Title	Lighting Analysis of Photometric Stereo for High-fidelity Shape Reconstruction
Author(s)	郭, 亨
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/88151
rights	
Note	

Lighting Analysis of Photometric Stereo for High-fidelity Shape Reconstruction

Heng Guo

To my parents.

List of Publications

A. Journal Paper

1. Heng Guo, Zhipeng Mo, Boxin Shi, Feng Lu, Sai Kit Yeung, Ping Tan, and Yasuyuki Matsushita: "Patch-based Uncalibrated Photometric Stereo under Natural Illumination," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (Sep. 2021).

2. Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita: "Multispectral Photometric Stereo for Spatially-Varying Spectral Reflectances," International Journal of Computer Vision (IJCV) (Minor Revision).

3. Heng Guo, Shuaicheng Liu, Shuyuan Zhu, Heng Tao Shen, and Bing Zeng: "View-consistent Meshflow for Stereoscopic Video Stabilization," IEEE Transactions on Computational Imaging (TCI) (Aug. 2018).

4. Heng Guo, Shuaicheng Liu, Tong He, Shuyuan Zhu, Bing Zeng, and Moncef Gabbouj: "Joint Video Stitching and Stabilization from Moving Cameras," IEEE Transactions on Image Processing (TIP) (Sep. 2016).

B. International Conference Papers

1. Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita: "Multispectral Photometric Stereo for Spatially-Varying Spectral Reflectances: A Well-posed Problem?" In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Jun. 2021).

2. Heng Guo, Shuaicheng Liu, Shuyuan Zhu, and Bing Zeng: "Joint Bundled Camera Paths for Stereoscopic Video Stabilization." In Proceedings of IEEE International Conference on Image Processing (ICIP) (Sep. 2016)

C. Domestic Conference Paper

1. Heng Guo, Boxin Shi, Michael Waechter, Yasuyuki Matsushita: "Self-calibrating Near-light Photometric Stereo under Anisotropic Light Emission," 画像の認識・理解シンポジウム（MIRU）(2020 年 8 月).

D. Awards

1. Heng Guo, Boxin Shi, Michael Waechter, Yasuyuki Matsushita: MIRU 学生優秀賞，第 23 回画像の認識・理解シンポジウム（MIRU）(2020 年 8 月)

Acknowlegements

First, I would like to express my greatest gratitude to my supervisor, Prof. Yasuyuki Matsushita, for his continuous support, thoughtful comments, and encouragement throughout my PhD study. Prof. Yasuyuki Matsushita's supervision helped me derive ideas, schedule research plans, make presentations and write academic papers. During the stressful period in 2020 when I had no research achievements, Prof. Matsushita gave me full support and helped me walk through the difficulties. His wisdom and kindness will always inspire me.

I am deeply grateful to Prof. Boxin Shi who introduces photometric stereo to me. He is always enlightening and supportive throughout my Ph.D training, and his insights and suggestions help me a lot in my research projects.

I was lucky to work with my collaborators, and would like to thank them for their invaluable advice and help: Prof. Boxin Shi for all my three projects (Chapter 2, Chapter 3, and Chapter 4); Dr. Michael Waechter for my first projects (Chapter 2); Dr. Zhipeng Mo, Prof. Feng Lu, Prof. Sai-Kit Yeung, and Prof. Ping Tan for my second projects (Chapter 3); Prof. Fumio Okura, Prof. Takuya Funatomi, and Prof. Yasuhiro Mukaigawa for my third projects (Chapter 4);

I would like to show my sincere thanks to my colleagues and friends: Mr. Xu Cao, Dr. Hiroaki Santo, Mr. Feiran Li, Dr. Zhipeng Mo, Dr. Guanying Chen, Mr. Zhuoyu Yang, Mr. Kenji Enomoto for their infectious enthusiasm in helping me in both studying and living. I really enjoy the collaborations and insightful discussions with them. Besides, I would like to acknowledge Dr. Jinglei Shi, Mr. Junming Fan, Mr. Shihong Liu, Mr. Shuhao Tao, Mr. AshiRiga, Mr. Junfeng Zhou for their accompanies and

supports during my doctoral study.

I would like to express my gratitude to Ms. Mihoko Kaneda and the Graduate School of Information Science and Technology officers for their support in my daily life at Osaka University.

Last, I want to thank my parents and beloved one Dan Wang for their unconditional love, support, and trust.

Abstract of thesis entitled

"Lighting Analysis of Photometric Stereo for High-fidelity Shape Reconstruction"

Submitted by

Heng Guo

for the degree of Doctor of Information Science and Technology

at Osaka University

in December, 2021

High-fidelity 3D shape reconstruction aims to infer a scene's detailed geometry from one or several images, which is desired in various applications, including virtual reality (VR), digital heritage preservation (e-Heritage), image re-lighting, and re-texturing. Photometric stereo is known to be effective for detailed shape recovery. Given image observations captured from a fixed camera under varying illuminations, photometric stereo estimates the orientations of a surface, *i.e.*, surface normals. However, classical photometric stereo techniques assume the lighting condition to be a single distant light with a calibrated lighting direction. Also, the image observations under varying illuminations are captured via multiple shots. The above lighting assumption requires a large space to place light sources sufficiently far away from a target scene, an additional effort for lighting direction calibration, a darkroom to capture images under single illumination, and a static target object during the capturing. It is important to develop photometric stereo methods with analyzing more practical lighting conditions as it enables applications including accurate shape estimation under the illumination of practical near LEDs, detailed shape recovery in the natural environment, and dynamic shape reconstruction. To achieve that, this thesis tackles three practical lighting conditions for photometric stereo: Anisotropic near-point light, uncalibrated natural light, and multispectral light.

The first part of this thesis achieves photometric stereo under near point light,

relaxing the infinitely distant light setting. Real-world point light sources (*e.g.* LEDs) have anisotropic radiant intensity distribution (RID), *i.e.*, the radiant intensity is varying from emission angle. We propose a spherical harmonic-based RID model to represent the real-world anisotropic light emission, which shows high representation power for diverse RIDs. We further propose a self-calibrating near-light photometric stereo method that formulates a linear system to estimate the RID, surface normal, and depth simultaneously. By accurately modeling the near point light, we show our method outperforms existing approaches on surface shape recovery under practical near light illuminations.

The second part of this thesis achieves photometric stereo under uncalibrated natural light, removing the requirement of the darkroom and light calibration. We take image observations under unknown natural illuminations and output the surface normal map. To achieve that, we introduce an equivalent directional lighting model to estimate the local surface normal up to an orthogonal ambiguity. Then we propose a graph-based patch merging method to solve the per-patch ambiguities. Compared to existing methods on uncalibrated natural illumination, we obtain more accurate results without requiring a fine initial guess of the target shape.

The third part of this thesis considers the problem of single-shot shape recovery by photometric stereo under multispectral light, also called multispectral photometric stereo (MPS). It relaxes the requirement of multiple shots to stack images under varying illuminations, which assumes a target object to be static during the data capturing. In MPS, observations are stacked into different spectral bands of a single-shot multispectral image. In this way, MPS can recover shape from a single image and further achieve dynamic shape reconstruction from a multispectral video sequence. However, solving surface shape from MPS is known to be ill-posed. We show that such a problem can be made well-posed for surfaces with uniform chromaticity but spatially-varying albedos based on our proposed formulation. We further introduce linear bases extracted from a spectral reflectance database and formulate shape from MPS under spatially-varying chromaticities and albedos as a well-posed problem.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1   Motivation

3D reconstruction from one or several 2D images is a fundamental inverse problem in computer vision. The recoveries of the shape and reflectance of real objects enable applications in a wide variety of fields. For example, virtual shopping has found success across retail industries due to the COVID-19 pandemic. By extracting products' 3D shape and reflectance information, customs can access and choose goods online without going to physical stores. On the other hand, a great number of cultural heritage and arts have been in danger of destruction and extinguishment over the past years. With 3D reconstruction, valuable heritage and cultural relics can be preserved in digital archives and passed down to the future generation.

There are roughly two categories of approaches to solve the 3D reconstruction. Geometric-based methods such as structure from motion [4] and structure light [5] recover multi-view coarse shape based on the triangulation of matching points between images. On the other hand, photometric-based approaches such as shape from shading [6] and photometric stereo [7] give detailed single view shape based on the shading variation. In this thesis, we focus on the technique of photometric stereo to achieve high-fidelity 3D reconstruction.

As shown in Fig. 1.1, classical photometric stereo, originally proposed by Wood-
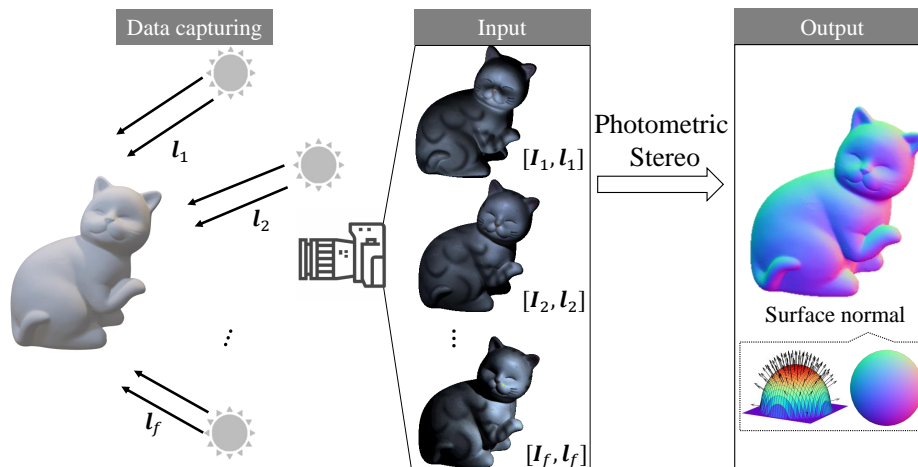
Fig. 1.1 Given multiple images of a static object captured under different light directions, photometric stereo can estimate object's surface normal map, which is visualized by a pseudo color map.

ham [7] and Silver [8], captures image observations from a fixed camera under varying lighting directions. Taking the image observations and the corresponding lighting directions as input, photometric stereo aims at recovering the surface normal map of an object. A surface normal is a unit vector that is perpendicular to the tangent plane to the surface, encoding the detailed shape information of the surface. Conventionally, surface normal map is visualized by a pseudo color map. As highlighted in the dashed box of Fig. 1.1, we show the color mapping of surface normal vectors on a hemisphere.

Photometric stereo is effective for detailed shape reconstruction. However, the light setting of photometric stereo is restrictive, limiting its application in real-world scenarios. As shown in Table 1.1, classical photometric stereo [7, 8] assumes the illumination to be distant light with uniform emission. The lighting direction and intensity are assumed to be known via calibration. The image observations under varying illuminations are captured via multiple shots at different timestamps. For each image observation, the scene is assumed to contain a single illumination. These lighting assumptions restrict photometric stereo to a laboratory environment, as shown in Fig. 1.2(a). In the following, we will discuss the detail for each light assumption and show how it affects photometric stereo to be applied in the wild environment.

Table 1.1 Lighting analysis of photometric stereo (PS). The marker ✓ and × label whether the lighting assumptions shown in the header are required or not in classical photometric stereo [8] and each chapter of this thesis.

Method	Distant light with uniform emission	Light calibration		Multiple shots to record varying lights	Single illumination
		Position/direction	Intensity		
Classical PS [8]	✓	✓	✓	✓	✓
Chapter 2[a]	×	✓	×	✓	✓
Chapter 3[b]	✓	×	×	✓	×
Chapter 4[c]	✓	✓	×	×	✓

[a] Photometric stereo under anisotropic near-point light
[b] Photometric stereo under uncalibrated natural light
[c] One-shot photometric stereo under multispectral light

Distant light with uniform emission   The distant light model assumes the point light source is infinitely far from the target scene so that the same lighting direction and intensity illuminate all the scene points of a target surface. However, this can only be approximately realized by finding a large space to place light sources sufficiently far from a small target object (*e.g.* cat in Fig. 1.2(a)). In such a case, the irradiance received by the surface could be low due to the light attenuation, increasing the noise level of captured images. Besides, the uniform emission requires the point light source equally emitting lights at varying emission angles, which is unpractical for real-world LEDs.

Light calibration   The lighting directions and intensities are assumed to be known in classical photometric stereo. To calibrate the lighting direction and intensities, additional calibration objects are required, such as the mirror ball [9] shown in Fig. 1.2(a) for measuring the lighting direction, a uniform Lambertian board (*e.g.*, Macbeth white balance chart [9]) for calibrating the lighting intensity. Therefore, the light calibration needs additional efforts and devices, which is not friendly for non-experts. Besides, as light calibration itself is an ongoing research problem [10], the calibration process could cause errors and propagate to the photometric stereo.

Multi-shot to record varying lights   Photometric stereo requires image observations under varying illuminations as inputs. This image sequence is recorded via

(a) Laboratory environment   (b) Natural environment
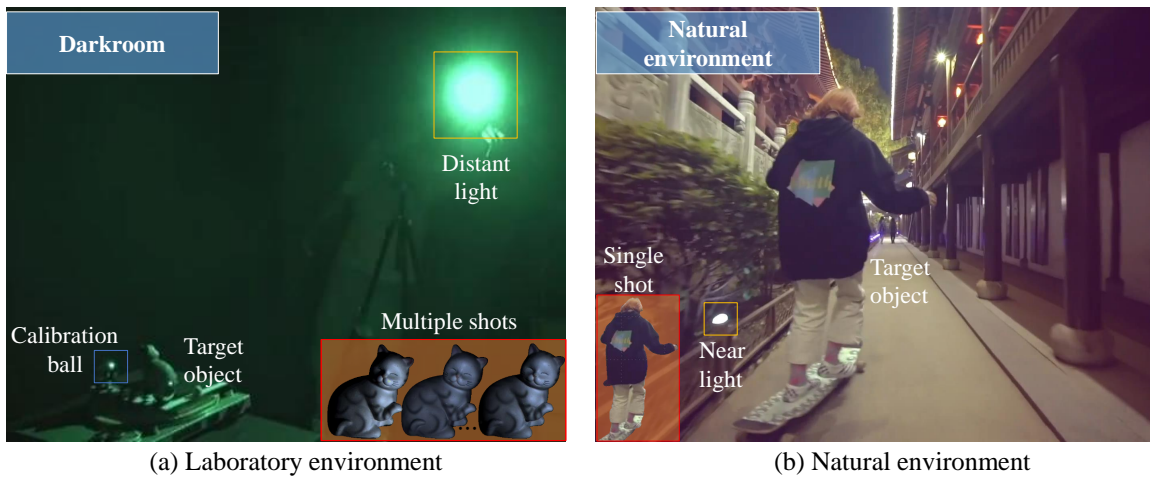
Fig. 1.2 Our motivation is extending photometric stereo from laboratory setup (**left**) to practical wild environment (**right**).

multiple shots at different timestamps, as shown in Fig. 1.2(a). During the data capturing, a target object needs to keep static w.r.t. the camera location, limiting the application of photometric stereo on dynamic scenes.

Single illumination   The single illumination requires an input image observation recording a scene only illuminated by a single light source. To achieve that, a darkroom shown in Fig. 1.2(a) is required to block illuminations from the environment. Restricted by the single illumination assumption, photometric stereo cannot walk out of the darkroom to recover a detailed shape of an object in the natural environment.

This thesis aims to bring photometric stereo out of the laboratory light setting shown in Fig. 1.2(a) and make photometric stereo capable of dynamic shape recovery. As shown on the right side of Fig. 1.2, we empower photometric stereo to be more practical under the following three aspects:

Near point light with anisotropic emission   Unlike classical photometric stereo assuming distant light, we propose a method in Chapter 2 to improve photometric stereo to be applicable under near-point light such as the street lamps shown in the golden box of Fig. 1.2(b). Besides, we take the anisotropic radiant intensity distribu-

tion (RID) of real-world LEDs into consideration, where the light emission at varying emission directions could be non-uniform. However, the challenge for photometric stereo under anisotropic near-point light is that the lighting directions and intensities cannot be treated as the same for varying scene points. As shown in Fig. 1.2(b), light directions are determined by the relative positions between scene points of the target object (girl on the skateboard) and the near light source, and the lighting intensity is influenced by the light falloff and anisotropic RIDs. Both vary from different positions of the target object.

Uncalibrated natural light Unlike classical photometric stereo assuming each image is only illuminated by a single light source with calibrated/known lighting direction, we propose a method in Chapter 3 to improve photometric stereo to be applicable under uncalibrated natural light, which considers all the light sources in the surrounding environment, as shown in Fig. 1.2(b). In this way, we require no darkroom and recover detailed surface shape from image observations captured in a natural environment. Also, we remove the potential error caused by light calibration and reduce the effort for non-experts. However, the challenge for photometric stereo under uncalibrated natural light is that the illumination for a single image contains all light sources from the surrounding environment. Therefore, the lighting directions and intensities for the surface are both spatially-varying, which is more complex compared to the darkroom setup where the whole surface is illuminated by a single light. Furthermore, as the environment light is assumed to be uncalibrated, the spatially-varying light directions and intensities are unknown, making the problem even challenging.

Single-shot shape recovery Unlike classical photometric stereo stacking image observations under varying illuminations via multiple shots at different timestamps, we propose a photometric stereo method in Chapter 4 to recover shape from a single-shot multispectral image, which can be applied in the scenarios of dynamic shape reconstruction such as the moving skateboard girl shown in Fig. 1.2(b). As classical photometric stereo requires multiple shots, a target object needs to keep static during

the data capturing. By using multispectral light sources and a multispectral camera, image observations can be stacked into a multispectral image via a single shot, *i.e.*, observations at varying light directions are encoded into different spectral bands. By solving photometric stereo on a multispectral video sequence, we can obtain the dynamic shape recovery. However, the challenge for photometric stereo under multispectral light, also called multispectral photometric stereo, is that the problem is ill-posed due to the multispectral reflectance.

To summarize, the goal of this thesis is to extend photometric stereo from laboratory setup to practical wild environment. We approach this goal by solving the challenges in photometric stereo under practical anisotropic near-point light and uncalibrated natural light, and making multispectral photometric stereo to be well-posed for single-shot shape recovery.

1.2 Contributions

The main contributions of this thesis can be summarized as follows:

- A self-calibrating near light photometric stereo method. We introduce a spherical harmonic based model for representing the local anisotropic radiant intensity distribution (RID) of the near point light, which is flexible to model diverse real-world RIDs compared to existing methods. We formulate a linear system with the proposed RID representation model to simultaneously solve the local RID calibration and surface shape recovery. Preliminary results of this research have been published in [11].

- A patch-based uncalibrated photometric stereo method under natural illumination. We introduce an equivalent directional lighting model to solve surface normal in a local patch up to an orthogonal ambiguity. To build connections between patches, we introduce a consistent surface normal clustering method guided by intensity profiles under natural light. Based on the patch connections, we further propose a graph-based patch merging method to solve the per-patch

ambiguities and merge local surface normals to a complete surface normal map.
Preliminary results of this research have been published in [12].

- Well-posed formulations of multispectral photometric stereo (MPS) for single-shot shape recovery. We show that MPS can be formulated to be well-posed under monochromatic surfaces with image cues only. We further introduce linear bases extracted from a spectral reflectance database and formulate MPS as a well-posed problem under more general spatially-varying chromaticities and albedos. Preliminary results of this research have been published in [13].

1.3   Outline of thesis

The remainder of this thesis is organized as follows.

Chapter 2   This chapter addresses the problem of photometric stereo under near-point light with anisotropic emission. We first explain our local radiant intensity distribution (RID) representation model based on spherical harmonic bases, and show its strength over previous RID models on the fitting accuracy of the real-world RIDs. Then we present our self-calibrating near light photometric stereo method to jointly recover the object shape and the anisotropic RID. Specifically, we take image observations, point light positions, and sparse depth of the target object as input and formulate a homogeneous linear system for the RID estimation and the sparse shape recovery. The dense shape is then iteratively estimated by surface normal integration and near-light shading model. With the proposed self-calibrating near-light photometric stereo, we show more accurate shape estimation can be obtained under the illumination of real-world near LEDs. Experimental results on both synthetic and real data demonstrate the effectiveness of our method.

Chapter 3   This chapter addresses the problem of photometric stereo under uncalibrated natural light. We show uncalibrated photometric stereo under environment

light can be solved in a "divide and conquer" manner. First, an equivalent directional lighting model for local surface patches consisting of slowly varying normals is introduced to solve per-patch surface normals up to an arbitrary orthogonal ambiguity. Then we explain our graph-based patch merging method to merge local surface normals to a complete surface normal map. Specifically, we build the patch connections by extracting consistent surface normal pairs via spatial overlaps among patches and intensity profiles. Guided by these connections, the local ambiguities are unified to a global orthogonal one through Markov Random Field optimization and rotation averaging. After applying the integrability constraint, we show our solution for surface normal contains only a binary ambiguity, which could be easily removed. Extensive experiments using both synthetic and real-world datasets show our method provides even comparable results to calibrated methods.

Chapter 4 This chapter addresses the problem of single-shot shape recovery from photometric stereo under multispectral light. We first introduce multispectral photometric stereo (MPS) to recover the surface normal of a scene measured under multiple light sources with different wavelengths. While it opens up a capability of a single-shot measurement of surface normal, the problem has been known ill-posed. Then we show the MPS problem becomes well-posed for surfaces with uniform chromaticity but spatially-varying albedos based on our new formulation. Specifically, if at least three (or two) scene points share the same chromaticity, the proposed method uniquely recovers their surface normals with the illumination of no less than four (or five) spectral lights in a closed-form. After that, we further show that MPS under a more general setting, spatially-varying both chromaticities and albedos, can become well-posed if the light spectra and camera spectral sensitivity are calibrated. For this general setting, we derive a unique and closed-form solution for MPS using the linear bases extracted from a spectral reflectance database. Experiments on both synthetic and real captured data with spatially-varying reflectance demonstrate our method's effectiveness and the potential for dynamic shape reconstruction.

Chapter 5   This chapter concludes the thesis by summarizing the algorithms developed in this dissertation and discussing the future work of photometric stereo under practical illumination.

Chapter 2

Photometric Stereo under Anisotropic Near-point Light

2.1 Introduction

3D reconstruction from photos is a topic almost as old as photography itself [14]. In the last decades, it has become an attractive research field within computer vision and has become important in countless industrial as well as end-user applications. Besides geometric approaches such as structure from motion and multi-view stereo, photometric approaches have also been investigated [7, 8] because of their capability of recovering detailed shape. The strength of photometric stereo is its ability to recover surface orientations of very high resolution as well as albedos.

However, in contrast to geometric approaches, photometric stereo typically requires much more controlled setups: A fixed camera viewpoint, fixed focal length and exposure, and images. It further requires taking multiple pictures under varying lighting directions. Early photometric stereo approaches [15, 16] assumed the light to be from a distant light source (*e.g.* sunlight), yielding parallel, uniform light, as shown in Fig. 2.1(a). This is however hard to perfectly achieve in real experiments using LEDs as light sources. Since setups whose light is not perfectly distant suffer from the effects of light fall-off and spatially varying lighting directions, algorithms that
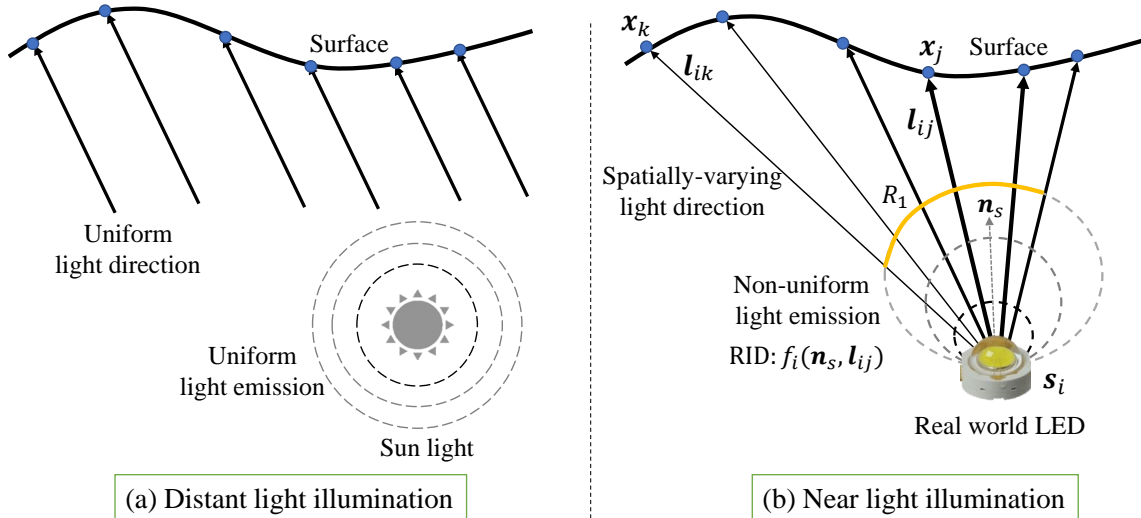
Fig. 2.1 Distant light illumination vs near light illumination. (a) Distant light (*e.g.*, sun light) has uniform light emission at varying angles, the lighting directions for varying scene points are the same. (b) Different scene points have varying lighting directions under the illumination of a near point light located at $\mathbf{s}_i$. The light emissions is modeled by the radiant intensity distribution (RID), as visualized by the equal radiant emission contours shown in dashed line. The RID illuminating a surface is located at a local smooth area R_1 shown in orange.

assume parallel light produce reconstruction errors. Near-light photometric stereo algorithms [17, 18], on the other hand, do account for these effects and thereby allow working with much more compact experiment setups. As shown in Fig. 2.1(b), different scene points receive varying lighting directions and intensities determined by the relative position between LED and the surface.

To achieve accurate shape recovery, existing near-light photometric stereo methods require precise calibration of involved point light sources, which includes their 3D positions and their radiant intensity distributions (RIDs) [19, 20]. The RID represents the radiant intensity of a light source under different emitted directions. As an anisotropic RID example shown in Fig. 2.1(b), the radiant at emitted directions $\mathbf{l}_{ik}$ attenuate quicker than that of $\mathbf{l}_{ij}$. Usually, different point light sources in the real world have different RIDs. To address the attenuation caused by RID when estimating an object shape, usually a calibration object is required to measure the position and intensity distribution of point light sources represented by a parametric model, before

applying near-light photometric stereo algorithms. Existing calibration approaches have some disadvantages:

1. A reference object with known geometry such as a plane with uniform albedo and Lambertian reflectance is required for the RID calibration [19, 20], which is not readily available in many real-world scenarios;

2. The radiance emitted by a point light typically changes over time due to ramp-up time and camera auto-exposure [21], leading to different photometric observations between the RID calibration step and the following image acquisition step;

3. Current RID representation models [19, 20] used in RID calibration and near-light photometric stereo [18, 21] are limited to radially symmetric RIDs, which is not flexible enough to handle a diversity of RIDs in the real world.

To address these problems, we propose a self-calibrating photometric stereo framework in which object shape reconstruction and RID calibration are conducted simultaneously without the use of a reference object. Our key observation is that the observed RID highlighted in orange curve in Fig. 2.1(b) is located in a local region when illuminating an object. We assume this local RID is smooth and can be interpolated by spherical harmonic bases. Based on this RID representation, we aim to estimate the shape of a Lambertian object with given calibrated point light positions and an initial estimate of a sparse depth map for the target object.

This chapter makes the following contributions.

- We propose a near-light photometric stereo method with RID auto-calibration without using a reference object.

- We propose an easy-to-fit RID representation in a local region for real-world anisotropic light emission.

- We show our self-calibrated near-light photometric stereo solution achieves higher accuracy on shape recovery than the state-of-the-art method.

2.2 Related work

Our work is related to point light source calibration and near-light photometric stereo.

Point light source calibration. The calibration of a point light source can be roughly separated into two tasks: light position calibration and the light radiant intensity distribution calibration. The former one has been well studied in recent years. Previous methods [22–25] triangulate multiple light positions from specular highlights on mirror objects with the known shape, such as a sphere or a plane. Recently, Bunteong and Chotikakamthorn *et al.* [26] propose a method that uses specular highlights and shadows to estimate point light positions and their directions. Santo *et al.* [10] calibrate the light positions from the shadow positions of pins onto a diffuse plane. With the small pinhead size, this method obtains higher accuracy in localizing the light source.

On the other hand, few works target the light radiant intensity distribution estimation. To model the RID of point lights, the most direct way is assuming the relative luminous flux in any direction is isotropic. This assumption is reasonable when light is far from the object. Under the near-light setting, the light attenuation caused by anisotropic RID of real-world point lights is non-negligible. Therefore more specific models should be adapted. Park *et al.* [19] and Visentini *et al.* [20] use a set of polynomials to approximate the radial symmetric RIDs. Another option is using an exponential function with cosine as bases to represent the RID. This model is widely used in existing near-light photometric stereo methods [18, 21, 27, 28], since the power parameter can be estimated from the datasheet of the point lights. To extend the representation power of the cosine-power based RID, Moreno *et al.* [29] use the sum of multiple Gaussian or cosine-power basis functions to describe a wide range of RIDs based on the light propagation characteristics. However, when given RID, there is no strategy to determine which basis function should be selected.

Given the RID representation model, how to estimate the parameter in the model becomes a more challenging problem. Existing methods [19, 20] solve the problem

with a planar Lambertian board with uniform albedo. Specifically, Park *et al.* [19] calibrate the point light RID based on the property that the shading on a Lambertian plane illuminated by a point light with radial symmetric RID is bilaterally symmetric. However, the optimized RID is sensitive to the expensive symmetry searching of the projected light pattern on the Lambertian plane. To avoid this problem, Visentini-Scarzanella and Kawasaki [20] find the symmetry axis by inspection of the global irradiance maxima, and a closed-form solution for both position and RID of point lights are also presented in this work. Different from these two methods, which rely on a textureless Lambertian plane and radial symmetric RID assumption, we use spherical harmonic bases to represent a general form of RID and have no requirement of any reference objects.

Near-light photometric stereo. Starting from the work of Iwahori *et al.* [30] and Clark [31], light propagation models used in photometric stereo have been extended from distant and directional lighting to nearby point light sources. Near-light photometric stereo is a non-linear problem since the distance and the anisotropic attenuation of point light sources are coupled with the local surface normals. To solve this non-linear problem, calibrated near-light photometric stereo methods [32–37] estimate the object shape and albedo in an alternative way, *i.e.*, estimating the surface normal and albedo from fixed lighting first and then integrating the depth from normal and updating the lighting from the current depth. The other way to solve the non-linearity in near-light photometric stereo is using variational methods [17, 18, 38, 39]. Such methods use image ratios to formulate the problem into a quasi-linear partial differential equation. More recently, Quéau *et al.* [18] review the existing methods and propose a provably convergent method of near-light photometric stereo.

In order to avoid the point light calibration, Papadhimitri *et al.* proposed an uncalibrated near-light photometric stereo method [40]. However, the point light RID is assumed as isotropic in this work. To make the problem more tractable, semi-calibrated photometric stereo approaches are proposed where light positions or directions are given, but the RIDs remain unknown. This notation was first proposed by

Cho *et al.* [41] under distant light setting, and then extended into near point light setup [18, 21, 27, 28]. However, prior information about the point light angular characteristics and specific reference objects are needed to recover the RID in [18, 21, 28]. In order to release this requirement, self-calibrating near-light photometric stereo tries to recover object shape and RID for each point light automatically. Recently, Xie *et al.* [42] propose an auto-calibration method to estimate depth map and surface albedo under isotropic near point lights. Different from their method, our self-calibrating near-light photometric stereo method is able to handle more general anisotropic RIDs and obtain accurate RID calibration and object shape simultaneously.

2.3 Near-light image formation model

Before presenting our self-calibrating near light photometric stereo method under anisotropic light emission, we first introduce the image formation model under the near-light illumination and derive the problem setting in this chapter.

We assume a Lambertian object is captured by using a fixed perspective camera with a linear radiometric response and illuminated by anisotropic point light sources. In the following, we use i and j as indices for point light sources (or their corresponding images) and pixels (or their corresponding surface points), respectively.

Given the above assumptions, a measured irradiance $m_{ij} \in \mathbb{R}$ can be written as

$$m_{ij} = e_{ij}\mathbf{b}_j^\top \mathbf{l}_{ij}, \tag{2.1}$$

where $\mathbf{b}_j = \rho_j \mathbf{n}_j \in \mathbb{R}^3$ is an albedo-scaled surface normal vector with albedo defined as $\rho_j \in \mathbb{R}_+$ and surface normal denoted as $\mathbf{n}_j \in S^2 \subset \mathbb{R}^3$, $e_{ij} \in \mathbb{R}_+$ is the irradiance emitted from light i on a surface corresponds to pixel j, and $\mathbf{l}_{ij} \in S^2 \subset \mathbb{R}^3$ is a unit lighting direction vector. Further, we denote a 3D point light location as $\mathbf{s}_i \in \mathbb{R}^3$ and surface point position as $\mathbf{x}_j \in \mathbb{R}^3$. We then have

$$\mathbf{l}_{ij} = \frac{\mathbf{s}_i - \mathbf{x}_j}{\|\mathbf{s}_i - \mathbf{x}_j\|_2}, \tag{2.2}$$

$$e_{ij} = \frac{f_{ij}}{d_{ij}^2} = \frac{f_i(\mathbf{l}_{ij}, \boldsymbol{\omega}_s^i)}{\|\mathbf{s}_i - \mathbf{x}_j\|_2^2}, \tag{2.3}$$

where $f_i : S^2 \times S^2 \to \mathbb{R}$ is the i-th radiant intensity distribution function (RIDF), f_{ij} is the RID from point light $\mathbf{s}_i$ to scene point $\mathbf{x}_j$, $\boldsymbol{\omega}_s^i$ denotes the point light principal direction (visualized in Fig. 2.1)(b), and $1/d_{ij}^2$ accounts for the light fall-off due to the distance d_{ij} between j-th scene point and i-th light. Based on the perspective camera model, the relationship between 3D scene point $\mathbf{x}_j$ and its image coordinates $\mathbf{p}_j$ is

$$\mathbf{x}_j = z_j \mathbf{K}^{-1} \mathbf{p}_j, \tag{2.4}$$

where $\mathbf{K}$ is the camera intrinsic matrix, and z_j is the depth at j-th pixel. Combining Eq. (2.1) to Eq. (2.3), we have

$$\frac{m_{ij} d_{ij}^2}{f_{ij}} = \mathbf{b}_j^\top \mathbf{l}_{ij}. \tag{2.5}$$

In a matrix form for representing all p pixels and q lights, the near light image formation mode in Eq. (2.5) can be written as

$$\text{vec}(\mathbf{M} \odot \mathbf{D} \odot \mathbf{D} \oslash \mathbf{F}) = \mathbf{L}\,\text{vec}(\mathbf{B}^\top), \tag{2.6}$$

where $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{F}$ with the size of $q \times p$ are matrices that contain all m_{ij}, d_{ij}, and f_{ij}, respectively, and $\mathbf{B}$ is a matrix that stacks all albedo-scaled surface normal $\mathbf{b}_j$ horizontally. Operators $\odot$, $\oslash$, and $\text{vec}(\cdot)$ represent element-wise multiplication, element-wise division, and vectorization. Further, $\bar{\mathbf{L}}_j \in \mathbb{R}^{q \times 3}$ stacks all lighting directions on the j-th pixel and $\mathbf{L}$ is formed by the block diagonalization of $\bar{\mathbf{L}}_j$:

$$\mathbf{L} = \begin{pmatrix} \bar{\mathbf{L}}_1 & & & \\ & \bar{\mathbf{L}}_2 & & \\ & & \ddots & \\ & & & \bar{\mathbf{L}}_p \end{pmatrix}. \tag{2.7}$$

Given observations $\mathbf{M}$, the corresponding point light positions $\mathbf{s}_i$, camera parameter $\mathbf{K}$
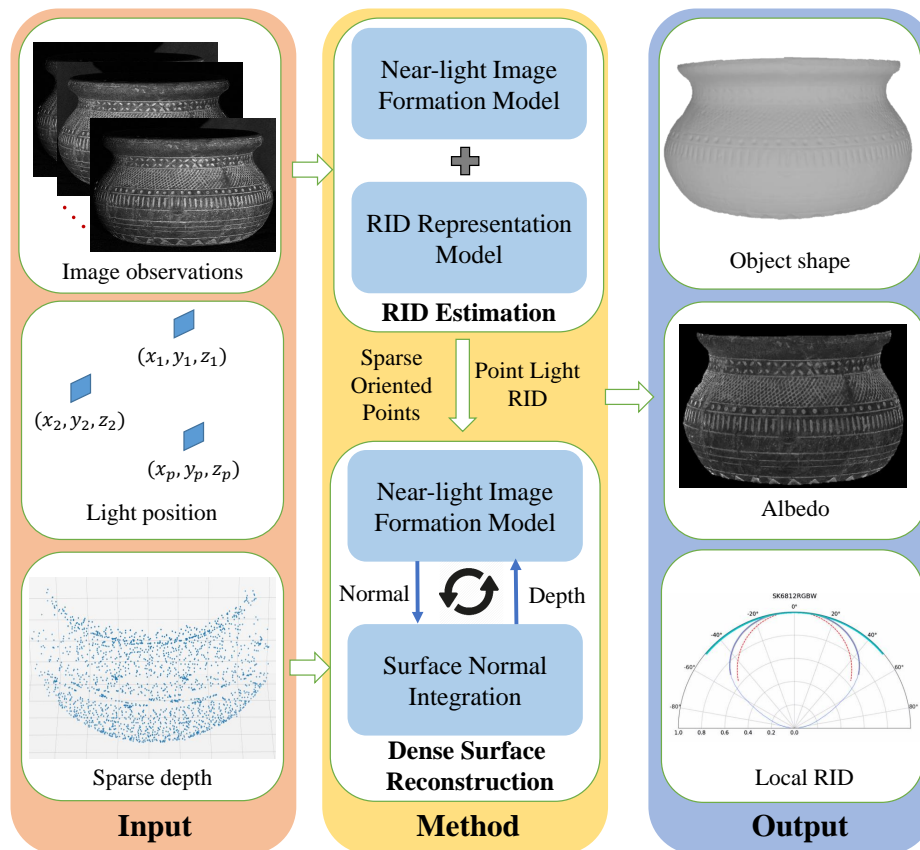
Fig. 2.2 The pipeline of our self-calibrating near-light photometric stereo.

and a sparse depth map $\hat{\mathbf{z}}$ of the object (we assume it is given for now, and introduce how we capture it in the experiment section), our goal is to simultaneously estimate object shape (dense normal and depth) and anisotropic radiant intensity distribution, namely the depth vector $\mathbf{z}$ of all z_j, the normal matrix $\mathbf{N}$ of all $\mathbf{n}_j$, the albedo vector $\boldsymbol{\rho}$ of all ρ_j, and all RIDFs f_i.

2.4 Self-calibrating near-light photometric stereo

The complete pipeline of our method is shown in Fig. 2.2. Our method takes near-light image observations, calibrated LED positions and the sparse depth map of an object as inputs, aiming at recovering the dense object shape, albedo, and the local RIDs for each point light source. There are two modules in our algorithm: Radiant intensity distribution (RID) estimation and dense surface reconstruction, *i.e.*, we first calculate
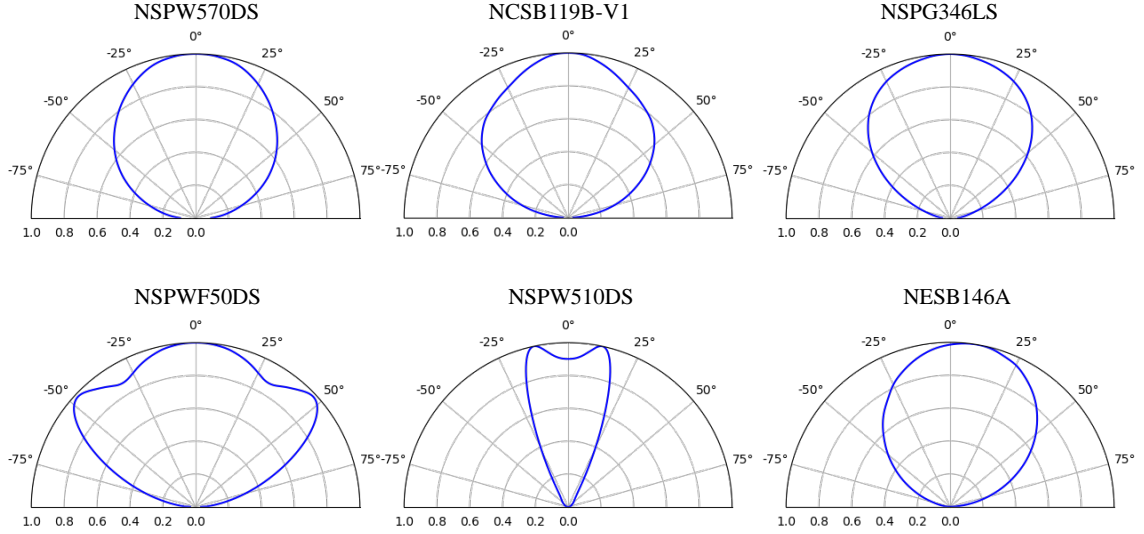
Fig. 2.3 RIDs of real point light sources visualized by polar graphs, where the polar axis represents the radiant intensity. The LEDs are from NACHIA Corp datasheet [1]with product names shown on the top. Notice that the RID of product "NESB146A" is not radially symmetric.

the light RID, then input it to the surface reconstruction process for obtaining dense depth and surface normals. Next, we will introduce the two modules in detail.

2.4.1 Radiant intensity distribution representation

We use RIDF to represent the RID in a parametric way. As visualized in Fig. 2.1(b), the RIDF could be defined with regard to the point light principal direction $\boldsymbol{\omega}_s^i$ and lighting direction $\mathbf{l}_{ij}$, *i.e.*, $f_i(\mathbf{l}_{ij}, \boldsymbol{\omega}_s^i)$. For isotropic point light source, the RIDF is a constant as the light is emitted equally in all directions. However, RIDs for most real-world LEDs are anisotropic, as shown in Fig. 2.3. To model the anisotropic RIDF, existing methods [18, 27] adopt an empirical cosine-power based function, *i.e.*

$$f_i(\mathbf{l}_{ij}, \boldsymbol{\omega}_s^i) = \cos^{\mu_i}(\mathbf{l}_{ij}^\top \boldsymbol{\omega}_s^i), \tag{2.8}$$

where μ_i is a hyper-parameter representing the radient intensity attenuation. The value of μ_i varies from different LED products. As both μ_i and $\boldsymbol{\omega}_s^i$ are unknown, previous methods [18, 27] need to calculate the attenuation parameter μ from the LED
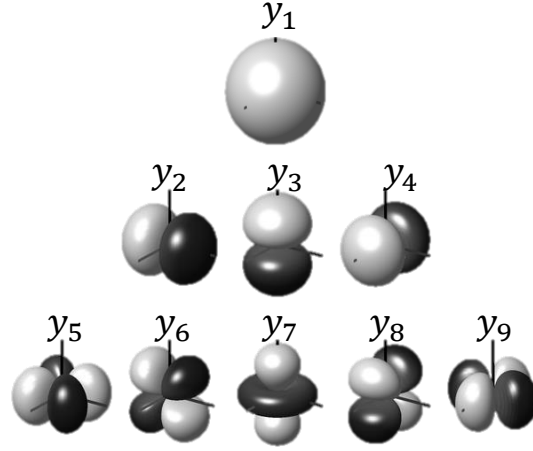
Fig. 2.4 Visualization of spherical harmonic basis function $y_k(\cdot)$ at different orders in the 3D Cartesian coordinates, where white means positive values and black is associated with negative values.

datasheet provided by the manufactures and calibrate the LED principal direction $\boldsymbol{\omega}_s^i$ with a flat Lambertian calibration board. Although the cosine-power based model is simple, it has limited representation power for real RIDs, *e.g.*, the radial asymmetric RID of "NESB146A".

Different from the cosine-power based RID model, our RID model is based on spherical harmonic (SH). As shown in Fig. 2.4, spherical harmonics are a frequency-space basis for representing functions defined over the sphere. They are the spherical analogue of the 1D Fourier series. We observed that RIDF is a function defined in the spherical coordinate system, therefore it is convenient to represent it with the linear combination of a set of SH bases. Besides, as the RIDFs under varying principal direction $\boldsymbol{\omega}_s^i$ are always spherical functions, we can omit the $\boldsymbol{\omega}_s^i$ and take the lighting direction as the only variable in our RIDF, *i.e.*,

$$f_i(\mathbf{l}_{ij}, \boldsymbol{\omega}_s^i) = f_i(\mathbf{l}_{ij}) = \sum_{k=1}^{h} c_k^i y_k(\mathbf{l}_{ij}). \tag{2.9}$$

where c_k^i and $y_k(\cdot)$ denote the coefficient and the basis of the k-th spherical harmonics of i-th point light. Compared to the cosine-power based RID model, we avoid the esti-

---

[1]`https://www.nichia.co.jp/en/product/led.html`

mation of the light principal direction $\boldsymbol{\omega}_s^i$ and the attenuation parameter μ. Besides, as SH basis can approximate any functions in the spherical coordinate system, our RID model is adapt to various RIDs which even do not hold the radial symmetry (*e.g.*, "NESB146A" in Fig. 2.3).

On the other hand, the lighting direction from point light to the object falls into a limited region R_1 as shown in Fig. 2.1(b). For photometric stereo, we mainly care about the radiant intensity distribution in this local region. It is reasonable to assume the RID in this patch is smooth; therefore the local RID can be sufficiently represented by a low-order SH basis. Actually, based on our experiments, we found the top-10 coefficients of SH are enough to represent the local RID.

In a matrix form, Eq. (2.9) can be written as

$$
\text{vec}(\mathbf{F}) = \Big[\mathbf{Y}_1\big|\cdots\big|\mathbf{Y}_p\Big]^\top \text{vec}(\mathbf{C}),
\tag{2.10}
$$

where $\mathbf{C}$ is the matrix containing all coefficient c_k^i and $\mathbf{Y}_j$ is defined as,

$$
\mathbf{Y}_j = \begin{pmatrix} y_1(\bar{\mathbf{L}}_j) & & & \\ & y_2(\bar{\mathbf{L}}_j) & & \\ & & \ddots & \\ & & & y_h(\bar{\mathbf{L}}_j) \end{pmatrix}.
\tag{2.11}
$$

Given the SH-based RID model, we will show in the following section that the unknown SH coefficients $\mathbf{C}$ can be self-calibrated in near-light photometric stereo.

2.4.2 Radiant intensity distribution estimation

In this section, we will show how to determine the albedo scaled-normal map and the RID for each point light. By merging the near-light image formation model Eq. (2.6) and the RID representation model Eq. (2.10), we derive our energy function including

a shading term $\mathcal{E}_s$ and RID term $\mathcal{E}_r$ as follows

$$\underset{\tilde{\mathbf{C}},\mathbf{G},\mathbf{B}}{\operatorname{argmin}} \underbrace{\left\|\mathbf{Q}\text{vec}(\mathbf{G}) - \mathbf{L}\text{vec}(\mathbf{B}^\top)\right\|_2^2}_{\mathcal{E}_s} + \lambda \underbrace{\left\|\text{vec}(\mathbf{G}) - \left[\mathbf{Y}_1|\cdots|\mathbf{Y}_w\right]^\top \text{vec}(\tilde{\mathbf{C}})\right\|_2^2}_{\mathcal{E}_r}, \quad (2.12)$$

where $\mathbf{Q} = \text{diag}(\text{vec}(\mathbf{M} \odot \mathbf{D} \odot \mathbf{D}))$, and $\text{diag}(\cdot)$ denotes the diagonalization operator. λ is the weight to balance the two terms (fixed as 0.027 in our experiment), $\mathbf{G}$ is the inverse RID matrix with its element equal to $1/f_{ij}$ and $\tilde{\mathbf{C}}$ is the spherical harmonic coefficient matrix for the inverse RID. Notice that it is reasonable to assume the inverse RID can be interpolated by SH basis since it is also defined on spherical coordinate. Since we replace $\mathbf{F}$ with inverse RID matrix $\mathbf{G}$, the above energy function can be formulated as a homogeneous linear system:

$$\underbrace{\begin{pmatrix} \mathbf{L} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & [\mathbf{Y_1}|\cdots|\mathbf{Y_w}]^\top \end{pmatrix}}_{\mathbf{A} \in \mathbb{R}^{2p\hat{q} \times (3p + p\hat{q} + h\hat{q})}} \underbrace{\begin{pmatrix} \text{vec}(\mathbf{B}^\top) \\ \text{vec}(\mathbf{G}) \\ \text{vec}(\tilde{\mathbf{C}}) \end{pmatrix}}_{\mathbf{y}} = \mathbf{0}, \quad (2.13)$$

where $p, \hat{q}$ and h represent the number of images, sparse points with known depth and SH bases. To obtain a non-trivial solution from Eq. (2.13), $\mathbf{A}$ should have 1D nullspace, which leads to the following inequality:

$$2p\hat{q} \geq 3p + p\hat{q} + h\hat{q} - 1,$$
$$h \leq p + \frac{1 - 3p}{\hat{q}}. \quad (2.14)$$

By applying singular value decomposition on $\mathbf{A}$, the albedo-scaled surface normal $\mathbf{B}$, inverse RID $\mathbf{G}$ and the SH coefficient $\tilde{\mathbf{C}}$ can be solved up to a scale ambiguity. We then conduct per-pixel normalization on $\mathbf{B}$ to obtain surface normal and albedo. In this way, we achieve near-light photometric stereo with self-calibrating the RID.

Note that in the optimization of Eq. (2.13), the depth should be known so that we can compute the the lighting directions $\mathbf{L}$. However, it is not necessary to know the complete depth map for the purpose of computing the RID coefficient. Therefore, we

only require a sparse depth map $\hat{\mathbf{z}}$ (*e.g.* from a stereo camera) as input to calibrate the RID for each point light and solve the corresponding sparse surface normals $\hat{\mathbf{N}}$. In the next section, we will show how to recover a complete depth map with obtained RIDs and sparse surface normals.

Algorithm 1: Self-calibrating Near-light Photometric Stereo.

Input :
Point light locations $\mathbf{s}$
Sparse depth map $\hat{\mathbf{z}}$
Image observations $\mathbf{M}$
MaxIters T, Threshold τ, λ, η
Solve $\tilde{\mathbf{C}}$, $\hat{\mathbf{B}}$ and $\mathbf{G}$ based on Eq. (2.12);
Estimate the initial complete normal map $\mathbf{N}^{(0)}$ from sparse normal $\hat{\mathbf{N}}$
following [43];
while convergence **do**
   Update depth $\mathbf{z}^{(t)}$ from normal $\mathbf{N}^{(t)}$ by surface normal integration in Eq. (2.16)
   Update the normal $\mathbf{N}^{(t+1)}$ and albedo $\boldsymbol{\rho}^{(t+1)}$ from $\mathbf{z}^{(t)}$ by minimizing the shading
   term $\mathcal{E}_s$ in Eq. (2.12).
   if $\|\mathbf{z}^{(t+1)} - \mathbf{z}^{(t)}\| \leq \tau$ **then**
      break
   end if
end while
Output :
Dense surface normal map $\mathbf{N}$, depth map $\mathbf{z}$, albedo $\boldsymbol{\rho}$;
Coefficients $\tilde{\mathbf{C}}$ of the inverse RIDF for each point lights;

2.4.3   Dense surface reconstruction

In this section, we follow the idea of [43] to first recover the initial complete surface normal map, and then iteratively refine the complete depth and surface normal map guided by the shading cues and normal integration. Following [43], with sparse depth $\hat{\mathbf{z}}$ and sparse normal $\hat{\mathbf{N}}$ from the last section, the complete surface normal field can be calculated on the distinct patches containing such sparse oriented points. We refer to the obtained dense surface normal as the initial surface normal map $\mathbf{N}^{(0)}$ and use it in computing the complete depth map.

As stated in [44], the relationship between depth and surface normal under the

perspective projection can be written as,

$$
\begin{cases}
\nabla \mathbf{z}_l = [\boldsymbol{\psi}, \boldsymbol{\gamma}]^\top, \\
\boldsymbol{\psi}(u, v) = -\dfrac{n_1}{un_1 + vn_2 + f_l n_3}, \\
\boldsymbol{\gamma}(u, v) = -\dfrac{n_2}{un_1 + vn_2 + f_l n_3},
\end{cases}
\tag{2.15}
$$

where $\mathbf{z}_l = \ln \mathbf{z}$ denotes the depth in logarithmic scale, f_l is the camera focal length, u and v represent the 2D image coordinate, $\mathbf{N}(u, v) = [n_1, n_2, n_3]^\top$.

Based on the depth and surface normal relationship, we obtain the complete depth map by surface normal integration [44, 45] with considering the prior of known sparse depth $\hat{\mathbf{z}}$, *i.e.*,

$$
\begin{pmatrix} \eta \mathbf{I} \\ \Delta \end{pmatrix} \mathbf{z}_l = \begin{pmatrix} \eta \hat{\mathbf{z}}_l \\ \frac{\partial \psi}{\partial u} + \frac{\partial \gamma}{\partial v} \end{pmatrix},
\tag{2.16}
$$

where η defines the weight for the sparse depth constraint, Δ refers to a Laplacian operator. With obtained log-scaled depth $\mathbf{z}_l$, we recover the depth map by $\mathbf{z} = e^{\mathbf{z}_l}$.

With initial dense surface normal $\mathbf{N}^{(0)}$ and sparse depth $\hat{\mathbf{z}}$, we recover a complete depth map $\mathbf{z}^{(0)}$ from Eq. (2.16). By minimizing the near-light shading term $\mathcal{E}_s$ shown in Eq. (2.12), we can further obtain a refined surface normal $\mathbf{N}^{(1)}$ with fixed depth map $\mathbf{z}^{(0)}$ and RIDs as input. The above two steps formulate an alternative optimization process summarized in Algorithm. 1. Based on our experiment, accurate surface normal and depth can be obtained up to 5 iterations.

2.5   Experiments on synthetic dataset

We evaluate our method on a synthetic dataset in the setting of anisotropic near-light illumination. We first present a RID representation power comparison between the cosine-power model used in [18] and our spherical harmonic based model. After that, we compare with [18] on the accuracy of surface normal and depth estimation.
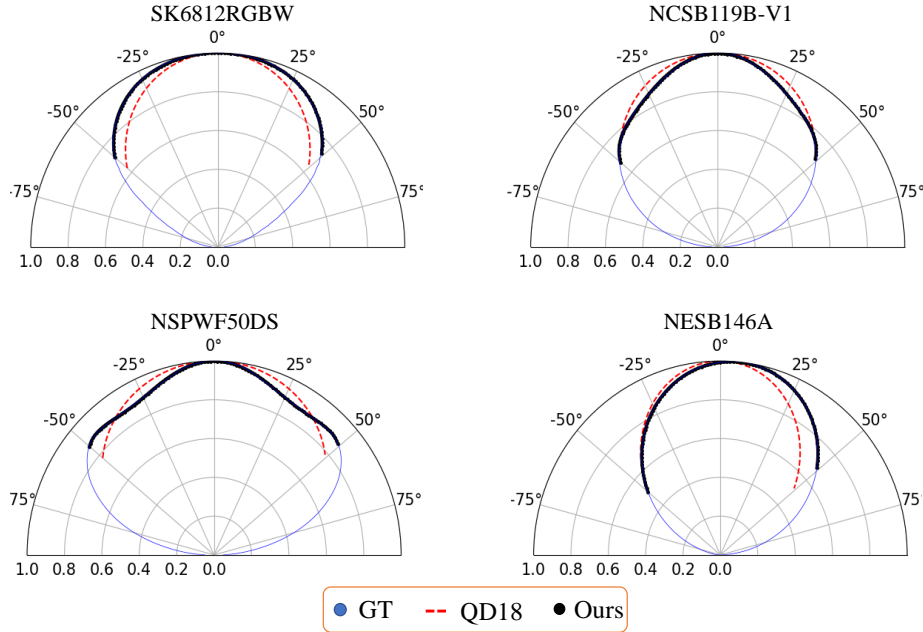
Fig. 2.5 The representation power comparison of different approaches on four real world RIDs. The LED product name of the point lights are shown on the top of each RID plot.

2.5.1 Evaluation on RID representation power

In Fig. 2.5, we fit four real-world RIDs at a local region defined in the range of 50 degrees deviated from the point light principal direction. Under this local region, we show the fitting comparison with the cosine-power based model used in QD18 [18]. The hyper-parameter μ of the cosine-power based model is extracted from the LED datasheet. The number of SH basis used in our RID model is 9. From Fig. 2.5, it is clear that our method gets better RID fitness. We further give a quantitative evaluation of the local RID fitting on real-world RIDs, where the fitting accuracy is defined by the mean squared error between the ground-truth and fitted RIDs. As shown in Fig. 2.6, we collect 204 real-world RIDs from LED manufactures and show the fitting error of our method and the cosine-power based model in log scale. On diverse LED products, our RID model achieves better RID fitness in the local region, demonstrating the representation power of our spherical harmonic based RID model over the previous method.
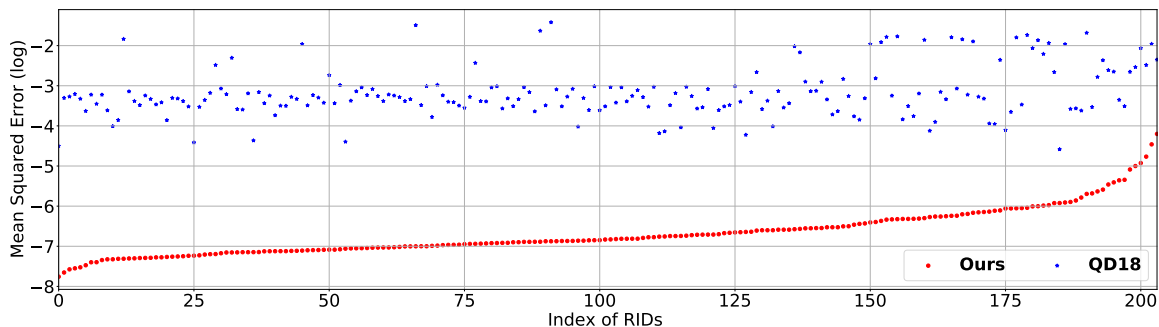
Fig. 2.6 Quantitative comparison of the RID representation power on 204 real RIDs from LED manufactures.

We also evaluate our RID representation model under a different number of bases and varying local RID regions. As shown in Fig. 2.7, for the same range of RID, the fitted RIDs from our method deviate from the ground truth (GT) by decreasing the number of spherical harmonic bases. On the other hand, under the same number of bases, when increasing the local RID region from $-50° \sim 50°$ to $-80° \sim 80°$, the fitting error from our RID model increases as the target RID curve becomes more complex. From a practical point of view, it is rare in photometric stereo that the observed RID covers the range of $-80° \sim 80°$. In such a case, the LED is put pretty close to the surface. However, even for that local RID range, we show that the target RID curve can still be approximated in a relatively high accuracy from our RID model with at least 10 bases.

2.5.2   Comparison on shape estimation

We compare our method with the semi-calibrated photometric stereo algorithm QD18 [18], in which the cosine-based model is used for RID representation. The inputs for QD18 [18] include the principal lighting directions and the hyper-parameter μ for the RID model, and a rough distance between object and camera for generating the initial depth map. As stated in the paper, the optimized shape output is sensitive to this initial depth map. To get a fair comparison, we use our initial depth $\mathbf{z}^{(0)}$ obtained from Sec. 2.4.3 as the initial depth input for QD18 [18].
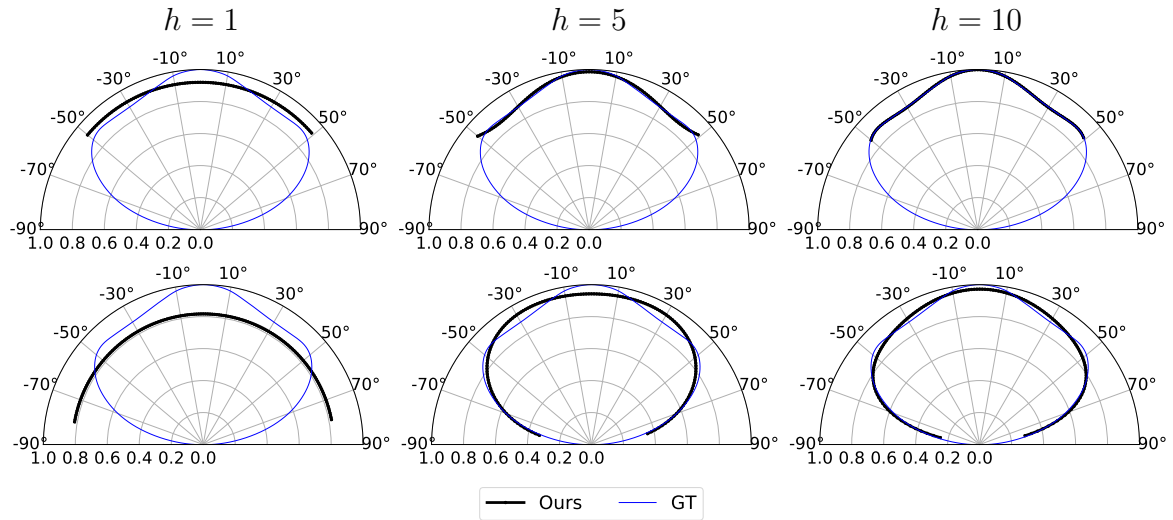
Fig. 2.7 Local RID fitting using different number (*i.e.*, $h = 1, 5, 10$) of spherical harmonic bases. The top and bottom rows show the RID fitting results in the local range of $-50° \sim 50°$, and $-80° \sim 80°$.)

We render three synthetic objects: BUNNY, FATTONY and JULIA with the size of 14 cm, as shown in the left side of Fig. 2.8. The camera is located at the world center with the focal lens and image resolution as 85 mm and 960×640. We locate the object 50 cm far from the camera. For illumination, we use 49 point light sources arranged in a 7×7 regular grid and capture 49 images by lighting on/off the LEDs one after another. The light source locations are in the range of $(X, Y, Z) = (\pm 21, \pm 21, 10)$, with the unit of centimeter. We assign the same RID to all the point lights on the board. On the right side of Fig. 2.8, we show the qualitative results of the estimated depth, normal and their corresponding error maps under the RID of "NCSB119B-V1." The error distributions show the influence of inaccurate modeling of the anisotropic RIDs. As our RID model is flexible for diverse RIDs, the surface normal and depth estimates from our method is more accurate than QD18 [18], and the corresponding error maps are uniform. We also provide quantitative comparison results on depth, surface normal and RID between our method and QD18 [18], as shown in Table 2.1 to Table 2.3. To calculate the accuracy of depth and normal, we calculate the mean absolute error and mean angular error between the estimation and the ground truth. Since there is a global scale ambiguity between our estimated RID and the ground truth RID, we
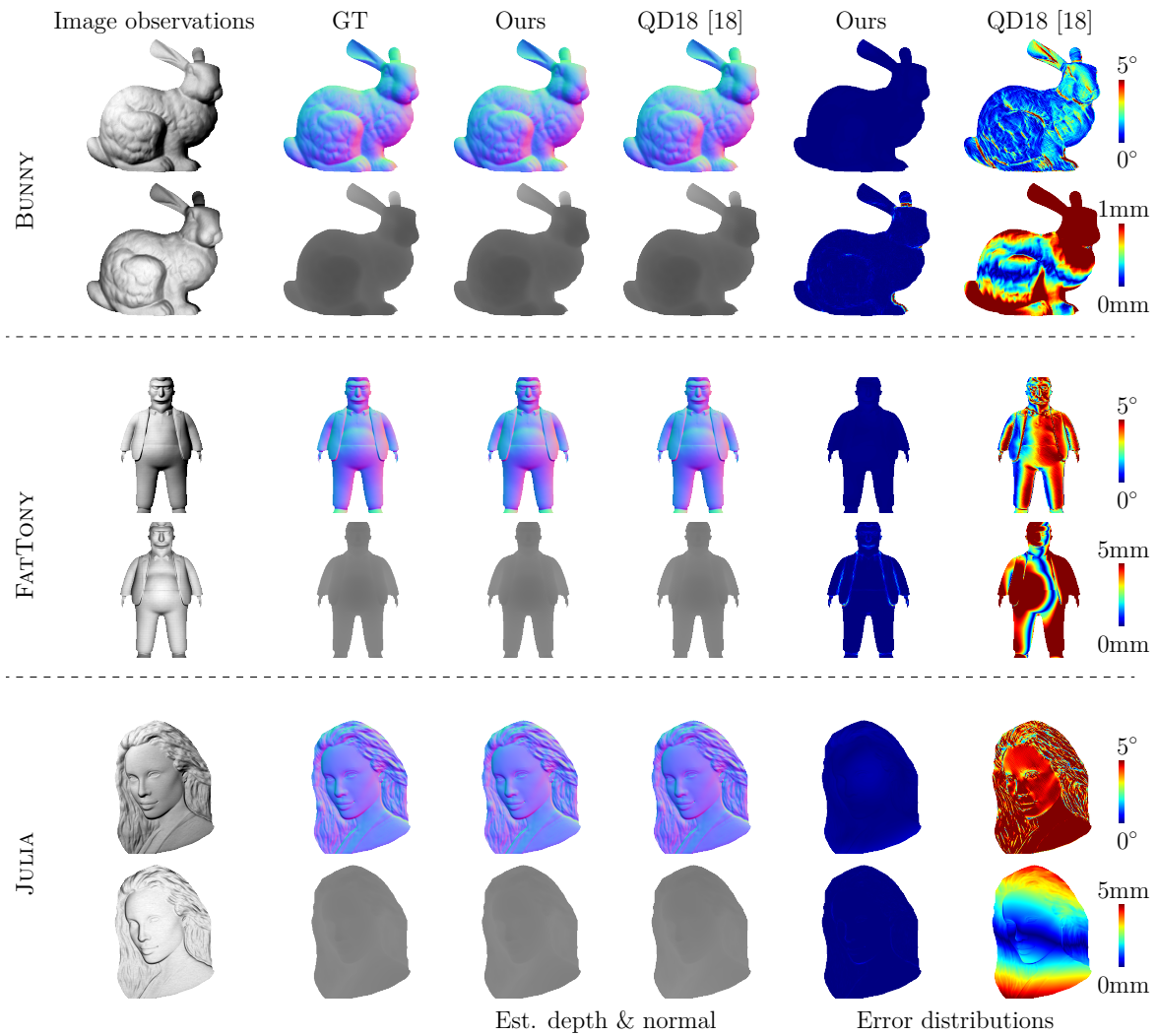
Fig. 2.8 Evaluation on shape estimation accuracy using synthetic dataset. The left side shows the near-light image observations. The right side shows the GT and estimated depth and surface normals, the corresponding absolute error maps and the angular error maps. Even rows give the surface normal estimates and odd rows give the depth estimates.

first calculate this global scale to align the RID estimation with the ground-truth and then evaluate the mean squared error on scaled RID. From the quantitative evaluation results, our self-calibrating photometric stereo method outperforms QD18 [18] on both shape and RID estimation.

Table 2.1 Comparison on depth estimation evaluated by mean absolute error in mm

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**0.031**	**0.034**	**0.025**
	NSPWF50DS	**0.016**	**0.017**	**0.015**
	SK6812RGBW	**0.016**	**0.017**	**0.016**
QD18 [18]	NCSB119B-V1	0.767	1.171	2.104
	NSPWF50DS	1.381	0.530	1.154
	SK6812RGBW	1.309	1.301	1.376

Table 2.2 Comparison on normal estimation evaluated by mean angular error in degree.

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**0.02**	**0.04**	**0.10**
	NSPWF50DS	**0.24**	**0.01**	**0.28**
	SK6812RGBW	**0.10**	**0.19**	**0.32**
QD18 [18]	NCSB119B-V1	1.49	3.51	5.19
	NSPWF50DS	2.25	2.02	3.01
	SK6812RGBW	2.51	3.03	3.45

Table 2.3 Comparison on RID estimation evaluated by mean squared error.

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**1.02e-6**	**1.23e-6**	**7.96e-6**
	NSPWF50DS	**2.02e-5**	**1.29e-7**	**1.41e-5**
	SK6812RGBW	**1.33e-5**	**3.68e-5**	**3.47e-5**
QD18 [18]	NCSB119B-V1	4.72e-5	3.38e-4	3.14e-4
	NSPWF50DS	1.56e-4	6.13e-5	8.42e-5
	SK6812RGBW	1.57e-4	2.59e-4	1.60e-4
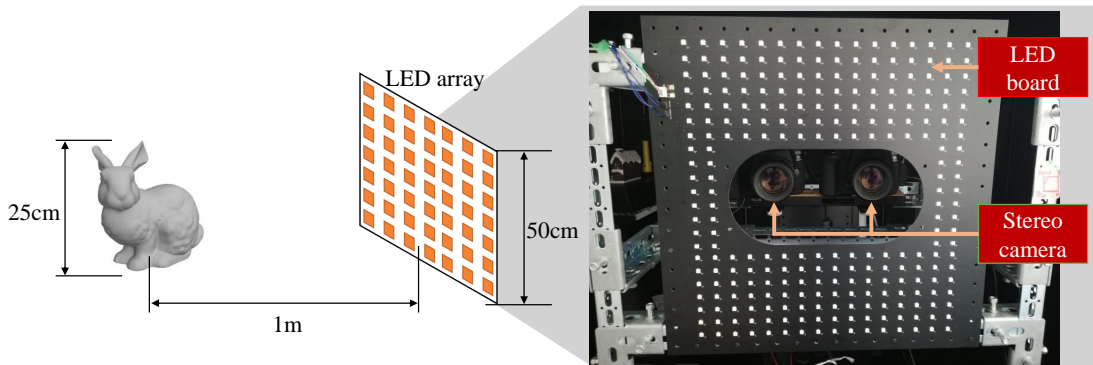
Fig. 2.9 Hardware setup for capturing near-light image observations.

2.6 Experiments on real dataset

As shown in Fig. 2.9, our capture setup includes two Canon EOS 5D Mark IV cameras with 85mm lens and a LED array board containing 256 point light sources with their positions calibrated. We carefully design LED distributions and the distance to a target object to maintain a near-light illumination. Specifically, the LED board has the size of 50×50 cm and we put it 1 m far from the target object, whose size is roughly in the range of 25×25 cm so that different scene points are illuminated by varying lighting directions, fall-offs and radiant intensities. The baseline between two cameras is 16.5 cm. We apply a semi-global stereo matching algorithm [46] to calculate the disparity between stereo image pairs and further obtain an initial sparse depth map of the target. Our LED board is equipped with LED diode "SK6812RGBW" from OPSCO Optoelectronics Corp[2]. The product's theoretical RID is shown in Fig. 2.5.

To build our real dataset, we choose two objects: Barrel and Pot as shown on the left side of Fig. 2.9. The object size is in the range of 25×25 cm so that the illumination can be treated as near-light. For each object, we first capture stereo image pairs under the natural light for computing the initial depth via stereo matching [46]. Then we capture 256 image observations under the point light illumination as the input of our near-light photometric stereo. As shown on the right side of Fig. 2.10, the estimated shape, albedo, and surface normal from our method and previous work

---

[2]Normand SC6812RGBW LED, `http://www.normandled.com/Product/view/id/883.html`. Retrieved Nov. 16, 2021
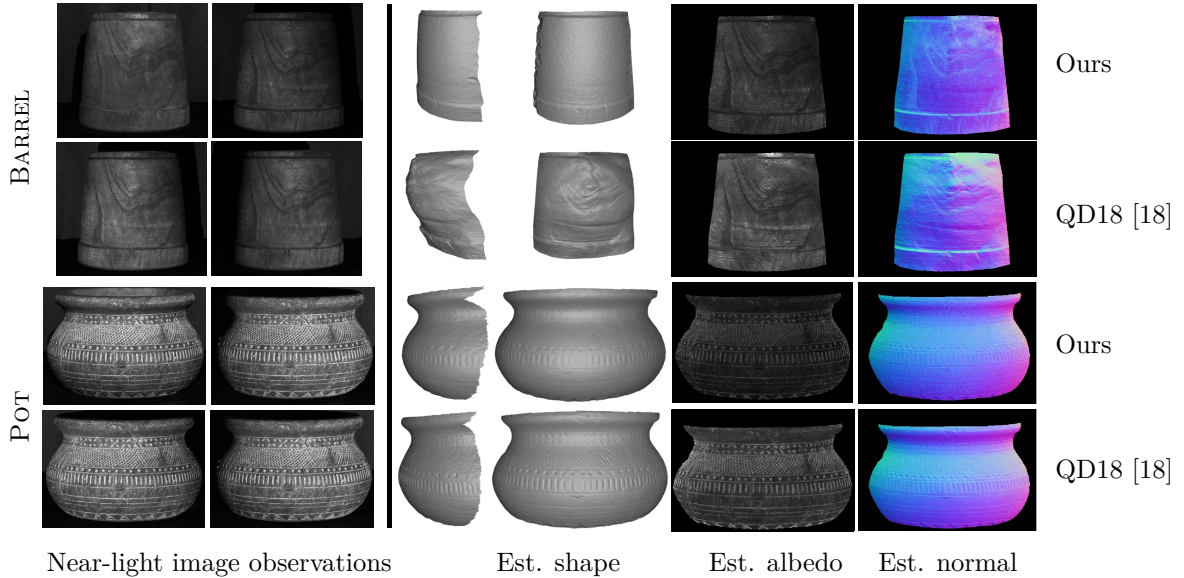
Fig. 2.10 Comparison on shape and albedo estimation between our method and QD18 [18] on real dataset. Our method achives more reasonable results from the side view of the shape estimation.

QD18 [18] are presented. From the side view of the shape estimation, our method achieves more reasonable shape reconstructions, especially for the BARREL object. Therefore, with the proposed flexible RID model and the self-calibrating near-light photometric stereo, we can obtain more accurate shape estimation under practical anisotropic point light illumination.

2.7 Conclusion

In this chapter, we propose a self-calibrating near-light photometric stereo method to handle shape recovery under anisotropic point lights. Our key observation is that the RID is defined in a spherical coordinate, and the interest RID in the photometric stereo task is located in a local region, which derives our RID representation model based on spherical harmonic bases. Given the sparse depth as input, we first conduct RID calibration with the proposed RID model, then iteratively recover complete object shape based on near-light image formation model and the surface normal integration.

Limitations   Our method is limited to Lambertian reflectance and known sparse depth map. To handle general surface reflectance, a possible solution is extracting the low-frequency components from image observations, which are close to Lambertian reflectance as stated in [47]. The sparse depth map obtained from stereo matching contains noise in practice. That influences the optimization of RID calibration and shape reconstruction. In our future work, we will apply the variational approach used in [18, 27] to improve the robustness of our method on noisy sparse depth input.

Besides, when the local RID has large variations, our RID representation with limited order of spherical harmonic bases becomes less accurate. In such cases, we can increase the number of spherical harmonic bases to improve the representation power. However, more image observations and sparse depth points are required to estimate additional unknowns from the basis coefficients in our RID model, as shown in Eq. (2.14). Therefore, our method cannot fit complex local RIDs accurately when we have limited input images and sparse points with known depth.

Finally, in addition to the qualitative comparison shown in Fig. 2.10, it is better to conduct a quantitative evaluation on real data. To achieve that, the ground-truth surface normal, depth, and albedo corresponding to the near-light images are required. In our future work, we plan to build a real-world dataset for near-light photometric stereo, from which we can compare existing methods with ours quantitatively.

Chapter 3

Photometric Stereo under Uncalibrated Natural light

3.1 Introduction

Given an image sequence of a Lambertian object illuminated by three non-coplanar directional lights, surface normals of the object could be estimated by photometric stereo [7]. The pixel-level details of surface normal estimates are of great interest for applications in 3D computer vision such as visual inspection [48] and augmented reality [49].

The classic photometric stereo setup has two assumptions on lighting: directional and calibrated lighting, which restrict the applicability of conventional photometric stereo. The directional lighting model assumes a single point light source placed far away from the target object, and typically requires the data capture to be conducted in a dark lab setting, as shown in Fig. 3.1(a). The calibrated lighting assumption needs an additional step for measuring both lighting intensities and directions, and calibrating lighting itself is also an ongoing research problem [10]. If the former assumption is relaxed, the problem becomes calibrated photometric stereo under natural illumination, while relaxing the latter assumption leads to uncalibrated photometric stereo under directional lighting. A fully calibration-free method under general light-

(a) Observation under calibrated directional light   (b) Observation under uncalibrated natural light
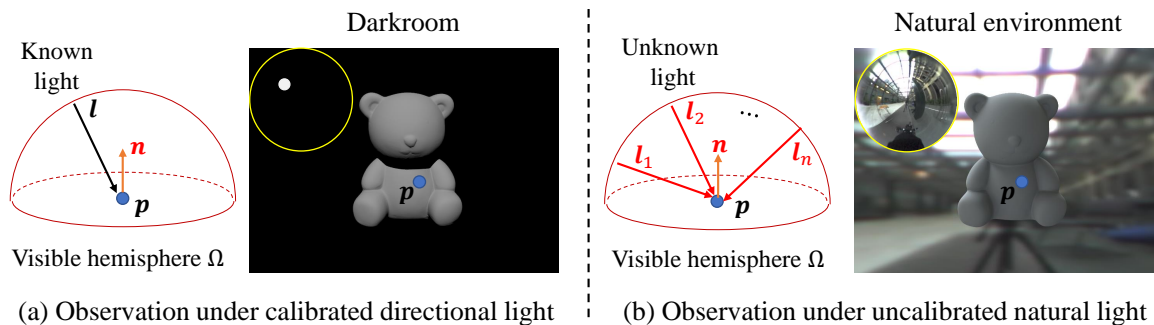
Fig. 3.1 Calibrated directional light vs. Uncalibrated natural light. The top left circle at each image observation shows the environment map for the scene. The visible hemisphere illustrates the illumination difference for one scene point under two light setting, where the unknowns are shown in red.

ing is desired because it will push photometric stereo from the laboratory setup to the practical wild environment and simplify the effort of 3D scanning for non-experts.

Generalizing calibrated and directional lighting assumptions at the same time will make the problem rather complicated. As shown in Fig. 3.1(a), we show an image observation under a single calibrated directional light, where the environment map shown in the top left circle records the lighting directions and intensities of the surrounding illumination. For one scene point, the illumination could come from all possible directions in the visible hemisphere determined by its surface normal direction (lights with directions out of the visible hemisphere are blocked, leading to attach shadows). However, in the case of single directional light in the darkroom, the environment map contains only a single point with its position and pixel value indicating the lighting direction and intensity. If we generalize the single directional illumination to the natural illumination as shown in Fig. 3.1(b), the environment map becomes more complicated as the surrounding environment contains multiple direct or indirect light sources. Therefore, one scene point is illuminated by varying lighting directions and intensities in the domain of the visible hemisphere. Besides, If we further assume an uncalibrated natural light setting, all the lighting directions and their corresponding intensities are unknown, making the problem even more challenging.

To solve photometric stereo with uncalibrated natural illumination, existing meth-
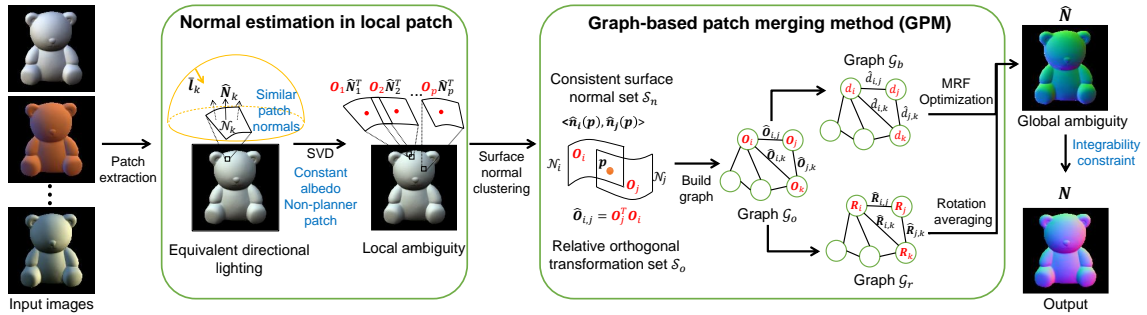
Fig. 3.2 Pipeline of patch-based uncalibrated photometric stereo. Variables shown in red represent the unknowns in our method. The assumptions in different stage of our method is shown in blue.

ods [50–53] require a rough shape of the target object. Although the initial shape can be obtained from multiview geometry [53], object shape prior (*e.g.*, face [51]) or RGBD cameras [3], it either needs extra system setup or restricts the application into pre-defined shapes. Recent work [54] uses a balloon-like perspective depth map for the shape initialization and estimate the surface shape and reflectance by an end-to-end variational optimization framework. However, the recovered shape accuracy after the optimization is still sensitive to depth initialization. On the other hand, Jung *et al.* [55] restrict the natural illumination as the skylight and solve the outdoor photometric stereo based on the prior of skylight distribution. Brahimi *et al.* [56] provide a closed-form solution for photometric stereo under general unknown lighting and perspective camera projection. However, the environment lighting is approximated by global first-order spherical harmonics, which has a gap with real-world natural illumination [57]. In summary, existing methods for uncalibrated photometric stereo under natural illumination are still limited due to the requirement of initial shapes and restrictive lighting approximation models.

In this chapter, we propose a photometric stereo method for uncalibrated natural illumination that relieves the requirements used in existing methods. We develop a "divide and conquer" approach to first "divide" the problem into tractable sub-problems with locally-resolvable ambiguity, and then "conquer" them jointly by merging all sub-results as a complete solution. Our key observation is that for a small surface patch

with similar surface normal directions, the visible hemisphere of an environment map also shows smooth changes. Therefore, the environment lighting for that local patch could be approximated as equivalent directional lighting by summing up all samples on the visible hemisphere. Based on the above observations, we present our method as shown in Fig. 3.2, where the "divide" and "conquer" processes are corresponding to the two modules in the pipeline.

In the "normal estimation in local patch" stage, we assume the surface normals in an extracted patch have similar directions so that the patch illumination can be approximated by equivalent directional lights. Following directional lighting-based uncalibrated photometric stereo techniques [58], we further assume the local patch has uniform albedo and non-planar shape, then the surface normals can be recovered up to an orthogonal ambiguity. In the "graph-based patch merging" stage, we resolve the orthogonal ambiguity in each patch and merge local shapes to a complete surface. Specifically, we first cluster consistent (equal) surface normal pairs and use them to calculate relative orthogonal transformations, which describe the geometry relationship among patches. Then an orthogonal ambiguity graph $\mathcal{G}_o$ is constructed with nodes and edges being set to the unknown patch-wise orthogonal ambiguities and the known relative orthogonal transformations. As the 3×3 orthogonal ambiguity can be decomposed into a binary part and a 3D rotation part, we divide the orthogonal ambiguity graph into a binary ambiguity graph $\mathcal{G}_b$ and a rotation ambiguity graph $\mathcal{G}_r$. We formulate the binary ambiguity estimation on $\mathcal{G}_b$ as a per-patch labeling problem, and solve it by a Markov Random Field (MRF) optimization [59]. Guided by the relative rotations (edges of the rotation ambiguity graph $\mathcal{G}_r$), we solve the rotation ambiguities in each patch by introducing rotation averaging [60] algorithms which has been widely applied in structure from motion framework [61]. After the patch merging stage, the unknown patch-wise orthogonal ambiguities can be determined up to a global orthogonal ambiguity. By further assuming the whole surface to be integrable, this global orthogonal ambiguity can be finally reduced to a convex/concave ambiguity.

An earlier version of this work appeared in [62]. Different from the graph-based

patch merging method presented in this chapter, the patch merging process in [62] takes consistent surface normal pairs as constraints and constructs an angular distance matrix with the element calculated by propagating angular distance along the shortest path between any two surface normal directions. Then the complete surface normal map is solved up to a global orthogonal ambiguity by conducting matrix factorization on this angular distance matrix. We refer this method as matrix-based patch merging method (MPM) corresponding to our newly proposed graph-based patch merging method (GPM) in this chapter. Compared with [62], this work improves the surface normal estimation accuracy by replacing MPM with GPM, and provides analysis of surface normal clustering under natural illumination via *consistent orthogonality condition*. To demonstrate the effectiveness of our new method, additional experiments on both synthetic and real data are also presented. To summarize, the main contributions of our work are as follows:

1. We explore the equivalent directional lighting model to solve patch-wise surface normal up to local ambiguities, bypassing the explicit requirement of global information of environment maps.

2. We extend the surface normal clustering via intensity profiles from directional lighting to general lighting case, and propose a consistent orthogonality condition to extract consistent surface normal pairs.

3. We introduce a rotation averaging and an MRF optimization to solve patch-wise orthogonal ambiguities and merge local surface normal solutions to a complete surface normal map up to a global orthogonal ambiguity.

Our output surface normal map only contains a concave/convex binary ambiguity. As proved in [56], it is an inherent ambiguity in uncalibrated photometric stereo under orthogonal camera projection and cannot be solved with image cues only. However, it can be either manually removed with little effort or resolved with shape prior.

3.2 Related work

There are two major restricting assumptions that need to be relaxed for photometric stereo [7] to be applied to practical applications – calibrated directional lighting assumption and Lambertian reflectance assumption. Correspondingly, to make photometric stereo work in more realistic scenes, there are two directions to generalize the conventional approach – generalization of lighting assumption and generalization of the reflectance model. This chapter focuses on the former problem, thus both calibrated and uncalibrated photometric stereo methods with non-Lambertian objects (*e.g.*, [63–67]) are beyond the scope, and we refer the readers to [68] for a comprehensive review and comparison of non-Lambertian photometric stereo methods.

3.2.1 Calibrated, directional lighting

The calibrated Lambertian photometric stereo with directional lighting assumption is the most classic setup. The first photometric stereo work [7] and its robust extensions rely on these assumptions. Various robust approaches have been proposed to eliminate deviations from the classic model by treating the corrupted measurements as outliers, such as Random Sample Consensus (RANSAC) [69, 70], median-based approach [71], low-rank matrix factorization (Robust-PCA) [72], and expectation maximization [73]. Wu *et al.* [74] formulate the calibrated dense photometric stereo problem as a Markov network. The per-pixel surface normal, encoded in the graph node, is optimized by minimizing the surface geometry smoothness (smoothness term) and the distance between its normal initialization (data term). Similar to their method, our proposed GPM also formulates the patch merging problem to a graph structure, but with different graph content and optimization scheme. As we have no light calibration or initial normal map, our optimization target encoded in the node is the patch-wise orthogonal ambiguity constrained by the relative orthogonal transformations assigned to the graph edges. The optimizations for the binary ambiguity and the rotation ambiguity are solved with MRF and rotation averaging, separately.

3.2.2   Calibrated, natural lighting

Natural illumination can be calibrated directly by using a mirror sphere as a light probe or indirectly by approximating sunlight as a dominant directional source. With mirror sphere measured environment maps, Yu *et al.* [2] show photometric stereo results by directly sampling the captured natural illumination. Ackermann *et al.* [75] implement photometric stereo for outdoor webcams using a time-lapse video, and Abrams *et al.* [76] show the necessity of using images taken over many months (thousands of images) for sufficiently observing illumination variations. Jung *et al.* [77] develop parameterized sun and sky lighting models to apply photometric stereo under outdoor illumination. Their latter work [55] refines the sky model and obtains better normal estimates on cloudy days. Shen *et al.* [78] provide an analysis about the limitation of point light source modeling for 1-day outdoor photometric stereo. Hold-Geoffroy *et al.* [1] show that outdoor observations recorded within a few hours could constrain a reliable normal estimation.

3.2.3   Uncalibrated, directional lighting

Photometric stereo without calibrated lighting as known input is called uncalibrated photometric stereo. Even if the lighting assumption is directional lighting, the solutions to both surface normal and lighting are not unique due to some inherent ambiguities. The shape (or lighting) can be estimated up to a 3×3 linear ambiguity [58]. When the surface is integrable, this ambiguity further reduces to a 3-parameter Generalized Bas-Relief (GBR) ambiguity under orthographic projection [79, 80] and vanishes under perspective camera projection [81]. Existing methods focus on the estimation of the 3 unknowns in GBR ambiguity to recover the normal estimates by using priors on albedo [82, 83], detecting local maximum diffuse points [84], or reflectance symmetry [85, 86]. If multiview inputs are available, the directional lighting directions could also be indirectly estimated, and photometric constraints are used to refine the shape [87, 88]. The lighting can also be semi-calibrated with directions being provided

Table 3.1 Summary of uncalibrated photometric stereo methods under natural illumination, where f, o, p, and k represent the numbers of images, spherical harmonic (SH) lighting basis, valid pixels, and extracted patches, respectively. Our method solves uncalibrated photometric stereo under a moderately flexible lighting model without requiring a initial shape prior.

Method	Initial shape	Lighting model	Lighting parameter	Representation power
[56]	None	Global SH	$f \times 4$	Weak
[3, 52]	Depth sensor	Global SH	$f \times o$	Weak
[54]	Visual hull [90]	Global SH	$f \times o$	Weak
[53, 91]	MVG[a]	Global SH	$f \times o$	Weak
[92]	Planar[c]	SV-SH[b]	$f \times 3 \times p$	Strong
[93]	Depth sensor	SV-SH	$f \times o \times k$	Strong
Ours	None	SV-directional	$f \times 3 \times k$	Moderate

[a] Multiview geometry
[b] Spatially-varying (SV) spherical harmonic (SH) lighting
[c] The method [92] is validated by near-planar objects.

and intensities remaining unknown [89].

3.2.4 Uncalibrated, natural lighting

This is the most challenging category of lighting conditions since it is general and unknown. Table 3.1 summarizes existing uncalibrated photometric stereo methods under natural illumination. Brahimi *et al.* [56] approximate the shading of the whole surface with a first-order SH globally, where the number of lighting parameters to be optimized is $4f$. With the integrability constraint, they show that the uncalibrated natural light photometric stereo under perspective camera projection is well-posed. However, the first-order SH is a simplified natural lighting representation. For the second-order SH representation ($o = 9$), there is a 9×3 ($= 27$ unknowns) linear ambiguity in estimated surface normals [94]. Unfortunately, this high-dimensional ambiguity cannot be completely removed without additional information. Existing methods require initial shapes from depth sensor [3, 52], or multiview geometry [95] to make the problem solvable. Recently, Haefner *et al.* [54] propose a variational optimization framework to recover shape, reflectance, and illumination jointly. Although their method auto-

matically initializes shape from silhouette [90], the embedded non-convex optimization framework is still sensitive to the initialization of depth, albedo, and lighting vectors. Compared to modeling the natural illumination for the whole surface with a global SH lighting, existing methods [92, 93] propose spatially-varying spherical harmonic (SV-SH) lighting models. Maier *et al.* [93] divide a surface into k patches and model the per-patch SH lighting independently. Quéau *et al.* [92] further optimize the per-pixel SV-lighting direction directly. Although the SV-SH model can accurately represent the natural light, the uncalibrated photometric stereo becomes highly ill-posed due to numerous unknown lighting parameters to be optimized. To make the problem solvable, these methods require dedicated shape initialization [92, 93] and non-physical lighting regularization [92].

Similar to Maier *et al.* [93], our method models the natural illumination with SV-directional lighting in local patches. As shown in Table 3.1, compared to the global SH lighting approximation, our lighting model has a stronger representation power for real-world natural illumination. Compared to the SV-SH model applied in Maier *et al.* [93], the SV-directional model is less flexible. However, with this moderate flexibility for natural lighting approximation and the physical assumptions on patches (uniform albedo, non-planar shape), we can obtain patch surface normals without requiring a shape initialization used in [93].

3.3 Normal estimation in local patch

Our method is based on the Lambertian image formation model under natural light. We assume direct illumination (*i.e.*, ignoring cast shadow, inter-reflection) and assume the camera is radiometrically calibrated or has a linear response, *i.e.*, the pixel brightness equals to the scene radiance. Let us consider a photometric stereo image sequence illuminated by f different environment maps. In default, for each valid pixel, we extract a patch $\mathcal{N}_k$ ($k = \{1, 2, \ldots, p\}$, where p is the total number of pixels) centered at the pixel location. In the following, we first approximate the illumination at a local
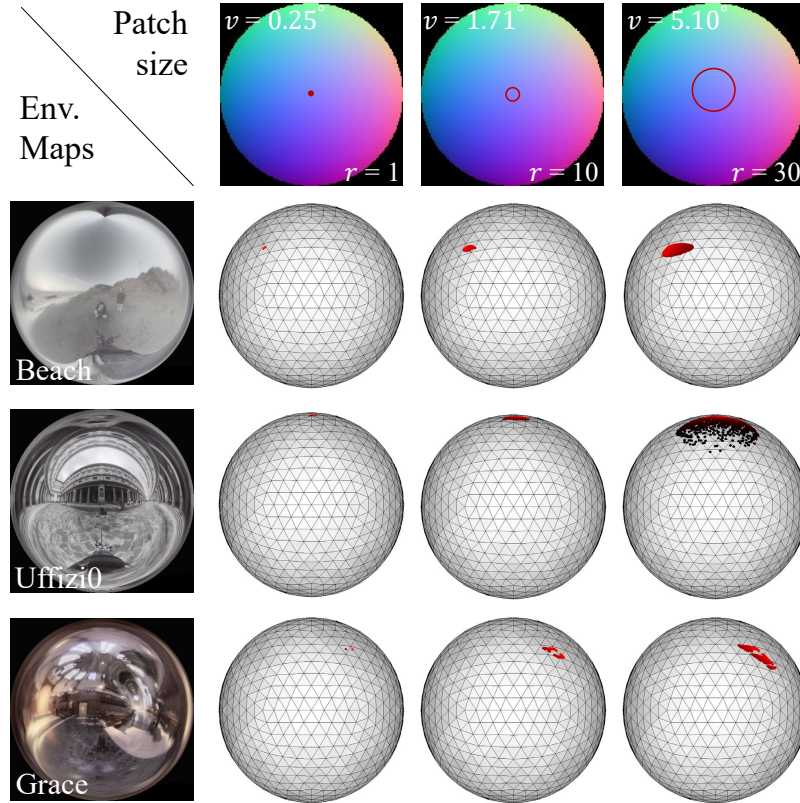
Fig. 3.3 Illustration of environment lighting approximation. Patches highlighted by concentric red circles contain varying radii r and mean angular difference of surface normal v. For each patch, we draw the equivalent directional lighting directions (dot on spheres) and intensities (red means strong while black means weak) for three different environment maps (figures courtesy of [96]).

patch with an equivalent directional lighting model, and then estimate the surface normals within the patch by conventional uncalibrated photometric stereo algorithms.

3.3.1 Equivalent directional lighting model

Given a scene point with Lambertian albedo ρ and surface normal $\mathbf{n} \in \mathbb{S}^2 \subset \mathbb{R}^3$, its pixel brightness is written as

$$b = \int_{\Omega} \rho L(\boldsymbol{\omega}) \max((\mathbf{n}^{\top}\boldsymbol{\omega}), 0)\mathrm{d}\boldsymbol{\omega}, \tag{3.1}$$

where $\boldsymbol{\omega} \in \mathbb{S}^2 \subset \mathbb{R}^3$ is a unit vector in the visible hemisphere Ω, and $L(\boldsymbol{\omega})$ is the environment lighting intensity from direction $\boldsymbol{\omega}$.

For any surface normal vector $\mathbf{n}_k$, it uniformly receives illumination from direction $\boldsymbol{\omega}$ sampled on the visible hemisphere $\Omega_k = \{\boldsymbol{\omega} \mid \mathbf{n}_k^\top \boldsymbol{\omega} \geq 0\}$ of the environment map. Then for any $\boldsymbol{\omega} \in \Omega_k$ we may perform the spherical integration over Ω_k to obtain the pixel brightness:

$$b_k = \rho \mathbf{n}_k^\top \int_{\Omega_k} L(\boldsymbol{\omega}) \boldsymbol{\omega} \mathrm{d}\boldsymbol{\omega} = \rho \mathbf{n}_k^\top \bar{\mathbf{l}}_k, \tag{3.2}$$

where $\bar{\mathbf{l}}_k$ denotes an *equivalent directional lighting* as the integral of all samples in Ω_k, and the subscript k indicates that for different surface normals, they face different visible hemispheres and therefore correspond to different equivalent directional lighting. Note here $\mathbf{n}$ is a unit vector, but $\bar{\mathbf{l}}$ is not necessary of length one since it encodes intensity scaled directional lighting direction.

We assume the surface normals in a small patch have similar directions. In this way, the natural illumination does not show abrupt changes for scene points within the patch. Given two surface normals within patch $\mathcal{N}_k$, we measure their *angular difference* as $\langle \mathbf{n}_{k,i}, \mathbf{n}_{k,j} \rangle = \arccos(\mathbf{n}_{k,i}, \mathbf{n}_{k,j})$. To evaluate surface normal's variation in a patch, we define the *mean patch angular difference* by $v_k = \frac{1}{p_k} \sum_i \langle \mathbf{n}_{k,i}, \mathbf{n}_{k,c} \rangle$, where c is the patch center index and p_k is the number of scene points in that patch. Then for a surface patch with small mean angular difference v_k, all surface normals should share approximately the same visible hemisphere Ω_k as well as $\bar{\mathbf{l}}_k$, so their brightness could be modeled by a single directional light as illustrated in Fig. 3.2. Similar lighting representation has been applied in [1, 92].

We illustrate and verify our lighting assumption using a synthetic experiment. Given a surface normal, we calculate its equivalent directional lighting by summing up all samplings on its visible hemisphere of the environment map, and draw the intensity and direction of such a lighting vector on the sphere as shown in Fig. 3.3. We use a sphere normal map of 256×256 pixels (the radius of the sphere is 128 pixels in the image domain) and calculate the equivalent lighting under three light probes

from [96]. By selecting central patches with the radius of $\{1, 10, 30\}$ pixels (indicated as red circles), the mean angular difference of surface normals increases from 0.25° to 5.10°, leading to more scattered equivalent directional lighting distributions. For relatively smaller patches (radius ≤ 10, around 300 pixels) whose surface normals having smaller variation, the corresponding lighting vectors are highly concentrated. In such case, it is safe for us to apply directional lighting assumptions in a patch-wise manner. In the following computation, we neither know the direction and intensity of equivalent directional lighting nor solve them explicitly, while we develop an uncalibrated photometric stereo method to solve the surface normal directly.

3.3.2   Uncalibrated photometric stereo in local patch

Assume a local surface patch $\mathcal{N}_k$ is illuminated by f different equivalent directional lighting $\mathbf{L}_k = [\bar{\mathbf{l}}_{k,1}, \bar{\mathbf{l}}_{k,2}, \cdots, \bar{\mathbf{l}}_{k,f}]$. Denote the matrix stacking all surface normal vectors $\mathbf{n}^\top$ in patch $\mathcal{N}_k$ in a row-wise manner as $\mathbf{N}_k$, denote the patch albedo as $\boldsymbol{\rho}_k = [\rho_1, \cdots, \rho_{p_k}]$, then the image brightness of this patch, denoted as $\mathbf{B}_k$, could be written as follows

$$\mathbf{B}_{k|p_k \times f} = \mathrm{diag}(\boldsymbol{\rho}_k)\mathbf{N}_{k|p_k \times 3}\mathbf{L}_{k|3 \times f}, \tag{3.3}$$

where $\mathrm{diag}(\cdot)$ is a diagonalization operator and p_k is the total number of pixels in patch $\mathcal{N}_k$. This representation is different from spherical harmonics for natural light, where a high-dimensional matrix decomposition (9D decomposition for a second order spherical harmonics) exists with unknown lighting [53, 94].

According to Eq. (3.3), for each patch the equivalent directional lighting model relieves the problem to be the Lambertian photometric stereo under unknown directional lighting, which is a well-studied research area with tractable solutions. So we perform SVD on $\mathbf{B}_k$, as it was done in classic uncalibrated photometric stereo methods [58]. The SVD decomposition gives us $\mathbf{B}_k = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, wherein ideal case $\boldsymbol{\Sigma}$ only contains three non-zero diagonal elements. We further denote $\tilde{\mathbf{N}}_k = \mathbf{U}\sqrt{\boldsymbol{\Sigma}}$ and $\tilde{\mathbf{L}}_k = \sqrt{\boldsymbol{\Sigma}}\mathbf{V}^\top$, where $\tilde{\mathbf{N}}_k$ and $\tilde{\mathbf{L}}_k$ are pseudo surface normals and pseudo equivalent directional light-

ing for each patch. Here, both the normal and lighting solutions contain an unknown 3×3 linear ambiguity, denoted as $\mathbf{Q}_k$, since any invertible matrix can be inserted between $\tilde{\mathbf{N}}_k$ and $\tilde{\mathbf{L}}_k$ to maintain the equality.

As we work on small patches, it is safe to assume a piece-wise uniform albedo. Suppose pixels within the patch $\mathcal{N}_k$ have the same albedo α_k, *i.e.* $\boldsymbol{\rho}_k = \alpha_k \mathbf{1}$, the pseudo surface normals for this patch should satisfy

$$\|\tilde{\mathbf{n}}_{k,i} \mathbf{Q}_k\|_2^2 = \tilde{\mathbf{n}}_{k,i} \mathbf{Q}_k^\top \mathbf{Q}_k \tilde{\mathbf{n}}_{k,i}^\top = \alpha_k, \tag{3.4}$$

where $\tilde{\mathbf{n}}_{k,i} \in \mathbb{R}^3$ is the i-th ($i = \{1, 2, \ldots, p_k\}$) row vector of $\tilde{\mathbf{N}}_k$. Without losing generality, we set $\alpha_k = 1$. As $\mathbf{Y}_k = \mathbf{Q}_k^\top \mathbf{Q}_k$ is an symmetric matrix, we can solve it if the local patch contains at least 6 pixels with varying surface normals:

$$\underbrace{\left[\mathrm{tri}(\tilde{\mathbf{n}}_{k,1}\tilde{\mathbf{n}}_{k,1}^\top) \quad \cdots \quad \mathrm{tri}(\tilde{\mathbf{n}}_{k,p_k}\tilde{\mathbf{n}}_{k,p_k}^\top)\right]^\top}_{\mathbf{E}} \underbrace{\mathrm{tri}(\mathbf{Y}_k)}_{\mathbf{y}} = \mathbf{1}, \tag{3.5}$$

where $\mathrm{tri}(\cdot)$ operator extracts the upper triangle matrix elements as a vector. The residue $\|\mathbf{Ey} - \mathbf{1}\|_2^2$ is recorded as e_a^k to measure the reliability of uniform albedo assumption. We conduct SVD on $\mathbf{Y}_k$ such that $\mathbf{Y}_k = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^\top$ and assign $\hat{\mathbf{Q}}_k$ as $\sqrt{\tilde{\mathbf{\Sigma}}}\tilde{\mathbf{U}}^\top$. Then we obtain pseudo surface normal map as $\hat{\mathbf{N}}_k = \tilde{\mathbf{N}}_k \hat{\mathbf{Q}}_k$. It has been proved in [79, 83] that the uniform albedo constraint reduces the 3×3 linear ambiguity in $\tilde{\mathbf{N}}_k$ to an orthogonal one in $\hat{\mathbf{N}}_k$ such that

$$\mathbf{B}_k = \hat{\mathbf{N}}_k \mathbf{O}_k^\top \mathbf{O}_k \hat{\mathbf{L}}_k, \tag{3.6}$$

where $\mathbf{O}_k \in O(3)$ is the orthogonal ambiguity that varying from patch to patch. $\hat{\mathbf{N}}_k$ and $\hat{\mathbf{L}}_k$ are the pseudo surface normals and equivalent directional lighting up to an orthogonal ambiguity w.r.t. their corresponding ground truth, *i.e.*,

$$\begin{aligned} \mathbf{N}_k^\top &= \mathbf{O}_k \hat{\mathbf{N}}_k^\top, \\ \mathbf{L}_k &= \mathbf{O}_k \hat{\mathbf{L}}_k. \end{aligned} \tag{3.7}$$

Non-uniform albedo across patches. As shown in Fig. 3.4(a), patches across the boundary of different albedos cannot keep uniform albedo assumption. Also, the natural illumination within patches near the feet and the ear part of the Bunny object cannot be treated as equivalent directional lighting since the surface normals at these regions vary significantly. Therefore, patch-wise surface normal estimates in these regions are inaccurate, as visualized in Fig. 3.4(b). Here we first fit an orthogonal matrix to align the patch-wise pseudo surface normals with the corresponding ground truth. Then the mean angular error between aligned pseudo surface normals and the truth surface normals are calculated to measure the patch-wise surface normal estimation accuracy. Hereafter we denote this error map as *patch surface normal error map*.

To reduce the influence of these inaccurate local surface normal estimates in the following patch merging process, we define a confidence metric to measure the reliability of normal estimation. For a surface patch $\mathcal{N}_k$, we evaluate the equivalent directional lighting approximation by defining a normalized patch re-rendering error $e_r^k = \|\mathbf{B}_k - \hat{\mathbf{N}}_k \hat{\mathbf{L}}_k\|_F^2 / \|\mathbf{B}_k\|_F^2$ and test the uniform albedo assumption by the residue e_a^k calculated from Eq. (3.36). Based on these two metrics e_a^k and e_r^k, we define the surface normal estimation confidence of patch $\mathcal{N}_k$ as follows,

$$c_k = e^{-(\beta e_r^k + \gamma e_a^k)}, \tag{3.8}$$

where β and γ are the coefficients used to balance e_r^k and e_a^k, and we set them as 5 and 0.5 empirically. As shown in Fig. 3.4(b-c), the confidence values of all patches are consistent with the patch surface normal angular error map.

3.4 Graph-based patch merging method

For each patch, now we have estimated pseudo surface normal $\hat{\mathbf{N}}_k$ up to an orthogonal ambiguity $\mathbf{O}_k$, with surface normal confidence measured by c_k. In this section, we will discuss how to merge all the patches into an entire surface. Specifically, we first
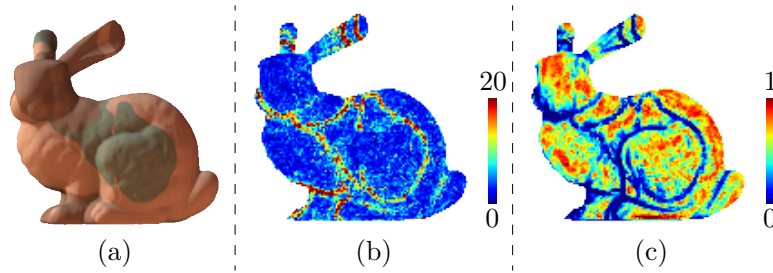
Fig. 3.4 An example of non-uniform albedo causing large errors across patches. (a) Image observation of the Bunny object with non-uniform albedo. (b) Mean angular errors (degree) of the patch-wise pseudo surface normals w.r.t. the true surface normals. Each pixel value encodes the mean angular error of the estimated surface normals for the patch centered at that pixel location. (c) Confidence map of patch surface normal estimation.

show consistent surface normal pair extraction from patch overlapping regions and intensity profiles, followed by the calculation of relative orthogonal transformations among patches. Taking relative orthogonal transformations as constraints, we introduce MRF optimization and rotation averaging to solve the patch-wise orthogonal ambiguities and merge the whole surface normal up to a global orthogonal ambiguity. This global ambiguity is finally reduced to a concave/convex ambiguity by addressing integrability.

MPM proposed in previous work [62] conducts this step by taking consistent surface normal pairs as constraints and creating an angular distance matrix with its element filled by propagating angular distance along the shortest path between every surface normal pairs. The whole surface is then obtained by matrix factorization on this angular distance matrix. Please refer to the original paper in [62] for details. However, MPM suffers from error accumulation during the propagation process. Also, the angular distance between two surface normals could be constrained by all possible paths connecting the corresponding scene points, only selecting the shortest path to constraint surface normals cannot guarantee a globally optimized result. As discussed below, the newly proposed GPM avoids the accumulative error in MPM [62] and optimizes all the patch connections simultaneously.

(a)

(b)   (c)
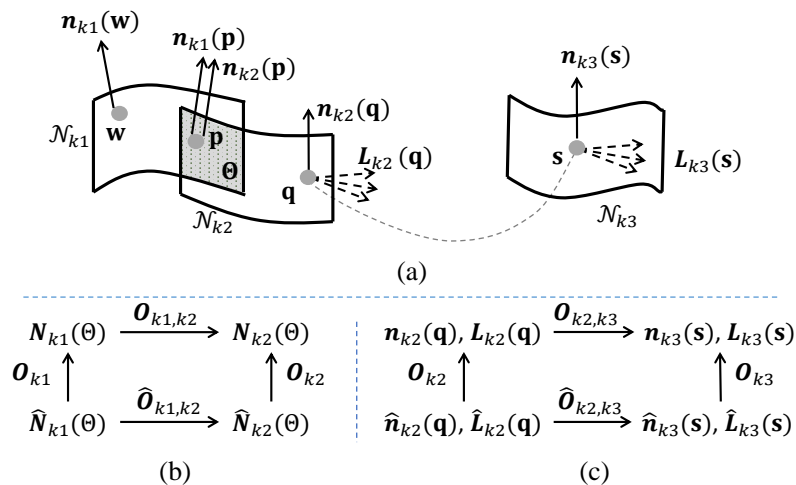
Fig. 3.5 Illustration of consistent surface normal clustering. (a) Consistent surface normal pairs are extracted from overlapping patch region Θ and scene point pair $(\mathbf{q}, \mathbf{s})$ satisfies consistent orthogonality condition. (b) The relative orthogonal transformation between patch $\mathcal{N}_{k1}$ and $\mathcal{N}_{k2}$ is calculated following the relationship between surface normals in the overlapping patch region Θ (Eqs. (3.9) and (3.10)). (c) The relative orthogonal transformation between patch $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$ is extracted from the consistent orthogonality condition between scene points $\mathbf{q}$ and $\mathbf{s}$ (Eqs. (3.13) and (3.14)).

3.4.1 Consistent surface normal clustering

As shown in Fig. 3.5(a), we provide three surface patches $\mathcal{N}_{k1} \sim \mathcal{N}_{k3}$ covering scene points $\{\mathbf{w}, \mathbf{p}, \mathbf{q}, \mathbf{s}\}$. For any scene point $\mathbf{p}$ located at the overlapping region $\Theta = \mathcal{N}_{k1} \cup \mathcal{N}_{k2}$ shown in the highlight area, the true surface normals from different patches at this point are consistent, *i.e.*, $\langle \mathbf{n}_{k1}(\mathbf{p}), \mathbf{n}_{k2}(\mathbf{p}) \rangle = 0$. In Fig. 3.5(b), we show the relationship of surface normals in the overlapping region between two patches. The unknown orthogonal ambiguities in patch $\mathcal{N}_{k1}$ and $\mathcal{N}_{k2}$ are denoted as $\mathbf{O}_{k1}$ and $\mathbf{O}_{k2}$. Since the true surface normals in the overlapping region Θ are consistent, *i.e.*, $\mathbf{O}_{k1,k2} = \mathbf{I}$, the pseudo surface normals of two patches in this region can be aligned by

$$\hat{\mathbf{O}}_{k1,k2} = \mathbf{O}_{k2}^{\top} \mathbf{O}_{k1,k2} \mathbf{O}_{k1} = \mathbf{O}_{k2}^{\top} \mathbf{O}_{k1}. \tag{3.9}$$

Obviously $\hat{\mathbf{O}}_{k1,k2} \in O(3)$ encodes the relationship between unknown orthogonal ambiguities of the two patches. We name it *relative orthogonal transformation* and it can

be solved by aligning pseudo surface normals in the overlapping regions, *i.e.*,

$$\hat{\mathbf{O}}^*_{k1,k2} = \underset{\hat{\mathbf{O}}_{k1,k2}}{\operatorname{argmin}} \|\hat{\mathbf{O}}_{k1,k2}\hat{\mathbf{N}}^\top_{k1}(\Theta) - \hat{\mathbf{N}}^\top_{k2}(\Theta)\|^2_F,$$
$$\text{s.t.} \quad \hat{\mathbf{O}}_{k1,k2} \in O(3). \tag{3.10}$$

Equation (3.10) is an Orthogonal Procrustes problem and we follow Gower *et al.* [97] to solve $\hat{\mathbf{O}}_{k1,k2}$.

Besides finding consistent surface normal pairs via spatial overlaps, the existing method [98] shows that pixels with strong correlation in their intensity profiles (an ordered sequence of scene irradiance at a pixel across images) have the same surface normals. This observation is proved to be valid under distant directional lighting. However, given natural illumination, correlated intensity profiles do not necessarily lead to consistent surface normals. A counter-example is a constant environment map, *i.e.*, $L(\boldsymbol{\omega}) = c$, under which all surface normals have correlated intensity profiles.

To approximately extend intensity profile constraint to natural lighting, we propose a *consistent orthogonality condition*. For two disconnected scene points $\mathbf{q}$ and $\mathbf{s}$ as shown in Fig. 3.5(a), *if their surface normals and equivalent directional lighting can be transformed by an orthogonal matrix simultaneously, i.e.,*

$$\mathbf{O}_{k2,k3} \begin{bmatrix} \mathbf{n}_{k2}(\mathbf{q}) & \mathbf{L}_{k2}(\mathbf{q}) \end{bmatrix} = \begin{bmatrix} \mathbf{n}_{k3}(\mathbf{s}) & \mathbf{L}_{k3}(\mathbf{s}) \end{bmatrix}, \tag{3.11}$$

where $\mathbf{O}_{k2,k3} \in O(3)$ is the orthogonal transformation, *then $\mathbf{O} = \mathbf{I}$ and both surface normals and equivalent directional lighting for $\mathbf{q}$ and $\mathbf{s}$ should be consistent.* Please refer to Sec. 3.6.1 for a detailed analysis.

However, what we know from Sec. 3.3.2 are pseudo equivalent lighting and surface normals of scene points, with unknown orthogonal ambiguities to the corresponding ground truth. So we extend the consistent orthogonality condition to the pseudo normal and lighting case. Assume the pseudo equivalent directional lighting and surface

normals of $\mathbf{q}$ and $\mathbf{s}$ can be aligned by an orthogonal matrix $\hat{\mathbf{O}}_{k2,k3}$, *i.e.*,

$$\hat{\mathbf{O}}_{k2,k3} \begin{bmatrix} \hat{\mathbf{n}}_{k2}(\mathbf{q}) & \hat{\mathbf{L}}_{k2}(\mathbf{q}) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{n}}_{k3}(\mathbf{s}) & \hat{\mathbf{L}}_{k3}(\mathbf{s}) \end{bmatrix}. \tag{3.12}$$

Following the relationship shown in Fig. 3.5(c), the truth surface normals and equivalent lighting between $\mathbf{q}$ and $\mathbf{s}$ can be simultaneously aligned by

$$\mathbf{O}_{k2,k3} = \mathbf{O}_{k3}\hat{\mathbf{O}}_{k2,k3}\mathbf{O}_{k2}^{\top} \in O(3), \tag{3.13}$$

where $\mathbf{O}_{k2}$ and $\mathbf{O}_{k3}$ are the orthogonal ambiguities of the surface patches $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$ covering point $\mathbf{q}$ and $\mathbf{s}$. Since Eq. (3.13) makes the consistent orthogonality condition true ($\mathbf{O}_{k2,k3} \in O(3)$), the truth surface normals at scene points $\mathbf{q}$ and $\mathbf{s}$ are consistent, *i.e.*, $\mathbf{O}_{k2,k3} = \mathbf{I}, \langle \mathbf{n}_{k2}(\mathbf{q}), \mathbf{n}_{k3}(\mathbf{s}) \rangle = 0$. Therefore, Eq. (3.12) is an extended consistent orthogonality condition to cluster consistent surface normals from pseudo surface normals and pseudo equivalent directional lighting.

Based on Eq. (3.3), if the surface normals and equivalent lighting of two scene points fit to the consistent orthogonality condition, their intensity profiles are correlated. Therefore, to cluster consistent normals on the whole surface, we first filter scene point pairs with correlated intensity profiles, and then check whether the pseudo surface normals and equivalent directional lighting of each filtered point pair satisfy Eq. (3.12).

Similar to $\hat{\mathbf{O}}_{k1,k2}$, $\hat{\mathbf{O}}_{k2,k3} = \mathbf{O}_{k3}^{\top}\mathbf{O}_{k2}$ also encodes the relationship of orthogonal ambiguities between two surface patches. To calculate relative orthogonal transformation between patch $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$, we minimize the following energy function as an Orthogonal Procrustes problem [97].

$$\hat{\mathbf{O}}_{k2,k3}^{*} = \underset{\hat{\mathbf{O}}_{k2,k3}}{\operatorname{argmin}} \|\hat{\mathbf{O}}_{k2,k3}\mathbf{D}_{k2}(\mathbf{q}) - \mathbf{D}_{k3}(\mathbf{s})\|_{F}^{2},$$
$$\text{s.t.} \quad \hat{\mathbf{O}}_{k2,k3} \in O(3), \tag{3.14}$$

where $\mathbf{D}_{k2}(\mathbf{q}) = [\hat{\mathbf{n}}_{k2}(\mathbf{q}) \quad \hat{\mathbf{L}}_{k2}(\mathbf{q})]$ and $\mathbf{D}_{k3}(\mathbf{s})$ follows the same definition.

To summarize, we collect consistent surface normal pairs from overlapping patch regions and scene points satisfying consistent orthogonality conditions. Based on these

consistent surface normal pairs, we extract the relative orthogonal transformations which describe the relationship between unknown patch-wise orthogonal ambiguities. All relative orthogonal transformations form a set $\mathcal{S}_o$, which will be used as edges in the following orthogonal ambiguity graph building process.

3.4.2   Constructing orthogonal ambiguity graph

We create an orthogonal ambiguity graph $\mathcal{G}_o = \{\mathcal{V}, \mathcal{E}\}$ (where $\mathcal{V}$ is the set of all nodes and $\mathcal{E}$ is the set of all edges connecting nodes) to build the connections among patches. As shown in Fig. 3.2, the nodes of a orthogonal ambiguity graph are filled with the unknown orthogonal ambiguities $\mathbf{O}_k$ from all patches. The relationship between orthogonal ambiguities can be represented by relative orthogonal transformations, as shown in Eq. (3.9). Therefore we apply all the elements in relative orthogonal transformation set $\mathcal{S}_o$ to build the edges $\mathcal{E}$ of $\mathcal{G}_o$. Given surface normal estimation confidence c_i and c_j of patch $\mathcal{N}_i$ and $\mathcal{N}_j$ calculated from Eq. (3.8), we further define the edge confidence as $c_{i,j} = c_i c_j$. Intuitively, if the normal estimations of two patches are reliable, we tend to trust the relative orthogonal transformation between them.

Based on the orthogonal ambiguity graph, we optimize the patch-wise orthogonal ambiguities via the following minimization:

$$
\mathbf{O}_1^*, \cdots, \mathbf{O}_p^* = \underset{\mathbf{O}_1, \cdots, \mathbf{O}_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} \mu(\mathbf{O}_j^\top \mathbf{O}_i - \mathbf{O}_{i,j}),
$$
$$
\text{s.t.} \quad \mathbf{O}_i \in O(3). \tag{3.15}
$$

where $\mu(\cdot)$ is a distance measure between two orthogonal matrices in $O(3)$. Directly solving Eq. (3.15) is non-trivial, so we decompose the orthogonal ambiguity $\mathbf{O}$ into two parts: binary ambiguity $d = |\mathbf{O}| \in \{+1, -1\}$ and rotation ambiguity $\mathbf{R} \in SO(3)$. Correspondingly, the orthogonal ambiguity graph can also be divided into binary ambiguity graph $\mathcal{G}_b$ and rotation ambiguity graph $\mathcal{G}_r$ as shown in Fig. 3.2. Based on these two graphs, we recover the patch-wise orthogonal ambiguities by solving their binary ambiguity part and rotation ambiguity part one after another.

3.4.3   Optimizing binary ambiguity graph

In binary ambiguity graph $\mathcal{G}_b$, the node value d_i and the edge value $d_{i,j}$ are calculated from the determinate of the orthogonal ambiguity $\mathbf{O}_i$ and the relative orthogonal transformation $\mathbf{O}_{i,j}$, respectively. Following Eq. (3.15), the binary ambiguities existing in nodes should satisfy

$$\{d_1^*, \cdots, d_p^*\} = \underset{d_1, \cdots, d_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} (d_i d_j - d_{i,j})^2,$$

$$\text{s.t.}\ \ d_i \in \{-1, 1\}. \tag{3.16}$$

Equation (3.16) can be interpreted as assigning each node of the undirected graph $\mathcal{G}_b$ a label defined on $\{-1, 1\}$. Therefore we formulate the problem as maximizing a posteriori estimation of binary MRF [99], with the energy function defined as

$$E(d) = \sum_{i \in \mathcal{V}} E_1(d_i) + \eta \sum_{i,j \in \mathcal{E}} E_2(d_i, d_j),$$

$$\text{s.t.}\ \ d_i \in \{-1, 1\}, \tag{3.17}$$

where coefficient η is used to balance the data term E_1 and the smoothness term E_2, i and j represent node indexes. We define the node with maximum degrees in $\mathcal{G}_b$ as the root node and set its binary ambiguity as 1, then our data term is defined as

$$E_1(d_i) = \begin{cases} \infty & d_i = -1, i = r \\ 0 & \text{others} \end{cases}, \tag{3.18}$$

where r is the index of the root node. Following Eq. (3.16), we define the smoothness term as

$$E_2(d_i, d_j) = \begin{cases} \infty & d_i d_j \neq d_{i,j} \\ 1 - c_{i,j} & d_i d_j = d_{i,j} \end{cases}, \tag{3.19}$$

where $c_{i,j}$ is the confidence of the edge connecting i-th and j-th node. Given the definition of the data term and the smoothness term, we minimize the energy function Eq. (3.17) with TRW-S algorithm [100]. Note that, since the binary ambiguity in the

root node could be either -1 or 1, the solved binary ambiguities in all nodes can only be optimized up to a global binary ambiguity.

3.4.4   Optimizing rotation ambiguity graph

With binary ambiguity solved, the orthogonal ambiguity in each node is reduced to rotation ambiguity. Guided by Eq. (3.15), we solve the rotation ambiguity via the following optimization:

$$\{\mathbf{R}_1^*, \cdots, \mathbf{R}_p^*\} = \underset{\mathbf{R}_1, \cdots, \mathbf{R}_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} \chi(\mu(\mathbf{R}_j^\top \mathbf{R}_i, \mathbf{R}_{i,j})),$$

$$\text{s.t.} \quad \mathbf{R}_i \in SO(3),$$

where $\mu(\cdot)$ is a distance measure between two rotations in $SO(3)$ and $\chi(\cdot)$ is a loss function defined over this distance measure. This optimization belongs to the rotation averaging problem [60]. Similar to Sec. 3.4.3, we fix the rotation ambiguity of the root node as identity, and follow Chatterjee *et al.* [101] to optimize the rotation ambiguity in each node. During the rotation averaging optimization, we apply geodesic distance measurement for $\mu(\cdot)$ and choose Cauchy loss function rather than ℓ_2 loss function for $\chi(\cdot)$ to improves the robustness when outliers exist in relative rotation transformation $\mathbf{R}_{i,j}$. Since the true rotation ambiguity of the root node is unknown, we can only solve per-patch rotation ambiguities up to a global rotation ambiguity.

After solving rotation ambiguities, we rotate all the patch-wise pseudo surface normals and average the normals in the overlapping regions to get a complete pseudo surface normal map $\hat{\mathbf{N}}$. Compared to the ground truth, pseudo surface normal map $\hat{\mathbf{N}}$ has two ambiguities left: a global binary ambiguity and a global rotation ambiguity. We combine the two ambiguities as a global orthogonal ambiguity $\mathbf{O}_g$.

So far, estimated $\hat{\mathbf{N}}$ contains only a global ambiguity w.r.t. the true surface normal map. This ambiguity can be reduced to a convex/concave ambiguity by forcing integrability constraint as suggested in [65]. The corresponding proof and the detailed steps for estimating the global ambiguity can be found in the Sec. 3.6.2. The
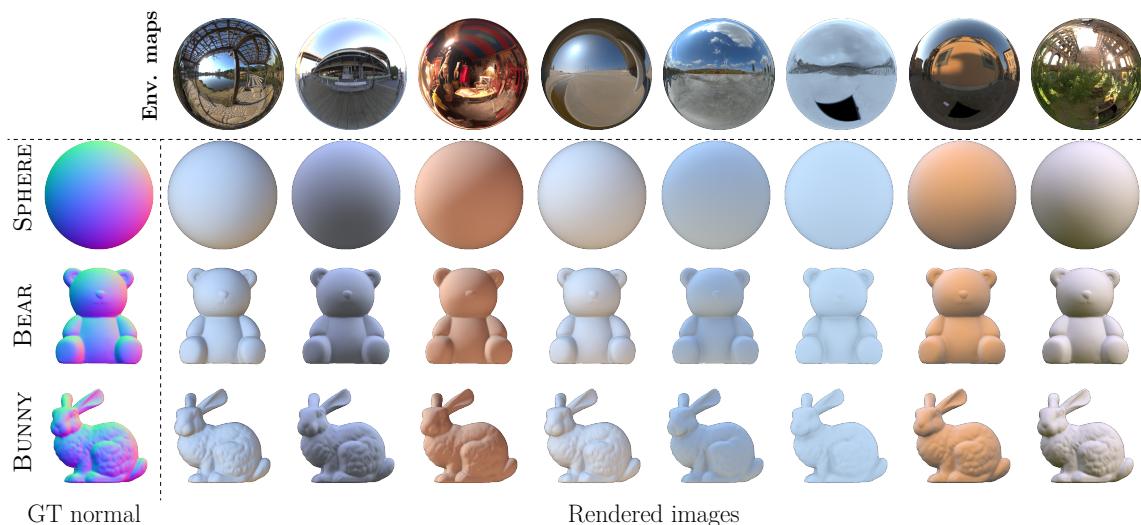
Fig. 3.6 Synthetic dataset. Environment maps (visualized as light probes) from sIBL Archive are shown in the top row. Below we show ground truth normals for three objects in the first column and examples of rendered images in other columns corresponding to the environment maps above.

remained binary convex/concave ambiguity in our surface normal estimation result could be easily removed manually.

3.5 Experimental results

We first use synthetic data to verify the quantitative accuracy of our method, followed by a comparison between the newly proposed graph-based patch merging method (GPM) and previous matrix-based patch merging method (MPM) [62]. Finally, we show a comparison with existing methods on real-world data.

3.5.1 Synthetic data setup

We collect 31 real-world environment maps from the sIBL Archive[1] as natural illumination sources, which include diverse natural illumination from both indoor and outdoor scenarios. We use Blender [102] as rendering engine and choose three objects – SPHERE, BEAR (from [68]) and BUNNY (with increasing geometric complexity) – to

---

[1]http://www.hdrlabs.com/sibl/archive.html
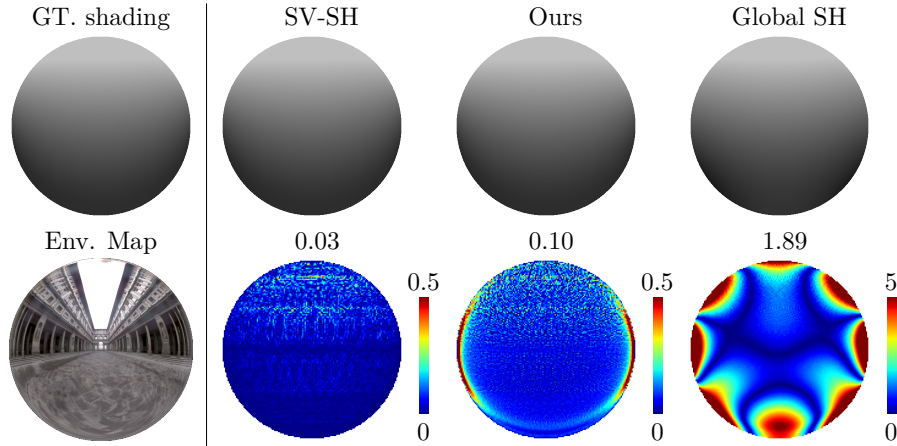
Fig. 3.7 Comparisons between different environment lighting approximation model shown in Table 3.1. The top row shows the shading maps and the bottom row provides the absolute error maps and the mean absolute error value of approximated shadings.

render Lambertian reflectance with white albedo under natural illumination. The image resolution of the three objects is fixed to 160×160. Ground truth surface normals and sample images in our synthetic dataset are shown in Fig. 3.6.

3.5.2   Representation power of lighting model

As shown in Table 3.1, the lighting models used in existing methods include global SH [54] and SV-SH [93]. We compare the representation power of these two models with our SV-directional equivalent lighting model. Figure 3.7 shows the comparison on an example environment map and its corresponding shading under Lambertian reflectance. Taking the ground-truth shading and surface normal as input, we extract 3×3 patches and calculate our equivalent lighting direction for each patch. Then we assign it as our approximated lighting direction at the patch center. To compare with SH-based lighting models, we render the shading map with our approximated lighting directions as shown in the third column of Fig. 3.7. At the same time, we calculate the second-order global SH lighting coefficients to approximate the shading given the ground truth surface normal. We also divide the image into patches and estimate SV-SH lighting to approximate the patch shading in a similar manner to our

SV-directional model. The absolute error maps w.r.t. the ground-truth shading are shown in the second row of Fig. 3.7. For representing the natural light in a local patch, the SV-SH model [93] applies spherical harmonic lighting, which is more flexible than a single directional light [56] used in our SV-directional model. Therefore, the SV-SH model has relatively higher approximation accuracy compared to ours. On the other hand, both models are more accurate than the global SH model [54] as the natural illumination is spatially-varying.

3.5.3   Lighting model verification

The local surface normal estimation in our method requires the illumination on a patch to be directional light. Theoretically, if the surface normals within a patch have the same direction, its natural illumination is equivalent to a single directional light. However, when surface patches contain diverse normal directions, it is unclear whether a single equivalent lighting direction represents the patch illumination accurately. In Sec. 3.3.1, we have defined the mean angular difference of surface normals (denoted as $v_{\mathbf{n}}$) to evaluate the normal variations in a local patch. Similarly, we can also define the mean angular difference of equivalent lighting directions (denoted as $v_{\bar{\mathbf{l}}}$) corresponding to the patch surface normals. This metric can be seen as the error using a single equivalent lighting direction to approximate the natural illumination within the patch. As shown in Fig. 3.8, we provide a statistic analysis for $v_{\bar{\mathbf{l}}}$ w.r.t. that of surface normals $v_{\mathbf{n}}$ on local patches. The green bar indicates the median value, the top and bottom bounds of the black box indicate the first and third quartile values, and the top and bottom ends of the vertical blue line indicate the minimum and maximum mean angular difference of equivalent lighting direction. Generally, if a local patch has a larger normal variation, its illumination is less accurately approximated by a single equivalent lighting direction.

We also investigate the influence on local surface normal estimation accuracy when we treat the patch illumination as an equivalent directional light. Given a surface patch, we first approximate the patch illumination with the equivalent lighting direc-
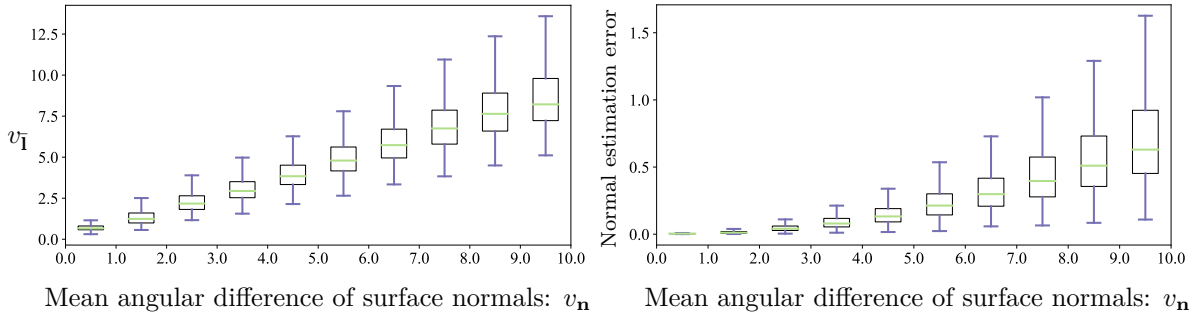
Fig. 3.8 Evaluation of equivalent lighting model. Left row provides the mean angular difference of equivalent lighting directions $v_{\bar{\mathbf{l}}}$ within a 3×3 patch w.r.t. that of surface normals $v_{\mathbf{n}}$ in the corresponding range shown in x-axis. The right row provides the mean angular error of patch normal estimation w.r.t. $v_{\mathbf{n}}$.

tion of the patch center, then estimate the patch surface normals with this approximated lighting and calculate the mean angular error w.r.t. the ground truth. The second row of Fig. 3.8 shows the statistic summary of patch surface normal estimation error w.r.t. the mean angular difference of patch surface normals. Although larger surface normal variation will make our equivalent lighting model approximation more difficult, the patch surface normal estimation errors from the approximated lighting direction remain at a low level ((mean angular error $< 1.5°$).

3.5.4   Performance under varying lighting conditions

We provide the evaluation of Ours (MPM) [62] and Ours (GPM) on synthetic data under varying numbers of environment lights. As shown in Fig. 3.9, we select 10 and 15 subsets of environment maps out of 20 in sIBL dataset to test how the normal estimation accuracy varies with lighting conditions. From the angular error distributions of the SPHERE object, by increasing the image observations under varying natural lights, the surface normal estimation errors become smaller. The table shown in Fig. 3.9 further provides the evaluation of MPM [62] and GPM on all three synthetic objects. The error values become larger for the BEAR and BUNNY compared to the smooth shape of SPHERE, which is caused by the difficulty in approximating equivalent directional lighting on shape patches with rapid normal variation. The

Object	Method	$f = 10$	$f = 15$	$f = 20$
Sphere	Ours (MPM)	8.87	6.42	3.97
	Ours (GPM)	4.47	3.29	**2.75**
Bear	Ours (MPM)	17.13	15.20	14.78
	Ours (GPM)	8.93	8.39	**7.89**
Bunny	Ours (MPM)	18.09	16.78	15.21
	Ours (GPM)	10.44	9.01	**8.69**
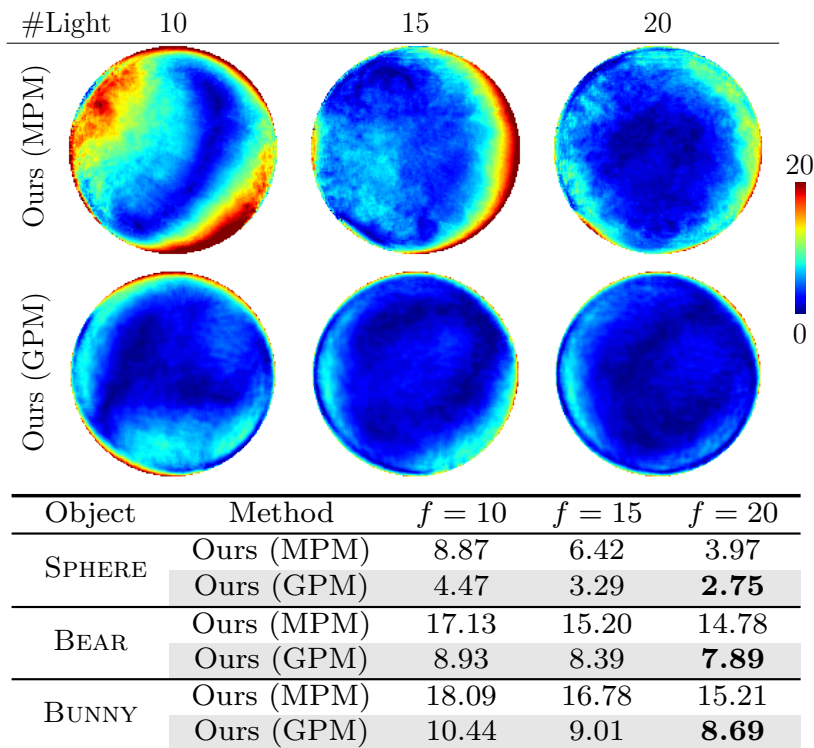
Fig. 3.9 Comparison between different patch merging methods (MPM & GPM) under varying numbers of lights (10, 15, and 20). The top two rows show the angular error distributions from Ours (MPM) and Ours (GPM) of the Sphere object. The table below provides the mean angular errors (in degree) w.r.t. the ground truth.

mean angular errors shown in the table tell that generally a larger number of input images and more diverse lighting distributions lead to more accurate surface normal recoveries. We have also tried further increasing the number of environment maps up to 31, but the improvement is rather unobvious, so we fix the number of input images as 20 for the experiments on synthetic data hereafter.

We also observe that under varying lighting conditions and object shapes, Ours (GPM) has a smaller mean angular error compared to Ours (MPM) [62]. It verifies that compared to the local shortest path searching strategy used in MPM, GPM's global optimization on all connections among patches via MRF optimization and rotation averaging can achieve more accurate surface normal estimation results.
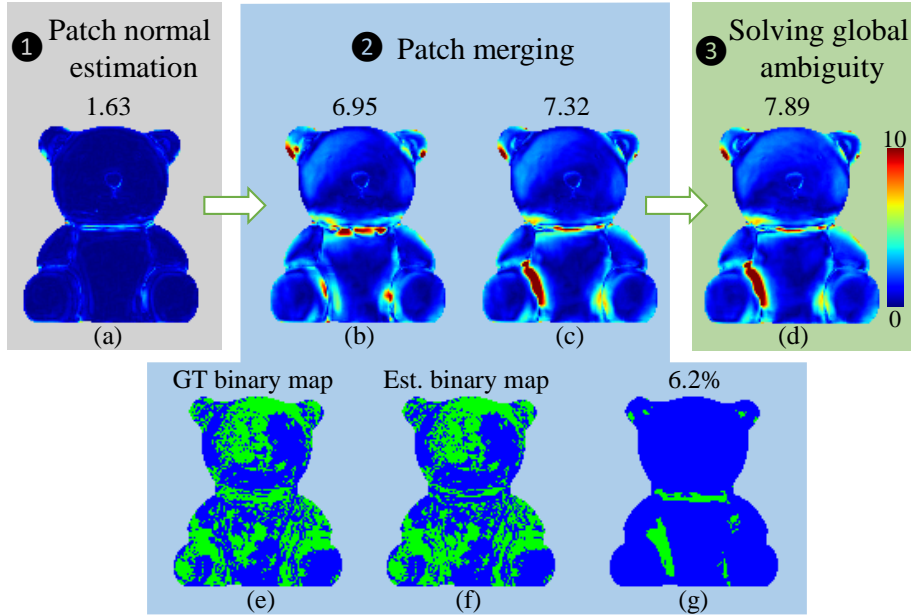
Fig. 3.10 Ablation study on the BEAR case. The top row shows the error map of normal estimates at each stage of Ours (GPM). The bottom row gives the binary ambiguity estimation, where the blue and green pixels correspond to binary ambiguity $\{1, -1\}$, respectively. The difference between (e) and (f) is shown in (g). About 6.2% patches have wrong binary ambiguity estimation.

3.5.5 Ablation study

As shown in Fig. 3.2, our method mainly contains three stages: local surface normal estimation, patch merging including MRF optimization and rotation averaging, and global ambiguity determination. Taking the BEAR as an example, we analyze the error of estimated surface normal maps from each stage.

In the first stage, the patch-wise surface normals are solved up to local ambiguities. We first resolve these orthogonal ambiguities by aligning the estimated surface normal to the ground truth in each patch and then merge aligned patch normals to build a complete surface normal map. The angular error map of this surface normal map shown in Fig. 3.10(a) verifies that normal estimation error brought by the first stage is 1.63°. We can see that inaccurate local surface normal estimates mainly occur at regions with large normal variations, such as the neck and leg parts of the BEAR, where the equivalent lighting approximations are inaccurate. These inaccurate

local surface normal estimates further influence the rotation ambiguity and binary ambiguity estimation in the second stage. Figure 3.10(b) and (c) show the angular error maps of GPM's surface normal estimation (up to a global ambiguity) with estimated and the ground truth binary ambiguities shown in Fig. 3.10(e) and (f). After resolving the patch-wise orthogonal ambiguities with MRF optimization and rotation averaging, surface normal estimation error has increased to 7.32°, in which MRF optimization contributes 0.37° and rotation averaging contributes 5.32°. As shown in Fig. 3.10(g), 6.2% of the patches have wrong binary ambiguity estimation and they are mainly distributed at regions with inaccurate local normal estimates. Our complete solution achieves 7.89°, where the error brought by solving the global ambiguity with integrability is 0.57°.

3.5.6   Comparison with existing methods

Based on the synthetic dataset shown in Fig. 3.6, we compare our method with existing methods [54] (denoted as "HY19") and [3] (denoted as "HP19") on surface normal estimation. HY19 [54] is the state-of-the-art method for uncalibrated photometric stereo under natural lighting. HP19 [3] also recovers detailed shape from a rough depth map taking image observations under unknown natural light as reference. As both HY19 [54] and HP19 [3] require depth initialization, we apply OT12 [90] to generate rough shapes from input images and use them to initialize HY19 [54] and HP19 [3]. As shown in Fig. 3.11, the surface normal estimates from HP19 [3] are influenced by its initial shape generated by visual hull [90]. HY19 [54] returns better results from the shape initialization. However, its normal estimates' accuracy is still limited by the global SH lighting approximation. In comparison, Ours (GPM) adopts a more flexible SV-directional equivalent lighting model and merges local patches by optimizing all patch connections, therefore achieves the most accurate surface normal recoveries among all the methods.
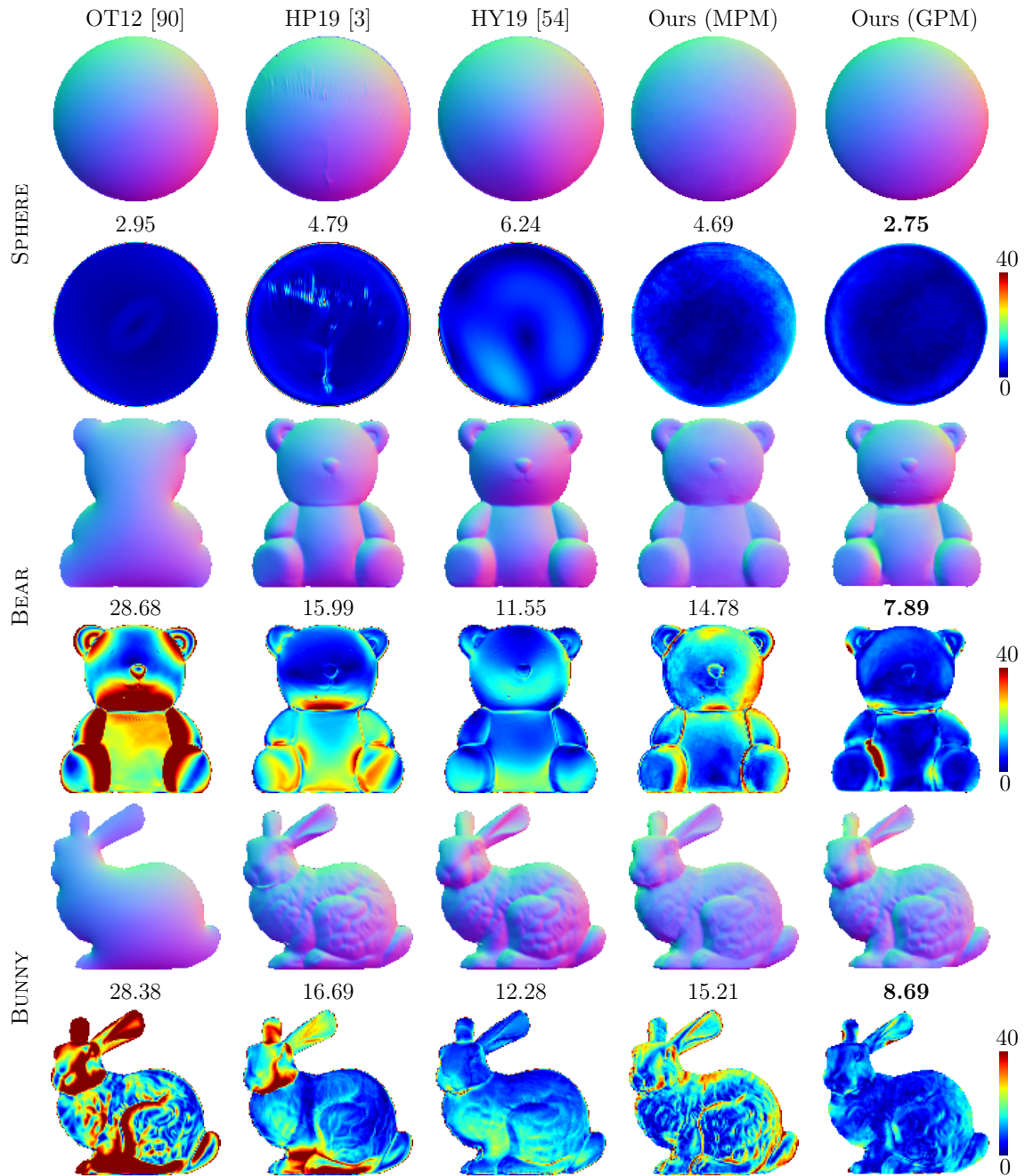
Fig. 3.11 Comparison with existing methods on synthetic data shown in Fig. 3.6. Even rows show the estimated surface normal maps, and odd rows show the corresponding angular error maps and the mean angular error values in degree.

3.5.7   Influence of SV-albedos and shadows

We have tested our method on objects with uniform albedo. Our method can also be applied to objects with spatially-varying albedos as long as abrupt albedo changes are not observed within the patch. In Fig. 3.12, the image observations of the first two columns are rendered with non-uniform albedo maps. Compared to the normal estimates for uniform albedo shown in Fig. 3.11, the piecewise constant albedo distribution only increases the estimation error from 8.69° to 9.31°. These errors are mainly brought by the patches across the albedo variation edges. Our method fails to output accurate surface normal estimates (error becomes 34.70°) for general spatially-varying albedo distribution as uniform albedo assumption is not valid for all patches.

We also evaluate the influence of cast shadows. Comparing the error distribution with and without the cast shadow, we observe less accurate surface normal estimates (mean angular error increases to 11.28°) around the Bunny's neck and foot regions, where cast shadows bring errors to the local surface normal estimation. On the other hand, when the environment maps include abrupt changes such as high-frequency light sources, the lighting directions for a surface patch have more variations as the visibility hemispheres of two surface normals may include/exclude a high-frequency (the extreme case is a single point light source) light source. As shown in the last column of Fig. 3.12, we add 100 small synthetic point light sources to all of the 20 environment maps. As the illumination for a local patch cannot be treated as equivalent directional lighting, the normal estimation error has increased from 8.69° to 14.59°.

3.5.8   Time and memory consuming

Table 3.2 shows the memory and time usage of our method with two different patch merging strategies on the three synthetic data shown in Fig. 3.6. Given p nodes and q edges of the orthogonal ambiguity graph, the memory usage in the GPM is $9(p+q)$ because the 3D orthogonal matrix has 9 elements. Since the nodes are locally connected in our orthogonal ambiguity graph, it's reasonable to assume $q = kp, k \ll p$,
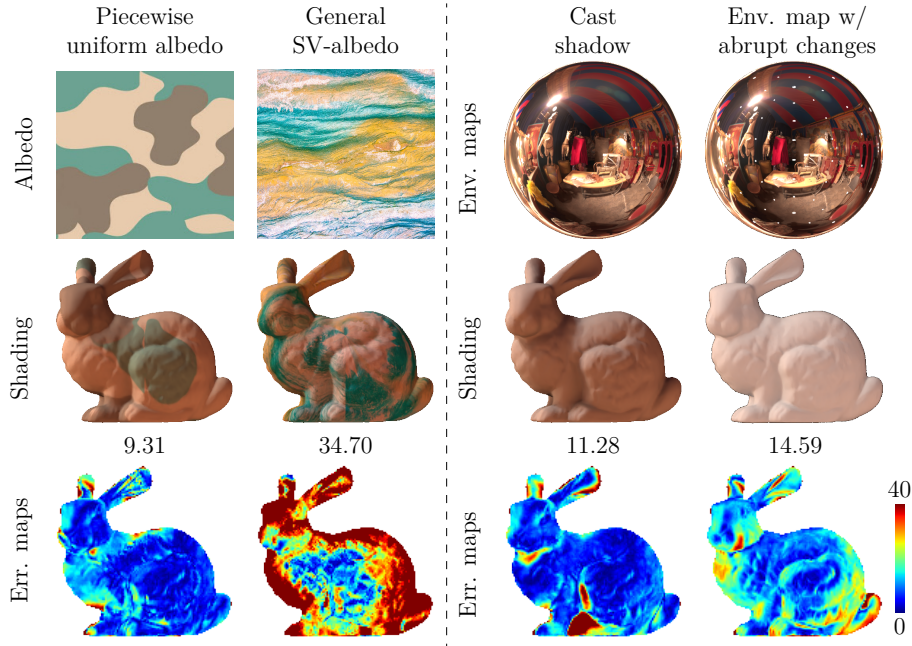
Fig. 3.12 The accuracy of Ours (GPM) is influenced by non-uniform albedos, cast shadows and environment maps with abrupt changes.

therefore the memory complexity in GPM is belong to $\mathcal{O}(9p + 9kp) = \mathcal{O}(p)$. On the other hand, given p pixels, the angular distance propagation matrix has a dimension of $p \times p$, which leads the memory complexity in MPM [62] to be $\mathcal{O}(p^2)$. Therefore, our GPM is more memory efficient compared to MPM [62]. Besides, the angular propagation matrix in MPM is built from the shortest path searching between every surface normal pairs, which leads to the time complexity $C_p^2 \mathcal{O}(p \log(p)) = \mathcal{O}(p^3 \log(p))$. The time complexity in GPM is related to the iterations of rotation averaging and MRF optimization. Therefore it is hard to represent the time cost in a theoretical way. From the experiments on the three objects shown in Table 3.2, GPM runs about 3 times faster than MPM [62] in average.

3.5.9 Experiments on real-world dataset

In this section, we evaluate and compare our method using real-world data. We first introduce three real datasets and their data capturing process for recording image data under varying natural illuminations. Then we give the shape estimation results from

Table 3.2 Comparison on Memory and Time Usage.

	Method	SPHERE	BUNNY	BEAR	Average
Memory Usage (MB/pix)	Ours (GPM)	**0.029**	**0.041**	**0.043**	**0.038**
	Ours (MPM)	0.157	0.189	0.175	0.174
Time Usage (ms/pix)	Ours (GPM)	**11.64**	**11.14**	**11.16**	**11.31**
	Ours (MPM)	52.35	27.20	28.07	35.87

existing methods and ours.

Real dataset   We choose three real-world dataset from [2] (denoted as "YY13"), [1] (denoted as "HJ15") and HP19 [3]. To obtain image data under varying natural illuminations, YY13 [2] fixes the relative position between the target objects and the camera while moving the whole setup to different places with different natural environments; HJ15 [1] captures object appearance within one day (6 hours) in an outdoor environment; HP19 [3] records images under an uncalibrated indoor environment and changes the natural lighting condition by additionally moving a point light source. The datasets from YY13 [2] and HJ15 [1] have a mirror sphere to calibrate the environment maps, but such information is not used in our method.

Experimental results   Surface normal estimation results using OWL (66 images) object from [1] is shown in Fig. 3.13. We use the ground truth normal provided by the authors and make a quantitative comparison with existing methods. Surface normal estimates from HP19 [3] and HY19 [54] have large mean angular errors due to the inaccurate initial shape from OT12 [90]. The result obtained from MPM [62] generally looks noisier, but it is quantitatively better than the calibrated result from HJ15 [1], especially in local regions near the OWL's eyes where HJ15 [1] shows large errors. Compared to MPM [62], GPM further improves the surface normal estimation accuracy at the body and contour region of the OWL, since all the local patch connections are globally optimized during the MRF optimization and rotation averaging.

We show the shape estimation results using HORSEHEAD (7 images), CHEF (multi-albedo, 7 images), and MOTHER&BABY (10 images) objects from YY13 [2] in Fig. 3.14.
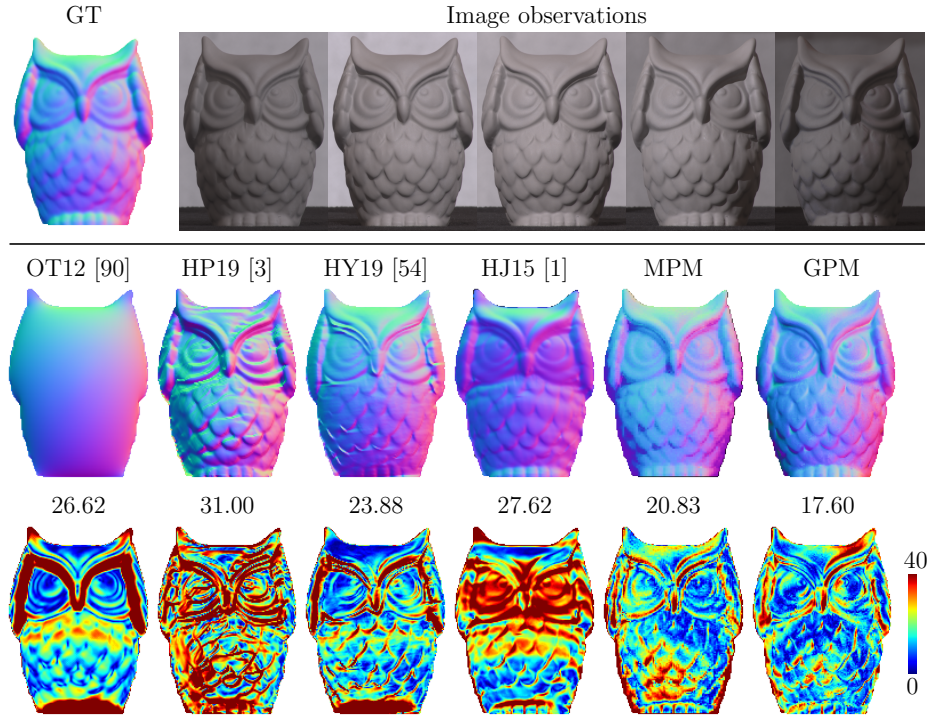
Fig. 3.13 Quantitative comparison on real data from HJ15 [1]. The numbers on the top of error maps are mean angular errors.

Since we do not have the ground truth for these data, we can only qualitatively compare our results with them by integrating estimated normal fields to the depth map with [44]. Referring to the geometry recovering results, the shape of the HORSEHEAD and the CHEF recovered from YY13 [2] are near flat, and the estimated head part of the MOTHER&BABY by YY13 [2] is distorted. Compared to YY13 [2], both MPM and GPM produce more visually plausible shape estimates without knowing anything about the lighting conditions. Besides, on the neck part of the HORSEHEAD, the body part of the CHEF and the arm and baby part of the MOTHER&BABY, GPM further outperforms MPM [62].

Figure 3.15 shows the shape recovery results of 8 challenging scenes from HP19 [3], where each data includes 20 high resolution (1280×720) images captured under uncalibrated daylight and moving point light sources. PH17 [52], HP19 [3] and HY19 [54] directly produce depth output with given initial object shape, and achieve visually plausible shape recoveries on all the eight scenes. Compared with these three meth-
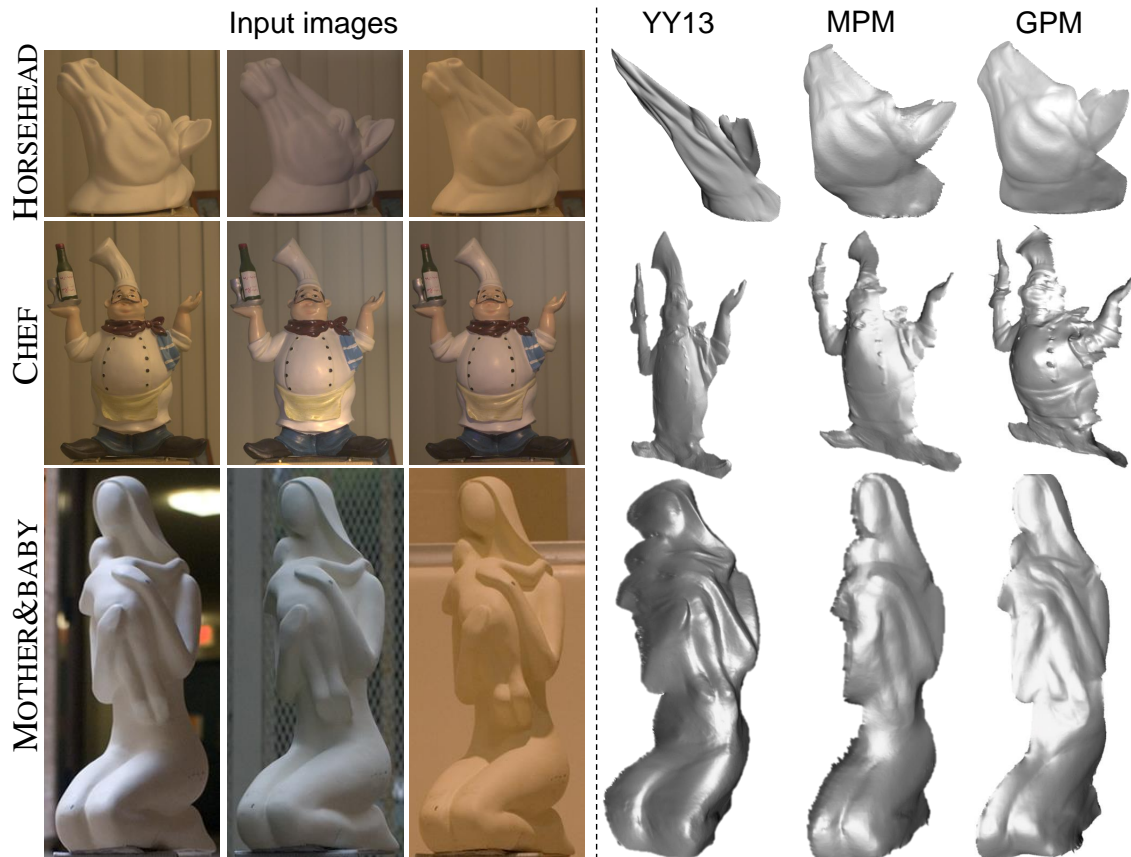
Fig. 3.14 Qualitative comparison on real data from YY13 [2]. Viewpoints are adjusted to emphasize the shape distortions.

ods, MPM and GPM are free of depth initialization. Following Quéau *et al.* [44], we integrate estimated surface normal from GPM and MPM to the depth. The results show that both methods obtain reasonable results on FACE1, FACE2 and BACKPACK. Compared to the state of the art HY19, our GPM produces comparable shape estimation in the case of SHIRT and even better result on the VASE by providing richer geometry details. The TABLETCASE and OVENMITT are two flat objects with SV-albedos. Surface normals of these two objects have nearly the same directions within a local patch, which is a degenerate case for solving uncalibrated photometric stereo (Sec. 3.3.2). Also, the complex shape of RUCKSACK brings large surface normal variations in local patches, which violates the equivalent directional lighting assumption. Therefore, both GPM and MPM fail on these three scenes.
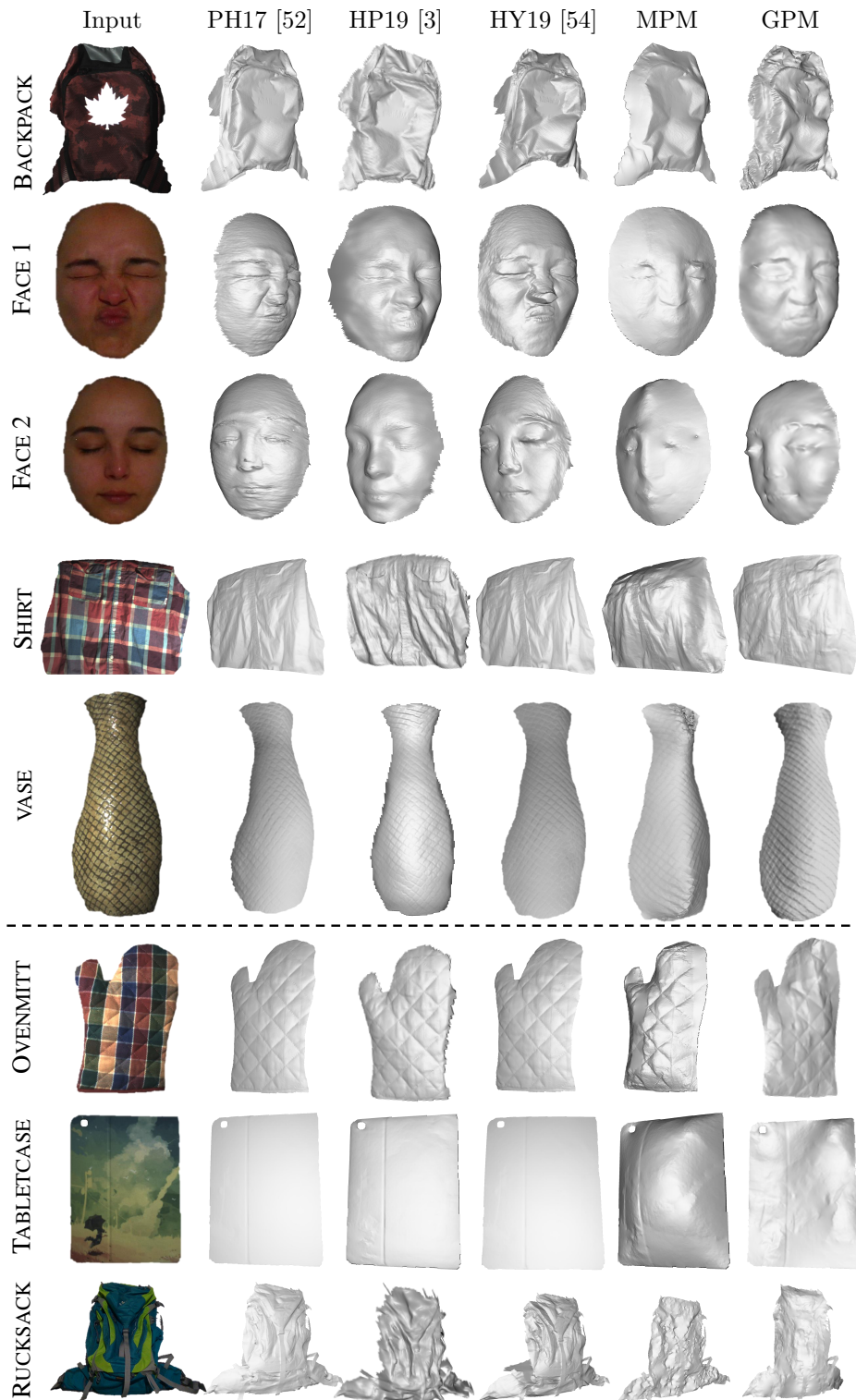
Fig. 3.15 Qualitative comparison on real data from HP19 [3]. GPM achieves comparable results with PH17 [52], HY19 [3] and HP19 [54] on scenes above the dotted line. The three scenes below the dotted line are the failure cases for both GPM and MPM. Note that PH17 [52], HY19 [3] and HP19 [54] require a shape prior as initialization, while GPM and MPM avoid that requirement and directly estimate surface normal from images.

3.6   Discussion

In this section, we will present the details of consistent surface normal clustering via intensity profiles, method of resolving the global orthogonal ambiguity, and the robust strategy for local surface normal estimation of planar patches.

### 3.6.1   Consistent orthogonal condition

Assume surface normals at two scene points $\mathbf{q}$ and $\mathbf{s}$ are $\mathbf{n}(\mathbf{q})$ and $\mathbf{n}(\mathbf{s})$. The equivalent lighting of these two scene points under f environment maps are

$$
\begin{aligned}
\bar{\mathbf{l}}^t(\mathbf{q}) &= \int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega}, \\
\bar{\mathbf{l}}^t(\mathbf{s}) &= \int_{\Omega(\mathbf{s})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega},
\end{aligned}
\tag{3.20}
$$

where $L^t(\boldsymbol{\omega}) : \mathbb{R}^3 \to \mathbb{R}$ represents light intensity of t-th environment map under a unit direction $\boldsymbol{\omega} \in \mathbb{S}^2 \subset \mathbb{R}^3$, $\Omega(\mathbf{q})$ and $\Omega(\mathbf{s})$ are the visible hemispheres corresponding to the two surface normals,

$$
\begin{aligned}
\Omega(\mathbf{q}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\boldsymbol{\omega} \geq 0\}, \\
\Omega(\mathbf{s}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{s})\boldsymbol{\omega} \geq 0\}.
\end{aligned}
\tag{3.21}
$$

As shown in Fig. 3.16(d-e), Since $\mathbf{n}(\mathbf{q})$ and $\mathbf{n}(\mathbf{s})$ are two unit direction, there must exists an orthogonal matrix $\mathbf{O} \in O(3)$ such that $\mathbf{O}\mathbf{n}(\mathbf{q}) = \mathbf{n}(\mathbf{s})$. As a result, the two corresponding visible hemispheres can also be aligned by $\mathbf{O}$,

$$
\begin{aligned}
\Omega(\mathbf{s}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{s})\boldsymbol{\omega} \geq 0\} \\
&= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\mathbf{O}^\top\boldsymbol{\omega} \geq 0\} \\
&= \{\mathbf{O}\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\boldsymbol{\omega} \geq 0\}.
\end{aligned}
\tag{3.22}
$$

In other word, for any $\boldsymbol{\omega} \in \Omega(\mathbf{q})$, $\mathbf{O}\boldsymbol{\omega} \in \Omega(\mathbf{s})$. If the consistent orthogonality condition is satisfied on these two scene points, both surface normal and equivalent lighting can
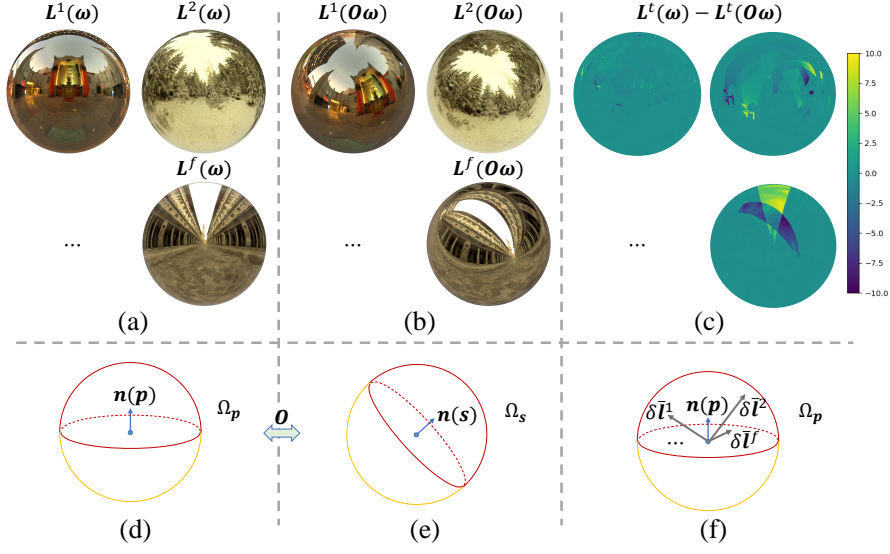
Fig. 3.16 (a) and (b) visualize the f environment maps of two visible hemispheres under distinct surface normals shown in (d) and (e). (f) shows the equivalent differential lighting $\delta\bar{\mathbf{l}}^1 \sim \delta\bar{\mathbf{l}}^f$ defined by the spherical integral of environment lighting intensity difference illustrated in (c).

be aligned by the same orthogonal matrix, *i.e.*,

$$
\begin{cases}
\mathbf{On}(\mathbf{q}) = \mathbf{n}(\mathbf{s}) \\
\mathbf{O}\bar{\mathbf{l}}^t(\mathbf{q}) = \bar{\mathbf{l}}^t(\mathbf{s}) \quad \forall t \in (1, f) \cdot \\
\mathbf{O}^\top \mathbf{O} = \mathbf{I}
\end{cases}
\tag{3.23}
$$

With consistent surface normals and equivalent lighting, the intensity profiles of scene point $\mathbf{q}$ and $\mathbf{s}$ are correlated. Therefore, correlated intensity profiles between scene points are the necessary condition of the consistent orthogonality condition.

Combining Eq. (3.20) with Eq. (3.23) we have

$$
\begin{aligned}
\mathbf{O}&\bar{\mathbf{l}}^t(\mathbf{q}) - \bar{\mathbf{l}}^t(\mathbf{s}) \\
&= \mathbf{O} \int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} - \int_{\Omega(\mathbf{s})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} \\
&= \mathbf{O} \int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} - \int_{\Omega(\mathbf{q})} L^t(\mathbf{O}\boldsymbol{\omega})(\mathbf{O}\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \\
&= \mathbf{O} \int_{\Omega(\mathbf{q})} [L^t(\boldsymbol{\omega}) - L^t(\mathbf{O}\boldsymbol{\omega})]\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} \\
&= \mathbf{O}\delta\bar{\mathbf{l}}^t = \mathbf{0}.
\end{aligned}
\tag{3.24}
$$

We denote $\delta\bar{\mathbf{l}}^t$ as *equivalent differential lighting*, which represents the spherical integral of the differential environment lighting intensity over the visible hemisphere $\Omega(\mathbf{p})$, as shown in Fig. 3.16. Since $\mathbf{O}$ is an invertible matrix, the consistent orthogonality condition leads to zero equivalent differential lighting.

Obviously, when surface normals at $\mathbf{q}$ and $\mathbf{s}$ are consistent, zero equivalent differential lighting can be satisfied as they have the same equivalent lighting. Under the case of $\mathbf{n}(\mathbf{q}) \neq \mathbf{n}(\mathbf{s})$, the consistent orthogonality condition requires all the f environment maps illuminating these two scene points satisfying the following condition:

$$\int_{\Omega(\mathbf{q})}[L^t(\boldsymbol{\omega}) - L^t(\mathbf{O}\boldsymbol{\omega})]\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} = \mathbf{0}, \quad \forall t \in (1, f). \tag{3.25}$$

It is difficult to analytically prove that unequal surface normal pairs cannot satisfy the consistent orthogonality condition, since Eq. (3.25) is related to the light intensity $L^t(\boldsymbol{\omega})$ from f unknown environment maps. As shown in Fig. 3.16, real-world environment maps are natural HDR images without following any regular distribution. Also, as we increase the environment lighting amount f, Eq. (3.25) becomes more difficult to achieve. Therefore, we provide a statistical analysis on real-world environment maps to verify whether inconsistent surface normal paris may satisfy the consistent orthogonality condition.

To begin with, we first define a "consistent orthogonal error" $d(\mathbf{q}, \mathbf{s})$ to evaluate whether the light and surface normal of scene points $\mathbf{q}$ and $\mathbf{s}$ fit to the consistent orthogonal condition. It can be defined as the mean angular error between vector set $[\mathbf{n}(\mathbf{s}), \bar{\mathbf{l}}^1(\mathbf{s}), \bar{\mathbf{l}}^2(\mathbf{s}), \cdots, \bar{\mathbf{l}}^f(\mathbf{s})]$ and $[\mathbf{On}(\mathbf{q}), \mathbf{O}\bar{\mathbf{l}}^1(\mathbf{s}), \mathbf{O}\bar{\mathbf{l}}^2(\mathbf{s}), \cdots, \mathbf{O}\bar{\mathbf{l}}^f(\mathbf{s})]$, where $\mathbf{O}$ is an orthogonal matrix that aligning the equivalent lights and surface normals between the two scene points. We consider the two scene points satisfy the consistent orthogonal condition if their consistent orthogonal error is less than a setting threshold.

We collect 31 real-world environment maps from sIBL Archive, and uniformly sample 151256 distinct surface normal directions from a sphere and pre-compute the equivalent lightings for each normal direction of all the 31 environment maps. We set the threshold for consistent orthogonal error as 0.01° and summarize the mean angular
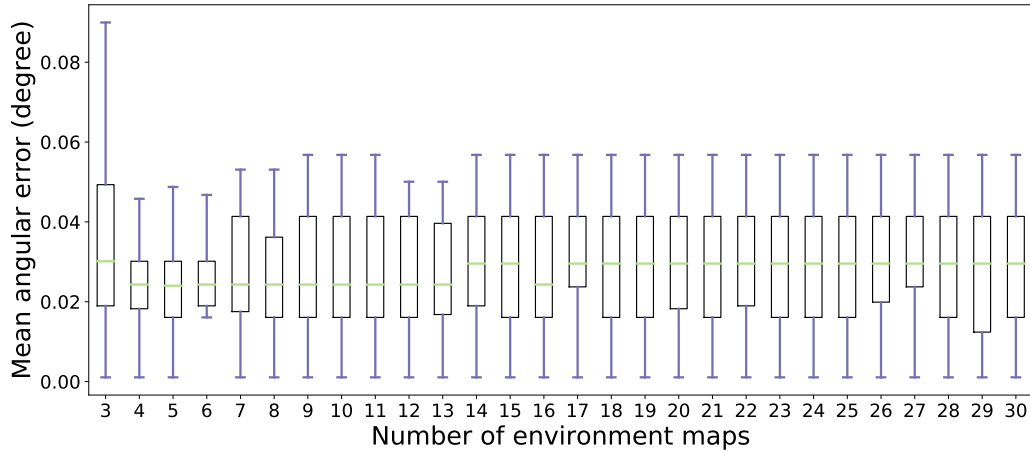
Fig. 3.17 Surface normal clustering error w.r.t. different environment lighting numbers f, with the consistent orthogonality condition satisfied. The statistics of angular errors are displayed using the box-and-whisker plot: The green bar indicates the median value, the top and bottom bounds of the black box indicate the first and third quartile values, and the top and bottom ends of the vertical blue line indicate the minimum and maximum errors.

error of surface normals satisfying the consistent orthogonal condition w.r.t. different numbers of environment lights in Fig. 3.17. As an example, in the case of 15 natural lightings, we randomly sample 15 out of 31 environment maps for 20 times to obtain 20 different environment map groups. For each group, we first extract scene point pairs whose surface normal and equivalent lighting directions satisfy the consistent orthogonality condition. Then we record the surface normal angular error of matched scene point pairs. The mean angular error on all the 20 groups are only 0.027 degrees, which are quite small numbers.

From the statistic in Fig. 3.17, the surface normal clustering errors with the consistent orthogonality condition on different number of environment maps are near zero. Therefore, From a practical point of view, it is sufficiently safe to say that under real-world natural lighting, if surface normals and the equivalent distant lighting directions of two scene points satisfy the consistent orthogonality condition, the two surface normals and the equivalent lightings have the same directions.

3.6.2   Resolving the orthogonal ambiguity via integrability

After local surface normal estimation and patch merging process, we can obtain a complete surface normal map up to a global orthogonal ambiguity. As discussed in [65, 79, 84], this global orthogonal ambiguity can be reduced to a convex/concave binary ambiguity by addressing the surface integrability constraint. In the following, we first give the proof and then present the steps to solve the global orthogonal ambiguity.

Proof of resolving orthogonal ambiguity   After merging patch surface normals by optimizing the binary ambiguity graph and the rotation ambiguity graph, we obtain a complete surface normal map $\hat{\mathbf{N}}$ up a global orthogonal ambiguity $\mathbf{O}_g$ such that

$$\mathbf{N} = \mathbf{O}_g\hat{\mathbf{N}}. \tag{3.26}$$

Following Belhumeur *et al.* [79], if the surface normal $\mathbf{n}$ satisfies the integrability constraint,then

$$\frac{\partial}{\partial x}\left(\frac{n_2}{n_3}\right) = \frac{\partial}{\partial y}\left(\frac{n_1}{n_3}\right),$$
$$n_3\frac{\partial n_2}{\partial x} - n_2\frac{\partial n_3}{\partial x} = n_3\frac{\partial n_1}{\partial y} - n_1\frac{\partial n_3}{\partial y}. \tag{3.27}$$

Denoting the three rows of the orthogonal ambiguity $\mathbf{O}_g$ as $\mathbf{o}_1, \mathbf{o}_2,$ and $\mathbf{o}_3$, Substituting Eq. (3.26) to Eq. (3.27), we obtain the following equality:

$$\begin{pmatrix} \hat{n}_3\hat{n}_{2y} - \hat{n}_2\hat{n}_{3y} \\ \hat{n}_1\hat{n}_{3y} - \hat{n}_3\hat{n}_{1y} \\ \hat{n}_2\hat{n}_{1y} - \hat{n}_1\hat{n}_{2y} \\ \hat{n}_2\hat{n}_{3x} - \hat{n}_3\hat{n}_{2x} \\ \hat{n}_3\hat{n}_{1x} - \hat{n}_1\hat{n}_{3x} \\ \hat{n}_1\hat{n}_{2x} - \hat{n}_2\hat{n}_{1x} \end{pmatrix}^\top \begin{pmatrix} \mathbf{o}_1 \times \mathbf{o}_3 \\ \mathbf{o}_2 \times \mathbf{o}_3 \end{pmatrix} = 0, \tag{3.28}$$

where $\times$ denotes the cross product operator, the subscript x, y represent the partial derivatives in two directions. As the pseudo surface normal for the complete surface

is known, we stack Eq. (3.28) for all scene points and obtain a homogeneous linear system $\mathbf{Ax} = \mathbf{0}$. The non-trivial solution of $\mathbf{x}$ is then obtained via SVD on $\mathbf{A}$, result in the cross product estimates $\mathbf{c}_{13}$ and $\mathbf{c}_{23}$ up to a scale k, *i.e.*

$$\begin{aligned} \mathbf{o}_1 \times \mathbf{o}_3 &= k\mathbf{c}_{13} \\ \mathbf{o}_2 \times \mathbf{o}_3 &= k\mathbf{c}_{23}. \end{aligned} \tag{3.29}$$

On the other hand, as the rows of orthogonal ambiguity $\mathbf{O}_g$, $\mathbf{o}_1 \sim \mathbf{o}_3$ are unit vectors. Therefore, we can solve the scale k up to a sign ambiguity. As the determinant of $\mathbf{O}_g$ can be either 1 or -1, there are 4 candidates of $\mathbf{O}_g$ satisfying both orthogonal constraint and integrability constraint, denoted as $\mathbf{O}_{g1} \sim \mathbf{O}_{g4}$ below:

$$\mathbf{O}_{g1} = \frac{1}{k} \begin{pmatrix} \mathbf{c}_{23}^\top \\ -\mathbf{c}_{13}^\top \\ -\mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}, \mathbf{O}_{g2} = \frac{1}{k} \begin{pmatrix} -\mathbf{c}_{23}^\top \\ \mathbf{c}_{13}^\top \\ -\mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix},$$

$$\mathbf{O}_{g3} = \frac{1}{k} \begin{pmatrix} -\mathbf{c}_{23}^\top \\ \mathbf{c}_{13}^\top \\ \mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}, \quad \mathbf{O}_{g4} = \frac{1}{k} \begin{pmatrix} \mathbf{c}_{23}^\top \\ -\mathbf{c}_{13}^\top \\ \mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}. \tag{3.30}$$

Obviously, we have $\mathbf{O}_{g1} = -\mathbf{O}_{g3}, \mathbf{O}_{g2} = -\mathbf{O}_{g4}$. As the ground truth surface normal should have the same direction of camera view, two of the four orthogonal ambiguity candidates are chosen to guarantee the recovered shape to be in front of the camera. Therefore, the global orthogonal ambiguity can be resolved up to a binary choice of the remained two candidates. In the geometry side, this binary ambiguity corresponds to the classical concave-convex ambiguity that occurs in shape from shading [103]. The conclusion is also consistent with previous methods [65, 79, 84].

Solving the orthogonal ambiguity Ideally, the estimated $\mathbf{c}_{13}$ and $\mathbf{c}_{23}$ from the homogeneous linear system derived from Eq. (3.28) should comply with the constraint that $\mathbf{c}_{13}^\top \mathbf{c}_{23} = 0$. Due to the error introduced by the finite difference and the inaccurate pseudo surface normals included in Eq. (3.28), the above constraint cannot be satisfied.

Therefore, we formulate an optimization to address both orthogonal constraint of $\mathbf{O}_g$ and the integrability constraint:

$$
\begin{aligned}
&\underset{\mathbf{x}}{\arg\min} \, \|\mathbf{A}\mathbf{x}\|_2^2, \\
s.t. \quad &\mathbf{x}^\top \mathbf{C}_1^\top \mathbf{C}_1 \mathbf{x} = 1, \\
&\mathbf{x}^\top \mathbf{C}_2^\top \mathbf{C}_2 \mathbf{x} = 1, \\
&\mathbf{x}^\top \mathbf{C}_1^\top \mathbf{C}_2 \mathbf{x} = 0,
\end{aligned}
\tag{3.31}
$$

where $\mathbf{C}_1 = [\mathbf{I}, \mathbf{0}]$, $\mathbf{C}_2 = [\mathbf{0}, \mathbf{I}]$, and $\mathbf{I} \in \mathbb{R}^3$ is an identity matrix. However, the above minimization is non-convex and hard to optimize. As the $O(3)$ group is compact, we solve the orthogonal ambiguity in a discrete Hypothesis-and-Test manner.

As the integrability constraint is invariant from the concave/convex ambiguity, we can first solve the rotation ambiguity from the global orthogonal ambiguity and then choose the correct one from Eq. (3.30). Without loss of generality, we decompose the rotation ambiguity matrix $\mathbf{R}$ into three sub-rotations

$$
\mathbf{R} = \mathbf{R}_z \mathbf{R}_y \mathbf{R}_x,
\tag{3.32}
$$

where $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ are rotation matrices along x, y, z axes, corresponding to the rotation angle $\theta_x, \theta_y, \theta_z$, respectively. Suppose the ground-truth rotations of $\mathbf{R}_x, \mathbf{R}_y$ are known, we can get pseudo surface normal $\tilde{\mathbf{N}}$ up to an rotation along the z-axis such that

$$
\mathbf{N} = \mathbf{R}_z \mathbf{R}_y \mathbf{R}_x \hat{\mathbf{N}} = \mathbf{R}_z \tilde{\mathbf{N}} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tilde{\mathbf{N}}.
\tag{3.33}
$$

Combine Eqs. 3.28 with Eqs. 3.33, we have

$$
\begin{pmatrix} \tilde{n}_2 \tilde{n}_{3y} - \tilde{n}_3 \tilde{n}_{2y} - \tilde{n}_3 \tilde{n}_{1x} + \tilde{n}_1 \tilde{n}_{3x} \\ \tilde{n}_2 \tilde{n}_{3x} - \tilde{n}_3 \tilde{n}_{2x} - \tilde{n}_1 \tilde{n}_{3y} + \tilde{n}_3 \tilde{n}_{1y} \end{pmatrix}^\top \begin{pmatrix} \sin\theta_z \\ \cos\theta_z \end{pmatrix} = 0.
\tag{3.34}
$$

Stacking Eq. (3.34) for all pixels, we obtain a constrained optimization system about

θ_z as follows

$$\underset{\mathbf{y}}{\operatorname{argmin}} \|\mathbf{B}\mathbf{y}\|_2^2, \qquad s.t. \qquad \|\mathbf{y}\|_2^2 = 1, \qquad (3.35)$$

where $\mathbf{y} = [\sin \theta_z, \cos \theta_z]^\top$. The above optimization can be formulated as a generalized Eigenvalue problem which has a unique solution [104].

Therefore, we sample different pairs of rotations $\mathbf{R}_x$ and $\mathbf{R}_y$, and then solve $\mathbf{R}_z$ with the integrability constraint. For each group of $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$, we record integrability cost as $\|\mathbf{B}\mathbf{y}\|_2^2$ from the optimization in Eq. (3.35). The rotation ambiguity is corresponding to the group with the smallest integrability cost. As shown in Algorithm 1, we summarize how to resolve the rotation ambiguity based on the integrability constraint.

With estimated rotation ambiguity $\mathbf{R}$, we can now build the four possible orthogonal ambiguity candidates shown in Eq. (3.30). Given the prior that the ground-truth surface normals have positive z elements in the viewer-oriented coordinate system, we remove two of the four candidates, and the remaining orthogonal ambiguity candidates correspond to the convex/concave ambiguity.

Algorithm 2: Solve rotation ambiguity with integrability

    Input : Max rotation angles θ_x^m, θ_y^m along x and y axes
    Output: Rotation ambiguity $\mathbf{R}$
    Initialization: Initial best cost c_s

1 **for** $\theta_x \in (-\theta_x^m, \theta_x^m)$ **do**
2   **for** $\theta_y \in (-\theta_y^m, \theta_y^m)$ **do**
3   **for** $\dot{\mathbf{N}} \in \{\hat{\mathbf{N}}, -\hat{\mathbf{N}}\}$ **do**
4   Calculate $\mathbf{R}_x, \mathbf{R}_y$ from θ_x, θ_y;
5   Rotate pseudo surface normal to $\tilde{\mathbf{N}} = \mathbf{R}_y \mathbf{R}_x \dot{\mathbf{N}}$;
6   Calculate $\mathbf{R}_z$ and record the integrability cost c from Eq. (3.35);
7   **if** $c < c_s$ **then**
8             $c_s = c$;
9             $\mathbf{R} = \mathbf{R}_z \mathbf{R}_y \mathbf{R}_x$;
10   **end**
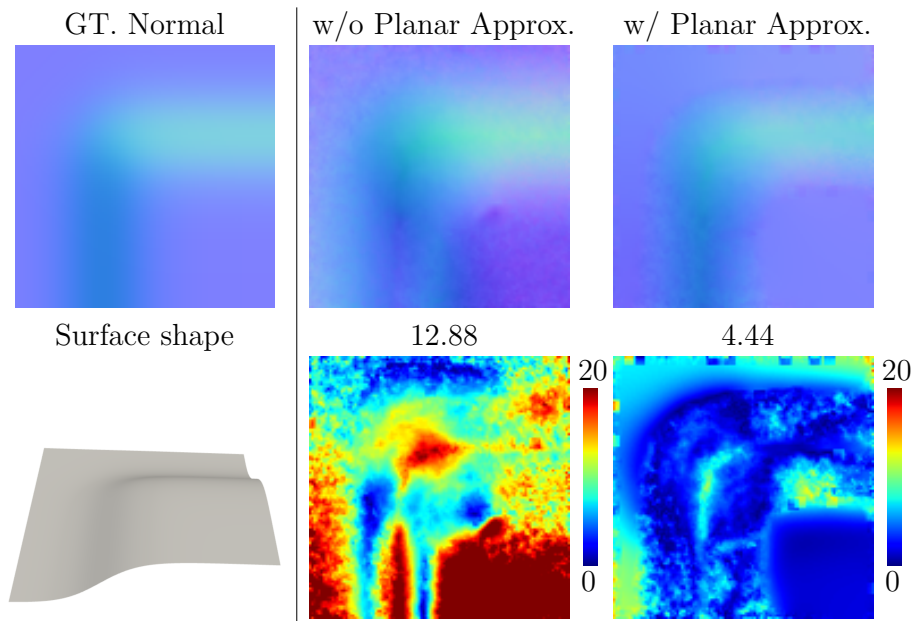11   **end**
12   **end**
13 **end**

Fig. 3.18 Normal estimation for surface with planar patches as shown in the left column. The middle and right columns show the surface normal estimates, mean angular errors, and the error distributions w/o and w/ "planar approximation" strategy.

3.6.3 Local surface normal estimation for planar patch

In Sec. 3.3.2, we have shown that surface normal can be solved up to an orthogonal ambiguity if there are at least 6 scene points within the patch sharing the same albedo but distinct surface normals. In this way, there exists a unique solution for the linear system shown below,

$$\underbrace{\left[\mathrm{tri}(\tilde{\mathbf{n}}_{k,1}\tilde{\mathbf{n}}_{k,1}^{\top}) \quad \cdots \quad \mathrm{tri}(\tilde{\mathbf{n}}_{k,p_k}\tilde{\mathbf{n}}_{k,p_k}^{\top})\right]^{\top}}_{\mathbf{E}} \underbrace{\mathrm{tri}(\mathbf{Q}_k^{\top}\mathbf{Q}_k)}_{\mathbf{y}} = \mathbf{1}, \tag{3.36}$$

where $\tilde{\mathbf{n}}_k$ is the pseudo surface normal up to a linear ambiguity $\mathbf{Q}_k$ in k-th surface patch. When there are no more than 6 diverse surface normals on the patch (*e.g.* planar surface), $\mathbf{E}$ becomes rank deficient, which reveals the degeneration in our local surface normal estimation. Such case implies the corresponding surface patch is near flat especially for large patch size (*e.g.* $5 \times 5, 7 \times 7$). Therefore, we use a "planar approximation" strategy to force the pseudo surface normals of the patch sharing the

same direction rather than following the estimation in Sec. 3.2 for degenerate patches. As the ground truth patch surface normal map is also near-planar, the pseudo surface normal map as a plane can be approximately aligned to its ground truth with an orthogonal matrix. Therefore, we can still apply our GPM to solve the per-patch orthogonal ambiguity including degenerate patches.

As shown in Fig. 3.18, we shown an object with planar local patches. For surface normal estimates labeled by "w/ Planar Approx.", we first detect degenerate patches by checking whether $\mathbf{E}$ in Eq. (3.36) is rank-deficient. After that we use "planar approximation" strategy to generate the pseudo surface normals for degenerate patches and apply Ours (GPM) to obtain the complete surface normal estimates. From the error maps shown in the bottom row, the "planar approximation" strategy enables our method handle the degenerate cases and obtain more accurate surface normal estimates at the near-planar surface regions.

3.7 Conclusion

We propose a uncalibrated photometric stereo method under unknown natural illumination. Our method simplifies the natural illumination using the equivalent directional lighting model that is valid for local patches. We then solve each patch up to an arbitrary orthogonal ambiguity. The patches are further unified through a graph-based patch merging method (GPM), which introduces MRF optimization and rotation averaging to solve the patch-wise ambiguities up to a global orthogonal ambiguity for the whole surface. Finally, we resolve the global ambiguity to become a concave/convex ambiguity, which could be easily removed manually. We believe such a method has great potential to bring photometric 3D modeling techniques from lab setup with controlled lighting to wild and large datasets on the Internet.

Limitation. The limitation of our method mainly lies in the local surface normal estimation process, in which three assumptions need to be satisfied: uniform albedo, patch surface normals having small angular difference, and non-planar surface with

6+ distinct surface normals. Our method cannot handle complex albedo maps (*e.g.*, general SV-albedo shown in Fig. 3.12 and Tabletcase in Fig. 3.15) when the uniform albedo assumption becomes invalid for most patches. Also, if the surface normals vary significantly or cast shadow exists in a local patch, the equivalent directional lighting approximation is less accurate, which leads to wrong shape estimation results such as the Rucksack as shown in Fig. 3.15. When given near planar surface such as the case of Tabletcase and Ovenmitt, our method is also not reliable due to the degeneration in the local surface normal estimation.

Future work. We hope to develop a more robust local surface normal estimation method that can handle more complex illumination, texture and cast shadows. To deal with the degenerate cases of near-planar local patches, we can detect them by calculating the numerical rank as discussed in Sec. 3.6.3. As this detection process is not stable due to the shadow and noise in the real-captured images, we left it as one of our future works. Besides, to resolve the linear ambiguity in local patches as discussed in Sec. 3, an alternative way is applying integrability rather than uniform albedo constraint to solve the pseudo normals up to a GBR ambiguity [79, 80]. But how to solve the patch-wise GBR ambiguities up to a global GBR ambiguity is beyond the scope of current method. Finally, our patch-wise processing shares similar spirits with local shading analysis in [105], where they find local shapes have simple parametric approximation under directional lighting. In contrast, we explore how local shapes simplify natural illumination representation. It might be interested to combine local shape constraints in [105] to further narrow the solution space.

Chapter 4

Photometric Stereo under Multispectral Light

4.1 Introduction

Photometric stereo is effective for the detailed recovery of three-dimensional (3D) surfaces. Conventional photometric stereo (CPS) methods, originally proposed by Woodham [7] and Silver [8], use images captured from a fixed camera under varying lighting directions, which are commonly obtained at different timestamps. Since CPS methods stack images with *time-multiplexing*, the target surface has to be static during
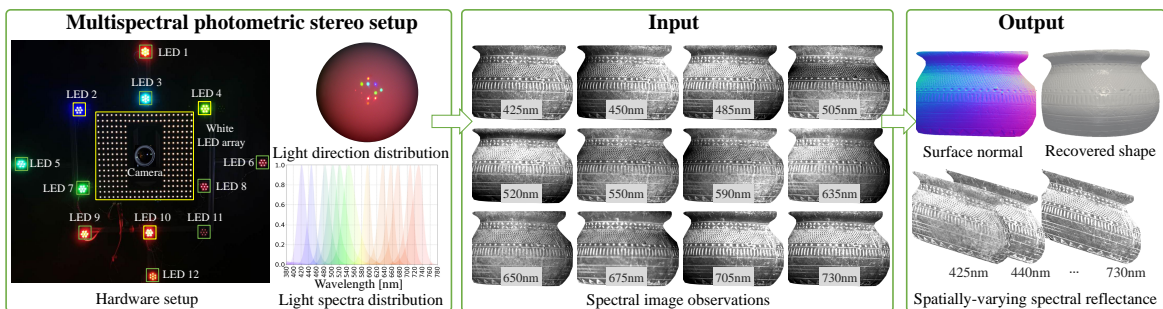

Fig. 4.1 Our multispectral photometric stereo setup with 12 narrow-band spectral LEDs placed at different locations. Taking the spectral image observations as input, our method outputs a closed-form unique solution of both surface normal and spatially-varying spectral reflectance.

the multiple shots. Also, as the inputs of CPS are gray scale or RGB images[1], the spectral property of the material is omitted in the estimated reflectance.

With multispectral photometric stereo (MPS) [106], detailed 3D shape and the corresponding spectral reflectance can be jointly recovered from a one-shot multispectral image via *spectral-multiplexing*. In this way, we can achieve detailed shape and spectral reflectance reconstruction for dynamic objects. However, unlike conventional time-multiplexing photometric stereo, MPS with Lambertian surfaces remains an ill-posed problem. In this chapter, we propose a method to make the problem tractable and provide a unique solution for MPS under spatially-varying reflectance, making the MPS approaches applicable for dynamic shape reconstruction.

An input image for MPS encodes observations under different lighting directions in different spectral bands, conveying the information about the surface normals and spectral reflectances. Figure 4.1 shows our MPS setup, which contains a fixed-viewpoint camera and 12 narrow-band spectral light sources located at different positions. From the input spectral image observations, our goal is to estimate both object shape and spectral reflectance simultaneously. However, under the illumination of f spectral lights, there are $f + 2$ unknowns (f for the reflectance of the spectral bands, and 2 for the surface normal). Since only f observations for each scene point are given, MPS is inherently underconstrained.

To make the problem tractable, existing methods use additional priors, *e.g.*, initial shape [107, 108], trained neural networks [109–111], or local smoothness regularization [112]. However, these priors are rather restrictive and may not always comply with actual scenes. Without these priors, existing methods make assumptions on the spatial distribution of spectral reflectance to solve the problem. Specifically, the spectral reflectance is decomposed/normalized into two parts: Chromaticity and albedo, as visualized on the top of Fig. 4.2. Based on the spatial distribution of the chromaticity and albedo, we can obtain four spectral reflectance types (SRT), where an example object for each type is shown on the bottom of Fig. 4.2.

---

[1]When RGB images are used in conventional (single-band) photometric stereo, they are turned into gray scale images for computing surface normal.
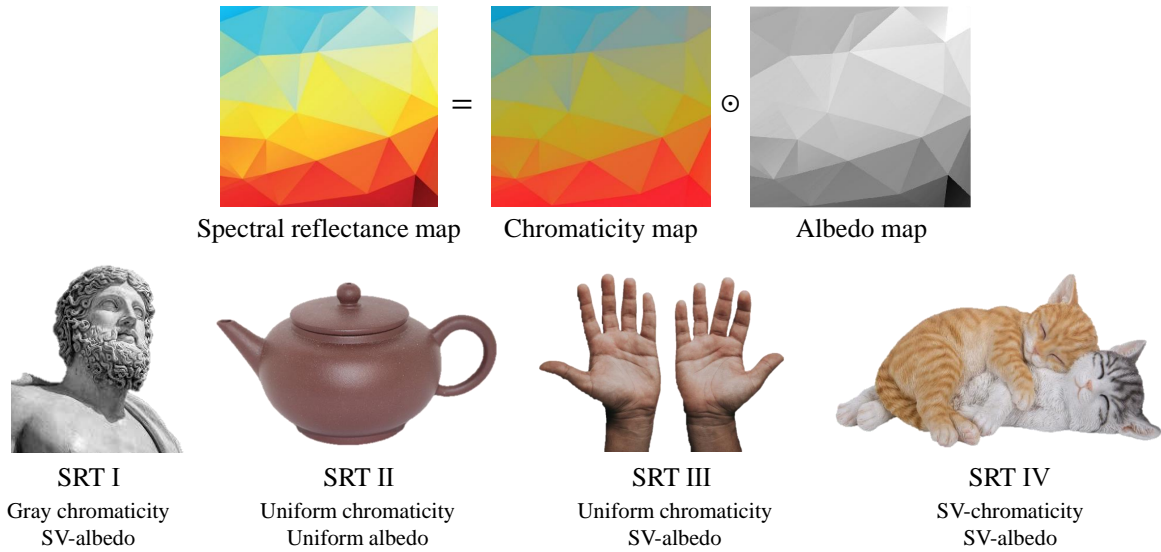
Fig. 4.2 (**Top**) Visualization of the spectral reflectance decomposition in the RGB space, where $\odot$ represents elementwize multiplication. The chromaticity is reated to the hue and the albedo is related to the saturation. (**Bottom**) Example objects of 4 spectral reflectance types (SRT) categoried by the spatial distribution of the chromaticity and the albedo.

SRT I limits the surface to be gray-chromatic (*e.g.*, statue in Fig. 4.2). MPS under such assumption is identical to well-posed classical photometric stereo [8]. By assuming SRT II where the object is monochromatic (not limited to gray) with uniform albedo (*e.g.*, pot in Fig. 4.2), existing methods [113, 114] provide a unique solution for MPS. However, SRT II is also restrictive for real-world scenes. As shown in [113, 114], incorrect surface normals are estimated at regions with roughly constant chromaticity but continuously changing albedos. With the relaxation of piece-wise constant chromaticities and albedos, previous methods [113, 114] also investigated MPS for spatially-varying chromaticities and albedos (SRT IV) such as the cat case in Fig. 4.2). However, the spatial clustering of the uniform spectral reflectance regions is not only cumbersome but also fragile to outliers, such as shadows and specular highlights.

In this chapter, we make MPS work well under spatially-varying spectral reflectances. Given a multispectral image under calibrated lighting directions, we first provide a closed-form MPS solution for surfaces with uniform chromaticity but spatially-varying albedos (SRT III in Fig. 4.2), without relying on any additional priors. We

further extend our method to deal with the surface with spatially-varying chromaticities and albedos (SRT IV in Fig. 4.2) by additionally calibrating the light spectra and camera spectral sensitivity.

Specifically, for SRT III surfaces, we treat the estimation of spectral reflectance and surface normal as a bilinear optimization problem. We show that the problem can be turned into a homogeneous system of linear equations, where the surface normal and spectral reflectances are jointly estimated. Given observations of SRT IV surfaces and calibrated light spectra and camera spectral sensitivity, we show that closed-form solutions for both surface normal and spectral reflectance are given in a per-pixel manner. We achieve this by expressing the spectral reflectance with linear bases, which are extracted from a material database of bidirectional reflectance distribution functions (BRDFs) [115]. Unlike previous methods that are restricted to three spectral channels [107–109, 113, 114], our method allows the use of arbitrarily many spectral channels. As a side-bonus of this input property, we can also rely on the off-the-shelf four or more source photometric stereo methods to deal with outliers, such as shadow and specular highlights, making our methods for both SRT III and IV more robust than existing RGB-based MPS methods.

To summarize, the primary contributions of our work are as follows.

- We show that MPS for monochromatic surfaces with spatially-varying albedos (SRT III) can be solved in a closed-form without introducing any external priors, and we derive the minimal conditions based on the number of spectral lights and scene points for the problem to have a unique solution.

- We introduce a basis representation for the spectral reflectance and present a closed-form MPS solution for surfaces with spatially-varying chromaticities and albedos if the light spectra and camera spectral sensitivity are calibrated.

- Our methods for both SRT III and SRT IV are robust to outliers, such as shadows and specular highlights, because of its capability of applying robust estimation thanks to that our method can take arbitrary many spectral channels as input.
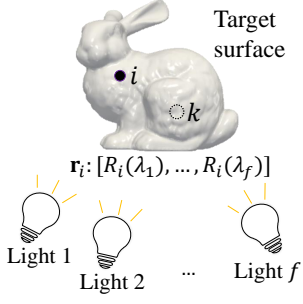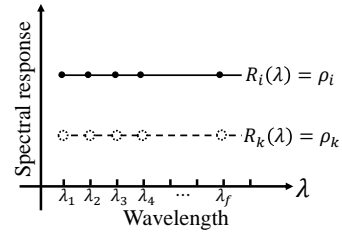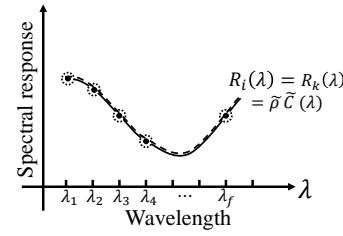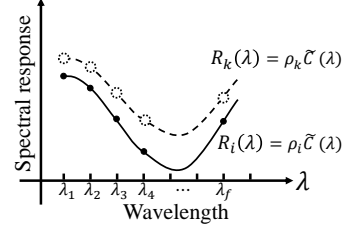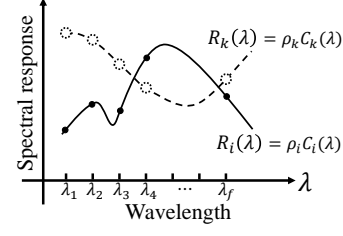
Spectral Reflectance Type (**SRT**)	**SRT I**	**SRT II**
$R_i(\lambda) = \rho_i C_i(\lambda), \quad i \in (1, p)$	Gray-chromaticity: $C_i(\lambda) = C_k(\lambda) = 1$ SV-albedo: $\rho_i, \rho_k \in \mathbb{R}_+$	Mono-chromaticity: $C_i(\lambda) = C_k(\lambda) = \tilde{C}(\lambda)$ Uniform albedo: $\rho_i = \rho_k = \tilde{\rho}$

Fig. 4.3 Visualization of four spectral reflectance types (SRT) categorized by the spatial distribution of the chromaticity $C(\lambda)$ and the albedo ρ, where $\tilde{C}(\lambda)$ and $\tilde{\rho}$ represents the uniform chromaticity and albedo shared by different scene points. The color maps provide spatial distribution examples of chromaticities and albedos for each SRT in the RGB space. Solid and hollow dots show the spectral reflectances of two scene points at f wavelengths $\mathbf{r} = [R(\lambda_1), \cdots, R(\lambda_f)]$. This chapter presents unique and closed-form solutions for both SRT III and IV.

4.2 Related works

As described in previous works [116, 117], the material spectral reflectance $R(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ can be decomposed into two parts: *Chromaticity* $C(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ and *albedo* $\rho \in \mathbb{R}_+$, such that $R(\lambda) = C(\lambda)\rho$, where λ represents wavelength. As shown in Fig. 4.3, based on the spatial distribution of chromaticity and albedo for a surface, we categorize 4 different surface spectral reflectance types (SRT) and order them in a way from simple to complex. In this section, we introduce existing methods based on their assumptions on SRT and list their properties for the comparison in Table 4.1.

Table 4.1 Comparison of MPS methods. Ours$_{\text{III}}$ provides a unique solution for a relatively general spectral reflectance (SRT III) without additional priors. Ours$_{\text{IV}}$ solves MPS for general spatially-varying spectral reflectance with less restrictive calibration.

SRT	Method	Input	# Lights	Additional priors
I	[8]	MSI[1]	$f \geq 3$	None
II	[106]	RGB	$f = 3$	Surface integrability
II	[118]	RGB	$f = 3$	Surface integrability
II	[119]	RGB	$f = 3$	Irradiance-normal mapping[2]
II	[113, 114]	RGB	$f = 3$	None
III	[117]	RGB	$f = 3$	Initial coarse shape Pixels with uniform albedo
IV	[113, 114]	RGB	$f = 3$	Reflectance quantization Piece-wise constant reflectance
IV	[109–111, 120]	RGB	$f = 3$	Fixed lighting direction
IV	[107, 108]	RGBD[3]	$f = 3$	Piece-wise constant chromaticity
IV	[112]	MSI	$f \geq 3$	Reflectance smoothness Surface normal smoothness
IV	[121]	MSI	$f \geq 5$	Spectral reflectance basis[1]
III	Ours$_{\text{III}}$	MSI	$f \geq 4$	None
IV	Ours$_{\text{IV}}$	MSI	$f \geq 4$	Calibrated light and camera spectrum Basis expression of the reflectance

[1] Multispectral image   [2] Scene-dependent calibration   [3] RGB + depth

SRT I   If the surface has gray chromaticity, *i.e.*, the chromaticity remains constant w.r.t. varying wavelength, MPS is identical to classical photometric stereo. Therefore, given 3 or more spectral bands, a closed-form solution for surface normal can be obtained without ambiguity [8].

SRT II   For monochromatic surfaces with uniform albedo, *i.e.*, all the scene points share the common chromaticity $\tilde{C}(\lambda)$ and albedo $\tilde{\rho}$, previous methods [106, 118] show surface normal can be estimated from a single RGB image up to a rotation ambiguity. The correct rotation was approximated by imposing an additional integrability condition. Hernández *et al.* [119] establish a one-to-one linear mapping between pixel measurements and surface normals to reconstruct the deformable cloth shape. This unknown linear mapping is calibrated via a planar board with a cloth sample fixed in the center. If the crosstalk between spectral channels is negligible, existing methods [113, 114] provide a unique solution for surface normals. However, their methods

are restricted to RGB 3-channel input and cannot be expanded to more channels.

SRT III Few methods focus on the monochromatic surfaces with spatially-varying albedos, which is commonly seen in natural objects (*e.g.*, wood and rocks) and human skins. Vogiatzis *et al.* [117] assume the spectral reflectance of the human face follows SRT III and obtain detailed reconstructions of faces in real-time. However, their surface normal estimation results rely on the accuracy of initial geometry and detection of equal-albedo pixels.

SRT IV If the chromaticity and albedo are both spatially-varying, MPS from a single multispectral image is ill-posed. Existing methods apply additional regularizations and provide numerical solutions for MPS. Chakrabarti *et al.* [114] and Ozawa *et al.* [113] relax the spatially-varying spectral reflectance to be piece-wise constant. Since their methods are both based on 3-channel RGB inputs, they discretize the spectral reflectance in a 3D space to cluster pixels with equal chromaticities and albedos so that they can turn the problem into a set of SRT II subproblems. The normal map is then estimated in each surface region that is predicted as having the same spectral reflectance. The method by Anderson *et al.* requires a coarse shape from depth map [107] or stereo pairs [108] and uses it to guide the chromaticity segmentation and the surface normal estimation. Similar to [113, 114], the piece-wise constant spectral reflectance assumption restricts the flexibility of the target surface's reflectance. The normal estimation accuracy is also influenced by the errors introduced by the reflectance clustering step.

Some recent methods directly take an RGB image as input and apply deep neural networks to predict the surface normal [109, 110, 120]. However, the lighting directions are required to be consistent between the training and test procedures. Miyazaki *et al.* [112] recover surface normals from a multispectral image with more than three channels. However, their recovered shape tends to be over-smoothed due to the spatial smoothness assumption on both surface normal and the reflectance. Fyffe *et al.* [121] assume the spectral reflectance lies in a low-dimensional space and represent it with a

statistical basis set. However, their spectral reflectance bases are scene-dependent and need to be calibrated with the known surface normal and reflectance pairs. Besides, the optimization of this method is non-convex and requires a good initialization.

Our method Taking a multispectral image with an arbitrary number of channels as input, we first formulate MPS for monochromatic surfaces with spatially-varying albedos (SRT III) as a well-posed problem, and estimate surface normal without introducing external priors [13]. We further show that MPS under SRT IV can be made tractable if the light spectra and camera spectral sensitivity are calibrated. Different than existing works [108, 113, 114], we avoid both the piece-wise uniform spectral reflectance restriction and the reflectance clustering steps by introducing a basis representation of the per-pixel spectral reflectance. Compared with Fyffe *et al.* [121], our formations for both SRT III and SRT IV are convex, and our extracted spectral reflectance bases are shown to be scene independent based on the real data experiments.

4.3 MPS under monochromatic reflectance

Given a multispectral camera with a linear radiometric response and f (geometrically but not spectrally) calibrated spectral directional lights, we capture a multispectral image of p scene points on a Lambertian surface by turning on all the spectral lights. If the crosstalk between spectral bands is negligible, *i.e.*, the observation under each spectral light is only observed in its corresponding camera channel, observations $\mathbf{m}_i \in \mathbb{R}_+^f$ for the i-th pixel can be written as follows

$$\mathbf{m}_i = \mathrm{diag}(\mathbf{t}_i)\{\mathbf{L}\mathbf{n}_i\}_+, \tag{4.1}$$

where $\mathbf{n}_i \in S^2 \subset \mathbb{R}^3$ represents the unit surface normal vector, $\mathbf{L} \in \mathbb{R}^{f \times 3}$ stacks all the light directions. We use $\mathrm{diag}(\cdot)$ as a diagonalization operator and $\{\cdot\}_+$ as a non-negative operator, which accounts for attached shadows. For simplicity, we omit this operator $\{\cdot\}_+$ in the following explanation. Here, $\mathbf{t}_i \in \mathbb{R}_+^f$ is related to the

camera spectral sensitivity, light source spectra and the surface spectral reflectance at f spectral bands. Its element follows

$$t_{ij} = \int_{\lambda \in \Omega_j} E_j(\lambda) R_i(\lambda) S_j(\lambda) d\lambda, \tag{4.2}$$

where Ω_j is the wavelength range of the j-th spectral band, $E_j(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ denotes the spectra of the j-th light, $S_j(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ defines the camera spectral sensitivity at j-th channel, and $R_i(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ is the material spectral reflectance of the i-th scene point. The problem of general MPS is to estimate $f + 2$ unknowns including $\mathbf{t}$ and surface normal $\mathbf{n}$ from f-element measurement vector $\mathbf{m}$, which is unfortunately an ill-posed problem.

We turn the MPS to be well-posed by assuming the surface following SRT III: The material spectral response can be decomposed into a uniform chromaticity $\tilde{C}(\lambda)$ and spatially varying albedos ρ_i, such that

$$R_i(\lambda) = \rho_i \tilde{C}(\lambda). \tag{4.3}$$

Combing Eqs. (4.2) and (4.3), we rewrite the spectral image observations for a scene point of the SRT III surface as

$$\mathbf{m}_i = \mathrm{diag}(\mathbf{q}) \rho_i \mathbf{L} \mathbf{n}_i, \tag{4.4}$$

where $\mathbf{q} \in \mathbb{R}_+^f$ is the uniform reflectance devoid of spatially-varying albedos, whose elements are

$$q_j = \int_{\lambda \in \Omega_j} E_j(\lambda) \tilde{C}(\lambda) S_j(\lambda) d\lambda. \tag{4.5}$$

With the uniform chromaticity $\tilde{C}(\lambda)$, $\mathbf{q}$ remains constant over the surface since both light spectra and camera spectral sensitivity are independent of the scene points. With the surfaces of SRT III, we found the minimum conditions to yield a unique MPS solution for surface normal are as follows.

Theorem 1 *Given f spectral observations under varying lighting directions of p scene points known to share the same chromaticity $\tilde{C}(\lambda)$, their surface normals can be uniquely determined if either one of the minimal conditions for the number of lightings and pixels is satisfied:*

- *Minimal pixel condition (MPC): $p = 2, f \geq 5$,*

- *Minimal lighting condition (MLC): $f = 4, p \geq 3$.*

In other words, if two scene points share the same chromaticity but varying surface normals, their surface normals can be uniquely determined given 5 or more lighting directions. On the other hand, if we know 3 or more scene points sharing the same chromaticity and their surface normals are non-coplanar, we can recover their normal directions with 4 or more spectral light sources. In the following subsections, we present the unique solution for SRT III and provide the proof for minimal solvable conditions MPC and MLC.

4.3.1   Unique solution for SRT III

Suppose a surface with p scene points sharing the same chromaticity, by representing all pixels and lighting directions in a matrix form, we rewrite Eq. (4.4) as

$$\mathbf{M} = \mathbf{Q}\mathbf{L}\mathbf{N}^\top\mathbf{P}, \tag{4.6}$$

where $\mathbf{Q} = \mathrm{diag}(\mathbf{q})$ is an $f \times f$ diagonal matrix, $\mathbf{M} \in \mathbb{R}_+^{f \times p}$ records the image observations of p scene points under f lights, $\mathbf{N} \in \mathbb{R}^{p \times 3}$ stacks all the surface normals in a row-wise manner, $\mathbf{P}$ is a $p \times p$ diagonal matrix with its diagonal element defined by pixel-wise spatially-varying albedos.

The above spectral image formation model has a similar structure with semi-calibrated photometric stereo (SCPS) [41]. However, the task and physical image formation model between SCPS [41] and our method for SRT III are different. SCPS [41]

denotes $\mathbf{q}$ as light intensities and aims at solving conventional photometric stereo without calibrating the light intensity, whereas ours focuses on the use of relatively general reflectance assumption (SRT III) and multispectral image cues to formulate MPS as a well-posed problem without additional priors. The unknown $\mathbf{q}$ in our method encodes the integral of the light spectra, camera spectral sensitivity, and the chromaticity shared by the scene points, as shown in Eq. (4.5), which is different from the light intensity notation in SCPS [41].

Given image observations $\mathbf{M}$ and the calibrated lighting directions $\mathbf{L}$, we recover uniform reflectance devoid of albedos $\mathbf{Q}$, surface normal $\mathbf{N}$, and albedo $\mathbf{P}$ by minimizing the following energy function:

$$\{\mathbf{Q}^*, \mathbf{N}^*, \mathbf{P}^*\} = \underset{\mathbf{Q}, \mathbf{N}, \mathbf{P}}{\operatorname{argmin}} \left\| \mathbf{M} - \mathbf{Q}\mathbf{L}\mathbf{N}^\top\mathbf{P} \right\|_F^2, \tag{4.7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. We define $\mathbf{B} = \mathbf{P}^\top\mathbf{N} \in \mathbb{R}^{p \times 3}$ as albedo-scaled surface normals. Here, $\mathbf{Q}$ is invertible since its diagonal elements are non-zero. Then we rewrite Eq. (4.6) as

$$\mathbf{Q}^{-1}\mathbf{M} - \mathbf{L}\mathbf{B}^\top = \mathbf{0}. \tag{4.8}$$

After vectorizing the unknown parameters $\mathbf{Q}^{-1}$ and $\mathbf{B}^\top$, we obtain

$$(\mathbf{I}_p \otimes \mathbf{L})\operatorname{vec}(\mathbf{B}^\top) - [\operatorname{diag}(\mathbf{m}_1) \cdots \operatorname{diag}(\mathbf{m}_\mathrm{p})]^\top \mathbf{Q}^{-1}\mathbf{1} = \mathbf{0}, \tag{4.9}$$

where $\operatorname{vec}(\cdot)$ and $\otimes$ represent vectorization and Kronecker product operators. $\mathbf{I}_p \in \mathbb{R}^{p \times p}$ is an identity matrix, $\mathbf{1}$ is a all-one f-dimensional vector, $\mathbf{m}_i$ is the i-th column vector of the image observations $\mathbf{M}$, indicating the measurement at the i-th pixel position.

By concatenating all unknowns of Eq. (4.9) into a vector, we obtain a homogeneous system of linear equations:

$$\underbrace{\left[-\mathbf{I}_p \otimes \mathbf{L} \mid [\operatorname{diag}(\mathbf{m}_1) \mid \cdots \mid \operatorname{diag}(\mathbf{m}_\mathrm{p})]^\top \right]}_{\mathbf{D}} \underbrace{\begin{bmatrix} \operatorname{vec}(\mathbf{B}^\top) \\ \mathbf{Q}^{-1}\mathbf{1} \end{bmatrix}}_{\mathbf{x}} = \mathbf{0}, \tag{4.10}$$

where $\mathbf{D} \in \mathbb{R}^{pf \times (3p+f)}$, and the unknown vector $\mathbf{x}$ has the dimension of $3p + f$. If $\mathbf{D}$ has 1d right nullspace, the solution of $\mathbf{x}$ is obtained up to a scale via a factorization of $\mathbf{D}$ by singular value decomposition (SVD). Based on the prior knowledge that surface normal has a unit norm, we normalize albedo-scaled surface normals $\mathbf{B}$ in $\mathbf{x}$ to finally obtain a unique surface normal estimation.

4.3.2 Minimal conditions for a unique solution

As discussed before, to obtain a non-trivial solution of the homogeneous system in Eq. (4.10), the right nullspace of $\mathbf{D}$ should be one dimension. Therefore, we have

$$pf \geq 3p + f - 1. \tag{4.11}$$

This solvable condition can be interpreted in another way. Given p pixels observed under f spectral bands, the total number of measurements is pf. Since we assume a monochromatic surface with spatially-varying albedos, we only need to know the uniform reflectance devoid of albedos $\mathbf{q}$ for one pixel, whose number of unknowns is f. For the remaining $(p-1)$ pixels, we need to know albedos with the number of unknowns $(p-1)$. Besides, for each pixel, the surface normal has 2 degrees of freedom. There are thus $2p$ unknowns for surface normal. Totally, the number of unknowns is $f + (p-1) + 2p = 3p + f - 1$. Since the number of measurements needs to be no less than the number of unknowns, we obtain the minimal solvable condition of Eq. (4.11).

To further analyze the minimal requirement for the number of lighting directions and pixels, we rewrite Eq. (4.11) as

$$(f - 3)(p - 1) \geq 2. \tag{4.12}$$

Therefore, the minimal requirements for the number of input lighting directions and pixels to obtain a unique solution for SRT III surfaces are

$$\begin{cases} p = 2, f \geq 5, \\ f = 4, p \geq 3, \end{cases} \tag{4.13}$$

which correspond to MPC and MLC in Theorem 1.

4.4 MPS under spatially-varying reflectance

As discussed in the previous sections, general MPS for a surface with spatially-varying reflectance (both chromaticities and albedos) is ill-posed. In this section, we show that the MPS under this SRT IV is tractable if the light sources' spectra E and camera spectral sensitivity S are calibrated in the form of a vector of their products $\mathbf{e} = [E_1(\lambda_1)S_1(\lambda_1), \cdots, E_f(\lambda_f)S_f(\lambda_f)]^\top$ for f distinct spectral bands. By denoting the material reflectances of corresponding spectral bands as $\mathbf{r} = [R(\lambda_1), \cdots, R(\lambda_f)]^\top$, then the image formation model for a pixel under f lights can be written as

$$\mathbf{m} = \text{diag}(\mathbf{e})\text{diag}(\mathbf{r})\mathbf{Ln}. \tag{4.14}$$

Given the calibrated $\mathbf{e}$, we compute the normalized image observations $\hat{\mathbf{m}}$ for a pixel by $\hat{\mathbf{m}} = \mathbf{m} \oslash \mathbf{e}$, where $\oslash$ denotes element-wise division. Then MPS for the SRT IV surface can be formulated as a bilinear optimization of per-pixel surface normal $\mathbf{n}$ and material spectral reflectance $\mathbf{r}$:

$$\{\mathbf{n}^*, \mathbf{r}^*\} = \underset{\mathbf{n}, \mathbf{r}}{\text{argmin}} \|\hat{\mathbf{m}} - \text{diag}(\mathbf{r})\mathbf{Ln}\|_2^2. \tag{4.15}$$

The problem still has f constraints with $f + 2$ unknowns. We now show how this can be further made tractable by introducing the basis representation of the material reflectances in the next section.

4.4.1 Unique solution for SRT IV

To reduce the number of unknowns in Eq. (4.15) and make the problem well-posed, a more compact representation for the spectral reflectance $\mathbf{r}$ is needed.

We assume the spectral reflectance $\mathbf{r}$ is non-zero anywhere and define an inverse spectral reflectance as $\hat{\mathbf{r}} = \mathbf{1} \oslash \mathbf{r}$. Then the normalized image observations for one

pixel satisfy

$$\text{diag}(\hat{\mathbf{m}})\hat{\mathbf{r}} - \mathbf{L}\mathbf{n} = \mathbf{0}. \tag{4.16}$$

In this expression, the inverse spectral reflectance $\hat{\mathbf{r}}$ lies in a f-dimensional space. We approximate it with k $(< f)$ independent linear basis to reduce the unknowns, *i.e.*,

$$\hat{\mathbf{r}} = \mathbf{B}\mathbf{c}, \tag{4.17}$$

where $\mathbf{B} \in \mathbb{R}^{f \times k}$ is a basis matrix stacking k basis vectors, $\mathbf{c} \in \mathbb{R}^{k}$ is the unknown basis coefficients. Combing Eqs. (4.16) and (4.17), we formulate the bilinear optimization of Eq. (4.15) as a homogeneous linear system,

$$\underbrace{\left[-\mathbf{L}|\text{diag}(\hat{\mathbf{m}})\mathbf{B}\right]}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{n} \\ \mathbf{c} \end{bmatrix}}_{\mathbf{y}} = \mathbf{0}, \tag{4.18}$$

where $\mathbf{A} \in \mathbb{R}^{f \times (3+k)}$, and $\mathbf{y}$ has the dimension of $3 + k$. Similar to Ours$_{\text{III}}$ discussed in Sec. 4.3, if $\mathbf{A}$ has one-dimensional right nullspace, we can obtain a unique solution $\mathbf{y}$ up to a scale by SVD. The estimated $\mathbf{y}$ are chosen as the right-singular vector corresponding to the smallest singular value of $\mathbf{A}$. By incorporating the unit norm constraint for the surface normal, we can finally resolve the scale ambiguity and uniquely obtain the estimation of per-pixel surface normal and spectral reflectance.

4.4.2 Spectral reflectance basis extraction

Previous methods conduct linear analysis on MERL BRDF dataset [122] and express the reflectances by the small number of coefficients associated with the basis vectors. However, their extracted bases are not suitable for MPS as the spectral information is omitted. In this chapter, we provide spectral reflectance bases extracted from a spectral BRDF database.

Dupuy *et al.* [115] provided a measured spectral BRDF dataset for 62 materials

with 195 equi-spaced spectral bins covering the $360 \sim 1000$ [nm] range. For each material, spectral responses for 8192 incident-outgoing direction samples are provided. Since we assume the Lambertian model, the spectral reflectances of 8192 directional samples for one material are treated as that of 8192 Lambertian materials independently. By stacking the spectral response of all materials at one wavelength as a row vector, we build a spectral material database $\mathbf{G} \in \mathbb{R}^{195 \times 507904(=62 \times 8192)}$.

With the wavelengths of f spectral lights calibrated, we obtain the corresponding spectral material database $\tilde{\mathbf{G}} \in \mathbb{R}^{f \times 507904}$ by sampling the rows of $\mathbf{G}$. To extract bases for the inverse spectral reflectance $\hat{\mathbf{r}}$, we remove the materials with near-zero spectral responses at any of the f wavelengths in $\tilde{\mathbf{G}}$ and conduct SVD on $\hat{\mathbf{G}} = \mathbf{1} \oslash \tilde{\mathbf{G}}$ as

$$\hat{\mathbf{G}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\top}, \tag{4.19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right orthogonal singular vectors, and $\boldsymbol{\Sigma}$ is a $f \times f$ diagonal matrix containing the singular values in a descending order. The column vectors of $\mathbf{U}$ provide orthogonal bases for the inverse spectral reflectance $\hat{\mathbf{r}}$.

Determining the number of bases   Following the Eckart–Young theorem [123], we select the first k columns of $\mathbf{U}$ as the basis matrix $\mathbf{B} \in \mathbb{R}^{f \times k}$ to approximate the inverse reflectance $\hat{\mathbf{r}}$. To obtain a non-trivial solution of Eq. (4.18), the number of independent basis vectors k should be selected to make $\mathbf{A} \in \mathbb{R}^{f \times (k+3)}$ has a one-dimensional right nullspace. Therefore, the rank of $\mathbf{A}$ should satisfy

$$\text{rank}(\mathbf{A}) = k + 2 < f. \tag{4.20}$$

We calculate the numerical rank of $\mathbf{A}$ following the threshold strategy suggested in [124] and iteratively increase the number of bases in $\mathbf{B}$ from 1 to $f - 3$ until $\mathbf{A}$ satisfies the rank requirement. Since our basis extraction is based on measured spectral BRDF dataset [115] containing various spectral reflectance candidates in the real world, the obtained basis $\mathbf{B}$ is expected to fit diverse scenes, as we will demonstrate it in the real data experiments.
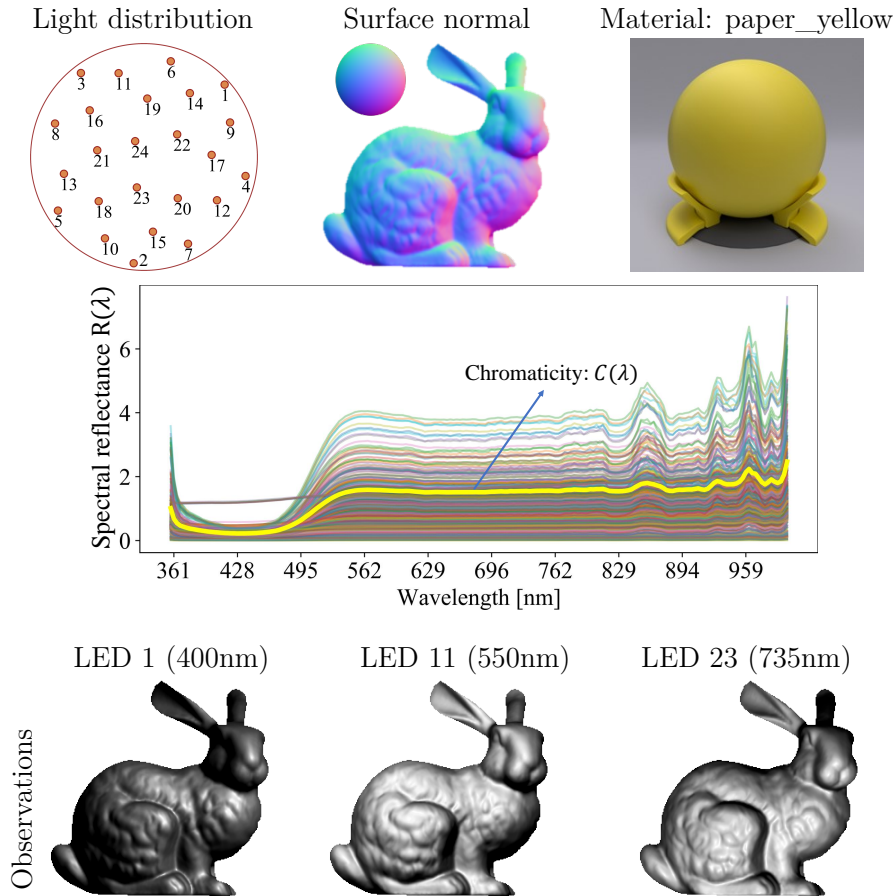
Fig. 4.4 Synthetic multispectral image rendering of a Bunny surface. The measured spectral BRDF "paper_yellow" roughly follows SRT III since its spectral response $R(\lambda)$ under varying groups of surface normal and light directions can be represented by a common chromaticity $C(\lambda)$ with varying scales (albedos).

4.5   Experiments on synthetic dataset

We here introduce experimental results on synthetic datasets. We first describe the details of synthetic data creation and the baseline settings. Then we compare Ours_{III} and Ours_{IV} with the existing MPS methods.

4.5.1   Experimental settings

Synthetic dataset   In our previous work [13], we have verified that Ours_{III} can accurately recover the surface normal on synthetic surfaces rendered with ideal SRT III

Material distribution | Material appearance

Image observation | paper_yellow | acrylic_felt_purple

Spectralon | Colodur_kalahari_2a

Fig. 4.5 Synthetic rendering for the SRT IV surface. The spectral reflectance contains 4 materials as labeled by the material distribution mask. The material appearances are visualized under natural illumination [115].

reflectances. This chapter gives a more realistic synthetic dataset with measured spectral reflectances. Similar to the synthetic shape and lighting direction distribution in [13], we choose BUNNY as our target shape and regularly 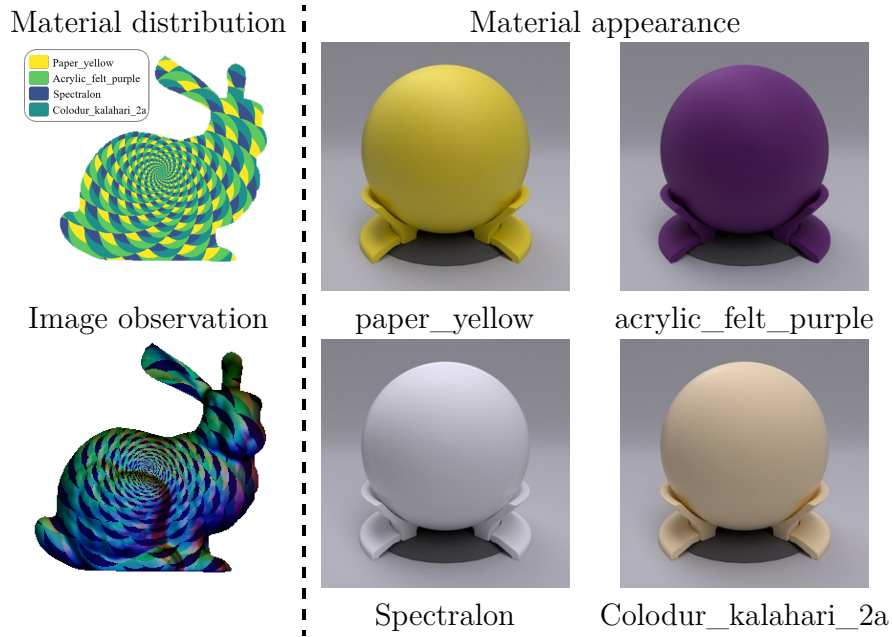sampled 24 synthetic light directions on a hemisphere with the elevation angle larger than 45°. The light spectra of the LEDs are narrow-band with the central wavelengths distributed evenly in the range between 400 ∼ 750 [nm].

To render the reflectance with SRT III, we choose a measured spectral BRDF "paper_yellow" [115], whose appearance is visualized under a natural illumination in Fig. 4.4. As shown in the middle row of the figure, we plot part of the spectral reflectance curves $R(\lambda)$ of the material under varying groups of surface normals and light directions. It is clear that most reflectance curves can be approximated by scaling the thick yellow curve labeled as chromaticity $C(\lambda)$, except for a few curves. Therefore, surfaces rendered with "paper_yellow" roughly have a uniform chromaticity but spatially-varying albedos (SRT III). Following the above rendering setting, we generate a synthetic multispectral image with 24 channels. The observations under

LEDs 1, 11, and 23 are visualized in the bottom row.

To render the reflectance with SRT IV, we select 4 different measured spectral BRDFs as shown in Fig. 4.5. The material distribution labels in the left-top indicate which BRDF to be applied to the regions on the Bunny surface. We render a synthetic multispectral image under the 24 lights and visualize it by concatenating the spectral channels illuminated by LEDs 1, 11, and 23, as shown at the left-bottom of the figure.

Baselines   As the baseline of the experiments, we selected two state-of-the-art MPS methods: CS16 [114] and OS18 [113], where we implemented OS18 [113] and used released code of CS16 [114] for evaluation. Since both methods take a 3-channel (*i.e.*, RGB) image as input, we selected 3 out of 24 spectral observations to mimic the 3-channel input image, as shown in Fig. 4.4. To verify the MLC, we tested our method for SRT III surfaces by assigning the spectral channels recording the observations under LEDs 1, 11, 21, and 23, which cover the observations used in OS18 [113] and CS16 [114] for comparison. The number of piece-wise constant chromaticities need to be set manually in CS16 [114]. To make a fair comparison, we set the number of chromaticities to be 1 and evaluate their method, Ours_{III}, and the SRT II module of OS18 [113] in the experiments of SRT II and III surfaces. When making comparisons on SRT IV surfaces, we use the default number of chromaticity clusters to 100 in CS16 [114], and compare it with Ours_{IV} and the SRT IV module of OS18 [113]. Besides, in the synthetic experiments, we remove the materials used in the test data from the spectral reflectance database when extracting the bases for Ours_{IV}.

In the following, Ours_{III} and Ours_{IV} are given observations under all 24 lights by default. $\text{Ours}_{\text{III}}(f_4)$ denotes our method for SRT III surfaces under MLC.

4.5.2   Surface normal estimation under SRT III

Using the ground-truth surface normal, we evaluated surface normal estimation accuracy by mean angular errors (MAE) in degree. Figure 4.6 shows the results of surface normal estimation for a synthetic SRT III surface. Ours_{III} achieves the smallest an-
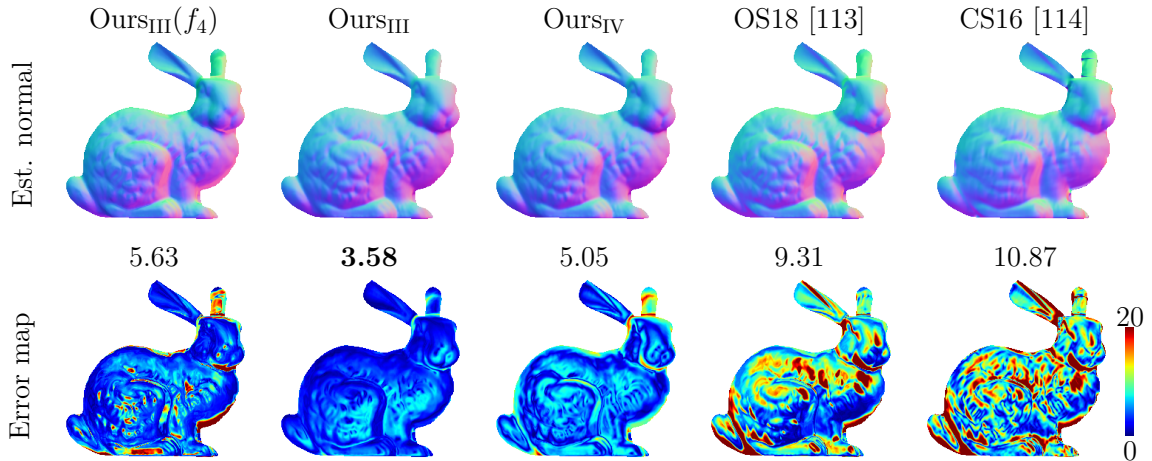
Fig. 4.6 Surface normal estimation results for an SRT III surface shown in Fig. 4.4.

gular error compared to the other methods. The estimation errors of OS18 [113] and CS16 [114] are mainly caused by their SRT II assumption and shadows. Also, the local polynomial shape regularization used in CS16 [114] additionally brings in errors in regions with large surface normal variations. $\text{Ours}_{\text{III}}(f_4)$ under MLC is less accurate than Ours_{III} due to the influence of shadows. However, compared with OS18 [113] and CS16 [114], $\text{Ours}_{\text{III}}(f_4)$ achieves higher accuracy with only one additional spectral observation appended to the input. This result demonstrates the effectiveness of our method on SRT III surfaces. In this setting, Ours_{IV} is less accurate compared to Ours_{III} due to its flexible representation power for this restricted setting.

4.5.3   Surface normal estimation under SRT IV

Figure 4.7 shows the surface normal estimation results of a surface with spatially-varying spectral reflectance (SRT IV). Ours_{IV} can handle spatially-varying chromaticities and albedos, therefore producing more accurate surface normal recovery compared to Ours_{III} that assumes the uniform chromaticity. Compared to OS18 [113] and CS16 [114], Ours_{IV} obtains the smallest angular error since we do not assume piece-wise constant spectral reflectances and require no reflectance clustering. From the error map shown in Fig. 4.7, the error distribution of Ours_{IV} is more uniform and
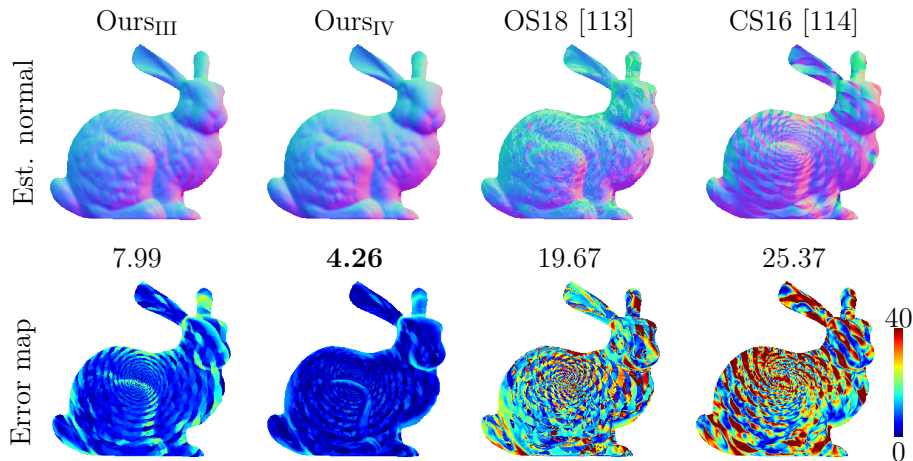
Fig. 4.7 Surface normal estimation comparison on the SRT IV surface shown in Fig. 4.5.

has less correlation to the material distribution compared to the other methods. This result shows the strength of $Ours_{IV}$ on surfaces with spatially-varying reflectances.

4.6 Experiments on real dataset

To assess the effectiveness of the proposed methods, we built a multispectral photometric stereo setup to conduct experiments on real data.

4.6.1 Hardware setup

Figure 4.1 (left) shows our multispectral photometric stereo setup, lighting direction and light spectra distributions. Our setup consists of 12 narrow-band spectral light sources and a monochromatic camera (FLIR Blackfly S). The light sources are fixed on a metal frame rig and distributed uniformly around the camera's optical axis to avoid biased light distributions. We calibrated the light directions with a monochromatic mirror ball following the method by Shi *et al.* [9]. The central wavelength of our spectral light sources uniformly spans in the range of $400 \sim 750$ [nm], and they are measured by a spectrometer Sekonic C-800. To verify our method without the influence of crosstalk across wavelength channels, we captured multiple images with a monochromatic camera by turning on each spectral light source one after another.

Spectral observations under LEDs 2, 4, and 10 with the central wavelength 450nm, 550nm, and 650nm are selected to mimic the RGB input for existing 3-channel MPS methods. We used 4 spectral observations under the illumination of LEDs 2, 4, 9, and 11 to verify the MLC of our method for SRT III surfaces (Ours$_{\text{III}}$).

To obtain the baseline surface normal (we call it the ground-truth (GT) surface normal hereafter), we additionally put an LED board that contains 256 white light sources sharing the same spectrum, in a similar manner to CS16 [114]. The GT surface normal is estimated using a conventional Lambertian least-squares photometric stereo [7], and we use it for quantitatively assessing the MPS results.

Spectral calibration For Ours$_{\text{IV}}$, light sources' spectra $E_1, \ldots, E_f$ and camera spectral sensitivity $S_1, \ldots, S_f$ need to be calibrated in the form of a vector of their products $\mathbf{e} = [E_1(\lambda_1)S_1(\lambda_1), \cdots, E_f(\lambda_f)S_f(\lambda_f)]^\top$. For the calibration, we use a Mac-Beth ColorChecker board [125] consisting of 24 patches of uniform spectral reflectances $R_1, \ldots, R_{24}$. Based on the image formation model of Eq. (4.14), the ratio of the vector $\mathbf{e}$'s elements at neighboring spectral channels follows

$$\frac{e_{j+1}}{e_j} = \frac{m_{j+1}}{m_j} \frac{R(\lambda_j)}{R(\lambda_{j+1})} \frac{\mathbf{l}_j^\top \mathbf{n}}{\mathbf{l}_{j+1}^\top \mathbf{n}}. \tag{4.21}$$

For a scene point on the ColorChecker board, the spectral reflectance ratio $\frac{R(\lambda_j)}{R(\lambda_{j+1})}$ under different wavelengths is known from measured spectral reflectance curves [126]. The surface normal $\mathbf{n}$ of the ColorChecker board can be estimated by the detected image corners and camera intrinsics [127]. With calibrated lighting directions $\mathbf{L}$ and the multispectral observations $\mathbf{m}$, we estimate the elements of $\mathbf{e}$ up to scale by solving the homogeneous system of equations derived from Eq. (4.21) using all the 24 monochromatic patches of the ColorChecker board. Since we can only recover $\mathbf{e}$ up to scale, the spectral reflectance estimation by Ours$_{\text{IV}}$ naturally has a scale ambiguity, but that does not influence the recovery of surface normals.
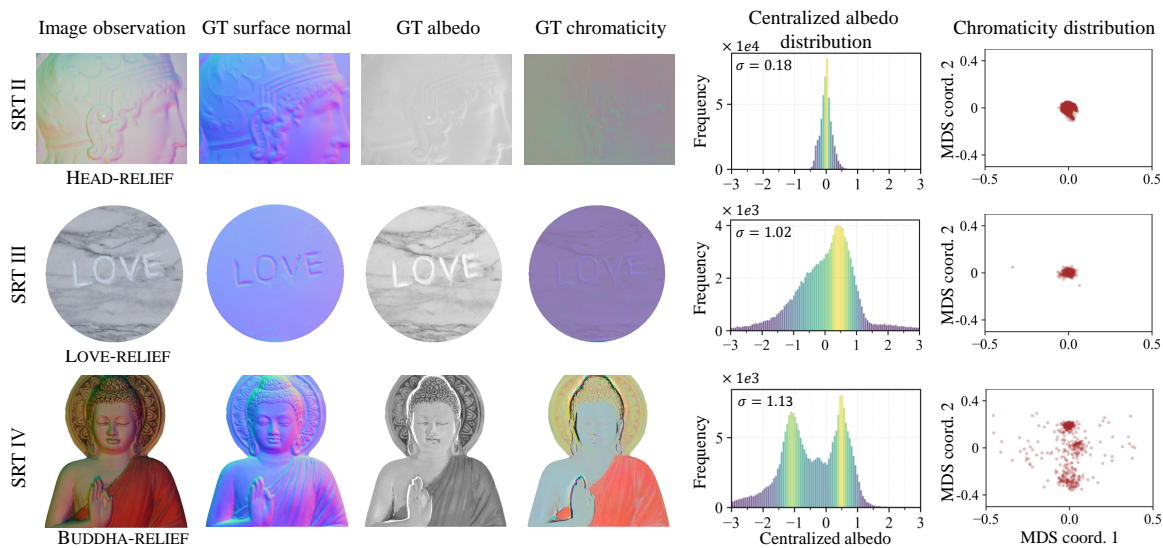
Fig. 4.8 Ground-truth surface normal, chromaticity, and albedo of three real objects: Head-relief, Love-relief and Buddha-relief, where the chromaticity is visualized by mapping 450nm, 550nm and 650nm responses to BGR color channels, respectively. The spectral reflectances for the three reliefs can be categorized as SRT II to IV from top to bottom, as seen by their centralized albedo histograms and the distributions of the chromaticities projected to the 2D space via MDS [128].

4.6.2   Real data setup

Based on our hardware setup, we capture a variety of objects for real data experiments. Prior to the experiment, we examine the SRTs of the scenes by analyzing their spectral reflectance distributions, as shown in Fig. 4.8. With calibrated $\mathbf{e}$, known light directions $\mathbf{L}$ and the ground-truth surface normal $\mathbf{n}$, we compute the spectral reflectance $\mathbf{r}$ based on the spectral image formation model shown in Eq. (4.14). The estimated reflectance $\mathbf{r}$ is further decomposed into the albedo and chromaticity by taking its norm as albedo and its direction as chromaticity as depicted in Fig. 4.8 as GT albedo and GT chromaticity, respectively. The chromaticity is visualized by mapping the responses at 450nm, 550nm, and 650nm to BGR color channels, respectively.

The last two columns of Fig. 4.8 show the histogram of centralized albedo by subtracting the mean value, and low-dimensional visualization of chromaticity distributions via multidimensional scaling (MDS) [128], respectively. The Head-relief has a relatively uniform albedo compared to the Love-relief and Buddha-relief

since its standard derivation σ of albedos is smaller than the other two. This is also consistent with the image observations shown in the first column. On the other hand, the chromaticity distribution of BUDDHA-RELIEF is more diverse than those of HEAD-RELIEF and LOVE-RELIEF, which indicates the spatially-varying chromaticity distribution in BUDDHA-RELIEF. As such, the spectral reflectances of the three real reliefs roughly follow SRT II, III, and IV.

We also observed that piece-wise constant spectral reflectance assumption used in [108, 113, 114] is relatively unpractical to approximate the general SRT IV surfaces. Although BUDDHA-RELIEF seems contain three piece-wise monochromatic regions from the image observation under natural illumination, it actually has diverse chromaticities, making the monochromatic region clustering [108, 113, 114] unstable.

4.6.3   Surface normal estimation results on real data

Based on the captured multispectral image data, we then access the surface normal estimation accuracy of our method.

Surface normal estimation under SRT III   As shown in Fig. 4.9, we compare our methods with baselines on three objects: HEAD-RELIEF, LOVE-RELIEF, and MOAI STATUE. The HEAD-RELIEF scene follows SRT II, and LOVE-RELIEF and MOAI STATUE follow SRT III. Since both existing methods [113, 114] and our methods (Ours$_{\text{III}}$, Ours$_{\text{IV}}$) can handle SRT II, the accuracy of recovered surface normals are comparable.

We observed large normal estimation errors by CS16 [114] and OS18 [113] on the LOVE-RELIEF and the MOAI STATUE, since the spatially-varying albedos violate the assumptions made in their methods. The error maps of CS16 [114] and OS18 [113] on the LOVE-RELIEF highlight the error regions due to the non-uniform albedo distribution. On the other hand, Ours$_{\text{III}}$ yields more accurate surface normal estimation results, which verifies our method's strength on SRT III surfaces. Under minimal solvable lighting conditions (MLC), the estimation errors of Ours$_{\text{III}}(f_4)$ increase compared
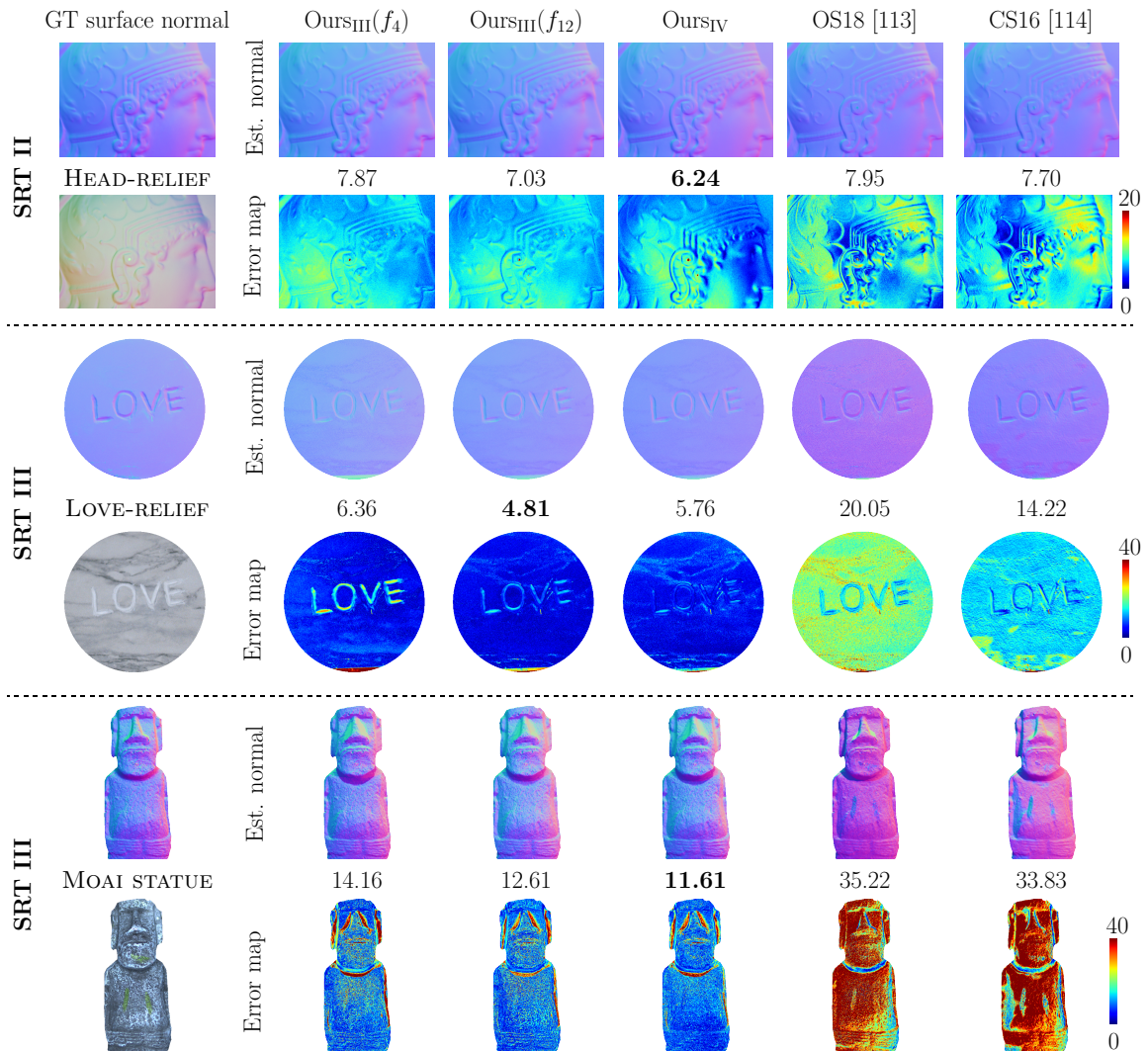
Fig. 4.9 Surface normal estimation results for real-world objects with SRT II (Head-relief) and SRT III (Love-relief and Moai statue).

to using all the 12 lights (Ours$_{III}(f_{12})$, which is mainly caused by the shadows at the concave regions.

Ours$_{IV}$ provides comparable results with Ours$_{III}$ on both SRT II and III. However, Ours$_{IV}$ requires the spectral calibration of both lights and camera as well as the spectral reflectance bases. Therefore, it is preferred to apply Ours$_{III}$ for monochromatic surfaces.

Surface normal estimation under SRT IV   Figure 4.10 shows surface normal estimation results of three SRT IV surfaces: Buddha-relief, Lion, and Puppy. CS16 [114] and OS18 [113] assume the surface contains a limited number of regions with uniform spectral reflectances. However, based on the distribution of albedos and chromaticities shown in Fig. 4.8, such assumption is invalid in the Buddha-relief. Also, it is difficult to infer the number of distinct albedos and chromaticities in the Lion and Puppy from the image observation. Therefore, both methods results in inaccurate surface normal estimates for these scenes.

Ours$_{\mathrm{III}}$ cannot handle spatially-varying chromaticities and outputs large errors on both scenes as well. On the other hand, the proposed method Ours$_{\mathrm{IV}}$ achieves accurate results because it explicitly accounts for the SRT IV surfaces. From the error map, it is seen that inaccurate surface normal estimates are mainly located at the regions where shadows are observed, and the surface normal estimation accuracy is not influenced by the spatially-varying reflectances in the results of Ours$_{\mathrm{IV}}$.

4.6.4   Discussion

In this section, we discuss our method's robustness against outliers and applicability to dynamic scene reconstruction.

4.6.5   Robustness against outliers

Although previous methods [113, 114] provide a unique solution for SRT II without external priors, their input is restricted to 3-channel RGB image and cannot take more bands (see appendix). On the other hand, our methods for both SRT III and SRT IV surfaces can handle multispectral images with 4 or more spectral channels. This capability of taking many spectral channels allows us to use a *robust* estimation approach in MPS, in a similar spirit to four or more source photometric stereo methods [47, 129, 130], to make our method robust against shadows and specular highlights. Namely, having more spectral channels allows us to discard some of them
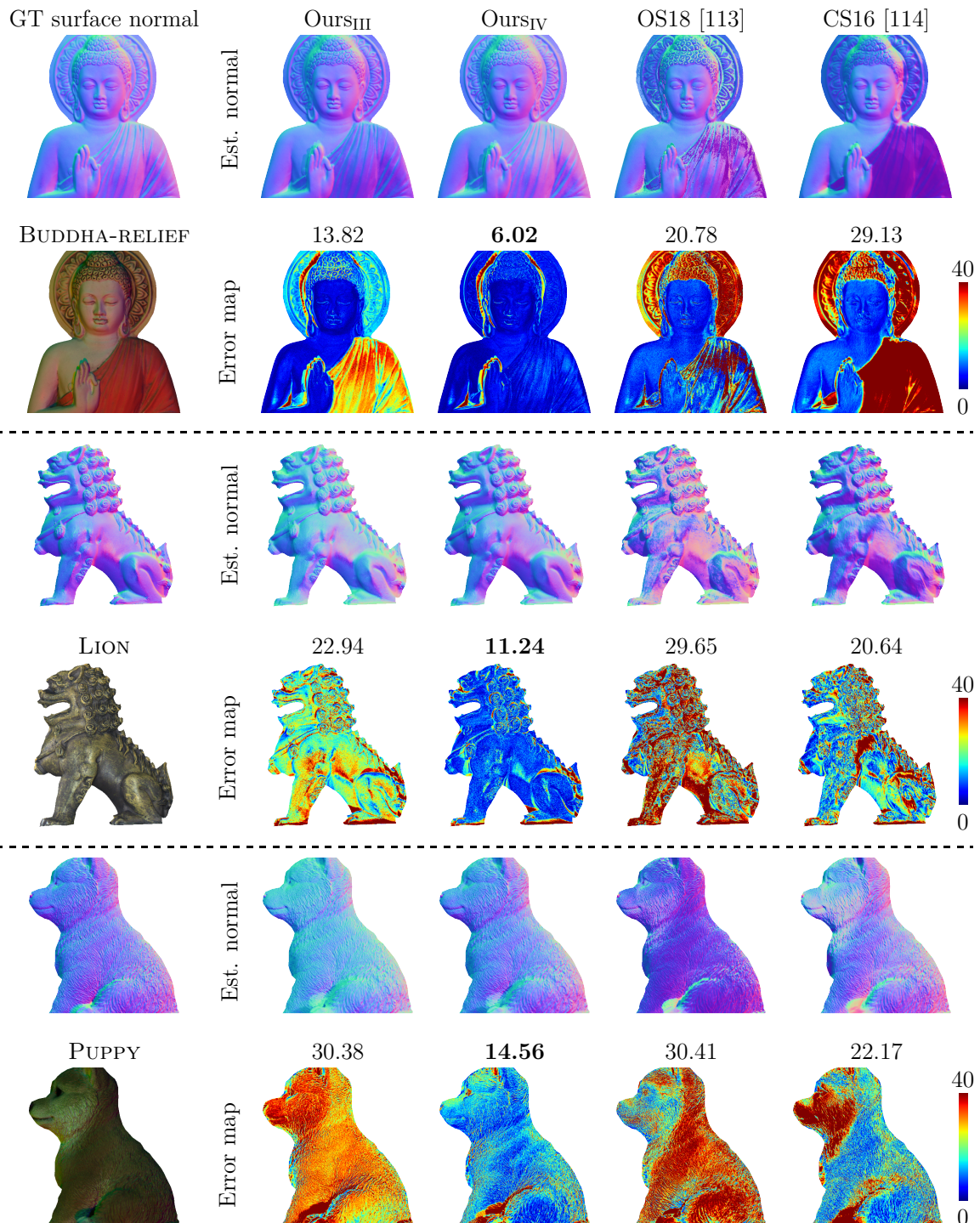
Fig. 4.10 Surface normal estimation results for surfaces with spatially-varying chromaticities and albedos (SRT IV): BUDDHA-RELIEF, LION, and PUPPY.
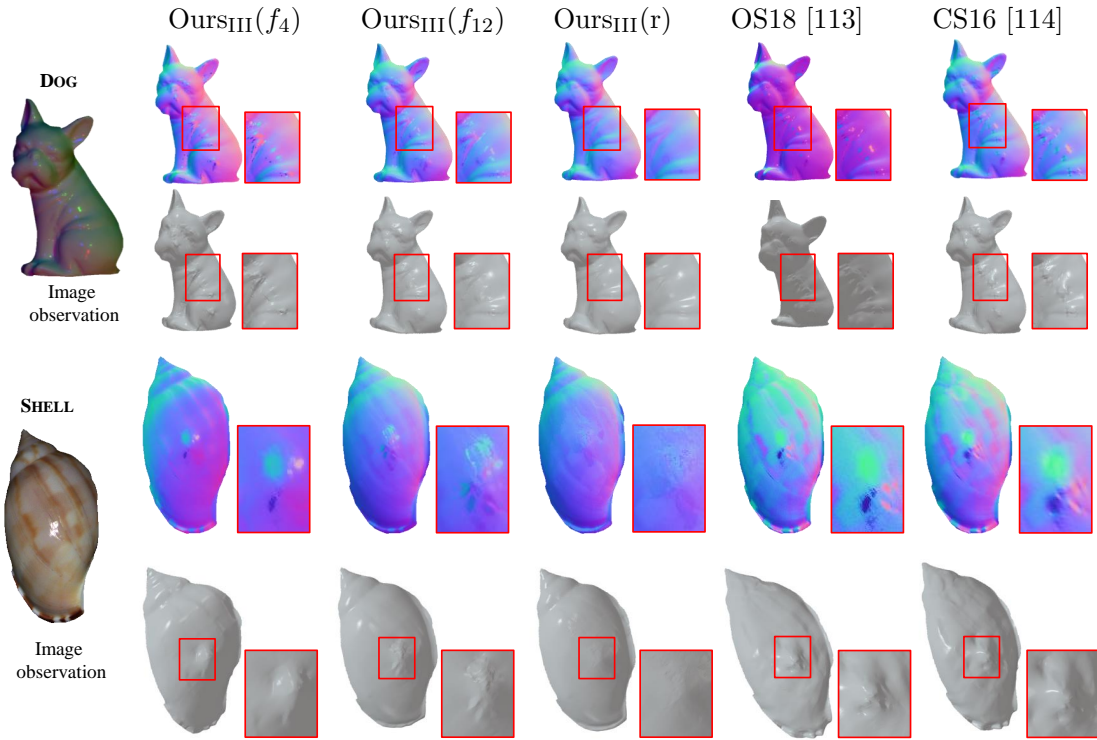
Fig. 4.11 Shape estimation results of two shiny objects, where DOG is monochromatic and SHELL has spatially-varying chromaticities. Even rows show estimated surface normals. Odd rows provide reconstructed surfaces integrated from the surface normal maps. Closed-up views show the artifacts caused by the specular highlights.

that are corrupted by outliers.

To demonstrate this capability, we use a simple position thresholding strategy used in [9, 47] to reject outliers in the input observations. Specifically, for each pixel, we sort the observations and discard shadows and specular highlights as outliers that correspond to dark and bright observations. The surface normal and the spectral reflectance can then be estimated using the inlier image observations. In the following, we denote the robust version of our SRT III method as "$\text{Ours}_{\text{III}}(\text{r})$," and our SRT IV method as "$\text{Ours}_{\text{IV}}(\text{r})$."

In Fig. 4.11, we test this robust estimation method in comparison to our non-robust version and previous methods on objects with shiny surfaces: DOG and SHELL. Since the reflectances of the two objects significantly deviate from the Lambertian reflectance, we cannot trust the surface normal estimated from conventional least-

squares photometric stereo [7] as the ground truth. Therefore, instead of comparing the surface normal maps, we applied a surface normal integration method [45] to reconstruct 3D shapes from estimated surface normals for a qualitative comparison.

As shown in Fig. 4.11, the recovered surface shape from few spectral image observations is heavily influenced by specular highlights (OS18 [113], CS16 [114] taking 3 spectral bands, and our methods (Ours$_{\text{III}}(f_4)$, Ours$_{\text{IV}}(f_4)$ taking 4 spectral bands). We also observe shape distortions at the middle region of Shell in the result of OS18 [113] and CS16 [114]. These are caused by the inaccurate chromaticity clustering for the spatially-varying reflectances. By adding more spectral bands under varying lighting directions as input (Ours$_{\text{III}}(f_{12})$ and Ours$_{\text{IV}}(f_{12})$), shape recovery becomes more plausible. However, artifacts caused by specularities still remain. By further discarding outlier of specular highlights, more convincing shape reconstruction results are obtained from the robust version of our method (Ours$_{\text{III}}(r)$ and Ours$_{\text{IV}}(r)$).

4.6.6   Dynamic shape recovery

We further test the applicability of our method to dynamic scenes using an industrial multispectral camera IMEC-SM-VIS[2], with which image observations at different spectral bands are obtained at once in one shot. As shown in Fig. 4.12, we estimate the dynamic shape of a deformable SRT IV surface in motion[3] and compare the result with OS18 [113] and CS16 [114]. We choose four pairs of spectral lights and camera channels having the strongest response at 480nm, 520nm, 590nm, and 635nm to obtain the multispectral input. Three out of the four channels at 480nm, 520nm, and 635nm are used as the input for OS18 [113] and CS16 [114]. The recovered shapes of OS18 [113] and CS16 [114] are influenced by the spatially-varying reflectance, especially at the boundary of multi-chromatic regions, as shown in the close-up views. Also, the shape details are lost in CS16 [114] due to the polynomial local shape constraint. On the other hand, the surface normal estimates of Ours$_{\text{IV}}$ are unaffected by

---

[2]`https://www.argocorp.com/cam/special/IMEC/IMEC_snapshot.html`. Retrieved Mar. 11, 2021

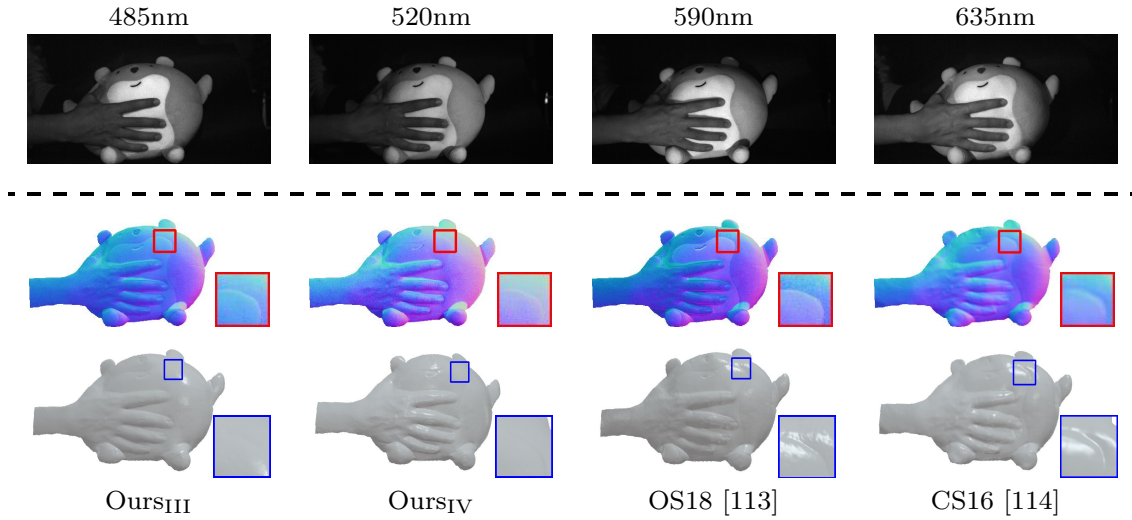[3]Please refer to the supplementary video.

Fig. 4.12 Dynamic shape recovery of a deforming surface with SRT IV. The first row shows the image observations of a multispectral video frame at varying bands. The last two rows provide the estimated surface normals and integrated surfaces. Closed-up views highlight shape distortions caused by spatially-varying reflectance.

the spatially-varying spectral reflectance. As a result, Ours$_{IV}$ achieves more reasonable dynamic shape recovery results on the deformable SRT IV surface. The dynamic shape recovery from our method has a potential to capture 3D movement and gesture of the human body, which may benefit the preservation of intangible cultural heritages such as traditional dances.

4.7   Conclusion

In this chapter, we show that MPS can be turned into a well-posed problem and provide unique solutions for surface normals under two general spectral reflectance types. Specifically, if the surface has uniform chromaticity but spatially-varying albedos (SRT III), we show that surface normal can be uniquely determined from 4+ spectral observations without introducing external priors. By further calibrating the light spectra and the camera spectral sensitivity, we present a closed-form solution of surface normal and spectral reflectance for surfaces with spatially-varying chromaticities and albedos (SRT IV), using a low-rank basis representation of the spectral

reflectance. Since our methods can take more than 4 spectral channels, our method can rely on outlier rejection strategies in the MPS setting to effectively remove shadows and specular highlights. Based on the experiments on synthetic and real datasets, we demonstrate the effectiveness of our MPS method on both SRT III and SRT IV.

Future work To obtain surface shape from a single-shot image, we encode image observations under varying illuminations at different spectral bands. Compared to the setting in CPS, we require a negligible crosstalk effect [113, 114], *i.e.*, each spectral channel only records the image measurement under the corresponding spectral light. From a practical viewpoint, it is interesting to solve MPS with the crosstalk effect being considered, which alleviates the requirement of the hardware setting in MPS. Besides, our MPS method is based on Lambertian reflectance assumption and treats specular highlights as outliers. It is also interesting to explore the MPS solution under general non-Lambertian spectral reflectances.

Chapter 5

Conclusions

This thesis focuses on lighting analysis of photometric stereo for high-fidelity shape reconstruction, aiming to bring photometric stereo out of laboratory light settings. Back to Table 1.1 and Fig. 1.2 in Chapter 1, classical photometric stereo assumes distant light with uniform light emission, requires calibration of the lighting direction and intensity, requires multiple shots to record image observations under varying illuminations, and assumes each image observation is illuminated by a single light source. In contrast, the goal of practical photometric stereo is removing all the above lighting assumptions so that detailed shapes can be recovered from a one-shot image observation captured under an uncalibrated natural environment, where the scene illumination from real-world near-light sources is accurately modeled. However, removing all the lighting assumptions to achieve practical photometric stereo is too challenging. Instead, this thesis divides practical photometric stereo into three sub-problems and presents photometric stereo solutions under the following three lighting conditions to approach the goal:

- Near point light with anisotropic emission (Chapter 2)

- Uncalibrated natural light (Chapter 3)

- Multispectral light for one-shot shape recovery (Chapter 4).

A brief summary of the algorithms and techniques introduced is given below.

To remove the requirement of infinite distant light, we analyze photometric stereo under practical anisotropic near point light in Chapter 2. We introduce a flexible radiant intensity distribution (RID) model based on the observation that local RID is smooth and can be represented by spherical harmonic bases. The proposed RID model shows higher fitting accuracy on real world RIDs compared to existing methods. We further propose a self-calibrating near light photometric stereo which solves the RID and surface shape jointly via a linear system. Promising results have been achieved on both synthetic and real data, which clearly demonstrate the effectiveness of the proposed approach on practical near point light illumination.

To remove the requirement of darkroom and light calibration, we analyze photometric stereo under uncalibrated natural light in Chapter 3. We introduce an equivalent directional lighting model to solve the local surface normal up to an orthogonal ambiguity under the unknown environment light. We further propose a consistent orthogonal condition to cluster the consistent surface normal pairs and build the connections between patches. Based on the relationship between local surface patches, a graph-based patch merging algorithm is introduced to solve the per-patch orthogonal ambiguity. Results on diverse real datasets clearly show that our method outperforms previous uncalibrated natural light photometric stereo methods without requiring a fine initial shape.

To enable photometric stereo for dynamic shape recovery, we analyze photometric stereo under multispectral light, also called multispectral photometric stereo (MPS), in Chapter 4. MPS is known as an ill-posed problem. We show that the MPS can be made well-posed for monochromatic surfaces based on our newly proposed formulation, and we derive a unique and closed-form solution for the surface normal estimation in MPS. In addition, we show that a more general setting of spatially-varying both chromaticities and albedos can become well-posed if the light spectra and camera spectral sensitivity are calibrated. For this general setting, we also give a unique and closed-form solution for MPS using the linear bases extracted from a spectral reflectance database. Experiments on both synthetic and real datasets demonstrate

that our method recovers surface shape from a single-shot multispectral image, and achieves dynamic shape recovery from a multispectral video sequence.

Future Work Although the methods proposed in this thesis are novel and achieve promising results on their specific tasks, there are rooms for each chapter to achieve the goal of practical photometric stereo. Specifically, Chapter 2 models the real-world near-light emission, but it still requires light position calibration and capturing data in a darkroom via multiple shots. Chapter 3 is free of the light calibration and darkroom, but it assumes a distant light model and requires multiple shots for image observations under varying natural environments. The varying natural light can be obtained from moving point light sources in an indoor environment [3] or varying daylight illuminations at different moments in a day [2]. Chapter 4 removes the requirement of multiple shots and achieves dynamic shape recovery, but it assumes distant light and requires darkroom and light direction calibration. It is better to further reduce the lighting assumptions and hardware requirements in each chapter to finally achieve practical photometric stereo. Therefore, our future works are listed as follows:

- Multispectral photometric stereo under near-point light

  Our multispectral photometric stereo method assumes the illumination to be distant spectral light. It is very interesting to combine the methods in Chapter 2 and Chapter 4 to achieve dynamic shape recovery under practical anisotropic near point lights.

- Uncalibrated multispectral photometric stereo & near-light photometric stereo

  In Chapter 2 and Chapter 4, we assume the point light positions and spectral lighting directions are calibrated. It is useful to develop the corresponding algorithms without light calibration since it requires additional efforts and is not friendly for the non-expert.

- Photometric stereo under non-Lambertian reflectance

In this thesis, we assume the surface follows Lambertian reflectance and mainly make efforts on photometric stereo under the practical light setting. However, real world surfaces have more general non-Lambertian reflectance. By considering the general reflectance in photometric stereo, the proposed methods in this thesis could output more accurate surface normal estimation in real scenarios.

References

[1] Y. Hold-Geoffroy, J. Zhang, P. F. U. Gotardo, and J.-F. Lalondey, "x-hour photometric stereo," in *3DV*, 2015.

[2] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan, "Outdoor photometric stereo," in *ICCP*, 2013.

[3] B. Haefner, S. Peng, A. Verma, Y. Quéau, and D. Cremers, "Photometric depth super-resolution," *TPAMI*, 2019.

[4] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *SIGGRAPH*, 2006, pp. 835–846.

[5] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[6] B. K. Horn, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," *Technical Report 232*.

[7] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, 1980.

[8] W. M. Silver, "Determining shape and reflectance using multiple images," Ph.D. dissertation, Massachusetts Institute of Technology, 1980.

[9] B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo," *TPAMI*, 2019.

[10] H. Santo, M. Waechter, M. Samejima, Y. Sugano, and Y. Matsushita, "Light structure from pin motion: Simple and accurate point light calibration for physics-based modeling," in *ECCV*, 2018.

[11] H. Guo, B. Shi, M. Waechter, and Y. Matsushita, "Self-calibrating near-light photometric stereo under anisotropic light emission," in *Image Recognition and Understanding*, 2020.

[12] H. Guo, Z. Mo, B. Shi, F. Lu, S. K. Yeung, P. Tan, and Y. Matsushita, "Patch-based uncalibrated photometric stereo under natural illumination," *TPAMI*, no. 01, pp. 1–1, 2021.

[13] H. Guo, F. Okura, B. Shi, T. Funatomi, Y. Mukaigawa, and Y. Matsushita, "Multispectral photometric stereo for spatially-varying spectral reflectances: A well posed problem?" in *CVPR*, 2021, pp. 963–971.

[14] A. Meydenbauer, "Die Photogrammetrie," *Wochenblatt des Architektenvereins zu Berlin*, vol. Jahrgang I, no. 49, pp. 471–472, 1867.

[15] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Robust photometric stereo using sparse regression," in *CVPR*, 2012.

[16] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan, "Self-calibrating photometric stereo," in *CVPR*, 2010.

[17] R. Mecca, Y. Quéau, F. Logothetis, and R. Cipolla, "A single-lobe photometric stereo approach for heterogeneous material," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1858–1888, 2016.

[18] Y. Quéau, B. Durix, T. Wu, D. Cremers, F. Lauze, and J.-D. Durou, "Led-based photometric stereo: Modeling, calibration and numerical solution," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 3, pp. 313–340, 2018.

[19] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. So Kweon, "Calibrating a non-isotropic near point light source using a plane," in *CVPR*, 2014, pp. 2259–2266.

[20] M. Visentini-Scarzanella and H. Kawasaki, "Simultaneous camera, light position and radiant intensity distribution calibration," in *Image and Video Technology*. Springer, 2015, pp. 557–571.

[21] F. Logothetis, R. Mecca, and R. Cipolla, "Semi-calibrated near field photometric stereo," in *CVPR*, 2017, pp. 941–950.

[22] M. W. Powell, S. Sarkar, and D. Goldgof, "A simple strategy for calibrating the geometry of light sources," *TPAMI*, vol. 23, no. 9, pp. 1022–1027, 2001.

[23] K. Hara, K. Nishino *et al.*, "Light source position and reflectance estimation from a single view without the distant illumination assumption," *TPAMI*, vol. 27, no. 4, pp. 493–505, 2005.

[24] K.-Y. K. Wong, D. Schnieders, and S. Li, "Recovering light directions and camera poses from a single sphere," in *ECCV*. Springer, 2008, pp. 631–642.

[25] T. Takai, A. Maki, K. Niinuma, and T. Matsuyama, "Difference sphere: an approach to near light source estimation," *CVIU*, vol. 113, no. 9, pp. 966–978, 2009.

[26] A. Bunteong and N. Chotikakamthorn, "Light source estimation using feature points from specular highlights and cast shadows," *International Journal of Physical Sciences*, vol. 11, no. 13, pp. 168–177, 2016.

[27] Y. Quéau, T. Wu, and D. Cremers, "Semi-calibrated near-light photometric stereo," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2017, pp. 656–668.

[28] F. Logothetis, R. Mecca, Y. Quéau, and R. Cipolla, "Near-field photometric stereo in ambient light," in *BMVC*, 2016.

[29] I. Moreno and C.-C. Sun, "Modeling the radiation pattern of leds," *Optics express*, vol. 16, no. 3, pp. 1808–1819, 2008.

[30] Y. Iwahori, H. Sugie, and N. Ishii, "Reconstructing shape from shading images under point light source illumination," in *ICPR*, vol. 1. IEEE, 1990, pp. 83–87.

[31] J. J. Clark, "Active photometric stereo," in *CVPR*. IEEE, 1992, pp. 29–34.

[32] J. Ahmad, J. Sun, L. Smith, and M. Smith, "An improved photometric stereo through distance estimation and light vector optimization from diffused maxima region," *Pattern Recognition Letters*, vol. 50, pp. 15–22, 2014.

[33] A. Bony, B. Bringier, and M. Khoudeir, "Tridimensional reconstruction by photometric stereo with near spot light sources," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[34] T. Collins and A. Bartoli, "3d reconstruction in laparoscopy with close-range photometric stereo," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2012, pp. 634–642.

[35] X. Huang, M. Walton, G. Bearman, and O. Cossairt, "Near light correction for image relighting and 3d shape recovery," in *2015 Digital Heritage*, vol. 1. IEEE, 2015, pp. 215–222.

[36] Y. Nie and Z. Song, "A novel photometric stereo method with nonisotropic point light sources," in *ICPR*. IEEE, 2016, pp. 1737–1742.

[37] C.-K. Yeh, N. Matsuda, X. Huang, F. Li, M. Walton, and O. Cossairt, "A streamlined photometric stereo framework for cultural heritage," in *ECCV*. Springer, 2016, pp. 738–752.

[38] R. Mecca, A. Wetzler, A. M. Bruckstein, and R. Kimmel, "Near field photometric stereo with point light sources," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2732–2770, 2014.

[39] R. Mecca, E. Rodolà, and D. Cremers, "Realistic photometric stereo using partial differential irradiance equation ratios," *Computers & Graphics*, vol. 51, pp. 8–16, 2015.

[40] T. Papadhimitri and P. Favaro, "Uncalibrated near-light photometric stereo," in *BMVC*. BMVA Press, 2014.

[41] D. Cho, Y. Matsushita, Y. W. Tai, and I. S. Kweon, "Semi-calibrated photometric stereo," *TPAMI*, 2018.

[42] L. Xie, Y. Xu, X. Zhang, W. Bao, C. Tong, and B. Shi, "A self-calibrated photo-geometric depth camera," *The Visual Computer*, vol. 35, no. 1, pp. 99–108, 2019.

[43] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the Eurographics symposium on Geometry processing*, vol. 7, 2006.

[44] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Variational methods for normal integration," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 609–632, 2018.

[45] W. Xie, Y. Zhang, C. C. Wang, and R. C.-K. Chung, "Surface-from-gradients: An approach based on discrete geometry processing," in *CVPR*, 2014, pp. 2195–2202.

[46] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *TPAMI*, vol. 30, no. 2, pp. 328–341, 2007.

[47] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "Bi-polynomial modeling of low-frequency reflectances," *TPAMI*, 2014.

[48] A. R. Farooq, M. L. Smith, L. N. Smith, and S. Midha, "Dynamic photometric stereo for on line quality control of ceramic tiles," *Computers in industry*, vol. 56, no. 8-9, pp. 918–934, 2005.

[49] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, "Augmented reality technologies, systems and applications," *Multimedia tools and applications*, vol. 51, no. 1, pp. 341–377, 2011.

[50] J. Ackermann, M. Ritz, A. Stork, and M. Goesele, "Removing the example from example-based photometric stereo," in *ECCV*, 2010.

[51] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *TPAMI*, vol. 33, no. 2, pp. 394–405, 2010.

[52] S. Peng, B. Häfner, Y. Quéau, and D. Cremers, "Depth super-resolution meets uncalibrated photometric stereo," in *ICCV Workshops*, 2017.

[53] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi, "Photometric stereo using internet images," in *3DV*, 2014.

[54] B. Haefner, Z. Ye, M. Gao, T. Wu, Y. Quéau, and D. Cremers, "Variational uncalibrated photometric stereo under general lighting," in *ICCV*, 2019.

[55] J. Jung, J.-Y. Lee, and I. S. Kweon, "One-day outdoor photometric stereo using skylight estimation," *IJCV*, vol. 127, no. 8, pp. 1126–1142, 2019.

[56] M. Brahimi, Y. Quéau, B. Haefner, and D. Cremers, "On the well-posedness of uncalibrated photometric stereo under general lighting," in *Advances in Photometric 3D-Reconstruction.* Springer, 2020, pp. 147–176.

[57] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *TPAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[58] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *Journal of the Optical Society of America*, vol. 11, pp. 3079–3089, 1994.

[59] S. Z. Li, "Markov random field models in computer vision," in *ECCV*, 1994.

[60] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *IJCV*, vol. 103, no. 3, pp. 267–305, 2013.

[61] S. Ullman, "The interpretation of structure from motion," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 203, no. 1153, pp. 405–426, 1979.

[62] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita, "Uncalibrated photometric stereo under natural illumination," in *CVPR*, 2018.

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

[64] T. Okabe, I. Sato, and Y. Sato, "Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions," in *ICCV*, 2009.

[65] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "Uncalibrated photometric stereo for unknown isotropic reflectances," in *CVPR*, 2013.

[66] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "Bi-polynomial modeling of low-frequency reflectances," *TPAMI*, vol. 36, no. 6, pp. 1078–1091, 2014.

[67] L. Chen, Y. Zheng, B. Shi, A. Subpa-Asa, and I. Sato, "A microfacet-based refectance model for photometric stereo with highly specular surfaces," in *ICCV*, 2017.

[68] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo," *TPAMI*, 2018.

[69] Y. Mukaigawa, Y. Ishii, and T. Shakunaga, "Analysis of photometric factors based on photometric linearization," *Journal of the Optical Society of America*, vol. 24, no. 10, pp. 3326–3334, 2007.

[70] K. Sunkavalli, T. Zickler, and H. Pfister, "Visibility subspaces: uncalibrated photometric stereo with shadows," *ECCV*, 2010.

[71] D. Miyazaki, K. Hara, and K. Ikeuchi, "Median photometric stereo as applied to the segonko tumulus and museum objects," *IJCV*, vol. 86, no. 2, pp. 229–242, 2010.

[72] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *ACCV*, 2010.

[73] T. Wu and C. Tang, "Photometric stereo via expectation maximization," *TPAMI*, vol. 32, no. 3, pp. 546–560, 2010.

[74] T.-P. Wu, K.-L. Tang, C.-K. Tang, and T.-T. Wong, "Dense photometric stereo: A markov random field approach," *TPAMI*, 2006.

[75] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele, "Photometric stereo for outdoor webcams," in *CVPR*, 2012.

[76] A. Abrams, C. Hawley, and R. Pless, "Heliometric stereo," in *ECCV*, 2012.

[77] J. Jung, J.-Y. Lee, and I. S. Kweon, "One-day outdoor photometric stereo via skylight estimation," in *CVPR*, 2015.

[78] F. Shen, K. Sunkavalli, N. Bonneel, S. Rusinkiewicz, H. Pfister, and X. Tong, "Time-lapse photometric stereo and applications," vol. 33, no. 7, pp. 359–367, 2014.

[79] P. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *IJCV*, vol. 35, no. 1, pp. 33–44, 1999.

[80] A. Yuille, D. Snow, R. Epstein, and P. Belhumeur, "Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability," *IJCV*, vol. 35, no. 3, pp. 203–222, 1999.

[81] T. Papadhimitri and P. Favaro, "A new perspective on uncalibrated photometric stereo," in *CVPR*, 2013.

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

[83] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan, "Self-calibrating photometric stereo," in *CVPR*, 2010.

[84] T. Papadhimitri and P. Favaro, "A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima," *IJCV*, vol. 107, no. 2, pp. 139–154, 2014.

[85] P. Tan, S. P. Mallick, L. Quan, D. Kriegman, and T. Zickler, "Isotropy, reciprocity and the generalized bas-relief ambiguity,," in *CVPR*, 2007.

[86] Z. Wu and P. Tan, "Calibrating photometric stereo by holistic reflectance symmetry analysis," in *CVPR*, 2013.

[87] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Multiview photometric stereo," *TPAMI*, vol. 30, no. 3, pp. 548–554, 2008.

[88] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi, "A hand-held photometric stereo camera for 3-D modeling," in *ICCV*, 2009.

[89] D. Cho, Y. Matsushita, Y.-W. Tai, and I. S. Kweon, "Photometric stereo under non-uniform light intensities and exposures," in *ECCV*, 2016.

[90] M. R. Oswald, E. Töppe, and D. Cremers, "Fast and globally optimal single view reconstruction of curved objects," in *CVPR*, 2012.

[91] L. Shen and P. Tan, "Photometric stereo and weather estimation using internet images," in *CVPR*, 2009.

[92] Y. Quéau, F. Lauze, and J.-D. Durou, "A l1-tv algorithm for robust perspective photometric stereo with spatially-varying lightings," in *International Conference on Scale Space and Variational Methods in Computer Vision*, 2015, pp. 498–510.

[93] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner, "Intrinsic3d: High-quality 3d reconstruction by joint appearance and geometry optimization with spatially-varying lighting," in *ICCV*, 2017.

[94] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general unknown lighting," *IJCV*, vol. 72, no. 3, pp. 239–257, 2007.

[95] J. Ackermann, M. Ritz, A. Stork, and M. Goesele, "Removing the example from example-based photometric stereo," in *ECCV Workshops*, 2010.

[96] P. E. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *SIGGRAPH*, 1998.

[97] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[98] S. J. Koppal and S. G. Narasimhan, "Clustering appearance for scene analysis," in *CVPR*, 2006.

[99] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *TPAMI*, vol. 30, no. 6, pp. 1068–1080, 2008.

[100] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *TPAMI*, vol. 28, no. 10, pp. 1568–1583, 2006.

[101] A. Chatterjee and V. M. Govindu, "Robust relative rotation averaging," *TPAMI*, vol. 40, no. 4, pp. 958–972, 2017.

[102] B. O. Community, "Blender - a 3D modelling and rendering package," Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. [Online]. Available: http://www.blender.org

[103] J. Oliensis, "Uniqueness in shape from shading," *IJCV*, vol. 6, no. 2, pp. 75–104, 1991.

[104] B. Ghojogh, F. Karray, and M. Crowley, "Eigenvalue and generalized eigenvalue problems: Tutorial," *arXiv preprint arXiv:1903.11240*, 2019.

[105] Y. Xiong, A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. Zickler, "From shading to local shape," *TPAMI*, vol. 37, no. 1, pp. 67–79, 2015.

[106] L. Kontsevich, A. Petrov, and I. Vergelskaya, "Reconstruction of shape from shading in color images," *Journal of the Optical Society of America*, no. 3, pp. 1047–1052, 1994.

[107] R. Anderson, B. Stenger, and R. Cipolla, "Augmenting depth camera output using photometric stereo." *MVA*, pp. 369–372, 2011.

[108] ——, "Color photometric stereo for multicolored surfaces," in *ICCV*, 2011, pp. 2182–2189.

[109] Y. Ju, L. Qi, H. Zhou, J. Dong, and L. Lu, "Demultiplexing colored images for multispectral photometric stereo via deep neural networks," *IEEE Access*, pp. 30 804–30 818, 2018.

[110] Y. Ju, X. Dong, Y. Wang, L. Qi, and J. Dong, "A dual-cue network for multispectral photometric stereo," *Pattern Recognition*, p. 107162, 2020.

[111] Y. Ju, L. Qi, J. He, X. Dong, F. Gao, and J. Dong, "MPS-Net: Learning to recover surface normal for multispectral photometric stereo," *Neurocomputing*, pp. 62–70, 2020.

[112] D. Miyazaki, Y. Onishi, and S. Hiura, "Color photometric stereo using multi-band camera constrained by median filter and occluding boundary," *Journal of Imaging*, no. 7, p. 64, 2019.

[113] K. Ozawa, I. Sato, and M. Yamaguchi, "Single color image photometric stereo for multi-colored surfaces," *Computer Vision and Image Understanding*, pp. 140–149, 2018.

[114] A. Chakrabarti and K. Sunkavalli, "Single-image rgb photometric stereo with spatially-varying albedo," in *3DV*, 2016, pp. 258–266.

[115] J. Dupuy and W. Jakob, "An adaptive parameterization for efficient material acquisition and rendering," *TOG*, vol. 37, no. 6, pp. 1–14, 2018.

[116] C. Hernández, G. Vogiatzis, and R. Cipolla, "Overcoming shadows in 3-source photometric stereo," *TPAMI*, no. 2, pp. 419–426, 2010.

[117] G. Vogiatzis and C. Hernández, "Self-calibrated, multi-spectral photometric stereo for 3d face capture," *IJCV*, no. 1, pp. 91–103, 2012.

[118] M. S. Drew and L. L. Kontsevich, "Closed-form attitude determination under spectrally varying illumination," in *CVPR*, 1994, pp. 985–990.

[119] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla, "Non-rigid photometric stereo with colored lights," in *CVPR*, 2007, pp. 1–8.

[120] D. Antensteiner, S. Stolc, and D. Soukup, "Single image multi-spectral photometric stereo using a split U-Shaped CNN," in *CVPR Workshops*, 2019, pp. 481–483.

[121] G. Fyffe, X. Yu, and P. Debevec, "Single-shot photometric stereo by spectral multiplexing," in *ICCP*, 2011, pp. 1–6.

[122] W. Matusik, "A data-driven reflectance model," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.

[123] R. M. Johnson, "On a theorem stated by eckart and young," *Psychometrika*, vol. 28, no. 3, pp. 259–263, 1963.

[124] W. H. Press, H. William, S. A. Teukolsky, A. Saul, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: The art of scientific computing*, 2007.

[125] C. S. McCamy, H. Marcus, J. G. Davidson *et al.*, "A color-rendition chart," *J. App. Photog. Eng*, vol. 2, no. 3, pp. 95–99, 1976.

[126] M. Mohammadi, M. Nezamabadi, and R. Berns, "A prototype calibration target for spectral imaging," *Association Internationale de la Couleur (AIC)-International Color Association*, 2005.

[127] Z. Zhang, "A flexible new technique for camera calibration," *TPAMI*, vol. 22, no. 11, pp. 1330–1334, 2000.

[128] M. A. Cox and T. F. Cox, "Multidimensional scaling," in *Handbook of data visualization*, 2008, pp. 315–347.

[129] S. Barsky and M. Petrou, "The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows," *TPAMI*, no. 10, pp. 1239–1252, 2003.

[130] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *ACCV*, 2010.