by Prof. Aizawa was very comfortable for me, making everyone creative and positive for the research and study. I also want to thank Prof. Toshihiko Yamasaki for not only advising me about the research direction but also chatting with me and inviting me for the dinner. Talking with Prof. Yamasaki always encouraged me.

Dr. Yasuyuki Matsushita and Dr. David P. Wipf provided me precious collaboration during my internship in Microsoft Research Asia (MSRA) and even after I left there thorough exciting skype meeting. I had always the pleasure of discussing with them not only about the collaborating project but also about many other topics.

Dr. Jiho Cho also provided me exciting collaboration. I was also happy to be a friend of him. Unfortunately, our collaborative work was not included in this dissertation, however the research mind I received from him exists in this dissertation.

I stayed a majority of my PhD life in Aizawa Yamasaki Laboratory and therefore was very encouraged by all lab members and Ms. Saiko Akabane. Though unfortunately there is no space to list all of members here, I wish to thank all of them for their support in numerous ways.

Finally, I want to express my gratitude to my family They gave me sincere support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

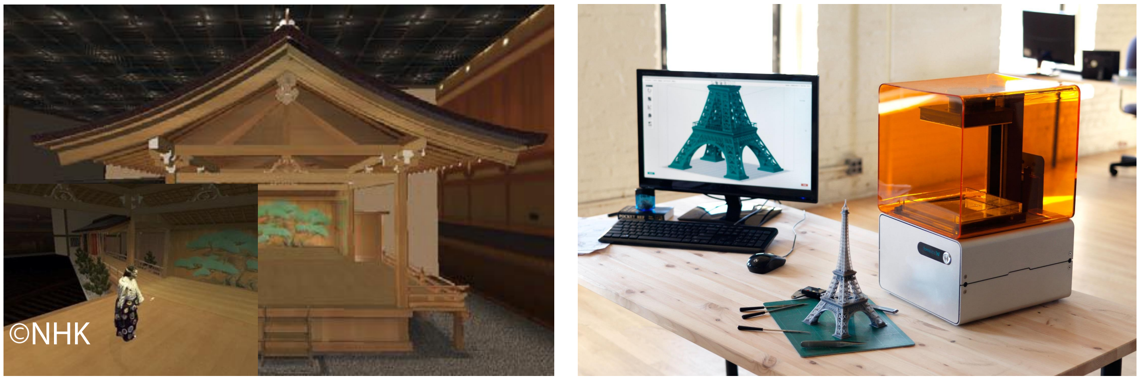


Figure 1-1: Examples of applications using reconstructed 3D model. (Left) dynamic 3D reconstruction of a human actor. (Right) The 3D printing technology is used for both prototyping and distributed manufacturing with applications in architecture, industrial design, automotive, aerospace, engineering, dental and medical industries, biotech (human tissue replacement), fashion,education, geographic information systems, food, and many other fields.

1.1 3-D Reconstruction

3-D reconstruction, the process of capturing the shape and/or appearance of real-world objects has drawn attention of researches in the computer vision area. The reconstructed 3-D information are typically reused for computer graphics, architecture, industrial design, robotics, medical applications and many other fields. Furthermore, the 3-D printing becomes one of the most attractive applications for taking advantages of reconstructed shape model (Fig. 1-1).

So far various kind of 3-D reconstruction techniques have been proposed to date, which are

Figure 1-2: Examples of appearances under different illumination. Left and right images are captured under a point light source placed at different positions. The appearance varies with the material and the shape of each object in the scene.

generally categorized into two approaches *i.e.*, passive and active, depending on whether the sensor actively emits the radiation which is directed toward the target or passively measures the energy which are reflected from it.

The active approach measures the distance from the camera to the objects directly using active infrared illumination. The most traditional sensors in this approach are perhaps laser range sensors which have been widely used for both research and industrial applications. While reliable, their prohibitively high cost and poor frame-rate (*e.g.*, 0.1Hz) have limited their usage. Therefore, depth sensors such as time-of-flight (ToF) and structured light sensors (*e.g.*, Microsoft Kinect) have recently been becoming popular alternatives to laser range sensors due to their low-price and high frame rate.

On the other hand, the passive approach is also expressed as *image-based* approach since the passive sensors such as digital cameras typically utilize 2-D images to reconstruct 3-D information as human biological systems do. Two of most traditional and representative techniques in the passive approach may be (multi-view) stereo [92, 95] and structure-from-motion [102] which reconstruct a 3-D scene by multiple image correspondences and triangulation analogously to the human 3-D perception system using the binocular disparity or the motion parallax. Those two approaches are generally referred to as the *geometric* passive approach since they rely on the multi-view geometry as a cue of reconstruction [50].

In contrast to the geometric approach, the *photometric* approach is another representative passive technique which explicitly models the natural light transport and inversely traces the image formation to recover the surface structure. The light transport and its derivative image formation involve the surface geometry (*i.e.*, generally represented as a surface normal), illumination and re-

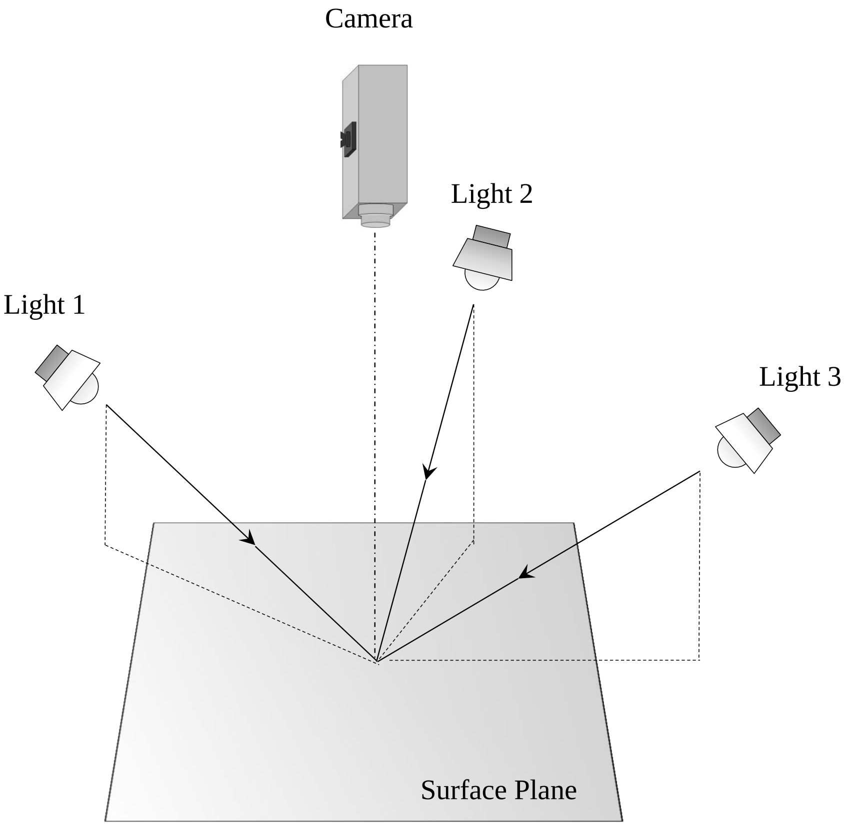


Figure 1-3: Typical setup of the image acquisition in the photometric stereo problem. Three and more light sources independently illuminate the scene, where they are placed at different positions with a sufficient distance from the target object to assume the directional lighting (we further discuss about the lighting setup in Section 2.2.1).

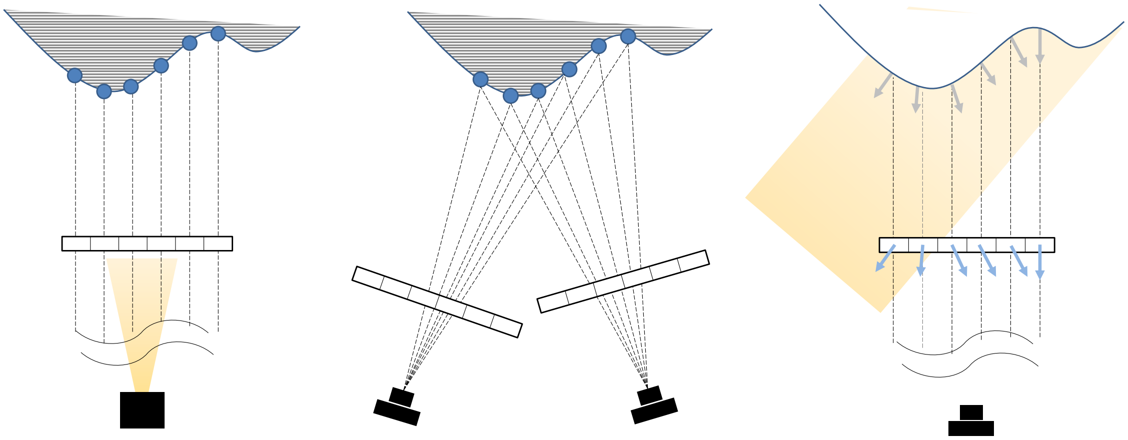
flection on the object surface, so any change of them leads to different appearances. The typical techniques in this approach are called by different names depending on whether the illumination is static or dynamic. If the image is captured under a single light source or static natural illumination, the reconstruction technique is called *shape-from-shading* [132]. When images are captured under a variational illumination (*i.e.*, generally multiple directional light sources as is shown in Fig. 1-2), the technique is referred to as the *photometric stereo* [122], which is a main interest in this dissertation. While the image acquisition setup is simple, the shape-from-shading technique is generally applicable to the scene whose reflectance is known (and often limited to Lambertian) since no appearance variation leads to the underdetermined problem. On the other hand, the image acquisition setup of the photometric stereo is generally complicated. At least three different kinds of illumination are required to constrain the problem even with the simplest Lambertian scene [122] as shown


(a) (b) (c)

Figure 1-4: Examples of a surface reconstruction using the photometric stereo technique. (a) Images are captured under varying illumination. (b) Surface normals are recovered by inversely solving the image formation model. In this figure, the surface normal is mapped by color, where each channel represent the each element of a three dimensional normal vector. (c) The surface mesh can be recovered from the estimated surface normals by using existing mesh reconstruction techniques *e.g.*, a poisson solver [4].

in Fig. 1-3. However, comparing with the shape-from-shading technique and the other 3-D reconstruction techniques including the active approach, the estimation accuracy of the geometry and the variation of available materials are very impressive (We show an example of the photometric stereo result in Fig. 1-4).

We illustrate the analytical comparison among passive approaches (photometric and geometric) and active approach in Fig. 1-5 and Table 1.1. The main advantage of the photometric approach is that it can recover much more accurate high-frequency details than the geometric passive approach and the active approach. Triangulation-based stereo techniques in the geometric approach on the other hand provide robust low frequency structure, but suffer from high frequency structure due to inaccurate correspondences and quantization errors. Also, a depth map captured by a ToF camera, the typical active sensor, has very low-resolution (*e.g.*, 176×144 in Swiss Ranger SR4000) and depth maps acquired by these sensors often contain unreliable areas where the scene contains occluded regions or objects with low reflectivity. Secondly, since especially the photometric stereo technique explicitly models the reflection of light, it is capable of handling wide variety of untextured and non-Lambertian surfaces. On the other hand, the multi-view stereo does not work for those materials since it assumes that the appearances from different viewpoints are same. The scalability of the technique is also very important factor for the 3-D reconstruction. While the active depth sensors only work for a certain range of size or distance of the target, passive approaches usu-



(a) Active approach (b) Geometric passive approach (c) Photometric passive approach

Figure 1-5: Illustration of 3-D reconstruction approaches. (a) The active sensor (approach) emits the infrared illumination and allocate the measured distances between the sensor and discrete surface points to a 2-D image so called depth map. (b) The geometric passive approach finds the correspondences over images, then recover the surface point using the triangulation (c) The photometric passive approach recovers the surface normal of the scene by inversely solving the image formation model which relates the illumination, the geometry and the reflection.

ally cover the range from micro to macro as long as the camera can well capture the radiance of the scene. We should note that the photometric stereo has a large advantage in recovering microscopic structures since the camera is static in the photometric stereo technique However, on the other hand the triangulation for those kind of surfaces are very problematic since the stereo cameras can not assure the enough base-lines for the triangulation. We should also note that the triangulation has advantages in the macro scene since the control of lighting environment is very difficult in such kind of scenes for the photometric stereo technique.

Several disadvantages of the photometric technique mainly lie in the image acquisition setting. While other passive approaches and active sensors perform well under the ambient illumination[1], the photometric approach (especially, the photometric stereo) is generally disrupted under the ambient illumination. Furthermore, photometric approach usually requires high dynamic range (HDR) image(s) to capture the reflections on dark to bright surfaces and therefore we need to capture the same scene under varying exposure of the camera unlike other passive approaches and active sensors. Another difficulty is caused by the calibration of lightings. In the photometric stereo, except for uncalibrated photometric stereo problems such as [38], lightings must be known in advance (see

---

[1] Active sensors actually does not work in the scene with the strong ambient illumination such as the outdoor environment.

Table 1.1: Comparison of pros. and cons. of shape reconstruction techniques.

Criterion	Photometric approach	Geometric approach	Active Sensors
Shape Recovery	normal	depth	depth
High Frequency Details	yes	no	no
Spatial Resolution	high	high	low
Camera	1+ fixed	2+ fixed or 1+ moving	1
Un-textured Region	ok	problematic	ok
Calibration	camera + lighting	camera	camera
Illumination	important	minor importance	minor importance
Object Movement	problematic	problematic	minor problematic
Non-Lambertian Surfaces	possible	challenging	possible
Scalability	yes	yes	no
Reflectance Recovery	possible	no	no

Section 2.2.1 in details). Therefore, in addition to the camera calibration which is also required to other passive approaches, the photometric stereo further requires a lighting calibration where the position and the intensity of the emission of the light sources are estimated using some calibration aperture.

In the next section, we will discuss about the photometric stereo problem which is a main interest in this dissertation. We will then more precisely define the inverse problems we wish to solve followed by detailed descriptions of several difficulties that are accompanied with these problems. We will conclude by providing an outline of the remainder of this thesis.

1.2 Photometric Stereo as Regression Analysis

Photometric stereo is an inverse problem of recovering surface normals of a scene from appearance variations under different lightings. Nearly all current approaches to the photometric stereo problem can be viewed as performing the regression analysis which finds the relationship between actual observation and unknown model parameters (*i.e.*, surface normals and unknown parameters of the image formation model) by minimizing the residual between the observation and estimated model output. For instance, the Lambertian diffuse reflectance model has three degree of freedom (DOF) (two for the unit surface normal vector and one for the surface albedo) assuming the lightings are known a priori, and its derivative image formation model under i-th lighting is simply represented as [2]

$$I_i = \rho \boldsymbol{n}^T \boldsymbol{l}_i = \rho n_1 l_i^1 + \rho n_2 l_i^2 + \rho n_3 l_i^3 = n'_1 l_i^1 + n'_2 l_i^2 + n'_3 l_i^3, \tag{1.1}$$

where I is an irradiance reflected at the point on a surface whose unit normal vector is $\boldsymbol{n} \triangleq [n_1 \ n_2 \ n_3]^T$, ρ is a surface albedo, $\boldsymbol{l} \triangleq [l_i^1 \ l_i^2 \ l_i^3]^T$ is a normalized lighting vector and $n'_k \triangleq \rho n_k$ $(k = 1, 2, 3)$. Given pairs of the observed intensity for a surface point and the lighting vector, the Lambertian photometric stereo is equivalent to performing a linear regression analysis whose regression function $f = n'_1 x_1 + n'_2 x_2 + n'_3 x_3$ relates the dependent variable $y \triangleq I$ to independent variables $[x_1 \ x_2 \ x_3]^T \triangleq [l_i^1 \ l_i^2 \ l_i^3]^T$. The unique and optimal solution exists when three and more pairs are given and all of them obey the Lambertian image formation model [122] *i.e.*, there exists a four dimensional plane passing through the origin and all 4-D points of $[x_1 \ x_2 \ x_3 \ y]$. However, the image formation process of the real world scene involves more complex interactions between scene shape, reflectance and illumination. In this situation, most widely used regression algorithm, the least-squares fitting, suffers from the violation of the assumption underlying a regression model, leading to a severely more complicated estimation task since there are now an infinite number of solutions that could have produced the observed intensity I with equal likelihood.

We assume now that appearances of the target object can be approximated by a reflectance function [3] which relates the observed intensity I at a given point on the object to the associated surface normal $\boldsymbol{n} \in \mathbb{R}^3$, the incoming lighting direction $\boldsymbol{l} \in \mathbb{R}^3$, and the outgoing viewing direction $\boldsymbol{v} \in \mathbb{R}^3$ (fixed as $\boldsymbol{v} = [0, 0, 1]^T$ in this dissertation), then the image formation model under i-th lighting is described as

$$I_i = f(\boldsymbol{n}, \boldsymbol{l}_i, \boldsymbol{v}, \boldsymbol{\mu}) + \epsilon_i. \tag{1.2}$$

Here $\boldsymbol{\mu}$ is a vector of model parameters derived from the reflectance model and ϵ is an additive corruption applied to the ideal reflectance[4].

Suppose we are presented with m images that are captured under varying directional lighting sources, the photometric stereo problem is to recover a normal vector as a part of unknown model parameters and additive corruptions, which is equivalent to solving a regression analysis, where the regression function f relates the dependent variable I to independent variables $[\boldsymbol{n} \ \boldsymbol{\mu}]$ in the absence of corruptions $\boldsymbol{\epsilon}$. The most ubiquitous estimator used for this task is one that maximizes

---

[2] Here we assume a single, orthographic camera whose response curve is linear. However, the photometric stereo problem can also be formulated under more general condition as will be discussed in Section 2.2.2.

[3] Note that BRDF including a Lambertian reflectance model is a special class of this function.

[4] In this thesis, we assume other phenomenon than local illumination such as global illumination, model outliers, calibration errors and sensor noises are modeled by this term.

the likelihood of the data, which is equivalent to the least-squares solution as

$$\min_{\boldsymbol{n},\boldsymbol{\mu},\boldsymbol{\epsilon}} \sum_{i=1}^{m} \|I_i - f(\boldsymbol{n}, \boldsymbol{l}_i, \boldsymbol{v}, \boldsymbol{\mu}) - \epsilon_i\|_2^2, \tag{1.3}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is a vector of additive corruptions. Note that the Lambertian photometric stereo presented above is a special case of Eq. (1.3) where $f = \rho \boldsymbol{n}^T \boldsymbol{l}_i$ and $\epsilon_i = 0 \ \forall i$. The simultaneous estimation of $\boldsymbol{n}$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ in Eq. (1.3) is generally intractable since there are multiple feasible solutions (*i.e.*, the number of unknown variables exceeds the number of independent equations). A statistical remedity to this indeterminacy neglects uninfulential observations as outliers that allows us to narrow the space of candidate solutions in a manner consistent with the image formation model [41, 35, 88, 35, 39, 25]. However, in general this is not easy since we know of little practical method identifying outliers when the regression function is non-linear or a cause of outliers is a mixture and/of multiple distributions derived from different phenomena *e.g.*, shadows, sensor noises, interreflections and so on.

Another issue is that approximating a surface reflectance with a specific parametric model inevitably limits the variations of target material its applications. From the statistical viewpoint, non-parametric regression might be helpful for the analysis [40], which uses fewer assumptions about the data (*e.g.*, the inlier model is represented as a sum of continuous basis functions weighted by coefficients). However, a critical issue to use the non-parametric representation in the photometric stereo problem is that the complex interaction between scene shape, reflectance and illumination is hard to be encoded in the nonparametric manner. Even if an appropriate model is constructed, a large number of predictors that may interact with the surface normal nonlinearly, make the inference difficult or even impossible without an alternative optimization scheme [9].

In response to the discussions above, the question is "how can we resolve a constant struggle between model complexity and tractability", implying that increasing the flexibility and robustness is required to handle wide variety of real scenes, however the complexity must be controlled to make the problem tractable.

In this dissertation, *the constrained regression analysis* is explored for tackling these issues in the photometric stereo problem. We show several results where some specific penalties or constraints applied to the least-squares regression analysis (Eq. (1.3)) (a) disambiguate multiple feasible solution in the robust regression methods and (b) resolves the non-linearity of the problem, which contributes to simplifying the non-linear least-squares problem into the constrained linear system.

Firstly, we prove that the minimization of a sparsity penalty applied to ϵ disambiguates the infinity of feasible solutions to Eq. (1.3), which is henceforth called photometric stereo using *sparse regression* as [43]. In the sparse regression, we do not assume the specific distribution of ϵ instead we assume its *sparsity* that the number of inlier data in observations is larger than one outliers. Under the assumption, the regression analysis is formulated as finding a solution so that as much as possible number of data is represented by the inlier regression model (*i.e.*, the image formation model without additive corruptions). The derivative ℓ_0-norm minimization entails a difficult, combinatorial optimization problem that must be efficiently solved. In Chapter 3, we consider two alternatives to brute force exhaustive search: A relaxation replacing the discontinuous, non-convex ℓ_0 norm with the convex surrogate ℓ_1 norm, and a simple hierarchical Bayesian approximation called sparse Bayesian learning (SBL) [111].

Secondly, we prove that a specific monotonicity assumption of the non-linear reflectance model, and its derived *inverse* representation of the image formation model can be used to separate the surface normal from other unknown parameters in the model, which allows us to relax the non-linear, non-parametric regression analysis to the linear system with the monotonicity constraint, leading to efficient and stable inference. This theory is utilized to realize the system of ”non-linear diffuse reflectance model” with ”sparse regression” in the Chapter 4 and quadratic formulation of simultaneous estimation of both surface normal end reflectance of general isotropic materials in Chapter 5. Note that the constrained regression with a single predictor have recently been well explored, that with *multiple predictors* are less explored and if any, estimating a multivariate regression function subject to shape restrictions with compact support is challenging and usually very time consuming [117]. Therefore, the selection of the basis functions for the non-parametric regression is critically important. In Chapter 4, we take advantages of the simplicity of the piecewise linear function, leading to a natural generalization of the Lambertian reflectance model while allowing us to handle more general non-linear diffuse reflections where needed. In Chapter 5, we explore the Bernstein polynomials approximation which naturally selects smooth functions with little computational effort unlike other non-parametric regression function (*e.g.*, smoothing spline [15]) and the shape-restricted regression function estimate is shown to be the solution of a quadratic programming problem [19, 117]; making it computationally attractive.

1.3 Thesis Outline

The reminder of this thesis is organized as follows. In Chapter 2 we derive the image formation model from the standpoint of radiometry and surveys the existing photometric stereo works. In Chapter 3 we propose a sparse regression based robust photometric stereo algorithm and provide a detailed, comparative evaluation with other Lambertian based robust algorithms. Chapter 4 extends the sparse regression based approach to handle non-Lambertian diffuse reflections. Chapter 5 switches the goal and provides a more general photometric stereo algorithm which works for general isotropic materials. We finally conclude this thesis in Chapter 6.

Chapter 2

Preliminaries

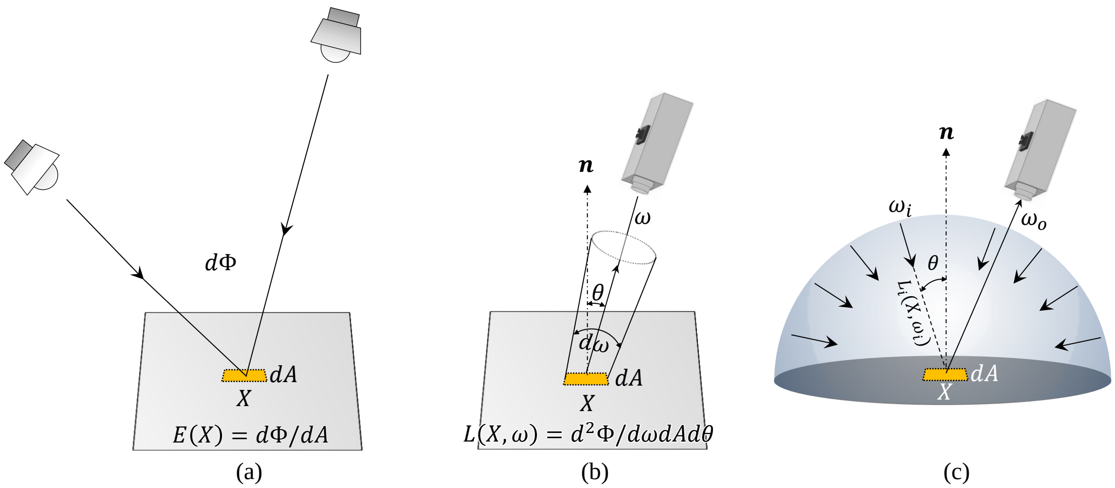
In this chapter, we will provide a light transport analysis to derive a unified image formation model underlying the photometric stereo problem. Then, a brief survey of existing photometric stereo algorithms is provided as well as a typical setups or modeling of the camera, the illumination and the reflectance. Finally, we close this chapter by summarizing the major symbols and assumptions which will appear in the overall dissertation.

2.1 Image Formation Model

Understanding photometric stereo problem requires some basic knowledge about the physical light transport process underlying the image formation. In this section, we will firstly provide a derivation of the image formation model that explains how an image is generated in the complicated interaction of the camera, illumination and surface geometry. Next we go into the simplified models that are widely assumed within the photometric stereo literature since the rigorously physical light transport model is too complex to trace inversely, leading to the intractable photometric stereo problem. This section borrows heavily from Kurachi [64], which the reader may wish to reference for a more comprehensive review of the radiometry.

We firstly review the definition of *irradiance* and *radiance* as illustrated in Fig. 2-1. In the radiometry, the irradiance $E(X)$ at a point X on a surface measures the temporal change of radiant energy received on the local small region dA containing X, thus expressed as

$$E(X) = \frac{d\Phi}{dA}, \tag{2.1}$$



(a) (b) (c)

Figure 2-1: Definition of irradiance and radiance. (a) The irradiance is the temporal change of the flux which hits the differential surface area. (b)The radiance is the flux in a differential cone from a differential area. (c) The irradiance at a surface point is a summation of incoming radiance directed from all points on the hemisphere Ω, and outgoing radiance to ω_o is a summation of the reflection of all incident lightings to the direction.

where Φ is the radiant energy per time (called as *flux*). While irradiance does not depend on the direction in which the light is coming, the radiance $L(X, \omega_i)$ is the radiant power exiting from a point X in a direction ω, which is described as

$$L(X, \omega) = \frac{d^2\Phi}{d\omega dA cos\theta}, \tag{2.2}$$

where θ is the angle between surface normal and the incident lighting direction.

In real environments, incident radiance comes from light reflected off other objects as well as light sources which emit the energy, which is called as *indirect illumination*. Therefore, the irradiance at a surface point is computed by integrating the cosine weighted incident radiance in all directions above the surface as

$$E(X) = \int_\Omega L_i(X, \omega_i)(\boldsymbol{n}^T \omega_i) d\omega_i. \tag{2.3}$$

Here, $\boldsymbol{n}$ is the unit normal vector (*i.e.*, $\boldsymbol{n}^T \omega_i = cos\theta$) and ω_i is the incoming direction. Ω represents a hemisphere above the surface at X on which the radiance over the set of directions to the source are integrated. Multiplication by $\boldsymbol{n}^T \boldsymbol{\omega_i}$ essentially undoes the projective foreshortening and thereby accounts for the spreading out of flux coming in at an angle.

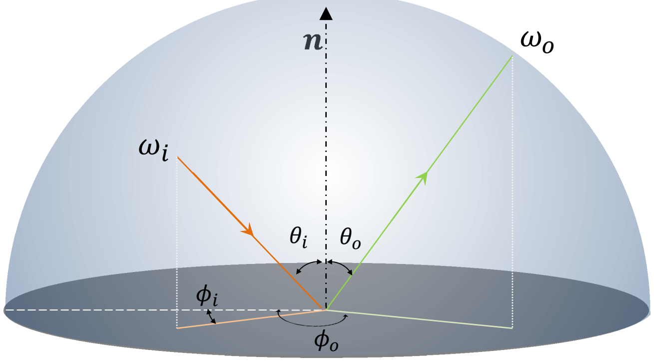


Figure 2-2: Incoming and outgoing vectors in the spherical coordinate system.

Outgoing radiance at a surface point X directing to ω_o involves integrating the reflection of incoming radiance in all directions. Assuming that the light is reflected at the point where the light is entered [1], the behavior of the reflection is modeled by a bidirectional reflectance distribution function $\rho(X, \omega_i, \omega_o)$ (BRDF), which determines how much light coming in from direction ω_i is reflected out in direction ω_o as

$$L_r(x, \omega_o) = \int_\Omega \rho(X, \omega_i, \omega_o) L_i(X, \omega_i)(\boldsymbol{n}^T \omega_i) d\omega_i, \tag{2.4}$$

A BRDF is defined as the ratio of a temporal variation of the outgoing radiance $L_r(X, \omega_o)$ to one of the surface irradiance at X as

$$\rho(X, \omega_o, \omega_i) = \frac{d^2 L_r(X, \omega_o)}{dL_i(X, \omega_i) \cos\theta d\omega_i}, \tag{2.5}$$

where $L_r(X, \omega_o)$ is the out going radiance in the direction pointing at ω_o. BRDF is typically defined relative to the local coordinate system of a given point on a surface. Using spherical coordinates, the direction of outgoing lightings are represented by two variables as $\omega_i = (\theta_i, \phi_i)$ and $\omega_o = (\theta_o, \phi_o)$ as shown in Fig. 2-2. Therefore, Eq. (2.5) is expressed as a function of four variables in the spherical coordinates *i.e.*, $\rho(X, \omega_o, \omega_i) = \rho(X, \theta_i, \phi_i, \theta_o, \phi_o)$. We note that the definition given in Eq. (2.5) is abstract and an explicit representation for BRDF (*e.g.*, Cook-Torrance model [29]) is usually required to render the scene in the computer graphics area[2]. However, this is not the case of the photometric stereo problem as we will further discuss in Section 2.2.3.

---

[1] In the natural environments, some of the incident light penetrates the object, bounces around below the surface, then emerges at some different point. This phenomenon is called as *subsurface scattering*.

[2] Some of concrete representations of BRDF are shown in the Appendix 2.

It is common observation that the physically plausible BRDFs commonly satisfy following properties:

(a) Non-negativity: A BRDF is never negative, while it can get arbitrarily large.

(b) Bidirectionality: The value of a BRDF remains the same even if the incoming and out going directions are exchanged.

(c) Energy conservation: The total power output computed by integrating the cosine-weighted BRDF value in all directions on the hemisphere cannot exceed the power coming in from ω_i.

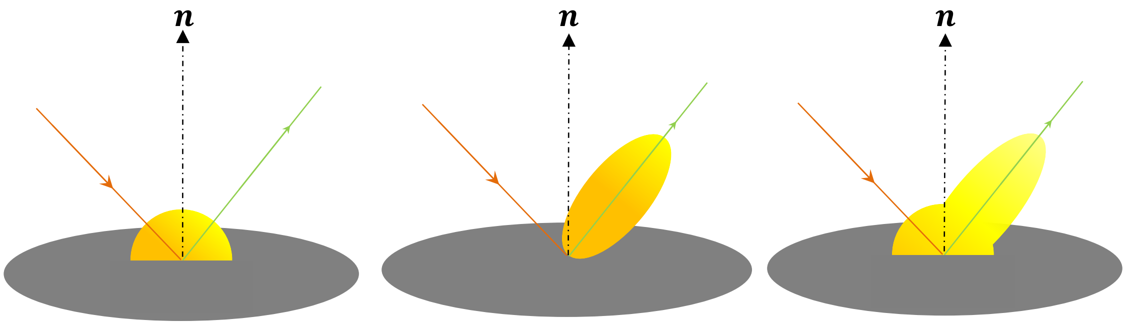
Those properties perhaps less intuitive, since a BRDF is complex and high-dimensional. Fortunately, we have a commonly used way of visualizing BRDF to fix the surface point x and the incoming direction ω_i, and consider the behavior of ρ as a function of only the outgoing direction ω_o. The typical illustration of BRDF is shown in Fig. 2-3. A BRDF is generally represented as a *lobe*, that is the relative distribution of the surface reflection of light from the fixed direction. Highly smooth surfaces (*e.g.*, metallic) exhibit only the specular reflections that are represented by long thin lobes centered near the mirror direction of ω_o, while diffusive surfaces have more uniform BRDF lobes (*e.g.*, rubber). It is worth mentioning that the real BRDF has usually more complex structures where some of lobes are complicatedly mixed together because surfaces are neither perfectly smooth nor entirely clean [23]. Furthermore, the peak in the distribution of reflected light is not always in the specular direction (*i.e.*, known as *off-specular peak*), leading to more complex image formation model which is hard to trace inversely in the photometric stereo problem[3]. In Chapter 5 of this dissertation, we will tackle the photometric stereo problem where the surface reflectance is represented by a sum of unknown lobes.

While Eq. (2.4) is useful for rendering scenes under known illumination environment, unfortunately, however, inversely tracing incoming lightings in all directions is almost impossible since this problem is highly ill-posed even with a simple BRDF such as Lambertian model [122]. Therefore, simplified model of Eq. (2.4) is generally analyzed in the most photometric stereo literature. Concretely, it is generally assumed that a direction ω_i corresponds to the primary light source (*e.g.*, a point light source placed in the scene) and the value of L_i is the radiance emitted by the source (*i.e.*, no indirect illumination exists). Under these assumptions, Eq. (2.4) is simplified as

$$L_r(X, \omega_o) = \rho(X, \omega_i, \omega_o) L_i(X, \omega_i) \boldsymbol{n}^T \omega_i. \tag{2.6}$$

---

[3] Though there are several causes that provide off-specular peaks, the most well-known phenomena is perhaps the Fresnel effect [42].



(a) Diffuse (b) Specular (c) Diffuse + Specular

Figure 2-3: Examples of the BRDF lobe. The yellow shaded areas represent the radiance power emitted from the center of the lobe as for a certain outcoming direction. Here we present three typical types of lobes that express (a) diffuse reflection, (b) specular reflection, (c) specular plus diffuse reflection.

Suppose the light source is modeled by the directional light[4], the radiance emitted by the source (L_i) no longer depends on the position of surface point leading to the position-independent value of $L_i(\omega_i)$. When we observe a surface point from a direction $\boldsymbol{v}$ where the incident light coming from $\boldsymbol{l}$ is reflected on the surface, the radiance in outgoing direction (*i.e.*, $\omega_o = \boldsymbol{v}$) is computed as,

$$L_r(X, \boldsymbol{v}) = \rho(X, \theta_i, \phi_i, \theta_o, \phi_o) L_i(\boldsymbol{l})(\boldsymbol{n}^T \boldsymbol{l}). \tag{2.7}$$

While Eq. (2.7) is much simpler than Eq. (2.4), the high dimensionality of the BRDF makes it difficult to use the model in the photometric stereo problem. For simplifying the BRDF analysis (and the formulation of the photometric stereo problem), a common assumption is that the BRDF is invariant to rotations and reflections about the surface tangent plane, which is called as *isotropic BRDF*. An isotropic BRDF is a function of three variables, that is $\rho(X, \theta_i, \theta_o, \|\phi_i - \phi_o\|)$ in the spherical coordinate. On the other hand, the original four dimensional BRDF is distinctively called as *anisotropic BRDF* (The examples of both isotropic and anisotropic material are shown in Fig. 2-4). It is worth mentioning that a majority of materials in the real world such as plastic, polished metal, paper, human skin, and most painted surfaces exhibit isotropic reflections, while the variation of anisotropic materials is, if any, very limited (*e.g.*, brushed or milled metal).

In the photometric stereo problem, we recover the normal of a surface point not from the radiance but the intensity (or RGB color) values at a pixel in *2-D images* recorded by a camera, therefore we also need to consider how the camera capture radiometric information of the scene. In most cam-

---

[4] We will further discuss the light model in Section 2.2.1.


(a) Isotropic (b) Anisotropic

Figure 2-4: Examples of the isotropic and anisotropic material.

eras, there exists a nonlinear relationship between the image intensities output from the camera and the scene radiance that enters the imaging system to compress the dynamic range of scenes and to account for nonlinearities in display systems [70] as

$$I(X) = g(L_r(X, \boldsymbol{v})). \tag{2.8}$$

Here $I(X)$ is a pixel intensity corresponding to a surface point X, g is a non-linear function which is generally called as *radiometric response function* of the camera.

Furthermore, the position of each pixel is defined on the image plane whose coordinate is different from the object coordinate whose center is placed at the object, where the surface point X, the surface normal, and the BRDF are defined. Therefore, we also need to define a mapping from 3-D surface point onto the 2-D image plane $\pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$ where a pixel on the image plane is projected from X as $[p\ q] = \pi(X)$. Considering this mapping, the image formation model is described as follow,

$$I(p, q) = I(\pi(X)) = g(L_r(\pi(X), \boldsymbol{v})) = g\left(\rho(\pi(X), \theta_i, \theta_o, \|\phi_i - \phi_o\|)L_i(\boldsymbol{l})(\boldsymbol{n}^T\boldsymbol{l})\right). \tag{2.9}$$

A reasonable approximation of the mapping from 3-D coordinate onto the 2-D image plane is the perspective projection model, where 3D points are mapped onto the 2D camera sensor by dividing the first two coordinates by the third one. However, because of the nonlinearity of this mapping, a simple orthographic projection is generally assumed in the photometric stereo problem, where a 3-D surface point $X = (x, y, z)$ is mapped onto a pixel at $[p\ q]$ on the image plane as $[p\ q] = [x\ y]$. In addition that, the radiometric camera response function is also commonly assumed as a linear function *i.e.*, $g(L_r) = rL_r$, and the scale factor r is usually dropped without losing generality.

Finally, the image formation model under the linear response function and orthographic camera model is constructed as follow,

$$\tilde{I}(p, q, \boldsymbol{l}) = \tilde{\rho}(p, q, \theta_i, \theta_o, \|\phi_i - \phi_o\|)(\boldsymbol{n}^T\boldsymbol{l}). \tag{2.10}$$

where $\tilde{I}$ is an observed intensity that is normalized by the radiance power of the incident light *i.e.*, $\tilde{I}(p, q, \boldsymbol{l}) = I(p, q)/L_i(\boldsymbol{l})$ and $\tilde{\rho}$ is the BRDF defined on the image space.

We should note that the linear camera response function and orthographic camera model are not necessary assumed in the photometric stereo literature. For example, Shi *et al.* [98] presented the algorithm that automatically recover the unknown radiometric response function by using the color profile which are obtained registered pixels across images. And Papadhimitri *et al.* [84] developed the uncalibrated photometric stereo[5] under the perspective projection camera model. While weaker assumptions on the image formation model contributes to simplify the image acquisition process, more complex image formation model leads to difficulties in handling wide variety of materials. Actually, both [98] and [84] limit on objects whose reflections are models by a simple Lambertian reflectance model[6].

For simplifying following discussions, we henceforth express the normalized intensity $\tilde{I}$ by I and BRDF $\tilde{\rho}$ multiplied by $\boldsymbol{n}^T\boldsymbol{l}$ as $f(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v})$. Then, modeling any kind of corruptions that violate Eq. (2.10) by ϵ (*e.g.*, sensor noises, shadows, inter-reflections) leads the general image formation model for the photometric stereo problem in this dissertation that is already presented in Eq. (1.2).

As we have already mentioned, the photometric stereo problem is to recover the surface normal of a point by inversely solving Eq. (1.2) given a collection of pairs of I and $\boldsymbol{l}$. In the next section, we will briefly review existing photometric stereo algorithms as well as show some typical configurations of image acquisition setup.

2.2 Taxonomy of Photometric Stereo Algorithms

In this section, existing photometric stereo algorithms are categorized according to three fundamental components that differentiate the major algorithms: (1) the lighting setup, (2) the camera setup, and (3) the reflectance modeling, where each component is further divided into some properties.

---

[5] In the literature of the photometric stereo, *uncalibrated* generally means that the lighting is unknown.

[6] Further discussions of the camera model assumption in the photometric stereo problem will be presented in Section 2.2.2.

Table 2.1: Summary of the lighting setup.

Typical Property	Major Condition	Minor Condition	This Dissertation
Direction and Intensity	known	unknown	known
Type	parallel	nonparallel	parallel
Number of sources	five+	three or four	five+
Configuration	random	special	random
Ambient illumination	neglected	considered	neglected
Shadows	neglected	considered	partially considered[1]
Inter-reflections	neglected	considered	neglected

[1] In Chapter 3 and Chapter 4 we reject shadows by using the robust sparse regression, however in Chapter 5 we assume that shadows are removed in advance.

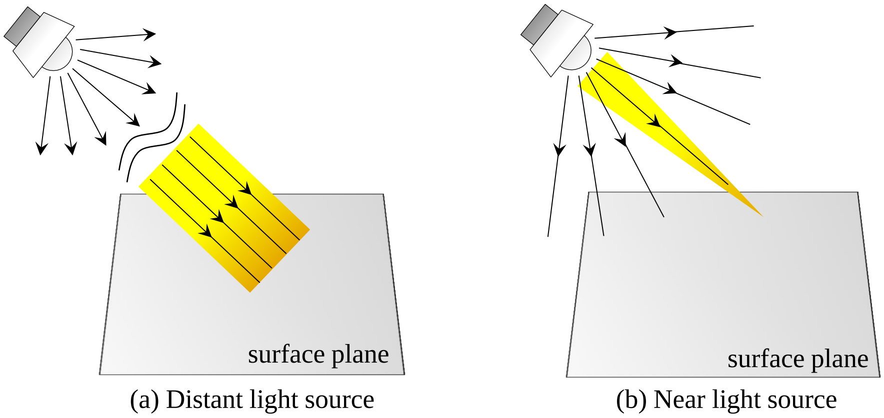
2.2.1 Lighting Setup

The lighting setup determines the illumination environment where images captured captured. As for this component, the photometric stereo algorithms are generally categorized by responding to these seven questions as shown in Table 2.1: (a) Calibrated or uncalibrated? (Are the direction and intensity of a light source known or not?) (b) What is the type of light source? (which illumination model is assumed for a light source?) (c) How many light sources are placed? (How many images are captured under varying illumination?) (d) How is the configuration of light source? (Where each light source is placed?) (e) How is the ambient illumination treated in the reconstruction? (f) How are cast/attached shadows treated in the reconstruction?, and (g) How are interreflections treated in the reconstruction?

(a) Prior information about the lighting direction and intensity

When the lighting directions and intensities are known, the problem is called as *calibrated* photometric stereo problem (*e.g.*, [122, 46, 9, 26, 55]). On the other hand, when no prior knowledge about the illumination is available, the problem is generally called as *uncalibrated* photometric stereo problem. Uncalibrated photometric stereo problem has been addressed with a variety of techniques that make explicit or implicit use of the generalized bas-relief (GBR) ambiguity, where surface normals of a Lambertian object can only be determined up to a linear ambiguity [51]. Therefore, assuming the Lambertian surfaces, many researchers tackled the uncalibrated photometric stereo problem as the disambiguation task of this GBR ambiguity [8, 98, 32, 6, 108, 135, 38, 104, 22, 84].

For instance, Alldrin *et al.* [8] solved the GBR ambiguity by minimizing the entropy of the albedo distribution under the assumption that albedos are based on a few intensity values. A similar approach is to group the normal-albedo distribution based on color appearance [98]. Different from



(a) Distant light source (b) Near light source

Figure 2-5: Two different behaviors of a point light source based on its distance from the surface. (a) when a point light is placed distant enough from the surface, the point light is approximated as a directional (parallel) light, where emitted radiance intensity and direction are constant over the surface plane. (b) when a point light is placed nearby the surface, its emitted radiance intensity and direction are varying over the surface plane.

a group of works which exploit the priors on surface albedos, some methods resolved the GBR ambiguity by analyzing reflectance properties. For instance, specularities of glossy surfaces [32], isotropy and reciprocity of the isotropic BRDF [6, 108], Helmholtz reciprocity principle[135], the maxima of the Lambertian diffuse reflectance component [38] have been incorporated in this problem. Other than these strategies, shadows [104], inter-reflections [22] and perspective geometry [84] have been used to resolve the GBR ambiguity.

While uncalibrated photometric stereo techniques are very practical since measuring the direction and intensity of light sources are time-consuming and often requires additional calibration object, it is worth noting that they usually give less accurate estimation than the calibrated photometric stereo techniques especially when the reflectance (or surface albedos) are spatially varying and more complex than a simple Lambertian reflectance model even though some recent methods have tried to handle general isotropic materials by analyzing the relationship between the intensity distribution of an intensity profile [72] or using the half vector symmetry of the isotropic material[127].

(b) The type of light sources

Generally, each light source is assumed to be a directional (parallel) light in the photometric stereo problem (*e.g.*, [7, 8, 6, 46, 26, 124, 59]), which is practically realized by placing a point light very far away the object *i.e.*, the working distance of an illuminator to an object surface is more than five times of the maximum dimension of the light emitting area [12].

Although that is a common case in the photometric stereo problem, there are some scenarios


(a) Image acquisition setup (b) Recovered surface normal map

Figure 2-6: Example of the surface geometry recovery using a screen illumination [5]. A nearly planar scene is captured by a static camera under the illumination of a common computer screen. Detailed surface geometry and reflectance are then recovered from them. Pictures here are from [5].

where the distant lighting assumption is invalid *e.g.*, there is no place to largely distribute the light sources in the indoor scenes. Therefore, there exist, if little, some photometric stereo algorithms which assumed other types of the light source. For instance, a group of works tackled the photometric stereo problem in the situation that the light source is placed near the target object [62, 103, 63]. Since the incident angles at each scene point are different as shown in Fig. 2-5, the distribution of radiance at every pixels (often called as *a light field*) need to be estimated in advance. Kim *et al.* [62] modeled the light field of a near light source considering the incident intensity drop-off passing along the light path, while Sun *et al.* [103] recovered the non-uniform radiance distribution by making use of the distribution of light irradiance sampled from a flat reference surface. While it makes the system more compact, the other benefit of the near light configuration is that the radiance fall off in distance from the light source can be used as a cue to estimate depths in the scene [63].

Different from the near point light, recently screen lighting is attracting the attention due to its practical setup (*e.g.*, the ordinal computer screen is available). For instance, Aittala *et al.* [5] recovered spatially-varying reflectance and small geometric variations utilizing portable laptop screen for illumination. They displayed continuous Fourier basis illumination patterns windowed by a Gaussian on the screen and the geometry and reflectance are estimated by inversely solving per-pixel image formation model based on the reflectance lobes approximated as mixtures of Gaussian (MoG) in the illumination screen's domain.

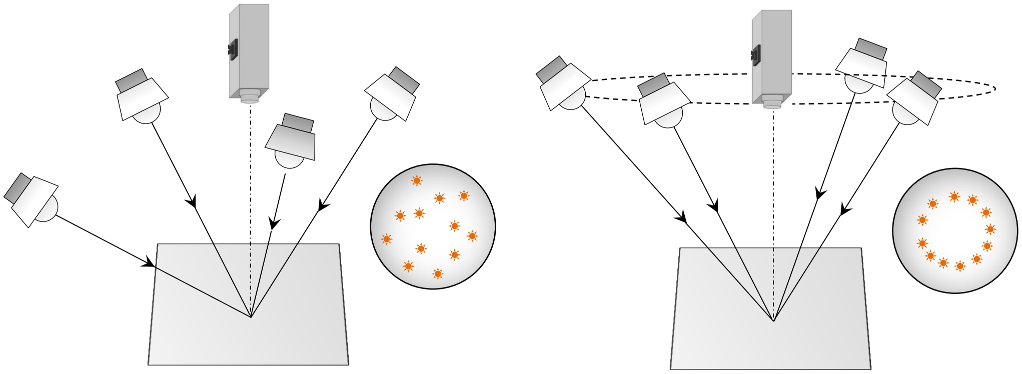
(c) The number of light sources

Besides the shape-from-shading (SFS) which considers only one image [36], a minimum number of light sources required to solve the photometric stereo problem is three (*i.e.*, 1DOF albedo + 2DOF surface normal) for the Lambertian surfaces [122] and if the surface albedos are assumed to be constant over pixels, the number is shrinked to two (*i.e.*, 2DOF surface normal). However, those requirements are easily disrupted in the presence of non-Lambertian corruptions such as shadows, specularities and image noises. Therefore most photometric stereo algorithms postulate a larger number of images is given to take account of the robust least-squares solution or apply a simple thresholding to reject shadows or specularities (Detailed discussions about the robust approach are presented in Section 1.2 and Section 2.2.3). There is, however, a particular situation where only a small number of images are available, therefore some previous works have tried to achieve the stable estimation even though only four images are given [28, 131, 14, 20]. Assuming that three of four observations are represented by a Lambertian model, an outlier could be detected by checking the consistency of all the possible triplets of lights [28, 131, 14]. More recently, Chandraker *et al.* [20] have presented an iterative MRF formulation for detecting shadows and exploiting it as a surface integration constraint. While effective, they still requires a minimum of four lights. On the other hand, Hernández *et al.* [55] have recently proposed a 3-light Lambertian photometric stereo algorithm in the presence of shadows and varying albedo (therefore there is at least 3DOF in the model) by exploiting photometric stereo constraints with only two lights while imposing smooth shape priors.

(d) The configuration of light sources

While a vast majority of photometric stereo algorithms assume that each light source is distributed randomly on the hemisphere centered at the object, the configuration of the light sources, responding to where each light source is placed at, affects the performance of the algorithm. For instance, Drbohlav and Chantler [32] revealed that the optimal light configuration for the three-light calibrated photometric stereo problem is an arbitrary orthogonal triplet which includes the constant slant solution, and that for the calibrated photometric stereo with $n > 3$ light sources should be equally spaced over $360/n$ degrees in its azimuth angle (*i.e.*, uniform distribution), which is extended to the optimal light configuration for the uncalibrated photometric stereo problem [56].

In addition, a special configuration of the light sources often provides additive information which constraints the under-determined photometric stereo problem. For instance, it is known that a ring light configuration where a scene is illuminated by directional lights located on a view centered cone (See Fig. 2-7), resolves the ambiguity of the photometric stereo problem [7, 133, 21]. Alldrin and Kriegman [7] presented that a circle of known light source is enough to detect the symmetry in the emitted radiance function, which is possible to be used for recovering the azimuth angle of any



(a) Random light configuration (b) Ring light configuration

Figure 2-7: Illustration of the light configurations. (a) In the random light configuration, each light source is placed at randomly above the surface. (b) On the other hand, in the ring light configuration, each light source is placed on a circle whose center corresponds to the optical axis of the camera.

isotropic surfaces. Zhout *et al*. [133] exploited a ring-light configuration as a constraint in the uncalibrated photometric stereo for a Lambertian scene, and this theory was extended by Chandraker *et al*. [21] which recovered the surface geometry of general isotropic surfaces by incorporating the spatial and temporal image derivatives of two differential pairs of light sources, at unknown positions on a circle.

(e) The treatment of the ambient illumination

The majority of photometric stereo algorithms generally assume that images are captured in the well-controlled dark room with a unique light source for each image acquisition. However, without the dark room, a complete control of the ambient illumination is difficult, and there exists the ambient illumination in the scene as shown in Fig. 2-8, leading to more complex image formation model than one of the controlled illumination environment. So far, some researches have tackled the photometric stereo problem under the ambient illumination [59, 58, 90, 97, 3, 130]. The simplest strategy to handle ambient illumination is to model the ambient illumination as an additive constant value in the image formation model (called as *the perfect ambient illumination*) and estimate the value with the surface normal [59, 58].

Unfortunately, however, the perfect ambient illumination is rarely observed in the complex natural environment (*e.g*., outdoor scenes), therefore more complex modeling of the ambient illumination is required. Traditionally, Sato and Ikeuchi [90] applied the photometric stereo technique to recover the geometry of outdoor scenes by separating specular and diffuse components as well as sun and skylight contributions. Recently, Shen *et al*. [97] have presented a photometric stereo


(a) Only direct illumination (b) Direct + ambient illumination

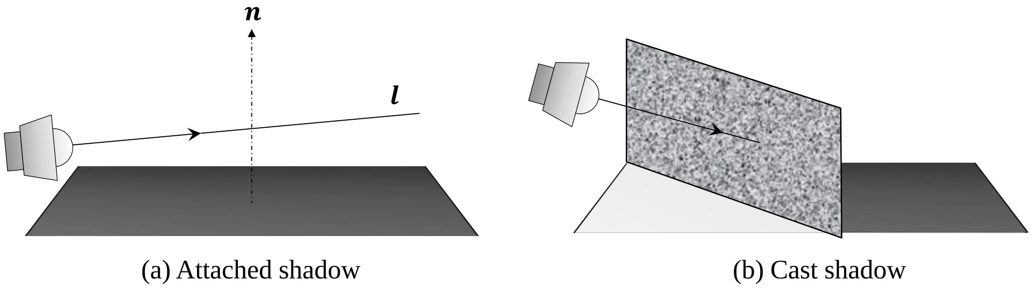
Figure 2-8: Examples of the ambient illumination. Images are captured under (a) a directional illumination with a fixed point light source, or (b) the ambient illumination in addition to the same directional illumination.

algorithm which is capable of uncontrolled, internet images where general lighting is considered as on a distant sphere represented by the spherical harmonics under the Lambertian assumption. While the similar manner of the outdoor illumination modeling is also adopted in [123, 48], they are limited on the Lambertian scenes. Recently, Ackermann *et al*. [3] have proposed a photometric stereo algorithm for general outdoor scenes. They modeled the solar illumination in the similar manner with [90] and represented the non-Lambertian material as a combination of a few basis BRDFs in the similar manner with Goldman *et al*. [46]. While photometric stereo algorithms under the ambient illumination often extend their applications, it is worth mentioning that the perfect separation of the ambient illumination in outdoor images are still difficult especially when the surface is not represented by a simple Lambertian model and/or the indirect illumination complicatedly communicates with the scene.

(f) The treatment of cast/attached shadows

In the natural scene, there observed two different kinds of shadow, that are called as *attached shadow* and *cast shadow* [37]. An attached shadow occurs when a surface curves away from a light source *i.e*., it appears where $\boldsymbol{l}^T\boldsymbol{n} \leq 0$. On the other hand, a cast shadow occur when a non-convex surface blocks a light as illustrated in Fig. 2-9.

Both kinds of shadow generally violate the image formation model underlying the photometric stereo problem. Therefore, a majority of the photometric stereo algorithms consider shadows as outliers and perform shadow detection and removal techniques that are mainly categorized into two classes: one uses the explicit shadow removal pre-processing before the surface normal estimation, and others dynamically neglect shadows by applying robust regression methods such as RANdom SAmple Consensus (RANSAC) [39].



(a) Attached shadow (b) Cast shadow

Figure 2-9: Illustration of attached and cast shadows. (a) An attached shadow appears where $\boldsymbol{n}^T \boldsymbol{l} \leq 0$ and (b) a cast shadow appears where the incident light does not reach because it is blocked by non-convex surfaces.

In the former class, the most traditional yet still widely used technique for avoiding the effects of shadows is intensity thresholding because of its simplicity. To handle spatially varying albedo or sensor saturations, some recent approaches incorporate classification methods for detecting shadows [14, 11, 3], or exploit graph-based visibility estimation for the shadow removal [20].

Though RANSAC based robust approach [79, 54, 105] in the latter class is effective when the number of observations are relatively small, it becomes computationally intractable with a large number of images since a huge number of the random sampling is required to get the stable estimation [7]. On the other hand, Wu and Tang [126] have used an Expectation-Maximization approach starting from the initial surface normal that are obtained from ratio images. Spatial constraints are also exploited to effectively remove outliers. Tang *et al.* [109] use an Markov random field (MRF) for imposing discontinuity preserving smoothness using belief propagation. A similar setting has also been used in [125] where graph cuts is used for deriving the optimal solution. While the MRF-based approaches preserve discontinuities, they tend to over-smooth the surface normal map. Spatial information is also incorporated in filtering approaches [76, 129]. Miyazaki *et al.* [76] use a filter-based method where an inaccurate surface normal is refined via the median estimate of neighboring pixels. Yu *et al.* [129] propose a pixelwise scheme to find a maximum feasible subset of Lambertian observations via a Big-M algorithm.

While most algorithms consider the shadows as *corruptions* (*i.e.*, outliers in the image formation model), some recent works positively take account of shadows for recovering surface geometry of the complex material [26, 82] since shadows are generated from the radiometric interaction between the geometry and the illumination, not affected by the reflections on the surface. For instance, Chung

---

[7] The RANSAC-based outlier rejection algorithms will be evaluated in Section 4.4.

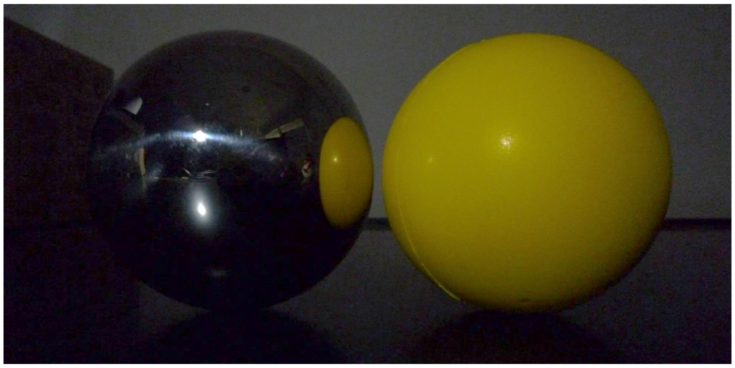


Figure 2-10: Example of interreflections in the scene. When the highly reflective surfaces are included in the scene, the interreflection occurs between these surfaces. Different from other kinds of global illumination (*e.g.*, the ambient illumination), it can not be avoided how carefully the illumination environment is controlled (*e.g.*, images are captured in the dark room).

and Jia [26] detected shadows using a Lambertian board placed behind the object, and estimated the surface normal using cast shadows by accounting shadow boundaries. On the other hand, Okabe *et al*. [82] encoded surface points via attached shadows observed under different light source directions and then estimated surface normals on the basis of the similarity of the attached shadow codes.

(g) The treatment of interreflections

Even though the surface is illuminated by a single light source with no ambient illumination, *interreflections* occur when concavity exists in the scene. Photometric stereo formulations usually do not account for the existence of indirect lighting, thus, producing incorrect shape and reflectance estimates in the presence of interreflections. For tackling this problem, some prior works have tried to separate the interreflections from the direct lighting [33, 60, 80, 94, 67, 52]. For instance, assuming that the target objects have Lambertian surfaces of uniform color, Liao *et al*. [67] separated the inter-reflection from observations by incorporating multi-spectral illumination. A different approach was proposed by Herbort *et al*. [52] that roughly reconstructs a surface with at first unknown indirect reflections and then, refines the initial surface iteratively regarding interreflections.

2.2.2 Camera Setup

The camera setup determines how outgoing radiance from the surface is geometrically and radiometrically recorded on each pixel in the image plane of the optical sensor. As for this component, the photometric stereo algorithms are generally categorized by responding to these five questions as shown in Table 2.2: (a) What is the model of camera projection? (b) Is the camera radiometric

Table 2.2: Summary of the camera setup.

Typical Property	Major Condition	Minor Condition	This Dissertation
Projection Model	orthographic	perspective	orthographic
Response Function	known	unknown	known
Color Space	gray-scale	color	gray-scale
Number of Cameras	one	two+	one

response function known? (c) Is the captured color image converted into a gray-scale? and (d) How many cameras are used to capture images?

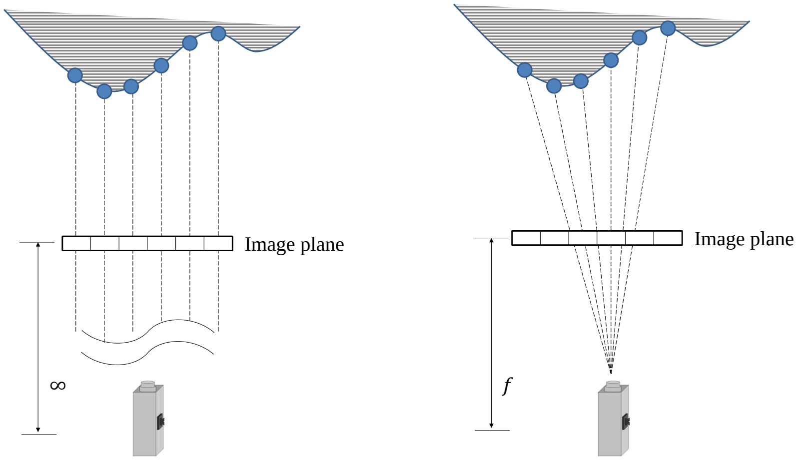
(a) The camera projection model

A majority of photometric stereo algorithms assume that the geometric projection from a 3-D surface point onto a 2-D pixel in the image plane is modeled by the orthographic projection due to its simplicity, which is practically equivalent to capturing the object near the center of the image plane using a lens with a long (theoretically infinite) focal length to capture the scene. However, if an object is placed near the object, the orthographic camera model becomes difficult to be realized. For handling this situation, some recent algorithms have built the image formation model in the photometric stereo problem based on the perspective projection of a pinhole camera [50] instead of the orthographic camera projection, which relates a 3-D point to a 2-D pixel with the non-linear projection [110, 59, 34, 75, 84] as illustrated in Fig. 2-11. The perspective camera model contributes to make the system more compact [59] and often provide an additional information with the photometric stereo problem. For instance, Higo *et al.* [59] presented a hand-held multi-view/photometric stereo system where a near point light source is attached to a hand-held camera to add a photometric constraint to the multi-view stereo problem. On the other hand, Papadhimitri and Pavaro [84] recently proved that the perspective projection model disambiguates the GBR ambiguity (See Section 2.2.1) in the uncalibrated photometric stereo problem.

However, non-linearity in the perspective projection inevitably leads to a complex image formation model as shown in Eq. (2.9). Therefore, the perspective projection of a pinhole camera is hard to be embedded in the plausible image formation model for the photometric stereo problem with more complex BRDF than a simple Lambertian reflectance model, which makes the algorithm less flexible about the target material.

(b) Prior information about the camera response function

The intensity/color values which are recorded on the image sensor is generally transformed in the imaging process, that is modeled by a non-linear function called as *a radiometric camera response*



(a) Orthographic projection (b) Perspective projection

Figure 2-11: Illustrations of the orthographic and the perspective camera models. Theoretically, the orthographic projection of a pinhole camera is realized by using a lens whose focal length is infinity However, practically, it is approximated by a telephoto lens.

function. We show examples non-linearly transformed images in Fig. 2-12. Here we perform a basic gamma correction to the original image ($\gamma = 1$), where an image intensity I is replaced by I^γ. Though the image formation model becomes incorrect without considering this non-linear transformation, most photometric stereo algorithms assume the camera response function is linear by following two options in the practical setups. One uses *camera raw data* which contains minimally processed data from the image sensor before the non-linear transformation is applied. While its simplicity, the raw data is available on limited and usually expensive single-lens reflex digital cameras. The second option is to independently apply the radiometric camera calibration which estimates the radiometric camera response function from images [47, 68, 66]. Once the camera response function is estimated, the image is easily linearized using the inverse of the camera response function.

While effective, the radiometric calibration is not always possible since it involves an additional data measurement. Therefore, some recent work has achieved the estimation of the surface normal under the *unknown* camera response function[98, 77, 2]. However, we note that the estimation of additional unknown variables (*i.e.*, the radiometric camera response function) usually sacrifices the model simplicity, and therefore the simultaneous estimation of the response function and the surface normal of the object whose material is more complex than a Lambertian reflectance model is still problematic.

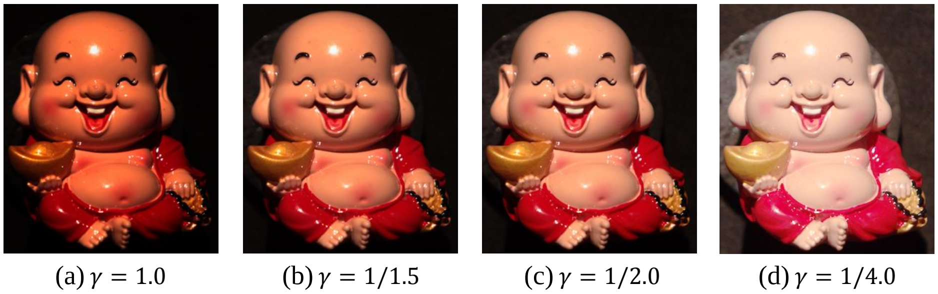

(a) $\gamma = 1.0$ (b) $\gamma = 1/1.5$ (c) $\gamma = 1/2.0$ (d) $\gamma = 1/4.0$

Figure 2-12: Images under varying camera response functions. (a)-(d) We vary the image gamma from 1.0 to $1/4.0$, where an intensity value I is transformed to I^{γ}.

(c) Color space of the image

While the reflectance function and its derived imaging model for a material varies over wavelength, a vast majority of the existing photometric stereo methods deal with gray-scale images since averaging multi-spectral images increases the signal-to-noise ratio of input images, leading to the robust estimation of the surface normal. However, some algorithms try to explicitly use color information in the photometric stereo problem mainly for two reasons.

The first reason is to recover the BRDF of the scene in addition to the surface geometry [46, 9, 5]. For instance, Alldrin *et al*. [9] succeeded to render the image from a noble viewpoint using the non-parametric representation of BRDF, where each cell in the BRDF matrix and the elevation angle of the scene are alternatively estimated with known lightings and azimuth angles of the scene.

The second reason is to achieve the single-shot recovery of the geometry for handling the dynamic scene [31, 10, 44]. While the most traditional color-based photometric stereo assumes that the materials in the scene have constant chromaticity, meaning that the spectral distribution of the surface reflectance varies only by a uniform scale factor, Decker *et al*. [31] relaxed the constant chromaticity restriction, enabled by the addition of time multiplexing. However, this system requires at least two additional frames of input to produce one frame of geometry, halving the temporal resolution of their results. On the other hand, Anderson *et al*. [10] avoided the time multiplexing by presenting a novel calibration technique for multispectral photometric stereo that can be applied to objects with multiple piecewise constant chromacities, and Fyffe *et al*. [44] also relaxed the constant chromaticity restriction, by capturing images with a greater number of spectrally distinct color channels. While effective, the current problem is the color-based photometric stereo often requires more complex light calibration and assumes that the scene contains a small number of distinct chromaticities. Furthermore, the scene is mostly approximated by a Lambertian reflectance model.

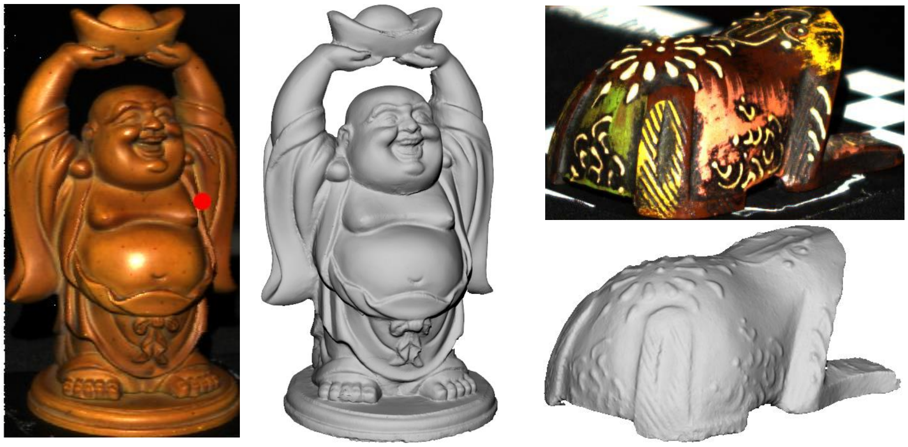


Figure 2-13: Results of multi-view photometric stereo algorithm [134]. Combining multi-view information, the metric reconstruction of the complete surface geometry is achieved. These images are from [134].

(d) Single-view or Multi-view

While images are generally captured by a single static camera in the photometric stereo problem, two and more cameras are often used for recovering the surface normals. They are distinctively called the *multi-view* photometric stereo algorithms [53, 116, 34, 128, 123, 134, 85].

The main advantage of the multi-view photometric stereo algorithms is that it enables the metric reconstruction (recovery of the depth information) of the complete shape with a help of multi-view geometry while the single view photometric stereo techniques recover the surface normals of the front face. For example, Hernández *et al*. [53] used multi-view silhouette and shading cues to recover the complete geometry of the target object lying on a turntable, and Yoshiyasu and Yamazaki [128] extended this work to handle topologically complex object captured by a moving single flash camera. Vlasic *et al*. [116] combined the photometric stereo problem with a binocular stereo of perspective cameras to achieve the metric reconstruction even though its narrow baseline does not allow to recover a full face of the target object. On the other hand, Wu *et al*. [123] and Zhou *et al*. [134] successfully integrated multi-view stereo and the photometric stereo algorithms to acquire the detailed and complete structure of the target object. While effective, the multi-view photometric stereo problem generally requires the feature matching to merge recovered surface normals of cameras based on the multi-view geometry [50], leading to the time-consuming optimization problem and the degeneration of the surface normal quality due to the quantization errors and/or incorrect feature matching. To tackle these difficulties, Park *et al*. [85] have recently presented a framework to utilize the planar mesh parametrization where each 3-D surface point is projected on to the 2-D space based on the 3-D to 2-D image warping algorithm. The key ingredient in the framework is that they merge all viewpoints before the photometric stereo technique is applied, therefore there is no apprehension of incorrect surface normal merging.

While the traditional multi-view photometric stereo techniques are capable of Lambertian surfaces, Zhou *et al*. [134] have recently proposed the multi-view photometric stereo algorithm for handling objects spatially varying isotropic reflectance by constructing a sparse point cloud by the structure-from-motion algorithm and propagating the depth information along with estimated iso-depth contours. However, it is worth mentioning that the multi-view photometric stereo algorithms are theoretically problematic when the appearance of the object is dramatically varying over viewpoints and/or there is no texture on the surface since multi-view integration based on the multi-view feature matching becomes prohibitedly difficult.

2.2.3 Reflectance modeling

Since Woodham [122] firstly introduced the photometric stereo technique for Lambertian scene, the extension of its work to non-Lambertian materials has been a problems of significant interest. Photometric stereo approaches to dealing with non-Lambertian effects are mainly categorized into four classes based on the way of reflectance modeling: (a) Lambertian photometric stereo with outlier rejection, (b) example-based reflectance modeling, (c) reflectance modeling with non-Lambertian BRDF, and (d) reflectance modeling with common BRDF properties.

Lambertian photometric stereo with outlier rejection

A large amount of photometric stereo algorithms recover surface normals of a scene via a simple Lambertian reflectance modeling while non-Lambertian corruptions such as shadows and interreflections are carefully detected and removed by various kind of corruption-specific algorithms or as we have repeatedly mentioned in and Section 2.2.1. While each corruption is reasonably discarded by those techniques, it will become computationally expensive to perform each algorithm independently on real images where various corruptions are complicatedly intermingled. Therefore, some approaches do not distinguish each corruption and reject them all as outliers. The most reasonable strategy may come from RANdom SAmple Consensus (RANSAC) scheme [79, 54, 105], which repeatedly and randomly take samples from observations to find the set of maximum Lambertian inliers. While RANSAC scheme is effective when the number of observations are relatively small, it becomes computationally unstable and expensive with a large number of images as we have already mentioned in Section 2.2.1. We remind that we have also already presented a plenty amount of outlier rejection based algorithms in Section 2.2.1 as ones that robustly reject attached/cast shadows [125, 109, 114, 126, 76, 129], therefore we recommend the reader to refer the section. In addition to them, Wu *et al.* [124] have recently proposed the rank-minimization based approach to decompose observations into the rank-3 Lambertian structure and non-Lambertian corruptions. While the method is stable and theoretically capable of non-Lambertian corruptions such as specular highlights, it is actually problematic in the presence of shadows and sensor satiations. More importantly, it is not guaranteed that the recovered low-rank matrix is derived from the rank-3 Lambertian structure.

In summary, most algorithms which assume underlying Lambertian reflectance model are robust to various non-Lambertian corruptions by rejecting them outliers. Because of simplicity of the model, some of them require neither the initialization nor the complex optimization that are usually required in algorithms which use complex non-Lambertian reflectance models. However, the fundamental restriction is that if it were not for the dense Lambertian structure in observations, the estimation is disrupted.

Example-based reflectance modeling

A few amount of photometric stereo algorithms are distinctively called *example-based* approach, which take advantages of surface reflectance of objects with known shape, captured under the same illumination environment with the target material.

The earliest example-based approach [101] requires a reference object whose material is exactly same with that of target object and the reference object is required to be captured in the same known lighting condition, which considerably limit the domain of capable materials. Recently, Hertzmann *et al.* [57] have eased these restrictions to handle uncalibrated scene and spatially varying materials by assuming that materials in natural world can be expressed as a small number of basis materials. However, it still requires a couple of reference objects whose geometry is known.

Reflectance modeling with non-Lambertian BRDF

Unless using Lambertian reflectance model nor the reference objects, various model-based approaches are applicable, which arrange the parametric or non-parametric models of material reflectance and iteratively optimize either surface normals and model parameters. So far, various physically inspired parametric models of the bidirectional reflectance distribution (BRDF) such as Torrance-Sparrow [112] and Ward [118] have been used in photometric stereo problem to account for specularities [45, 27]. Materials without any Lambertian structure are reasonably addressed in this approach, however, they are incapable of handling spatially varying materials since model parameters are usually optimized over images.

Therefore, in recent years, there has been an emphasis on representing a target material with a small number of fundamental materials (*e.g.*., two) to handle wide variety of spatially varying materials [46, 9]. Goldman *et al.* [46] have approximated each fundamental material by the Ward model [118] and both parameters in fundamental BRDFs and per-pixel weight maps are recovered with surface normals. While effective, their method has restriction on the size of specular lobes and shadows, and its complex and nonlinear structure requires careful tuning of model parameters and initialization. More recently, Alldrin *et al.* [9] have introduced non-parametric representation of the fundamental material and simultaneously recovered geometry and BRDFs by bi-variate approximations of isotropic reflectance functions, which increase computational stability. This algorithm reduces some restrictions in [46], however it also requires shadow-free observations and azimuth angle maps acquired in advance. While model-based approaches are potentially capable of wide variety of materials, the high-dimensional ill-posed problem and complex optimization scheme cause the instability and inefficiency of the estimation. This algorithm is later extended by [134] to achieve the multi-view photometric stereo algorithm for non-Lambertian scenes as we have already presented in Section 2.2.2.

To simplify the model and optimization, Shi *et al.* [99] proposed a compact biquadratic representation of isotropic BRDF. While the work does not require any complex parameter tuning and global optimization, however, it assumes that both high-frequency and shadowed observations are removed.

Reflectance modeling with common BRDF properties

The last group in the non-Lambertian photometric stereo neither uses reference objects nor explicitly models material reflectance model, instead take advantages of properties of general reflectance shared among real materials which are being revealed by comprehensive analysis [107, 23]. The main advantage of using those shared properties is that the algorithms are capable of handling wide varieties of scene since any parametric reflectance model is not required to solve the photometric stereo problem. One attractive property is isotropy and reciprocity symmetry which appears in isotropic materials [74], which is widely exploited surface normal reconstruction [106, 7, 58]. Monotonicity constraints on general reflectance is also used in some works [58, 100]. Other than these properties, radiance similarity [89], image deviates [21], diffusive maxima [38] and the half

Table 2.3: Definitions of major symbols used in the dissertation.

Symbol	Name
m	Number of images
$\boldsymbol{n}$	Surface normal direction (unit vector)
$\boldsymbol{l}$	Incoming lighting direction (unit vector)
L	Lighting matrix ($L \triangleq [\boldsymbol{l}_1, \cdots, \boldsymbol{l}_m]$)
$\boldsymbol{v}$	Outgoing viewing direction (generally fixed by $[0\ 0\ 1]^T$)
I	Image intensity (three channels of r,g,b values are converted into a gray-scale value)
$\boldsymbol{o}$	Intensity vector for a surface point ($\boldsymbol{o} \triangleq [I_1, \cdots, I_m]$)
ρ	Isotropic BRDF for 3 variables θ_i, θ_o, $\|\phi_i - \phi_o\|$
f	Reflectance function ($f \triangleq \rho(\theta_i, \theta_o, \|\phi_i - \phi_o\|)(\boldsymbol{n}^T\boldsymbol{l})$)
ϵ	Corruption in the image formation model *i.e.*, $I = f + \epsilon$

vector symmetry of isotropic BRDF [127] are also exploited to add further constraint on the ill-posed problem and increase the accuracy of surface normal recovery.

In summary, most algorithms in this category can recover surface normals of complex materials even though there is no underlying Lambertian structure. Furthermore, sum of them can also recover detailed material reflectance as well as the geometry of target. However, many algorithms in this category require either/both an initial normal estimate which is usually recovered by Lambertian photometric stereo, tuning of a large number of parameters and a complex non-linear optimization, where even one of them may cause instability of the estimation. In addition, most of them are incompetent for effects such as shadows and image noises which are not included in the model.

2.3 Definitions and Assumptions

Before stepping into the next chapter, we summarize the major symbols which will appear in this dissertation in Table 2.3 and indicate the assumptions that we henceforth rely on in the dissertation:

1. Object is illuminated by a point light source at infinity from varying and known directions, and no indirect light source exists.
2. The light entered a surface point is reflected out from the same point, without any transmission or subsurface scattering.
3. Camera view is orthographic, and the radiometric response function is known (and linear).
4. Relative position between the camera and the object is fixed across all images.

Chapter 3

Robust Lambertian Photometric Stereo Using Sparse Regression

In this chapter, an alternative photometric stereo algorithm is presented, for stably and accurately estimating the surface normals of a scene in the presence of various non-Lambertian effects.

In the photometric stereo literature, an algorithm is generally categorized into the *robust* approach when the surface normal and other model parameters are recovered using an estimator that is robust to outliers. Given sufficient amount of inliers, algorithms in this approach potentially work for any kind of outliers even when the data distribution of each corruption is unknown. There have been many robust algorithms that assume a basic Lambertian model but augmented with outlier detection for handling all non-Lambertian regions of the scene [73, 79, 129, 105, 124] (Details were presented in Section 2.2.3). While this strategy is numerically stable and relatively insensitive to initializations, it may be computationally expensive since a large number of images is required for robust outlier rejection.

In this chapter and Chapter 4, we will prove a formulation of the photometric stereo problem based on the alternative robust regression technique called *sparse regression*, where a sparsity penalty is applied to the corruption ϵ in the image formation model Eq. (1.2) to reasonably disambiguate the problem of estimating both surface normal and model corruptions simultaneously. Though the concept of the sparse regression is general and not restricted by the form of the regression function, there are many computational issues since its derivative optimization problem includes a difficult ℓ_0 minimization, which becomes almost intractable in the presence of non-linearity on unknown variables. Therefore, in this chapter, we will begin by assuming a basic Lambertian

BRDF in the image formation model and considering other non-Lambertian effects as outliers as

$$I_i = \rho \boldsymbol{n}^T \boldsymbol{l}_i + \epsilon_i \quad i = 1, \cdots, m. \tag{3.1}$$

Given m observations, the goal is to recover ρ, $\boldsymbol{n}$ and ϵ_i by inversely solving Eq. (3.1).

For this purpose, a hierarchical Bayesian model is developed that automatically decomposes observed appearances into a Lambertian diffuse component and a sparse, non-Lambertian component for capturing shadows, specularities, and other corruptions. Optimization and inference is accomplished using a robust majorization-minimization technique akin to the popular EM algorithm, with desirable convergence properties and quantifiable advantages over standard convex estimators.

Of course we do not know a priori where non-Lambertian appearances will be located, and thus the overall estimation problem is underconstrained even with many available images. Thus, our hierarchical Bayesian model attempts to maximize the number of observations that can be adequately explained via the Lambertian diffuse reflectance function (the inlier model) while treating the remaining observations, including specularities, shadows, and sensor saturations, as non-Lambertian elements with separate, unknown variances learned from the data (the outlier model). The partitioning between inliers and outliers is estimated simultaneously with the normal vectors and surface albedos using a principled variational Bayesian technique.

The proposed framework benefits from simple, efficient pixelwise optimization, which is easily amenable to parallel processing. Moreover it does not require the pre-processing of specularities/shadows, careful initialization strategies, or typical smoothness constraints for both object structure and reflectance, which can disrupt the recovery of fine details. Consequently, we do not suffer from numerical instability even with relatively few input images (*e.g.*, 20) and we do not have difficulty handling spatially varying albedos with high-frequency structures. Extensive experiments in Section 3.2 show that our implementation produces more stable, accurate, and efficient surface normal estimates than other robust algorithms such as [79, 124].

3.1 Modeling Lambertian-based Image Formation with Outliers

Woodham et al. [122] revealed that the intensity I of a point in a Lambertian scene under a lighting direction $\boldsymbol{l} \in \mathbb{R}^3$ is expressed as follow,

$$I = \rho \, \boldsymbol{n}^T \boldsymbol{l}, \tag{3.2}$$

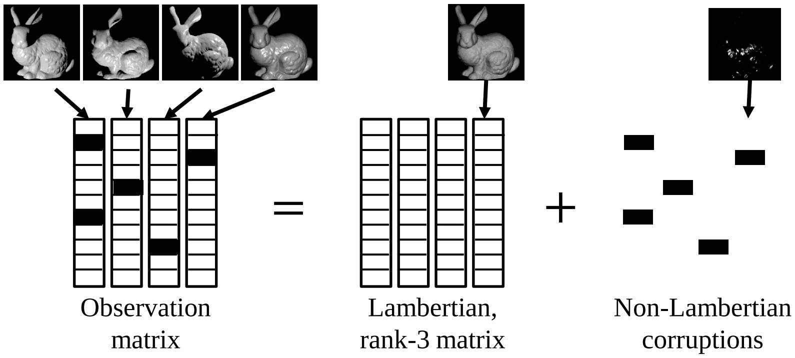


Figure 3-1: Illustration of the image formation of Lambertian scenes. An observation matrix O is composed of the rank-3 Lambertian matrix ($D = N^T L$) and the non-Lambertian corruption matrix E.

where ρ is the diffuse albedo, and $\boldsymbol{n} \in \mathbb{R}^3$ is the surface normal at the point. This image formation model is equivalent to the case where the BRDF is a constant value *i.e.*, $\rho(\theta_i, \theta_o, \|\phi_i - \phi_o\|) = \rho$. We should note that the actual Lambertian reflectance model accounts the intensity of a light, which is omitted here since it is assumed that the appearances are normalized by the intensity of lightings as Eq. (2.10).

Given m images with $k \triangleq h \times w$ pixels, we define an observation matrix by aligning each image as a vector:

$$O \triangleq [(\boldsymbol{I}_1)| \dots |(\boldsymbol{I}_m)] \in \mathbb{R}^{k \times m}, \tag{3.3}$$

where $\boldsymbol{I}_p \triangleq [I_p(1), ..., I_p(k)]^T$ for $p = 1, ..., m$, and

$$I_p(j) = \rho_j \ \boldsymbol{n}_j^T \boldsymbol{l}_p. \tag{3.4}$$

Therefore, the observations in a Lambertian scene can be expressed via the rank-3 expression

$$O = N^T L, \tag{3.5}$$

where $N = [\rho_1 \boldsymbol{n}_1 \ | \dots | \ \rho_k \boldsymbol{n}_k] \in \mathbb{R}^{3 \times k}$ and $L = [\boldsymbol{l}_1 \ | \dots | \ \boldsymbol{l}_m] \in \mathbb{R}^{3 \times m}$. This rank-3 structure of the Lambertian appearance is solid since it does not depend on neither the direction of incident lighting nor the surface geometry. In real scenes, however, various effects beyond the Lambertian formulation are observed, *e.g.*, specularities, shadows, image noise and so on, which destroy the

Lambertian rank-3 structure. In the robust approach, these effects are interpreted as additive corruptions $E \in \mathbb{R}^{m \times n}$ applied to an otherwise ideal Lambertian scene leading to the image formation model as

$$O = N^T L + E. \tag{3.6}$$

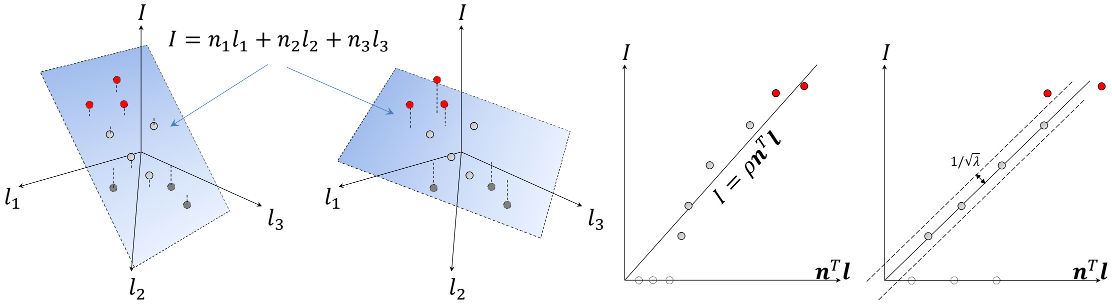
Given observed images (O) and lighting directions (L), the goal is to recover surface normals (N) as a part of the Lambertian diffusive component ($N^T L$) in the presence of non-Lambertian corruptions (E). For reference, we illustrate the image formation of the Lambertian scene in Fig. 3-1.

3.1.1 Robust Photometric Stereo Using Sparse Regression

Eq. (3.6) is an under-constrained problem since the number of unknowns exceeds the number of linear equations *i.e.*, assuming the varying surface albedos, the total unknown variables are $3k + mk$ (*i.e.*, $3k$ DOF in N and mk DOF in E). While most previous methods deal with corruptions by applying traditional outlier removal techniques assuming the type of them and then recovering surface normals from the resulting purified Lambertian component [79, 129, 76], the distribution-dependent approach is always in danger of removing inliers as well as outliers due to lack of prior knowledge about the potential outlier distribution. Alternatively, we try to recover N without explicitly removing corruptions in a separate step with a help of the sparse regression technique. An essential ingredient is a sparsity penalty applied to E, whose minimization disambiguates the infinity of feasible solutions to Eq. (3.6). This penalty quantifies the reasonable observation that non-Lambertian effects emerge primarily in limited areas of each image. For example, specularities surround the spot where the surface normal is oriented halfway between lighting and viewing directions, while shadows are created only when $L^T N \leq 0$ (attached shadow) or when a non-convex surface blocks the light (cast shadow). Strictly speaking, we assume that the optimal feasible solution to Eq. (3.6) produces a sparse error matrix. Reflecting this assumption, our estimation problem can be formulated as

$$\min_{N,E} \|E\|_0 \ \text{ s.t. } \ O = N^T L + E. \tag{3.7}$$

Here, $\|\cdot\|_0$ is an ℓ_0-norm penalty, which counts the number of non-zero entries in the matrix. To reiterate, Eq. (3.7) builds on the assumption that images are captured under known lighting conditions and any non-Lambertian corruptions have sparse structure. If these assumptions are not true (*e.g.*, because of imperfect lighting calibration, non-sparse specularities, etc.), then the hard constraint in Eq. (3.7) is no longer appropriate. To compensate for more diffuse modeling errors,



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

Figure 3-2: Illustration of the sparse regression. (a) ℓ_2-based regression gives inaccurate plane fitting in 4-D space of $[l_1, l_2, l_3, I]$, however (b) ℓ_0-based regression successfully finds the plane by reasonably neglecting outliers. (c) The Lambertian-based sparse regression is also illustrated as a regression analysis in 2-D space of $[\boldsymbol{n}^T\boldsymbol{l}, I]$ whose regression function is represented as $y = \rho x$. (d) the soft constraint in Eq. (3.12) is considered as a margin where a data point is considered as lying on the function".

we relax the hard constraint via an additional model mismatch penalty giving

$$\min_{N,E} \|O - N^T L - E\|_2^2 + \lambda \|E\|_0, \tag{3.8}$$

where λ is a nonnegative trade off parameter balancing data fit with sparsity. Note that in the limit as $\lambda \to 0$, the two problems are equivalent (the limit must be taken outside of the minimization).

Since Eq. (3.7) and Eq. (3.8) decouples, we can consider instead an equivalent series of separate, pixel-wise optimizations problems of the canonical form

$$\min_{\boldsymbol{n},\boldsymbol{e}} \|\boldsymbol{e}\|_0 \text{ s.t. } \boldsymbol{I} = L^T\boldsymbol{n} + \boldsymbol{e}, \tag{3.9}$$

$$\min_{\boldsymbol{n},\boldsymbol{e}} \|\boldsymbol{I} - L^T\boldsymbol{n} - \boldsymbol{e}\|_2^2 + \lambda\|\boldsymbol{e}\|_0 \ . \tag{3.10}$$

where the column vector $\boldsymbol{I}$ denotes an arbitrary transposed row of O, $\boldsymbol{n}$ is the associated unknown normal vector, and $\boldsymbol{e}$ is the sparse error component *i.e.*, $\boldsymbol{e} \triangleq [\epsilon_1, \cdots, \epsilon_m]$ (we have omitted pixel-wise subscripts for simplicity). For intuitively understanding Eq. (3.10), we equivalently express it in the similar manner with Eq. (1.3) as

$$\min_{\boldsymbol{n},\boldsymbol{e}} \|\boldsymbol{e}\|_0 \text{ s.t. } \forall i \ \ I_i = \boldsymbol{n}^T\boldsymbol{l}_i + \epsilon_i, \tag{3.11}$$

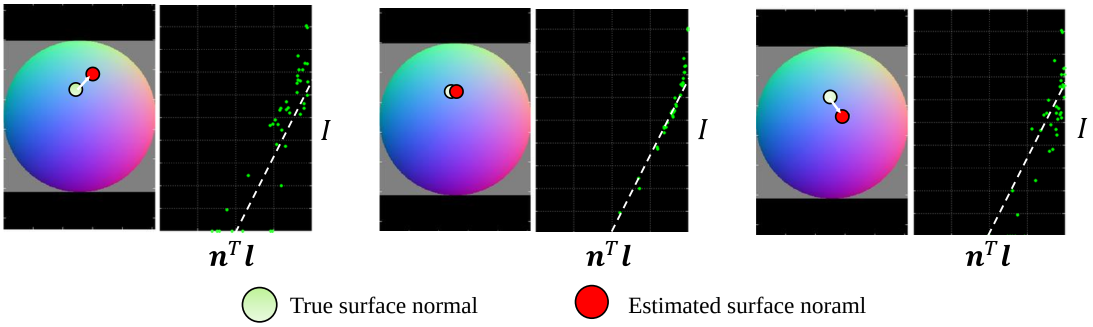


Figure 3-3: Illustration of 2-D plots with varying surface normal.

$$\min_{\boldsymbol{n},\boldsymbol{e}} \sum_{i=1}^{m} \|I_i - \boldsymbol{n}^T \boldsymbol{l}_i - \epsilon_i\|_2^2 + \lambda \|\boldsymbol{e}\|_0 \; . \tag{3.12}$$

What do these transformation tell as? To answer the question, we illustrate the concept the sparse regression analysis in Fig. 3-2. As was mentioned in Section 1.2, the standard Lambertian photometric stereo problem (without a ℓ_0 penalty) can be viewed as the regression analysis which finds the relationship between observation I and unknown model parameters ($\boldsymbol{n} = [n'_1, n'_2, n'_3]^T$) by minimizing the sum of ℓ_2 residual between the observation and estimated model output $f(l_1, l_2, l_3) = n'_1 l_1 + n'_2 l_2 + n'_3 l_3$. It is equivalent to finding a plane in the four dimensional space of l_1, l_2, l_3, I such that the sum of ℓ_2 distance between the plane and each 4-D plots $[l_1^i, l_2^i, l_3^i, I_i]$ is minimized (See Fig. 3-2-(a)). On the other hand, solving Eq. (3.11) is equivalent to finding a plane such that as much as possible 4-D plots $[l_1^i, l_2^i, l_3^i, I_i]$ are *exactly* lying on the regression plane (See Fig. 3-2-(b)). In other words, the model parameters of the regression plane is estimated such that the number of outliers (*i.e.*, deviations from the regression model) is minimized, that is why this kind of regression is called a *sparse* regression. It will make sense to state that Eq. (3.11) and Eq. (3.12) are also viewed as a regression analysis to find the linear function of $y = \rho x$ in the 2-D space whose coordinates are $\boldsymbol{n}^T \boldsymbol{l}$ and I (See Fig. 3-2-(c),(d)). Note that in this regression, both the slant of a linear function ρ and the position of 2-D data $[\boldsymbol{n}^T \boldsymbol{l}, I]$ are varying over unknown variables. Here we can consider that the soft constraint of Eq. (3.12) allows each 2-D plot of $[\boldsymbol{n}^T \boldsymbol{l}, I]$ not to exactly lying on the linear function (the sparsity constraint is relaxed by a merging $1/\sqrt{\lambda}$). The concrete examples of 2-D plots of $[\boldsymbol{n}^T \boldsymbol{l}, I]$ in the presence of some non-Lambertian corruptions are illustrated in Fig. 3-3. We observe that if the surface normal is close to the ground truth, many samples are lying on the estimated linear model (*i.e.*, many observations are represented by a Lambertian model).

Eq. (3.9) and Eq. (3.10) entail difficult, combinatorial optimization problems that must be efficiently solved at every pixel. Here we consider two alternatives to brute force exhaustive search. First, in the machine learning and statistics literature, it is common to replace the discontinuous, non-convex ℓ_0 norm with the convex surrogate ℓ_1 norm. (The ℓ_1 norm of a vector $\boldsymbol{z}$ is given by $\sum_i |z_i|$, which constitutes the tightest convex approximation to the ℓ_0 norm.) In certain situations the resulting estimate will closely match the solution to Eq. (3.7) and/or Eq. (3.8); however, in the context of photometric stereo this substitution may not always be adequate (see Section 3.1.4 for more details). Secondly, we can apply a simple hierarchical Bayesian approximation to estimate $\boldsymbol{n}$ while simultaneously accounting for $\boldsymbol{e}$. This formulation, called SBL, is described in detail next.

3.1.2 Recovery of normals and corruptions via iterative reweighted ℓ_1 estimator

Recently, there has been growing interest in finding sparse signal representation from redundant dictionaries solving

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_0. \tag{3.13}$$

Here, Φ is a matrix whose columns represent an overcomplete or redundant basis, $\boldsymbol{x}$ is a vector of unknown coefficients to be learned, $\boldsymbol{y}$ is the signal vector, and λ is a trade-off parameter which is balancing the diffuse and sparse terms. While an exhaustive search for the optimal solution is prohibitively expensive, recently many sparse approximation algorithms have been proposed, which rely on iterative reweighting schemes that produce more focal estimates as optimization progress [18, 24, 30] (a comprehensive survey is conducted by Wipf and Nagarajan [119]). In the iterative scheme, the $(k+1)$-th iteration is given by,

$$\boldsymbol{x}^{(k+1)} \rightarrow \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\sum_i \omega_i^{(k)} x_i^2 = W^{(k)}\Phi^T\left(\lambda I + \Phi W^{(k)}\Phi^T\right)^{-1}\boldsymbol{y}, \tag{3.14}$$

where $W^{(k)}$ is a diagonal weighting matrix from the k-th iteration with i-th diagonal element $1/\omega_i^{(k)}$ that is potentially a function of all $\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k)}$. In the iterative reweighting schemes, different methods are distinguished by the choice of $\omega^{(k)}$, which determines the surrogate cost function for enforcing sparsity that is being minimized. One such variant is presented by Candés *et al.* [18], where an iterative ℓ_1 reweighting update rule of ω_i is provided as

$$\omega_i^{(k+1)} \rightarrow \left[|x_i^{(k+1)}| + \epsilon\right]^{-1}, \tag{3.15}$$

where ϵ is generally chosen as a fixed (zero in our experiment). At a glance, the form of Eq. (3.10) and Eq. (3.13) look different, however Eq. (3.13) can be derived from Eq. (3.10) defining $\Phi = [L^T \ I_{(m)}]$, $\boldsymbol{x} = [\boldsymbol{n}^T \ \boldsymbol{e}^T]^T$ and $\boldsymbol{y} = \boldsymbol{I}$, where $I_{(p)}$ is a p dimensional identity matrix. Note that since only the sparsity of $\boldsymbol{e}$, not $\boldsymbol{n}$ should be enforced, the weighting matrix W is defined as $\text{diag}[W_n, W_e]$, where the weighting matrix for $\boldsymbol{n}$ is fixed as $W_n \triangleq \alpha I_{(3)}$ (sufficiently large α allows $\boldsymbol{n}$ not to be zero without penalty). In contrast, $W_e \triangleq \text{diag}[\boldsymbol{\omega}^{-1}]$ is a fully-parameterized, diagonal matrix, where $\boldsymbol{\omega} \triangleq [\omega_1, ..., \omega_k]^T$ is a non-negative vector of variances in one-to-one correspondence with elements of $\boldsymbol{e}$ (sufficiently large ω_i enforces the sparsity of x_i). Then, the update rules of $\boldsymbol{n}$ and $\boldsymbol{e}$ are now represented as follow

$$\begin{aligned} \omega_i^{(k+1)} &\rightarrow |\epsilon_i^{(k)}|^{-1}, \ \forall i, \quad W_e^{(k+1)} = \text{diag}[\boldsymbol{\omega}^{(k+1)^{-1}}], \\ \begin{bmatrix} \boldsymbol{n}^{(k+1)} \\ \boldsymbol{e}^{(k+1)} \end{bmatrix} &\rightarrow \begin{bmatrix} \alpha L \\ W_e^{(k+1)} \end{bmatrix} \left(\lambda I_{(m)} + \alpha L^T L + W_e^{(k+1)}\right)^{-1} \boldsymbol{I}. \end{aligned} \tag{3.16}$$

Note that only ω_i and $\boldsymbol{e}$ need to be updated until the convergence and then a fixed point of $\boldsymbol{e}$ gives $\boldsymbol{n}$. These expressions require the calculation of a $m \times m$ inverse matrix, whose computational complexity is proportional to m. Fortunately, the computation of inverse matrix becomes independent from m by taking advantage of the matrix inversion lemma [113]. Recall that the Woodbury matrix identity is represented as

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}, \tag{3.17}$$

where A, U, C and V all denote matrices of the correct size. Comparing Eq. (3.16) and Eq. (3.17), the matrix inversion lemma is applicable to Eq. (3.16) by associating $A = \lambda I_{(m)} + \Gamma_e^{(k+1)}$, $U = L^T$, $C = \alpha I_{(3)}$ and $V = L$. Then update rules are transformed as

$$\begin{aligned} \omega_i^{(k+1)} &\rightarrow |e_i^{(k)}|^{-1}, \forall i, \quad W_e^{(k+1)} = \text{diag}[\boldsymbol{\omega}^{(k+1)^{-1}}], \\ \Psi^{(k+1)} &\rightarrow \left(\lambda I_{(m)} + W_e^{(k+1)}\right)^{-1}, \\ \begin{bmatrix} \boldsymbol{n}^{(k+1)} \\ \boldsymbol{e}^{(k+1)} \end{bmatrix} &\rightarrow \begin{bmatrix} \alpha L \\ \Gamma_e^{(k+1)} \end{bmatrix} \left(\Psi^{(k+1)} - \Psi^{(k+1)} L^T \left(\alpha^{-1} I_{(3)} + L \Psi^{(k+1)} L^T\right)^{-1}\right). \end{aligned} \tag{3.18}$$

In these expressions, the cost of the matrix inversion changes $O(m)$ to $O(3)$ since $\left(\alpha^{-1} I_{(3)} + L\Psi^{(k+1)} L^T\right) \in \mathbb{R}^{3\times 3}$. Note that the update of $\Psi^{(k+1)} \in \mathbb{R}^{m\times m}$ is not computationally expensive since both $I_{(m)}$

and $\Gamma_e^{(k+1)}$ are diagonal matrices.

3.1.3 Recovery of normals and corruptions via SBL estimator

An alternative penalty emerges from a dual-space view [121] of sparse Bayesian learning(SBL) [111], which is based on the notion of automatic relevance determination (ARD) [81]. SBL assumes the standard Gaussian likelihood function for the first-level, diffuse errors giving

$$p(\boldsymbol{I}|\boldsymbol{n},\boldsymbol{e}) = N(\boldsymbol{I}; L^T\boldsymbol{n} + \boldsymbol{e}, \lambda I), \tag{3.19}$$

We next apply an independent, zero-mean Gaussian prior distributions on both $\boldsymbol{n}$ and $\boldsymbol{e}$:

$$p(\boldsymbol{n}) = N(\boldsymbol{n}; \mathbf{0}, \Sigma_n), \quad p(\boldsymbol{e}) = N(\boldsymbol{e}; \mathbf{0}, \Gamma). \tag{3.20}$$

Σ_n describes the prior variance of the unknown normal vector $\boldsymbol{n}$ as $\Sigma_n = \sigma_n^2 I_{(3)}$; they are fixed to convey our lack of *a priori* certainty about $\boldsymbol{n}$. Thus the prior on $\boldsymbol{n}$ will be relatively uninformative. The values of σ_n^2 will be discussed further below. In contrast, $\Gamma \triangleq \mathrm{diag}[\boldsymbol{\gamma}]$ is a fully-parameterized, diagonal matrix, where $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_{m+1}]^T$ is a non-negative vector of variances in one-to-one correspondence with elements of $\boldsymbol{e}$. A large variance γ_i indicates that the corresponding ϵ_i is free to reflect the data, compensating for non-Lambertian effects (outliers), while a small or zero-valued variance implies that the associated error term is constrained near zero (inliers).

Combining the likelihood and prior using Bayes' rule leads to the posterior distribution $p(\boldsymbol{n},\boldsymbol{e}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\boldsymbol{n},\boldsymbol{e})p(\boldsymbol{n})p(\boldsymbol{e})$. To estimate $\boldsymbol{n}$, we may further marginalize over $\boldsymbol{e}$ to give

$$p(\boldsymbol{n}|\boldsymbol{I}) = \int p(\boldsymbol{n},\boldsymbol{e}|\boldsymbol{I}) d\boldsymbol{e} = N(\boldsymbol{n}; \mu, \Sigma), \tag{3.21}$$

with mean and covariance defined as

$$\begin{aligned} \mu &= \Sigma A^T \left(\Gamma + \lambda I\right)^{-1} \boldsymbol{I}, \\ \Sigma &= \left[\Sigma_x^{-1} + A^T \left(\Gamma + \lambda I_{(m)}\right)^{-1} A\right]^{-1}. \end{aligned} \tag{3.22}$$

Here, $A = L^T$. We now have a convenient closed-form estimator for $\boldsymbol{x}$ given by the posterior mean. The only issue then is the values for the unknown parameters Γ. Without prior knowledge as to the locations of the sparse errors, the empirical Bayesian approach to learning Γ is to marginalize the full joint distribution over all unobserved random variables, in this case $\boldsymbol{x}$ and $\boldsymbol{e}$, and then maximize the resulting likelihood function with respect to Γ [111]. Equivalently, we will minimize

$$\begin{aligned} L(\Gamma) &\triangleq -\log \int p(\boldsymbol{I}|\boldsymbol{n},\boldsymbol{e})p(\boldsymbol{n})p(\boldsymbol{e})d\boldsymbol{n}d\boldsymbol{e} \\ &\equiv \log|\Sigma_I| + \boldsymbol{I}^T\Sigma_I^{-1}\boldsymbol{I} \\ &\quad \text{with} \quad \Sigma_I \triangleq A\Sigma_x A^T + \Gamma + \lambda I_{(m)}, \end{aligned} \tag{3.23}$$

with respect to Γ. While $L(\Gamma)$ is non-convex, optimization can be accomplished by adapting a majorization-minimization approach from [119] to the photometric stereo problem. This technique essentially involves the construction of rigorous upper bounds (see below) on each of the two terms in Eq. (3.23) using auxiliary variables $\boldsymbol{z} \triangleq [z_1, \ldots, z_{m+1}]^T$ and $\boldsymbol{u} \triangleq [u_1, \ldots, u_{m+1}]^T$ (the EM algorithm can be viewed as a special case). For fixed values of $\boldsymbol{z}$ and $\boldsymbol{u}$, a closed form solution for Γ exists. Likewise, for a fixed value of Γ, the auxiliary variables can be updated in closed form to tighten the upper bound around the current Γ estimate.

While some details are omitted for brevity, using results from convex analysis it can be shown that for all $\boldsymbol{u} \geq 0$,

$$\begin{aligned} \log|\Sigma_y| &= \log|\Gamma| + \log\left|A\Sigma_x A^T + \lambda I_{(m)}\right| \\ &\quad + \log\left|\Gamma^{-1} + \left(A\Sigma_x A^T + \lambda I_{(m)}\right)^{-1}\right| \\ &\leq \log|\Gamma| + \log\left|A\Sigma_x A^T + \lambda I_{(m)}\right| \\ &\quad + \sum_i \frac{u_i}{\gamma_i} - h^*(\boldsymbol{u}) \\ &\equiv \sum_i \left(\frac{u_i}{\gamma_i} + \log\gamma_i\right) - h^*(\boldsymbol{u}), \end{aligned} \tag{3.24}$$

where $h^*(\boldsymbol{z})$ denotes the concave conjugate function [16] of $h(\boldsymbol{\beta}) \triangleq \log\left|\text{diag}[\boldsymbol{\beta}] + \left(A\Sigma_x A^T + \lambda I_{(m)}\right)^{-1}\right|$ and we have removed irrelevant factors independent from $\boldsymbol{u}$ or $\boldsymbol{\gamma}$. It can be shown that equality (and therefore the minimum of the right-hand side) is obtained in Eq. (3.24) if and only if

$$\boldsymbol{u} = \text{diag}\left[\Gamma^{-1} + \left(A\Sigma_x A^T + \lambda I_{(m)}\right)^{-1}\right]^{-1}. \tag{3.25}$$

In a somewhat related fashion, the second term in $L(\Gamma)$ can be upper-bounded via

$$\begin{aligned} \boldsymbol{I}^T \Sigma_I^{-1} \boldsymbol{I} &\leq (\boldsymbol{I} - \boldsymbol{z})^T \left(A \Sigma_n A^T + \lambda I_{(m)} \right)^{-1} (\boldsymbol{I} - \boldsymbol{z}) \\ &\quad + \sum_i \frac{z_i^2}{\gamma_i}, \end{aligned} \tag{3.26}$$

which holds for all $\boldsymbol{z}$, with equality if and only if

$$\boldsymbol{z} = \Gamma \Sigma_I^{-1} \boldsymbol{I}. \tag{3.27}$$

Now with $\boldsymbol{z}$ and $\boldsymbol{u}$ fixed, the overall upper bound on $L(\Gamma)$ decouples and we can solve for each γ_i individually by collecting the $\boldsymbol{\gamma}$-dependent terms from Eq. (3.24) and Eq. (3.26), leading to the problem

$$\min_{\gamma_i \geq 0} \frac{z_i^2 + u_i}{\gamma_i} + \log \gamma_i, \tag{3.28}$$

which has a simple closed-form solution.

Combining all of the above, and using matrix inversion formula to produce numerically efficient computations, produces update rules for the $(k+1)$-th iteration given by

$$\begin{aligned} \gamma_i^{(k+1)} &\leftarrow \left(z_i^{(k)} \right)^2 + u_i^{(k)}, \forall i, \quad \Gamma^{(k+1)} = \text{diag}[\boldsymbol{\gamma}^{(k+1)}] \\ \boldsymbol{z}^{(k+1)} &\leftarrow \Gamma^{(k+1)} \left(S^{(k+1)} \right)^{-1} \boldsymbol{y} \\ \boldsymbol{u}^{(k+1)} &\leftarrow \text{diag} \left[\Gamma^{(k+1)} - \left(\Gamma^{(k+1)} \right)^2 \left(S^{(k+1)} \right)^{-1} \right], \end{aligned} \tag{3.29}$$

where $S^{(k+1)}$ is computed via

$$\begin{aligned} S^{(k+1)} &= D - DA \left[\Sigma_x^{-1} + A^T D A \right]^{-1} A^T D \\ &\quad \text{and} \quad D \triangleq (\Gamma^{(k+1)} + \lambda I_{(m)})^{-1}. \end{aligned} \tag{3.30}$$

These expressions only require $O(n)$ computations and are guaranteed to reduce $L(\Gamma)$ until a fixed point Γ_* is reached. This value can then be plugged into Eq. (3.22) to estimate the unknown normal vector and model parameters. We denote this point estimator as $\boldsymbol{x}_{sbl}$. Here it is worth noting that the inverse matrix in Eq. (3.30) can also be efficiently computed by using the matrix inversion lemma in the similar manner with Eq. (3.18). If the variances Γ_* reflect the true profile of the sparse errors, then $\boldsymbol{x}_{sbl}$ will closely approximate the true surface normal. This claim will be quantified more explicitly in the next section.

We have thus far omitted details regarding the choice of λ, α and σ_n^2. The first one can be reasonably set according to our prior expectations regarding the magnitudes of diffuse modeling errors, but in practice there is considerable flexibility here since some diffuse errors will be absorbed into $\boldsymbol{e}$. In contrast, we can realistically set $\alpha, \sigma_n^2 \to \infty$, which implies zero precision or no prior information about the normal vectors and yet still leads to stable, convergent update rules. However we have observed that on certain problems a smaller selection for α, σ_n^2 can lead to a modest improvement in performance, presumably because it has a regularizing effect that smoothes the cost function and improves the convergence path of the update rules from Eq. (3.29) (perhaps counterintuitively, in certain situations it does not alter the globally optimal solution as discussed below). It is also possible to learn σ_n^2 using similar updates to those used for Γ, but this introduces additional complexity and does not improve performance.

3.1.4 Analytical evaluation

Previously we discussed two tractable methods for solving Eq. (3.10): a convex ℓ_1-norm-based relaxation and a hierarchical Bayesian model called SBL. This section briefly discusses comparative theoretical properties of these approaches relevant to the photometric stereo problem. To facilitate the analysis, here we consider the idealized case where there are no diffuse modeling errors, or that λ is small. In this situation, the basic problem from Eq. (3.10) becomes

$$\min_{\boldsymbol{n},\boldsymbol{e}} \|\boldsymbol{e}\|_0 \text{ s.t. } \boldsymbol{I} = L^T\boldsymbol{n} + \boldsymbol{e}, \tag{3.31}$$

which represents the pixel-wise analog of Eq. (3.7). If the lighting directions and sparse errors are in general position (meaning they are not arranged in an adversarial configuration with zero Lebesgue measure), then it can be shown that the minimizer of Eq. (3.31) denoted $\boldsymbol{n}_0$ is guaranteed to be the correct normal vector as long as the associated feasible error component $\boldsymbol{e} = \boldsymbol{I} - L^T\boldsymbol{n}_0$ satisfies $\|\boldsymbol{e}\|_0 < m - 3$. Therefore, a relevant benchmark for comparing photometric stereo algorithms involves quantifying conditions whereby a candidate algorithm can correctly compute $\boldsymbol{n}_0$.

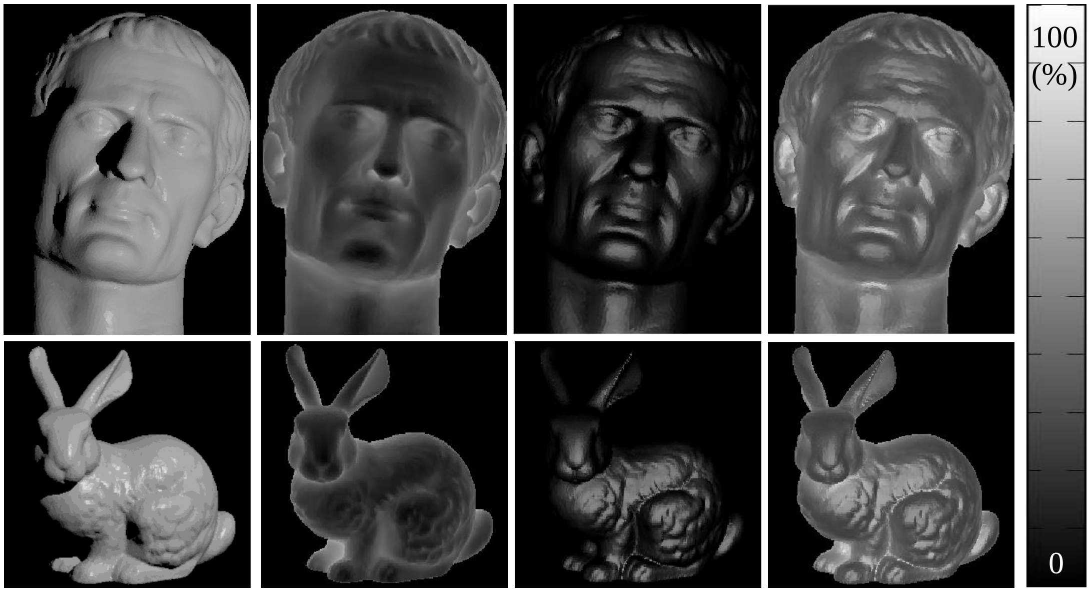
In this context, recent theoretical results have demonstrated that any minimizer $\boldsymbol{n}_1$ of the ℓ_1 relaxation approach will equivalently be a minimizer of Eq. (3.31) provided $\|\boldsymbol{e}\|_0$ is sufficiently small relative to a measure of the structure in columns of the lighting matrix L^T [17]. Unfortunately however, for typical photometric stereo problems the requisite equivalency conditions often do not hold (*i.e.*, $\|\boldsymbol{e}\|_0$ is required to be prohibitively small) both because of structure imposed by the lighting geometry and implicit structure that emerges from the relatively small dimensionality of the problem (meaning we do not benefit from asymptotic large deviation bounds that apply as m becomes large). Fortunately, SBL offers the potential for improvement over ℓ_1 via the following result.

Theorem: For all $\sigma_n^2 > 0$ (and assuming $\lambda \to 0$), if Γ_* is a global minimum of Eq. (3.23), then the associated estimator $\boldsymbol{n}_{sbl}$ will be a global minimum of Eq. (3.31). Moreover, for σ_n^2 sufficiently large it follows that: (i) Any analogous locally minimizing SBL solution is achieved at an estimate $\boldsymbol{n}_{sbl}$ satisfying $\|\boldsymbol{I} - L^T\boldsymbol{n}_{sbl}\|_0 \leq m - 3$, (ii) SBL can be implemented with a tractable decent method such that convergence to a minimum (possibly local) that produces an $\boldsymbol{n}_{sbl}$ estimator as good or better than the global ℓ_1 solution is guaranteed, meaning $\|\boldsymbol{I} - L^T\boldsymbol{n}_{sbl}\|_0 \leq \|\boldsymbol{I} - L^T\boldsymbol{n}_1\|_0$.

The proof is relatively straightforward and uses block-matrix inverse and determinant identities, as well as ideas from [17], to extend SBL properties derived in [120] to problems in the form of Eq. (3.31). We may thus conclude that SBL can enjoy the same theoretical guarantees as the ℓ_1 solution yet boosted by a huge potential advantage assuming that we are able to find the global minimum of Eq. (3.23) (which will always produce an $\boldsymbol{n}_{sbl} = \boldsymbol{n}_0$, unlike ℓ_1). There are at least two reasons why we might expect this to be possible based on insights drawn from [120]. First, as discussed previously, L^T will necessarily have some structure unlike, for example, high dimensional random matrices. In this environment SBL performance is often vastly superior to ℓ_1 because it can be shown to be implicitly based on an A-dependent sparsity penalty that can compensate, at least in part, for structure in $A = L^T$. Secondly, the sparse errors $\boldsymbol{e}$ will likely have substantially different magnitudes depending on image and object properties (meaning the non-zero elements of $\boldsymbol{e}$ will not all have the same magnitude), and it has been shown that in this condition SBL is more likely to converge to the global minimum [120].

3.2 Experimental Results

In this section, we quantitatively evaluate our method on synthetic and real image data. All experiments were performed on an Intel Core2 Duo E6400 (2.13GHz, single thread) machine with 4GB RAM and were implemented in MATLAB. For the SBL- and ℓ_1-based methods we used $\lambda = 1.0^{-6}$ in the synthetic experiments with no additive image noise and perfect lighting calibrations, and



(a) Object (b) Shadow (c) Specularity (d) Shadow + Specularity

Figure 3-4: Corruptions in *Caesar* and *Bunny* datasets. We illustrate the frequency $(\%)$ (*i.e.*, $\{$The number of corruptions appeared at a pixel$\}/\{$Total number of images$\}$) of (a) shadow, (b) specularity and (c) shadow and specularity.

$\lambda = 10^{-2}$ for the other cases. We set $\alpha = \sigma_n^2 = 10^6$ for all experiments.

3.2.1 Datasets

In synthesic experiments, we generate 32-bit HDR gray-scale images of two objects called *Bunny* (256×256) and *Caesar* (300×400) with foreground masks under different lighting conditions whose directions are randomly selected from a hemisphere with the object placed at the center. Specular reflections are attached using the Cook-Torrance reflectance model [29] [1] and cast/attached shadows are synthesized under each lighting condition[2]. Note that when in use (as defined for each experiment), the shadow mask is applied equivalently to all algorithms. To increase statistical reliability, all experimental results are averaged over 20 different sets of 40 input images. The average ratio of specularities in *Bunny* and *Caesar* are 8.4% and 11.6% and that of cast/attached shadows are 24.0% and 27.8% respectively (See Fig. 3-4). To quantitatively evaluate the performance, we compute the angular error between the recovered normal map and the ground truth. In this experiment, our meth-

---

[1] See details about the Cook-Torrance model in the Appendix B.

[2] Rendering the scene with cast shadows is actually not easy since they are the product of the complex interaction of geometry and illumination in the scene. Fortunately, many free ray-tracing softwares are available and in this dissertation, we used *blender* [1] software for synthesizing cast shadows.

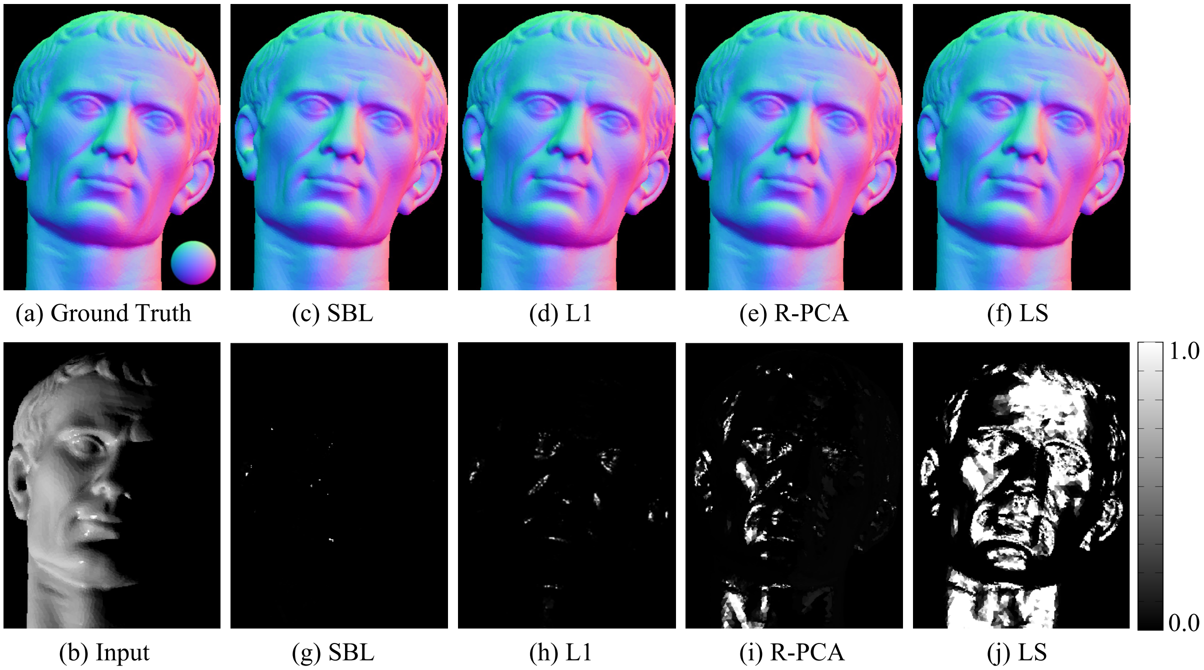


Figure 3-5: Recovery of surface normals from 40 images of *Caesar* (300 × 400) with explicit shadow removal. (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

ods via sparse regression are implemented by both SBL and a convex ℓ_1-norm based relaxation (L1). We compare our methods with the R-PCA-based method proposed by Wu et al.[124] (using a fixed trade off parameter) and the standard least squares (LS)-based Lambertian photometric stereo [122] estimate obtained by solving

$$\min_{N} \|O - N^T L\|_2^2. \tag{3.32}$$

The latter is equivalent to minimizing Eq. (3.8) with λ set to a large value. Note that the method from [124] introduce the sparsity penalty applied to non-Lambertian corruptions, however these two methods differ in the point that our method strictly constrains Lambertian rank-3 structure, while the R-PCA-based method does not always recover exactly rank-3 Lambertian matrix.

3.2.2 Quantitative Evaluation with Synthetic Images

We begin with the evaluation of performance using the synthesized images, which has Lambertian diffuse reflections along with non-Lambertian specularities and shadows.

We change experimental conditions with regard to the number of images, surface roughness (*i.e.*, the ratio of specularities), shadow removal (*i.e.*, whether or not a shadow mask is used to remove zero-valued elements from the observed images), and the presence of additional Gaussian noise. While the amount of cast/attached shadows are fixed to 24.0% and 27.8% of the number of all the pixels, respectively, the percentage of specularities depends on the surface roughness parameter that is independently selected in each experiment.

Valid number of images for efficient recovery in the presence of specularities

In this experiment, we vary the number of images to estimate the minimum number required for effective recovery when using the shadow mask with fixed surface roughness. The percentage of specular pixels in *Bunny* and *Caesar* are 8.4% and 11.6%, respectively.

Once 40 images are generated for each dataset, the image subset is randomly subsampled without replacement. The results are displayed in Table 3.2 and Fig. 3-5. We observe that the sparse-regression-based methods are significantly more accurate than both R-PCA and LS. It is also clear that SBL is more accurate than ℓ_1, although somewhat more expensive computationally.[3] Note that, although not feasible in general, when the number of images is only 5, the most accurate and efficient implementation for regression could be to just systematically test every subset of 3 images (*i.e.*, brute force search only requires 10 iterations at each pixel).

Robustness to shadows and image noise

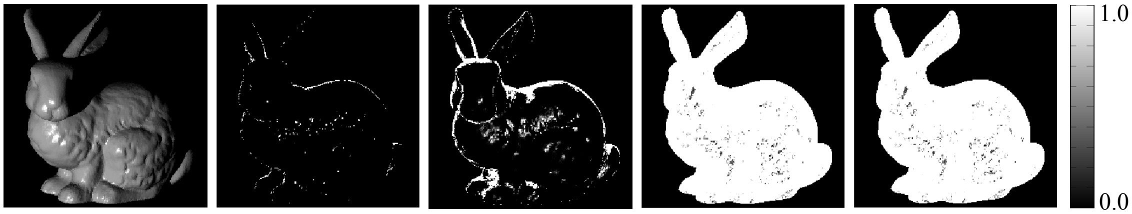
We now evaluate the robustness of our method against corruptions; shadows and image noise. We set two conditions for evaluating the effects of (i) shadows (fixed specularities, no shadow removal, no image noise), (ii) additive Gaussian image noise (fixed specularities, explicit shadow removal, varying amount of image noise). The ability to estimate surface normals without an explicit shadow mask is important, since in practical situations shadow locations are not always easy to be determined a priori. The number of images is 40 in (ii) and varying from 5 to 40 in (i). We use *Bunny* for evaluation, and the ratio of specularities is 8.4% in (i) and (ii), and image noise is 10% to 50% in (ii). Image noise obeys a zero-mean iid Gaussian distribution with $\sigma^2 = 0.1$.

The results are illustrated in Fig. 3-6, Fig. 3-7, Table 3.3 and Table 3.4. While performance of each method degrades when additional corrupted pixels (outliers) contaminate the estimation process, our sparse regression methods outperform both R-PCA and LS in accuracy and outperform R-PCA in efficiency in the presence of shadows and noise. We observe that the performance of

---

[3] The SBL convergence rate can be slower with fewer images because of the increased problem difficulty. This explains why the computation time may actually be shorter with more images.

(a) Ground Truth (c) SBL (d) L1 (e) R-PCA (f) LS



(b) Input (g) SBL (h) L1 (i) R-PCA (j) LS

Figure 3-6: Recovery of surface normals from 40 images of *Bunny*(256×256) without explicit shadow removal. (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

R-PCA degrades when a shadow mask is unavailable (*i.e.*, the positions of missing entries are unknown) while our sparse-regression based method automatically compensates for missing entries in the estimation process.

We also observe that SBL is more accurate than ℓ_1 in all conditions with slight more expense of computational effort. We further compare the results of SBL and ℓ_1 using the case where the number of images is 5 without removing shadows. The error maps and the numbers of corruptions per-pixel are displayed in Fig. 3-8. We observe that the ℓ_1 method typically suffer from shadows while SBL can find the correct solution in most pixels as long as the number of corruptions is less than 3, which is the theoretical limit when only 5 observations are given.

Sparseness of corruptions

While it does not often occur, our assumption that corruptions appear sparsely over images can be violated in some situations, *e.g.*, observations include wide-lobe specularities. In this experiment, we examine how our approach is affected by the sparseness of corruptions by changing the width of the specular lobe (*i.e.*, percentage of specularities appeared in observations). Here, we use *Bunny* dataset whose amount of specularities is varying from 10% to 60%, respectively, by appropriately choosing the model parameter of Cook-Torrance reflectance model. In this experiment, we remove shadows explicitly to factor out their effect.

The result is illustrated in Fig. 3-9. As expected, performance of both SBL and ℓ_1 degrades as

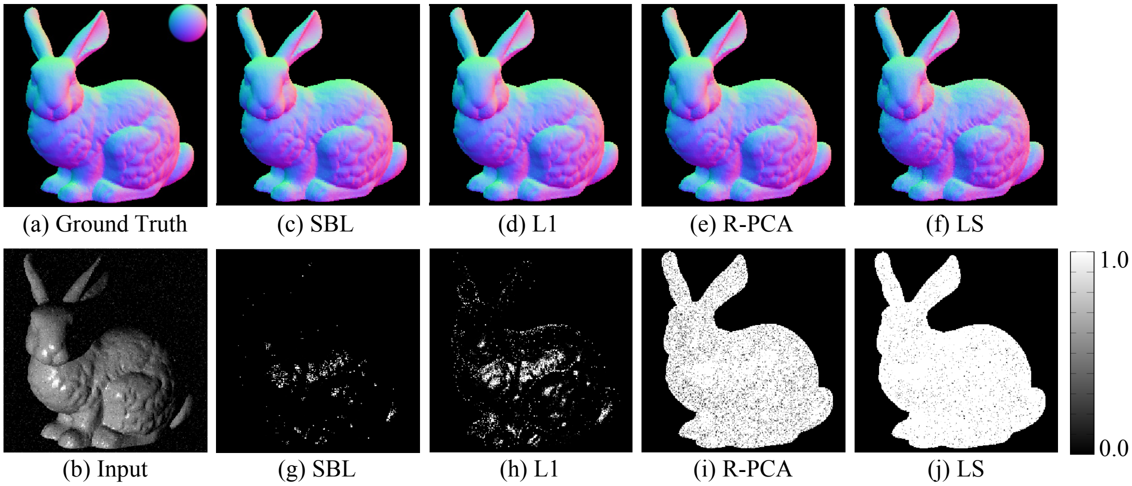


Figure 3-7: Recovery of surface normals from 40 images of *Bunny* with explicit shadow removal and additive Gaussian noises (30%). (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

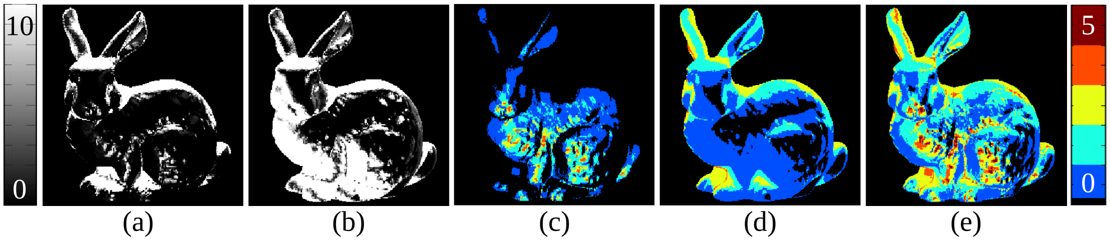


Figure 3-8: Comparison between SBL and ℓ_1-based method. Error maps of (a) SBL and (b) L1 (in degrees). The per-pixel number of (c) specularities, (d) shadows, (e) corruptions (The maximum is 5). These maps are illustrated based on one of twenty datasets.

the ratio of specular corruptions increases. However, even when the sparseness of outliers is difficult to be held, we observe that our sparse-regression based method still outperforms R-PCA and LS.

Comparison with RANSAC-based approaches

We also compared our approach with two different RANSAC-based methods. Here we use 40 images of *Bunny* dataset whose percentage of specular pixels is 8.4%. Shadows are removed and images noises are not added in this experiment.

The first one is a photometric linearization approach [79], which converts an input image into the ideal Lambertian image. It uses RANSAC for robustly identifying the basis images in the presence of non-Lambertian corruptions. Once linearized images are acquired, the standard Lambertian-

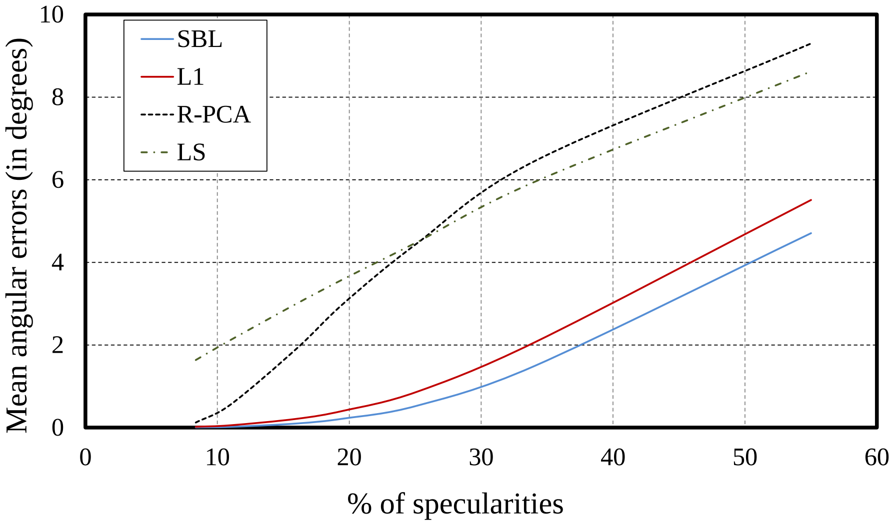


Figure 3-9: Experimental results of *Bunny* with varying amount of specularities. The x-axis and y-axis indicate the ratio of specularities and the mean angular error of normal map.

based approach [122] is applied with known lightings to estimate surface normals. The second one uses RANSAC more directly in Lambertian photometric stereo method like [54]. In this approach, three images are randomly sampled for each position independently, and the surface normal and albedo are estimated from them using Lambertian photometric stereo [122]. Then the Euclidean distances between observations and predicted intensities are computed, and the number of *inliers* whose distances are less than a threshold are counted. After the sampling process, the surface normal and albedo with the maximum number of inliers are adopted.

The results are summarized in Table 3.5 and Table 3.6, respectively. We have also included standard deviations in the tables for observing the estimation stability. We observe in Table 3.5 that although we use a large number of samples for RANSAC (*e.g.*, 2000) for the photometric linearization, this approach cannot always stably find the solution especially when the number of images is large. On the other hand, our method is able to successfully and efficiently finds the solution. We also observe in Table 3.6 that a large number of samples (*e.g.*, 200) gives very accurate estimation while it takes much time to compute it. The smaller number of samples gives more efficient but less accurate estimation. When the number of samples is 20, the computational cost is almost same as our SBL-based method, but the mean error is 50 times larger than ours, which also demonstrates the effectiveness of the proposed method.

It is also crucial to emphasize that the RANSAC algorithm scales very poorly as the number of

Table 3.1: Experimental results of *Bunny* with varying number of images.

No. of	Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)			
images	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
5	**6.0**	6.0	15.3	7.0	4.7	**4.3**	10.7	4.8	46.5	13.6	15.7	**4.6**
10	**0.09**	0.61	3.8	1.9	**0.27**	0.58	0.81	1.8	36.3	13.6	37.8	**5.9**
15	**0.076**	0.16	0.21	1.6	**0.052**	0.13	0.19	1.6	26.8	13.1	55.1	**6.3**
20	**0.033**	0.080	0.11	1.6	**0.022**	0.078	0.11	1.6	24.2	13.5	70.5	**6.9**
25	**0.018**	0.055	0.084	1.6	**0.010**	0.048	0.069	1.6	23.1	14.1	86.0	**7.6**
30	**0.012**	0.037	0.080	1.7	**0.0048**	0.032	0.065	1.7	23.1	14.2	121.0	**8.4**
35	**0.0057**	0.023	0.098	1.6	**0.0029**	0.019	0.093	1.6	22.7	14.6	161.3	**8.5**
40	**0.0039**	0.019	0.12	1.6	**0.0020**	0.015	0.12	1.6	22.6	15.0	200.7	**9.4**

Table 3.2: Experimental results of *Caesar* with varying number of images.

Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)			
SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
6.2	**6.0**	26.2	7.4	4.73	**4.63**	31.0	4.97	106.4	34.2	45.8	**15.9**
0.24	0.40	10.7	0.94	**0.19**	0.26	1.8	0.93	97.2	34.1	93.7	**19.2**
0.044	0.11	2.6	0.77	**0.047**	0.083	0.14	0.76	67.0	31.4	153.3	**22.0**
0.018	0.051	0.079	0.76	**0.015**	0.035	0.065	0.72	60.9	32.7	177.5	**23.3**
0.011	0.034	0.068	0.76	**0.0081**	0.023	0.059	0.76	57.9	34.3	196.9	**25.1**
0.0063	0.018	0.043	0.77	**0.0082**	0.018	0.043	0.77	58.1	33.2	231.7	**27.7**
0.0045	0.012	0.036	0.78	**0.0031**	0.0084	0.033	0.80	58.4	34.5	259.2	**29.4**
0.0031	0.0094	0.037	0.76	**0.0019**	0.0063	0.034	0.78	59.6	35.2	281.2	**31.5**

inlier model parameters is increased. Simply put, a larger number of inlier random samples is required to robustly estimate a larger number of model parameters, and obtaining such inlier samples becomes combinatorially more difficult in higher dimensions. Consequently, while RANSAC may work reasonably well here recovering surface normals under a Lambertian diffusive model (where the number of unknowns is effectively only 3), it will become intractable when using more complex reflectance models, which is a central concern herein. In contrast, our method scales linearly in the number of inlier parameters and can therefore be robustly expanded to handle non-Lambertian parameterized reflectance functions with outliers, e.g., the general diffusive model detailed in Chapter 4 or other more sophisticated extensions.

The comparison of ℓ_0 and ℓ_2 regressions

While the evaluation with average angular errors of recovered normal maps have indicated the outstanding performance of our method to the photometric stereo problem, the reason why the sparse regression outperforms least-squares regression in photometric stereo problem may not be intuitively understood. To clarify it, we randomly sampled a pixel from the overall image and visualized

Table 3.3: Results of *Bunny* without shadow removal.

No. of images	Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)			
	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
5	**5.2**	11.9	12.1	12.1	**5.0**	12.3	12.5	12.5	213.0	37.0	45.8	**5.1**
10	**2.8**	5.6	10.9	10.9	**2.3**	5.6	11.3	11.3	98.9	33.0	93.7	**6.0**
15	**1.9**	4.0	9.9	10.0	**2.3**	4.0	10.1	10.2	66.8	32.5	153.3	**7.4**
20	**1.2**	2.7	9.4	9.5	**1.0**	2.7	9.6	9.6	52.9	30.0	177.5	**7.6**
25	**0.81**	1.9	8.9	9.0	**0.69**	1.8	8.9	9.0	46.2	31.0	196.9	**9.1**
30	**0.62**	1.6	9.0	9.1	**0.61**	1.5	8.9	8.9	41.1	32.0	231.7	**9.4**
35	**0.59**	1.5	9.1	9.1	**0.58**	1.4	9.3	9.3	41.1	34.4	259.2	**11.0**
40	**0.53**	1.2	8.8	8.9	**0.58**	1.2	9.0	9.1	39.4	33.3	281.2	**10.7**

Table 3.4: Experimental results of *Bunny* with varying amount of additive Gaussian noises.

Dens. of noises (%)	Mean error (in degrees)				Median error (in degrees)			
	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
10	**0.0079**	0.040	0.16	3.3	**0.0060**	0.039	0.16	3.3
20	**0.021**	0.11	0.79	4.4	**0.019**	0.099	0.80	4.3
30	**0.068**	0.29	3.6	5.3	**0.060**	0.25	3.2	5.2
40	**0.21**	0.70	9.8	6.2	**0.18**	0.63	9.9	6.1
50	**0.58**	1.5	11.7	7.0	**0.53**	1.4	11.7	6.9

plots $[I, \boldsymbol{n}^T\boldsymbol{l}]$ for each lighting direction and a linear regression function in Fig. 3-10. Here, blue dots come from the estimated surface normal with LS(ℓ_2) / SBL(ℓ_0) and red dots come from the ground truth surface normal. Overlapping the blue dots onto red dots means the recovered surface normal is correct. We observe that while least-squares regression is easily disrupted by any corruptions beyond the Lambertian rule, SBL-based sparse regression robustly neglects the outliers.

3.2.3 Quantitative Evaluation with Inaccurate Lighting Directions

Other than specularities, shadows and image noises, calibration errors are also considered as corruptions in the diffusive image formation model. However, unfortunately, this type of errors violates our assumption that non-diffusive corruptions are sparse in observations. In this experiment, we evaluate the robustness to modeling errors by using synthesized 40 *Bunny* images and incorrect lighting directions (five degrees of angular error in random directions were added) to recover surface normals. In addition, we also attempt to refine lighting directions by iteratively recovering both surface normals and lighting directions based on the symmetrical structure of $\boldsymbol{n}^T\boldsymbol{l}$. First, we estimate surface normals using the given, errant lighting directions. Then, fixing recovered surface normals, we update the lighting directions using a least squares fit. We continue this process iter-

Table 3.5: Comparison with RANSAC based approach [79] with *Bunny* in dataset (A) (Number of samples is 2000).

No. of Images	Mean error		Median error		Standard deviation		Elapsed time	
	SBL	[3]	SBL	[3]	SBL	[3]	SBL	[3]
5	**6.0**	6.7	**4.7**	5.4	**4.1**	4.4	**46.5**	52.3
10	**0.09**	0.74	**0.27**	0.38	**0.35**	1.4	**36.3**	544.0
15	**0.076**	0.61	**0.052**	0.12	**0.059**	1.9	**26.8**	958.4
20	**0.033**	0.70	**0.022**	0.058	**0.027**	1.2	**24.2**	1048.9
25	**0.018**	1.0	**0.010**	0.063	**0.020**	2.6	**23.1**	1141.8
30	**0.012**	2.3	**0.0048**	0.042	**0.012**	4.0	**23.1**	1227.8
35	**0.0057**	3.2	**0.0029**	0.051	**0.0064**	9.2	**22.7**	1327.9
40	**0.0039**	2.1	**0.0020**	0.046	**0.0048**	4.3	**22.6**	1430.4

Table 3.6: Performance of RANSAC based approach [54] with *Bunny* in dataset (A) (Number of images is 40).

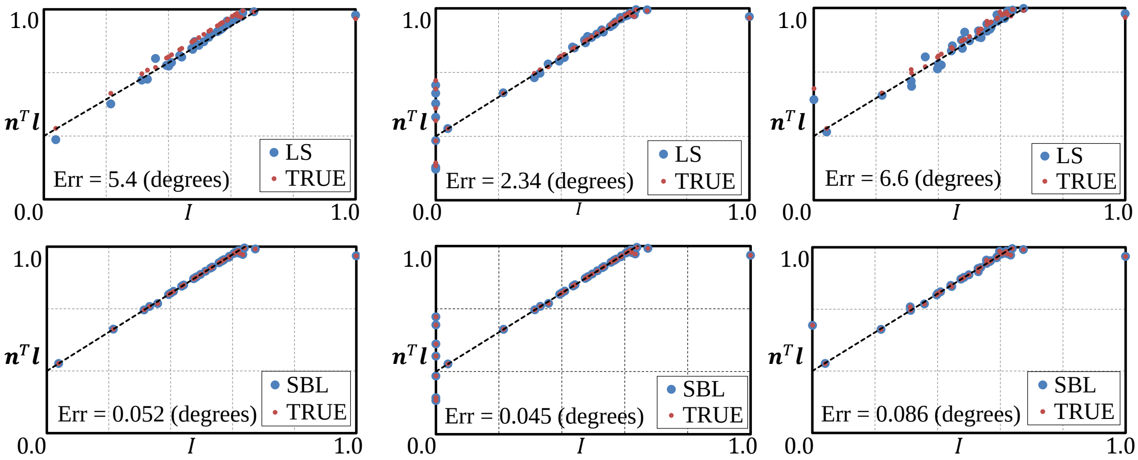
No. of Samples	Mean error	Median error	Standard deviation	Elapsed time
200	0.000019	7.9E-06	0.000027	283.0
100	0.0035	7.6E-06	0.31	142.4
80	0.0099	7.6E-06	0.70	115.2
60	0.019	7.6E-06	0.98	86.0
40	0.034	7.3E-06	1.2	57.9
20	0.22	7.4E-06	3.4	29.8
10	0.51	7.5E-06	5.0	15.6

atively until convergence. The experimental results are illustrated in Fig. 3-11. We observe that our method outperforms the other two methods even without refining the lighting directions; however, optimizing the lighting direction via a few iterations always improves the normal estimates. (Although the exact optimal number of iterations may be difficult to determine, a single iteration always has a substantial benefit).

3.2.4 Qualitative Evaluation with Real Images

We also evaluate our algorithm (only the SBL implementation) using real images.

We captured RAW images without gamma correction by Canon 30D camera with a 200 mm telephotolens and set it 1.5 m far from target object. Lighting conditions are randomly selected from a hemisphere whose radius is 1.5 m with the object placed at the center (therefore, the intensity of each light source assumes to be constant). Since photometric stereo techniques recover surface normals from the appearance variations, both dark and bright areas require to be preserved in the image. To prevent clipping of bright and dark regions of the scene, we combined multiple



(a) Lambertian + Specularities (b) Lambertian + Specularities + Shadows (c) Lambertian + Specularities + Noises

Figure 3-10: Comparison between ℓ_2 and ℓ_0 regression. 2-D plots of $(I, \boldsymbol{n}^T\boldsymbol{l})$ with recovered surface normal are illustrated (blue plots). If blue and red (true) plots place at the same position, it means the estimation is succeeded.

low dynamic range (LDR) images captured with different exposure into a single high dynamic range (HDR) image. To eliminate ambient illumination (which is present even in a dark room to some degree), we acquired ambient images by occluding the light source relative to the target object (*i.e.*, we blocked the light source so that the target object was in shadow, so that only ambient light illuminated the object). Ambient images are subtracted from the target images for removing the effect of ambient illumination. In addition to HDRI compositing, the estimation of lighting direction is required before performing each algorithm. In this dissertation, we simply put a glossy sphere in the same scene with the target object so that images of both calibration sphere and target object are captured under the same illumination environment. Once the region of the perfect specularity (*i.e.*, mirror direction) is extracted, the lighting direction is calculated based on the radius of the sphere and the distance between the region and the center of the calibration sphere. We show some examples of LDR images captured under the dark room and images for the ambient/lighting calibration in Fig. 3-12. We use a set of 25 images of *Chocolate bear* (261x421), and 40 images each of *Doraemon* (269x420) and *Fat guy* (293x344). We evaluate the performance by visual inspection of the output normal maps, elevation angle maps (orientations between normals and a view direction) and azimuth angle maps (normal orientation on the x-y plane) that are illustrated in Fig. 4-10. We observe that our method can estimate smoother and more reasonable normal maps in the presence of a large amount of specularities.

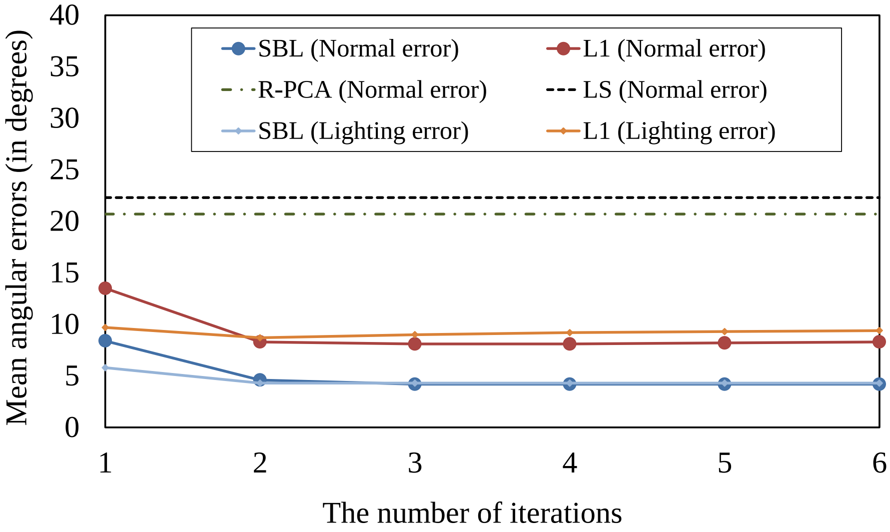
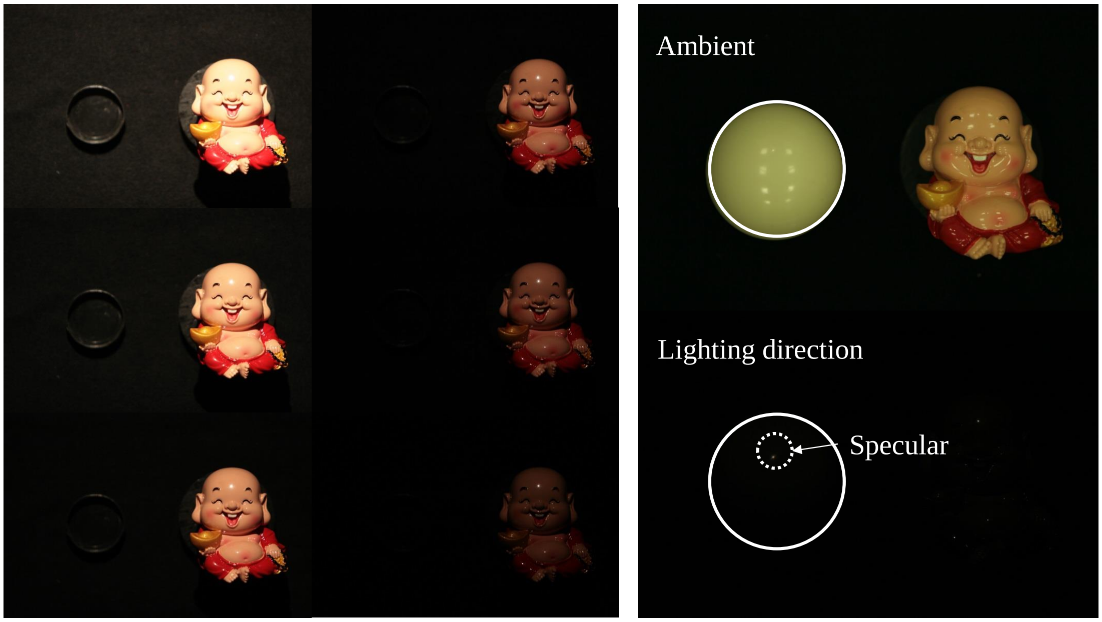


Figure 3-11: Experimental results of *Bunny* in dataset (A) under incorrect lighting directions to the 6-th iteration.

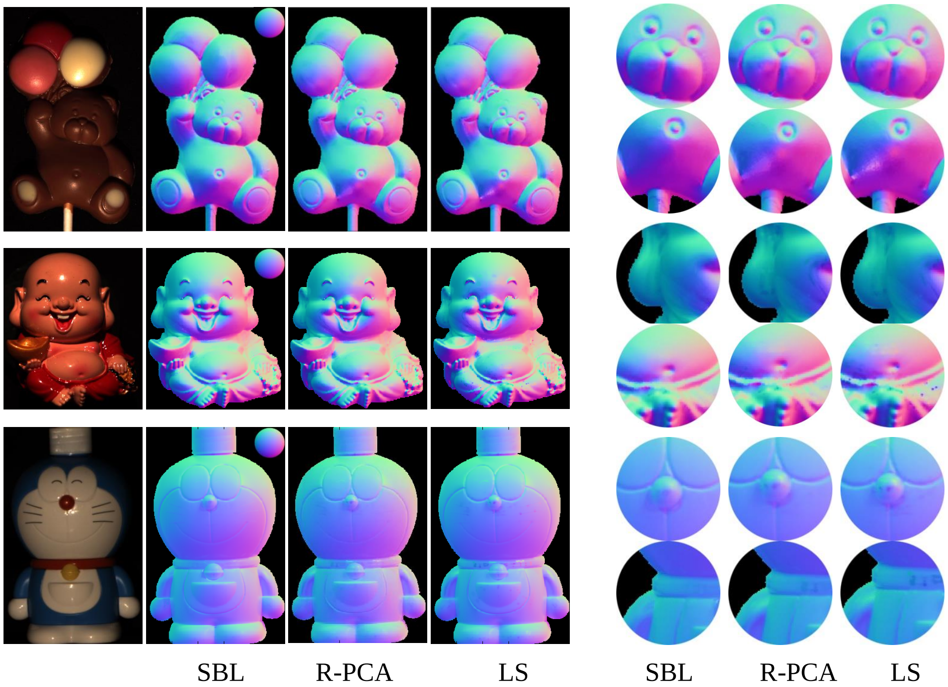
3.3 Conclusion

Herein we have demonstrated the superior performance of our sparse regression approach to photometric stereo problem (especially for the Lambertian + sparse corruption scene) through extensive analyses and experimentation. In particular, our method gives more accurate estimates of surface normals than previous least squares and R-PCA approaches while remaining computationally competitive. Regarding competing sparse regression techniques, SBL is both theoretically and empirically superior to ℓ_1-based estimates but requires a modest increase in computational cost. A limitation of the current algorithm is that we assume the diffusive component of the reflectance is Lambertian. Therefore non-Lambertian diffusive objects or specularities whose lobe is considerably large can potentially be problematic, although this affect is partially mitigated by the diffuse and sparse error terms built into our model. Therefore, in the next chapter, we will follow the sparse regression approach but extend the algorithm so that our model accepts non-Lambertian diffusive components coupled with sparse corruptions to recover surface normals robustly across a wider range of materials.

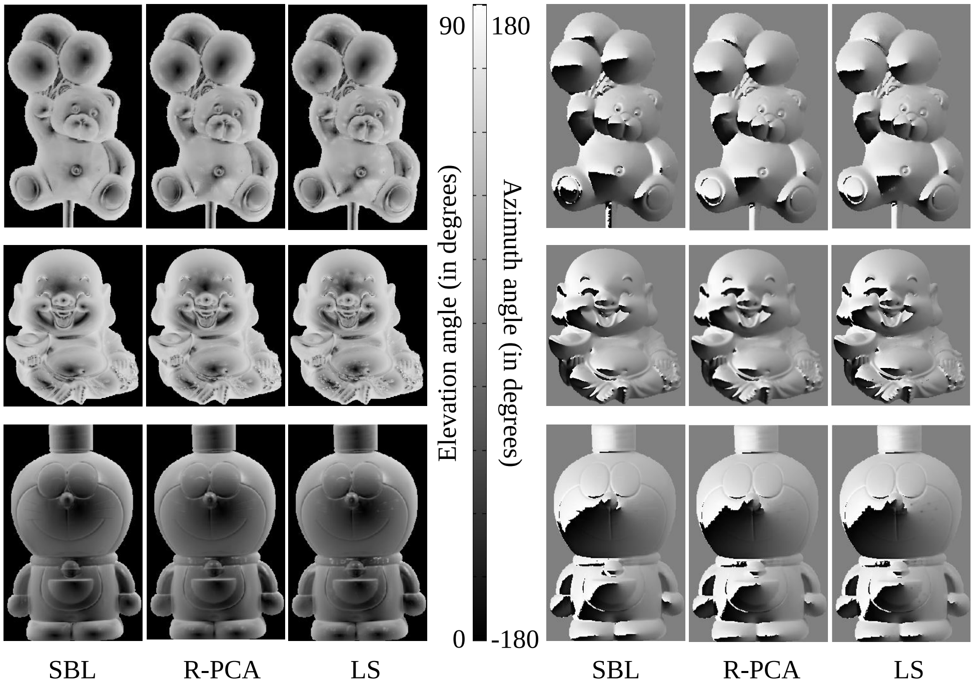


(a) HDRI acquisition (b) Calibration

Figure 3-12: Illustration of the HDRI acquisition and lighting calibration. (a) Images with varying exposure are merged into the HDRI. (b) Ambient illumination is measured from an image that is captured in the absence of the point light source and lighting direction is calibrated using a highly reflective sphere.



(a) Input (b) Normal map (c) Close-up



(d) Elevation angle map (e) Azimuth angle map

Figure 3-13: Experimental results with real datasets. We used three kind of datasets called *Chocolate bear* (25 images with 261x421), *Fat guy* (40 images with 293x344) and *Doraemon* (40 images with 269x420). (a) Example of input images. (b),(C) Recovered surface normals and close-up views. (d) Elevation angles of recovered surface normals. (e) Azimuth angles of recovered surface normals.

Chapter 4

Robust Photometric Stereo using Sparse Regression for General Diffuse Surfaces

Previously, we have argued that the sparse regression framework is particularly well-suited for finding maximally sparse Lambertian observations in the presence of various kind of non-Lambertian corruptions, demonstrating that it outperforms the state-of-the art robust algorithm for both synthetic and real datasets. On the other hand, however, the limitation is that the performance is degraded in the absence of dense, dominant Lambertian diffusive reflections in the observation.

In this chapter, we provide an extension of the Lambertian sparse regression based robust photometric stereo algorithm for stably and accurately estimating the surface normals of a scene in the presence of dense, dominant *non-Lambertian* diffusive reflections. For this purpose, a hierarchical Bayesian model is developed that automatically decomposes observed appearances into a dense diffuse component and a sparse, non-diffuse component for capturing shadows, specularities, and other corruptions. In the similar manner with the previous chapter, optimization and inference is accomplished using a robust majorization-minimization technique akin to the popular EM algorithm, with desirable convergence properties and quantifiable advantages over standard convex estimators.

However, simply introducing non-linear diffusive BRDF in the imaging model as Eq. (1.2) poses a difficult optimization challenge due to discrete, combinatorial nature of ℓ_0-norm minimization. For tackling this difficulty, we develop our diffuse component by assuming, in the absence of non-diffuse corruptions, that pixelwise appearances are well-approximated by a monotonic (and therefore invertible) function of the dot-product between the surface normal and the lighting direction. We may then consider the inverse representation of the image formation process, where the unknown normal vector is now separated from the unknown monotonic inverse reflectance function. By parameterizing the latter using a non-linear, nonparametric approximation (*e.g.*, piecewise linear approximation or smoothing splines), we obtain a set of linear equations in both the surface normals and diffuse parameters, leading to simple, closed-form estimators. The method proposed in this chapter tries to capitalize on both of Lambertian robust approach, and non-Lambertian model-based approach which have been discussed in Section 2.2.3 in that it relaxes the often restrictive Lambertian reflectance model in the robust approach by using a non-linear and non-parametric representation of reflectances, while simultaneously performing robust estimation to avoid over-fitting. More specifically, our sparsity penalty effectively rejects outliers while the inverse non-linear representation of diffuse reflectances enables us to handle a wide variety of materials.

Note that although we mainly discuss about the piecewise linear function [115] as an approximation of general diffuse reflections because of its simplicity and representational capability, the other non-parametric representation is also applicable in our framework [1]. However, in the context of photometric stereo, the piecewise linear function is a natural extension of the Lambertian reflectance model, which directly corresponds to the case where there is only one linear segment. Later we will empirically demonstrate that additional piecewise linear diffuse segments (e.g., three) can effectively represent many complex non-Lambertian reflections.

In summary, our *non-Lambertian* sparse regression analysis attempts to maximize the number of observations that can be adequately explained via the piecewise linear inverse diffuse function (the inlier model) while treating the remaining observations, including specularities, shadows, and sensor saturations, as non-diffuse outliers with separate, unknown variances (the outlier model). The partitioning between inliers and outliers is estimated simultaneously with the normal vectors and model parameters using a principled variational Bayesian technique (*i.e.*, SBL).

Even though the model is extended to the general diffuse reflections, the proposed framework still benefits from simple, efficient pixelwise optimization, which is easily amenable to parallel processing. We remind that we also rely on the following assumptions:

1. Relative position between the camera and the object is fixed across all images.
2. Object is illuminated by a point light source at infinity from varying and known directions.
3. Camera view is orthographic, and the radiometric response function is linear.

---

[1] In Appendix A, we discuss about the non-Lambertian diffuse modeling based on the penalized least-squares using the B-spline approximation.

4.1 Problem Statement

As has been presented in Section 2.1, reflections on the surface of real world objects can be encoded by a BRDF, which relates the observed intensity I (outgoing radiance, strictly speaking) at a given point on the object to the associated surface normal $\boldsymbol{n} \in \mathbb{R}^3$, the incoming lighting direction $\boldsymbol{l} \in \mathbb{R}^3$, and the outgoing viewing direction $\boldsymbol{v} \in \mathbb{R}^3$ via

$$I = f(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}). \tag{4.1}$$

The dichromatic reflectance model [96] states that if the scene is illuminated by a single dominant point light source, the radiance is a linear combination of diffuse and specular reflectance via

$$I = f_d(\boldsymbol{n}, \boldsymbol{l}) + f_s(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}). \tag{4.2}$$

In practice, various additional effects are observed, including attached/cast shadows, image noise, inter-reflections, and so on. We can interpret these effects as additive corruptions e applied to the ideal scene leading to the image formation model giving

$$I = f_d(\boldsymbol{n}, \boldsymbol{l}) + f_s(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) + e. \tag{4.3}$$

The photometric stereo assuming the dichromatic reflectance model is a problem to recover surface normal $\boldsymbol{n}$ of a scene by inversely solving Eq. (5.1) from a collection of m observations under the unknown set of parameters (f_d, f_s, e). Note that except for uncalibrated photometric stereo problems such as [38] (See details in Section 2.2.1), $\boldsymbol{l}$ and $\boldsymbol{v}$ are usually known.

Chapter 3 and other early photometric stereo works [79, 124] assumed that the diffuse component of observation is represented by Lambertian reflectance model (*i.e.*, $f_d = \rho \boldsymbol{n}^T \boldsymbol{l}$, where ρ is a surface albedo) and discarded the non-Lambertian component $f_s + e$ as outliers. While the Lambertian assumption is effective for a certain materials, this strong assumption on the reflectance substantially limits the target objects. Instead, we introduce the general representation of the material diffuse function to handle non-Lambertian diffuse materials as follow,

$$f_d(\boldsymbol{n}, \boldsymbol{l}) = f(\boldsymbol{n}^T \boldsymbol{l}), \tag{4.4}$$

where we assume $f(\boldsymbol{n}^T \boldsymbol{l})$ passes through the origin, that is $f(0) = 0$. Note that Lambertian image

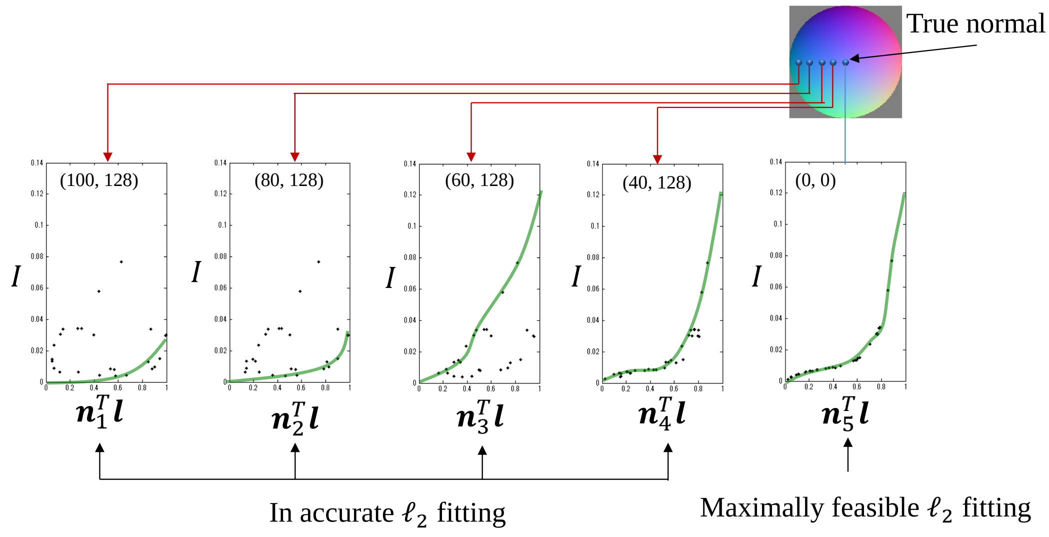


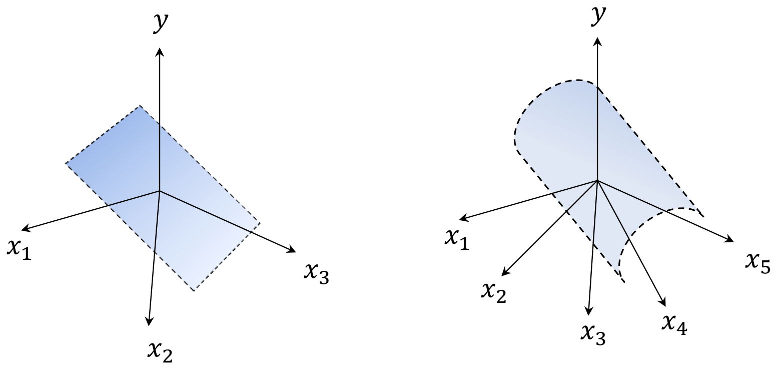
Figure 4-1: Illustration of the non-linear least-squares regression. Here we show 2-D plots of $[\boldsymbol{n}^T\boldsymbol{l}, I]$ with different $\boldsymbol{n}$ and estimated underlying regression model. We observe that the sum of residuals of the regression is minimized when the true surface normal is provided.

formation is a special case when f is a linear function. Then, we also merge f_s and e as deviations from the diffuse reflection and use the following imaging model,

$$I = \tilde{I} + e = f(\boldsymbol{n}^T\boldsymbol{l}) + e. \tag{4.5}$$

Here, $\tilde{I}$ is the diffuse component of I. Our goal is to recover unknown surface normal $\boldsymbol{n}$, diffuse reflectance function f and non-diffuse corruptions e from a collection of lighting directions $\boldsymbol{l}$ and associated appearances I. In other words from the least-squires regression analysis, the problem is to estimate the surface normal as a part of unknowns in the regression model so that the sum of each residual (*i.e.*, distance between each data and the regression function) is minimized (See Fig. 4-1).

However, there are two critical issues which must be solved: (a) the coincidence of unknown parameters $\boldsymbol{n}$ and f in the same term, and (b) an under-constrained problem since the number of unknowns (equal to $m + 2$ plus however many degrees of freedom are needed to describe f) always exceeds the number of equations (equal to the number of images m). We overcome these difficulties by combining a convenient, piece-wise linear inverse representation of the imaging model with a sparsity penalty applied through latent variables embedded in a robust hierarchical Bayesian framework.



(a) (b)

Figure 4-2: Comparison of linear and non-linear regression. While (a) a standard Lambertian photometric stereo problem is considered as the regression analysis where a maximally feasible four dimensional plane is searched over the 4-D space whose coordinates are $[x_1, x_2, x_3, y] = [l_1, l_2, l_3, I]$, (b) the non-linear photometric stereo problem is considered as the regression analysis where both the dimension of the search space and complexity of the regression function are increased.

4.2 Why is the Problem Difficult to Solve?

Before stepping in the proposed algorithm, we further explain why non-linear diffusive reflectance model is difficult to be handled. Suppose we know that the diffuse reflectance component is approximated by a 2-nd order polynomials passing though the origin as

$$f(\boldsymbol{n}^T\boldsymbol{l}) = a_2(\boldsymbol{n}^T\boldsymbol{l})^2 + a_1\boldsymbol{n}^T\boldsymbol{l}, \tag{4.6}$$

where a_2, a_1 are model parameters. Merging Eq. (4.6) into Eq. (4.5) with $\boldsymbol{n} \triangleq [n'_1 \ n'_2 \ n'_3]$ (See details in Section 1.2), the image formation model is now expressed as

$$I = a_2\left({n'_1}^2 l_1^2 + {n'_2}^2 l_2^2 + {n'_3}^2 l_3^2 + 2n'_1 l_1 + 2n'_2 l_2 + 2n'_3 l_3\right) + a_1\left(n'_1 l_1 + n'_2 l_2 + n'_3 l_3\right) + \epsilon. \tag{4.7}$$

Therefore, the regression analysis corresponding to Eq. (1.3) is now represented as follow,

$$\min_{\boldsymbol{n},\boldsymbol{a},\boldsymbol{\epsilon}} \sum_{i=1}^{m} \|I_i - a_2\left({n'_1}^2 {l_1^i}^2 + {n'_2}^2 {l_2^i}^2 + {n'_3}^2 {l_3^i}^2 + 2n'_1 l_1^i + 2n'_2 l_2^i + 2n'_3 l_3^i\right) - a_1\left(n'_1 l_1^i + n'_2 l_2^i + n'_3 l_3^i\right) - \epsilon_i\|_2^2, \tag{4.8}$$

where $\boldsymbol{a} \triangleq [a_2 \ a_1]^T$. Apparently, the optimization of Eq. (4.8) is under-determined and non-linear due to the coincidence of unknown parameters $\boldsymbol{n}$ and $\boldsymbol{a}$, and the 2-nd order term of them (In addition to increased number of unknowns, the underlying regression model is no longer plane in the high-dimensional space as shown in Fig. 4-2). Even though there are various non-linear optimization techniques [91], no closed-form solution exist and careful initialization is required to avoid the local minima. More seriously, there is almost no convenient solver to perform the sparse (ℓ_0) regression on the non-linear problem. In the next section, we prove that the non-linearity of the problem can be resolved by assuming the monotonicity of the diffuse BRDF, and difficult or even intractable non-linear ℓ_2,ℓ_0-norm minimization problems are simplified into corresponding linear problems.

4.3 Inverse Diffuse Reflectance Model

For simplicity, we first neglect the non-diffuse corruptions e in Eq. (4.5). Then, we assume the monotonicity of the diffuse reflectance function like [58, 100] which provides the following constraint on the function under two different lighting directions $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$,

$$\boldsymbol{n}^T\boldsymbol{l}_i > \boldsymbol{n}^T\boldsymbol{l}_j \leftrightarrow f(\boldsymbol{n}^T\boldsymbol{l}_i) > f(\boldsymbol{n}^T\boldsymbol{l}_j). \tag{4.9}$$

Under this assumption, the unique existence of the inverse function of f is guaranteed giving

$$f^{-1}(\tilde{I}) = g(\tilde{I}) = \boldsymbol{n}^T\boldsymbol{l}. \tag{4.10}$$

Now that we assume only diffuse reflections appear in the scene, *i.e.*, $\tilde{I} = I$, the following equation is acquired,

$$\boldsymbol{n}^T\boldsymbol{l} = g(I). \tag{4.11}$$

We call Eq. (4.11) the *inverse diffuse reflectance model*. The fundamental advantage of this model is that an unknown function $g(I)$ and a surface normal $\boldsymbol{n}$ are separated, which contributes to the simplicity of the problem. Eq. (4.11) suggests that the per-pixel collection of 2-D plot $(I_i, \boldsymbol{n}^T\boldsymbol{l}_i)$ corresponding to i-th image must draw a monotonic inverse diffuse function $g(I)$. While this relationship limits the solution space of both $\boldsymbol{n}$ and g, the problem is that there are still multiple feasible solutions of a pair of $\boldsymbol{n}$ and $g(I)$ as illustrated in Fig. 4-3, especially when m is small. To reduce inherent ambiguity of the problem, we further assume a parametric model of the general inverse diffuse function $g(I)$.

Given that the left-hand-side of Eq. (4.11) is linear in the unknown normal vector $\boldsymbol{n}$, for compu-

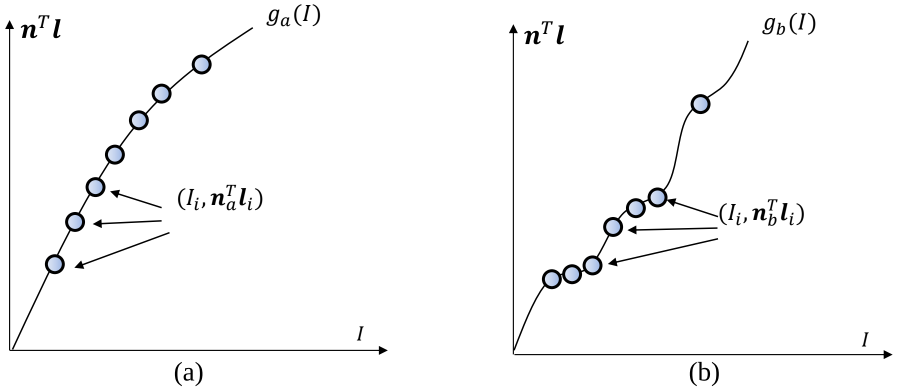


Figure 4-3: We can draw different monotonic curves which interpolate 2-D plots $(I_i, \boldsymbol{n}^T\boldsymbol{l}_i)$ derived from $(\boldsymbol{n}_a)$ in (a) and $(\boldsymbol{n}_b)$ in (b), which illustrates that there are multiple feasible solutions of a pair of $\boldsymbol{n}$ and $g(I)$ which satisfy the inverse diffuse reflectance model in Eq. (4.11).

tational simplicity we would like to impose similar linearity on the right-hand-side in our parameterized representation of $g(I)$. For this purpose, we then choose to express $g(I)$ as a summation over p fixed and known, non-linear basis functions $g_k(I)$ weighted by an unknown coefficient vector $\boldsymbol{a}$, leading to the representation

$$g(I) = \sum_{k=1}^{p} a_k g_k(I). \tag{4.12}$$

While non-linear in I, $g(I)$ is clearly linear in $\boldsymbol{a} = [a_1, \ldots, a_p]^T$. Choices for each g_k include polynomial, Gaussian, logistic, and spline functions as well as many others.

4.3.1 Selection of Basis Function

The specific choice of g_k is crucial in the non-parametric regression analysis [49] that explores a relation ship between two variables *i.e.*, I and $\boldsymbol{n}^T\boldsymbol{l}$. We here choose to adopt a piecewise linear representation [115][2] which is composed of multiple polylinear functions of the form

$$g_k(I) = \begin{cases} 0 & (0 \le I < b_{k-1}) \\ I - b_{k-1} & (b_{k-1} \le I < b_k) \\ b_k - b_{k-1} & (b_k \le I) \end{cases} \tag{4.13}$$

---

[2] A possible alternative choice of the basis function is smoothing splines with B-spline curves, which will be discussed in the Appendix A.

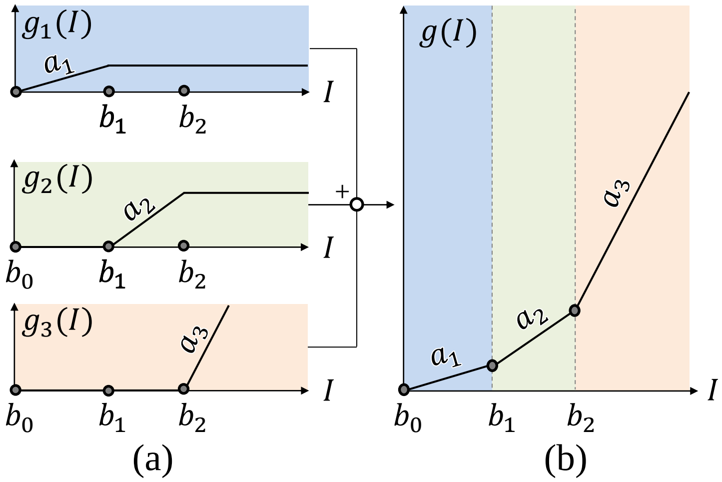


Figure 4-4: The illustration of the piecewise linear function. We show the case when the number of basis functions is three. (a) Each basis function is defined as a polylinear function which has a breaking point, and (b) the piecewise linear function is defined as the summation of these basis functions.

where each b_{k-1} denotes the point where the $(k-1)$-th linear segment ends and the k-th segment begins. See Fig. 4-4 for details. By construction, $g(I)$ will be a *continuous* piecewise linear function, meaning each adjoining linear segment connects to one another at the corresponding point b_{k-1}, regardless of the coefficients $\boldsymbol{a}$. We also assume that $b_0 = 0$, meaning $g(I)$ will intersect the origin. Remaining values of b_{k-1} are chosen such that each piecewise linear segment spans an equal-sized bin over the range of I.

This choice of basis function leads to a natural generalization of the Lambertian reflectance model, which is obtained when all elements of $\boldsymbol{a}$ are set to the same positive value. We thus preserve preferable properties of the Lambertian model where appropriate, while still allowing us to handle more general non-linear diffuse reflections where needed. In fact, even with p small, we may nonetheless approximate a wide variety of non-linear functionals, with monotonicity ensured whenever $\boldsymbol{a}$ has all non-negative elements (while we do not strictly enforce non-negativity, our learning procedure described in the following sections strongly disfavors any $a_k < 0$). Here we should note that our model relates to the recent bilinear BRDF model proposed by Romeiro and Zickler [87], which is represented as a linear combination of non-negative basis functions learned through non-negative matrix factorization of 100 materials in the MERL BRDF database [74]. However, there are two different points. Firstly, our basis functions are learned directly from the data itself (*i.e.*, linear segment ends in Eq. (4.13) are decided from samples at each pixel) while [87] learns basis functions from the external database. More importantly, we only model the diffusive component by a simple linear function which contributes to the computationally efficiency and stability, while [87] adapts basis functions represented as discretized two-dimensional matrices.

By substituting Eq. (5.7) into the inverse diffuse reflectance model, Eq. (4.11) becomes

$$\boldsymbol{n}^T \boldsymbol{l} = \sum_{k=1}^{p} a_k g_k(I). \tag{4.14}$$

Collecting variations of observation at the same pixel under different lighting directions, the equations can be merged into following linear problem,

$$\tilde{A}\boldsymbol{x} = 0, \tag{4.15}$$

where $\boldsymbol{x} \triangleq [n_x, n_y, n_z, a_1, a_2, \ldots, a_p]^T \in \mathbb{R}^{p+3}$, and n_x, n_y, n_z are the three elements of the surface normal. $\tilde{A} \in \mathbb{R}^{m\times(3+p)}$ is a data matrix whose j-th row is given by

$$\begin{aligned} \tilde{A}_j =& [-l_x^j, -l_y^j, -l_z^j, b_1 - b_0, \\ & \ldots, b_{k-1} - b_{k-2}, I_j - b_{k-1}, 0, \ldots, 0]. \end{aligned} \tag{4.16}$$

Here we assume $b_{k-1} \leq I_j < b_k$ and l_x^j, l_y^j, l_z^j are three elements of the j-th lighting direction.

Without loss of generality, we may avoid the degenerate $\boldsymbol{x} = 0$ solution to Eq. (4.15) by constraining $\sum_{k=1}^{p} a_k = 1$. For this purpose we replace $\tilde{A}$ by A where $[0, 0, 0, 1, \ldots, 1]$ is appended as the $(m+1)$-th row of $\tilde{A}$ and we define $\boldsymbol{y} \in \mathbb{R}^{m+1}$ as a vector of all zeros except for a one as the last element. Given the appearance variations $(I_1, I_2, \ldots, I_m)$ under different known lighting conditions $(\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_m)$, the optimal surface normal $(\boldsymbol{n})$ and model parameters $(a_1, a_2, \ldots, a_p)$ are recovered by solving the linear problem

$$\min_{\boldsymbol{x}} \|W(A\boldsymbol{x} - \boldsymbol{y})\|_2^2. \tag{4.17}$$

Here $W \triangleq \text{diag}[1, \cdots, 1, \infty] \in \mathbb{R}^{(m+1)\times(m+1)}$ is a diagonal weighting matrix designed to strictly enforce the constraint $\sum_{k=1}^{p} a_k = 1$. Note that $2 + p$ linearly independent images are sufficient for producing a unique solution to Eq. (4.17). We will refer to this photometric stereo method as *piecewise linear least-squares regression (PL-LS)*. In spite of its simplicity, PL-LS works for a wide variety of non-Lambertian diffuse materials (see our experimental results in Section 4.4.2). The problem of course is that real images are frequently contaminated with various non-diffuse effects as modeled in Eq. (4.5). The next section will focus on how to handle these corruptions within our inverse diffuse reflectance model.

4.3.2 Piecewise Linear Sparse Regression with Non-Diffuse Corruptions

By introducing additive non-diffuse corruptions $\boldsymbol{e}'$ in Eq. (4.14), the image formation model becomes

$$\boldsymbol{n}^T\boldsymbol{l} = \sum_{k=1}^{p} a_k g_k(I) + e'. \tag{4.18}$$

Note that although the value of e' in Eq. (4.18) is different from the e in Eq. (4.5), both factors are directed at the same observations and thus serve the same overall purpose (see Fig. 5-3). The standard least-squares based photometric stereo problem corresponding to Eq. (4.18) would involve solving

$$\min_{\boldsymbol{x},\boldsymbol{e}'} \|W(A\boldsymbol{x} + \boldsymbol{e}' - \boldsymbol{y})\|_2^2 \quad \text{s.t. } e'_{(m+1)} = 0, \tag{4.19}$$

where $\boldsymbol{e}' \triangleq [e'_1, e'_2, \ldots, e'_{m+1}]^T \in \mathbb{R}^{m+1}$. Note that $e'_{(m+1)}$ must be set to zero to maintain the constraint $\sum_{k=1}^{p} a_k = 1$ (see details in Section 4.3.1).

Given observations I and lighting directions $\boldsymbol{l}$, the ultimate goal is to recover surface normals and model parameters $\boldsymbol{x}$ and non-diffuse corruptions $\boldsymbol{e}'$. However, this is an under-constrained problem since the number of unknowns $p + 2 + m$ will always exceeds the number of independent equations m.

One solution to this ambiguity is to apply simple shadow/specular thresholding [46, 99] or a color channel transformation [73] as a preprocessing step, to obtain an estimate of $\boldsymbol{e}'$ and/or discard outliers. However, these types of heuristics may discard useful information at times and come with an additional computational expense. Moreover, graph-based approaches [125, 109, 114] and robust algorithms [79, 124] do not naturally embed within our framework since they may conflict with our inverse piecewise diffuse model and/or degrade the numerical stability.

As has been discussed in Chapter 3, we instead introduce a sparsity penalty applied to $\boldsymbol{e}'$, whose minimization disambiguates the infinity of feasible solutions to Eq. (4.18), meaning to perform the sparse regression. Remind that without the inverse representation of diffuse function, it would be almost impossible to perform the sparse regression analysis on the highly nonlinear image formation model. We also remind that this penalty quantifies the reasonable observation that objects in the natural world exhibit dominant diffuse reflections while non-diffuse effects emerge primarily

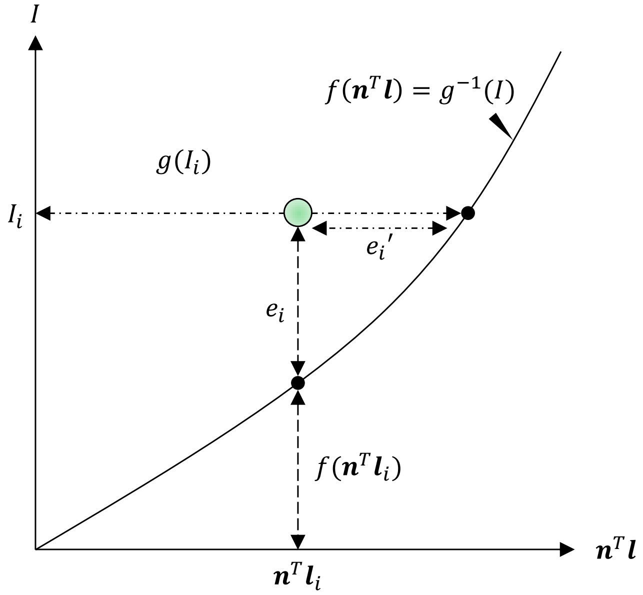


Figure 4-5: A 2-D point $(\boldsymbol{n}^T\boldsymbol{l}_i, I_i)$ can be represented by both the forward image formation model $(I_i = f(\boldsymbol{n}^T\boldsymbol{l}_i) + e_i)$ and the inverse image formation model $(\boldsymbol{n}^T\boldsymbol{l}_i = g(I_i) + e_i')$. The illustration indicates that they are uniquely convertible if the reflectance function is monotonic.

in limited areas of its appearance. Strictly speaking, we assume that the optimal feasible solution to Eq. (4.19) is acquired when the largest possible number of observations are lying on the piecewise linear diffuse reflectance function. Reflecting this assumption, our estimation problem can be formulated as

$$\min_{\boldsymbol{x},\boldsymbol{e}'} \|\boldsymbol{e}'\|_0 \quad \text{s.t.} \qquad \boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}',\ e'_{(m+1)} = 0. \tag{4.20}$$

Here, $\|\cdot\|_0$ represents the ℓ_0-norm, which counts the number of non-zero entries in a vector. To reiterate, Eq. (4.20) builds on the assumption that images are captured under known lighting conditions and any non-diffuse corruptions have sparse structure. If these assumptions are not true (*e.g.*, because of imperfect lighting calibration, no dominant diffuse structure, etc.), then the hard constraint in Eq. (4.20) is no longer appropriate. To compensate for more modeling errors, we relax the hard constraint via an additional model mismatch penalty giving

$$\min_{\boldsymbol{x},\boldsymbol{e}'} \|W(A\boldsymbol{x} + \boldsymbol{e}' - \boldsymbol{y})\|_2 + \lambda\|\boldsymbol{e}'\|_0 \quad \text{s.t.}\ \ e'_{(m+1)} = 0, \tag{4.21}$$

where λ is a nonnegative trade-off parameter balancing data fit with sparsity. Note that in the limit as $\lambda \to 0$, problems (4.20) and (4.21) are equivalent (the limit must be taken outside of the minimization). Similarly to Eq. (3.10), Eq. (4.21) also entails a difficult, combinatorial optimization problem that must be efficiently solved at every pixel. Here we only consider a hierarchical Bayesian approximation to estimate $\boldsymbol{x}$ while simultaneously accounting for $\boldsymbol{e}'$ since we have already revealed that ℓ_1-based relaxation is not well appropriate for the photometric stereo problem in Chapter 3. This formulation, a generalized version of sparse Bayesian learning (SBL) [111], is described in detail next. We note that most of the derivation are common with the formulation of the Lambertian sparse regression analysis, however there are some differences because there are large number of parameters ($\boldsymbol{x}$) in the piecewise linear model than a simple Lambertian model.

4.3.3 Recovery of Normals and Corruptions Via SBL

The derivation of the update rules of $\boldsymbol{n}$ and $\boldsymbol{e}'$ in SBL has similar manner with one of the robust Lambertian approach. We also assume the standard Gaussian likelihood function for the first-level, diffuse errors giving

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{e}') = N(\boldsymbol{y}; A\boldsymbol{x} + \boldsymbol{e}', \lambda W^{-1}), \tag{4.22}$$

Note that we define $W^{-1} \triangleq \text{diag}[1, \cdots, 1, 0] \in \mathbb{R}^{(m+1)\times(m+1)}$. We next apply an independent, zero-mean Gaussian prior distributions on both $\boldsymbol{x}$ and $\boldsymbol{e}'$:

$$p(\boldsymbol{x}) = N(\boldsymbol{x}; \boldsymbol{0}, \Sigma_x), \quad p(\boldsymbol{e}') = N(\boldsymbol{e}'; \boldsymbol{0}, \Gamma). \tag{4.23}$$

Σ_x describes the prior variance of the unknown normal vector $\boldsymbol{n}$ and model parameters $\boldsymbol{a}$ as $\Sigma_x = \text{diag}(\sigma_n^2 I_{(3)}, \sigma_a^2 I_{(p)})$ where $I_{(k)} \in \mathbb{R}^{k\times k}$ is the identity matrix; they are fixed to convey our lack of *a priori* certainty about $\boldsymbol{x}$. Thus the prior on $\boldsymbol{x}$ will be relatively uninformative (however, it is natural to assume the prior distribution of $\boldsymbol{n}$ and $\boldsymbol{a}$ independently). The values of σ_n^2 and σ_a^2 will be discussed further below. In contrast, $\Gamma \triangleq \text{diag}[\boldsymbol{\gamma}]$ is a fully-parameterized, diagonal matrix, where $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_{m+1}]^T$ is a non-negative vector of variances in one-to-one correspondence with elements of $\boldsymbol{e}'$. A large variance γ_i indicates that the corresponding e'_i is free to reflect the data, compensating for non-diffuse effects (outliers), while a small or zero-valued variance implies that the associated error term is constrained near zero (inliers). While we are ignorant of which observations are outliers, $\gamma_{(m+1)}$ is fixed to be zero because of the constraint regarding $e'_{(m+1)}$ in Eq. (4.21).

Combining the likelihood and prior using Bayes' rule leads to the posterior distribution $p(\boldsymbol{x}, \boldsymbol{e}'|\boldsymbol{y}) \propto$

$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{e}')p(\boldsymbol{x})p(\boldsymbol{e}')$. To estimate $\boldsymbol{x}$, we may further marginalize over $\boldsymbol{e}'$ to give

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x}, \boldsymbol{e}'|\boldsymbol{y}) d\boldsymbol{e}' = N(\boldsymbol{x}; \mu, \Sigma), \tag{4.24}$$

with mean and covariance defined as

$$\begin{aligned} \mu &= \Sigma A^T \left(\Gamma + \lambda W^{-1}\right)^{-1} \boldsymbol{y}, \\ \Sigma &= \left[\Sigma_x^{-1} + A^T \left(\Gamma + \lambda W^{-1}\right)^{-1} A\right]^{-1}. \end{aligned} \tag{4.25}$$

With the same derivation in Section 3.1.3, the update rules for the $(k+1)$-th iteration are given by

$$\begin{aligned} \gamma_i^{(k+1)} &\leftarrow \left(z_i^{(k)}\right)^2 + u_i^{(k)}, \forall i, \quad \Gamma^{(k+1)} = \text{diag}[\boldsymbol{\gamma}^{(k+1)}] \\ \boldsymbol{z}^{(k+1)} &\leftarrow \Gamma^{(k+1)} \left(S^{(k+1)}\right)^{-1} \boldsymbol{y} \\ \boldsymbol{u}^{(k+1)} &\leftarrow \text{diag}\left[\Gamma^{(k+1)} - \left(\Gamma^{(k+1)}\right)^2 \left(S^{(k+1)}\right)^{-1}\right], \end{aligned} \tag{4.26}$$

where $S^{(k+1)}$ is computed via

$$\begin{aligned} S^{(k+1)} &= D - DA\left[\Sigma_x^{-1} + A^T D A\right]^{-1} A^T D \\ &\quad \text{and} \quad D \triangleq (\Gamma^{(k+1)} + \lambda W^{-1})^{-1}. \end{aligned} \tag{4.27}$$

These expressions only require $O(n)$ computations and are guaranteed to reduce $L(\Gamma)$ until a fixed point Γ_* is reached. This value can then be plugged into Eq. (4.25) to estimate the unknown normal vector and model parameters. We denote this point estimator as $\boldsymbol{x}_{sbl}$. If the variances Γ_* reflect the true profile of the sparse errors, then $\boldsymbol{x}_{sbl}$ will closely approximate the true surface normal. This claim will be quantified more explicitly in the next section. We remind that the inverse matrix in Eq. (3.30) can also be efficiently computed by using the matrix inversion lemma in the similar manner with Eq. (3.18).

The choice of λ and σ_n^2 have been already presented in Section 3.1.3. Even if the image formation model is different from one in Chapter 3, the prior distribution of the surface normal $\boldsymbol{n}$ would not be change. On the other hand, the optimal λ can be changed since the distribution of outlines would change accompanied with the update of the image formation model, however actually the sufficiently large value of λ is usually enough since the sparse regression framework generally does not differentiate the type of outliers. Therefore, it can be reasonably set according to our prior expectations regarding the magnitudes of diffuse modeling errors, but in practice there is considerable flexibility here since some diffuse errors will be absorbed into $\boldsymbol{e}'$.

However, we need to discuss the effects of σ_a^2 on our algorithm. Since the value of σ_a^2 regularizes the shape of the piecewise linear function Eq. (4.13), smaller values will prevent the reverse of the sign of each linear segment and enforce the monotonicity of the function. While too small of a σ_a^2 may limit the generality of $g(I)$, it can be proven that as $\sigma_a^2 \to 0$ the model naturally reduces to a simple, Lambertian form for the diffuse component. We will empirically determine an appropriate value for σ_a^2 in Section 4.4.2.

4.4 Experimental Results

In this section, we quantitatively evaluate our method on synthetic and real image data. All experiments were performed on an Intel Core2 Duo E6400 (2.13GHz, single thread) machine with 4GB RAM and were implemented in MATLAB. For the SBL-based method we used $\lambda = 1.0^{-6}$ in the synthetic experiments with no additive image noise and perfect lighting calibrations (Section Section 4.4.2), and $\lambda = 10^{-2}$ for the other cases (Section 3.2.4). We set $\sigma_n^2 = 10^6$ for all experiments. As for σ_a^2, which can affect the solution when $p > 1$, we experimentally find the optimal value in Section 4.4.2-(a).

4.4.1 Datasets

For quantitatively evaluating our method, four different datasets are used. We generate 32-bit HDR images of two target scenes, *Bunny* (256 × 256) and *Caesar* (300 × 400), illuminated under random directional lightings. We use a few different BRDF settings for rendering; (A) combination of Lafortune diffuse reflection [65] and Cook-Torrance specularity[3], and (B) MERL BRDF database [74] [4]. Additionally, as the forth dataset, denoted (C), we record real images for qualitatively evaluating our method in a practical scenario. For the datasets (A), (B), both cast shadows and attached shadows are also rendered. To synthesize the cast shadow in the images rendered with MERL BRDF database, we firstly generate the ”cast shadow mask” using the ray-tracing software [1]. Then, both cast shadow mask and images rendered with MERL BRDF are merged.

---

[3] The details about these BRDF will be presented in Appendix B.

[4] A complete illustration of rendered images are given in Appendix C.

Note that though light calibration is noise-free, and shadowed pixel intensities are exactly zero in the synthesized datasets, some calibration errors and non-zero shadowed pixels exist in the real data (C).

4.4.2 Performance Evaluation with $p > 1$

We begin by the case $p > 1$ since the piecewise linear model with $p = 1$ is equivalent to the Lambertian reflectance model which has been already discussed in Chapter 3.

On the other hand, if the majority of observations are represented by non-linear diffuse reflections, then piecewise linear sparse regression with $p > 1$ basis functions or segments is expected to be considerably more effective. In this experiment, we evaluate the p-functions piecewise sparse linear regression to those complex objects by using the dataset (A) rendered with non-linear Lafortune diffuse reflectance model [65] and Cook-Torrance Reflectance model, and (B) rendered with one hundred BRDF functions from the MERL BRDF database [74].

Here, in addition to the SBL-based piecewise linear sparse regression (PL-SBL), we further implement the method which is solving Eq. (4.17) (PL-LS). We compare our methods with the R-PCA-based method [124] and a recent parametric non-Lambertian photometric stereo method with biquadratic reflectance model [99], which reasonably represents the low-frequency of non-linear reflectance though shadows and high-frequency observations must be removed in advance. The biquadratic model is a approximated BRDF as follow,

$$\begin{aligned} I = {} & \alpha_1(\boldsymbol{n}^T\boldsymbol{l})^2(\boldsymbol{l}^T\boldsymbol{h})^2 + \alpha_2(\boldsymbol{n}^T\boldsymbol{l})^2(\boldsymbol{l}^T\boldsymbol{h}) + \alpha_3(\boldsymbol{n}^T\boldsymbol{l})^2 \\ & + \alpha_4(\boldsymbol{n}^T\boldsymbol{l})(\boldsymbol{l}^T\boldsymbol{h})^2 + \alpha_5(\boldsymbol{n}^T\boldsymbol{l})(\boldsymbol{l}^T\boldsymbol{h}) + \alpha_6(\boldsymbol{n}^T\boldsymbol{l}) \\ & + \alpha_7(\boldsymbol{l}^T\boldsymbol{h})^2 + \alpha_8(\boldsymbol{l}^T\boldsymbol{h}) + \alpha_9, \end{aligned} \tag{4.28}$$

where $\boldsymbol{h}$ is a half vector directed at $\boldsymbol{l} + \boldsymbol{v}$ and $\boldsymbol{\alpha} \triangleq [\alpha_1, \alpha_2, \ldots, \alpha_9]$ are model parameters. In the implementation of [99], we used the non-shadowed pixels whose intensities are ranked below the 25%, and both surface normals and model parameters are iteratively updated initialized by the Lambertian photometric stereo method. Note that, shadowed pixels are also rejected, but all observations are taken into account in other methods.

Choosing σ_a^2 and p

Here we examine the choices for σ_a^2 and for the number of basis functions p in piecewise linear model by using dataset (B). Note that if p is large enough, we can essentially represent any complex non-linear function, although we run the risk of over-fitting when m is too small. However, we can always compensate to some extent via σ_a^2, since regardless of p, for σ_a^2 sufficiently small we approximate the standard Lambertian model with all linear segments having equal slope, and therefore equal diffuse albedo.

In this experiment, we vary individually σ_a^2 and p to find the optimal parameters. The results are illustrated in Fig. 4-6. Here, average angular errors in each normal map are further averaged over 100 materials. We observe that as expected, too small σ_a^2 deteriorates the performance, yet does not affect the performance when the value is sufficiently large. Therefore, in the following experiments, we fix σ_a^2 by 1.0. As for the number of basis functions, it appears that $p = 3$ is optimal in the case of dataset (B).

We also illustrate per-material angular errors with different p in Fig. 4-7. We observe that piecewise linear function with many basis functions, *e.g.*, 6, works very well under the material with narrow specular peaks, *e.g.*, for (1) specular-white-phenolic and (4) gray-plastic, while it suffers from over-fitting in polished metal with the broad specularity, *e.g.*, for (82) silver-metallic paint or materials with complex 2-lobe BRDF, *e.g.*, (53) natural-209 and (83) ipswich-pine-221), where we can hardly see the underlying monotone diffuse structure in observations.

Quantitative comparison with other methods

We evaluate of the performance of our method by a numerical comparison with other methods using datasets (A) and (B). Here, our methods (PL-SBL, PL-LS) are compared with the biquadratic photometric stereo method by Shi *et al.* [99] (Biquadratic) and LS.

First, we use dataset (A) to verify that our inverse piecewise linear diffuse model reasonably handles non-Lambertian diffuse reflections. In the Lafortune model [65], a general rotationally symmetric diffuse component ρ_d is written as $\rho_d = (\boldsymbol{n}^T\boldsymbol{l})^k(\boldsymbol{n}^T\boldsymbol{v})^k$. Here k is a model parameter which is fixed to 3.0 in our experiment. Note that specularities of dataset (A) are rendered with Cook-Torrance model. Therefore we can consider dataset (A) as a combination of dominant general diffuse reflections and sparse specular reflections.

The results are illustrated in Fig. 4-8 and Table 4.1. We present the results of PL-SBL with $p = 1$ and $p = 3$, and PL-LS with $p = 3$. Note that PL-LS with $p = 1$ is exactly same with LS. We observe that while non-Lambertian diffuse reflections degrade the performances of LS and PL-SBL($p = 1$), Biquadratic, PL-LS($p = 3$) and PL-SBL($p = 3$) work better since they are potentially

Table 4.1: Comparison among different methods with dataset (A).

No. of Images	Mean error (in degrees)				
	LS	BQ	PL-SBL (p=1)	PL-LS (p=3)	PL-SBL (p=3)
5	**4.9**	6.9	5.4	25.1	6.0
10	3.7	4.4	3.3	2.6	**1.5**
15	3.7	3.5	3.5	1.4	**0.80**
20	3.8	2.8	3.6	1.3	**0.67**
25	3.8	2.2	3.6	1.3	**0.58**
30	3.7	1.9	3.5	1.2	**0.49**
35	3.8	1.7	3.6	1.2	**0.47**
40	3.8	1.6	3.7	1.2	**0.49**

capable of non-Lambertian reflections. We also observe that PL-SBL($p = 3$) performs best since the estimation of Biquadratic and PL-LS are disrupted by specularities which are not included in each model while the sparsity penalty in PL-SBL($p = 3$) reasonably neglects them as the model outlier.

Secondly, we evaluated the performance of our method by using dataset (B). The results are illustrated in Fig. 4-9. We observe that while SBL($p = 1$) outperforms R-PCA and LS in most of materials, the average angular errors are large for various complex materials. In contrast, PL-SBL($p = 3$) works better for those kind of objects since the piecewise linear function with several basis functions can capture the nonlinearity of the non-Lambertian diffuse structure. We also observe that the Biquadratic model is the most effective method of all for dataset (B), however we emphasize that our method achieves competitive performance even without heuristically filtering high-frequency specularities. Consequently, on real images our approach produces qualitatively superior results (see Section 3.2.4 below). Finally, we can see from the comparison between PL-LS($p = 3$) and PL-SBL($p = 3$) that the sparsity penalty is also helpful for large p in our method, which can reject specularities efficiently.

We also compare our method with others in the presence of shadows and image noises. The results are illustrated in Table 4.2. We observe that our SBL-based methods still work under attached/cast shadows and image noises in contrast that Biquadratic and least-squares-based methods are easily disrupted by those corruptions.

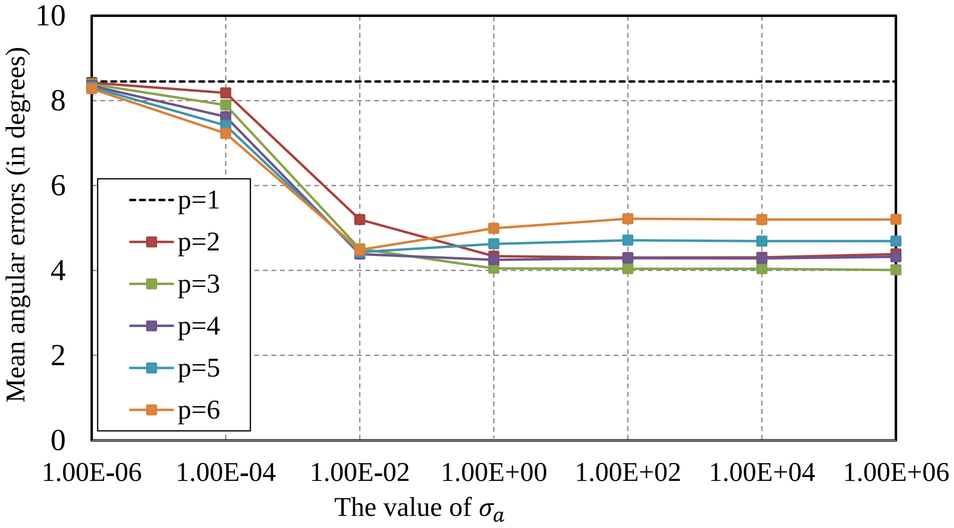


Figure 4-6: Experimental results of dataset (A) with different values of p and σ_a^2. The results are averaged over 100 different materials.

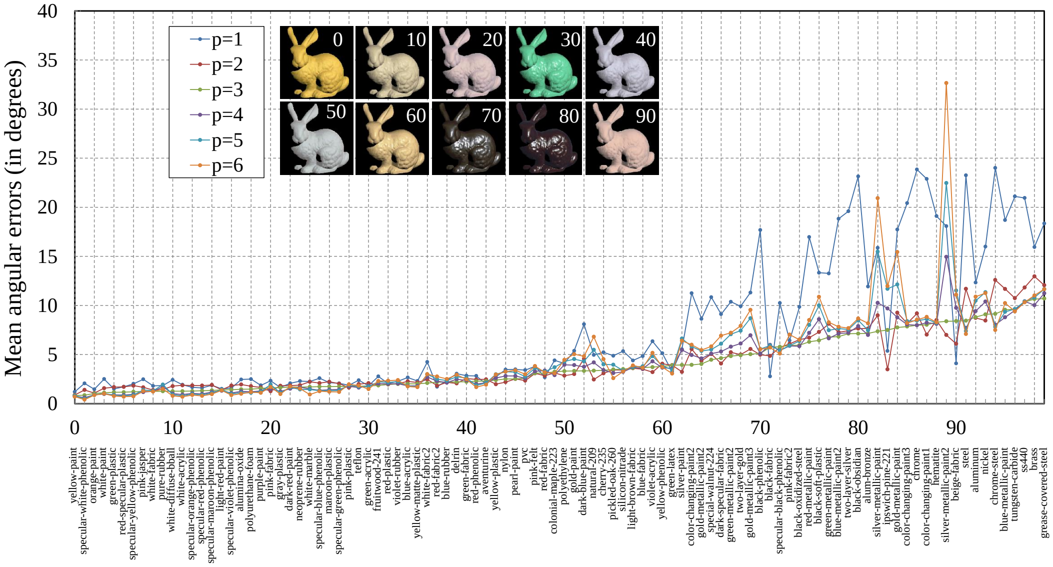
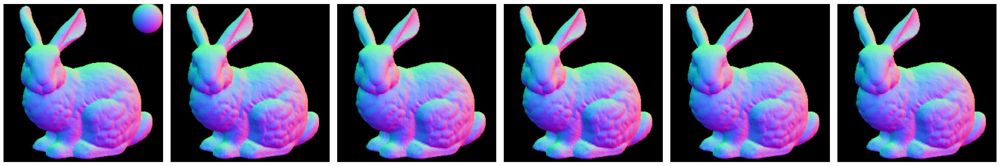
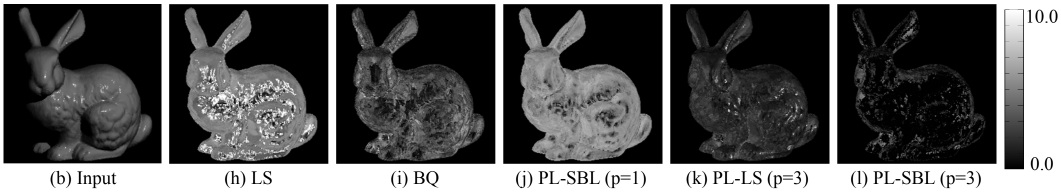


Figure 4-7: Experimental results of dataset (B) under the different number of basis functions [74]. We aligned results in ascending order of mean angular error of PL-SBL($p = 3$). Some rendered images corresponding to each material ID are also illustrated. A complete illustration of rendered images are shown in Appendix C.



(a) Ground Truth (c) LS (d) BQ (e) PL-SBL (p=1) (f) PL-LS (p=3) (g) PL-SBL (p=3)



(b) Input (h) LS (i) BQ (j) PL-SBL (p=1) (k) PL-LS (p=3) (l) PL-SBL (p=3)

Figure 4-8: Recovery of surface normals from 40 images of *Bunny* in dataset (A) with explicit shadow removal. (a) Input, (b) Ground truth, (c)-(l) Recovered surface normals and Error maps (in degrees).

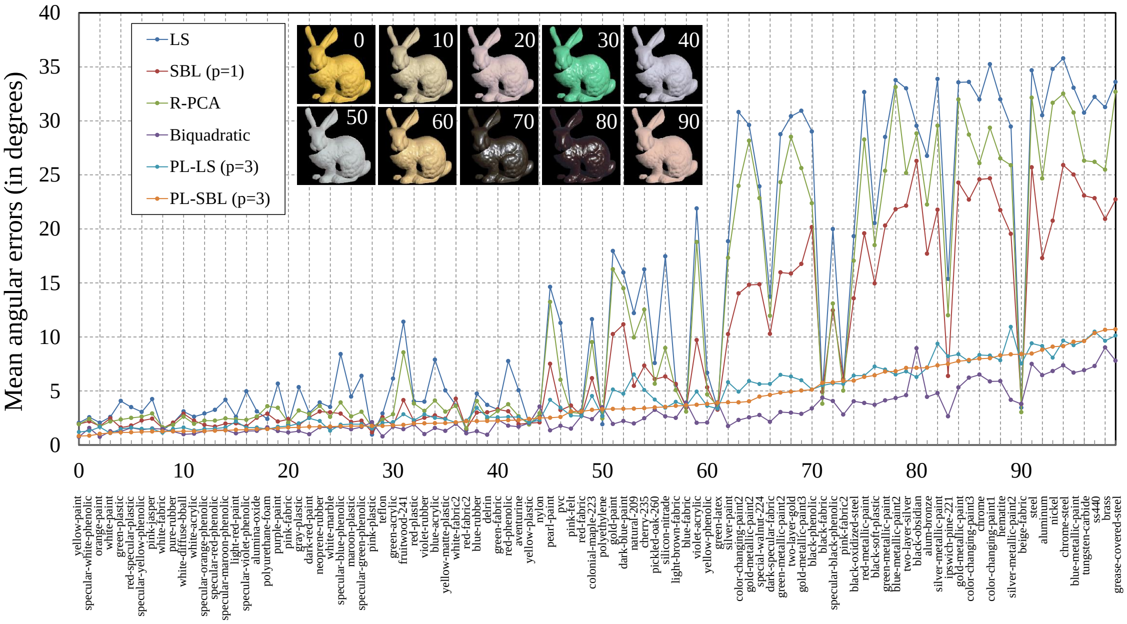


Figure 4-9: Comparison among different methods with dataset (B). We aligned results in ascending order of mean angular error of PL-SBL($p = 3$). Some rendered images corresponding to each material ID are also illustrated. A complete illustration of rendered images are shown in Appendix C.

Table 4.2: Experimental results of dataset (B) under the different kind of corruptions. This table illustrates average angular errors of surface normals for 100 materials.

	Mean error (in degrees)				
	LS	SBL (p=1)	Biquadratic	PL-LS (p=3)	PL-SBL (p=3)
w/o shadow and noise	13.9	8.9	**2.9**	4.4	4.1
w/ shadow	14.3	**7.7**	44.0	28.0	10.5
w/ noise	35.8	8.2	11.0	14.1	**4.8**

4.4.3 Qualitative Evaluation with Real Images

We also evaluate our algorithm (PL-SBL) using real images (dataset (C)). The dataset (C) is exactly same with ones in Section 3.2.4 (the detailed acquisition setup and calibration procedures are presented in Section 3.2.4). We briefly remind that we have captured RAW images without gamma correction by Canon 30D camera with a 200[mm] tele-photo lens and set it 1.5[m] far from target object. Lighting conditions are randomly selected from a hemisphere whose radius is 1.5[m] with the object placed at the center. For calibrating light sources, a glossy sphere was placed in the scene. We use a set of 25 images of *Chocolate bear* (261×421), and 40 images each of *Doraemon* (269×420) and *Fat guy* (293×344). Note that in this experiment, we did not remove shadows from images by zero-intensity thresholding since even shadowed pixels have non-zero values due to the presence of slight ambient illumination, sensor satiations, low signal-to-noise ratio, inter-reflections between the object and the floor and so on. Therefore, we sort all observations in increasing order and use lowest 25% of observations for the low-intensity condition of Biquadratic (low). We also applied the biquadratic reflectance model using all observations denoted as Biquadratic (all). We evaluate the performance of PL-SBL by visual inspection of the output normal maps and recovered surface meshes by a poisson solver [4], which recovers surface meshes from a gradient map.

The results are illustrated in Fig. 4-10. First, we observe that PL-SBL($p = 1, p = 3$) succeeds in efficiently rejecting specularities and estimating smoother and more reasonable normal maps. While Biquadratic (low) has produced the highest performance in Section 4.4(b), it has substantial difficulty with real images where calibration errors, shadowed pixels, and sensor saturations are often mis-classified as low-frequency reflectance, leading to unpredictable errors in practice (see Table 4.2). We note that increasing the threshold for shadow removal might solve this problem, but the optimal threshold selection remains hard and scene-dependent. Although including

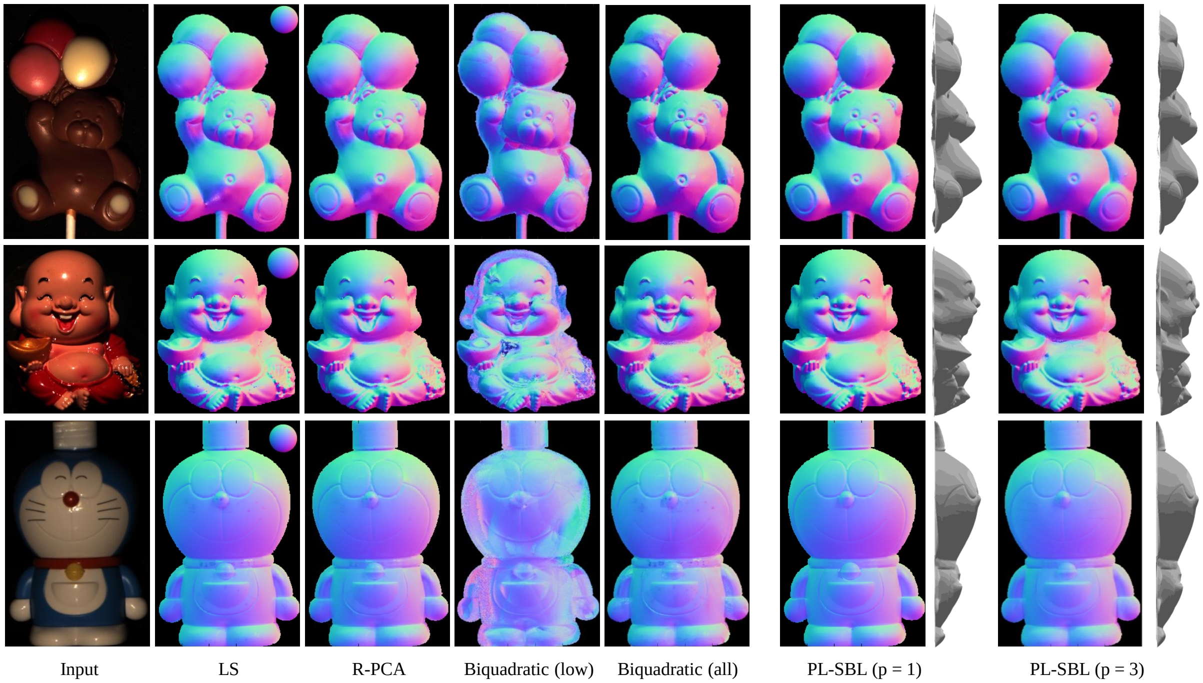


Figure 4-10: Experimental results with real datasets (C). We used three kind of datasets called *Chocolate bear* (25 images with 261×421), *Fat guy* (40 images with 293×344) and *Doraemon* (40 images with 269×420). We show example of input images, recovered surface normals and surface meshes (only for PL-SBL($p = 3$)).

higher frequency observations may sometimes give better results by diluting low-frequency corruptions with a larger number of samples (see Biquadratic (all)), accurate estimation still remains hard since the simple biquadratic model has difficulty representing complex, non-linear high-frequency observations (as is mentioned in [99]).

Second, we observe that reconstructed surface meshes by PL-SBL($p = 3$) are more reasonable than PL-SBL($p = 1$) *e.g.*, the stomach of *fatguy* recovered by PL-SBL($p = 1$) is shaper than that of PL-SBL($p = 3$), though we can not make the further quantitative comparison due to the lack of the ground truth. From those observations, we can say that our method is effective, especially when $p > 1$, in the practical situation.

4.5 Conclusion

In this chapter, we have demonstrated that the spare regression analysis presented in Chapter 3 is also performed on general, nonlinear diffuse reflectance model assuming the monotonicity of the diffuse reflectance function. The inverse diffuse reflectance model and its derivative piecewise lin-

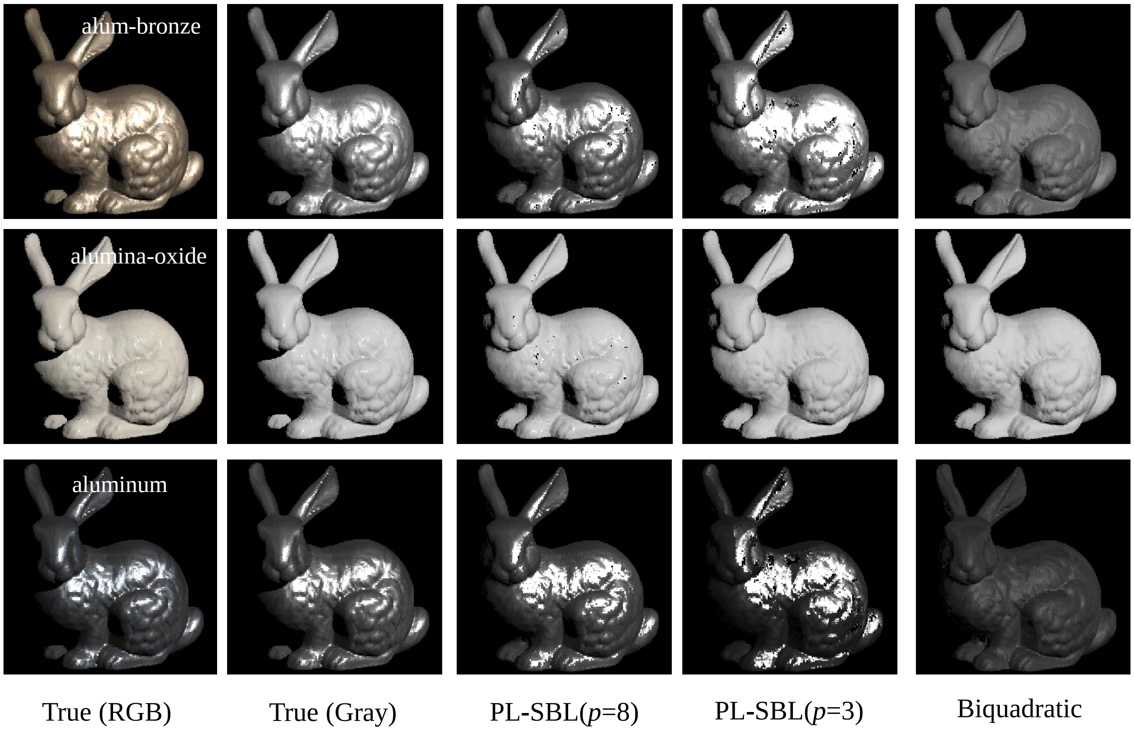


Figure 4-11: Image synthesis under a novel lighting using recovered reflectance of metallic objects. First two columns show input RGB and gray-scale images and following three columns show rendered images using recovered surface normal and reflectance function.

ear approximation of the diffuse reflectance function reasonably resolved the nonlinearity of the problem and the difficult, and complex ℓ_0-minimization problem is efficiently solved by a simple, hierarchical Bayesian approach called SBL in the quite similar manner with the Lambertian photometric stereo problem that was discussed in Section 3.

Our extensive evaluation with both synthetic and real images indicated that our method works even though the inlier reflectance model is not assumed to be Lambertian and in the presence of various non-diffuse corruptions, as well as our method performs well on real images where the shadow removal is not easy.

While we have presented in Chapter 3 and Chapter 4 that the robustness of sparse regression analysis for the photometric stereo problem. The limitation of ”general diffuse or Lambertian” + ”sparse non-diffuse outliers” model is that the regression fails in the absence of the unique dominant diffuse structure in observations. Therefore our method cannot potentially handle materials with unknown, non-diffusive dominant structure *e.g.*, materials with unknown BRDF where dominant structure is rough specularities or 2-lobe BRDF which is represented by a sum of several functions whose dominant directions are unknown *polyethylene and natural ones like fabrics*. For instance, metallic materials such as *alum-bronze*, *alumina-oxide* and *aluminum* in MERL BRDF database do not have dense, dominant diffusive structure, therefore problematic in our framework. We illustrate recovered reflectance functions by PL-SBL ($p = 3$,$p = 8$) and Biquadratic in Fig. 4-11. We observe that our piecewise linear diffuse function can not capture the diffuse reflections since the appearance of the metallic object is composed of *sparse* diffuse reflections and *dense* specular reflections. Note that the biquadratic model [99] well represent low-frequency diffuse reflections since it removes specular reflections before the regression analysis.

In the next chapter, we switch gears and examine how we can handle those those complex materials.

Chapter 5

Constrained Bivariate Regression for General Isotropic Surfaces

Previously, we have argued that the sparse regression analysis reasonably neglects various kind of artifacts which can not be represented by the inlier reflectance model. In addition, we have argued that the sparse Bayesian learning (SBL) framework is particularly well-suited for finding maximally sparse outliers, showing that it has stable and more focal estimate than a popular ℓ_1-norm based relaxation. Furthermore, we also proved that even if the original image formation model (*i.e.*, underlying regression model) is non-linear (therefore almost intractable to perform the sparse regression based on the ℓ_0 minimization.), the non-linearity can be resolved if the reflectance function is reasonably assumed to be monotonic (therefore invertible) and its inverse representation is linear. On the other hand, so far we have limited the form of reflectance function to be *bivariate i.e.*, reflectance of the target material is represented as a single lobe $f(\boldsymbol{n}^T\boldsymbol{\alpha})$, where its preferred direction $\boldsymbol{\alpha}$ is known (assumed to be the incident lighting vector in Chapter 3 and Chapter 4).

This chapter motivates a photometric stereo problem for accurately estimating the surface normals of a general isotropic scene whose reflectance lobe's number and preferred directions are unknown. For this purpose, we begin by representing general isotropic reflectance by using a convenient and flexible sum-of-lobe representation [23]. Then, we prove that if there is an additional constraint on lobe's preferred direction, pixelwise appearances are well-approximated by a bivariate monotonic, smooth function of the dot-products between the surface normal and the lighting direction, and between lighting and viewing directions. We may then consider the inverse representation of the image formation process, where the unknown normal vector is now separated from

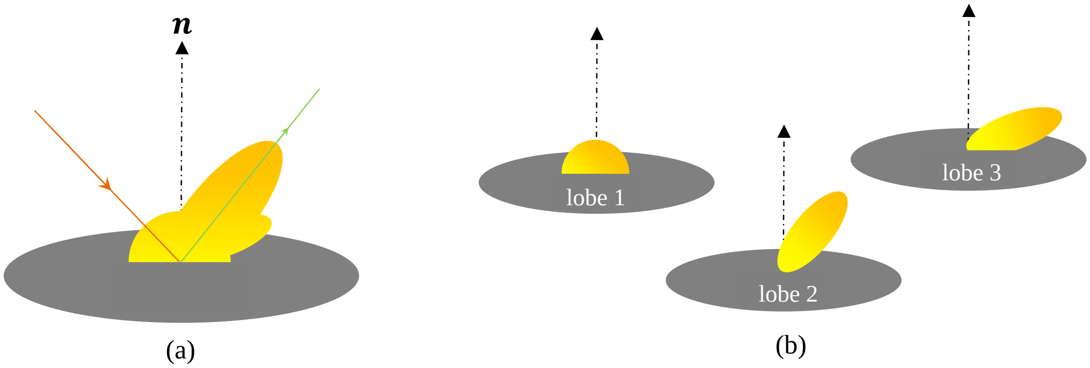


Figure 5-1: Illustration of k-lobe BRDF. The BRDF in the real world is generally composed of k different reflectance lobes which are pointing at different *preferred* directions. In the illustration, a k-lobe BRDF in (a) is a sum of three different basis BRDFs in (b).

the unknown monotonic inverse reflectance function. By parameterizing the latter using a Bernstein polynomials [71], we obtain a set of constrained linear equations in both the surface normals and reflectance parameters, leading to a simple, quadratic programming problem.

In similar to previous algorithms proposed in Chapter 3 and Chapter 4, the proposed framework benefits from the efficient pixelwise optimization, which is easily amenable to parallel processing and does not require typical smoothness constraints for both object structure and reflectance, which can disrupt the recovery of fine details.

5.1 Photometric stereo using constrained bivariate regression

In this section, we formulate the photometric stereo as a constrained bivariate regression problem. Henceforth we also rely on the following assumptions:

(1) relative position between the camera and the object is fixed across all images.

(2) object is illuminated by a point light source at infinity from varying and known directions.

(3) camera view is orthographic, and the radiometric response function is linear.

5.1.1 Problem Statement

As has been shown in Section 2.1, diverse appearances of real world objects can be encoded by a BRDF (ρ), which relates the observed intensity I at a given point on the object to the associated surface normal $\boldsymbol{n} \in \mathbb{R}^3$, the incoming lighting direction $\boldsymbol{l} \in \mathbb{R}^3$, and the outgoing viewing direction

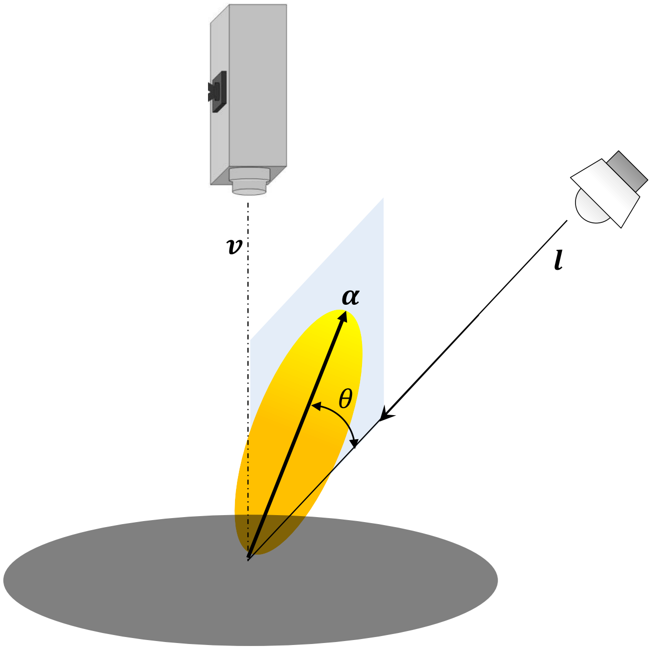


Figure 5-2: Illustration of the restriction on $\boldsymbol{\alpha}$. We assume that the preferred direction of each basis BRDF is given by a unit vector lying on the plane spanned by incident lighting direction and outgoing viewing direction. Introducing this constraint leads to 1-DOF representation of $\boldsymbol{\alpha}$ (*i.e.*, any $\boldsymbol{\alpha}$ is represented by θ).

$\boldsymbol{v} \in \mathbb{R}^3$ via

$$I = \rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) \max{(\boldsymbol{n}^T \boldsymbol{l}, 0)}, \tag{5.1}$$

where $\max{(\boldsymbol{n}^T \boldsymbol{l}, 0)}$ accounts for attached shadows. We note that in this chapter, we do not assume the existence of the additive corruption (ϵ in Eq. (1.2)) such as shadows and sensor noises for simplifying the discussion. The photometric stereo problem that we tackle here is to recover the surface normal $\boldsymbol{n}$ of a scene by inversely solving Eq. (5.1) from a collection of m observations under different, known lighting conditions.

Recently, it has been revealed by Chandraker and Ramamoorthi [23] that an isotropic BRDF consists of sum of lobes whose contribution to the reflected intensity decreases monotonically as the surface normal deviates away from the direction where the reflectance lobe is concentrated (*i.e.*, referred to as a *preferred direction*). The semi-parametric model of isotropic BRDF that is represented as a sum of K different univariate functions is presented as

$$\rho = \sum_{k=1}^{K} \rho_k(\boldsymbol{n}^T \boldsymbol{\alpha}_k). \tag{5.2}$$

Here ρ_k are (unknown) non-linear functions, and $\boldsymbol{\alpha}_k$ (*i.e.*, $||\boldsymbol{\alpha}_k|| = 1$) are called *preferred directions*, along which ρ_k are concentrated (A comprehensive illustration of k-lobe BRDF is shown in Fig. 5-1). It is known that physically valid reflectance functions satisfy following requirements:
(L1) Monotonicity: ${\rho_k}' > 0$.
(L2) Non-negativity: $\rho_k \geq 0$.
(L3) Passing thorough the origin: $\rho_k(0) = 0$.

It is shown that inversely solving Eq. (5.2) under known surface normals gives good estimation of wide varieties of isotropic BRDF without suffering from the curse of dimensionality [23]. Unfortunately, however, directly solving Eq. (5.2) in the context of the photometric stereo problem is prohibitively difficult since there are so many unknown parameters, as well as some of them are coincident in the same term (*i.e.*, $\boldsymbol{n}, \boldsymbol{\alpha}_k, \rho_k$). Therefore, following this remark, the surface normal has been recovered utilizing monotonicity of the reflectance function under the assumption that the number of lobe is one and its preferred directions is known (*e.g.*, lighting direction in [58] and half vector in [9, 100]). While effective, these methods are highly disruptive when approximation of lobes are incorrect or the reflectance function is composed of two or more lobes. Furthermore, to our best knowledge, there is no work which has achieved the simultaneous estimation of both azimuth and elevation angles by enforcing the monotonicity of the reflectance function whose preferred direction is different from the lighting vector (Note that [9, 100] assume the azimuth angle of the surface normal is known). Although Alldrin *et al.* [9] proposed a photometric stereo method which works for unknown lobes, they assume that the azimuth angle of the surface normal is known.

Instead, we only assume that the preferred direction ($\boldsymbol{\alpha}_k$) of each function (ρ_k) is lying on the plane spanned by a lighting direction $\boldsymbol{l}$ and a viewing direction $\boldsymbol{v}$ as

$$\boldsymbol{\alpha}_k = \frac{p_k\boldsymbol{l} + q_k\boldsymbol{v}}{||p_k\boldsymbol{l} + q_k\boldsymbol{v}||}, \tag{5.3}$$

where p_k and q_k are non-negative unknown values (*i.e.*, $p_k \geq 0$ and $q_k \geq 0$). Note that the degree of freedom of $\boldsymbol{\alpha}_k$ is actually one because $||\boldsymbol{\alpha}_k|| = 1$ (Illustration of this constraint is shown in Fig. 5-2). We also note that this assumption does not violate most existing physically-based models as well as measured isotropic BRDF (*e.g.*, *blue-fabric* in MERL BRDF database [74] has two directions, that are $4\boldsymbol{l} + 3\boldsymbol{v}$ and $3\boldsymbol{l} + 4\boldsymbol{v}$ as reported in [23]). Then, this assumption provides us following important result.

Theorem: Suppose there is no shadow at a surface point (*i.e.*, $\forall i\ \boldsymbol{n}^T\boldsymbol{l}_i,\ \boldsymbol{l}_i^T\boldsymbol{v} \geq 0$ and $I_i \geq 0$, where i is the index of the light) and $\rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v})$ in Eq. (5.1) has a form of Eq. (5.2), whose parameters satisfy the requirements of physically valid BRDF **(L1)**-**(L3)** and Eq. (5.3). Then, it is guaranteed that there exists at least one continuous bivariate function $f(x, y)\ \forall x, y \in [0, 1]$, which satisfies $f \geq 0, \partial f/\partial x > 0, \partial f/\partial y \leq 0$ and $\forall i\ I_i = f(\boldsymbol{n}^T\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v})$.

Proof: From Eq. (5.3), $\boldsymbol{n}^T\boldsymbol{\alpha}_k$ is transformed into

$$\boldsymbol{n}^T\boldsymbol{\alpha}_k = \frac{p_k\boldsymbol{n}^T\boldsymbol{l} + q_k\boldsymbol{n}^T\boldsymbol{v}}{\sqrt{p_k^2 + q_k^2 + 2p_kq_k\boldsymbol{l}^T\boldsymbol{v}}}. \tag{5.4}$$

Here we used $||\boldsymbol{l}|| = ||\boldsymbol{v}|| = 1$. Eq. (5.4) illustrates that $\boldsymbol{n}^T\boldsymbol{\alpha}_k$ is non-decreasing for $\boldsymbol{n}^T\boldsymbol{l}$ with fixed $\boldsymbol{l}^T\boldsymbol{v}$ and non-increasing for $\boldsymbol{l}^T\boldsymbol{v}$ with fixed $\boldsymbol{n}^T\boldsymbol{l}$ since p, q are non-negative constant values and $\boldsymbol{n}^T\boldsymbol{v}$ is constant over different lightings. From **(L1)**, it is guaranteed that each $\rho_k(\boldsymbol{n}^T\boldsymbol{\alpha}_k)$ is also nondecreasing/non-increasing for $\boldsymbol{n}^T\boldsymbol{l}$ and $\boldsymbol{l}^T\boldsymbol{v}$ when either of them is fixed. Integrating these results into Eq. (5.1) and Eq. (5.2), it is proved that I is monotonic increasing for $\boldsymbol{n}^T\boldsymbol{l}$ with fixed $\boldsymbol{l}^T\boldsymbol{v}$ and non-increasing for $\boldsymbol{l}^T\boldsymbol{v}$ with fixed $\boldsymbol{n}^T\boldsymbol{l}$, which implies we can always define continuous functions $f(x, y)$ which satisfy $f \geq 0,\ \partial f/\partial x > 0,\ \ \partial f/\partial y \leq 0$ and $\forall i\ I_i = f(\boldsymbol{n}^T\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v})$ since $I_i \geq 0$ and $\forall i\ 0 \leq \boldsymbol{n}^T\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v} \leq 1$.

We illustrate this theorem in Fig. 5-3-(a). We note that assuming $f(x, y)$ is always passing through the y-axis (*i.e.*, $f(0, y) = 0$) does not limit any kind of isotropic BRDF represented by Eq. (5.1) because $I = \boldsymbol{n}^T\boldsymbol{l}\rho$.

Following this theorem, we formulate the photometric stereo as a constrained bivariate regression problem whose goal is to recover a combination of an unknown surface normal $\boldsymbol{n}$ and a continuous bivariate function f from a collection of lighting directions $\boldsymbol{l}_i$ and associated appearances I_i $(i = 1, \ldots, m)$, which satisfies following equations and constraints,

$$I_i = f(\boldsymbol{n}^T\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v}) \quad i = 1, \ldots, m, \tag{5.5}$$

(L4) Monotonicity (x): $\partial f/\partial x > 0$.

(L5) Monotonicity (y): $\partial f/\partial y \leq 0$.

(L6) Non-negativity: $f \geq 0$.

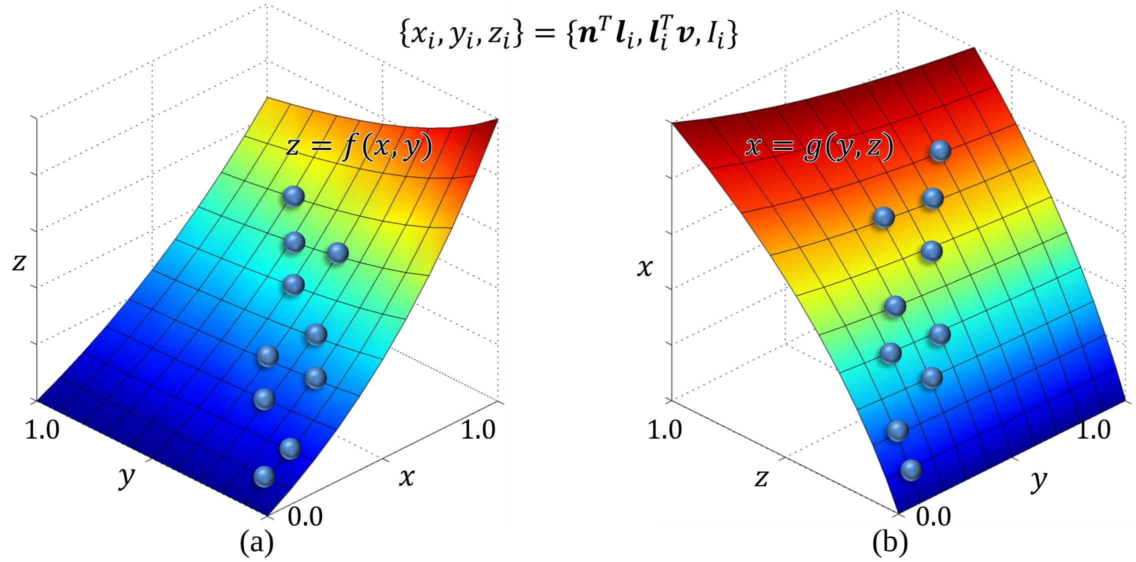


Figure 5-3: (a) Collections of $\{x, y, z\} = \{\boldsymbol{n}^T\boldsymbol{l}, \boldsymbol{l}^T\boldsymbol{v}, I\}$ are lying on a continuous function of $z = f(x, y)$ which satisfies $\partial f/\partial x > 0$, $\partial f/\partial y \leq 0$ and $f(0, y) = 0$. (b) 3-d points lying on f are also lying on a inverse function ($x = g(y, z)$). The illustration indicates that they are uniquely convertible if $\forall x\ \partial f/\partial x > 0$.

(L7) Passing through y-axis: $f(0, y) = 0$.

We call Eq. (5.5) the *forward bivariate reflectance model*. The major benefit of this problem formulation is that we do not need to explicitly approximate the number of lobes K and their preferred directions $\boldsymbol{\alpha}$, that means any kind of isotropic materials are tractable. However, there is a critical issue which must be solved: the coincidence of unknown parameters $\boldsymbol{n}$ and f in the same term. We overcome this difficulty by a convenient, inverse representation of the imaging model applied through a constrained bivariate regression framework.

5.1.2 Inverse Bivariate Reflectance Model

Strict monotonicity of $f(x, y)$ **(L4)** guarantees the unique existence of the function giving $x = g(y, f(x, y)) = g(y, z)$, which obeys following requirements:

(L8) Monotonicity (y): $\partial g/\partial y \geq 0$.

(L9) Monotonicity (z): $\partial g/\partial z > 0$.

(L10) Non-negativity: $g \geq 0$.

(L11) Passing through y-axis: $g(y, 0) = 0$.

The proof, which has been omitted for brevity, is obvious by seeing Fig. 5-3-(b). From the definition, each 3-d point of $\{x, y, z\} = \{\boldsymbol{n}^T\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v}, I_i\}(i = 1, \ldots, m)$ lying on f is also lying on g as follow,

$$\boldsymbol{n}^T\boldsymbol{l}_i = g(\boldsymbol{l}_i^T\boldsymbol{v}, I_i) \quad i = 1, \ldots, m. \tag{5.6}$$

In contradiction to Eq. (5.5), we call Eq. (5.6) the *inverse bivariate reflectance model*. Our goal is now updated to recover the surface normal $\boldsymbol{n}$ and a continuous bivariate function g with some shape restrictions **(L8)**-**(L11)**. The fundamental advantage of Eq. (5.6) is that unknown variables of $\boldsymbol{n}$ and g are separated, which contributes to simplifying the problem.

While constraints on g limit the solution space of Eq. (5.6), there are still multiple feasible solutions of a pair of $\boldsymbol{n}$ and g since $\{\boldsymbol{l}_i^T\boldsymbol{v}, I_i\}$ are sparsely distributed on the valid range of $\{y, z\}$. To reduce inherent ambiguity of the problem, we further assume a parametric model of the inverse bivariate reflectance function $g(y, z)$. Given that the left-hand-side of Eq. (5.6) is linear in the unknown normal vector $\boldsymbol{n}$, for computational simplicity we would like to impose similar linearity on the right-hand-side in our parameterized representation of $g(\boldsymbol{l}^T\boldsymbol{v}, I)$ (we have omitted subscripts for simplicity). For this purpose, we then choose to express $g(\boldsymbol{l}^T\boldsymbol{v}, I)$ as a summation over p fixed and known, non-linear basis functions $g_k(\boldsymbol{l}^T\boldsymbol{v}, I)$ weighted by an unknown coefficient vector $\boldsymbol{\beta}$, leading to the representation

$$g(\boldsymbol{l}^T\boldsymbol{v}, I) = \sum_{k=1}^{p} \beta_k g_k(\boldsymbol{l}^T\boldsymbol{v}, I). \tag{5.7}$$

While non-linear in $\boldsymbol{l}^T\boldsymbol{v}$ and I, $g(\boldsymbol{l}^T\boldsymbol{v}, I)$ is clearly linear in $\boldsymbol{\beta} \triangleq [\beta_1, \ldots, \beta_p]^T$. We need to choose g_k carefully since estimating a multivariate regression function subject to shape restrictions with compact support is challenging and usually very time consuming [117]. The choice of g_k will be discussed later.

By substituting Eq. (5.13) into the inverse bivariate reflectance model, Eq. (5.6) becomes

$$\boldsymbol{n}^T\boldsymbol{l}_i = \boldsymbol{\beta}^T G(\boldsymbol{l}_i^T\boldsymbol{v}, I_i) \quad i = 1, \ldots, m, \tag{5.8}$$

where $G \triangleq [g_1(\boldsymbol{l}_i^T\boldsymbol{v}, I_i), \cdots, g_p(\boldsymbol{l}_i^T\boldsymbol{v}, I_i)]^T$. Collecting variations of observation at the same pixel under different lighting directions, Eq. (5.8) can be merged into following linear problem,

$$L^T\boldsymbol{n} = G^T\boldsymbol{\beta}. \tag{5.9}$$

Here, $L \triangleq [\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m]$. By merging unknown variables $(\boldsymbol{n}, \boldsymbol{\beta})$, this problem is transformed as

$$P\boldsymbol{x} = [L^T - G^T]\boldsymbol{x} = 0, \tag{5.10}$$

where $\boldsymbol{x} \triangleq [n_x, n_y, n_z, \beta_0, \ldots, \beta_p]^T$ and n_x, n_y, n_z are the three elements of the surface normal. Without loss of generality, we may avoid the degenerate $\boldsymbol{x} = 0$ solution to Eq. (5.19) by constraining $\sum_i x_i = 1$, which implies $\boldsymbol{c}^T\boldsymbol{x} = 1$ where $\boldsymbol{c} = [1, \ldots, 1]^T$.

Given the appearance variations $(I_1, I_2, \ldots, I_m)$ under different known lighting conditions $(\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_m)$, the optimal surface normal $(\boldsymbol{n})$ and model parameters $(\boldsymbol{\beta})$ are recovered by solving the constrained linear problem,

$$\min_{\boldsymbol{x}} ||P\boldsymbol{x}||_2^2, \quad \text{s.t. } \textbf{(L8)} - \textbf{(L11)}, \;\; c^T\boldsymbol{x} = 1, \tag{5.11}$$

where **(L8)**-**(L11)** are summarized as monotonicity, non-negativity and boundary condition of g_k where the concrete, mathematical form of the constraints are determined when we choose the basis function g_k. Here, it is worth mentioning that if all of the constraints in Eq. (5.11) are represented by a *linear* form (*e.g.*, $A\boldsymbol{\beta} = \boldsymbol{b}$ or $A\boldsymbol{\beta} \leq \boldsymbol{b}$), Eq. (5.11) is known to be considered as a standard quadratic programming problem, where many convenient estimators have been proposed to date.

5.2 Selection of Basis Function

There has been increasing interest in estimating a multivariate regression function subject to various constraints, such as nonnegativity, monotonicity, convexity and concavity among many others. Performing such shape-restricted regression analysis is more challenging for multivariate independent variables, therefore the literature about the topic is relatively scarce (we refer the reader to the comprehensive survey about the shape restricted multi-variate regression by Wang [117]).

Here we adopt bivariate Bernstein polynomials [71], where the shape-restricted regression function estimate is shown to be the solution of a quadratic programming problem [19, 117]; making it computationally attractive. Furthermore, the Bernstein polynomials approximation naturally selects smooth functions with little computational effort unlike other non-parametric regression function (*e.g.*, smoothing spline [15]), which implicitly enforces the smoothness of BRDF like [9]. Bivariate

Bernstein polynomials [71] are composed of multiple basis functions of the form,

$$\begin{aligned} &b_{k_1,k_2}(x_1,x_2,N_1,N_2) = b_{k_1}(x_1,N_1)b_{k_2}(x_2,N_2), \\ &b_{k_i}(x_i,N_i) = \begin{pmatrix} N_i \\ k_i \end{pmatrix} x^{k_i}(1-x_i)^{N_i-k_i} \ \ (i=1,2), \end{aligned} \tag{5.12}$$

where $0 \le x_i \le 1$ and N_i is the order of the polynomial as for x_i, which will be chosen as a function of the sample size m (*e.g.*, $N_i = o(m_i^{\gamma})$ with $\gamma_i > 0$ suitably chosen via the popular V-fold cross-validation method as shown in [117]). We illustrate the Bernstein basis function in the case of $\{N_1 = N_2 = 2\}$ in Fig. 5-4. We transform $I_i (i = 1, \ldots, m)$ to lie in the unit $[0,1]$ via a simple linear equation as $I_i = I_i / \max(\boldsymbol{I})$. Note that $\boldsymbol{l}_i^T \boldsymbol{v}$ naturally lies in $[0,1]$ since we only consider the case $v = [0,0,1]^T$ and $l_z > 0$. Then, bivariate Bernstein polynomials approximation of g is represented as,

$$\begin{aligned} x = g(y,z) &= \boldsymbol{\beta}^T \boldsymbol{b}_{N_y,N_z}(y,z) \\ &= \sum_{k_y=0}^{N_y} \sum_{k_z=0}^{N_z} \beta_{k_y,k_z} b_{k_y,k_z}(y,z,N_y,N_z), \end{aligned} \tag{5.13}$$

where $\boldsymbol{b}_{N_y,N_z} \triangleq [b_{0,0}, \ldots, b_{N_y,N_z}]^T \in \mathbb{R}^{(N_y+1)(N_z+1)\times 1}$ and $\boldsymbol{\beta} \triangleq [\beta_{0,0}, \ldots, \beta_{N_y,N_z}]^T \in \mathbb{R}^{(N_y+1)(N_z+1)\times 1}$.

Unlike the B-splines procedure (which may require quadratic constraints on the coefficients) [15], shape restrictions (*e.g.*, monotonicity, non-negativity) on Eq. (5.13) are easily encoded via linear constraints, that is $A\boldsymbol{\beta} \ge 0$ and $C\boldsymbol{\beta} = 0$, where A, C are shape restriction matrices. Following [117], the shape restriction matrices required for our problem are defined as follows,

(1) **Monotonicity**: $\partial g/\partial y \ge 0$ and $\partial g/\partial z \ge 0$. **(L8),(L9)**
The first order partial derivatives of g with respect to y in Eq. (5.13) can be represented as

$$\begin{aligned} &\partial g(y,z)/\partial y \\ &= N_y \sum_{k_z=0}^{N_z} \sum_{k_y=0}^{N_y-1} (\beta_{k_y+1,k_z} - \beta_{k_y,k_z}) b_{k_y,k_z}(y,z,N_y-1,N_z). \end{aligned} \tag{5.14}$$

Hence the non-decreasing constraint (*i.e.*$\partial g/\partial y \ge 0$) is simply achieved by enforcing $\beta_{k_y,k_z} \le \beta_{k_y+1,k_z}$ for $k_y = 1, \ldots, N_y - 1$. The non-decreasing constraint with respect to z (*i.e.*$\partial g/\partial z \ge 0$) is also achieved in the same manner. The restriction matrix for a linear constraint $A_{mono}\boldsymbol{\beta} \ge 0$ is rep-

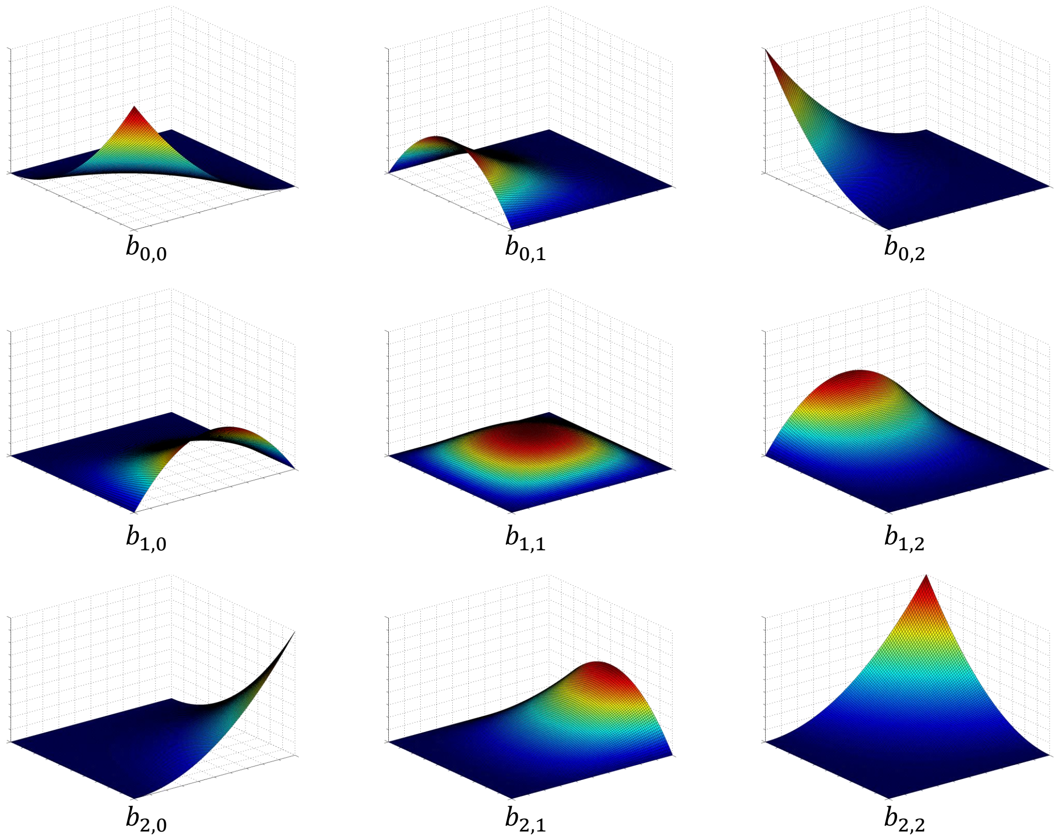


Figure 5-4: 2-nd order Bernstein basis function.

resented as $A_{mono} = [A_y^T A_z^T]^T$ that is composed of sub-matrices $A_y \in \mathbb{R}^{N_y(N_z+1)\times(N_y+1)(N_z+1)}$ and $A_z \in \mathbb{R}^{N_z(N_y+1)\times(N_y+1)(N_z+1)}$, where A_r ensures the monotonicity of the function with respect to r (The concrete form of the matrix is included in the supplementary). Note that the strict monotonicity constraint $\partial g/\partial z > 0$ is eased to $\partial g/\partial z \geq 0$ for computational simplicity. the restriction matrix for $\partial g/\partial y \geq 0$ and $\partial g/\partial z \geq 0$ is represented as $A_m = [A_y^T A_z^T]^T$ which is composed of sub-matrices $A_y \in \mathbb{R}^{N_y(N_z+1)\times(N_y+1)(N_z+1)}$ and $A_z \in \mathbb{R}^{N_z(N_y+1)\times(N_y+1)(N_z+1)}$, where A_r ensures the monotonicity of the function with respect to r. The concrete form of each sub-matrix is represented as follow,

$$A_y = \begin{pmatrix} -1 & 0 & \dots & 0 & 1 & & & & & \\ & -1 & 0 & \dots & 0 & 1 & & & & \\ & & & & & \ddots & & & & \\ & & & & & -1 & 0 & \dots & 0 & 1 \\ & & & & & & & & & \end{pmatrix}. \tag{5.15}$$

$$A_z = \begin{pmatrix} B & & & \\ & B & & \\ & & \ddots & \\ & & & B \end{pmatrix}, B = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & & & -1 & 1 \end{pmatrix}.$$

Note that there are N_z of 0 between -1 and 1 for each row of A_y. When we apply the retro-reflection detection which was presented in Sec. 3 of our submission, we firstly use A_y to solve the problem as described in Sec. 2.2 of our submission, and then we replace A_y in Eq. (5.15) by $A'_y = -A_y$ to capture the retro-reflective behavior in the observation.

(2) **Non-negativity**: $g \geq 0$. **(L10)**

From the definition in Eq. (5.12), it is easy to show that all Bernstein basis polynomials are non-negative with respect to $0 \leq y, z \leq 1$. Therefore, the non-negativity of g is guaranteed when $\forall i\ \beta_i \geq 0$. Henceforth, the restriction matrix for a linear constraint $A_{nonneg}\boldsymbol{\beta} \geq 0$ is as $A_{nonneg} \triangleq diag([1, \dots, 1]) \in \mathbb{R}^{(N_y+1)(N_z+1)\times(N_y+1)(N_z+1)}$.

(3) **Passing through y-axis**: $g(y, 0) = 0$. **(L11)**

From the definition in Eq. (5.12), $b_{k_y,k_z}(y, 0) = 0$ for all $k_z \neq 0$. Therefore $g(y, 0) = \sum_{k_y=0}^{N_y} \beta_{k_y,0} b_{k_y,0}(y, 0, N_y, N_z)$ becomes zero for all y when $\forall k_y\ \ \beta_{k_y,0} = 0$. This constraint is encoded via a linear constraint $C\boldsymbol{\beta} = 0$ where $C \in \mathbb{R}^{(N_y+1)(N_z+1)\times(N_y+1)(N_z+1)} \triangleq diag([1, 0, \dots, 1, 0, \dots, 1, \dots])$ with N_z of 0 between 1.

5.2.1 Solution Method

By substituting Eq. (5.13) into the inverse bivariate reflectance model, Eq. (5.6) becomes

$$\boldsymbol{n}^T \boldsymbol{l}_i = \boldsymbol{\beta}^T \boldsymbol{b}_{N_y,N_z}(\boldsymbol{l}_i^T \boldsymbol{v}, I_i) \quad i = 1, \dots, m, \tag{5.16}$$

where coefficients of Bernstein polynomials ($\boldsymbol{\beta}$) are restricted via following equations,

$$A\boldsymbol{\beta} = \begin{bmatrix} A_{mono} \\ A_{nonneg} \end{bmatrix} \boldsymbol{\beta} \geq 0,\ C\boldsymbol{\beta} = 0. \tag{5.17}$$

Collecting variations of observation at the same pixel under different lighting directions, Eq. (5.16) can be merged into following linear problem,

$$L^T \boldsymbol{n} = B^T \boldsymbol{\beta}. \tag{5.18}$$

Here, $B \triangleq [\boldsymbol{b}_{N_y,N_z}(\boldsymbol{l}_1^T \boldsymbol{v}, I_1), \ldots, \boldsymbol{b}_{N_y,N_z}(\boldsymbol{l}_m^T \boldsymbol{v}, I_m)]$ and $L \triangleq [\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m]$. By merging unknown variables $(\boldsymbol{n}, \boldsymbol{\beta})$, this problem is transformed as

$$P\boldsymbol{x} = [L^T - B^T]\boldsymbol{x} = 0, \tag{5.19}$$

where $\boldsymbol{x} \triangleq [n_x, n_y, n_z, \beta_{0,0}, \ldots, \beta_{N_y,N_z}]^T$ and n_x, n_y, n_z are the three elements of the surface normal. Given the appearance variations $(I_1, I_2, \ldots, I_m)$ under different known lighting conditions $(\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_m)$, the optimal surface normal $(\boldsymbol{n})$ and model parameters $(\boldsymbol{\beta})$ are recovered by solving the constrained linear problem,

$$\min_{\boldsymbol{x}} ||P\boldsymbol{x}||_2^2, \quad \text{s.t. } \tilde{A}\boldsymbol{x} \geq 0 \text{ and } \tilde{C}\boldsymbol{x} = 0, \tag{5.20}$$

where $\tilde{A} \triangleq [\mathbf{0}\ A]$ and $\tilde{C} \triangleq \begin{bmatrix} \boldsymbol{c}^T \\ \mathbf{0}\ C \end{bmatrix}$. Eq. (5.20) can be effectively solved by the general quadratic programming since both equality and inequality constraints are expressed as the linear equation.

5.3 Handling Retro-reflective Materials

Our inverse bivariate reflectance model in Eq. (5.6) which is derived from the sum-of-lobe representation of general isotropic BRDF [23] does not have the ability to represent the retro-reflective behavior in the observations, which is the phenomenon of light rays striking a surface and being redirected back to the source of light. We show an example of retro-reflective object in Fig. 5-5.

The main reason is because the sum-of-lobes model in [23] was originally proposed under the assumption that both lighting and viewing directions are static, so the dependency on $\boldsymbol{l}^T \boldsymbol{v}$ was not considered in the model. Therefore, our derived bivariate inverse model also has trouble with handling retro-reflections which violates the monotonicity assumption for $\boldsymbol{l}^T \boldsymbol{v}$ (*i.e.*, **(L5)** in our submission). To clarify this point, we categorized reflections into three classes *i.e.*, diffuse, specular (including off-specular reflection) and retro reflection as shown in Fig. 5-6-(a). A retro-reflection provides smaller luminance as the difference between incident and ongoing directions increases

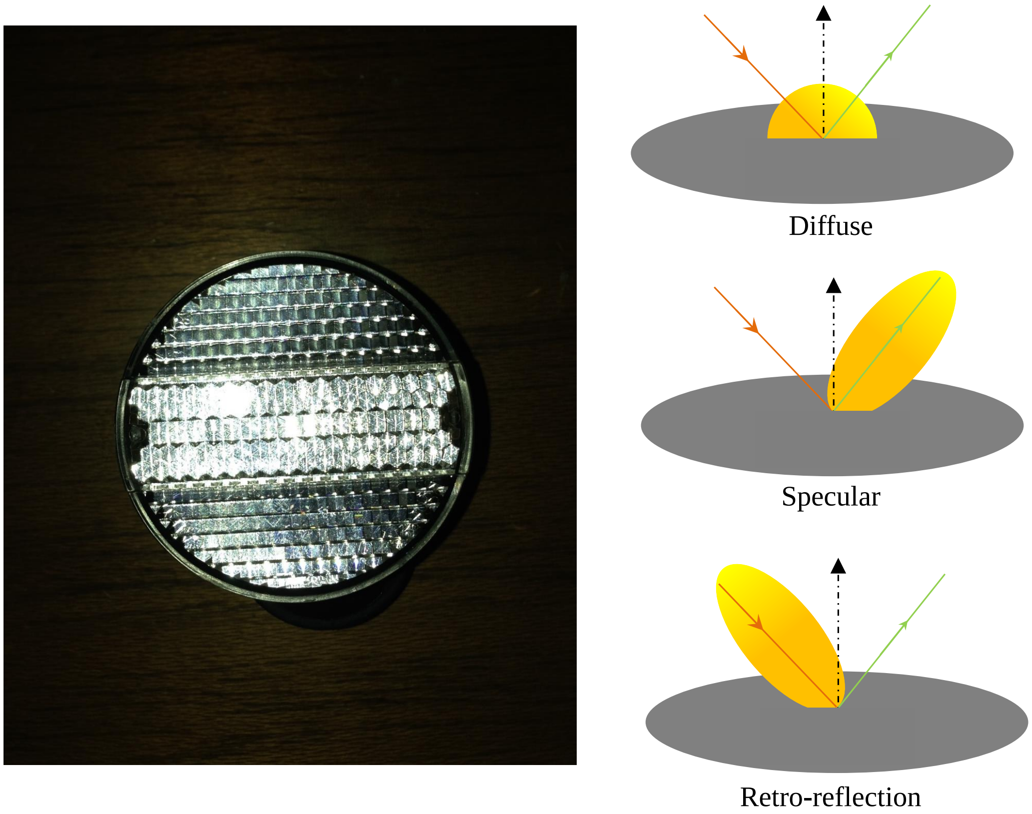


Figure 5-5: An example of the retro reflective object. (Left) Bicycle retroreflectors. Retroreflectors are devices that operate by returning light back to the light source along the same light direction. (Right) Illustration of reflectance lobes corresponding to (a) diffuse, (b) specular and (c) retro reflections.

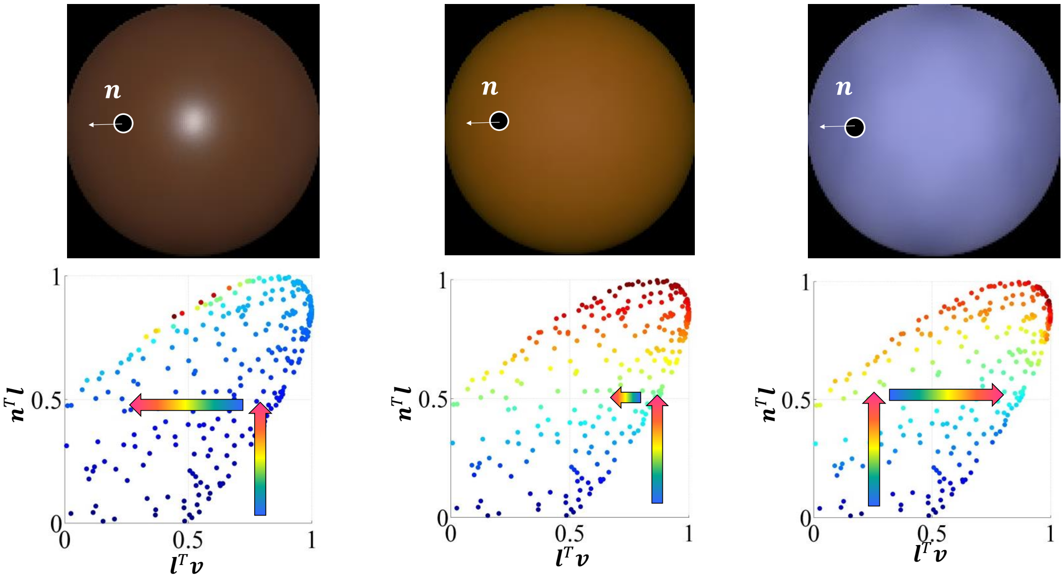
(*i.e.*, $I \propto \boldsymbol{l}^T\boldsymbol{v}$) in an opposite manner of other reflectance lobes. We confirm these phenomena by using some representative materials in MERL BRDF database [74]. In Fig. 5-6, we illustrated images which were rendered with BRDF of *ipswich-pine-221*, *orange-paint* and *blue-fabric* as examples of specular, diffuse and retro-reflective materials, respectively. In addition to them, plots of $(\boldsymbol{l}^T\boldsymbol{v}, \boldsymbol{n}^T\boldsymbol{l}, I)$ for fixed surface normal $\boldsymbol{n}$ and varying lightings $\boldsymbol{l}$ are illustrated by projecting them onto $\boldsymbol{l}^T\boldsymbol{v} - \boldsymbol{n}^T\boldsymbol{l}$ plane and expressing I using color (red plot has a large intensity). We observe that only *blue-fabric* violates our assumption *i.e.*, I is non-increasing for $\boldsymbol{l}^T\boldsymbol{v}$, which coincides with the observation that our naive method without retro-detection scheme had difficulty in *blue-fabric* as shown in Sec. 4.2 of our submission.

If the reflectance of a target object obeys Eq. (5.2), our method reasonably recovers the surface normal of the object by solving Eq. (5.20). However, one limitation of our method is that this assumption is not satisfied in the presence of retro-reflections which are often observed on rough surfaces since Eq. (5.2) does not have the ability to represent this kind of reflections. In the presence of retro-reflections, our surface normal estimation fails due to the violation of **(L5)** by the behavior of retro-reflections that the power of reflections increases as $\boldsymbol{l}^T\boldsymbol{v}$ increases. While it may limit available materials, fortunately we have found that our method practically handles those retro-reflective materials by simply reversing the direction of the monotonicity constraint on $\boldsymbol{l}^T\boldsymbol{v}$ in Eq. (5.20) since the retro reflections does not affect the monotonicity for $\boldsymbol{n}^T\boldsymbol{l}$ while unify the direction of the monotonicity for $\boldsymbol{l}^T\boldsymbol{v}$ over the BRDF space. The problem of course is that we do not know whether the material is retro-reflective or not. To overcome this difficulty, we present a practical approach to handle both non-retro-reflective and retro-reflective materials. The important observation is that when we incorrectly constrain the problem, the regression is usually failed. Therefore, once we got the regression outputs under both constrains, we can judge which constraint was optimal by examining regression errors.

However, we have empirically found that comparing regression errors in Eq. (5.20) does not work, instead we compute following linear regression error E for choosing the optimal solution:

$$\begin{aligned}\tilde{a} &= \arg\min_{a} \sum_{i=1}^{m} \|\hat{\boldsymbol{n}}^T\boldsymbol{l}_i - aI_i\|_2^2, && (5.21)\\ E &= \sum_{i=1}^{m} \|\hat{\boldsymbol{n}}^T\boldsymbol{l}_i - \tilde{a}I_i\|_2^2. && (5.22)\end{aligned}$$

Here, $\hat{\boldsymbol{n}}$ is a recovered surface normal by solving Eq. (5.20) under the monotonicity constrain for $\boldsymbol{l}^T\boldsymbol{v}$ in either of two directions. We simply choose the direction whose E is smaller than the other. This strategy is very simple but very efficient as we will show in Section 5.4. Assuming that the reflectance function as for fixed surface normal has one preferable monotonicity for $\boldsymbol{l}^T\boldsymbol{v}$ which would give more accurate estimation of the surface normal, the problem is, as described in our submission, how can we find the direction of monotonicity. We overcame this difficulty by three steps: (a) estimating surface normals under both constraints, (b) computing regression errors with recovered surface normals, (c) adopting the surface normal with smaller errors as the final estimation. As we have mentioned, we have examined two kind of regression errors from Eq. (5.20) and Eq. (5.22). In Fig. 5-7, we illustrate the regression results based on two constraints (*i.e.*, $\partial g/\partial y \geq 0$ or $\partial g/\partial y \geq 0$ in **(L9)** of our submission), where plots of $(\boldsymbol{l}^T\boldsymbol{v}, I, \boldsymbol{n}^T\boldsymbol{l})$ were overlaid with reconstructed inverse reflectance function represented by Bernstein polynomials (each

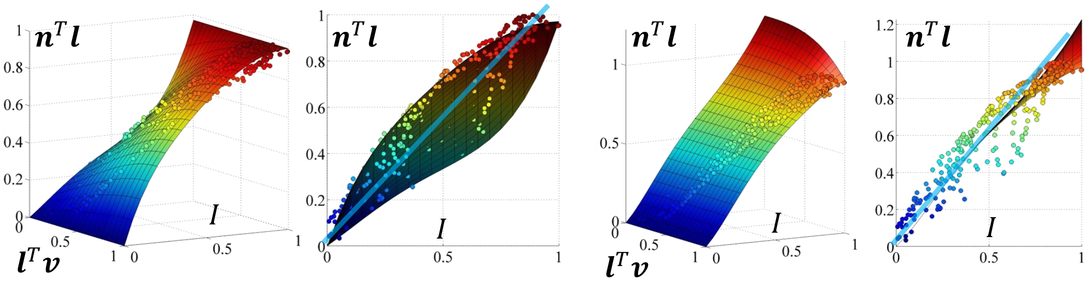


ipswich-pine-221 (specular) orange-paint (diffuse) blue-fabric (retro-reflective)

Figure 5-6: Illustration of 3-d plots of $(\boldsymbol{l}^T\boldsymbol{v}, \boldsymbol{n}^T\boldsymbol{l}, I)$ for three types of reflections. (Top) Rendered images of *ipswich-pine-221* (specular), *orange-paint* (diffuse) and *blue-fabric* (retro-reflective). (Bottom) 3-d plots of $(\boldsymbol{l}^T\boldsymbol{v}, \boldsymbol{n}^T\boldsymbol{l}, I)$ which are projected onto the $\boldsymbol{l}^T\boldsymbol{v} - \boldsymbol{n}^T\boldsymbol{l}$ space.

case is presented from two viewpoints). In addition, we also show regression errors from Eq. (5.20) and Eq. (5.22) for each material and an angular error of surface normal at the bottom of plots. As observed, regression errors from Eq. (15) do not work for determining the appropriate constraint in *blue-fabric* since the flexible Bernstein polynomials were well fitted to observations even though the constraint was not correct. On the other hand, the linear regression error from Eq. (5.22) reasonably increased in both *blue-fabric* and *ipswich-pine-221* when the constraint was incorrect. We observed this relationship for most materials in MERL BRDF database, that is why we adopted the linear regression error for our retro-reflection detection algorithm.

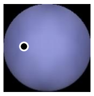
Note that this strategy is theoretically problematic in the case where *both* specularities and retro-refections are simultaneously observed *i.e.*, there is no monotonic variation of I in the $\boldsymbol{l}^T\boldsymbol{v}$ direction. However, we also note that these case are merely observed in the natural world, and if any, our retro-reflection detection algorithm still improves the result since there is usually one *dominant reflection* which provides more appropriate constraint for the problem. Furthermore, our regression scheme based on robust Bernstein polynomials usually suppresses the estimation errors caused by the non-dominant reflections (those observations are supported by our experimental results using MERL BRDF database [74] which will be shown in Section 5.4.2 and Appendix C.



R. Err = (0.44, 0.18), A. Err = 5.88

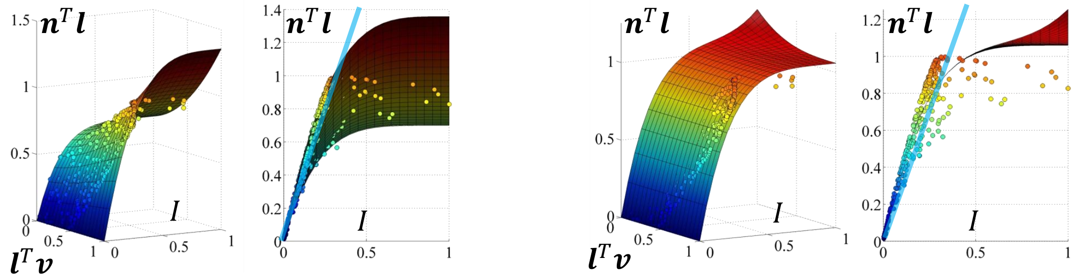
R. Err = (0.79, 0.12), A. Err = 1.06

(a) $\frac{\partial g}{\partial y} \geq 0$



blue-fabric

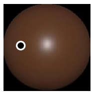
(b) $\frac{\partial g}{\partial y} \leq 0$



R. Err = (0.82, 0.23), A. Err = 0.96

R. Err = (0.86, 0.92), A. Err = 1.06

(c) $\frac{\partial g}{\partial y} \geq 0$



ipswich-pine-225

(d) $\frac{\partial g}{\partial y} \leq 0$

Figure 5-7: Illustration of regression results based on different constraints on $\boldsymbol{l}^T\boldsymbol{v}$. (a) $\partial g/\partial y \geq 0$ and (b) ($\partial g/\partial y \leq 0$) using *blue-fabric*, and (c) $\partial g/\partial y \geq 0$ and (d) ($\partial g/\partial y \leq 0$) using *ipswich-pine-225*. Each 3-d plot $(\boldsymbol{I}, \boldsymbol{l}^T\boldsymbol{v}, \boldsymbol{n}^T\boldsymbol{l})$ which were generated using recovered surface normal is overlaid with the reconstructed inverse bivariate reflectance function. At the second plots for each condition, we also overlaid the linear function fitted to recovered plots via a least-square regression. In addition, we show regression errors (R. Err) from Eq. (5.20) and Eq. (5.22) respectively with an angular error of recovered surface normal (A. Err).

5.3.1 Recovery of Spatially Varying Reflectance

While our main interest is to recover the normal of a surface point accurately, its spatially varying reflectance function can also be recovered once we got the converged surface normal estimation $\boldsymbol{n}^*$. We now consider the forward bivariate reflectance model Eq. (5.5) with the Bernstein polynomials approximation giving

$$I_i = f(\boldsymbol{n}^{*T}\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v}) = \boldsymbol{\alpha}^T\boldsymbol{b}_{N_x,N_y}(\boldsymbol{n}^{*T}\boldsymbol{l}_i, \boldsymbol{l}_i^T\boldsymbol{v}). \tag{5.23}$$

Collecting equations under different lighting directions, they are integrated into the constrained linear problem as

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{I} - B'\boldsymbol{\alpha}||_2^2, \quad \text{s.t. } A'\boldsymbol{\alpha} \geq 0 \text{ and } C'\boldsymbol{\alpha} = 0, \tag{5.24}$$

where $B' = [\boldsymbol{b}_{N_x,N_y}(\boldsymbol{n}^{*T}\boldsymbol{l}_1, \boldsymbol{l}_1^T\boldsymbol{v}), \ldots, \boldsymbol{b}_{N_x,N_y}(\boldsymbol{n}^{*T}\boldsymbol{l}_m, \boldsymbol{l}_m^T\boldsymbol{v})]$, and A', C' are shape restriction matrices which are defined as for **(L4)-(L7)**. Note that here we do not need to impose the sum constraint on $\boldsymbol{\alpha}$ (*i.e.*, $\sum_i \alpha_i = 1$). Once $\boldsymbol{\alpha}$ is recovered, the observation under the arbitrary lighting direction is given by solving Eq. (5.23).

We state that the perfect recovery of the reflectance is difficult by our method since we rely on the per-pixel sampling (*i.e.*, the reflectance function a surface point is *independently* recovered from the appearance variations on the surface point). While it becomes less problematic with sufficient number of samples (e.g., 300), merely observed phenomenon (e.g., highlight) is difficult to be encoded in the reflectance function with smaller samples since the smoothing effect of Bernstein polynomials robustly neglects outlier samples. However, we think it is not a critical limitation since even in that case, most observations are treated as inliers and normals are accurately recovered. We emphasize that our goal is to achieve pixel-wise and accurate normal recovery for various isotropic materials. It is worth mentioning that once the surface geometry is acquired, we can apply existing reflectance reconstruction techniques available under known geometry [23, 129].

5.4 Experimental Results

In this section, we evaluate our method on synthetic and real image data. All experiments were performed on an Intel Core i7-2640M (2.80GHz, single thread) machine with 8GB RAM and were implemented in MATLAB. For the quantitative evaluation, we generate 32-bit HDR images of sphere (256 × 256) with foreground masks under different BRDF settings; (A) common physical

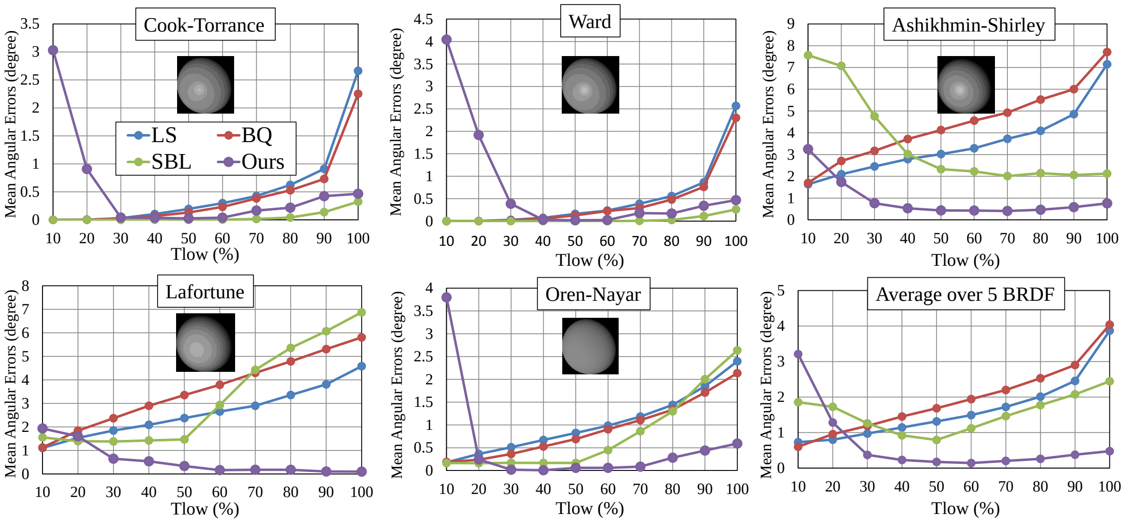


Figure 5-8: Experimental results of dataset (A) with five different BRDF with varying frequencies in observations.

or phenomenological BRDF (Cook-Torrance [29], Ward [118] , Lafortune [65], Oren-Nayar [83] and Ashikhmin-Shirley [13]) and (B) measured MERL BRDF database [74]. Lighting directions are randomly selected from a hemisphere with the object placed at the center. Additionally, for the third dataset, denoted (C), we use real images for qualitatively evaluating our method in the practical situations. For each dataset, shadows are removed via simple thresholding like other works such as [99]. Because ground truth surface normals are provided in dataset (A) and (B), we quantitatively evaluate our method by the angular error between recovered normal map and the ground truth when using these datasets.

5.4.1 Evaluation with Synthesized BRDF

We evaluate the performance using the synthesized images in dataset (A) generated under 100 different lightings using five common BRDF (detailed descriptions about each BRDF are presented in Appendix 1).

Here we compare our method with the standard Lambertian-based photometric stereo [122](LS) and a recent SBL-based robust approach [61] (SBL) (λ is fixed by 10^{-6}), which separates the non-Lambertian corruptions from observations via ℓ_0-based sparse regression. And our method is also compared with a recent parametric non-Lambertian photometric stereo method with the biquadratic reflectance model [99] (BQ). In this experiment, we fix (N_y, N_z) in Eq. (5.16) by $(1, 5)$ to examine the robustness of our method against those parameters.

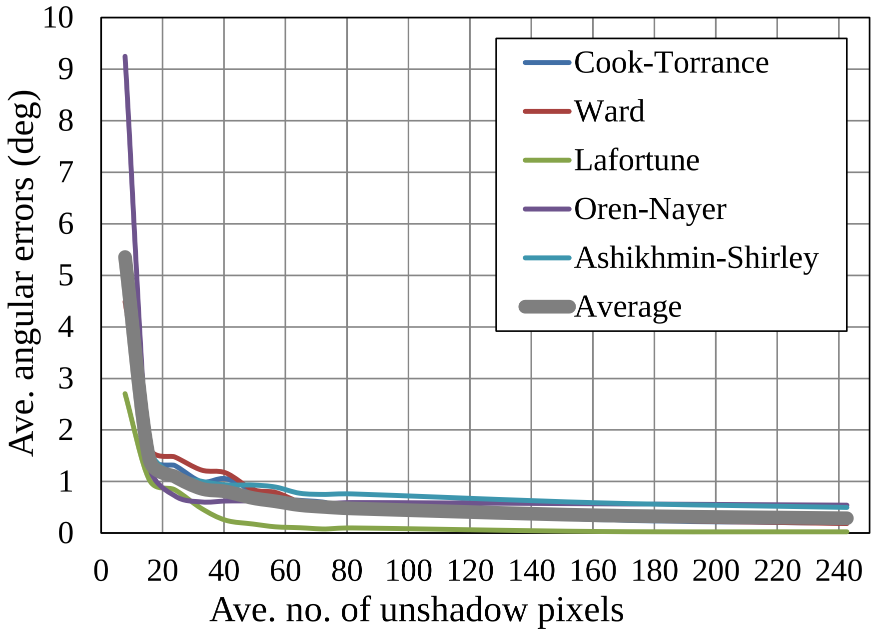


Figure 5-9: Experimental results of dataset (A) with varying number of images. For a fair evaluation, we display the average number of non-shadowed pixels which join our algorithm on x-axis instead of the number of images.

Evaluation with varying frequencies

We first evaluate our algorithm using observations with varying frequencies for presenting the flexibility of our model comparing with models used in previous works (*e.g.*, BQ[99]) which only work for low-frequency reflectances. Here high-frequency specularities are discarded by using the non-shadowed pixels whose intensities are ranked below the $T_{low}\%$ ($10 \leq T_{low} \leq 100$). The result is illustrated in Fig. 5-8. In totally, we observe that our method outperforms other algorithms in almost all frequencies (unfortunately, our method does not work when the number of images is very small as will be discussed in following sentence) for Lafortune model, Ashikhmin-Shirley model and Oren-Nayar model and performs competitively with SBL for Cook-Torrance and Ward model. Interestingly, our model works well when all of the frequencies are included unlike other method and also works for Oren-Nayar model which exhibit strong retro-reflective reflections due to our simple $\boldsymbol{l}^T\boldsymbol{v}$ constraint selection strategy.

Valid number of input images

We also evaluate our algorithm using varying number of images to find the valid number required for effective recovery. The results are displayed in Fig. 5-9. We observe that the minimum number

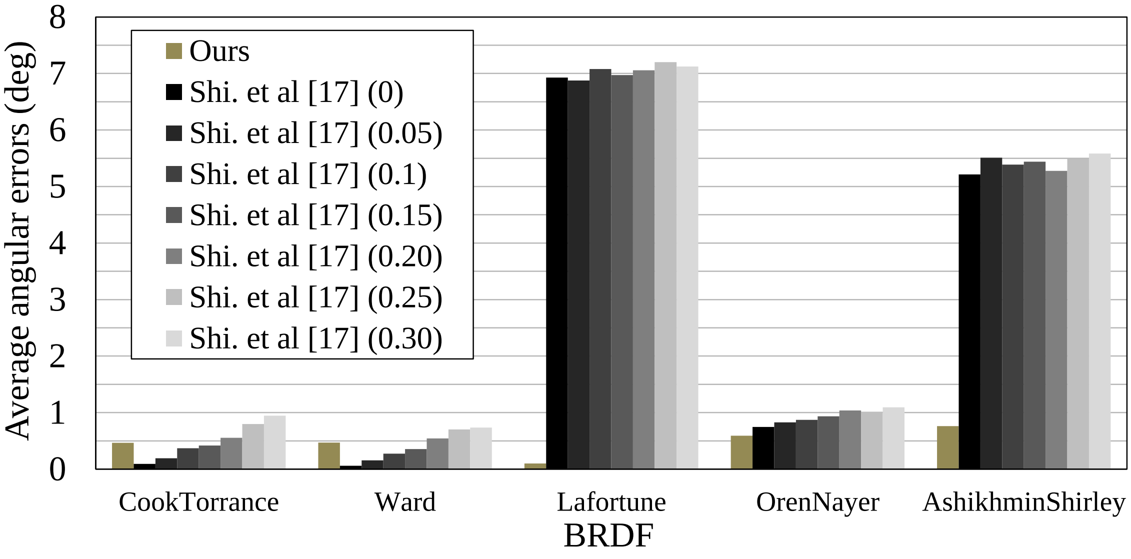


Figure 5-10: Comparison with Shi *et al*. [100]. The values in the bracket is the angular errors (deg) added into the azimuth angle map which is used in [100].

of images required to make the algorithm work is around 20 and more than 60 are required for the stable reconstruction. While these values may be relatively large in the photometric stereo literature but our method is applicable of various materials thanks to the complex constrained bivariate regression procedure.

Comparison with other monotonicity-based approach

In this section, we compared our method with recent elevation-angle estimation algorithm [100] assuming the dominant specular lobe pointing at the half vector direction. Since [100] requires an azimuth angle map as input while our method simultaneously recovers all elements in the normal, we compare our algorithm with [100] by gradually adding estimation errors in the azimuth angle map used as the input of [100]. The results are illustrated in Fig. 5-10. We observe that when the true azimuth angle is given, [100] outperforms our method in Cook-Torrance and Ward dataset. However, as the amount of errors increases, the differences become smaller and finally our method outperforms [100]. As we expected, [100] does not work for Lafortune and Ashikhmin-Shirley model since those model violate their assumption.

Evaluation of computational time

Here, we examine the computational time required for our computation. We tried various combinations of N_y and N_z and m in Eq. (5.13) and solved our optimization problem Eq. (5.16) using *lsqlin* function in Matlab. The evaluation results are illustrated in Fig. 5-11. Here we present a per-pixel

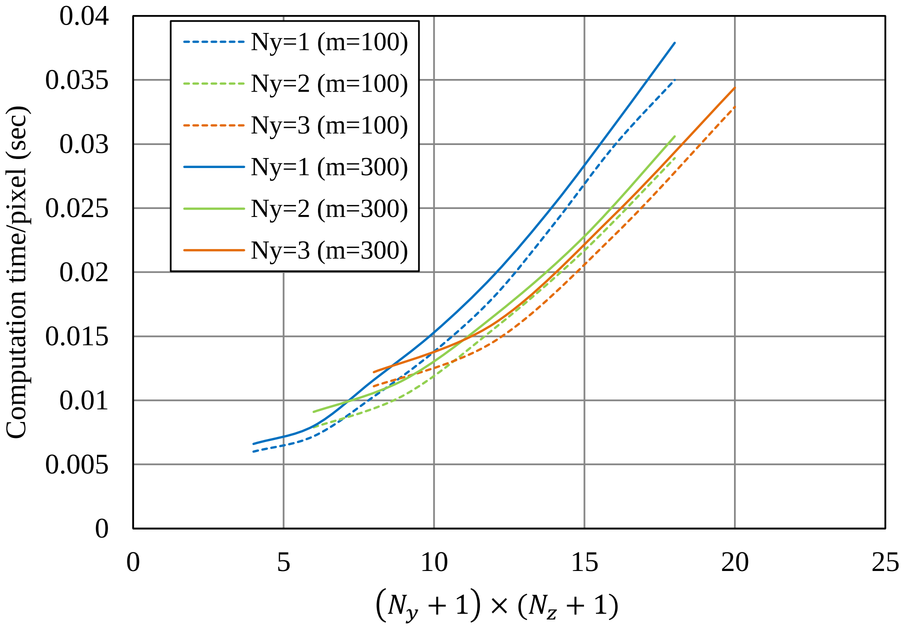


Figure 5-11: Evaluation of computational time. We illustrate the per-pixel computational time for each combination of the number of Bernstein basis functions and that of lightings.

computational time to solve *one* optimization problem. Therefore, the actual computational time is twice as large as one in the figure since we apply our algorithm twice to distinguish retro-reflective materials. We observe that the computational complexity depends on the number of basis functions rather than the number of lightings.

5.4.2 Evaluation with Measured BRDF

Here we evaluate the performance of our method to the dataset (B). We generate images under 300 different lightings for 100 different materials from the MERL BRDF database [74] (shadows are also removed in advance). In this experiment, our method is also compared with LS [122], SBL [61] and BQ [99]. LS SBL and our method use all the frequencies in observations (*i.e.*, $T_{low} = 100$) while only BQ is performed with both $T_{low} = 25$ and $T_{low} = 100$ since this model was originally designed for representing the low-frequency observations. In this experiment, we fixed $N_y = 3$ and $N_z = 5$ and performed our method with/without the retro-reflection detection algorithm described in Section 5.3 to verify the effectiveness of this process.

The results are illustrated in Fig. 5-12. We observe that our method with our efficient retro-reflection detection outperforms other algorithms for most of materials. While BQ ($T_{low} = 25$)

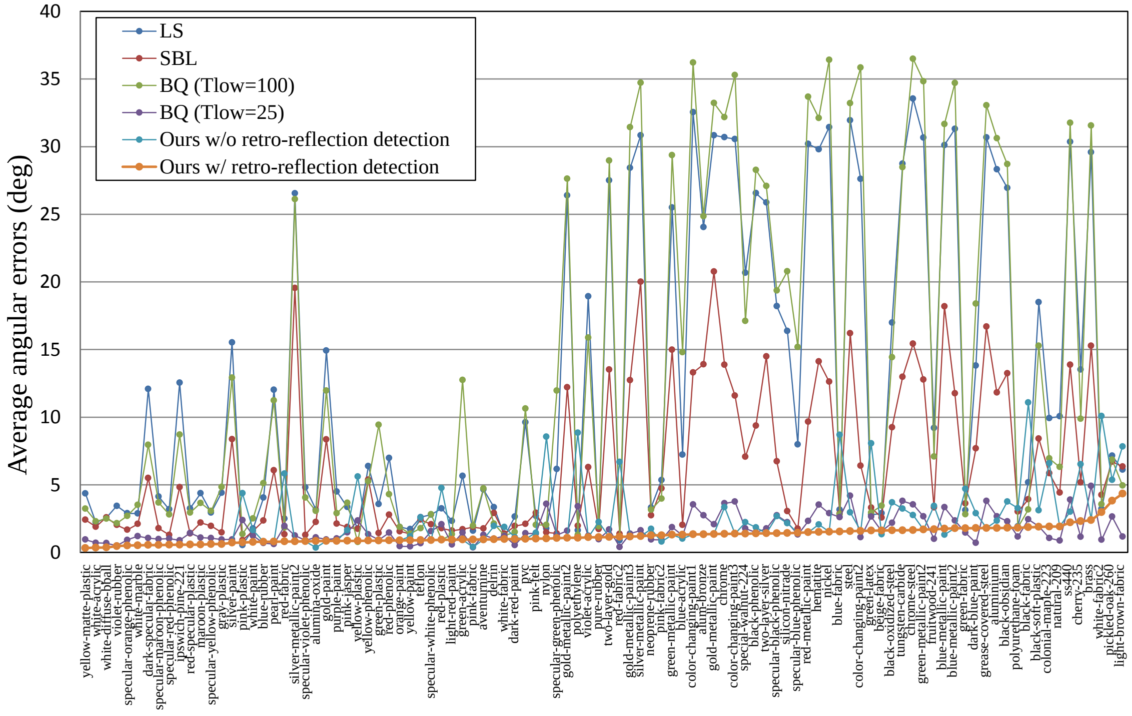


Figure 5-12: Comparison among different methods with dataset (B). We aligned results in ascending order of mean angular error of our method with the retro-reflection detection algorithm. The details of estimation results are presented in Appendix C.

is more effective for some materials, we emphasize that our method is capable of handling all frequencies in observations due to our flexible reflectance model while BQ ($T_{low} = 100$) does not work for most of materials. We also observe that angular errors of our method without retro-reflection detection are relatively large for materials which exhibit strong retro-reflections (*e.g.*, MERL fabrics), which indicates that our retro-reflection detection algorithm works pretty well for those materials. We note that average angular errors over 100 materials are 12.5 (LS), 6.2 (SBL), 13.1 (BQ, $T_{low} = 100$), 1.7 (BQ, $T_{low} = 25$), 2.4 (Ours w/o retro-reflection detection) and 1.2 (Ours w/ retro-reflection detection), respectively.

5.4.3 Qualitative Evaluation with Real Images

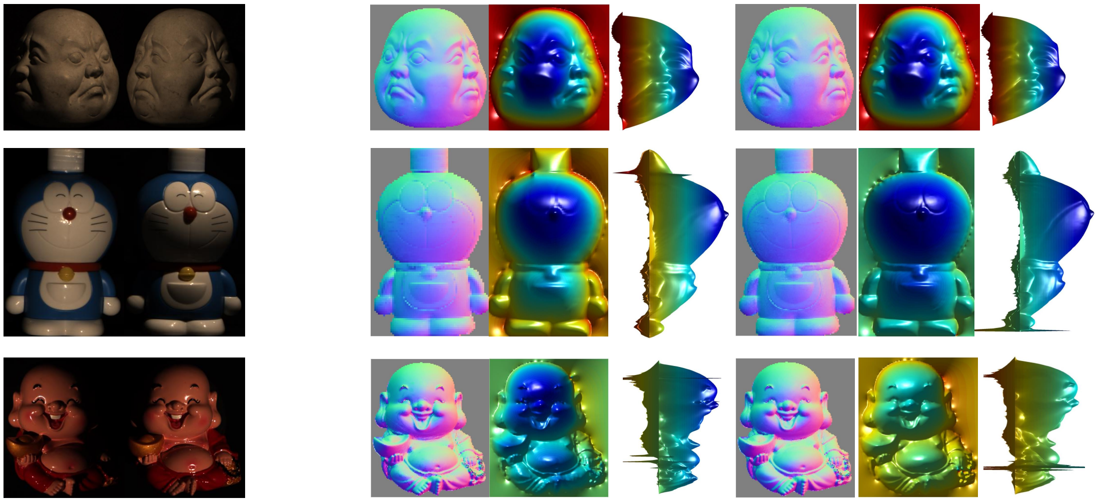
We also evaluate our algorithm using real images. We use three different datasets: a set of (a) 100 images of *two-face*, (b) 100 images of *doraemon*, and (c) 44 images of *fatguy*, (d) 112 images of *apple* and (e) 102 images of *gourd1* (these two datasets are from [9]). Note that shadows are removed by a simple thresholding. We evaluate the performance by visual inspection of the output normal maps and recovered surface meshes by a poisson solver [4], which recovers surface meshes from a gradient map. In addition, we also try the novel view synthesis using the shape and spatially varying reflectance functions recovered from *apple* and *gourd1* datasets, which have relatively large number of images. We use 111 of 112 images of *apple* and 101 of 102 images of *Gourd1* for the shape and reflectance estimation, and then render the image from the viewpoint of the remained image.

The results are illustrated in Fig. 5-13 and Fig. 5-14. By comparing our method with LS [122], we observe that our method succeeded to estimate smoother and more reasonable normal maps and surface meshes. We also observe that BQ ($T_{low} = 25$) worked poor for those datasets since shadows could not be completely removed by a simple thresholding, therefore the low-frequency component in the observation was still corrupted. In contrast to that, our method performed well since our method could account all observations without discarding the informative high-frequency component which allows us to estimate the surface normal of the scene robustly. As shown in Fig. 5-14, our algorithm reasonably synthesizes the novel view with small differences from real images even though some narrow specular highlights are not successfully produced because of the insufficiency of valid images as we have mentioned in Section 5.3.1.

5.5 Conclusion

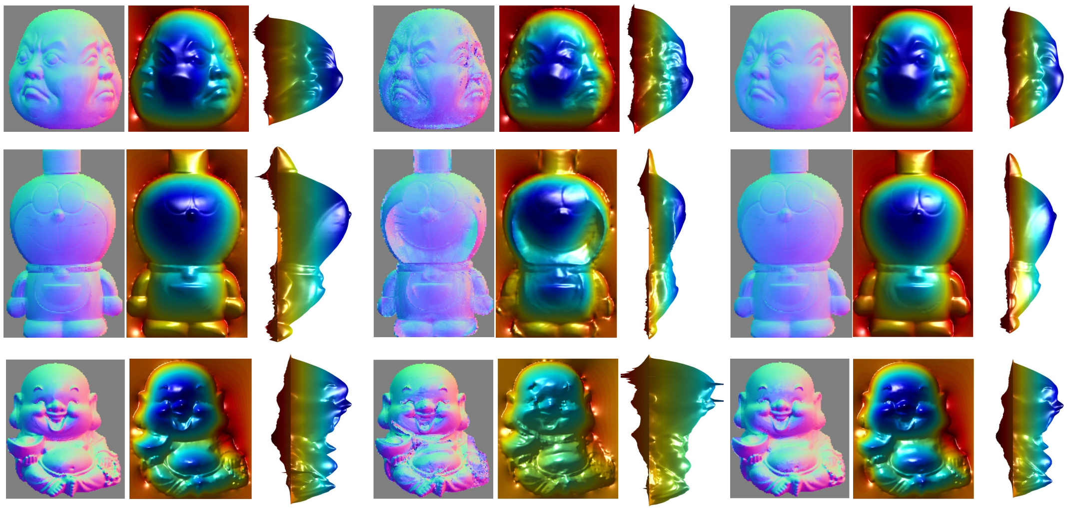
In this chapter, we have proposed the constrained bivariate regression based photometric stereo which works for various kind of isotropic surfaces by exploiting various conditions shared among physically valid BRDF. Approximating the inverse bivariate reflectance function by convenient Bernstein polynomials, we succeed to estimate the surface normal by solving standard quadratic programming problems. Our detailed experimental results have shown the state-of-the-art performance of our method for both synthetic and real data.

The current limitation is that we assume shadows or other corruptions (*e.g.*, inter-reflections) are discarded from images in advance, which may be impractical in real scenes. To ease this condition, we are very interested in incorporating data cleansing scheme with sparse regression in the similar manner with [61] to handle various kind of corruptions. Since our problem is described as linear equations, it seems that this extension is not impossible. Details about the limitation of our algorithm will be further discussed in the next chapter.



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

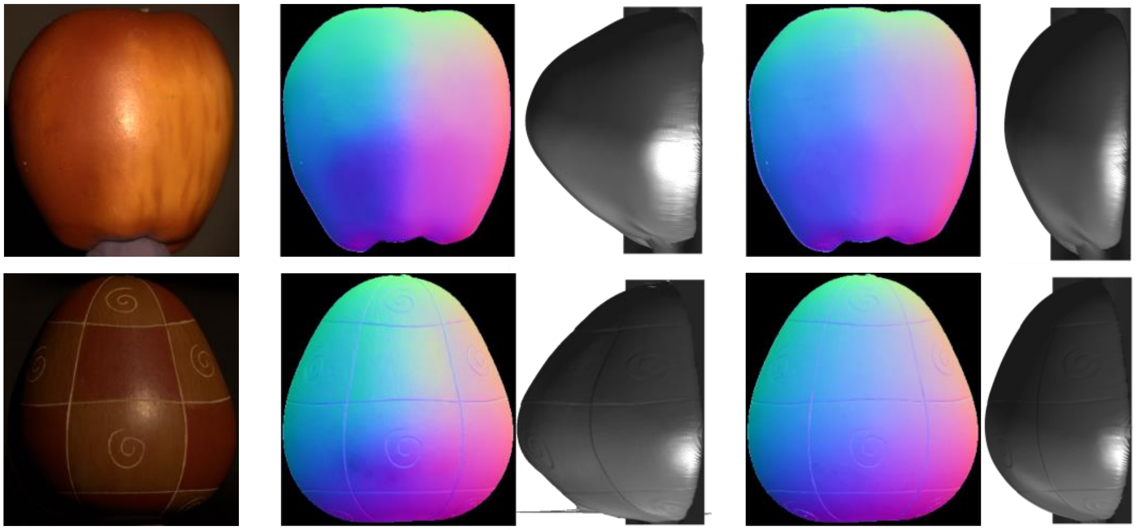
Input LS SBL



(f) (g) (h) (i) (j) (k)

BQ ($T_{low} = 100$) BQ ($T_{low} = 25$) Ours (w/ retro-reflection detection)

Figure 5-13: Experimental results using real data (*two-face, doraemon and fatguy*). We illustrate (a) input images, and normal maps recovered by (b) LS, (d) SBL, (f) BQ ($T_{low} = 100$), (h) BQ ($T_{low} = 25$) and (j) Ours (with a retro-reflective detection). We also show surface meshes generated from normal maps recovered by (c) LS, (e) SBL, (g) BQ ($T_{low} = 100$), (i) BQ ($T_{low} = 25$) and (k) Ours (with a retro-reflective detection).



(a) Input (b) LS (c) Ours

(e) Real image (f) Ours (g) Real image (h) Ours

Figure 5-14: Experimental results using real data (*apple, gourd1 and chocolate bear*). (a) We show example of input images of each dataset, and recovered surface normals and surface meshes by (b) LS and (c) our method. vel view synthesis with real datasets (C). (a,c) Images taken from the input dataset and (b,d) images rendered using shapes and spatially varying reflectance functions recovered by our method.

Chapter 6

Conclusion

In this dissertation, mainly two types of the constrained regression analysis are explored for estimating the surface normal from images captured under varying illumination under some challenging situations of the photometric stereo problem. First, we presented that a sparse regression analysis is well-suited for addressing various kind of model outliers that are simultaneously observed with inlier reflections. While shadow, sensor noises, specular highlight appear independently in the image, they are hard to represented by the low-order reflectance model such as Lambertian reflectance model [122] due to its physically complex behavior. In Chapter 3, we have proved that the sparsity penalty applied to the corruption in the image formation model of ”Lambertian diffuse reflection” plus ”sparse outliers”, reasonably neglect outliers with unknown variance while estimating surface normal. The sparse regression analysis is general, and possible to be performed on any photometric stereo problem with more complex image formation model than a simple Lambertian-based one.

However, unfortunately, there is a computational difficulty for applying the sparse regression on any non-linear modes that discontinuous and non-convex ℓ_0-norm of the regression error is hard to be minimized when the residual is represented by a non-linear function, that is a case where the inlier reflectance model is approximated by a non-linear parametric BRDF. To tackle this difficulty, we have took advantage of the inverse representation of a non-linear BRDF, that resolves the non-linearity of the problem assuming the monotonicity of the function on its independent variable. In Chapter 4, we have demonstrated the performance of our sparse regression based photometric stereo algorithm using various kind of datasets including both synthetic and real images, where some of materials do not have dominant, dense Lambertian rank-3 structure.

Second, we have proved that explicating the inverse representation of the image formation model not only useful for applying the sparse regression to the general diffusive reflections, but also can be used to separate the surface normal from other unknown parameters in BRDF whose number of lobes and preferred directions are unknown. In Chapter 5, we have formulated the photometric stereo problem for general, spatially varying isotropic surfaces as the constraint bivariate regression analysis where inverse reflectance function is approximated by a convenient Bernstein polynomials. We have proved that even though the number of lobes and their preferred directions are unknown, the surface normal can be recovered by solving a quadratic programming problem without any prior knowledge about the surface geometry like previous algorithm such as [9, 100]. In addition to that, we have also shown that complex retro-reflective reflections are also accounted by our model by utilizing the regression error for adaptively selecting the sufficient constraint for the regression analysis.

We believe that findings in our dissertation should not be limited on the photometric stereo problems that we have shown in this dissertation. For instance, the sparse regression analysis can also be performed on the general image formation model where BRDF is not limited on the diffusive function. Actually, one of our future works is applying the sparsity penalty to the derivation of the k-lobe BRDF presented in Chapter 5. Different from the regression analysis performed in Chapter 4, we have to consider how to combine the sparse regression with the *constrained* least-squares regression. However, we believe it is not impossible since all constraints are represented by a linear form thanks to the convenient Bernstein basis approximation.

We are also interested in performing the constrained regression on the *uncalibrated* photometric stereo problem, where lighting directions are unknown. Though our model is also affected by the bas-relief ambiguity, and require additional prior knowledge on the reflectance to resolve it, we believe that the inverse representation we have explored in this dissertation is still useful to solve the uncalibrated photometric stereo problem for *general* isotropic surfaces, that is apparently more difficult than most uncalibrated photometric stereo problems that were focused only on the Lambertian scene.

The accurate reflectance recovery would be one another important future direction of our dissertation. Actually, we have achieved spatially varying reflectance recovery in Chapter 5, however the reconstruction was imperfect since the high-frequency reflections were hard to be captured from the limited number of samples observed at a single surface point. To chive the accurate reconstruction, it may be inevitable to merge appearances over pixels. The problem is, however, gathering samples from isolated pixels lead to the global optimization problem, which may destroy the benefit of efficient per-pixel optimization framework of our method. Since our algorithm can accurately recover the surface normal from relatively smaller number of samples that are required for the SVBRDF estimation, it can be possible to alternatively solve *local* photometric stereo problem and *global* reflectance recovery problem for estimating both geometry and reflectance accurately.

A more specific direction for future work is to further analyze the optimal constraint selection for handling retro-reflective analysis that arises in Chapter 5. As shown in that chapter, the retro-reflective behavior is handled by reasonably changing the direction of the monotonicity constraint for $\boldsymbol{l}^T\boldsymbol{v}$. However, it requires to analyze the residuals of the regression analysis performed with two different constraints independently, leading to the increase of computational cost. Performing the regression analysis on both specular-reflective and retro-reflective materials without changing the constraints on the problem is a very important challenge.

Finally, the technique in Chapter 5 presents one possible way to address general isotropic materials, but there are still numerous materials that exhibit more complex reflections than the isotropic reflection (The full BRDF is actually four dimensional function, not three). Finding the solution to handle general *anisotropic* materials will increase potential applications, including recognition, industrial parts inspection, topological analysis, etc.

List of Publications

Related Papers

Journal Papers

- S. Ikehata, D. Wipf, Y. Matsushita and K. Aizawa, ”Photometric Stereo using Sparse Bayesian Regression for General Diffuse Surfaces”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. *(to appear)*

Reviewed Conference Papers

- S. Ikehata and K. Aizawa, ”Photometric Stereo using Constained Bivariate Regression for General Isotropic Surfaces”, *In Proc. of IEEE Computer Vision and Pattern Recognition*, 2014. *(to appear)*

- S. Ikehata, D. Wipf, Y. Matsushita and K. Aizawa, ”Robust Photometric Stereo using Sparse Regression”, *In Proc. of IEEE Computer Vision and Pattern Recognition*, 2012.

Non-reviewed Conference Papers

- S. Ikehata, D. Wipf, Y. Matsushita and K. Aizawa, ”Robust Photometric Stereo using Sparse Regression”, *In Proc. of Meeting on Image Recognition and Understanding(MIRU)*, 2012. (Japanese)

Other Publications

Reviewed Conference Papers

- S. Ikehata, J. Cho and K. Aizawa, ”DEPTH MAP INPAINTING AND SUPER-RESOLUTION BASED ON INTERNAL STATISTICS OF GEOMETRY AND APPEARANCE”, *In Proc. of IEEE International Conference on Image Processing (ICIP)*, 2013.
- J. Cho, S. Ikehata, H. Yoo, M. Gelautz and K. Aizawa, ”Depth Map Upsampling using Cost-Volume Filtering”, *In Proc. of IEEE IVMSP Workshop: 3D Image/Video Technologies and Applications*, 2013.
- S. Ikehata and K. Aizawa, ”Confidence-based Refinement of Corrupted Depth Maps”, *In Proc. of Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012.

Non Reviewed Conference Papers

- S. Ikehata, T. Yamasaki and K. Aizawa, Estimation and Fusion of Refined Depth maps based on Confidence Metrics, *InProc of Meeting on Image Recognition and Understanding (MIRU)*, 2011. (Japanese)
- S. Ikehata, T. Yamasaki and K. Aizawa, Dense Depth Map Estimation of Multiple-Wide-baseline Images with Confidnece Based Bundle Optimization, *IEICE general conference, 2011*. (Japanese)
- S. Ikehata, T. Yamasaki and K. Aizawa, ”The New Method for Calibrating Multi Hand-held Camera Using Images of the Scene for 3D Reconstruction of Non-rigid Objects”, *InProc of Meeting on Image Recognition and Understanding (MIRU)*, 2010. (Japanese)

Appendix A. Robust Photometric Stereo using Penalized Least Squares

Abstract

In this appendix, the nonparametric regression-based photometric stereo method is proposed, which does not explicitly assume the inverse reflectance model such as the piecewise linear model as shown in Chapter 4. More strictly speaking, I approximate the inverse diffusive reflectance function by the smooth spline function whose second derivatives are penalized by its smoothness.

A.1 Introduction

When we assume that the observation (I) is composed of dense diffusive component $f(\boldsymbol{n}^T\boldsymbol{l})$ and sparse non-diffusive corruptions (e), the appearance of a pixel is modeled as follow,

$$I_i = f(\boldsymbol{n}^T\boldsymbol{l}_i) + e_i. \tag{A.1}$$

Assuming that the diffusive reflectance function ($f(n^T\boldsymbol{l})$) is monotonic, the inverse representation of Eq. (A.1) is

$$\boldsymbol{n}^T\boldsymbol{l}_i = f^{-1}(I_i) + e'_i. \tag{A.2}$$

In Section 4, we approximated $f^{-1}(I_i)$ as a set of successive linear subfunctions (piecewise linear reflectance model) and our extensive discussions and experiments showed how reasonable they were. However, one defect is that it requires explicitly fixed number of sub-functions. So the important question is ”how can we find the way to recover surface normals without explicitly define the form of inverse reflectance model?”. For responding to this question, we state that an well-known nonparametric regression method called penalized least-squares regression may useful for photometric stereo problem due to its linearity and simplicity.

A.2 Penalized Least-squares for Photometric Stereo Problem

A.2.1 Penalized least-squares regression

Penalized least-squares regression [86] is a kind of nonparametric regression technique which finds a smooth function balancing between fitting the data closely and avoiding excessive roughness or rapid variation. Define x_i as a variable and y_i as the observation associated with x_i, we can fit the data using a nonparametric model as follows,

$$y_i = f(x_i) + e, \tag{A.3}$$

where f is an unknown function that is assumed to be reasonably smooth, e_i is an independent error. Of course, the functional space of $f(x)$ is so large that we can always find a function f that interpolates the data points. In order to obtain an estimate that fits the data well and has some degrees of smoothness, we can use the penalized least-squares method, which minimizes following cost function,

$$E = \frac{1}{n}\sum_{i=1}^{n} \|y_i - f(x_i)\|_2^2 + \int f''(x)dx. \tag{A.4}$$

This criterion trades off fidelity to the data (measured by the residual sum-of-squares) versus roughness of the function (measured by the penalty term). As is well known that a class of functions in the solution space is the cubic B- spline function (therefore it is also called as *smoothing spline regression*). Defining internal spline knots on each x_i and corresponding six external knots as $\xi_1 = \xi_2 = \xi_3 = x_1$ and $x_n = \xi_{n+4} = \xi_{n+5} = \xi_{n+6}$, the cost function Eq. (A.4) is now represented as follows,

$$E = (\boldsymbol{y} - G\beta)^T(\boldsymbol{y} - G\beta) + \lambda\beta^T K\beta, \tag{A.5}$$

where G is the cubic B-spline basis matrix (calculated by de Boor Cox equations), β is the coefficient matrix and K is the penalty matrix whose k_{ij} is defined as follow,

$$k_{ij} = \int \frac{d^2B_i(x)}{dx^2}\frac{d^2B_j(x)}{dx^2}dx. \tag{A.6}$$

Note that since second derivatives of a cubic spline function is linear, the k_{ij} is easily calculated from B-spline basis. From these representations, the hat matrix to minimize the cost function is now estimated as follow,

$$H = G(G^T G + \lambda K)^{-1}G^T, \tag{A.7}$$

where λ is fixed by a very small value (10^{-10} in my experiment). Now, it provides estimates of the function at $\boldsymbol{x}$ as

$$\hat{\boldsymbol{y}} = H\boldsymbol{y}. \tag{A.8}$$

The smoothing spline regression is convenient for the nonparametric photometric stereo problem in two reasons. First, G and K only depend on $\boldsymbol{x}$, not $\boldsymbol{y}$, which means that if we assume $\boldsymbol{x}$ as appearance ($\boldsymbol{I}$) and $\boldsymbol{y}$ as $L^T\boldsymbol{n}$ (L is a lighting matrix), the hat matrix is described only by $\boldsymbol{I}$. And the second reason is that as is indicated by Eq. (A.8), the estimation of the function is linear to the observations ($\boldsymbol{y}$), which enables us to easily extend nonparametric regression to *sparse* nonparametric regression. The details are described in the following section.

A.2.2 ℓ_2 / ℓ_0 minimization for photometric stereo

As is already mentioned, we can consider $L^T\boldsymbol{n}$ as $\boldsymbol{y}$ and $\boldsymbol{I}$ as $\boldsymbol{x}$ by comparing Eq. (A.3) with Eq. (A.2), which means that if we fix the surface normal $\boldsymbol{n}$, we can acquire a function $f(I)$ which smoothly interpolates a collection of 2-d plots $(I_i, \boldsymbol{n}^T\boldsymbol{l}_i)$. By using smoothing spline, the values of $f(I)$ for each lighting direction is estimated as follow,

$$\hat{y} = G(G^TG + \lambda K)^{-1}G^TL^T\boldsymbol{n}. \tag{A.9}$$

Note that the B-spline basis matrix B is calculated from a set of appearance $\boldsymbol{I}$ by using each I as knots.

In the photometric stereo problem, the goal is to recover the surface normal $\boldsymbol{n}$ so that as many as possible diffusive observations are lying on a unique, monotone function, which is expressed as

$$\min_{\boldsymbol{n}} \|L^T\boldsymbol{n} - G(G^TG + \lambda K)^{-1}G^TL^T\boldsymbol{n}\|_2^2. \tag{A.10}$$

This equation is transformed as,

$$\min_{\boldsymbol{n}} \|A\boldsymbol{n}\|_2^2, A \triangleq (E_n - G(G^TG + \lambda K)^{-1}G^T)L^T, \tag{A.11}$$

where $E_n \in \mathbb{R}^{n\times n}$ is an identity matrix. Assuming additive non-diffusive corruptions are included in appearance (*i.e.*$y_i = \hat{y}_i + e'_i$), the sparse regression problem for photometric stereo is now described as follow,

$$\min_{\boldsymbol{n},\boldsymbol{e}'} \|\boldsymbol{e}'\|_0. \quad \boldsymbol{e}' = A\boldsymbol{n}. \text{ s.t. } \boldsymbol{n}' \neq 0 \tag{A.12}$$

This optimization problem is almost same with a piecewise linear case, so it could be solved by any ℓ_0 optimization methods such as SBL.

A.2.3 Further constraints

Eq. (A.12) recovers the surface normal so that as many as possible observations are lying on the smooth spline function. In contrast to the piecewise linear sparse regression which needs the number of sub-functions, Eq. (A.12) can handle all appearances simultaneously without dividing observations. However, unfortunately, smoothing spline regression has two unpleased properties. The first is that it does not always pass the origin $(0, 0)$ where most reflectance functions should be satisfied. And second is that, even though the sparse regression is applied to find the surface normal, the smoothing spline function itself (Eq. (A.8)) is defined by ℓ_2 minimization. More intuitively speaking, when we have the true surface normal and a set of ground truth plots $(I_i, n^T\boldsymbol{l}_i)$, we assume that the diffusive reflectance function is the function where as much as possible ground truth plots are lying on the function. However, it is not guaranteed that the function is lying on observations since the fitting of smoothing function is based on ℓ_2 distance minimization penalized by its smoothness. To overcome those problems, two further modifications are proposed; the first is adding constraint that the function must pass the origin and the second is adding another sparse penalty to neglect observations in finding smooth function.

To enforce the recovered function to pass the origin, a knot on $(0, 0)$ is added to the data and fix its distance from the function by zero in the SBL-based optimization process. Strictly speaking, we modify the optimization as follow,

$$\begin{gathered}\min_{\boldsymbol{x},\tilde{\boldsymbol{e}'}} \|\tilde{\boldsymbol{e}'}\|_0. \quad \tilde{\boldsymbol{e}'} = \tilde{A}\boldsymbol{x}. \text{ s.t. } \boldsymbol{n}' \neq 0 \text{ and } \boldsymbol{e}'_0 = 0,\\ \tilde{A} = (E_{n+1} - G(G^TG + \lambda K)^{-1}G^T)\begin{bmatrix} 0 & 0 \\ 0 & L^T \end{bmatrix},\end{gathered} \tag{A.13}$$

where $\boldsymbol{x} = [0 \;\; \boldsymbol{n}^T]^T$ and $\tilde{\boldsymbol{e}}$ is a vector which has an additional element to the top of $\boldsymbol{e}'$.

In the second modification to find the smooth function by the ℓ_0 regression not by the ℓ_2 regression, I introduce the additional sparse error vector as follow,

$$\hat{\boldsymbol{y}} = H(\boldsymbol{y} - \boldsymbol{e}''). \tag{A.14}$$

Note that $\boldsymbol{y} - \boldsymbol{e}''$ with non-zero elements in $\boldsymbol{e}''$ are used to recover the spline function instead of $\boldsymbol{y}$ meaning that y_i with non-zero e_i is neglected in deciding the curve of smooth function, which enables us to neglect outliers. Considering this modification, the problem is updated as

$$\begin{aligned} &\min_{\boldsymbol{n},\boldsymbol{e}',\boldsymbol{e}''} \|\boldsymbol{e}'\|_0 + \|W\boldsymbol{e}''\|_0 \\ &s.t.\ \ \boldsymbol{e'} = L^T\boldsymbol{n} - \boldsymbol{e}'' - G(G^TG + \lambda K)^{-1}G^T(L^T\boldsymbol{n} - \boldsymbol{e}'') \\ &s.t.\ \ \boldsymbol{n} \neq 0. \end{aligned} \tag{A.15}$$

W is the weighing matrix, which is set by the identity matrix in the experiment. When we combine Eq. (A.13) and Eq. (A.15), the final nonparametric sparse regression-based photometric stereo problem is formulated as follow,

$$\begin{aligned} &\min_{\boldsymbol{n},\tilde{\boldsymbol{e}}',\boldsymbol{e}''} \|\tilde{\boldsymbol{e}}'\|_0 + \|W\boldsymbol{e}''\|_0 \\ &s.t.\ \ \tilde{\boldsymbol{e}}' = \tilde{A}\boldsymbol{x} - B\boldsymbol{e}'',\ \ B = E_{n+1} - G(G^TG + \lambda K)^{-1}G^T \\ &s.t.\ \ \boldsymbol{n} \neq 0. \end{aligned} \tag{A.16}$$

Note that a sum-constraint on this problem is used to avoid $\boldsymbol{n} = 0$ as has been mentioned in Section 4.3.1.

A.3 Experimental Results

In this section, the nonparametric regression-based photometric stereo methods are evaluated by the synthesized *bunny* dataset (256x256 pixels, one hundred 32-bit HDR images rendered with MERL BRDF functions). In this experiment, we focus on the per-pixel error analysis rather than the per-image mean angular errors evaluation to see how each regression works for each material. We compare the nonparametric regression-based photometric stereo methods from Eq. (A.11) (NL-LS), Eq. (A.12) (NL-SBL) and Eq. (A.16) (NL-SBL (modified)) with our piecewise linear sparse regression-based method (PL-SBL). In this experiment, we categorized 100 BRDFs into four classes and extract representative materials of all *i.e.* (a) dense diffusive reflections *e.g. yellow-paint and*

Table 6.1: Experimental results of *Bunny* with eight different BRDFs.

		Angular error (in degrees)			
	Material name	PL-SBL	NL-LS	NL-SBL	NL-SBL (modified)
1	yellow-paint	0.80	**0.21**	0.29	0.24
2	white-diffuse-paint	0.66	0.35	0.43	**0.32**
3	green-acrylic	**2.0**	3.6	4.2	4.0
4	pink-fabric	2.3	1.3	1.6	**0.85**
5	green-fabric	2.0	**0.90**	9.4	8.5
6	polyethylene	6.6	4.81	5.3	**4.0**
7	black-phenolic	1.4	**0.71**	23.5	7.9
8	gold-metallic-paint	**1.9**	23.1	26.5	27.7
Ave.		**2.2**	4.4	8.9	6.7

white-diffuse paint (b) dense diffusive reflections + sparse specularities *e.g. green-acrylic* (c) 2-lobe materials *e.g. pink-fabric, green-fabric and polyethylene* (d) metallic materials (sparse diffusive + dense specular) *e.g. black-phenolic and gold-metallic-paint.* In the experiment, observations with non-zero values are only used in estimation. In PL-SBL, the number of sub-functions are fixed by three and the deviation of line slants (σ_a) is fixed by 1.0.

The results are illustrated in Table 6.1 and Fig. 6-1. In Table 6.1, the angular errors for each method and material are displayed and Fig. 6-1 shows recovered smooth functions and 2-d plots $(I_i, \boldsymbol{n}^T\boldsymbol{l}_i)$ with ground truth green plots. Note that in this experiment, the observations are sampled from the same pixel for each method and each material (i.e. $[x,\ y] = [61, 80]$). We observe that non-parametric regression-based photometric stereo method (NL-LS, NL-SBL (modified)) works better than PL-SBL for most non-metallic objects. It seems difficult to conclude whether ℓ_2 or ℓ_1 regression is better. In contrast that, we can clearly see that nonparametric methods have difficulty in handling metallic objects while PL-SBL works much better than them. In metallic objects, there is little (sparse) diffusive reflections and most appearances come from specularities. Unfortunately, NL-LS or NL-SBL are over-fitted to those corruptions due to its high degree of freedom. In contrast that we believe the piecewise linear function can rather capture the sparse linear structure of appearance in metallic scene due to its strong assumption of the local linearity. We also observe that PL-SBL has larger robustness to corruptions such as additive noises.

A.4 Conclusion

In this appendix, we presented nonparametric regression-based photometric methods which try to represent the inverse diffusive reflectance function by nonparametric smooth B-spline function. The new method is purely data-driven and solved by closed-form equations which can capture complex diffusive structure of materials without any nonlinear optimization. Experimental results show that these methods can recover more accurately than piece-wise linear function for materials with dense diffusive structure. The current limitation of nonparametric regression-based photometric stereo is the over-fitting to corruptions. For example, metallic objects are incapable by nonparametric method due to its high degree of freedom while piecewise linear function can capture the sparse linear diffusive structure appeared in low-frequency domain of metallic objects. From current experimental results, we can say that the piecewise linear function is better than smooth B-spline function for photometric stereo problem in average. It may be because not only the small number of parameters of the piecewise linear function are favorable for avoiding over-fitting but also the partially linear structure of the function can reasonably represent the reflectance of natural materials.

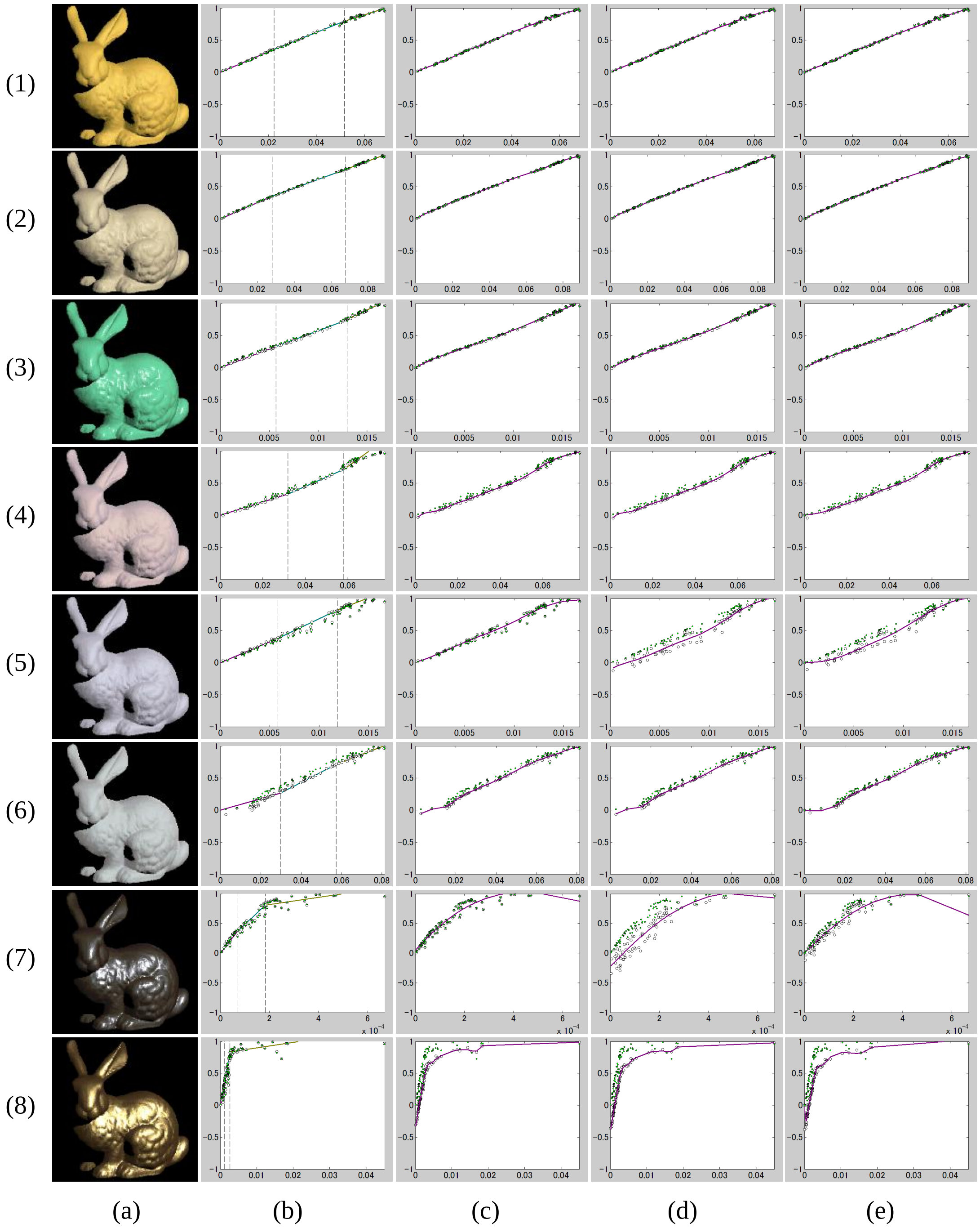


Figure 6-1: (a) Input images (only one pixel is sampled in the experiments), (b)-(e) Recovered 2-d plots $(I_i, \boldsymbol{n}^T\boldsymbol{l}_i)$ (blank circles), function (colored lines) of PL-SBL, NL-LS, NL-SBL and NL-SBL (modified version) with ground truth 2-d plots (green circles).

Appendix B: Details about common parametric BRDFs

This appendix expands on the detailed description about the isotropic BRDF which were used for rendering synthetic images in this dissertation. For further information of each BRDF, we recommend readers to refer a recent survey by Montes and Urena [78]. We show examples of rendered images with those BRDF in Fig. 6-2.

B.1 Cook-Torrance Model [29]

Cook-Torrance model represents the reflection using a combination of diffuse and specular parts as

$$\rho(\boldsymbol{n},\boldsymbol{l},\boldsymbol{v}) = \frac{k_d}{\pi} + k_s f_S(\boldsymbol{n},\boldsymbol{l},\boldsymbol{v},\lambda_1,\mu_1). \tag{B.1}$$

Here k_d and k_s are model parameters representing the strength of diffuse and specular terms respectively. f_S is a nonlinear function which is represented as

$$f_S(\boldsymbol{n},\boldsymbol{l},\boldsymbol{v},\lambda_1,\mu_1) = \frac{F(\boldsymbol{l},\boldsymbol{v},\lambda_1)}{\pi}\frac{D(\boldsymbol{n},\boldsymbol{l},\boldsymbol{v},\mu_1)G(\boldsymbol{n},\boldsymbol{l},\boldsymbol{v})}{(\boldsymbol{n}^T\boldsymbol{v})(\boldsymbol{n}^T\boldsymbol{l})}, \tag{B.2}$$

where F is the Fresnel factor, D is the microfacets distribution which is computed by the Beckmann distribution function and G is the geometric attenuation factor, respectively. We use the same D and G in [29] and use the Schlick approximation of the Fresnel term (F) [93] as follow:

$$F(\boldsymbol{l},\boldsymbol{v},\lambda_1) = \lambda_1 + (1-\lambda_1)(1-\boldsymbol{l}^T\boldsymbol{h})^5, \tag{B.3}$$

where λ is the reflection coefficient for light incoming parallel to the normal, and $\boldsymbol{h}$ is a half vector as $\boldsymbol{h} = (\boldsymbol{l} + \boldsymbol{v})/|\boldsymbol{l} + \boldsymbol{v}|$.

In the Cook-Torrance model, the diffuse reflection is represented by a linear Lambertian reflection [122] and the specular reflection is modeled by a specular lobe which is pointing at the normalized half vector (smaller roughness parameter μ_1 provides narrower specular highlight) and Fresnel effect, increasing the specular reflectivity as the surface turns away from the viewing direction. In this dissertation, we used $k_d = 0.9$, $k_s = 0.1$, $\lambda = 0.2$ and $\mu_1 = 0.2$, respectively.

B.2 Ward Model [118]

Ward model also represents the reflection in the same form of Eq. (B.1) in Cook-Torrance model, however this model only extracts the microfacets distribution D in Eq. (B.2) for representing the specular reflection as follow

$$f_S(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}, \mu_2) = \frac{1}{4\pi\mu_2^2\sqrt{(\boldsymbol{n}^T\boldsymbol{l})(\boldsymbol{n}^T\boldsymbol{v})}}\exp\left(\frac{1}{\mu_2^2}\left(1 - \frac{1}{\boldsymbol{n}^T\boldsymbol{h}^2}\right)\right), \tag{B.4}$$

where μ_2 is the roughness parameter which determines the size of specular highlight (smaller roughness parameter μ_2 provides narrower specular highlight). In the Ward model, the diffuse reflection is also represented by a linear Lambertian reflection [122]. In this dissertation, we used $\mu_2 = 0.2$. We note that specular highlights rendered with Cook-Torrance and Ward BRDF are mainly distributed in the high-frequency observations (*i.e.*, observations with large intensities), therefore, effects of those specularities are easily neglected by discarding input observations whose intensities are relatively large (*i.e.*, $T_{low} < 50\%$ in Section 5.4.

B.3 Lafortune Model [65]

Lafortune model is one of the most multifunctional BRDF models which is able to represent the data from real materials which have more than one lobe. In our experiment, we use this model to represent the general diffuse reflection which is more complex than a simple Lambertian reflection used in the Cook-Torrance and Ward BRDF.

As shown by Lin and Lee [69], a general rotationally symmetric diffuse component in the Lafortune model is represented as

$$\rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = (\boldsymbol{n}^T\boldsymbol{l})^k(\boldsymbol{n}^T\boldsymbol{v})^k, \tag{B.5}$$

where k is a model parameter which determines the non-linearity of the function. In this dissertation, we rendered images with only diffuse component without off-specular component in the Lafortune model to verify performance of each algorithm to handle complex non-linear diffuse reflections. In this dissertation, we used $k = 0.5$.

B.4 Oren-Nayar Model [83]

Oren-Nayar model is derived from Lambertian model [122] to explain the view dependency of the matte or rough surfaces with geometric optics. While this model and Oren-Nayar model both represent non-Lambertian diffuse reflections, this model can represent retro-reflective behavior of non-Lambertian diffusive objects unlike Lafortune BRDF.

The Oren-Nayar model is represented as follow

$$\rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = \frac{1}{\pi} \left(A + B \max\left(0, \cos(\phi_i - \phi_o)\right) \sin\alpha \tan\beta \right), \tag{B.6}$$

where ϕ_i and ϕ_o are differential angles as for $\boldsymbol{l}$ and $\boldsymbol{v}$ respectively, and a, b, A and B are written as

$$a = \max(\cos^{-1}(\boldsymbol{n}^T\boldsymbol{l}), \cos^{-1}(\boldsymbol{n}^T\boldsymbol{v})) \quad b = \min(\cos^{-1}(\boldsymbol{n}^T\boldsymbol{l}), \cos^{-1}(\boldsymbol{n}^T\boldsymbol{v})). \tag{B.7}$$

and

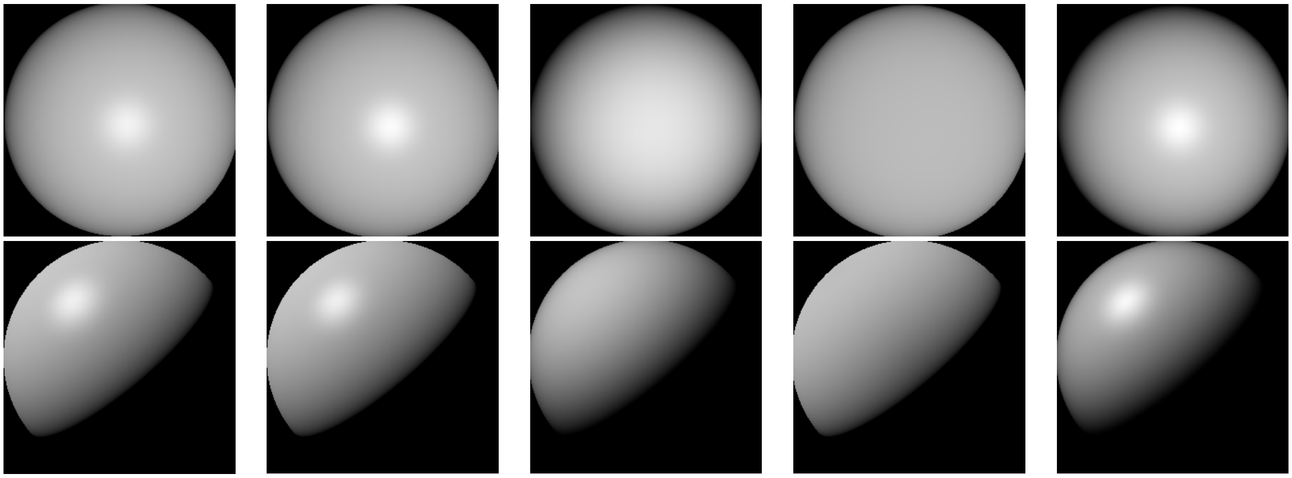
$$A = 1 - 0.5\frac{\alpha^2}{\alpha^2 + 0.33} \quad B = 0.45\frac{\alpha^2}{\alpha^2 + 0.09}. \tag{B.8}$$

Here, α determines the surface roughness and it is equivalent to the Lambertian model in the case of $\alpha = 0$. In this dissertation, we fixed the roughness parameter α by 0.5.

B.5 Ashikhmin-Shirley Model [13]

Ashikhmin-Shirely model is expressed as a weighted sum of the diffuse and specular term as

$$\rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = \frac{k_d}{\pi} f_D(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) + k_s f_S(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}, \lambda_2, \mu_3). \tag{B.9}$$



Cook-Torrance Ward Lafortune Oren-Nayar Ashikhmin-Shirely

Figure 6-2: Examples of images rendered with five different BRDFs.

The non-linear diffuse term f_D guarantees the reciprocity and energy conservation properties and is written as

$$f_D(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = \frac{28}{23}\left(1 - \left(1 - \frac{\cos(\boldsymbol{n}^T\boldsymbol{l})}{2}\right)^5\right)\left(1 - \left(1 - \frac{\cos(\boldsymbol{n}^T\boldsymbol{v})}{2}\right)^5\right). \tag{B.10}$$

The specular term uses the distribution over the half vector in the similar manner with Cook-Torrance model [29] as

$$f_S(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}, \lambda_2, \mu_3) = \frac{F(\boldsymbol{l}, \boldsymbol{v}, \lambda_2)}{\pi}\frac{D(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}, \mu_3)}{8\pi(\boldsymbol{h}^T\boldsymbol{l})\max(\boldsymbol{n}^T\boldsymbol{l}, \boldsymbol{n}^T\boldsymbol{v})}, \tag{B.11}$$

where F is the Schlick's approximation of Fresnel term and D is a microfacet distribution function described as

$$F(\boldsymbol{l}, \boldsymbol{v}, \lambda_2) = (\mu_3 + 1)(\boldsymbol{n}^T\boldsymbol{h})^{\mu 3}, \tag{B.12}$$

where μ_3 controls the roughness of the material which was fixed by 50.

Appendix C: Details of Results Using MERL BRDF Database

This appendix provides supplementary about our experiments using MERL BRDF database [74]. First, Fig. 6-3 illustrate synthesized images rendered with MERL BRDF database [74] which are used in Section 4.4.2. We also show details of Fig. 5-12 in Section 5.4 about input images, recovered surface normal maps and corresponding error maps for 100 materials in alphabetical order from Fig. 6-4 to Fig. 6-103. 3-d plots of $(\boldsymbol{l}^T\boldsymbol{v}, \boldsymbol{n}^T\boldsymbol{l}, I)$ for three different surface normals which were projected onto the $\boldsymbol{l}^T\boldsymbol{v} - \boldsymbol{n}^T\boldsymbol{l}$ space are also illustrated, whose plots were colored by its intensity. We have provided those information so that readers can examine how our method or other methods performed on each material in Fig. 5-12.

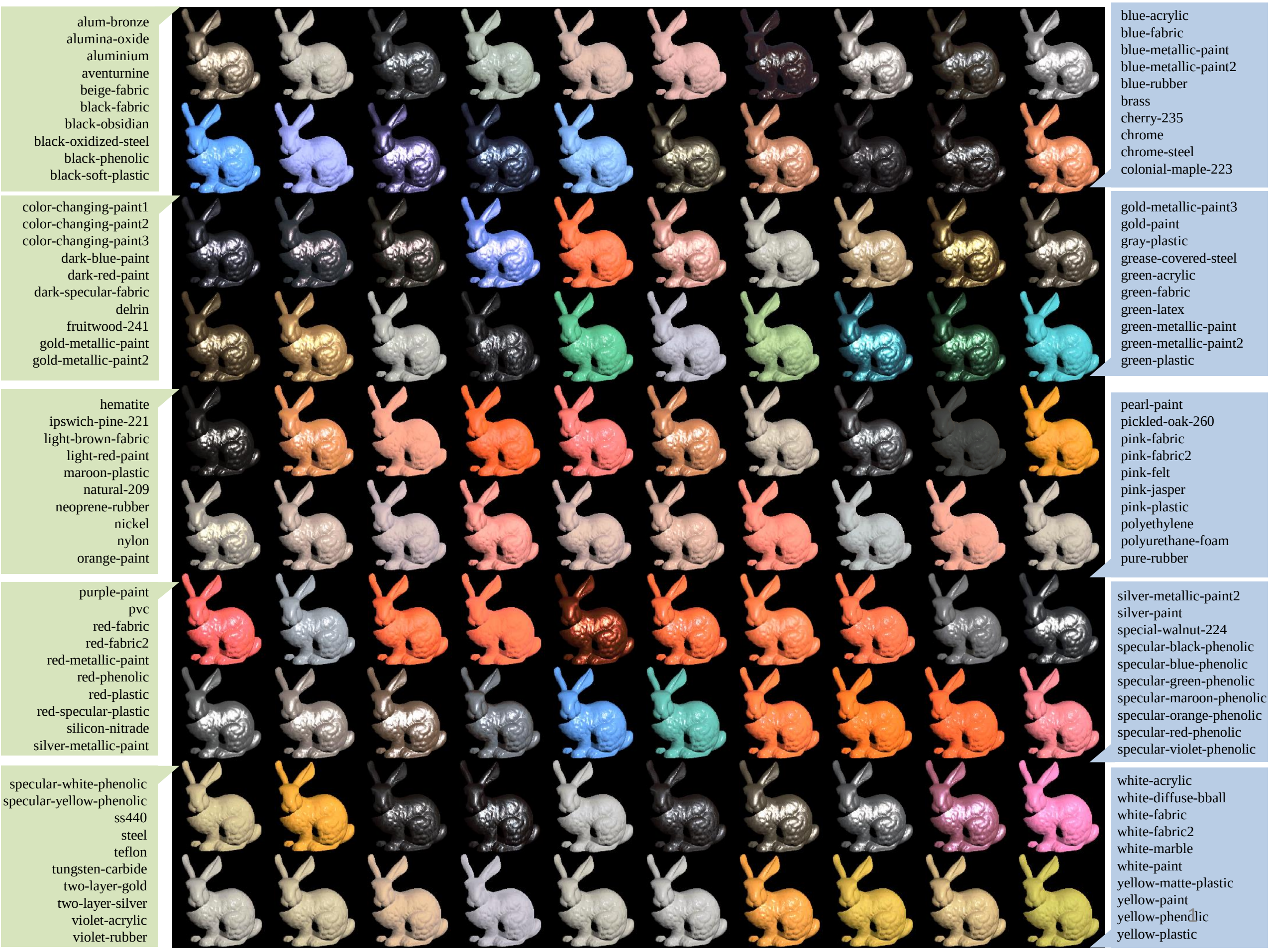


Figure 6-3: *Bunny* images rendered with MERL BRDF databased [74].

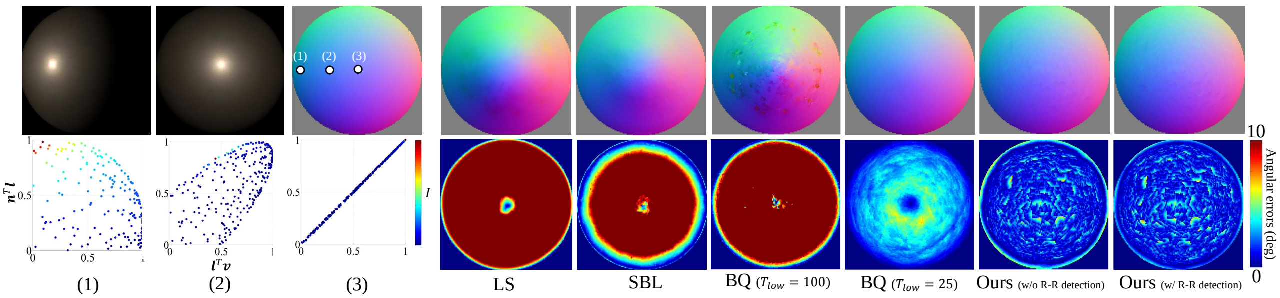


Figure 6-4: The results of alum-bronze.

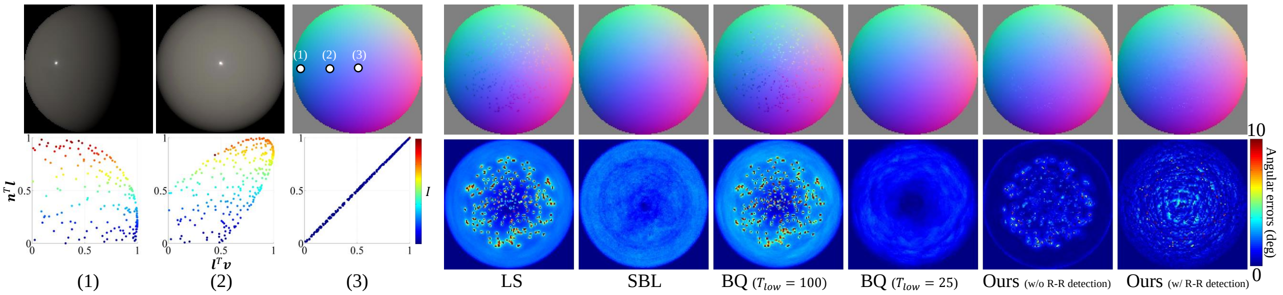


Figure 6-5: The results of alumina-oxide.

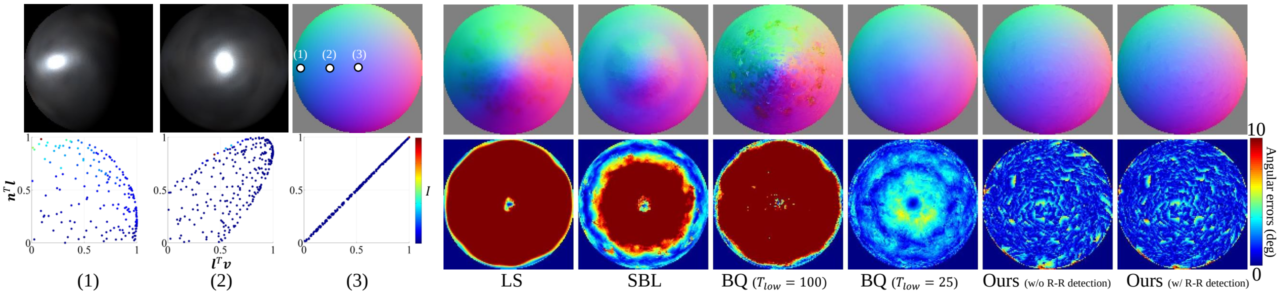


Figure 6-6: The results of aluminium.

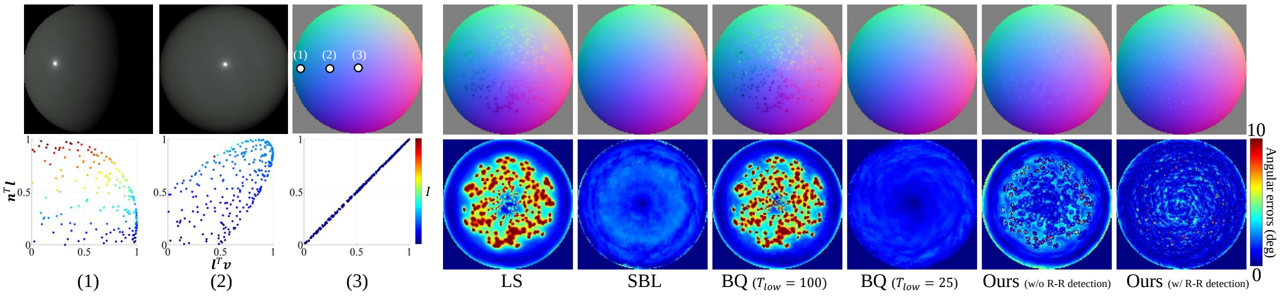


Figure 6-7: The results of aventurnine.

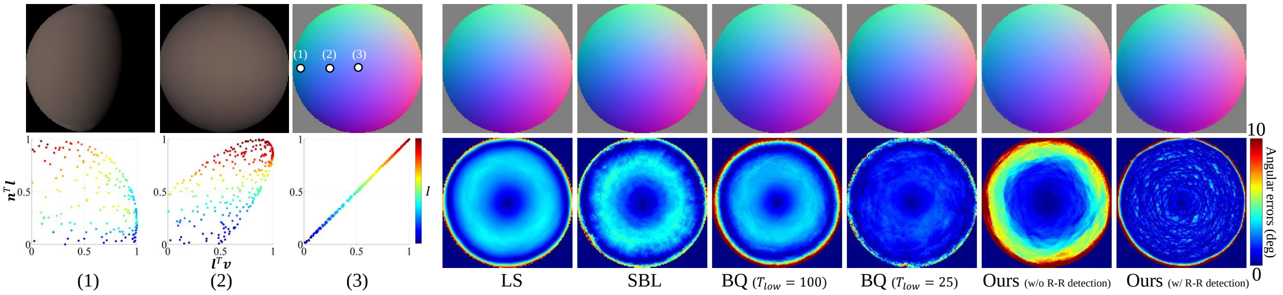


Figure 6-8: The results of beige-fabric.

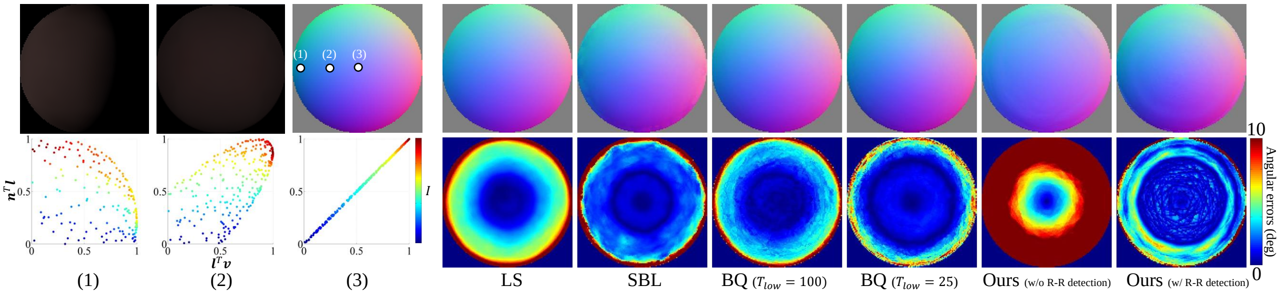


Figure 6-9: The results of black-fabric.

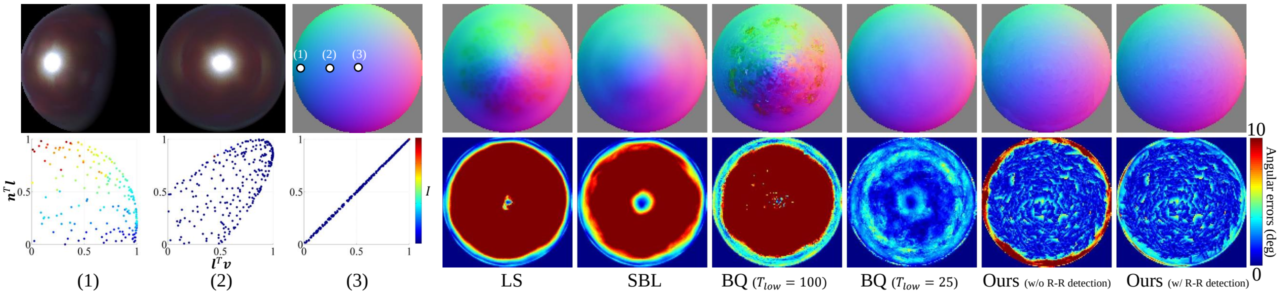


Figure 6-10: The results of black-obsidian.

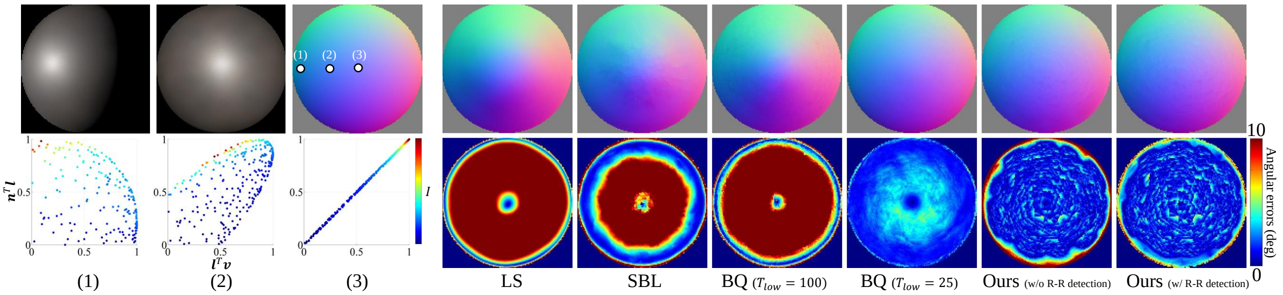


Figure 6-11: The results of black-oxidized-steel.

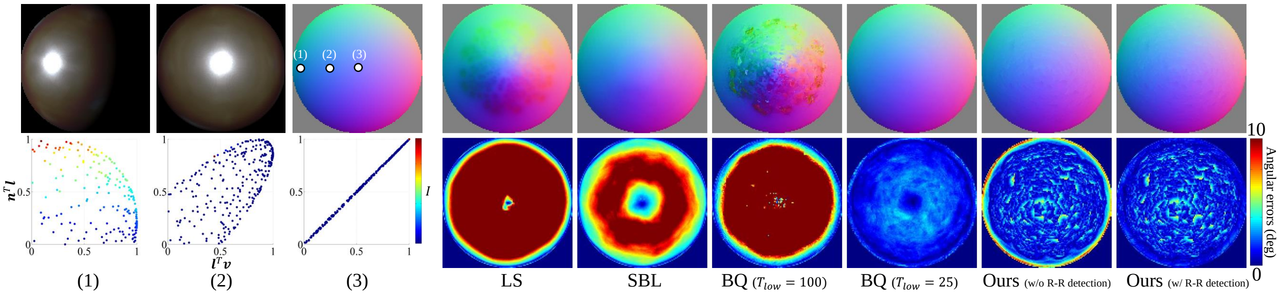


Figure 6-12: The results of black-phenolic.

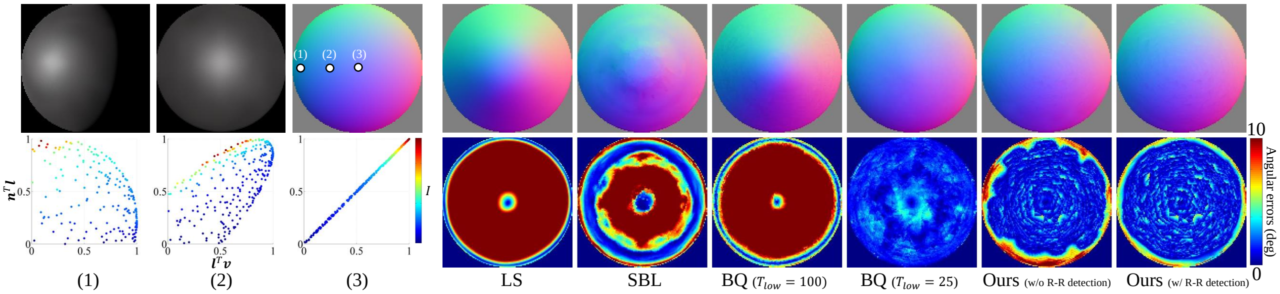


Figure 6-13: The results of black-soft-plastic.

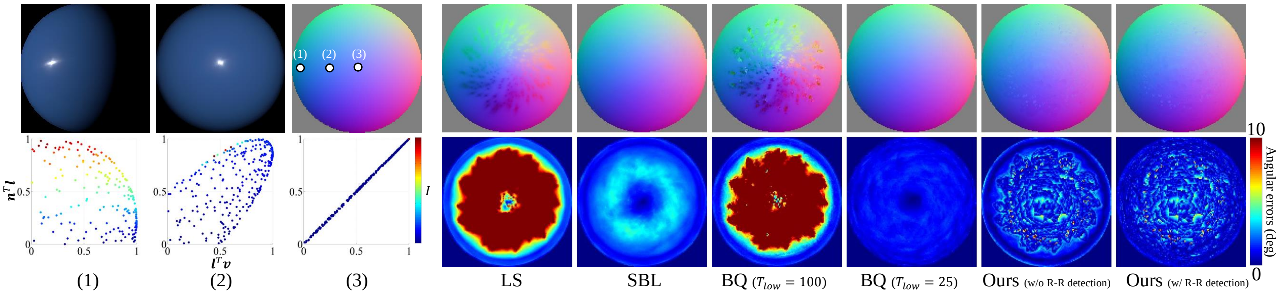


Figure 6-14: The results of blue-acrylic.

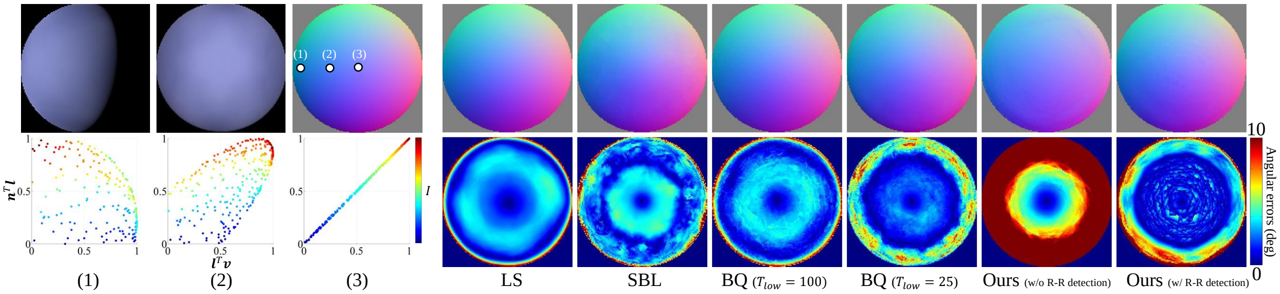


Figure 6-15: The results of blue-fabric.

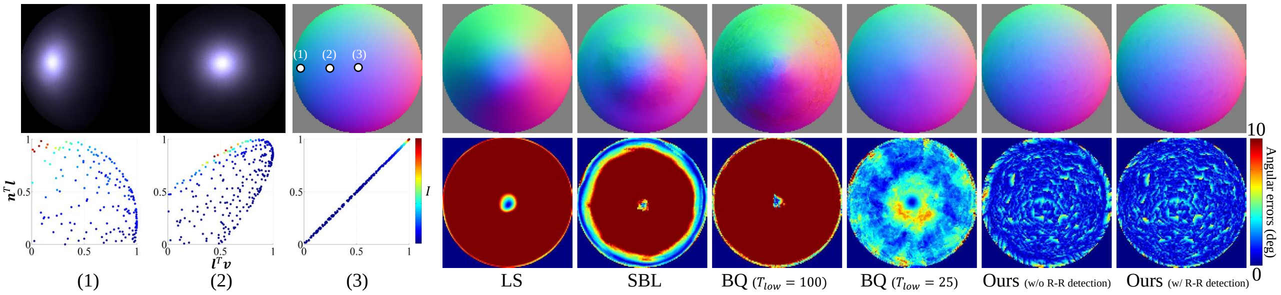


Figure 6-16: The results of blue-metallic-paint.

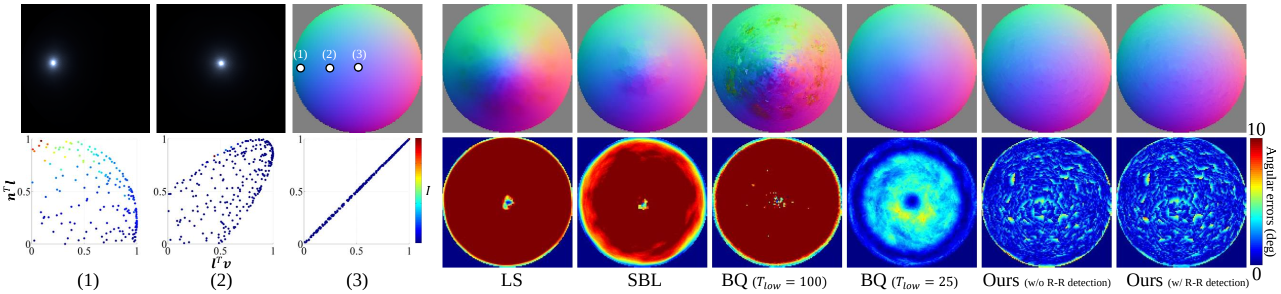


Figure 6-17: The results of blue-metallic-paint2.

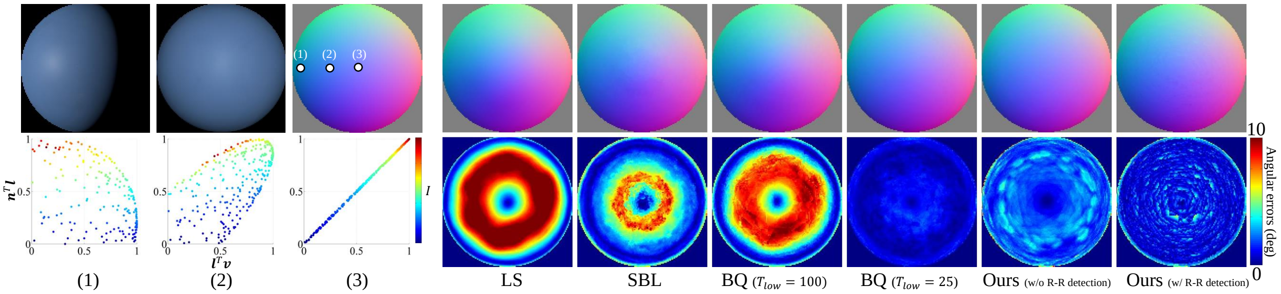


Figure 6-18: The results of blue-rubber.

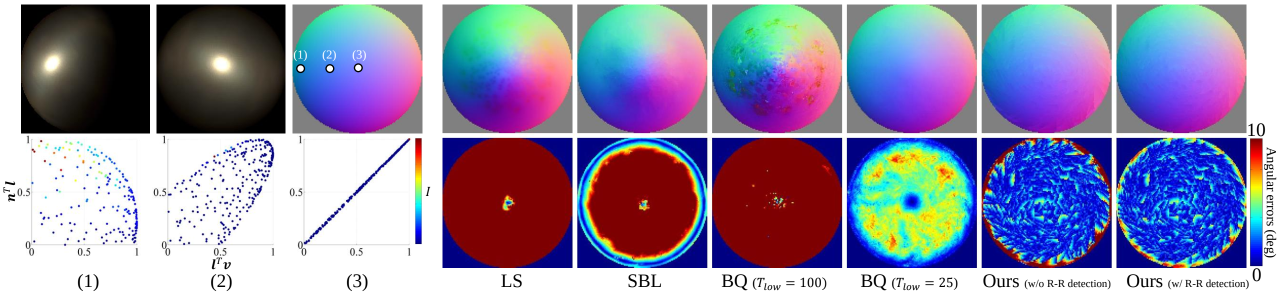


Figure 6-19: The results of brass.

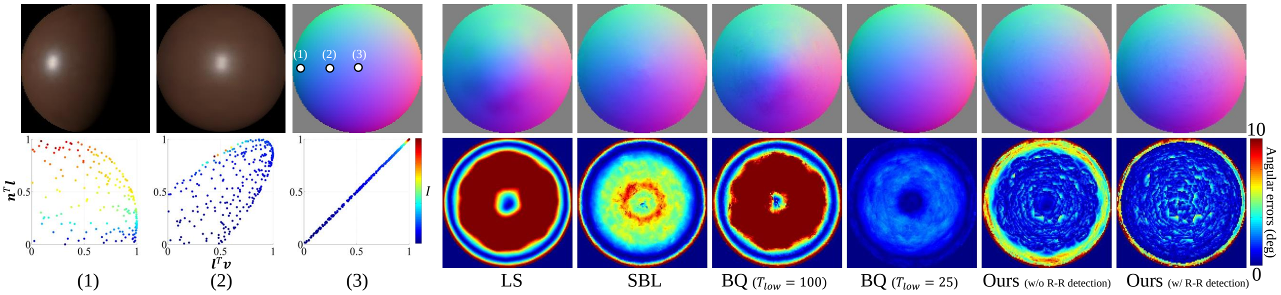


Figure 6-20: The results of cherry-235.

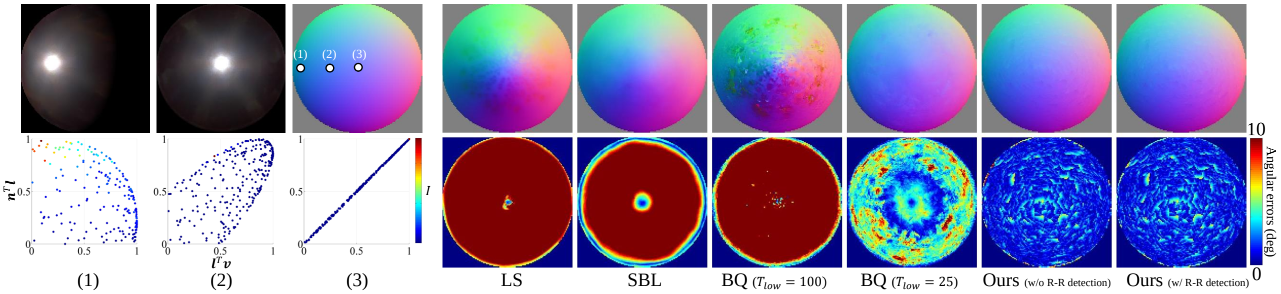


Figure 6-21: The results of chrome.

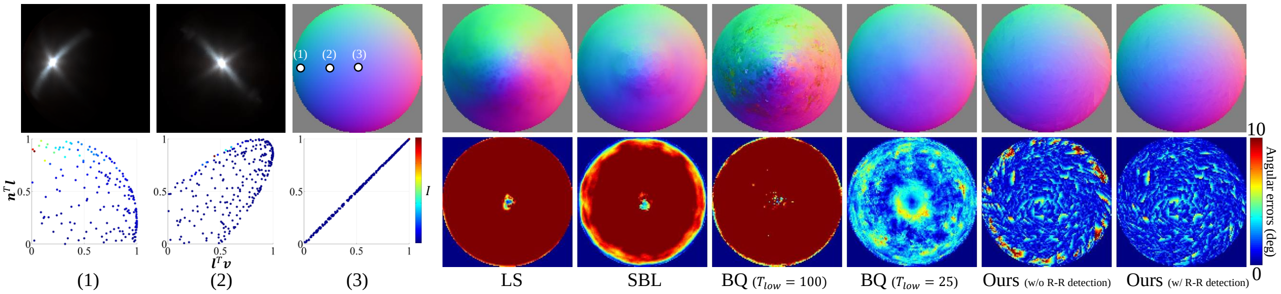


Figure 6-22: The results of chrome-steel.

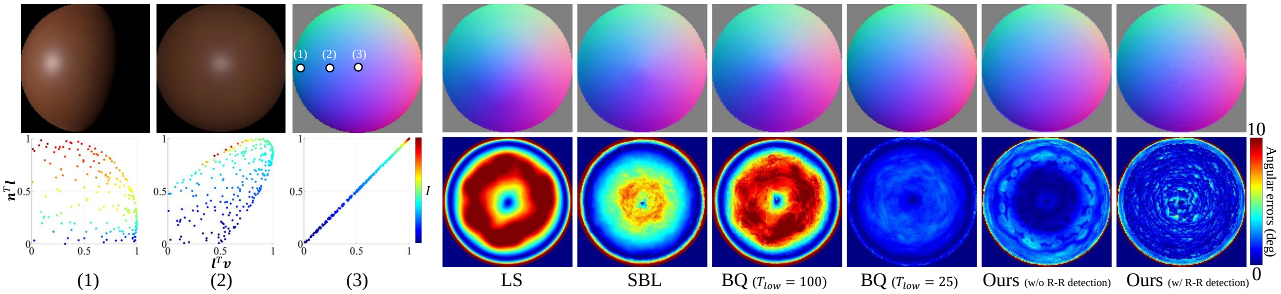
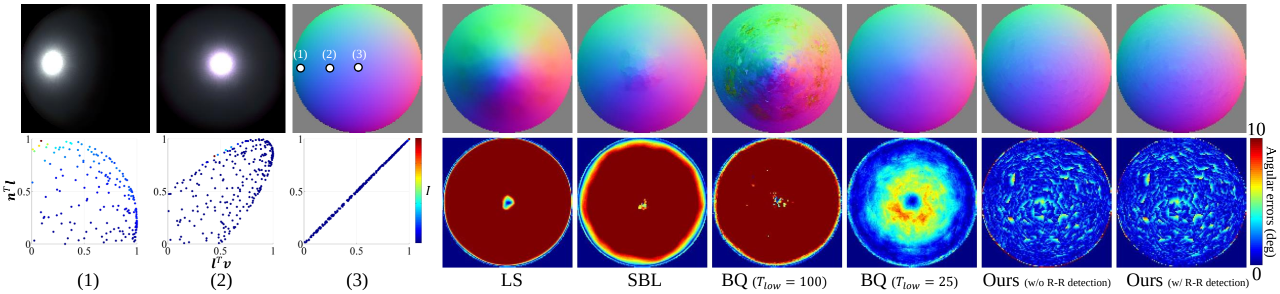
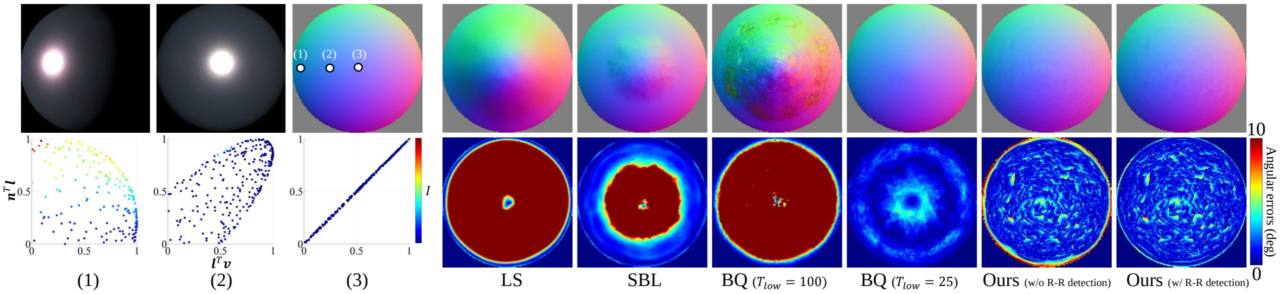


Figure 6-23: The results of colonial-maple-223.



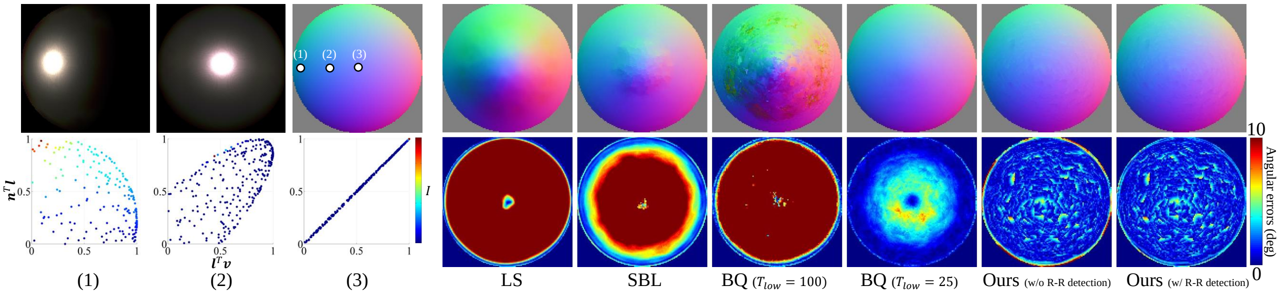
(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-24: The results of color-changing-paint1.



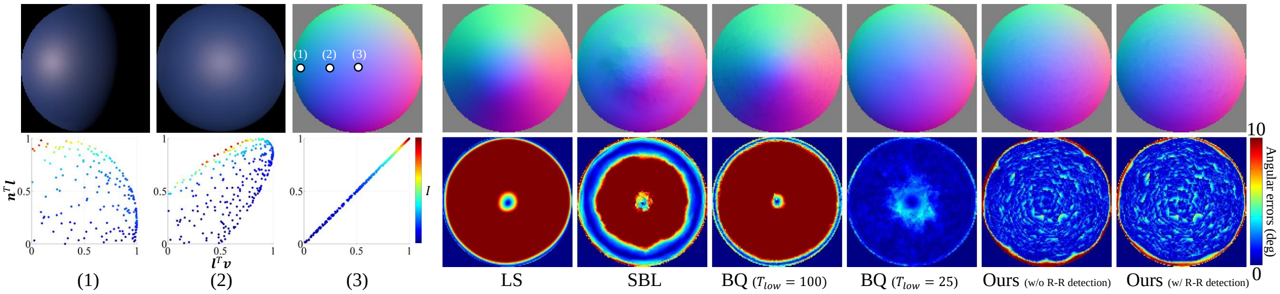
(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-25: The results of color-changing-paint2.



(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-26: The results of color-changing-paint3.



(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-27: The results of dark-blue-paint.

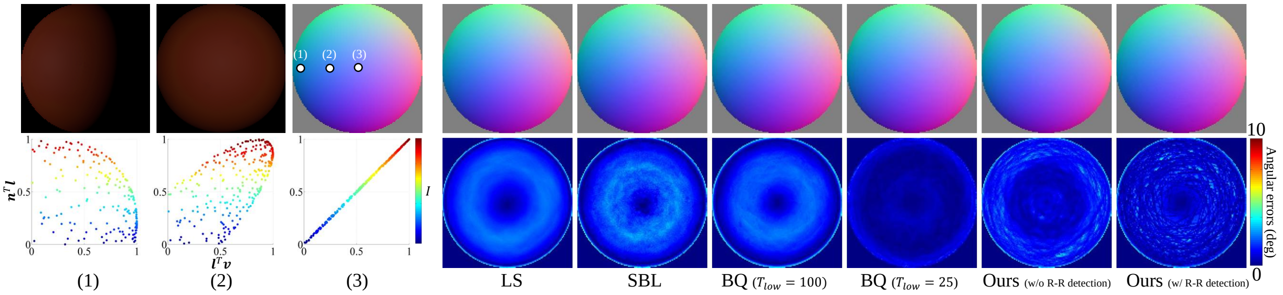


Figure 6-28: The results of dark-red-paint.

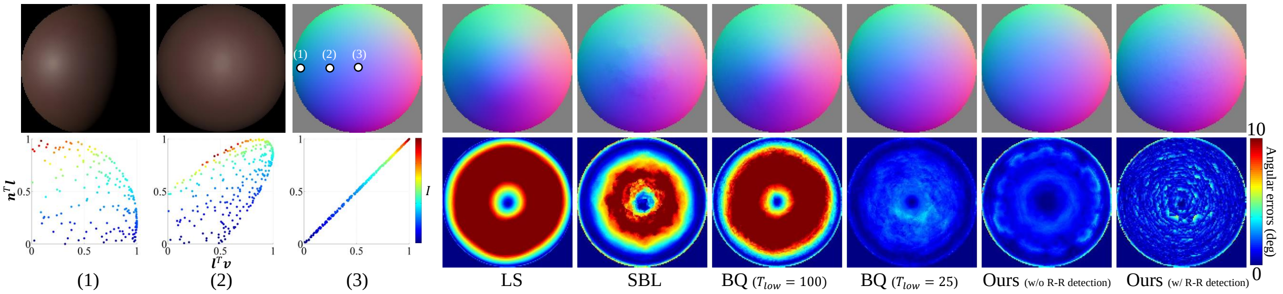


Figure 6-29: The results of dark-specular-fabric.

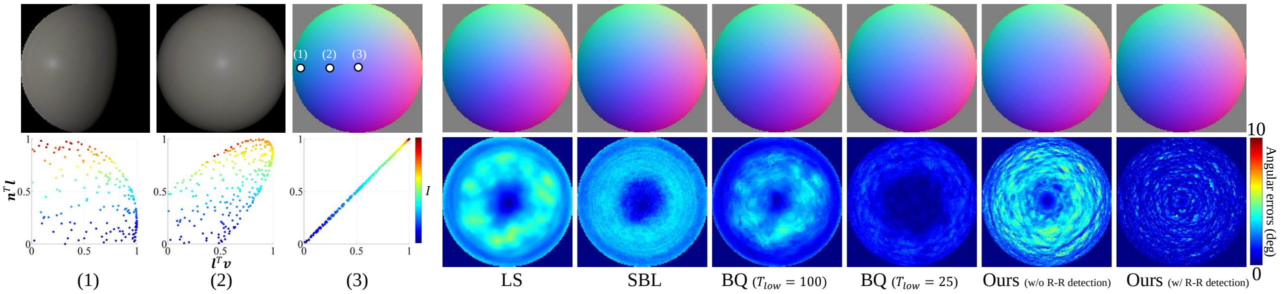


Figure 6-30: The results of delrin.

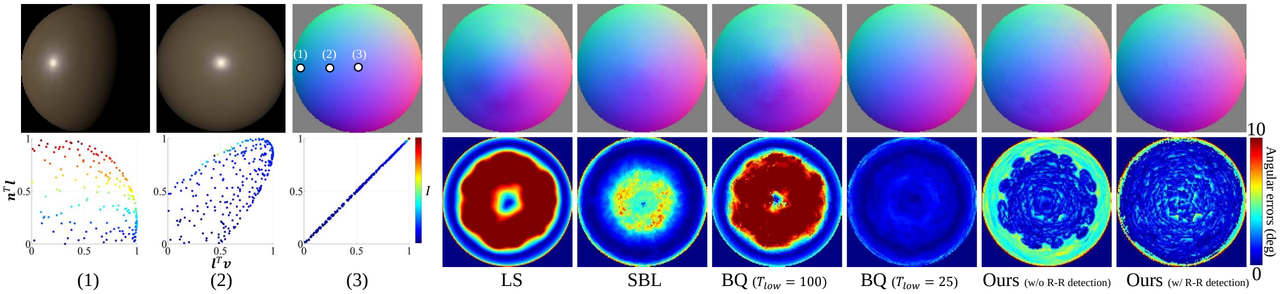


Figure 6-31: The results of fruitwood-241.

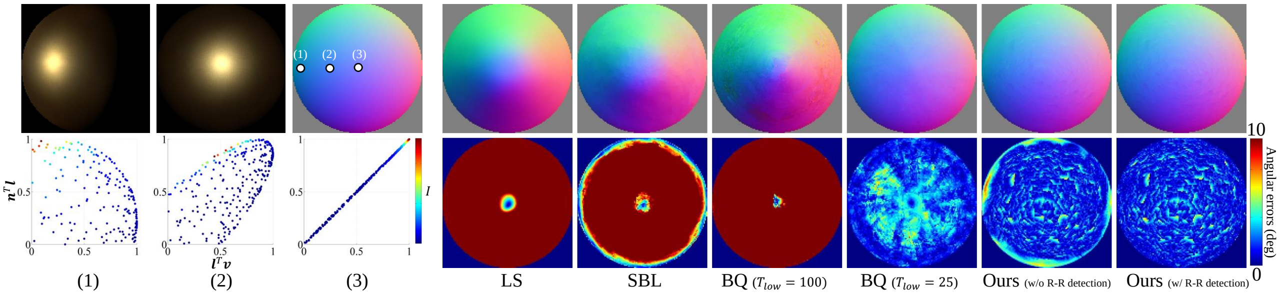


Figure 6-32: The results of gold-metallic-paint.

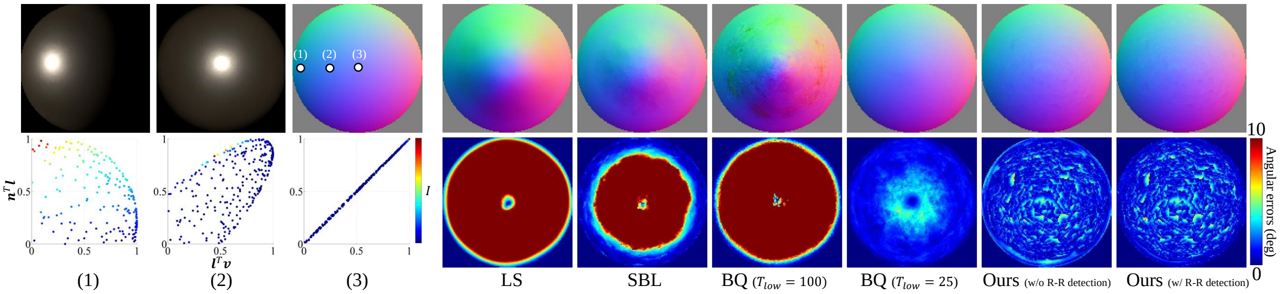


Figure 6-33: The results of gold-metallic-paint2.

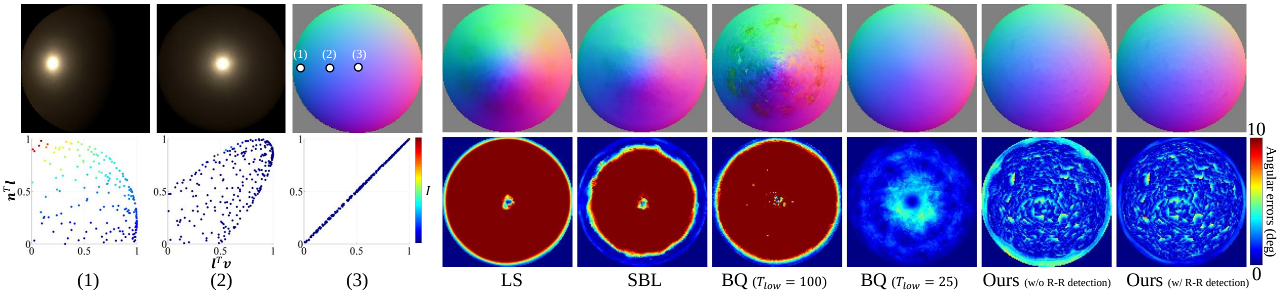


Figure 6-34: The results of gold-metallic-paint3.

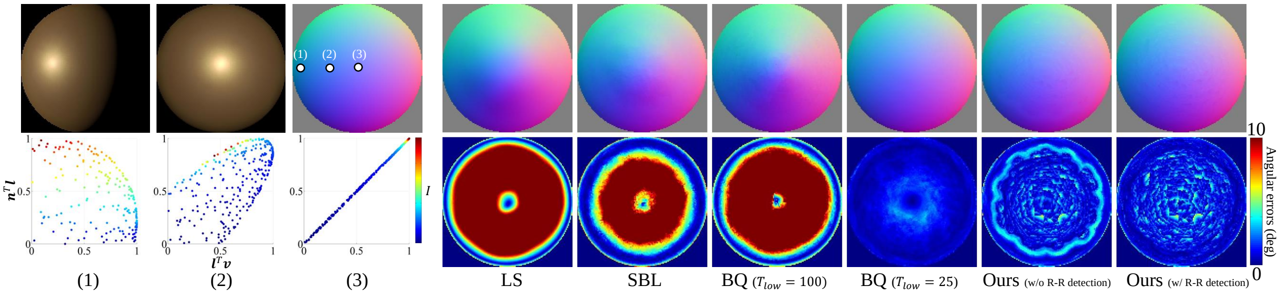


Figure 6-35: The results of gold-paint.

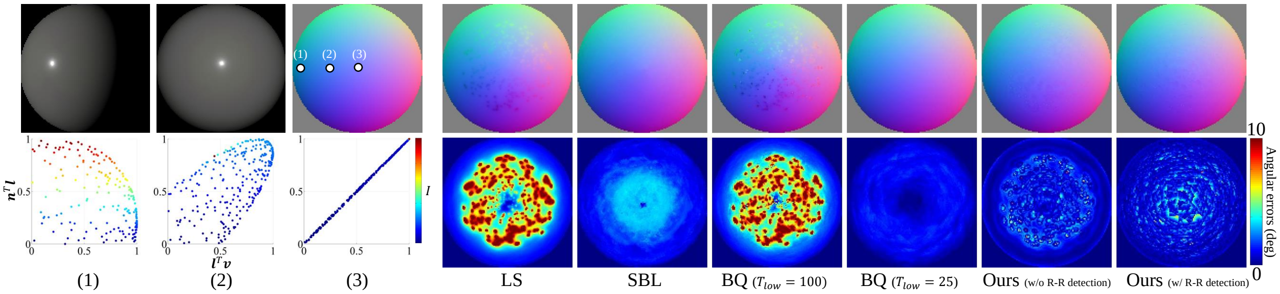


Figure 6-36: The results of gray-plastic.

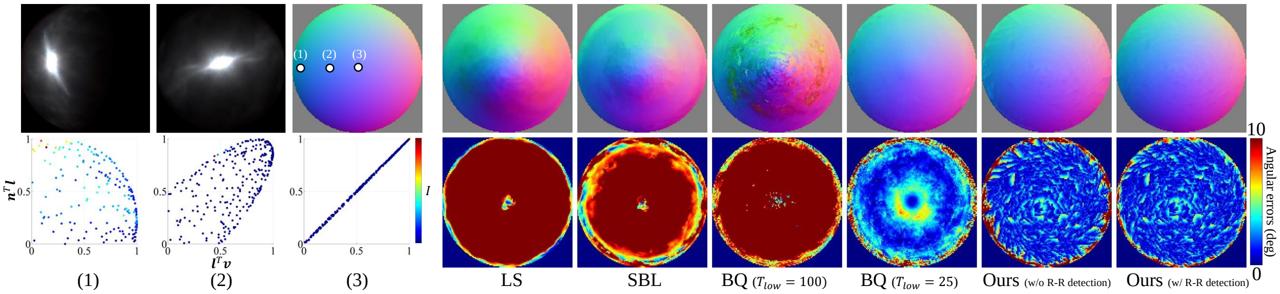


Figure 6-37: The results of grease-covered-steel.

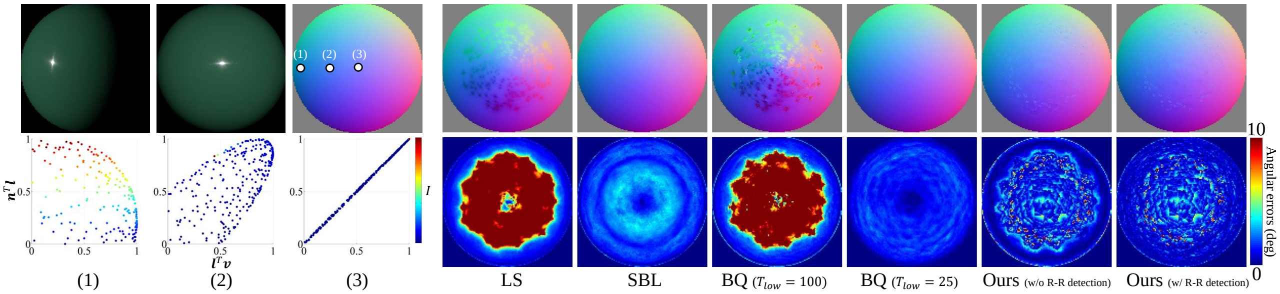


Figure 6-38: The results of green-acrylic.

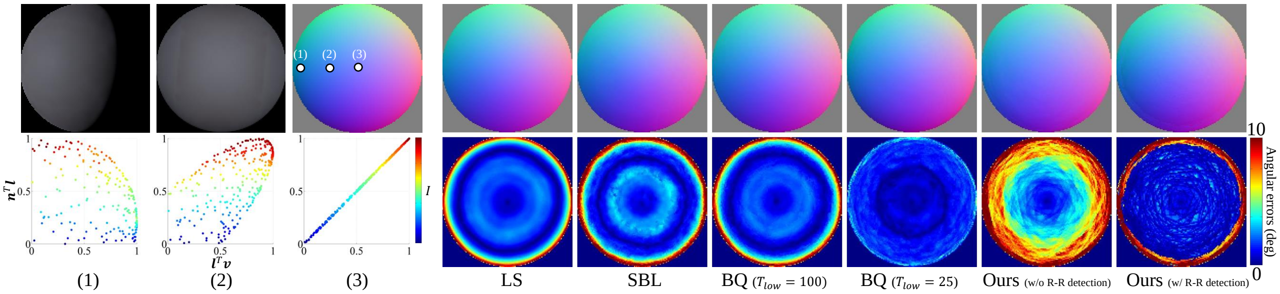


Figure 6-39: The results of green-fabric.

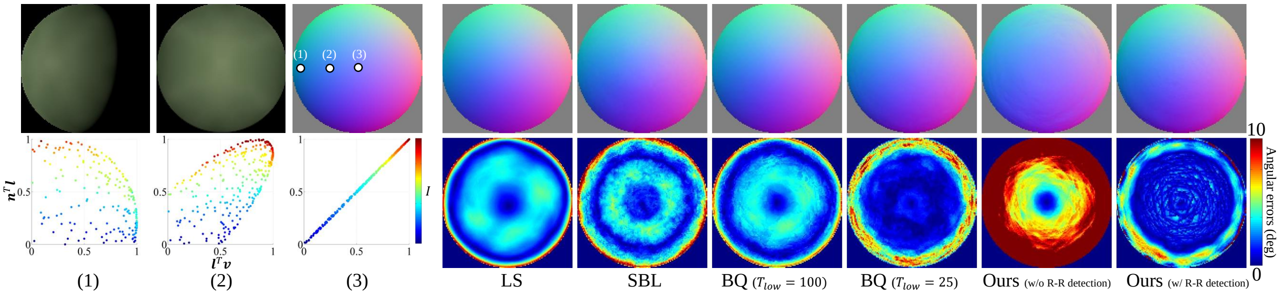


Figure 6-40: The results of green-latex.

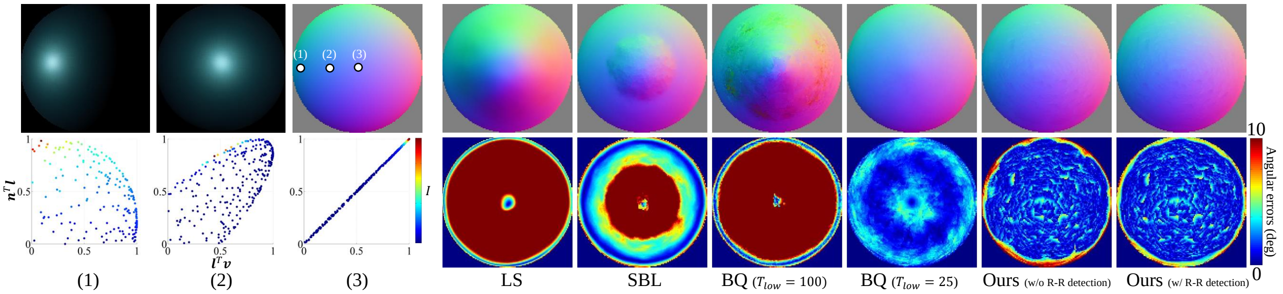


Figure 6-41: The results of green-metallic-paint.

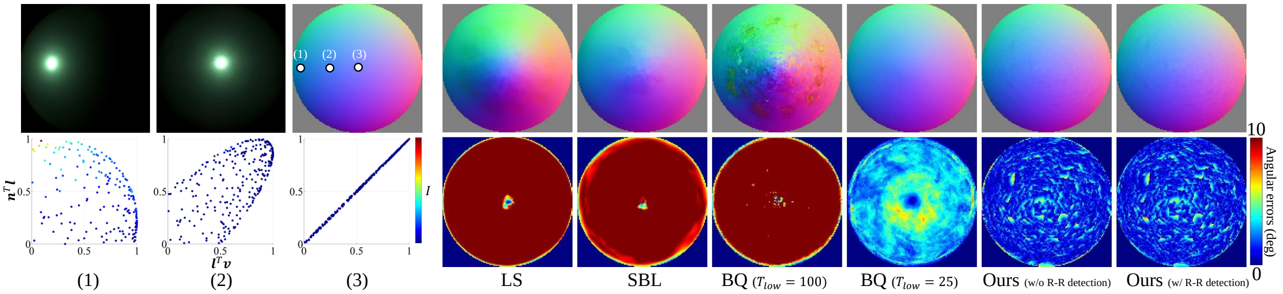


Figure 6-42: The results of green-metallic-paint2.

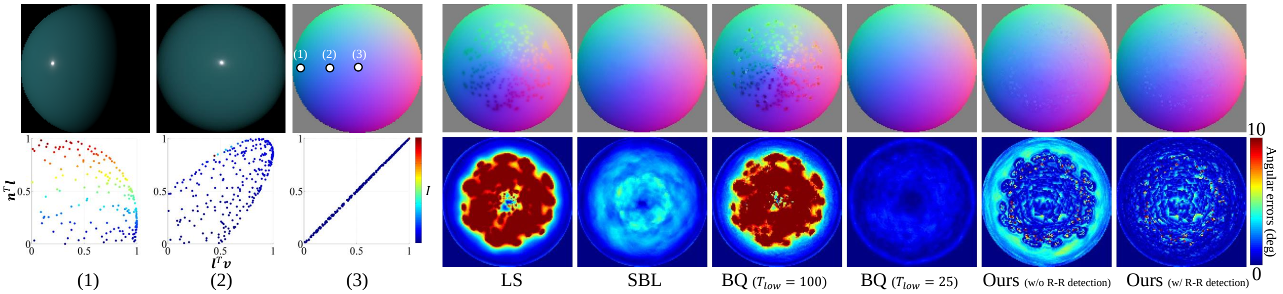


Figure 6-43: The results of green-plastic.

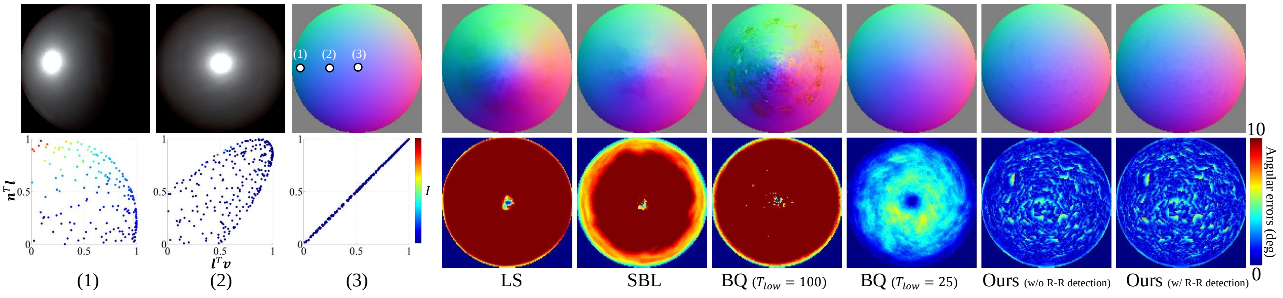


Figure 6-44: The results of hematite.

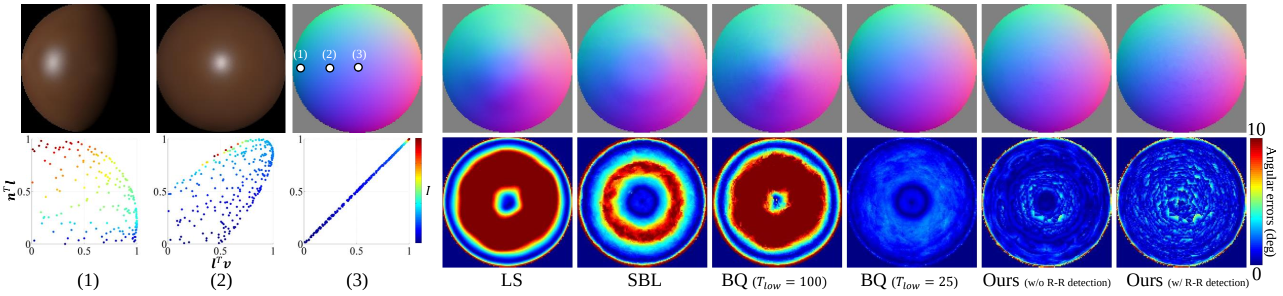


Figure 6-45: The results of ipswich-pine-221.

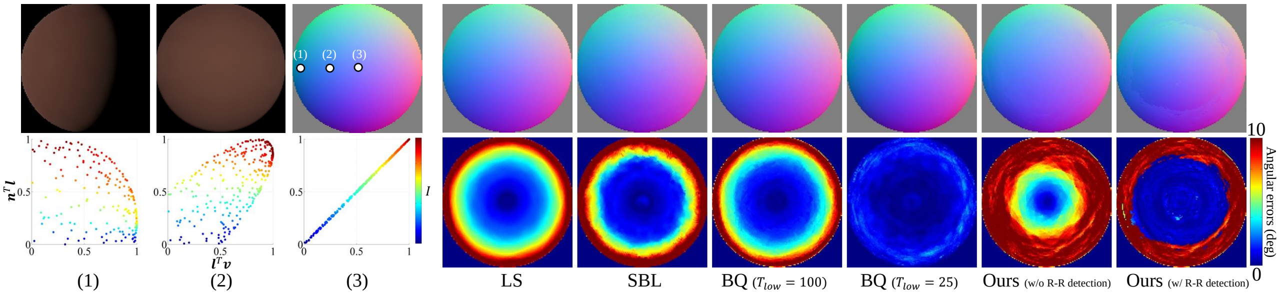


Figure 6-46: The results of light-brown-fabric.

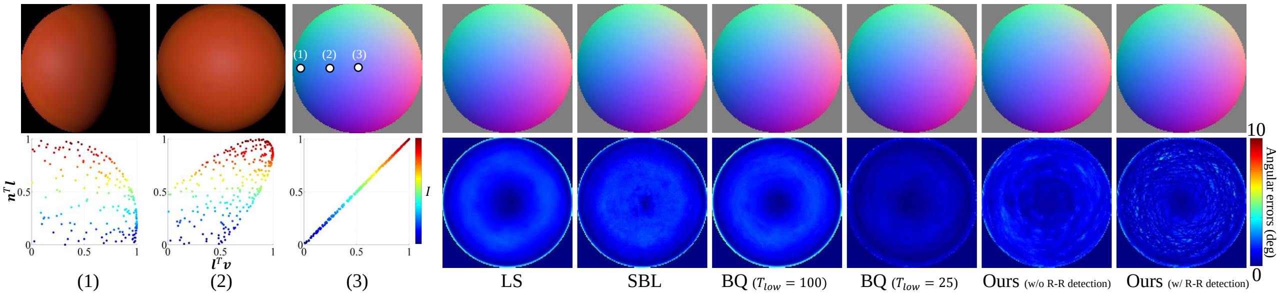


Figure 6-47: The results of light-red-paint.

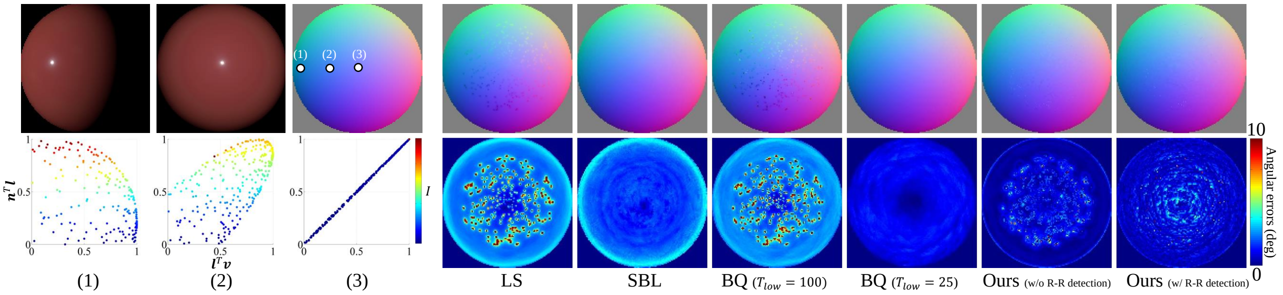


Figure 6-48: The results of maroon-plastic.

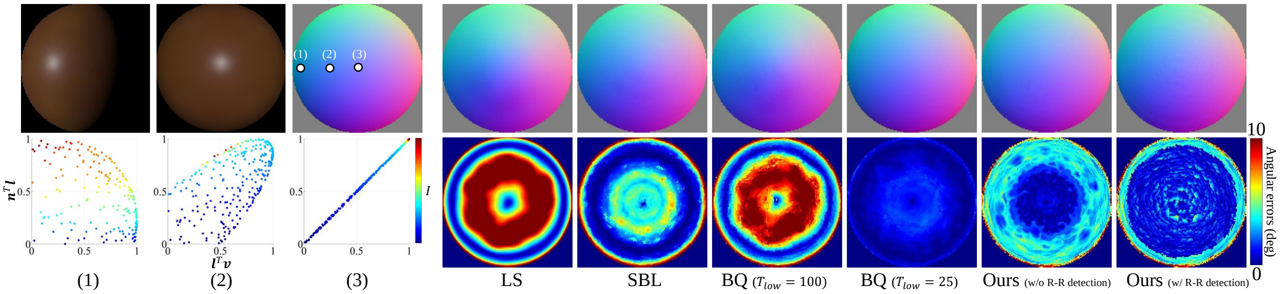


Figure 6-49: The results of natural-209.

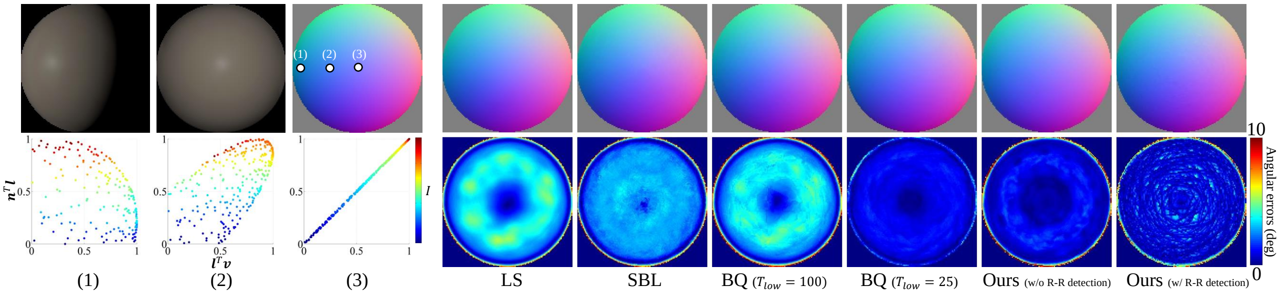


Figure 6-50: The results of neoprene-rubber.

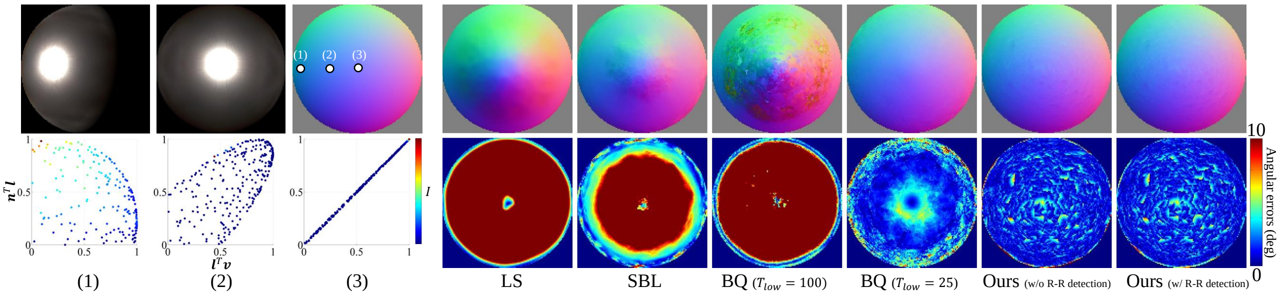
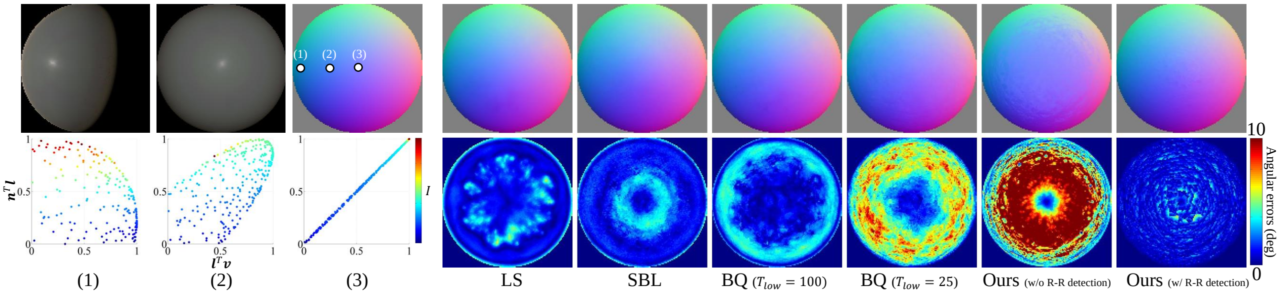
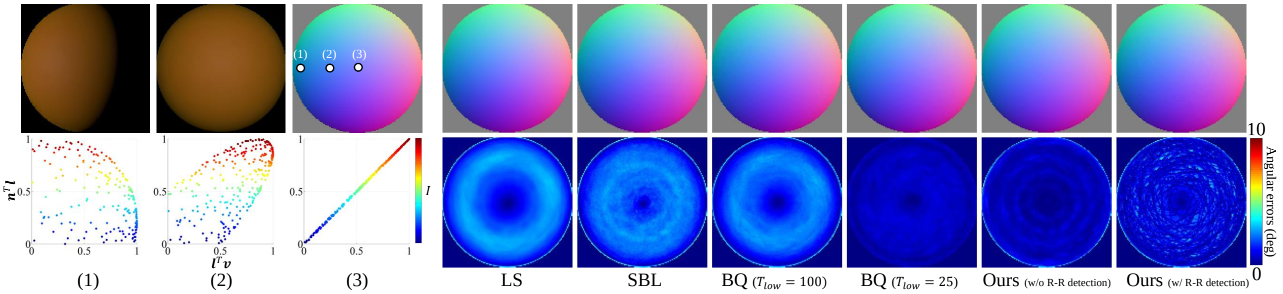


Figure 6-51: The results of nickel.



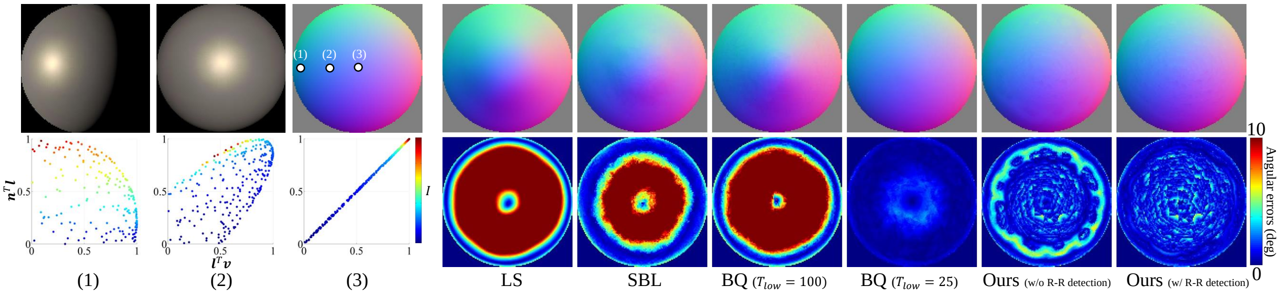
(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-52: The results of nylon.



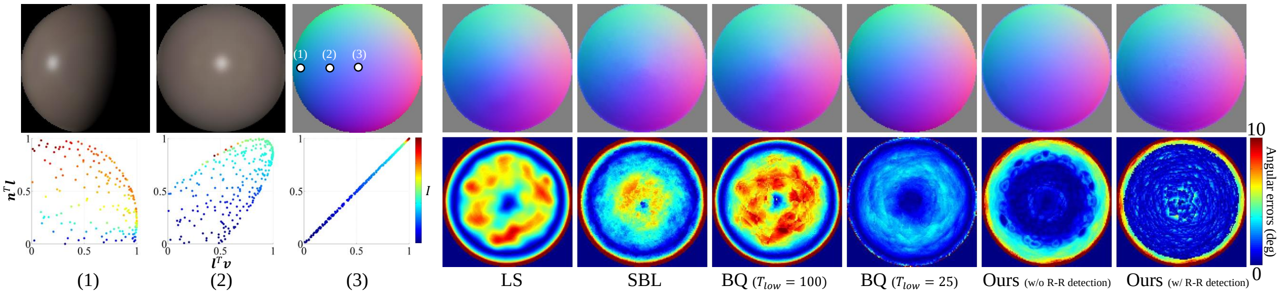
(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-53: The results of orange-paint.



(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-54: The results of pearl-paint.



(1) (2) (3) LS SBL BQ (T_{low} = 100) BQ (T_{low} = 25) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-55: The results of pickled-oak-260.

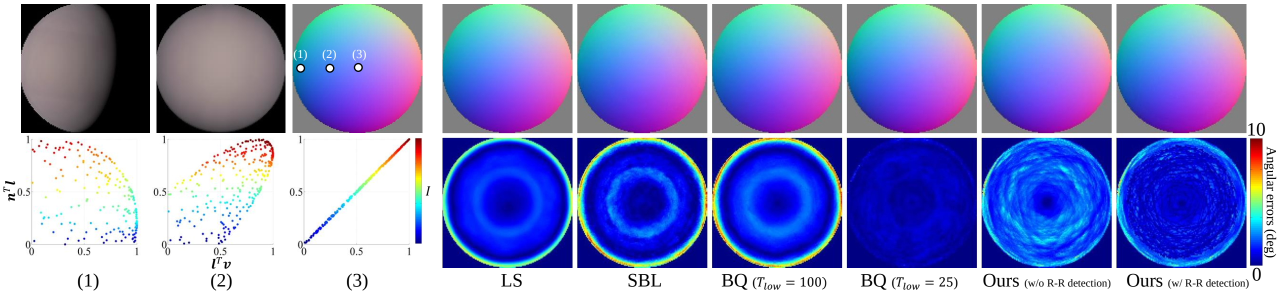


Figure 6-56: The results of pink-fabric.

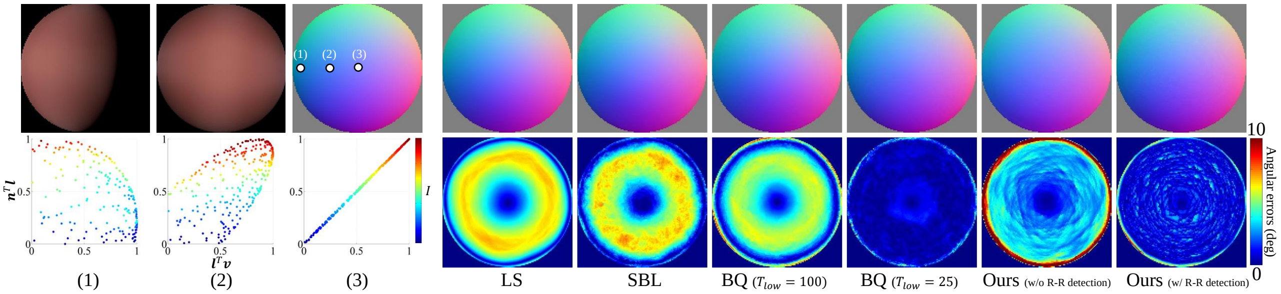


Figure 6-57: The results of pink-fabric2.

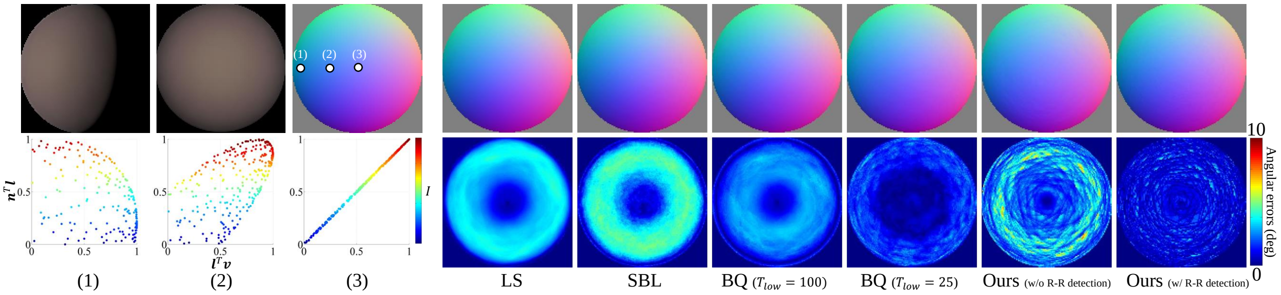


Figure 6-58: The results of pink-felt.

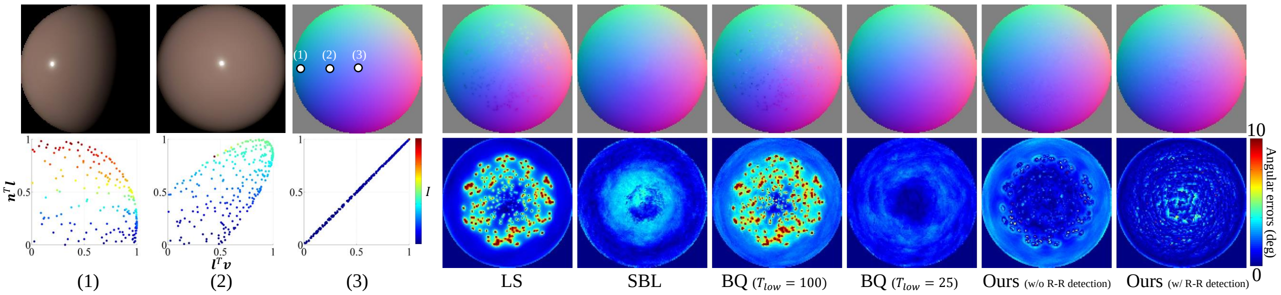


Figure 6-59: The results of pink-jasper.

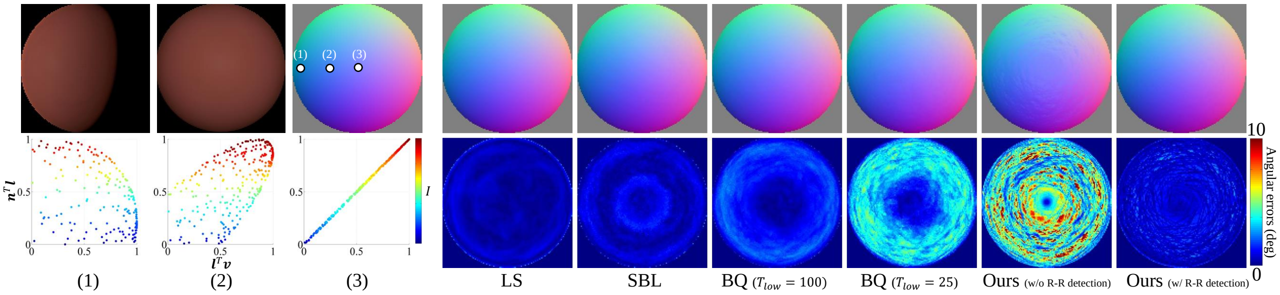


Figure 6-60: The results of pink-plastic.

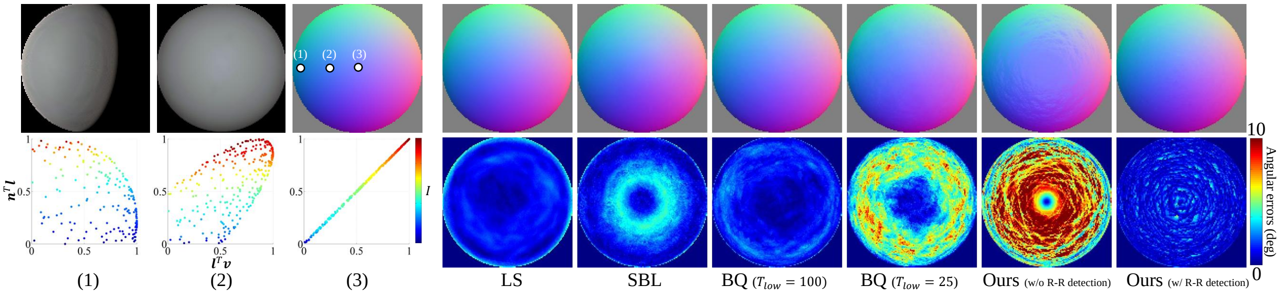


Figure 6-61: The results of polyethylene.

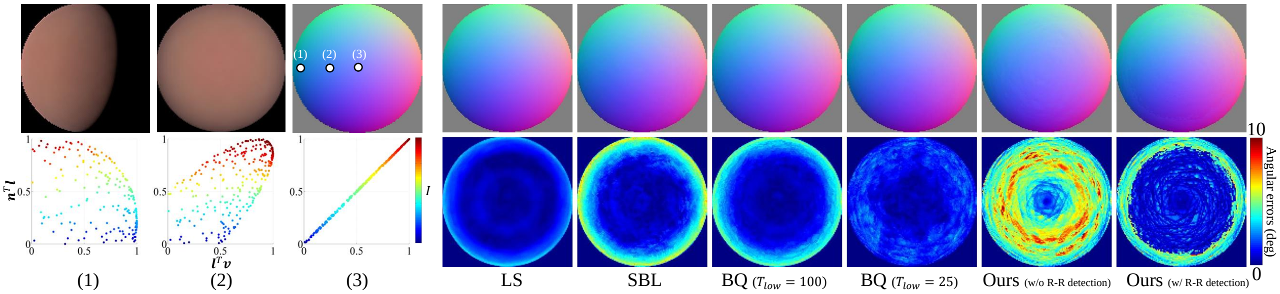


Figure 6-62: The results of polyurethane-foam.

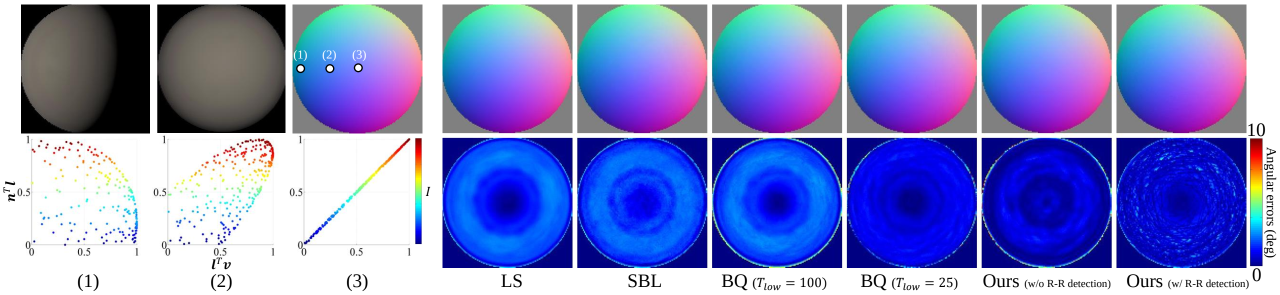


Figure 6-63: The results of pure-rubber.

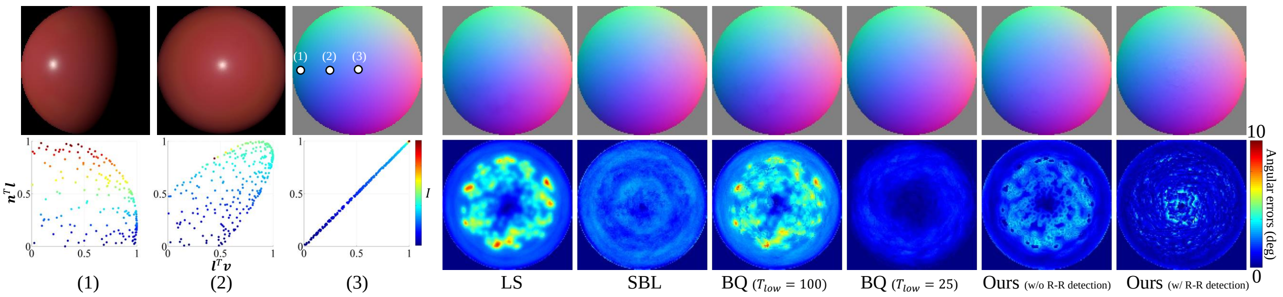


Figure 6-64: The results of purple-paint.

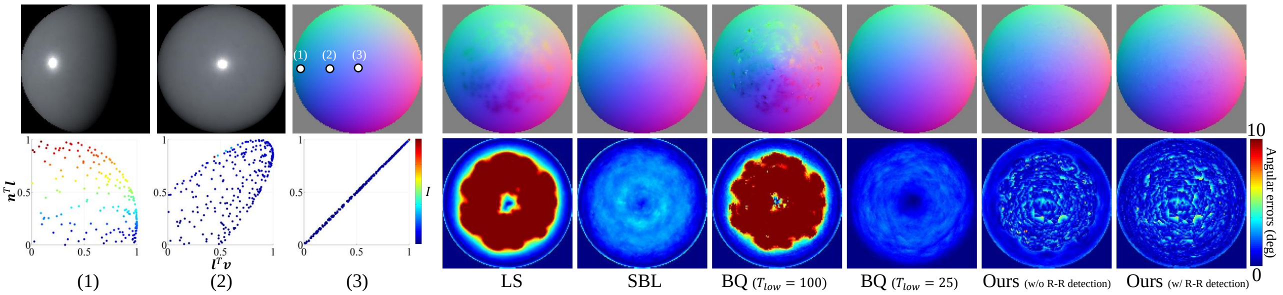


Figure 6-65: The results of pvc.

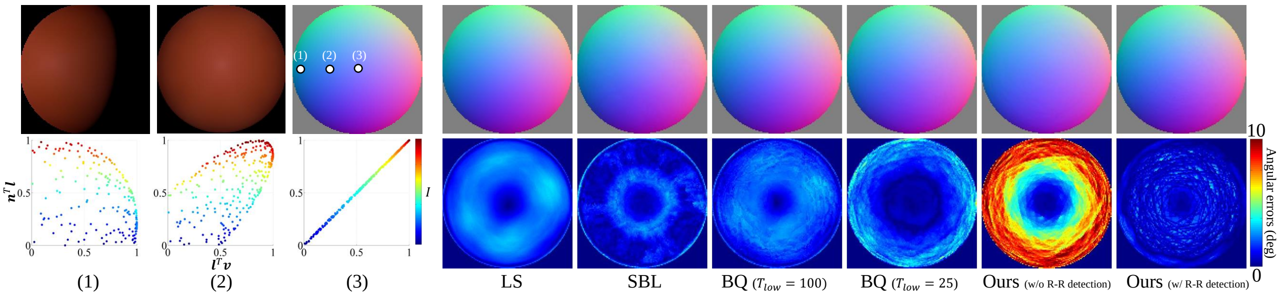


Figure 6-66: The results of red-fabric.

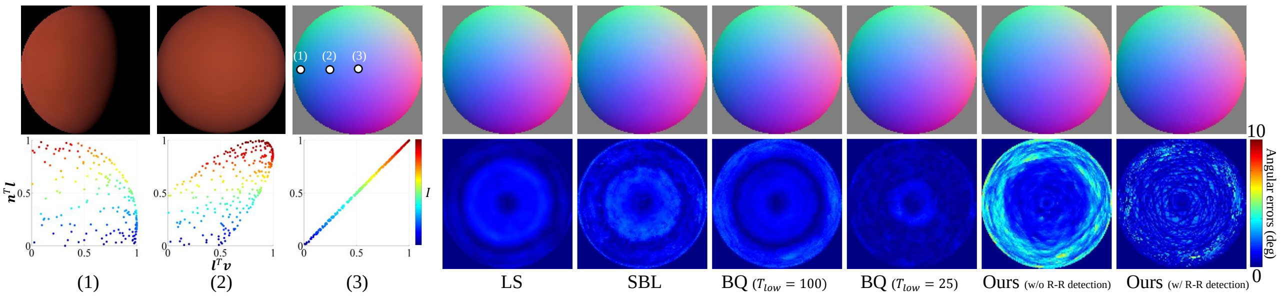


Figure 6-67: The results of red-fabric2.

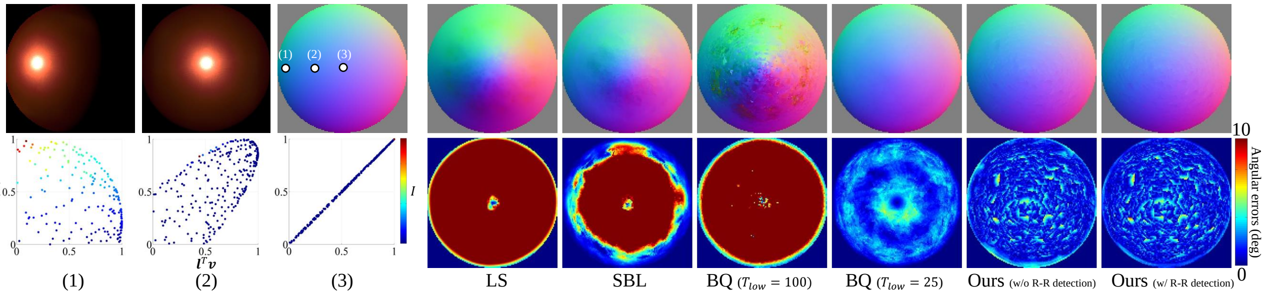

(1) (2) (3) LS SBL BQ ($T_{low} = 100$) BQ ($T_{low} = 25$) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-68: The results of red-metallic-paint.

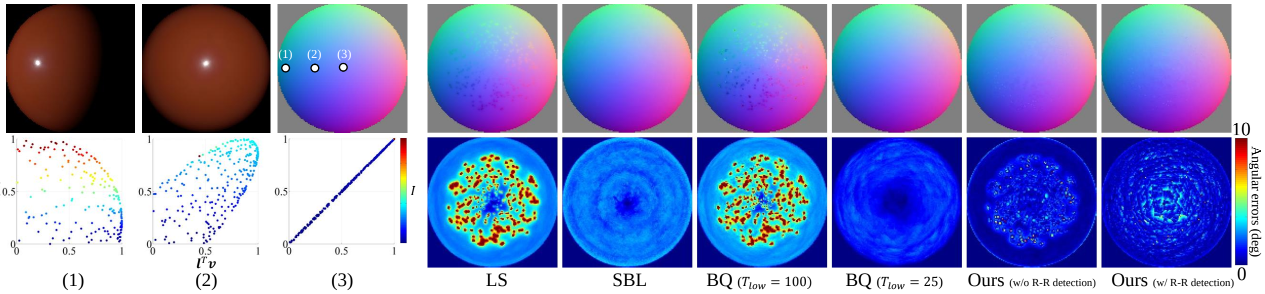

(1) (2) (3) LS SBL BQ ($T_{low} = 100$) BQ ($T_{low} = 25$) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-69: The results of red-phenolic.

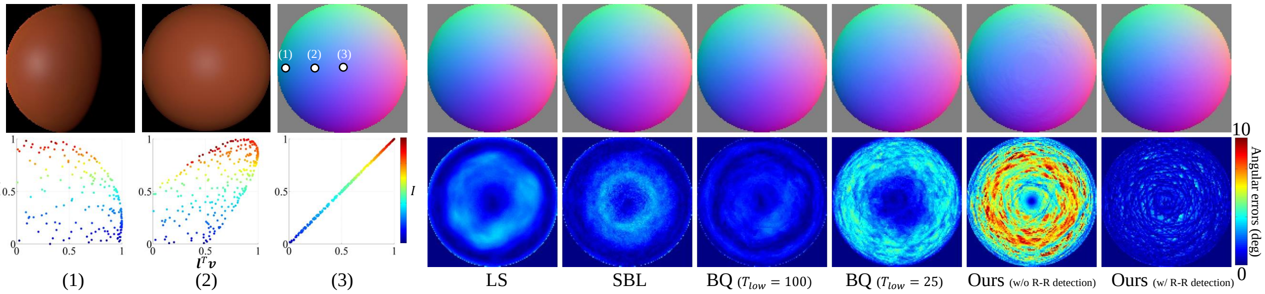

(1) (2) (3) LS SBL BQ ($T_{low} = 100$) BQ ($T_{low} = 25$) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-70: The results of red-plastic.

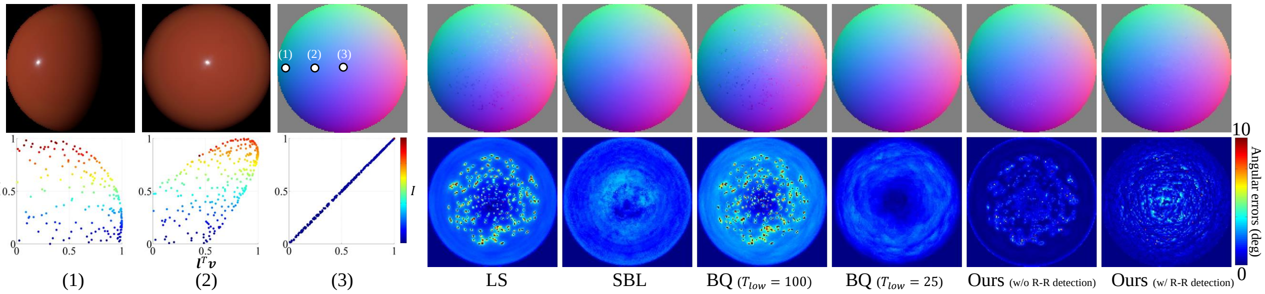

(1) (2) (3) LS SBL BQ ($T_{low} = 100$) BQ ($T_{low} = 25$) Ours (w/o R-R detection) Ours (w/ R-R detection)

Figure 6-71: The results of red-specular-plastic.

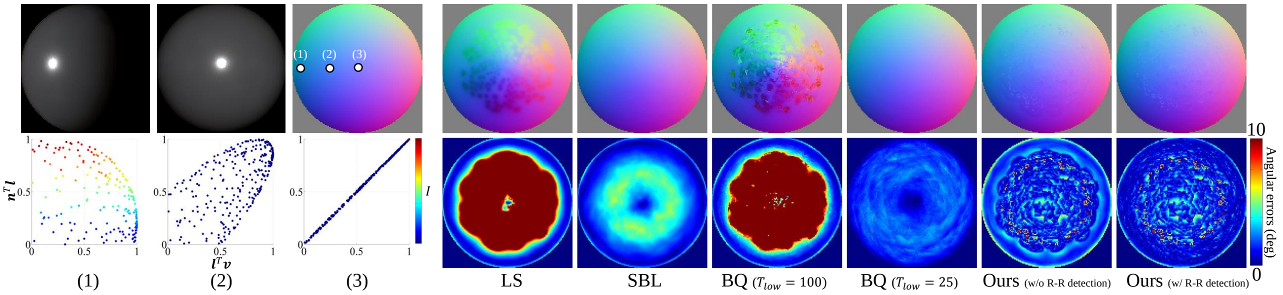


Figure 6-72: The results of silicon-nitrade.

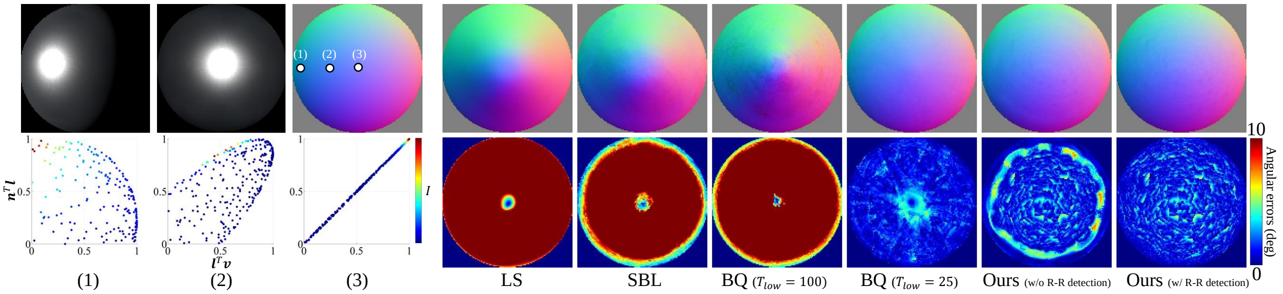


Figure 6-73: The results of silver-metallic-paint.

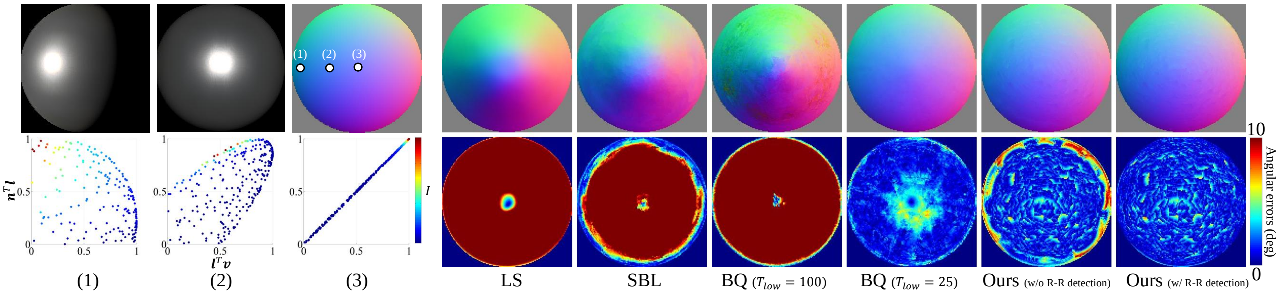


Figure 6-74: The results of silver-metallic-paint2.

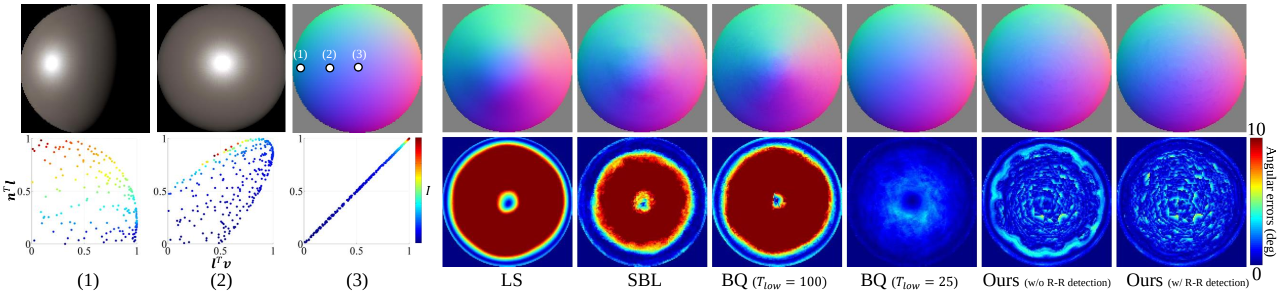


Figure 6-75: The results of silver-paint.

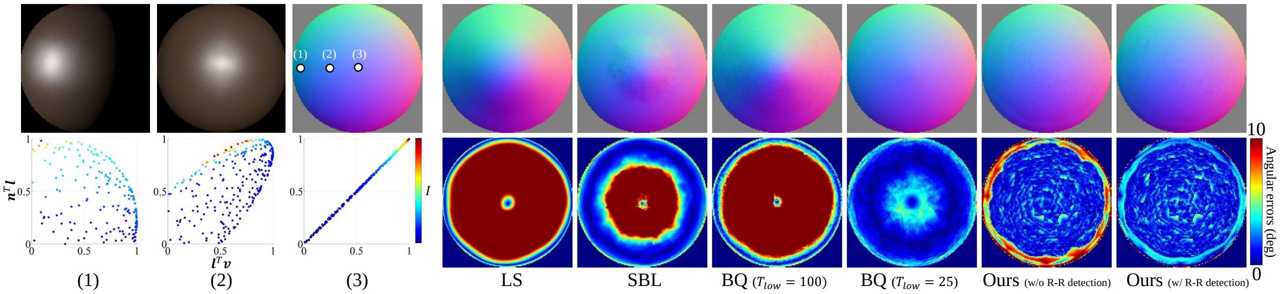


Figure 6-76: The results of special-walnut-224.

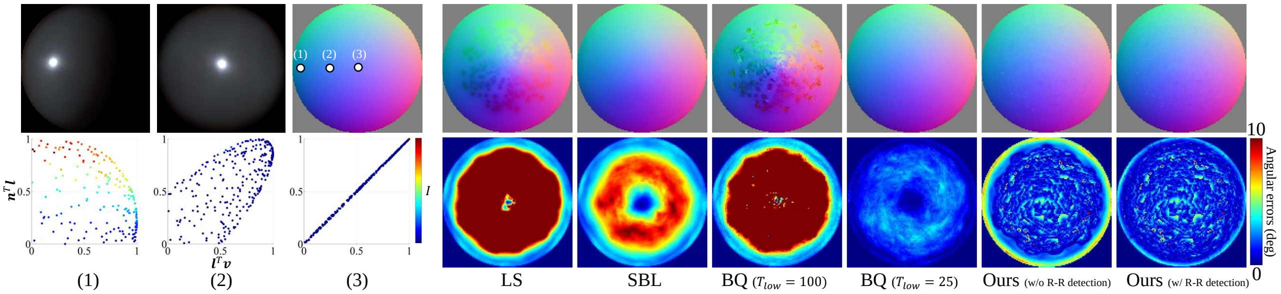


Figure 6-77: The results of specular-black-phenolic.

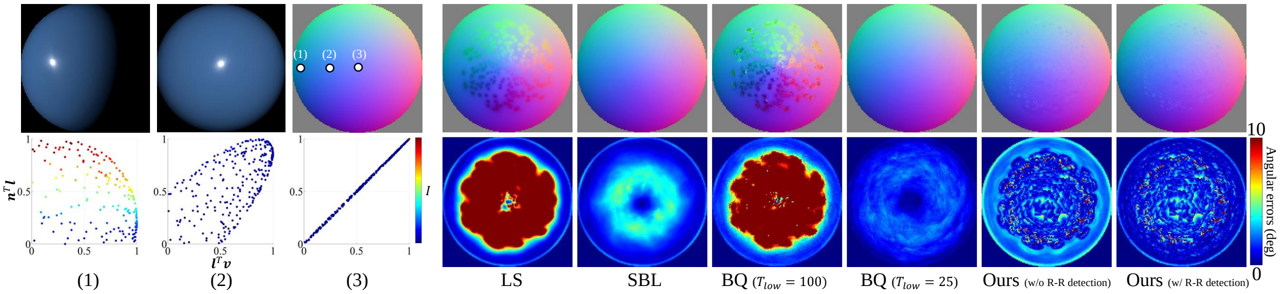


Figure 6-78: The results of specular-blue-phenolic.

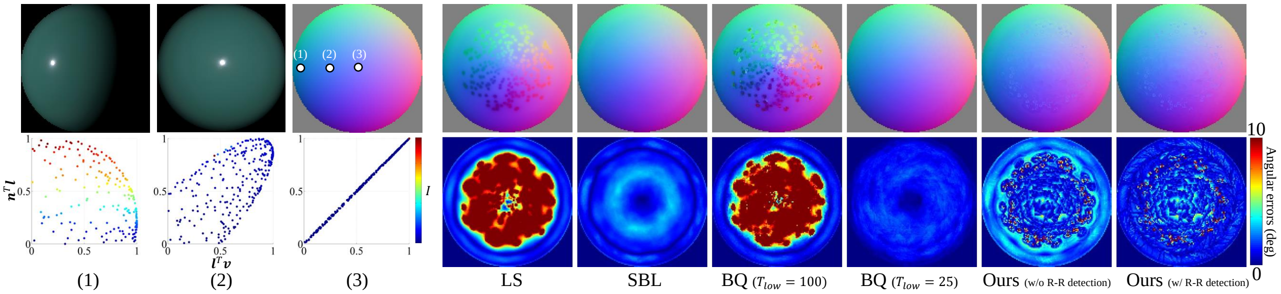


Figure 6-79: The results of specular-green-phenolic.

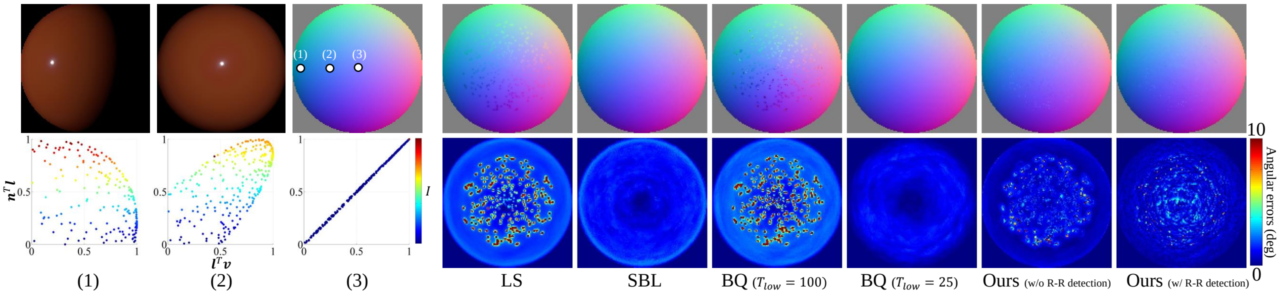


Figure 6-80: The results of specular-maroon-phenolic.

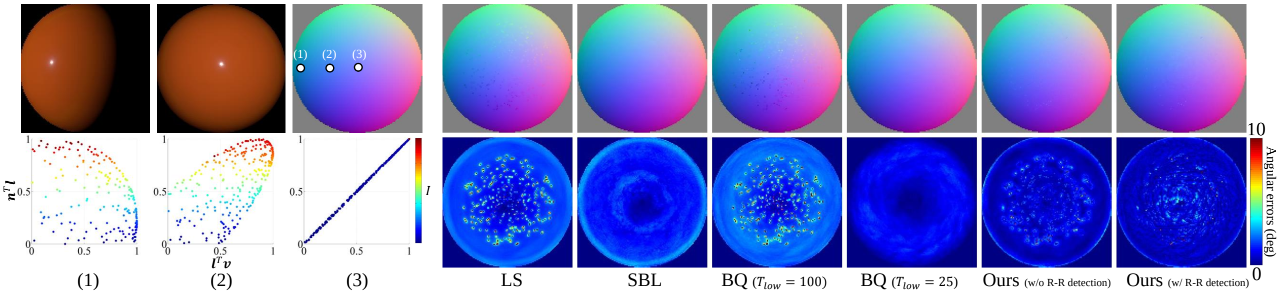


Figure 6-81: The results of specular-orange-phenolic.

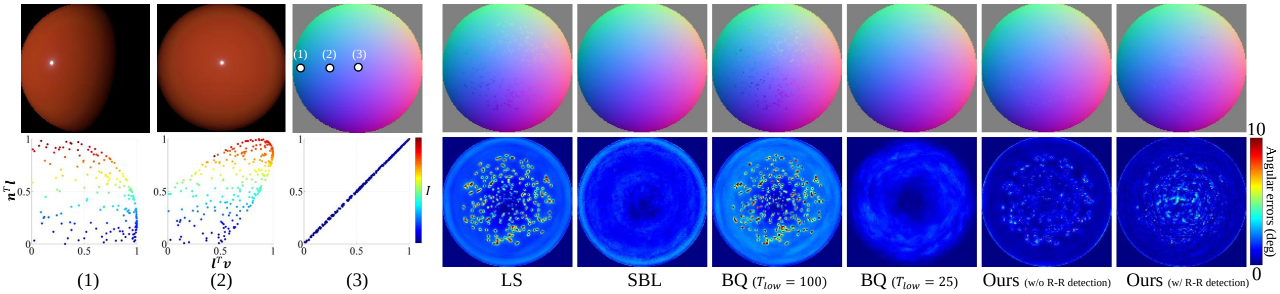


Figure 6-82: The results of specular-red-phenolic.

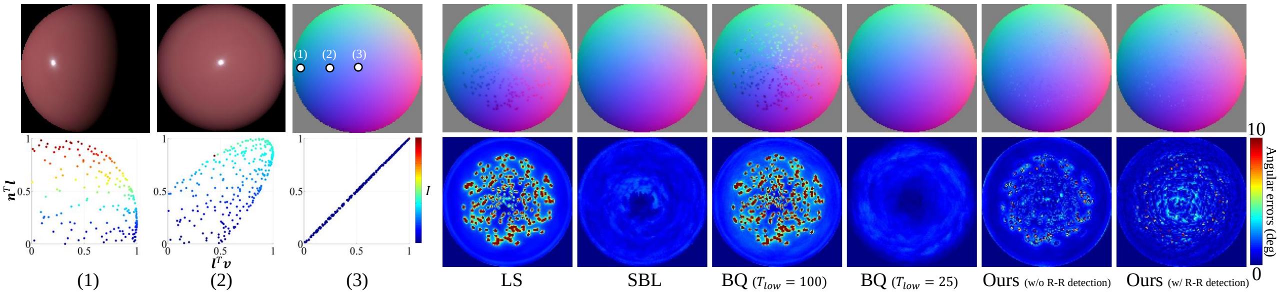


Figure 6-83: The results of specular-violet-phenolic.

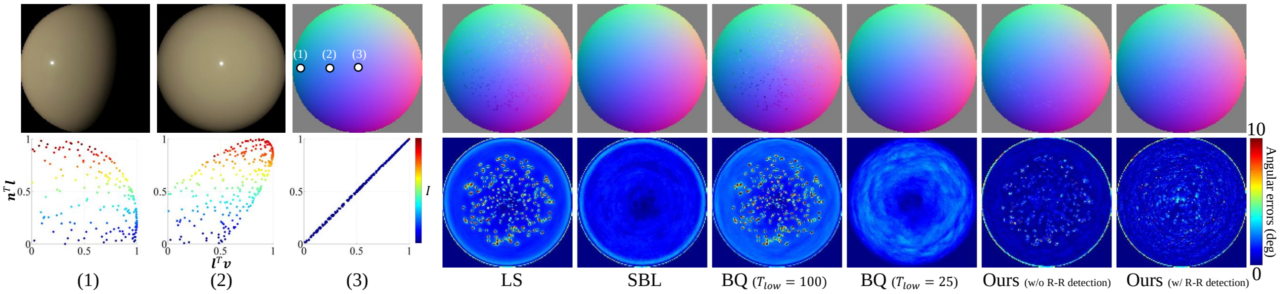


Figure 6-84: The results of specular-white-phenolic.

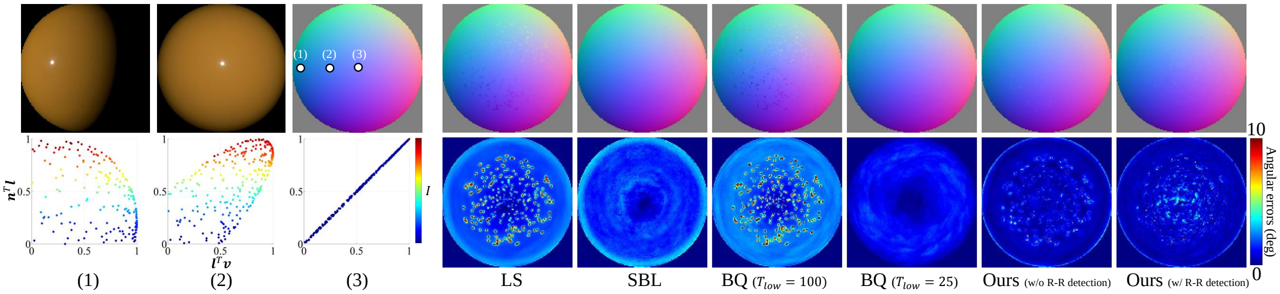


Figure 6-85: The results of specular-yellow-phenolic.

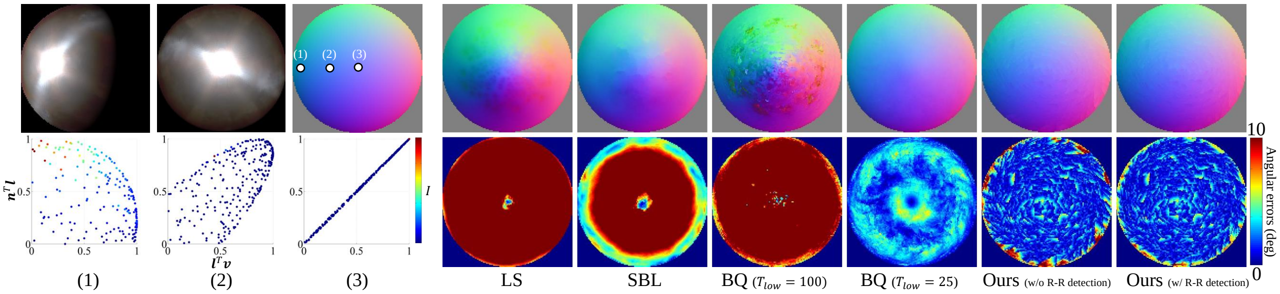


Figure 6-86: The results of ss440.

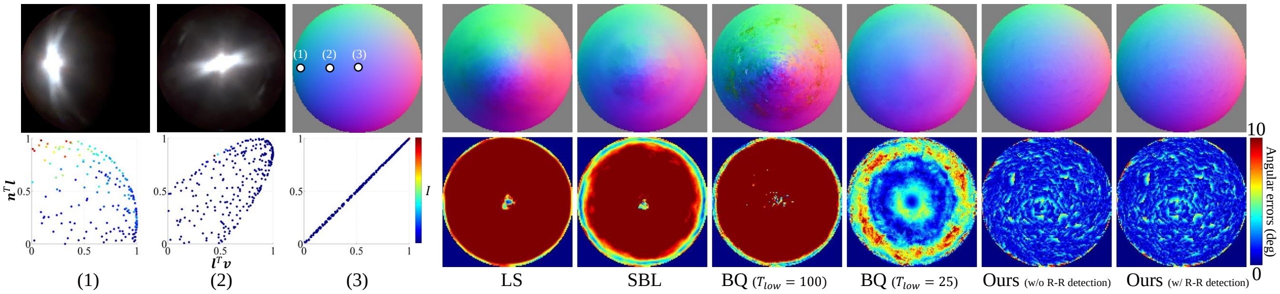


Figure 6-87: The results of steel.

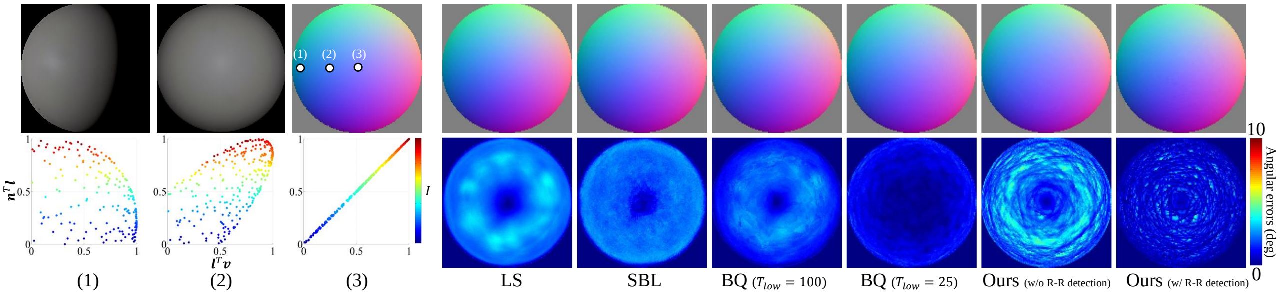


Figure 6-88: The results of teflon.

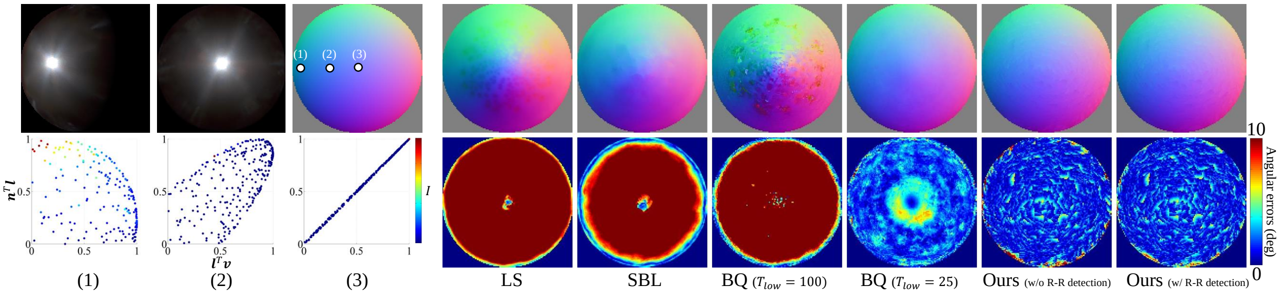


Figure 6-89: The results of tungsten-carbide.

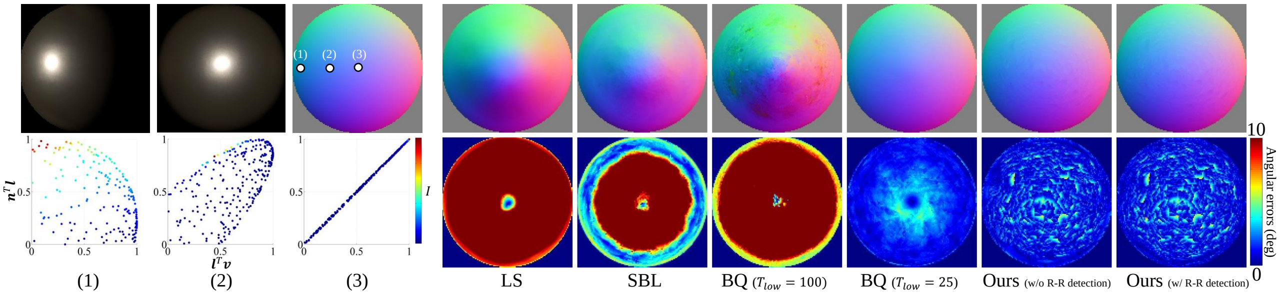


Figure 6-90: The results of two-layer-gold.

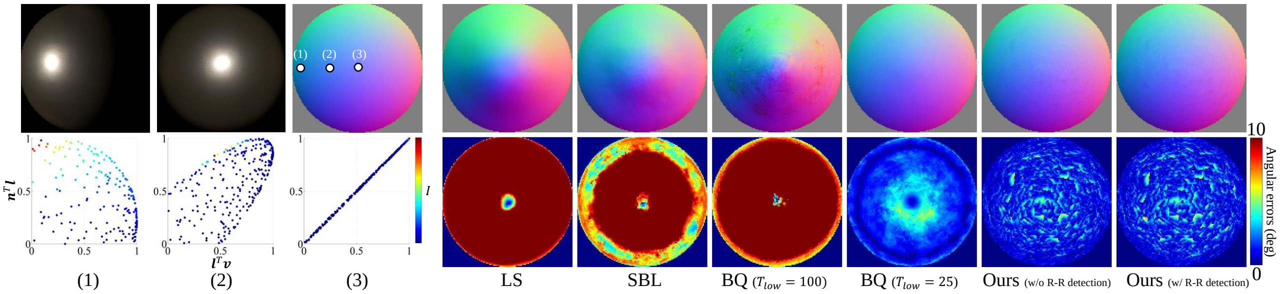


Figure 6-91: The results of two-layer-silver.

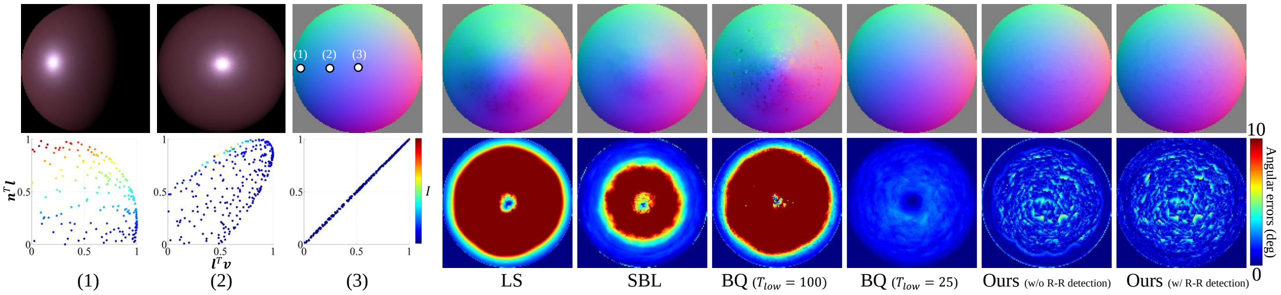


Figure 6-92: The results of violet-acrylic.

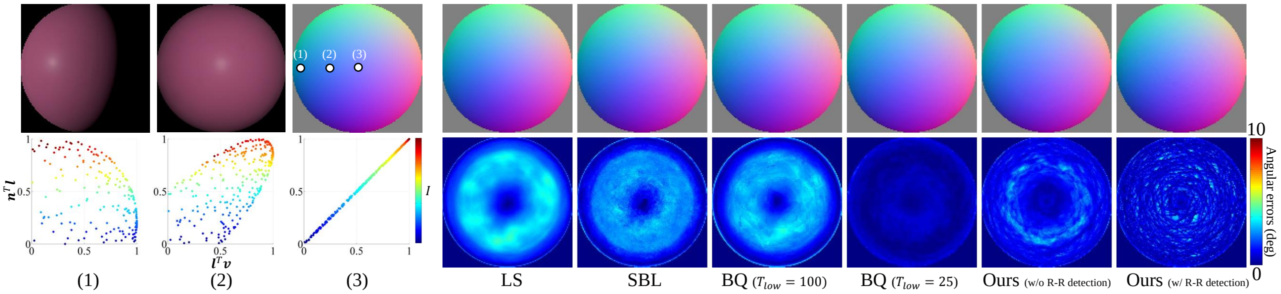


Figure 6-93: The results of violet-rubber.

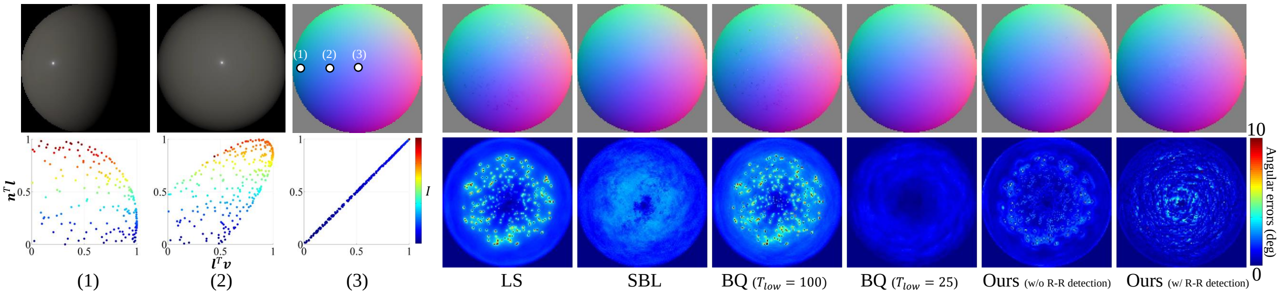


Figure 6-94: The results of white-acrylic.

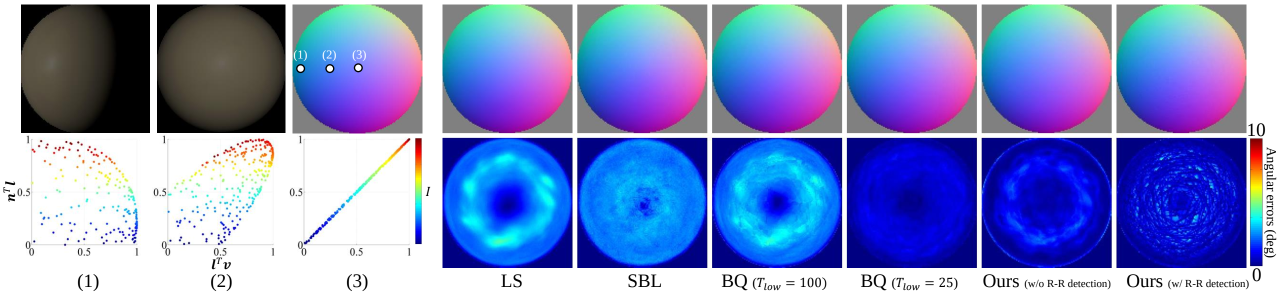


Figure 6-95: The results of white-diffuse-bball.

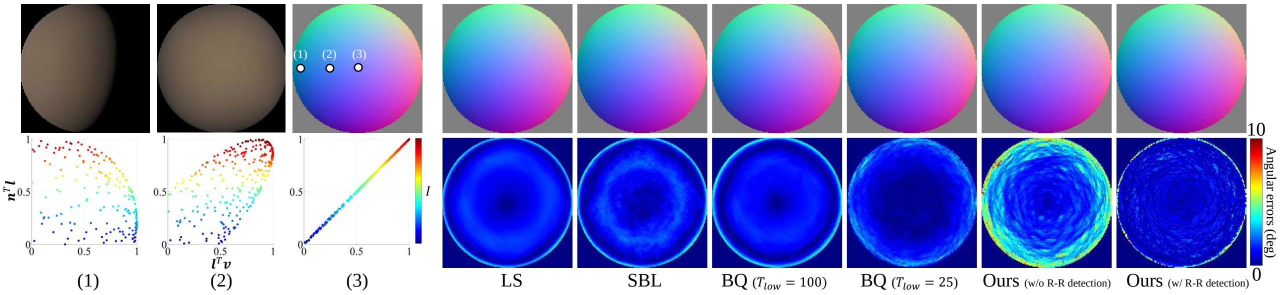


Figure 6-96: The results of white-fabric.

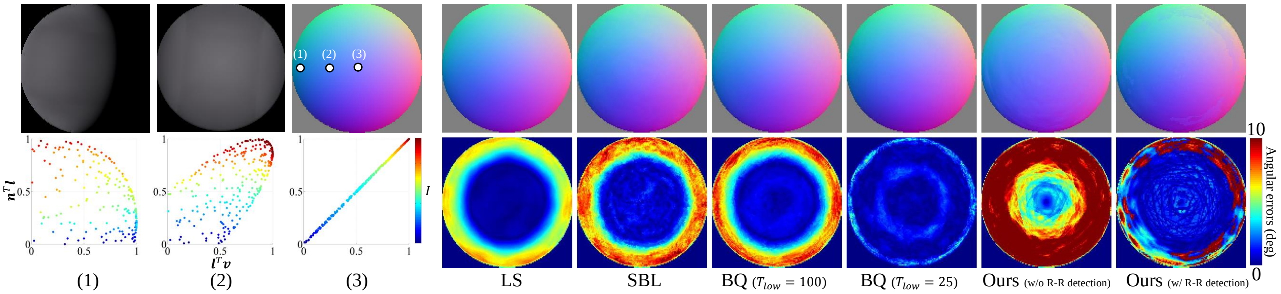


Figure 6-97: The results of white-fabric2.

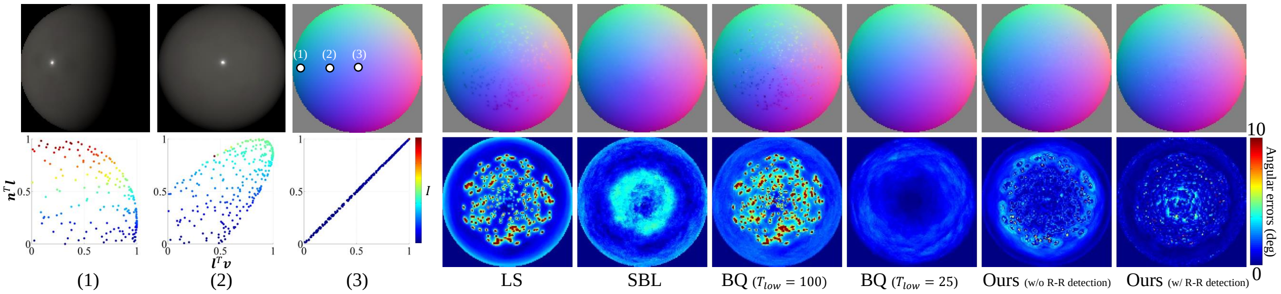


Figure 6-98: The results of white-marble.

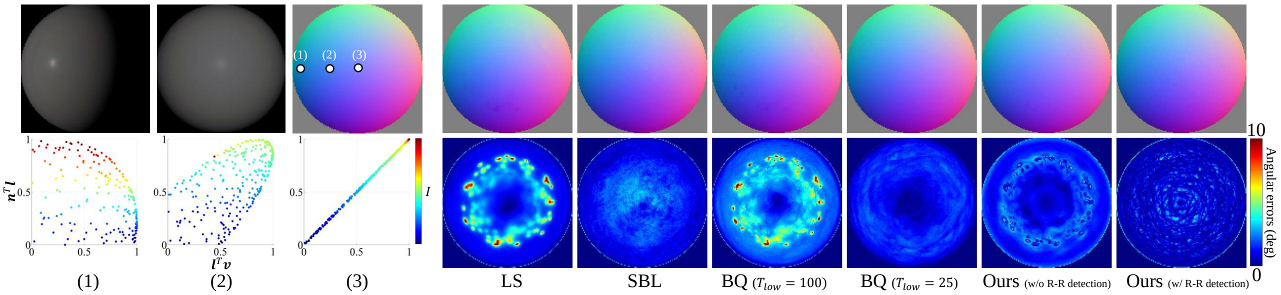


Figure 6-99: The results of white-paint.

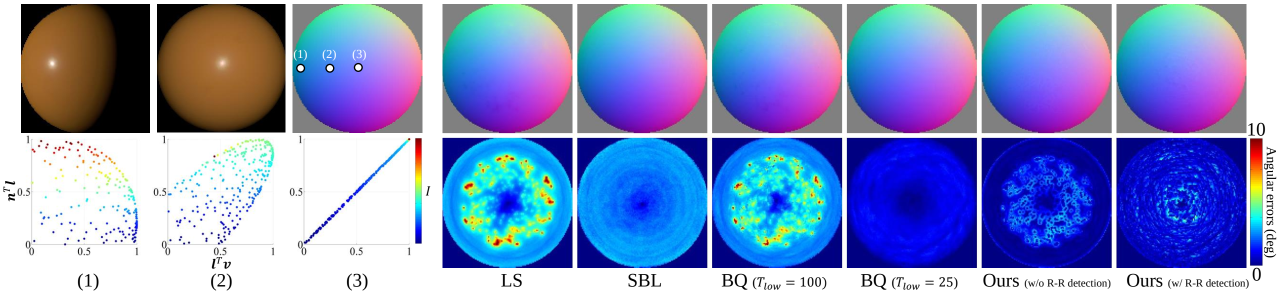


Figure 6-100: The results of yellow-matte-plastic.

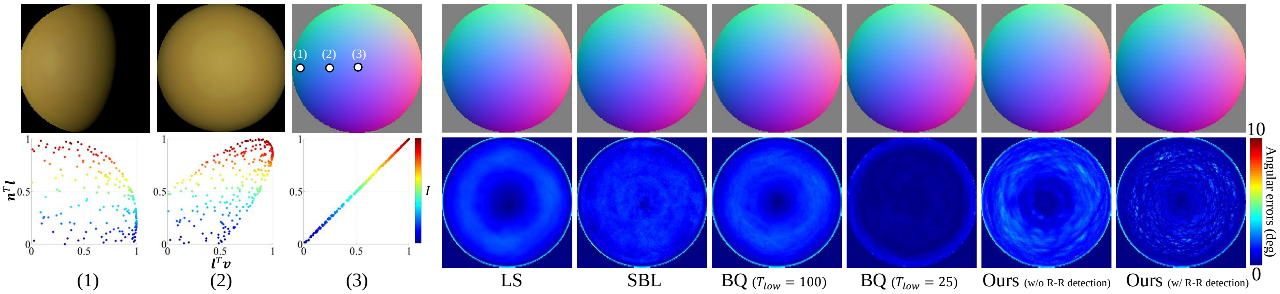


Figure 6-101: The results of yellow-paint.

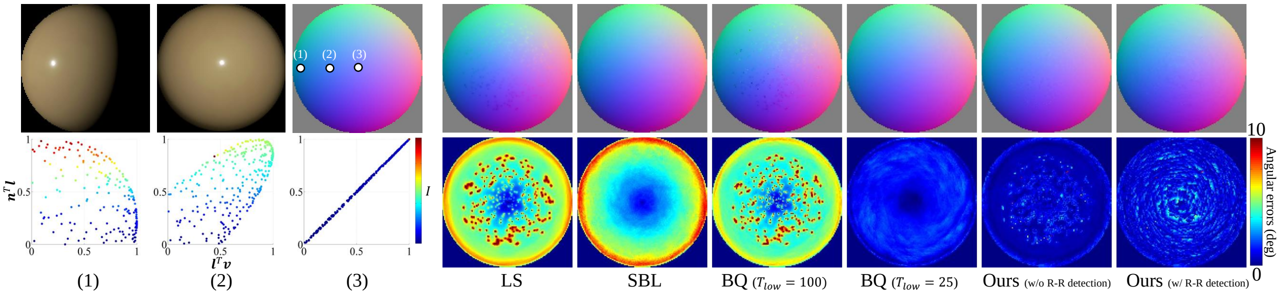


Figure 6-102: The results of yellow-phenolic.

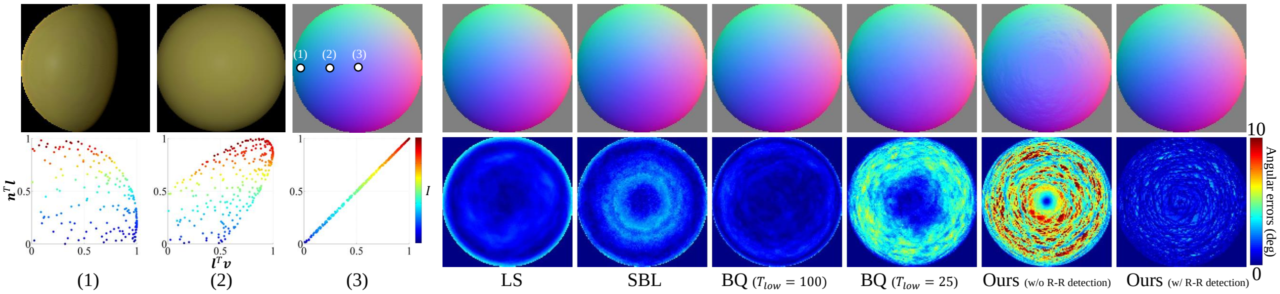


Figure 6-103: The results of yellow-plastic.

Bibliography

[1] http://www.blender.org/.

[2] A. Abrams, C. Hawley, and R. Pless. Heliometric stere: Shape from sun position. *Proc. ECCV*, 2012.

[3] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *Proc. CVPR*, 2012.

[4] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field ? In *Proc. ECCV*, 2006.

[5] M. Aittala, T. Weyrich, and J. Lehtinen. Practical svbrdf capture in the frequency domain. *Proc. ACM SIGGRAPH*, 2013.

[6] N. Alldrin and D. Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Proc. ICIP*, 2007.

[7] N. Alldrin and D. Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Proc. ICCV*, 2007.

[8] N. Alldrin, S. Mallick, and D. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. *Proc. CVPR*, 2007.

[9] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. CVPR*, 2008.

[10] R. Anderson, B. Stenger, and R. Cipolla. Color photometric stereo for multicolored surfaces. *Proc. ICIP*, 2011.

[11] V. Argyriou and M. Petrou. Recursive photometric stereo when multiple shadows and highlights are present. In *Proc. CVPR*, 2008.

[12] I. Ashdown. Near-field photometry: Measureing and modeling complex 3-d light sources. *Proc. ACM SIGGRAPH*, 1995.

[13] M. Ashikhmin and P. Shirley. An an isotropic phong brdf model. *Journal on Graphics Tools*, 5(2):25–32, 2000.

[14] S. Barsky. The 4-source photometri stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1239–1252, 2003.

[15] K. Bollaerts, P. H. Eilers, and I. Mechelen. Simple and multiple p-splines regression with shape constraints. *British J. Math. Statist. Psych.*, 59(2):451–469, 2006.

[16] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[17] E. Candés and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[18] E. Candés, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourie Anal. Appl*, 14(5):877–905, 2008.

[19] P. Chak. Semi-nonparametric estimation with bernstein polynomials. *Economics Letters*, 89(2):153–156, 2005.

[20] M. Chandraker, S. Agarwal, and D. Kriegman. Shadowcuts: Photometric stereo with shadows. In *Proc. CVPR*, 2007.

[21] M. Chandraker, J. Bai, and R. Ramamoorthi. A theory of differential photometric stereo for unknown brdfs. In *Proc. CVPR*, 2011.

[22] M. Chandraker, F. Kahl, and D. Kriegman. Reflections on the generalized bas-relief ambiguity. *Proc. CVPR*, 2005.

[23] M. Chandraker and R. Ramamoorthi. What an image reveals about material reflectance. In *Proc. ICCV*, 2011.

[24] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *Int. Conf. Accoust., Speech, Signal Process.*, pages 3869–3872, 2008.

[25] S. Choi and J. Kim. Robust regression to varying data distribution and its application to landmark-based localization. *Proc. IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2008.

[26] H. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. *Proc. CVPR*, 2008.

[27] H. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proc. CVPR*, 2008.

[28] E. Coleman and R. Jain. *Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry*. Jones and Bartlett Publishers Inc., 1992.

[29] R. Cook and K. Torrance. A reflectance model for computer graphics. *ACM Trans. on Graph.*, 15(4):307–316, 1981.

[30] M. Davies and R. Gribonval. Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p \geq 1$. Technical report, Technical Report IRISA, 2008.

[31] B. Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. *Proc. CVPR*, 2009.

[32] O. Drbohlav and M. Chantler. On optimal light configurations in photometric stereo. *Proc. ICIP*, 2005.

[33] M. Drew and B. Funt. Calculating surface reflectance using a single-bounce model of mutual reflection. *Proc. ICIP*, 1990.

[34] H. Du, D. Goldman, and S. Seitz. Binocular photometric stereo. *BMVC*, 2011.

[35] R. Duda and P. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[36] J. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *CVIU*, 109(1):22–43, 2008.

[37] J. Elder, S. Trithart, G. Pintilie, and D. MacLean. Rapid processing of cast and attached shadows. *Perception*, 33(11):1319–1338, 2004.

[38] P. Favaro and T. Papadhimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *Proc. CVPR*, 2012.

[39] M. Fischler and R. Bolles. Random sample consensus: Aparadigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24(6):381–395, 1981.

[40] J. Fox. *Nonparametric regression. An R and S-Plus companion to applied regression (Appendix)*. Thousand Oaks, CA, 2002.

[41] J. Fox. *Robust regression. An R and S-Plus companion to applied regression (Appendix)*. Thousand Oaks, CA, 2002.

[42] A. Fresnel. *The Wave Theory of Light*. American Book Company, 1819.

[43] J. Friedman. J. friedman. *International Journal of Forecasting*, 28(3):722–738, 2012.

[44] G. Fyffe, X. Yu, and P. Devebec. Single-shot photometric stereo by spectral multiplexing. *ICCP*, 2011.

[45] A. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. ICCV*, 2003.

[46] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Proc. ICCV*, October 2005.

[47] M. Grossberg and S. Nayar. Determining the camera response from images: What is knowable? *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(11):1455–1467, 2003.

[48] Y. Han, J. Lee, and I. Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. *Proc. ICIP*, 2013.

[49] W. Härdle. *Applied Nonparametric Regression*. Econometric Society Monographs, 1992.

[50] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[51] H. Hayakawa. Photometric stereo under a light souce with arbitary motion. *Journal of the Optical Soceity of America*, 11(11):3079–3089, 1994.

[52] S. Herbort, D. Schugk, and C. Wöhler. 3d reconstruction of interreflection-affected surface concavities using photometric stereo. *Proc. VISAPP*, 2013.

[53] C. Hernández, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):548–554, 2007.

[54] C. Hernández, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):548–554, 2008.

[55] C. Hernández, G. Vogiatzis, and R. Cipolla. Overcoming shadows in 3-source photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(2):419–426, 2011.

[56] F. Hernández-Rodríguez and M. Castelán. In search of an optimal set of images for uncalibrated photometric stereo. *Proc. Mexican International Conference on Artificial Intelligence*, 2008.

[57] A. Hertzmann and S. Seitz. Example-based photometric stereo: shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1254–1264, 2005.

[58] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *Proc. CVPR*, 2010.

[59] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. *Proc. ICIP*, 2009.

[60] J. Ho, B. Funt, and M. Drew. Separating a color signal into illumination and surface reflectance components: Theorey and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(10):966–977, 1990.

[61] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa. Robust photometric stereo using sparse regression. In *Proc. CVPR*, 2012.

[62] B. Kim and P. Depth. Shape from shading using the photometric stereo method. *CVGIP: Image Understanding*, 54(3):416–427, 1991.

[63] S. Koppal and S. Narasimhan. Novel depth cues from uncalibrated near-field lighting. *Proc. ICIP*, 2007.

[64] N. Kurachi. *The Magic of Computer Graphics*. CRC Press, 2011.

[65] E. Lafortune, S.-C. Foo, K. Torrance, and D. Greenberg. Non-linear approximation of reflectance functions. In *Proc. ACM SIGGRAPH*, 1997.

[66] J. Lee, B. Shi, Y. Matsushita, L. Kweon, and K. Ikeuchi. Radiometric calibration by transform invariant low-rank structure. *Proc. CVPR*, 2011.

[67] M. Liao, X. Huang, and R. Yang. Interreflection removal for photometric stereo by using spectrum-dependent albedo. In *Proc. CVPR*, 2011.

[68] S. Lin, J. Gu, S. Yamazaki, and H. Shum. Radiometric calibration from a single image. *Proc. CVPR*, 2004.

[69] S. Lin and S. Lee. A representation of specular appearance. *In Proc. ICCV*, 1999.

[70] S. Lin and L. Zhang. Determining the radiometric response function from a single grayscale image. *Proc. CVPR*, 2005.

[71] G. G. Lorentz. *Bernstein Polynomials*. Chelsea Publishing Company, New York, 1986.

[72] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato. Uncalibrated photometric stereo for unknown isotropic reflectance. *Proc. CVPR*, 2013.

[73] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: reconstructing specular surfaces using color. In *Proc. CVPR*, 2005.

[74] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. on Graph.*, 22(3):759–769, 2003.

[75] R. Mecca, G. Rosman, R. Kimmel, and A. Bruckstein. Perspective photometric stereo with shadows. In *Proc. of 4th International Conference on Scale Space and Variational Methods in Computer Vision*, 2013.

[76] D. Miyazaki, K. Hara, and K. Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2):229–242, 2010.

[77] W. Mongkulmann, T. Okabe, and Y. Sato. Photometric stereo with auto-radiometric calibration. *In Proc. of IEEE International Conference on Computer Vision Workshops*, 2011.

[78] R. Montes and C. Urena. An overview of brdf models. Technical report, LSI-2012-001 en Digibug Coleccion: TIC167 - Articulos, 2012.

[79] Y. Mukaigawa, Y. Ishii, and T. Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Soceity of America*, 24(10):3326–3334, 2007.

[80] S. Nayar, K. Ikeuchi, and T. Kanade. Proc. ICIP. *Shape from Interreflections*, 1991.

[81] R. Neal. *Bayesian Learning for Neural Networks*. New York: Springer-Verlag, 1996.

[82] T. Okabe, I. Sato, and Y. Sato. Attached shadow coding: Estimating surface nromals from shadows under unknown reflectance and lighting condition. *Proc. ICIP*, 2009.

[83] M. Oren and S. Nayar. Generalization of lambert's reflectance model. In *In Proc. of the 21st annual conference on Computer graphics and interactive tecniques*, 1994.

[84] T. Papadhimitri and P. Favaro. A new perspective on uncalibrated photometric stereo. In *Proc. CVPR*, 2013.

[85] J. Park, S. Sinha, Y. Matsushita, Y. Tai, and I. Kweon. Multiview photometric stereo using planar mesh parameterization. *Proc. ICIP*, 2013.

[86] C. H. Reinsch. Smoothing by spline function. *Numersche Mathematik*, 10:177–183, 1967.

[87] F. Romeiro and T. Zickler. Blind reflectometry. In *Proc. ECCV*, 2010.

[88] P. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880, 1984.

[89] I. Sato, T. Okabe, Q. Yu, and Y. Sato. Shape reconstruction based on similarity in radiance changes under varying illumination. In *Proc. ICCV*, 2007.

[90] Y. Sato and K. Ikeuchi. Reflectance analysis under solar illumination. *Proc. IEEE Worksho on Physics-Based Modeling and Computer Vision*, 1995.

[91] L. Scales. *Introduction to non-linear optimization*. Springer-Verlag New York. Inc, 1985.

[92] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *proc. of Workshop on Stereo and Multi-Baseline Vision*. IEEE, 2001.

[93] C. Schlick. *A fast alternative to phong's specular model*. Eds Paul Heckbert, Academic Press, 1994.

[94] S. Seitz, Y. Matsushita, and K. Kutulakos. A theory of inverse light transport. *Proc. ICIP*, 2005.

[95] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *proc. of Computer Vision and Pattern Recognition*. IEEE, 2006.

[96] S. Shafer. Using color to seperate reflection components. *COLOR research and application*, 10(4):210–218, 1985.

[97] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. *Proc. CVPR*, 2009.

[98] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Proc. CVPR*, 2010.

[99] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi. A biquadratic reflectance model for radiometric image analysis. In *Proc. CVPR*, 2012.

[100] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi. Elevation angle from reflectance monotonicity. In *Proc. ECCV*, 2012.

[101] W. M. Silver. *Determining shape and reflectance using multiple images*. Master's thesis, MIT, 1980.

[102] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *Proc. ACM SIGGRAPH*, 2006.

[103] J. Sun, M. Smith, L. Smith, and A. Farooq. Sampling light field for photometric stereo. *International Journal of Computer Theory and Engineering*, 5(1):14–18, 2013.

[104] K. Sunkavalli, T. Zickler, and H. Pfister. Visibility subspaces: Uncalibrated photometric stereo with shadows. *Proc. ECCV*, 2010.

[105] K. Sunkavalli, T. Zickler, and H. Pfister. Visibility subspaces: Uncalibrated photometric stereo with shadows. In *Proc. ECCV*, 2010.

[106] P. Tan, S. P. Mallick, L. Quan, D. Kriegman, and T. Zickler. Isotropy, reciprocity and the generalized bas-relief ambiguity. In *Proc. CVPR*, 2007.

[107] P. Tan, L. Quan, and T. Zickler. The geometry of reflectance symmetries. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2506–2520, 2011.

[108] P. Tan and T. Zickler. A projective framework for radiometric image analysis. *Proc. CVPR*, 2009.

[109] K. Tang, C. Tang, and T. Wong. Dense photometric stereo using tensorial belief propagation. In *Proc. CVPR*, 2005.

[110] A. Tankus and N. Kiryati. Photometric stereo under perspective projection. *Proc. ICIP*, 2005.

[111] M. Tipping. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learns. Res.*, 1:211–244, 2001.

[112] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1112, 1967.

[113] D. Tylavsky and G. Sohie. Generalization of the matrix inversion lemma. In *Proceedings of the IEEE*, volume 74, pages 1050–1052, 1986.

[114] F. Verbiest and L. V. Gool. Photometric stereo with coherent outlier handling and confidence estimation. In *Proc. CVPR*, 2008.

[115] E. Vieth. Fitting piecewise linear regression functions to biological responses. *Journal of Applied Physiology*, 67(1):390–396, 1989.

[116] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Propoviá, S. Rusinkiewicz, and W. Mtusik. Dynamic shape capture using multi-view photometric stereo. *SIGGRAPHASIA*, 2009.

[117] J. Wang. *Shape restricted nonparametric regression with Bernstein polynomials*. PhD thesis, North Carolina State University, 2012.

[118] G. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics*, 26(2):265–272, 1992.

[119] D. Wipf and S. Nagarajan. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):317–329, 2010.

[120] D. Wipf, B. D. Rao, and S. Nagarajan. Latent variable bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9), 2011.

[121] D. P. Wipf and S. Nagarajan. A new view of automatic relevance determination. *Adv. Neural Inf. Process. Syst.*, 20, 2008.

[122] P. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Engg*, 19(1):139–144, 1980.

[123] C. Wu, Y. Liu, and Q. Dai. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *IEEE Trans. on Visualization and Computer Graphics*, 17(8):1082–1095, 2011.

[124] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. ACCV*, 2010.

[125] T. Wu and C. Tang. Dense photometric stereo using a mirror sphere and graph cut. In *Proc. CVPR*, 2005.

[126] T. Wu and C. Tang. Photometric stereo via expectation maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010.

[127] Z. Wu and P. Tan. Calibrating photometric stereo by holistic reflectance symmetry analysis. *Proc. CVPR*, 2013.

[128] Y. Yoshiyasu and N. Yamazaki. Topology-adaptive multi-view photometric stereo. *Proc. CVPR*, 2011.

[129] C. Yu, Y. Seo, and S. Lee. Photometric stereo from maximum feasible lambertian reflections. In *Proc. ECCV*, 2010.

[130] L. yu, S. Yeung, Y. Tai, D. Terzopoulos, and T. Chan. Outdoor photometric stereo. *ICCP*, 2013.

[131] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. *Proc. CVPR*, 1997.

[132] R. Zhang, P. Tsai, J. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.

[133] Z. Zhou and P. Tan. Ring-light photometric stereo. *Proc. ECCV*, 2010.

[134] Z. Zhou, Z. Wu, and P. Tan. Multi-view photometric stereo with spatially varying isotropic surfaces. *Proc. CVPR*, 2013.

[135] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 2002.