Photometric Stereo for Micro-scale Shape Reconstruction

Boren Li

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Tomonari Furukawa, Chair
A. Lynn Abbott
Andrew J. Kurdila
Saied Taheri
Mehdi Ahmadian

December 8, 2016
Blacksburg, Virginia

Keywords: 3D Reconstruction, Photometric Stereo, Surface Normal Integration, Image Formation
Copyright 2016, Boren Li
Photometric Stereo for Micro-scale Shape Reconstruction

Boren Li

ABSTRACT

This dissertation proposes an approach for 3D micro-scale shape reconstruction using photometric stereo (PS) with surface normal integration (SNI). Based on the proposed approach, a portable cost-effective stationary system is developed to capture 3D shapes in the order of micrometer scale. The PS with SNI technique is adopted to reconstruct 3D microtopology since this technique is highlighted for its capability to reproduce fine surface details at pixel resolution. Furthermore, since the primary hardware components are merely a camera and several typical LEDs, the system based on PS with SNI can be made portable at low cost.

The principal contributions are three folds. First, a PS method based on dichromatic reflectance model (DRM) using color input images is proposed to generalize PS applicable to a wider range of surfaces with non-Lambertian reflectances. The proposed method not only estimates surface orientations from diffuse reflection but also exploits information from specularity owing to the proposed diffuse-specular separation algorithm. Using the proposed PS method, material-dependent features can be simultaneously extracted in addition to surface orientations, which offers much richer information in understanding the 3D scene and poses more potential functionalities, such as specular removal, intrinsic image decomposition, digital relighting, material-based segmentation, material transfer and material classification.

The second contribution is the development of an SNI method dealing with perspective distortion. The proposed SNI is performed on the image plane instead of on the target surface.
as did by orthographic SNI owing to the newly derived representation of surface normals. The motivation behind the representation is from the observation that spatially uniform image points are simpler for integration than the non-uniform distribution of surface points under perspective projection. The new representation is then manipulated to the so-called log gradient space in analogy to the gradient space in orthographic SNI. With this analogy, the proposed method can inherit most past algorithms developed for orthographic SNI. By applying the proposed SNI, perspective distortion can be efficiently tackled with for smooth surfaces. In addition, the method is PS-independent, which can keep the image irradiance equation in a simple form during PS.

The third contribution is the design and calibration of a 3D micro-scale shape reconstruction system using the derived PS and SNI methods. This system is originally designed for on-site measurement of pavement microtexture, while its applicability can be generalized to a wider range of surfaces. Optimal illumination was investigated in theory and through numerical simulations. Five different calibrations regarding various aspects of the system were either newly proposed or modified from existing methods. The performances of these calibrations were individually evaluated. Efficacy of the developed system was finally demonstrated through comprehensive comparative studies with existing systems. Its capability for on-site measurement was also confirmed.
Photometric Stereo for Micro-scale Shape Reconstruction

Boren Li

GENERAL AUDIENCE ABSTRACT

Shapes in our world are three-dimensional (3D). How to measure and digitize shapes in 3D into computer understandable virtual models using cameras is called 3D shape reconstruction in the field of computer vision. This dissertation concerns the problem of 3D shape reconstruction, while concentrates on recovering shapes at micro-meter scale, referred to as 3D micro-scale shape reconstruction. Quantifying 3D shapes at micro-scale is significant for both industry and academia. In industry, quantification of 3D shapes at micro-meter scale can be employed in precision parts manufacturing, industrial quality control and rapid prototyping, whilst in academia, even finer resolution may be required to study the microtopography of a surface, such as for the purpose of investigating the nature of friction between surfaces.

In this dissertation, a systematic solution is given for 3D micro-scale shape reconstruction using techniques called photometric stereo (PS) and surface normal integration (SNI) sequentially. PS estimates surface normals for each pixel-corresponding surface patch using images captured under various illumination directions from a fixed viewpoint. These surface normals are then integrated to reconstruct the surface in 3D via SNI. Based on these general principles, a prototype system was developed. The hardware of the system is simple, mainly contains a color digital single-lens reflex (DSLR) camera with a macro lens, multiple LEDs,
a control circuit and a cover. During operation, the LEDs are sequentially turned on and create different illuminations upon the surface of concern. The DSLR camera simultaneously captures images with one LED lit at a time. Having these images for the target surface under various illuminations, the 3D surface at micro-scale is reconstructed through post-processing by PS with SNI algorithms.

Three principal contributions are presented in this dissertation. First, a PS algorithm using color images is demonstrated to improve the shape reconstruction accuracy and its applicability for a wider range of surfaces with different reflectance properties. The proposed PS algorithm can also estimate material-dependent properties of the surface, making potential applications, such as material classification and inference, feasible. The second contribution is to improve the SNI algorithm to deal with the camera’s perspective distortion. Experimental results suggested that the algorithm has been successful in dealing with the distortion for smooth surfaces. The design and calibration of the prototype system are presented as the third contribution. The system can achieve high data acquisition rate due to its area scanning nature, dense measurements at micro-scale due to the PS with SNI approach, and low-cost due to the simple hardware configurations. Efficacy of the system was demonstrated through comprehensive comparative studies with existing systems. Its capability for on-site measurement was also proven.
Dedication

This dissertation is first dedicated to my beloved grandma, Shoufang Li, who accompanied me during my entire childhood and passed away in 2014 when I was in Virginia Tech for my doctoral degree. I was very indebted to her for her altruistic love and felt very guilty that I could not look after her during her last journey. The dissertation is also dedicated to my parents, Yi Li and Haiou Wang. Without their understanding and encouragement, I could not imagine that I could persist in finishing the degree.
Acknowledgments

To begin with, I sincerely appreciate Prof. Tomonari Furukawa for his guidance and support throughout my Ph.D. career. He brought me a very interesting research topic of photometric stereo that provided me an opportunity to investigate the question raised during my childhood: why the world look like this? Does the world appear the same in everyone’s eyes?

I am also indebted to Prof. Abbott, Prof. Taheri, Prof. Ahmadian and Prof. Kurdila for their services on my advisory committee and suggestions that brought new horizons on my work. I must also thank my masters advisor, Prof. Hooshang Hemami, for his guidance of how to become a good independent researcher.

I really appreciate the various funding sources to support my Ph.D., including Center for Tire Research (CenTire), Electrical and Computer Engineering (ECE) department in Virginia Tech, Railway Technologies Laboratory (RTL), and Tire Consortium. I also appreciate the help of Michael Durr and Joel Lazeration from GOODYEAR in acquisition of the ground truth for the asphalt road sample using their equipment.

I would like to express my sincere thanks to all the past and current CMS lab members for their continuous support and generous help, including Kunjin, Xianqiao, Bonson, Kuya, Stanislaw, Varun, Howard, Affan, Luan, Rich, Chris, Tian, Yazhe, Orson, Hangxin, Murat, Yoon, Mengyu, Mickey, George, Josiah, Cong, Jihong, Diya and Urvi.

Finally, I would like to thank all my great friends in Blacksburg, especially Yuantao Xie, Chuang Su, Dewei Wang and Qinan Li. Their accompanices make my life more colorful in Blacksburg. The happy moments we shared will never fade away and become my best assets that I will treasure most in my heart.
Contents

1 Introduction .................................................. 1
   1.1 Background .............................................. 1
   1.2 Objectives .............................................. 2
   1.3 Proposed Approach ...................................... 2
   1.4 Principal Contributions ................................. 3
   1.5 Publications ........................................... 4
   1.6 Organization ........................................... 5

2 Literature Review ........................................... 8
   2.1 Micro-scale shape reconstruction in 3D ...................... 8
   2.2 Non-Lambertian Photometric Stereo .......................... 11
   2.3 Perspective Surface Normal Integration ...................... 13
   2.4 Summary ................................................ 15

3 Image Formation: The Forward Problem ......................... 16
   3.1 Imaging Geometry ...................................... 17
   3.2 Imaging Photometry .................................... 19
       3.2.1 Radiometric Concepts .............................. 19
       3.2.2 Light Attenuation ................................. 21
       3.2.3 Surface Reflectance ............................... 22
       3.2.4 Image Irradiance Formation ....................... 28
       3.2.5 Image Intensity Formation ....................... 29
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3 Color Image Formation</td>
<td>31</td>
</tr>
<tr>
<td>3.4 Summary</td>
<td>34</td>
</tr>
<tr>
<td>4 DRM-based Color Photometric Stereo</td>
<td>35</td>
</tr>
<tr>
<td>4.1 Problem Formulation and Related Works</td>
<td>36</td>
</tr>
<tr>
<td>4.1.1 Generic Problem of Color PS</td>
<td>36</td>
</tr>
<tr>
<td>4.1.2 Color Image Formation Model</td>
<td>37</td>
</tr>
<tr>
<td>4.1.3 Existing Methods Using DRM</td>
<td>39</td>
</tr>
<tr>
<td>4.2 Color PS Using Diffuse-Specular Separation</td>
<td>42</td>
</tr>
<tr>
<td>4.2.1 Objective Function</td>
<td>42</td>
</tr>
<tr>
<td>4.2.2 Overview of the Proposed Color PS method</td>
<td>43</td>
</tr>
<tr>
<td>4.2.3 Diffuse-Specular Separation</td>
<td>45</td>
</tr>
<tr>
<td>4.2.4 Surface Normal Refinement</td>
<td>49</td>
</tr>
<tr>
<td>4.2.5 Remarks</td>
<td>51</td>
</tr>
<tr>
<td>4.3 Evaluations on Surface Orientation Estimation</td>
<td>52</td>
</tr>
<tr>
<td>4.3.1 Evaluations Using Synthetic Input Images</td>
<td>52</td>
</tr>
<tr>
<td>4.3.2 Evaluations Using Real Images</td>
<td>58</td>
</tr>
<tr>
<td>4.4 Implications</td>
<td>62</td>
</tr>
<tr>
<td>4.4.1 Specular Removal</td>
<td>63</td>
</tr>
<tr>
<td>4.4.2 Intrinsic Image Decomposition</td>
<td>64</td>
</tr>
<tr>
<td>4.4.3 Digital Relighting</td>
<td>65</td>
</tr>
<tr>
<td>4.4.4 Material Transfer</td>
<td>68</td>
</tr>
<tr>
<td>4.5 Conclusions and Future Works</td>
<td>70</td>
</tr>
<tr>
<td>5 Surface Normal Integration Dealing with Perspective Distortion</td>
<td>72</td>
</tr>
<tr>
<td>5.1 Orthographic Surface Normal Integration</td>
<td>73</td>
</tr>
<tr>
<td>5.2 SNI under Perspective Projection</td>
<td>74</td>
</tr>
<tr>
<td>5.2.1 Surface Normal on Image Plane</td>
<td>74</td>
</tr>
<tr>
<td>5.2.2 Perspective Surface Normal Integration</td>
<td>76</td>
</tr>
</tbody>
</table>
5.2.3 Uniqueness of Solution ........................................... 79
5.3 Performance Evaluation ........................................... 81
5.4 Conclusions and Future Works ................................... 85

6 Micro-scale Shape Reconstruction System Using Photometric Stereo 86
6.1 Problem Formulation ................................................... 87
   6.1.1 Microtexture Road Profiling Requirement ..................... 87
   6.1.2 Flows of Road Profiling Using Color PS and SNI ............. 88
   6.1.3 Light Direction Estimation from Light Position ............. 90
6.2 System Design ......................................................... 91
   6.2.1 Design objectives .............................................. 91
   6.2.2 System Overview .............................................. 91
   6.2.3 Optimal Illumination Configuration ............................ 92
   6.2.4 Design of Imaging System ................................... 96
   6.2.5 Design of Control Circuit ................................... 97
   6.2.6 System Integration and Specification .......................... 98
6.3 Calibrations ........................................................... 99
   6.3.1 Camera Geometrical Calibration ............................... 100
   6.3.2 Camera Radiometric Calibration ............................... 100
   6.3.3 Light Position Calibration .................................... 102
   6.3.4 Light Attenuation Calibration ................................. 106
   6.3.5 Specular Color Calibration ................................... 109
6.4 Experimental Results ................................................ 110
   6.4.1 Experiments on Optimal Illumination Configuration .......... 110
   6.4.2 Evaluation on Calibration Methods ............................ 114
   6.4.3 Evaluation of the Developed System ............................ 123
6.5 Conclusions and Future Works ..................................... 131

7 Conclusions and Future Works ................................. 135
List of Figures

3.1 Coordinate systems for imaging geometry modelling .......................... 17
3.2 Radiometric concepts illustration: surface patch illuminated by a point source 20
3.3 Light attenuation modelling ............................................................... 21
3.4 Normal-centered coordinate system for definition of BRDF ................. 24
3.5 100 materials in the MERL database: pictures of the spheres [69] ........ 27
3.6 Relationship between $\{N\}$ and $\{H\}$ (redrawn from [70]) .............. 28
3.7 General flow from image irradiance to image intensity for a color digital camera 29

4.1 Schematic diagram for generic color PS ........................................... 37
4.2 Modelling of light configurations in a circular pattern .......................... 38
4.3 Flows of the proposed color PS method using diffuse-specular separation .. 43
4.4 (a) Rendered image under the first illuminant; (b) Normal map of the scene; (c) Unit diffuse colors of the scene; (d) Angular error of the estimated unit diffuse colors under $T_d = 0.02$ .................................................. 54
4.5 Angular error of the estimated unit diffuse colors when $T_d$ is improperly chosen: (a) Under $T_d = 0.002$; (b) Under $T_d = 0.2$ ............................. 55
4.6 Angular error of the estimated surface orientations: (a) without surface normal refinement; (b) with surface normal refinement. (c) Improvement percentage of surface orientation estimation with surface normal refinement; (d) Histogram of improvement percentage of surface orientation estimation with surface normal refinement ........................................ 56
4.7 Box-and-whisker plot for angular error of surface orientations for 30 different dielectric materials ......................................................... 59
4.8 Evaluations on surface orientation estimation using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Ground truth of the normal map; (c) Angular error of the estimated surface orientations ................................. 60

4.9 Comparison of surface orientation estimation on six datasets using ten different PS methods .................................................................................................................................................. 61

4.10 A schematic illustration for the implications of the proposed DRM-based color PS method ........................................................................................................................................ 62

4.11 Specular removal using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Separated diffuse component; (c) Separated specular component ......................................................................................... 63

4.12 Intrinsic image decomposition using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Estimated albedo map; (c) Estimated shading map ........................................................................................................ 64

4.13 Digital relighting using DiLigenT database: (a) Captured image irradiance under the first illuminant; (b) Digital relighted image under the first illuminant; (c) Error of the digital relighting .................................................................................. 65

4.14 Shadow classification using digital relighting: (a) Captured image; (b) Relighted image; (c) Error of the relighted image; (d) Detected shadows before performing PS (bright pixel represents shadow); (e) Shadow classification results (green indicates cast-shadows and yellow implies self-shadows) ........ 67

4.15 Material-based segmentation based on extracted material-dependent features 67

4.16 Material transfer using BRDF database .............................................................................................................................. 68

4.17 Material transfer from the other captured scenes .............................................................................................................. 69

5.1 The curl-divergence space and the integrability constraint ........................................................................................................ 80

5.2 (a) Visible surface; (b) Normal map ..................................................................................................................................................... 82

5.3 (a) Distance error from OrthSNIG; (b) Distance error from PerSNILog; (c) Distance error from PerSNIG; (d) Histogram of distance error from OrthSNIG; (e) Histogram of distance error from PerSNILog; (f) Histogram of distance error from PerSNIG ................................................................................................................................. 83

5.4 Noise sensitivity of OrthSNIG, PerSNILog and PerSNIG .................................................................................................................. 84

6.1 Microtexture road profiling problem ........................................................................................................................................... 88

6.2 Hardware components and general flows of road profiling using color PS and SNI .................................................................................................................................................. 89
6.3 Hardware components of the developed system for online data capturing 91
6.4 Design and developed control circuit 98
6.5 Calibration sequence 99
6.6 Schematic diagram for explanation of the light direction calibration 103
6.7 Schematic diagram for explanation of the light position determination 105
6.8 Schematic diagram for explanation of the light strength calibration 107
6.9 (a) Normal map of the scene geometry; (b) An example of synthetic input image under \( \eta = 30^\circ \) 111
6.10 Optimal light zenith angle study for Lambertian surface 112
6.11 Optimal number of lights study for Lambertian surface 113
6.12 Optimal light zenith angle study for real material reflectances 114
6.13 Results of camera geometrical calibration: (a) Relative poses between the camera and the checker board; (b) Reprojection errors of checker board corners; (c) Radial distortion; (d) Tangential distortion; (e) Complete distortion 115
6.14 Captured images for the same scene under nine different shutter speed 116
6.15 Estimated camera radiometric response 117
6.16 Histograms of residuals obtained through: (a) original cost function for the red channel; (b) original cost function for the green channel; (c) original cost function for the blue channel; (d) modified cost function for the red channel; (e) modified cost function for the green channel; (f) modified cost function for the blue channel 118
6.17 (a) Image used for light direction calibration under the first illuminant; (b) Box-and-whisker plot of distances from triangulated light positions to each 3D line (light direction); (c) The calibrated light positions in \( \{C\} \) 120
6.18 Histogram of the rerendering percent error using the proposed light attenuation model 121
6.19 (a) Reconstructed surface with light attenuation correction; (b) Reconstructed surface without light attenuation correction 122
6.20 (a) Light SPD; (b) Camera’s spectral sensitivity for each color channel 123
6.21 (a) Measurements for sandpaper of Grit 80; (b) Measurements for sandpaper of Grit 120 124
6.22 Asphalt surface region 1: (a) Estimated normal map from PS; (b) Digital relighted image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM.

6.23 Asphalt surface region 2: (a) Estimated normal map from PS; (b) Digital relighted image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM.

6.24 Asphalt surface region 3: (a) Estimated normal map from PS; (b) Digital relighted image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM.

6.25 Asphalt surface region 4: (a) Estimated normal map from PS; (b) Relighted color image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM.

6.26 FFT analyses comparison between PS and GFM: (a) Region 1; (b) Region 2; (c) Region 3; (d) Region 4.

6.27 (a) The developed system with a cover for on-site measurement; (b) Locations of experiments.

6.28 Measurement results on Location 1-4: (a) Estimated normal map; (b) Relighted image; (c) Reconstructed surface; (d) Residual from PS; (e) Residual from SNI.

6.29 Measurement results on Location 5-8: (a) Estimated normal map; (b) Relighted color image; (c) Reconstructed surface; (d) Residual from PS; (e) Residual from SNI.
List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Radiometric concepts</td>
<td>19</td>
</tr>
<tr>
<td>4.1</td>
<td>Improvement on different-colored spheres</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Evaluation of the improvement for different reflectances</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison with other existing PS methods</td>
<td>61</td>
</tr>
<tr>
<td>4.4</td>
<td>RMS error of digital relighting</td>
<td>66</td>
</tr>
<tr>
<td>5.1</td>
<td>Camera intrinsic parameters</td>
<td>82</td>
</tr>
<tr>
<td>5.2</td>
<td>Mean computation time for each SNI method</td>
<td>84</td>
</tr>
<tr>
<td>6.1</td>
<td>Requirement for microtexture road profiling system</td>
<td>88</td>
</tr>
<tr>
<td>6.2</td>
<td>Estimated system specifications</td>
<td>99</td>
</tr>
<tr>
<td>6.3</td>
<td>Mean residuals during CRF estimation</td>
<td>118</td>
</tr>
<tr>
<td>6.4</td>
<td>Mean rerendering percent error</td>
<td>122</td>
</tr>
<tr>
<td>6.5</td>
<td>Measurement comparison on sandpapers</td>
<td>125</td>
</tr>
<tr>
<td>6.6</td>
<td>Measurement setting comparison between PS and GFM</td>
<td>125</td>
</tr>
<tr>
<td>6.7</td>
<td>Surface types of different measurement locations</td>
<td>132</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Background

Object shapes in our world are three-dimensional (3D). Digitizing real-world shapes in 3D into computer understandable virtual models using cameras is a fundamental problem in computer vision, referred to as 3D shape reconstruction. Recent years have witnessed growing interest within the design, manufacturing and graphics communities in constructing devices for applications using 3D shape reconstruction. At the large scale, applications may include architectural retrofits, virtual reality flythrough and cultural artifacts preservations. At smaller scales, applications using 3D shape reconstruction techniques have proliferated in product design, reverse engineering, museum archiving, movie making, video games and online shopping. Quantifying 3D shapes at micro-scale is also significant for both industry and academia. In industry, quantification of 3D shapes at micro-scale can be employed in precision parts manufacturing, industrial quality control and rapid prototyping, whilst in academia, even finer resolution is required to study the microtopography of a surface, such as for the purpose of investigating the nature of friction between surfaces. As digital camera
hardware evolves, capturing object details at the scale of micro-meters is more feasible using commercial-grade cameras than ever before, which opened the gate for more cost-effective solutions to reconstruct micro-scale shapes in 3D.

1.2 Objectives

The primary objectives of this dissertation are stated as follows:

- To propose a technique for 3D micro-scale shape reconstruction;
- To develop a portable cost-effective stationary system that is capable of capturing 3D shapes on the order of micrometer scale.

1.3 Proposed Approach

In order to achieve these objectives, techniques based on photometric stereo (PS) with surface normal integration (SNI) are adopted to reconstruct 3D microtopology of a surface. The geometrical cue in terms of surface orientation is encoded in every observed image intensity, referred to as shading. PS-based techniques, posed as inverse problems of image formation, utilize shadings to extract surface orientations from a set of images captured under various illuminations at a fixed viewpoint. These surface orientations are then integrated using SNI to reconstruct the surface. Among various 3D shape reconstruction technologies, PS with SNI is highlighted for its capability to reproduce fine details of the surface at pixel resolution. Due to its experimental settings, it is especially adaptable for stationary applications. Since the primary hardware components required by the PS-based systems are merely a camera and several typical light sources, these systems can be made portable at low cost. Therefore,
PS with SNI becomes our top candidate to achieve the objective of 3D micro-scale shape reconstruction.

1.4 Principal Contributions

The principal contributions of this dissertation are three folds:

- A PS method based on dichromatic reflectance model (DRM) using color images is proposed to generalize PS applicable to a wider range of surfaces with non-Lambertian reflectances. The proposed method not only estimates surface orientations from diffuse reflection but also exploits information from specularities owing to the proposed diffuse-specular separation algorithm. Using the proposed DRM-based color PS method, material-dependent features can be simultaneously obtained in addition to surface orientations, which offers much richer information in understanding the 3D scene and poses more potential functionalities, such as specular removal, intrinsic image decomposition, digital relighting, material-based segmentation, material transfer and material classification;

- An SNI method dealing with perspective distortion is developed. The proposed SNI method is performed on the image plane instead of on the target surface normally done by orthographic SNI owing to the newly derived representation of surface normals. The motivation behind the representation is by observing that the spatially uniform distribution of image points is simpler for integration than the non-uniform distribution of surface points under perspective projection. This new representation is then manipulated to the so-called log gradient space in analogy to the gradient space in orthographic SNI. Using the proposed SNI method, perspective distortion can be ef-
ficiently tackled with for smooth surfaces. In addition, the method is PS-independent, which can keep the image irradiance equation in a simple form during PS. Due to the derived analogy, the proposed method is capable of building on most past algorithms developed for orthographic SNI;

- A 3D micro-scale shape reconstruction system is developed using the proposed PS with SNI method. This system is originally designed to measure pavement microtexture on-site, while its applicability can be generalized to a wider range of surfaces. The optimal illumination configuration was investigated. Five different calibrations regarding various aspects of the system were either newly proposed or modified from existing methods. The performance of the developed system was comprehensively evaluated.

1.5 Publications


- Li, Boren and Furukawa, T., Microtexture road profiling using photometric stereo,
Li, Boren and Furukawa, T., DRM-based color photometric stereo using diffuse-specular separation for non-Lambertian surfaces, *IET Computer Vision*, 2017. (submitted)

Li, Boren and Furukawa, T., Design and calibration of an on-site microtexture road profiling system using color photometric stereo, *Journal of Transportation Engineering*, 2017. (submitted)

Li, Boren and Furukawa, T., Surface normal integration under perspective projection, 2017. (under preparation)


1.6 Organization

This dissertation is organized as follows:
• Chapter 2 surveys past works. Relevant 3D micro-scale shape reconstruction technologies are first classified, followed by the related works corresponding to each original contribution.

• Chapter 3 models image formation as the forward problem of PS. Two problems in image formation, imaging geometry and imaging photometry, are separately discussed. The chapter emphasizes on modelling of imaging photometry, including light attenuation, surface reflectance, image irradiance formation and image intensity formation. The color image formation model based on DRM is finally given.

• Chapter 4 presents the proposed DRM-based color PS method. The problem of color PS is first formulated and closely-related works using DRM are summarized. The proposed method is then elucidated. Comprehensive evaluation results of the method are demonstrated next, while implications are finally given.

• Chapter 5 describes the proposed SNI method dealing with perspective distortion. Orthographic SNI is first reviewed as fundamentals, followed by the elucidation of the proposed perspective SNI method. The efficacy of the method is then evaluated via comparative studies using numerical simulations.

• Chapter 6 presents the design and calibration of the micro-scale shape reconstruction system using PS. The problem is first formulated, followed by the design considerations of the system, including illumination configuration, imaging system and control circuit. Five different calibrations for the developed system are then presented, involving camera geometrical calibration, light position calibration, camera radiometric calibration, light attenuation calibration and specular color calibration. Numerical simulation results in determining optimal illumination configurations are first demonstrated, followed by individual evaluations of the five calibrations. Overall performances of the
developed system are then evaluated through comparative studies in the lab environment, whilst its capability for on-site measurements is finally demonstrated.

- Chapter 7 summarizes the major conclusions of the dissertation and proposes future works.
Chapter 2

Literature Review

This chapter reviews past work regarding each original contribution presented in this dissertation. Existing micro-scale 3D shape reconstruction systems are first reviewed and classified, which gives rise to the proposed approach in performing micro-scale shape reconstruction using photometric stereo (PS) with surface normal integration (SNI). Past works aiming at generalization of PS dealing with non-Lambertian reflectances are given next. Related works of SNI dealing with perspective distortion are finally reviewed.

2.1 Micro-scale shape reconstruction in 3D

Existing micro-scale shape reconstruction systems can be categorized into contact and non-contact types. Contact-type systems usually refer to the stylus profilometer, where a stylus tip touches the surface of interest and converts force measurements into displacements of elevation [1][2]. Owing to the small radius of the stylus tip, such systems can achieve extremely high resolution, less than 1\(\mu m\). Furthermore, a stylus profilometer is surface independent, making it applicable to a wide range of rigid surfaces. However, due to its contact and point-
scanning nature, the measurement speed of such system is low. In addition, its measurement range in the tri-dimensional space is very limited. Last but not least, this type of system is costly because of its dependency on complicated hardware and maintenance procedures.

Non-contact type systems typically utilize optical sensors. Based on their scanning methods, the systems can be further classified into three categories: point-scanning-based, line-scanning-based and area-scanning-based. At least three types of system exist with different principles of operation. The first type is based on time-of-flight, where a reference pulse from a laser is emitted to the target surface and the source-surface distance is derived from the received reflected pulse by comparing the time difference between the two pulses \[3\,4\]. System based on time-of-flight is reluctant to achieve high elevation resolution considering the high speed of light and the small elevation variation of the micro-scale surface, though recent advance \[5\] has shown the possibility. The second type of point-scanning-based system employs laser-dot triangulation method \[6\]. The emission source generates a laser beam perpendicular to the surface, while the receiver detects a fraction of the beam reflected by the surface and adopts the optical principle of triangulation to identify the receiver-surface distance. The major advantage of such system is the high accuracy. In addition, such system is relatively insensitive to illumination conditions and surface reflectances. However, in order to achieve higher elevation resolution by increasing the baseline distance between the emitter and receiver, effects of shadowing or masking become more obvious, which results in more blind spots from the measurements. The third type of point-scanning-based system relies on interferometric methods \[7\]. This type of system operates by projecting a spatially or temporally varying periodic pattern onto the surface, followed by mixing the reflected light with a reference pattern which demodulates the signal to reveal the variation of surface geometry. The elevation resolution for such system is very high since it is a fraction of laser wavelength, while the complicated hardware limits its on-site capability outside the
lab environment. Despite all the pros and cons for the above mentioned types of system, the point-scanning nature leads to the inherently low measurement speed which becomes the major bottleneck.

To compensate for the drawback of measurement speed, line-scanning-based and area-scanning-based systems were developed to measure micro-scale shapes in 3D. Line-scanning-based type of system usually refers to profilometers using light sectioning principle \[8\][9][10]. Such system utilizes triangulation similar to the laser-dot system, while instead of projecting a laser dot onto the surface, a light sectioning system lays a thin line or an intensively illuminated band with sharp light edges onto the surface instead. A video camera monitors the line or band from an angle in relation to the laser or light beam and the system extracts distance between the camera and the line or band through triangulation. The line-sectioning-based system can be viewed as an advanced version of the laser-dot-triangulation system due to its higher measurement speed.

In order to further enhance the measurement speed, area-scanning-based systems were proposed. One type of systems is built upon the structured light method which utilizes a camera in monitoring a projected bi-dimensional patterns of non-coherent light on the surface to derive distance information using triangulation. Projection strategies have been extensively studied, including the projection of grid patterns \[11\], dot patterns \[12\], multiple vertical slits \[13\], multi-color projection patterns \[14\] and fringe patterns \[15][16\]. Owing to its area scanning nature, structured light systems are even faster in data acquisition, while as all the triangulation-based methods, the blind-spot problem avoids such system more accepted. Furthermore, the generation of ideal micro-scale light patterns is challenging and not economical, which also limits the proliferation of such type of systems.
### 2.2 Non-Lambertian Photometric Stereo

Photometric stereo (PS) estimates surface orientations using images captured from a fixed viewpoint under various illuminations and is especially powerful in acquisition of fine surface details at pixel level [17]. Surface orientation itself is significant in a variety of fields, such as geometric segmentation for three-dimensional (3D) object recognition [18] and digital rerendering in computer graphics. Surface geometries, which can be obtained via integrating surface orientations, have also been proven useful for applications, such as industrial quality control and reverse engineering. Due to the strength of PS and the significance for surface orientation acquisition, PS has drawn increasing interest since its debut [19]. However, making PS for a general real scene remains challenging due to the diverse reflectance properties of different materials that appear non-Lambertian [20]. This has given rise to the need for reliable estimation of surface orientations for a wide range of non-Lambertian reflectance, which essentially requires a proper imaging photometry model characterizing the forward problem and a subsequent PS method that inversely derives surface orientations.

Existing PS methods dealing with non-Lambertian reflectance can be classified into three categories. The first approximates surface reflectance using analytical bidirectional reflectance distribution function (BRDF) [21] and formulates the estimation of surface orientations as a nonlinear fitting problem. Nayar et al. [22] derived surface orientations using the Torrance-Sparrow BRDF [23] and incorporated extended sources to ensure sufficient information from specularities. Georghiades [24] inverted the same BRDF to simultaneously estimate surface orientations and resolve the generalized bas-relief ambiguity. Goldman et al. [25] assumed that general material reflectance can be represented by a convex combination of fundamental materials characterized by the Ward BRDF [26] and recovered surface orientations, fundamental material BRDFs and weight maps simultaneously for further scene editing purpose. Methods in this category exploiting information from surface reflectance are
capable to derive not only surface orientations but also the other parameters in the analytical BRDFs, allowing more functionalities, such as digital relighting and material classification. However, due to the nonlinearity of analytical BRDFs and larger number of parameters to be estimated, these methods are sensitive to initializations, numerically unstable under heavily corrupted outliers (e.g. shadows), and infeasible with just a small number of observations.

The second category of methods infers surface orientations through adopting the general properties of BRDF, such as isotropy, monotonicity and reciprocity. Alldrin et al. [27] developed a non-parametric PS method using bi-variate approximation of the isotropy property. Higo et al. [28] analyzed the general BRDF constraints and employed the BRDF properties of monotonicity, isotropy and visibility to vote for the most possible surface orientations for single-lobed reflectance. More recently, Shi et al. [29] proposed a bi-polynomial representation for low-frequency reflectance that was especially adaptable to the inverse problem as PS, while Ikehata et al. [30] developed another PS method inverting general isotropic BRDF as sum of lobes with unknown center directions based on [31]. Methods in this category capitalize on the most fundamental properties of BRDF and therefore, have the potential to deal with a broader range of reflectance. However, these methods are only capable to derive surface orientations with limited other functionalities and require a even larger set of observations compared with the first category.

Methods in the third category assume that non-Lambertian effects appear sparsely among observations and treat them as outliers. A substantial corpus of methods rely on robust statistical techniques for outlier rejection. Earlier methods choose three optimal lights out of four where the surface appears mostly Lambertian to estimate normal [32][33]. When more input images are available, the subset of inliers can be determined using more robust manner by RANSAC [34], maximum-likelihood estimation [35], graph cuts [36]. More recently, Wu et al. [37] formulated the PS with outlier rejection as a global rank minimization
problem, while Ikehata et al. [38] employed sparse Bayesian regression instead. Barsky et al. [39] initiated another line of PS researches using color images where they first addressed the significance of using specular color in the dichromatic reflectance model (DRM) [40] for specularity rejection. Using the cue from known specular color under the same theoretical foundation, Zickler et al. [41] derived a PS method in a novel two-dimensional specular invariant color subspace. The major advantages of methods in the third category are their robustness and requirement for less images, while they are inefficient in the presence of dense non-Lambertian effects.

2.3 Perspective Surface Normal Integration

The need for an efficient method of a dense normal field is driven by several computer vision tasks, such as shape from shading [42], deflectometry [43] and PS [19]. Much research followed the seminal work of PS and one of the most fundamental assumptions was the orthographic imaging geometry. Two pioneering works in solving the surface normal integration (SNI) problem were from Horn et al. [44] and Frankot et al.. [45]. Horn et al. solved the problem of orthographic SNI using variational approaches, while Frankot et al. derived a solution using Fourier analysis. Since then, Simchony et al. [46] proposed a direct analytical method via solving the Poisson equations. The problem formulation was well-established after the late 1980s. Researches then concentrated on improving the efficiency of the algorithms and proposing methods to enforce the integrability constraints. Ho et al. [47] introduced a fast and efficient method for orthographic SNI via solving the Eikonal equation using modified fast marching method, followed by Galliani et al. [48] in modifying the framework to a discrete form. Agrawal et al. [49] proposed an algebraic approach to optimize the gradient field to be curl-free before SNI using graph theory, unlike most of the regularization methods,
such as [50][51]. Harker et al. [52][53] argued for a solution satisfying the Lyapunov equation instead of solving the Poisson equation directly to speed up SNI by three orders. Though these methods have been proven effective, systematic error from perspective distortion still occurs when the orthographic projection does not hold.

Existing PS with SNI algorithms dealing with perspective distortion can be largely divided into two categories:

- Include perspective projection in the image irradiance equation and solve perspective distortion during PS [54][55][56];
- Deal perspective distortion in SNI and make it PS-independent [57].

Methods falling into the first category attempted to model perspective projection into the image irradiance equation and tried to invert such models using PS. Lee et al. [54] might be the first to tackle the perspective distortion problem in shape from shading (SfS) and formulated the image irradiance equation with surface parameterized by triangular patches. Tankus et al. [55] parameterized the surface as a point cloud and showed that the image irradiance equation under Lambertian reflectance and perspective projection can be inverted by PS, similar to the orthographic case. Recently, Mecca et al. [56] proposed a direct differential approach to invert the image irradiance equation. These methods were effective in dealing with surfaces with Lambertian reflectance, whilst the image irradiance equation was formulated into a much complex form that was difficult to be inverted if more sophisticated BRDF should be applied with limited number of images.
2.4 Summary

This chapter has presented extensive efforts in the state-of-the-art investigations. Three reviews corresponding to each original contribution were conducted. Existing micro-scale shape reconstruction systems were first reviewed, which led to the proposed approach using PS with SNI throughout the dissertation. It was particularly relevant to Chapter 6 in design and calibration of such a system. PS methods dealing with non-Lambertian reflectance were then discussed, which gave rise to the proposed DRM-based color PS method presented in Chapter 4. SNI methods were finally reviewed and the need for perspective SNI was addressed. This review corresponded to the proposed SNI method given in Chapter 5.
Chapter 3

Image Formation: The Forward Problem

Photometric stereo (PS), in essence, is an inverse problem. In order to solve the problem, it is crucial to first understand the physics of the forward problem: image formation. Image formation consists of two major problems, imaging geometry and imaging photometry. Imaging geometry determines the correspondences between the 3D scene in the world to pixel locations on the image, whilst imaging photometry characterizes the amount of photons intercepted by each pixel-corresponding camera sensor cell reflected from the 3D scene. In this chapter, the imaging geometry model is first given, followed by the modelling of imaging photometry. The color image formation model using dichromatic reflectance theory is finally established.
3.1 Imaging Geometry

The pin-hole camera model with perspective projection is applied to characterize the imaging geometry. Figure 3.1 shows the coordinate systems used to model the perspective projection from a 3D surface point, $\tilde{X} = [X, Y, Z]^T$, to a 2D image point, $x = [x, y, 1]^T$. Four coordinate systems are established as shown by the figure, which are the 3D camera coordinate, $\{C\}$, the 2D pixel coordinate, $\{P\}$, the 2D image coordinate, $\{I\}$, and the 3D world coordinate, $\{W\}$. These coordinate setups follow the convention given in [58]. Assume the imaging system is properly focused, which implies that rays reflected from the target surface passing through the entrance aperture are deflected to meet within a single pixel. The distance from the camera center $C$ to the image plane is then the effective focal length, $f$. The perspective
projection from \( \{C\} \tilde{X} \) to \( \{I\} \mathbf{x} \) is, therefore, given by:

\[
\{I\} \mathbf{x} = \frac{1}{\{C\} Z} \hat{K}^{\{C\}} \tilde{X} = \frac{1}{\{C\} Z} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \{C\} \mathbf{e} X.
\] (3.1)

\( \{I\} \mathbf{x} \) is with the same unit as point in \( \{C\} \). In order to let the point with pixel unit, a transformation of this projected image point on the image plane from \( \{I\} \) to \( \{P\} \) is necessary and given by:

\[
\{P\} \mathbf{x} = \mathbf{M}_p \{I\} \mathbf{x} = \begin{bmatrix} m_x & \gamma & o_x \\ 0 & m_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \{I\} \mathbf{x},
\] (3.2)

where \( m_x \) and \( m_y \) are number of pixels per unit distance in the \( x \) and \( y \) directions, respectively. \( o_x \) and \( o_y \) are the principal point locations and \( \gamma \) is the skew factor. From equation (3.1) and (3.2), the relationship between a 3D surface point in \( \{C\} \) and a 2D pixel in \( \{P\} \) is established as:

\[
\{P\} \mathbf{x} = \frac{1}{\{C\} Z} \mathbf{M}_p \hat{K}^{\{C\}} \tilde{X} = \frac{1}{\{C\} Z} \mathbf{K}^{\{C\}} \tilde{X},
\] (3.3)

where \( \mathbf{K} \) is the camera intrinsics. The projected 2D points in \( \{P\} \) are rasterized into a regular grid in the camera as:

\[
\begin{bmatrix} j \\ i \end{bmatrix} = \text{ceil} \left( \begin{bmatrix} \{P\} x + 0.5 \\ \{P\} y + 0.5 \end{bmatrix} \right),
\] (3.4)

where \( \text{ceil} \) denotes the ceiling operator. \( \{i\}_{i \in [1,H]} \) and \( \{j\}_{j \in [1,W]} \) represent the row and column indices of an image. \( H \) and \( W \) denote number of pixels in \( \{P\} y \) and \( \{P\} x \) directions. Sometimes, surface points in \( \{C\} \) are required to be transformed to \( \{W\} \) whose \( X - Y \) plane
defines the reference plane. \( \{W\} \) and \( \{C\} \) are simply related by a 3D rigid transformation:

\[
\{C\}X = R_W \{W\}X + t_W,
\]

where \( R_W \) is a rotation matrix and \( t_W \) is a translation vector. \( R_W \) and \( t_W \) together are referred to as the camera extrinsics.

### 3.2 Imaging Photometry

#### 3.2.1 Radiometric Concepts

Fundamental radiometric concepts are introduced in the following subsection, which were defined in [59]. These concepts are cornerstones to understand surface reflectance introduced in Section 3.2.3. Table 3.1 lists the terms, symbols and unit dimensions of four basic radiometric concepts. Source radiant flux, \( \Phi \), is the power of the source electromagnetic radiation and is measured in watts (W). The source radiant intensity, \( J \), is the exitant flux per unit solid angle and is measured in watts per steradian (W \( \cdot \) sr\(^{-1}\)). The surface irradiance, \( E_i \), is the incident flux density measured in watts per square meter of the surface (W \( \cdot \) m\(^{-2}\)). The scene radiance, \( L_r \), is the flux emitted from the surface per unit foreshortened area per unit solid angle and measured in (W \( \cdot \) m\(^{-2}\) \( \cdot \) sr\(^{-1}\)).
Figure 3.2 illustrates an example to clarify these four basic radiometric concepts and shows a surface patch illuminated by a point source. \( \hat{n} \) denotes the unit surface normal. \((\cdot)\) represents unit vectors later on. Suppose a point source with radiant flux, \(d\Phi\), illuminating a surface patch with area of \(dA\), the source radiant intensity, \(J\), is computed using:

\[
J = \frac{d\Phi}{d\omega_i}, \tag{3.6}
\]

where \(d\omega_i\) is the solid angle subtended by \(dA\) viewing from the source. \(d\omega_i\) is computed using:

\[
d\omega_i = \frac{dA'}{R_i^2} = \frac{dA \cos \theta_i}{R_i^2}, \tag{3.7}
\]

where \(dA'\) is the foreshortened area, \(R_i\) is the source-surface distance and \(\theta_i\) is the incident angle. The flux intercepted by the surface patch is then:

\[
d\Phi = J \cdot dA \cdot \cos \theta_i/R_i^2. \tag{3.8}
\]

The surface irradiance, \(E_i\), can be, therefore, computed by:

\[
E_i = \frac{d\Phi}{dA} = J \cdot \cos \theta_i/R_i^2. \tag{3.9}
\]
The scene radiance, $L_r$, reflected from the surface patch is obtained using:

$$L_r = \frac{d\Phi_r}{(dA \cos \theta_r d\omega_r)},$$  \hspace{1cm} (3.10)

where $\theta_r$ is the exitant angle and $d\omega_r$ is the exitant solid angle. $\Phi_r$ is the reflected radiant flux.

### 3.2.2 Light Attenuation

Two types of light attenuation are modelled, which are light attenuation due to distance and radial attenuation. The models coincide with those given in [56]. Figure 3.3 shows the quantities necessary to establish these models.

Suppose a point source located at ${^C}\tilde{S}$ illuminates the surface at a point ${^C}\tilde{X}$. The
unit light direction, \( \hat{I} \), and the light-surface distance, \( R_i \), can be then computed as:

\[
\begin{align*}
R_i &= \|^{(C)}\mathbf{S} - (C)\mathbf{X} \| \\
\hat{I} &= \left(\frac{(C)\mathbf{S} - (C)\mathbf{X}}{R_i}\right)/R_i.
\end{align*}
\]  

(3.11)

According to equation (3.9), the surface irradiance, \( E_i \), is inverse proportional to \( R_i^2 \). This is the first type of light attenuation due to the light-surface distance.

Many existing light sources are directional, which means they are the brightest along a principle direction and become dimmer as the light direction deviates from the principle direction. Denote the unit principle light direction as, \( \hat{g} \). The angle between \( \hat{I} \) and \( \hat{g} \) is denoted as \( \theta_t \). The light radial attenuation behavior can be, therefore, modelled as \( \cos^\chi(\theta_t) \), where \( \chi \) is the attenuation coefficient reminiscent of \( \beta \) or \( \alpha \) in the specular reflectance models.

Considering both the light attenuation behaviors, the surface irradiance at \( ^{(C)}\mathbf{X} \) illuminated by the point source \( ^{(C)}\mathbf{S} \) can be modelled as:

\[
E_i = J \cdot \cos \theta_i \cdot \cos^\chi(\theta_t)/R_i^2.
\]  

(3.12)

### 3.2.3 Surface Reflectance

Surface reflection is a process by which electromagnetic flux incident on a stationary surface leaves the surface without a change in frequency [59]. Modelling surface reflectance is a challenging task because it exhibits significant difference among different materials. To make the inverse problem of PS tractable, it is necessary to simplify the reflectance models. In this dissertation, the following assumptions are made to limit the scope of surface reflectance modelling:

- Phenomena, such as transmission, absorption, spectral effects, polarization and fluo-
rescence, is negligible;

- The area for each pixel-corresponding surface patch is large compared with the wavelength of incident light and therefore, geometrical-optics is valid. Phenomena related to wave optics, e.g. interference and diffraction, is negligible;

- Light is immediately reflected when reaching the surface;

- Surface patches are homogeneous;

Having these assumptions, the surface reflectance can be characterized using the Bidirectional Reflectance Distribution Function (BRDF), denoted as \( f_r \). BRDF is a function of four variables and defined by the ratio of surface irradiance, \( dE_i(\theta_i, \phi_i) \), to the scene radiance, \( dL_r(\theta_r, \phi_r) \), as:

\[
f_r(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_r, \phi_r)}{dE_i(\theta_i, \phi_i)}, \tag{3.13}
\]

where two variables of zenith angle \( \theta_i(\theta_r) \) and azimuth angle \( \phi_i(\phi_r) \) are used to specify a particular direction. Figure 3.4 presents the normal-centered Cartesian coordinate system, \( \{N\} \), for clarifying the BRDF definition. The unit surface normal, \( \mathbf{n} \), aligns with \( \{N\}Z \) which is perpendicular to the plane spanned by \( \{N\}X - \{N\}Y \). \( (\theta_i, \phi_i) \) specifies the unit light direction (incoming direction), \( \mathbf{I} \), whilst \( (\theta_r, \phi_r) \) determines the unit viewer direction (outgoing direction), \( \mathbf{v} \). The unit of BRDF is \( sr^{-1} \).

BRDFs have some general properties, including isotropy, reciprocity, energy conservation and non-negativity. BRDF with isotropy is an important subset of BRDFs and valid for materials for which rotations about the surface normal can be ignored. Instead of the general BRDF with 4D, BRDF with isotropy is a 3D function denoted as \( f_r(\theta_i, \theta_r, |\phi_i - \phi_r|) \). Reciprocity states that light paths are reversible, which means by swapping the incident and exitant directions, \( f_r \) yields the same value. The energy conservation property depicts that
the energy received by the surface patch must be greater than the sum of energy emitted by
the surface. The non-negativity restricts $f_r$ to be non-negative since $E_i$ and $L_r$ must also be
non-negative values.

Analytical BRDFs

Analytical BRDFs have been extensively studied in both fields of computer graphics and
computer vision for decades and a variety of models were proposed. Existing analytical
BRDFs can be classified into two categories: empirical models and physically-based models.
Relevant BRDF models are briefly reviewed here. Comprehensive review of BRDFs can be
found in [60]

Empirical models attempt to approximate surface reflectance appearance without need
to follow the basic laws of physics. The first empirical model is the Lambertian, or ideal
diffuse, model which reflected the same radiance in all directions. The Lambertian BRDF is
represented by:

$$f_r = \frac{1}{\pi}.$$  (3.14)
The Lambertian BRDF is widely adopted because of its computational simplicity, while real materials usually deviate from the Lambertian model when \( \theta_i \) or \( \theta_r \) is greater than 60° [61]. Another popular empirical model is the Blinn-Phong BRDF [62] which is a linear combination of the diffuse and specular components and given by:

\[
   f_r = k_d + k_s \frac{(\cos \theta_h)^\beta}{\cos \theta_i},
\]

where \( k_d \) and \( k_s \) are respectively the diffuse and specular reflectance factors. \( \beta \) is the shininess coefficient. \( \theta_h \) is the halfway angle which is the angle between \( \mathbf{l} \) and the unit half vector, \( \mathbf{h} \), where \( \mathbf{h} \) is computed using:

\[
   \mathbf{h} = (\mathbf{v} + \mathbf{l}) / \|\mathbf{v} + \mathbf{l}\|,
\]

where \( \mathbf{v}^{i,j} \) is the unit viewer direction and determined by:

\[
   \begin{align*}
   \mathbf{v}^{i,j} &= \begin{bmatrix} j - o_x & i - o_y \\ m_x & m_y & f \end{bmatrix}^T \\
   \mathbf{v}^{i,j} &= \frac{\mathbf{v}^{i,j}}{\|\mathbf{v}^{i,j}\|}.
   \end{align*}
\]

The Blinn-Phong model provides more degrees of realism than the Lambertian model by additionally introducing the specular component characterizing the specular lobe in a simple form, while it is unphysical due to its violation of the energy conservation property. Though other empirical BRDF models, such as the Ward BRDF [26] and Ashikhmin-Shirley BRDF [63], were also proven useful with their respective advantages, these models are not reviewed here due to their irrelevance to the later proposed PS method in Chapter 4.

Physically-based BRDFs have been developing for decades in the computer graphics community and are mostly built upon the microfacet theory [23]. One of the most widely
accepted model is the Cook-Torrance BRDF [64] which is in the form of:

\[
f_r = \rho_d + \rho_s \frac{D \cdot F \cdot G}{\cos \theta_i \cos \theta_r},
\]  

(3.18)

where \( \rho_d \) and \( \rho_s \) are scaling factors of the diffuse and specular components, respectively. \( D \) is the Beckmann distribution [65] characterized by:

\[
D = \frac{1}{\alpha^2 \cos^4 \theta_h} \exp \left( -\frac{\tan^2 \theta_h}{\alpha^2} \right),
\]  

(3.19)

where \( \alpha \) encodes the surface roughness. When the refractive index, \( n_r \), and \( \alpha \) are small, \( \alpha \) can be approximated by the shininess coefficient in the Blinn-Phong model as:

\[
\alpha = \sqrt{\frac{2}{\beta + 2}}.
\]  

(3.20)

The Beckmann distribution is derived by assuming the variation in height of the surface is Gaussian, and then the slope distribution is in accordance with the Beckmann distribution. \( F \) is the Fresnel term that is often characterized by Schlick’s approximation and represented by [66]:

\[
F = F(0) + (1 - F(0)) (1 - \cos \theta_i)^5,
\]  

(3.21)

where \( F(0) \) is \( F \) when \( \theta_i = 0 \). \( G \) is the geometrical attenuation term and derived based on the V-groove analyses of shadowing and masking. \( G \) is represented by [67]:

\[
G = \min \left( 1, \frac{2 \cos \theta_h \cos \theta_r}{\cos \theta_{rh}}, \frac{2 \cos \theta_h \cos \theta_i}{\cos \theta_{rh}} \right).
\]  

(3.22)

where \( \theta_{rh} \) is the angle between \( \vec{v} \) and \( \vec{h} \). \( G \) is near-constant and varies smoothly over a large range of exitant angles [67]. Other important physically-based BRDFs, such as Oren-Nayar BRDF [68], will not be further reviewed here.
A Data-driven BRDF

A data-driven reflectance model for isotropic BRDF was proposed by Matusik [69]. He measured 100 isotropic BRDFs as shown by Figure 3.5 and all the data is publicly available in the MERL database. The 100 different materials are commonly used, such as rubber, wood, plastic, metal, phenolic, acrylic and etc. These real material BRDFs are used to render synthesized input images for evaluation of the later proposed PS method. Unlike the definition of BRDF in the normal-centered coordinate system, \( \{N\} \), BRDFs in the MERL database, referred to as MERL BRDFs, are stored in a different coordinate system introduced by Rusinkiewicz [70]. The unit half vector, \( \mathbf{h} \), aligns with the Z axis instead of the normal vector in \( \{N\} \) and therefore, the new coordinate system is called the H-centered coordinate.

![100 materials in the MERL database: pictures of the spheres](image)

Figure 3.5: 100 materials in the MERL database: pictures of the spheres [69]
system denoted as \( \{H\} \). The relationship between \( \{N\} \) and \( \{H\} \) is shown by Figure 3.6. By enforcing the reciprocity constraint:

\[ f_r (\theta_h, \theta_d, \phi_d) = f_r (\theta_h, \theta_d, \phi_d + \pi), \]  

the measured data size can be halved. Another consideration of using \( \{H\} \) instead of \( \{N\} \) is that the specular peaks of highly reflective materials are difficult to be represented in \( \{N\} \) due to the sampling problem.

### 3.2.4 Image Irradiance Formation

Image irradiance is the amount of flux density intercepted from the scene radiance and measured in \( W \cdot m^{-2} \). Assume that the imaging system is properly focused and the chromatic aberration can be ignored. Assume also that no vignetting occurs which implies that the entrance aperture is a constant circle whose diameter is denoted as \( d_l \). The image irradiance
is then modelled as [71]:

\[ e = \frac{\pi}{4} \left( \frac{d_l}{f} \right)^2 \cos^4 \theta_o L_r, \]  

(3.24)

where \( e \) is the image irradiance, \( f/d_l \) represents the f-number of the imaging system, and \( \theta_o \) denotes the off-axis angle. \( \theta_o \) can be computed using \( \theta_o = \arccos(v_z) \) and \( v_z \) is the third tuple of \( \vec{v} \). Equation (3.24) quantifies the relationship between the f-number of the lens and the image irradiance. The off-axis illumination due to \( \cos^4 \theta_o \) also explains the fade-off of brightness as pixel locations deviate more from the optical axis.

A variety of later works attempt to model and correct the vignetting effect. In this dissertation, the Kang-Weiss method [72] is adopted and the vignetting effect can be normalized.

### 3.2.5 Image Intensity Formation

Image intensity refers to the quantity of observation in an image in relation to the image irradiance. The general flow from image irradiance to image intensity for a color digital camera is given by Figure 3.7. The camera shutter speed (exposure time) controls the amount of light reaching the image sensor. The image irradiance is integrated over time
when the shutter is open as:
\[ h_e = \int_0^{t_e} e(\tau) \, d\tau, \]

where \( t_e \) is the exposure time and \( h_e \) is the image radiant exposure in the unit of Joule \( \cdot \) m\(^{-2}\). For a color digital camera, the grid of image sensors is arranged in accordance with a color filter array (e.g., Bayer array). The color filter array has three types of sensor cells (R,G,B) with their respective spectral sensitivities. For a particular type of sensor cell, the image radiant exposure, \( \{h_c\}_{c=R,G,B} \), can be represented as:
\[ h_c = \int_0^{t_e} \int_0^\infty e(\lambda, \tau) \, C_c(\lambda) \, d\lambda d\tau, \]

where \( \lambda \) is the wavelength and \( \{C_c\}_{c=R,G,B} \) represents the camera’s spectral sensitivity for a particular color channel. The image radiant exposure is then boosted by a sensor amplifier, referred to as the film speed (or ISO speed), as:
\[ i_c = \kappa_{iso} h_c, \]

where \( \{i_c\}_{c=R,G,B} \) is the ISO-amplified image radiant exposure and \( \kappa_{iso} \) is the scaling factor of ISO. Note that here, the dynamic range of the image sensor is assumed to be infinite such that sensor saturations are ignored. \( i_c \) is then digitized by an A/D converter to form the raw image readings. Through demosaic of the raw image, color image intensities, \( \{I_c\}_{c=R,G,B} \), are obtained with three-tuples on each pixel. Conventional procedures to achieve displayable images also include white balancing and gamma correction, details of which can be found in [73].

Assume the scene being exposed is static during exposure, the color image intensity is
then directly proportional to the color image irradiance as:

\[ I_c = \kappa_{iso} t_e \int_0^\infty e(\lambda) C_c(\lambda) d\lambda = \kappa_{iso} t e_c, \quad (3.28) \]

where \{e_c\}_{c=R,G,B} is the color image irradiance.

### 3.3 Color Image Formation

Having established the general imaging geometry and photometry model in Section 3.1 and 3.2, the color image formation using the dichromatic reflectance theory [40] is presented here.

Material reflectance may exhibit a combination of diffuse and specular reflection. For dielectric materials, there is a difference in the spectral selectivity for the different components of reflection [61]. The diffuse component may be spectrally selective, while the specular component is not. That is to say the wavelength distribution of the specular component is the same as the incident light. From this observation, Shafer et al. [40] proposed the dichromatic reflectance model (DRM). Following the assumptions listed below, the color image formation model under a point source is established.

- Material reflectance can be represented by a linear combination of diffuse and specular components;
- The wavelength distribution of the specular component is the same as the incident light, i.e. dichromatic reflectance [40];
- The wavelength distribution is not directionally selective such that the BRDF for each wavelength is the same, i.e. neutral interface [74];
• Inter-reflections between surface patches can be ignored such that surface irradiances only originate from the source;

• The surface patch being considered is shadow-free.

Denote \( \{ e_{d,c}^{i,j} \} = R,G,B \) and \( \{ e_{s,c}^{i,j} \} = R,G,B \) as the diffuse and specular components of the color image irradiance at pixel \((i,j)\), respectively. Then, the color image irradiance at pixel \((i,j)\), \( e_c^{i,j} \), is represented by:

\[
e_c^{i,j} = e_{d,c}^{i,j} + e_{s,c}^{i,j} = (\bar{d}_c^{i,j} f_{d}^{i,j} + \bar{s}_c^{i,j} f_{s}^{i,j}) \kappa_{n}^{i,j},
\]

where \( \{ \bar{d}_c^{i,j} \} = R,G,B \) and \( \{ \bar{s}_c^{i,j} \} = R,G,B \) are wavelength dependent factors that are determined by:

\[
\begin{align*}
\bar{d}_c^{i,j} &= \frac{1}{\kappa_{d}^{i,j}} \int_{0}^{\infty} J(\lambda) R^{i,j}(\lambda) C_{c}(\lambda) d\lambda, \\
\bar{s}_c^{i,j} &= \frac{1}{\kappa_{s}} \int_{0}^{\infty} J(\lambda) C_{c}(\lambda) d\lambda,
\end{align*}
\]

where \( J(\lambda) \) is the spectral power density (SPD) of the source and \( R^{i,j}(\lambda) \) is the spectral body reflectance. \( \kappa_{d}^{i,j} \) and \( \kappa_{s} \) are just to normalize \( \bar{d}_c^{i,j} = [\bar{d}_R^{i,j}, \bar{d}_G^{i,j}, \bar{d}_B^{i,j}]^T \) and \( \bar{s}_{RGB} = [\bar{s}_R, \bar{s}_G, \bar{s}_B]^T \) as unit vectors. \( f_{d}^{i,j} \) and \( f_{s}^{i,j} \) in equation (3.29) are geometrical scaling factors of the diffuse and specular components, respectively. They are in relation to the BRDF as:

\[
\begin{align*}
f_{d}^{i,j} &= \kappa_{d}^{i,j} f_{r,d}^{i,j} \cos \theta_{i}^{i,j}, \\
f_{s}^{i,j} &= \kappa_{s} f_{r,s}^{i,j} \cos \theta_{i}^{i,j},
\end{align*}
\]

where \( f_{r,d}^{i,j} \) and \( f_{r,s}^{i,j} \) are the diffuse and specular BRDF components, respectively. Since \( f_{r,d} \) and \( f_{r,s} \) have also their respective scaling factors according to equation (3.15) or (3.18), \( \kappa_{d}^{i,j} \) and \( \kappa_{s} \) can be incorporated inside the BRDF models. \( \kappa_{n}^{i,j} \) in equation (3.29) is a scaling
factor that encodes the light attenuations and off-axis illuminations. $\kappa_{n}^{i,j}$ is represented by:

$$
\kappa_{n}^{i,j} = \left( \cos^{\chi} \theta_{l}^{i,j} / (R_{i}^{i,j})^{2} \right) \left( \frac{\pi}{4} \left( \frac{d_{l}}{f} \right)^{2} \cos^{4} \theta_{o}^{i,j} \right). 
$$

(3.32)

By scaling the image irradiance, $\{e_{c}^{i,j}\}_{c=R,G,B}$ with $1/\kappa_{n}^{i,j}$, the corrected image irradiance is obtained and denoted as, $\{\tilde{e}_{c}^{i,j}\}_{c=R,G,B}$. Let the corrected color image irradiance vector at pixel $(i, j)$ be written as $\tilde{e}_{RGB}^{i,j} = [\tilde{e}_{R}^{i,j}, \tilde{e}_{G}^{i,j}, \tilde{e}_{B}^{i,j}]^{T}$. It can be then represented in the following form as:

$$
\tilde{e}_{RGB}^{i,j} = \tilde{d}_{RGB}^{i,j} f_{d}^{i,j} + \tilde{s}_{RGB} f_{s}^{i,j}.
$$

(3.33)

When the surface reflectance is characterized by the Blinn-Phong model as equation (3.15), $\{\tilde{e}_{c}^{i,j}\}_{c=R,G,B}$ is then represented as:

$$
\tilde{e}_{RGB}^{i,j} = \tilde{k}_{d}^{i,j} \tilde{d}_{RGB}^{i,j} \cos \theta_{i}^{i,j} + \tilde{k}_{s}^{i,j} \tilde{s}_{RGB} \left( \cos \theta_{h}^{i,j} \right)^{\beta^{i,j}}.
$$

(3.34)

where $\tilde{k}_{d}^{i,j} = k_{d}^{i,j} \kappa_{d}^{i,j}$ and $\tilde{k}_{s}^{i,j} = k_{s}^{i,j} \kappa_{s}$. When the surface reflectance is represented by the Cook-Torrance model as equation (3.18), $\{\tilde{e}_{c}^{i,j}\}_{c=R,G,B}$ is then written as:

$$
\tilde{e}_{RGB}^{i,j} = \tilde{\rho}_{d}^{i,j} \tilde{d}_{RGB}^{i,j} \cos \theta_{i}^{i,j} + \tilde{\rho}_{s}^{i,j} \tilde{s}_{RGB} \frac{D^{i,j} \cdot F^{i,j} \cdot G^{i,j}}{\cos \theta_{r}^{i,j}}.
$$

(3.35)

where $\tilde{\rho}_{d}^{i,j} = \rho_{d}^{i,j} \kappa_{d}^{i,j}$ and $\tilde{\rho}_{s} = \rho_{s} \kappa_{s}$. 
3.4 Summary

This chapter has presented how images are formed as the forward problem to PS. Two problems from image formation were separately discussed. The first problem of imaging geometry was briefly reviewed, whilst the second problem of imaging photometry was carefully modelled. The formation of color images using DRM was finally presented.
Chapter 4

DRM-based Color Photometric Stereo

This chapter presents a photometric stereo (PS) method based on the dichromatic reflectance model (DRM) using color images. The proposed method estimates surface orientations for surfaces with non-Lambertian reflectance using diffuse-specular separation and contains two steps. The first step, referred to as diffuse-specular separation, initializes surface orientations in a specular invariant color subspace and further separates the diffuse and specular components in the RGB space. In the second step, the surface orientations are refined by first initializing specular parameters via solving a log-linear regression problem owing to the separation and then fitting the DRM using constrained trust region algorithm. Since reliable information from diffuse reflection free from specularities is adopted in the initializations, the proposed method is robust and feasible with less observations. At pixels where dense non-Lambertian reflectances appear, signals from specularities are exploited to refine the surface orientations and the additionally acquired specular parameters are potentially valuable for more applications, such as digital relighting and material classification.

This chapter is organized as follows. The next section presents the problem formulation of color PS and related works. Section II first overviews the flow of the proposed PS
method and then, elucidates the specific two steps in subsequent subsections. It follows the comprehensive evaluations on surface orientation estimation using both synthetic and real images. The implications of the proposed method are discussed in the next section, while conclusions and future works are finally given.

4.1 Problem Formulation and Related Works

In this section, the generic color PS problem is first formulated, followed by the elucidation of color image formation model to be used. A brief review of two closely related past works are finally given to clarify the original contribution given in Section 4.2.

4.1.1 Generic Problem of Color PS

Figure 4.1 shows the schematic diagram of the hardware components of the sensor and necessary data for generic color PS. The hardware components are an RGB camera and a set of directional point sources whose positions are priorly known. Assume that ambient light is completely blocked, the color PS requires that the RGB camera captures a color image for the target surface with one source lit at a time. Given the \( N \) color images illuminated under \( N \) different sources, the objective of color PS is to identify surface orientations at each pixel. More specifically, given the corrected color image irradiance measurements at pixel \((i, j)\), \( \{\tilde{e}_{k}^{i,j}\}_{k \in [1,N]} \), under the unit illumination directions, \( \{\tilde{l}_{k}^{i,j}\}_{k \in [1,N]} \), estimated from the light positions, the color PS aims to derive the unit surface normal, \( \tilde{n}^{i,j} \), by minimizing the sum of \( N \) corresponding squared residuals:

\[
\epsilon^{i,j} = \sum_{k=1}^{N} \left[ (\tilde{e}_{k}^{i,j} - f(\tilde{n}^{i,j}, \tilde{l}_{k}^{i,j}))^{T} w - f(\tilde{n}^{i,j}, \tilde{l}_{k}^{i,j}) \right]^{2} \rightarrow \min_{\tilde{n}^{i,j}}.
\] 

(4.1)
where \( \mathbf{w} \) is a vector of weights for the color channels whereas \( f \) is a function of \( \mathbf{n}_{i,j} \) having \( \mathbf{l}_{k} \) as parameters. \((\cdot)^*\) means the measurement later on.

\[ e_{RGB}^{i,j} = f_{d,k} d_{RGB}^{i,j} + f_{s,k} s_{RGB}^{i,j} = k_{d} d_{RGB}^{i,j} \left( (\mathbf{n}_{i,j})^T \mathbf{l}_{k} \right) + k_{s} s_{RGB}^{i,j} \left( (\mathbf{n}_{i,j})^T \mathbf{h}_{k} \right)^{\beta_{i,j}}, \quad (4.2) \]

**4.1.2 Color Image Formation Model**

In Chapter 3, the color image formation using DRM was established. From the Blinn-Phong model as shown in equation (3.34), the corrected color image irradiance on pixel \((i, j)\) under the \( k^{th} \) illumination is written into a function of \( \mathbf{n}_{i,j} \) and \( \mathbf{l}_{k} \) as:

[Figure 4.1: Schematic diagram for generic color PS]
where \( \{ \hat{h}_{k}^{i,j} \}_{k \in [1,N]} \) is the \( k^{th} \) unit half vector that is computed according to equation (3.16). More specifically, it is written as:

\[
\hat{h}_{k}^{i,j} = \left( \bar{v}^{i,j} + \bar{r}_{k}^{i,j} \right) / \| \bar{v}^{i,j} + \bar{r}_{k}^{i,j} \|, \quad (4.3)
\]

where \( \bar{v}^{i,j} \) is the unit viewer direction and determined using equation (3.3) by assuming \( \gamma = 0 \) as:

\[
\begin{cases}
\bar{v}^{i,j} &= - \left[ \frac{j - o_{x}}{m_{x}}, \frac{i - o_{y}}{m_{y}}, f \right]^{T} \\
\bar{v}^{i,j} &= \bar{v}^{i,j} / \| \bar{v}^{i,j} \|. 
\end{cases} \quad (4.4)
\]

**Light configurations**

Figure 4.2: Modelling of light configurations in a circular pattern

Figure 4.2 shows the diagram for modelling of light configurations in a circular pattern.
Let the $k^{th}$ light position be $(C)\mathbf{S}_k = [(C)X_k, (C)Y_k, (C)Z_k]^T$. $(C)\mathbf{S}_k$ is represented in the spherical space as:

\[
\begin{align*}
(C)X_k &= r_k \sin \theta_k \cos \phi_k \\
(C)Y_k &= r_k \sin \theta_k \sin \phi_k \\
(C)Z_k &= r_k \cos \theta_k,
\end{align*}
\]

where $r_k$, $\theta_k$ and $\phi_k$ are the radius, zenith and azimuth angle, respectively. The $N$ lights are configured such that they have the same zenith angle and radius, and uniformly distributed azimuth angles around the optical axis, i.e. $(C)\theta_k = \theta$, $(C)r_k = r$ and $(C)\phi_k = \frac{360^\circ (k-1)}{N} + \phi_0$, where $\phi_0$ is the azimuth angle of the first light position in $(C)$.

Assume the optical axis is oriented perpendicularly to the reference plane, denoted as $\pi_r$. The camera reference plane, $\pi_c$, is, therefore, parallel to $\pi_r$, where $\pi_c$ is spanned by $(C)X - (C)Y$. The distance between $\pi_c$ and $\pi_r$ is denoted as $d_b$. The zenith angle of the $k^{th}$ light direction relative to the center of the field of view (FoV) on $\pi_r$ is denoted as $\eta_k$. $\eta_k$ is the same for each light ideally and denoted as $\eta$. $\eta$ is computed by:

\[
\eta = \arcsin \left( \frac{r}{\sqrt{r^2 + d_b^2 - 2rd_b \cos \theta / \sin \theta}} \right). 
\]

(4.6)

This circular pattern of light configurations will be used across several chapters. It is introduced here to clarify experimental parameters in Section 4.3.1 later.

### 4.1.3 Existing Methods Using DRM

Existing color PS methods using DRM contain two major variants, representative works of which were developed by Barsky et al. [39] and by Zickler et al. [41].
The PS method of Barsky derives the unit surface normal $\mathbf{n}^{i,j}$ such that the sum of squared residuals between the corrected color image irradiance projected along $\mathbf{d}^{i,j}_{RGB}$ and the diffuse geometrical scaling factor is minimized:

$$
\epsilon^{i,j} = \sum_{k=1}^{N_b} \left[ (\mathbf{e}^{i,j,*}_k)^T \mathbf{d}^{i,j}_{RGB} - f^{i,j}_{d,k} \right]^2 = \sum_{k=1}^{N_b} \left[ (\mathbf{e}^{i,j,*}_k)^T \mathbf{d}^{i,j}_{RGB} - \bar{k}^{i,j}_d (\mathbf{n}^{i,j})^T \mathbf{l}^{i,j}_k \right]^2, \quad (4.7)
$$

where $N_b (\leq N)$ is the number of specular-free observations. $N_b$ is used instead of $N$ because the Barsky’s PS method assumes that non-Lambertian reflectance appears sparsely among observations and they are rejected as outliers. As the first process of outlier rejection, the direct principal component analysis (DPCA) \[75\] on $\left[ \mathbf{E}^{i,j,*}_{RGB} \right]$ is performed to estimate $\mathbf{d}^{i,j}_{RGB}$, where $\left[ \mathbf{E}^{i,j,*}_{RGB} \right] = [\mathbf{e}^{i,j,*}_1, \mathbf{e}^{i,j,*}_2, ..., \mathbf{e}^{i,j,*}_N]^T$. Given the prior knowledge of $\mathbf{s}_{RGB}$ and the estimated $\bar{\mathbf{d}}^{i,j}_{RGB}$, the method then determines $f^{i,j}_{s,k}$ using:

$$
f^{i,j}_{s,k} = \frac{(\mathbf{e}^{i,j,*}_k)^T \mathbf{s}_{RGB} - (\mathbf{e}^{i,j,*}_k)^T \mathbf{d}^{i,j}_{RGB} \left( (\mathbf{d}^{i,j}_{RGB})^T \mathbf{s}_{RGB} \right)}{1 - \left( (\bar{\mathbf{d}}^{i,j}_{RGB})^T \mathbf{s}_{RGB} \right)^2}. \quad (4.8)
$$

A threshold on $f^{i,j}_{s,k}$ is used to determine specularities which are further rejected as outliers.

The objective function (4.7) of this PS method is not different from that of the conventional Lambertian-based PS method [19]. Given $\{\mathbf{I}^{i,j}_k\}_{k \in [1,N_b]}$ as a priori, $\mathbf{n}^{i,j}$ and $\bar{k}^{i,j}_d$ are determined analytically as equation (4.9) by minimizing the objective function (4.7).

$$
\begin{align*}
\bar{k}^{i,j}_d &= \left\| \left[ \mathbf{L}^{i,j} \right]^T \left[ \mathbf{L}^{i,j} \right]^{-1} \left[ \mathbf{L}^{i,j} \right]^T \left[ \mathbf{E}^{i,j,*}_{RGB} \right] \mathbf{d}^{i,j}_{RGB} \right\| \\
\mathbf{n}^{i,j} &= \frac{\left( \left[ \mathbf{L}^{i,j} \right]^T \left[ \mathbf{L}^{i,j} \right]^{-1} \left[ \mathbf{L}^{i,j} \right]^T \left[ \mathbf{E}^{i,j,*}_{RGB} \right] \mathbf{d}^{i,j}_{RGB} \right)}{\bar{k}^{i,j}_d}, \quad (4.9)
\end{align*}
$$

where $\left[ \mathbf{L}^{i,j} \right] = [\mathbf{I}^{i,j}_1, \mathbf{I}^{i,j}_2, ..., \mathbf{I}^{i,j}_N]^T$ and rank($[\mathbf{L}^{i,j}]$) = 3. While having exhibited the efficacy of specularity detection through the use of $\mathbf{s}_{RGB}$, the estimation of $\bar{\mathbf{d}}^{i,j}_{RGB}$ using DPCA is...
error-prone due to the existence of specularities. Furthermore, the specularity is rejected as outlier though useful information may be contained.

Unlike Barsky’s PS method, which performs PS in the RGB color space, the PS method of Zickler estimates surface orientations in the SUV color space. The objective function to be minimized is given by:

\[
\epsilon_{i,j} = \sum_{k=1}^{N} \left[ (\mathbf{e}_{i,j}^{*})^T \mathbf{d}_{i,j}^{*} - (\kappa_{z}^{i,j})^2 f_{d,k} \right]^2 = \sum_{k=1}^{N} \left[ (\mathbf{e}_{i,j}^{*})^T \mathbf{d}_{i,j}^{*} - (\kappa_{z}^{i,j})^2 \mathbf{k}_{d}^{i,j} (\mathbf{n}_{i,j})^T \mathbf{l}_{k}^{i,j} \right]^2,
\]

(4.10)

where \(\kappa_{z}^{i,j}\) is a scaling factor. The SUV color space is source dependent and determined based on the known \(\mathbf{s}_{RGB}\). To operate in the SUV space, the corrected color image irradiance is transformed from RGB to SUV as:

\[
\mathbf{e}_{SUV,k}^{i,j} = \mathbf{R}_{s} \mathbf{e}_{k}^{i,j} = \mathbf{d}_{SUV}^{i,j} f_{s,k}^{i,j} + \mathbf{s}_{SUV} f_{s,k}^{i,j},
\]

(4.11)

where \(\mathbf{d}_{SUV}^{i,j} = [\mathbf{d}_{U}^{i,j}, \mathbf{d}_{V}^{i,j}, \mathbf{d}_{S}^{i,j}]^T\), \(\mathbf{s}_{SUV} = [0, 0, 1]^T\), and \(\mathbf{R}_{s} \in SO(3)\) is any transformation that yields \(\mathbf{s}_{SUV} = \mathbf{R}_{s} \mathbf{s}_{RGB} = [0, 0, 1]^T\). \(UV\) forms a color subspace that is invariant to \(f_{s,k}^{i,j}\). The Zickler’s PS first estimates \(\mathbf{d}_{UV}^{i,j}\) of unit length using DPCA on \([\mathbf{E}_{UV}^{i,j}]^T [\mathbf{E}_{UV}^{i,j}]\). It then becomes the conventional Lambertian-based PS problem and the solution is obtained in the similar form as equation (4.9). Compared with the Barsky’s PS, the Zickler’s PS utilizes more observations (\(N\) instead of \(N_{b}\)) and performs PS in the specular-free color subspace, \(UV\). Shape information from specularities is, however, neglected again and the method can only provide diffuse component up to the normalized RGB space. The diffuse-specular separation in the RGB space remains open.
4.2 Color PS Using Diffuse-Specular Separation

While the primary objective is to derive unit surface normals from the observed RGB images, our color PS method based on DRM also aims at the separation of color diffuse and specular components for purposes such as specular removal [76] and intrinsic image decomposition [77]. In addition, parameters characterizing $f^{i,j}_{s,k}$ are desired where dense non-Lambertian reflectance appears for potential applications, such as digital relighting and material classification.

4.2.1 Objective Function

The generalised objective is: given $\{\tilde{e}^{i,j}_k\}_{k=1}^N$ illuminated by $N$ known $\{\tilde{l}^{i,j}_k\}_{k=1}^N$ and $\tilde{s}_{RGB}$, solve for $\tilde{n}^{i,j}$, $\tilde{e}^{i,j}_{d,k}$, $\tilde{e}^{i,j}_{s,k}$, $\tilde{d}^{i,j}_{d,RGB}$, $\tilde{k}^{i,j}_{d}$, $\tilde{k}^{i,j}_{s}$ and $\beta^{i,j}$, where $\tilde{e}^{i,j}_{d,k} = f^{i,j}_{d,k} \tilde{d}^{i,j}_{d,RGB}$ and $\tilde{e}^{i,j}_{s,k} = f^{i,j}_{s,k} \tilde{s}_{RGB}$.

The cost function for the generalised DRM-based color PS problem is given by:

$$
\epsilon^{i,j} = \sum_{k=1}^{N} \left( (\tilde{e}^{i,j}_k)_{RGB} - \tilde{k}^{i,j}_{d}(\tilde{n}^{i,j})_{RGB}^T (\tilde{d}^{i,j}_{d,RGB})^T \tilde{s}_{RGB} - \tilde{k}^{i,j}_{s}(\tilde{n}^{i,j})_{RGB}^T (\tilde{h}^{i,j}_k)^{\beta^{i,j}} \right)^2, \quad (4.12)
$$

where $\epsilon^{i,j}$ is the sum of squared residuals on pixel $(i,j)$. Here, the Blinn-Phong model is used to characterize the surface reflectance because of its computational efficiency and good performance in characterizing a wide range of isotropic material reflectances. Other reflectance model, such as the Cook-Torrance model as equation (3.35), can be also adopted in a similar manner as the cost function given by (4.12). Note that the $N$ image irradiances are assumed shadow-free and if shadows appear, the corresponding terms should be discarded from the cost function.
4.2.2 Overview of the Proposed Color PS method

Figure 4.3 shows the flow of the proposed DRM-based color PS method using diffuse-specular separation. The proposed method consists of a main process called the diffuse-specular separation and a subsidiary process of surface normal refinement. The main process identifies good initial estimates of surface orientations and the subsidiary process further refines them to maximize accuracy where dense non-Lambertian reflectances appear. These processes are marked in the red boxes.

The main process is further composed of three sub-processes: diffuse color estimation, diffuse-specular separation and PS in the UV color subspace. After completing shadow rejection as a preprocess and generating the shadow-free image irradiance matrix $[E_{RGB}]$, the main process begins with the diffuse color estimation and derives $\bar{d}_{RGB}^{ij}$ using robust principal component analyses (RPCA) and outliers from this process are regarded as the specularity map. The diffuse-specular separability is then determined for each pixel by
checking the diffuse-specular chromatic angle $\psi^{i,j} = \arccos \left( (\tilde{d}_{RGB}^{i,j})^T \bar{s}_{RGB} \right)$. The diffuse-specular separation by this check is the particular concept and process proposed in this paper. For separable pixels, the proposed method performs PS in the UV color subspace and derives the initial guess of unit surface normal $(\tilde{n}^{i,j})_0$ and the diffuse reflectance factor $(\tilde{\kappa}^{i,j}_d)_0$ for later refinement purpose. Simultaneously, the specular geometrical scaling factor, $(\tilde{f}^{i,j}_s)_0$, can be initialized using the estimated $\tilde{d}_{RGB}^{i,j}$, the known $\bar{s}_{RGB}$ and the specularity map.

In the subsidiary process of surface normal refinement, the initial guess of the specular parameters, $(\tilde{\kappa}^{i,j}_s)_0$ and $(\tilde{\beta}^{i,j})_0$, are derived from $(\tilde{n}^{i,j})_0$ and $(\tilde{f}^{i,j}_s)_0$ by solving a log-linear regression problem. Finally, the surface normal refinement identifies $\tilde{n}^{i,j}$ and $k^{i,j}_d$ with enhanced accuracy while additionally updating $k^{i,j}_s$ and $\beta^{i,j}$.

For pixels where sparse non-Lambertian reflectances appear, the first process is sufficient to reliably estimate the surface orientations. To tackle with the dense non-Lambertian reflectance problem, the second process is additionally required to exploit surface normal information from specularities in the DRM with the designed parameter initialization strategy. The rest of this section elucidates the proposed method in detail.
4.2.3 Diffuse-Specular Separation

Based on equation (4.2), \( \hat{E}_{i;j}^{RGB} \) is decomposed as:

\[
\hat{E}_{i;j}^{RGB} = f_{d;1}^{i;j} \bar{d}_{RGB}^i + f_{s;1}^{i;j} \bar{s}_{RGB} + \ldots + f_{d;M}^{i;j} \bar{d}_{RGB}^i + f_{s;M}^{i;j} \bar{s}_{RGB},
\]

where \( M \) is the number of shadow-free observations. Assume most entries in \( f_{s;1}^{i;j} \) are near-zero. The principal component of \( \hat{E}_{i;j}^{RGB} \) is thus the best estimate of \( \bar{d}_{RGB}^i \) if non-zero entries in \( f_{s;1}^{i;j} \) can be reliably rejected. The proposed RPCA algorithm is designed for this purpose and given as follows:

1. PCA for \( \hat{E}_{i;j}^{RGB} \) to estimate \( \bar{d}_{RGB}^i \);

2. Compute residual matrix: \( \mathbf{R}_{d}^{i;j} = \hat{E}_{i;j}^{RGB} \bar{d}_{RGB}^i - \hat{E}_{i;j}^{RGB} \)

3. Compute residual vector: \( r_{d}^{i;j} = \sqrt{(r_{d;1}^{i;j})^2 + (r_{d;2}^{i;j})^2 + (r_{d;3}^{i;j})^2} \), where \( \mathbf{R}_{d}^{i;j} = [r_{d;1}^{i;j}, r_{d;2}^{i;j}, r_{d;3}^{i;j}] \) and \( \circ2 \) denotes the Hadamard square.

4. if mean\( (r_{d}^{i;j}) \) is smaller than \( T_d \), terminate RPCA and output the current estimate of \( \bar{d}_{RGB}^i \) and the specularity map. Otherwise, find the index with max \( \left( \frac{r_{d}^{i;j} - \text{mean}(r_{d}^{i;j})}{\text{std}(r_{d}^{i;j})} \right) \) and label it in the specularity map. Remove the corresponding row in \( \hat{E}_{i;j}^{RGB} \) and repeat from step 1.

The RPCA algorithm estimates \( \bar{d}_{RGB}^i \) more accurately than the DPCA by rejecting specularities as outliers through residual analyses. Distinct from Barsky-PS that identifies specularities after \( \bar{d}_{RGB}^i \) estimation, the proposed algorithm is capable to derive \( \bar{d}_{RGB}^i \) and the
specularity map simultaneously.

The RPCA performance depends on $\mathbf{d}_{\text{RGB}}^{i,j}$ of the surface patch. Through projecting $\mathbf{E}_{\text{RGB}}^{i,j}$ along $\mathbf{s}_{\text{RGB}}$, equation (4.13) is rewritten as:

$$
\left[ \mathbf{E}_{\text{RGB}}^{i,j} \right] \mathbf{s}_{\text{RGB}} = f_d^{i,j} \cos(\psi^{i,j}) + f_s^{i,j}.
$$

(4.14)

The inlier in RPCA is $f_d^{i,j} \cos(\psi^{i,j})$, while $f_s^{i,j}$ is the outlier. Therefore, when $\psi^{i,j}$ is small, RPCA is more noise-robust due to the stronger inlier, while it is more challenging to identify specularities due to relatively weaker $f_s^{i,j}$. The threshold $T_d$ is used to balance this performance. If $T_d$ is too big, the RPCA is tolerant to specularities and performs similar to DPCA. By contrast, if $T_d$ is too small, the RPCA rejects innocent pixels as specularities and makes it sensitive to image noise.

Having estimated $\mathbf{d}_{\text{RGB}}^{i,j}$, it is then necessary to check whether the diffuse-specular separation is feasible using color. In fact, $\left[ \mathbf{E}_{\text{RGB}}^{i,j} \right]^T$ forms a 2D vector space spanned by $\mathbf{d}_{\text{RGB}}^{i,j}$ and $\mathbf{s}_{\text{RGB}}$. If $\mathbf{d}_{\text{RGB}}^{i,j} = \mathbf{s}_{\text{RGB}}$, the vector space reduces to 1D and no directional cue exists. In the SUV color space, the UV subspace contains no information and the decoupling fails. If $\psi^{i,j} \geq T_c$, $\mathbf{d}_{\text{RGB}}^{i,j}$ and $\mathbf{s}_{\text{RGB}}$ are sufficiently distinct and pixel $(i,j)$ is considered as separable. The choice of $T_c$ is affected by the performance of $\mathbf{d}_{\text{RGB}}^{i,j}$ estimation. Qualitatively speaking, when $T_c$ is chosen small, more non-separable pixels are falsely classified as separable, resulting in the corruption of later steps. If $T_c$ is too big, more separable pixels are misclassified, leading to the feasibility of surface normal estimations reduced to less pixel locations.

$(f_s^{i,j})_0$ is then initialized using equation (4.8) with the estimated $\mathbf{d}_{\text{RGB}}^{i,j}$ and the known $\mathbf{s}_{\text{RGB}}$ for later usage in the subsidiary process. $(f_s^{i,j})_0$ entries are clamped to zero if they are not in the specularity map. Unlike Barsky’s PS algorithm by having a threshold on the computed $(f_s^{i,j})$ to determine the specularity map, the proposed method derives the
specularity map in the diffuse color estimation process, which is intended to reduce the number of parameters to be used.

Image irradiances of separable pixels are transformed to the SUV color space as $\tilde{E}_{SUV}^{ij}$ using $R_s$. $\tilde{E}_{UV}^{ij}$ is then formed by picking the first two columns of $\tilde{E}_{SUV}^{ij}$. With the same $R_s$, the estimated $d_{RGB}^{ij}$ is also transformed to the SUV space as $\tilde{d}_{SUV}^{ij} = [\tilde{d}_u^{ij}, \tilde{d}_v^{ij}, \tilde{d}_s^{ij}]^T$, where $\tilde{d}_{SUV}^{ij}$ is a unit vector. Let $\tilde{d}_{UV}^{ij} = [\tilde{d}_u^{ij}, \tilde{d}_v^{ij}]^T$ as a unit vector which is defined as: $\tilde{d}_{UV}^{ij} = \kappa_z d_{UV}^{ij}$, where $\kappa_z^{ij} = \sqrt{(\tilde{d}_u^{ij})^2 + (\tilde{d}_v^{ij})^2}$. Since $\tilde{d}_{RGB}^{ij} \neq \tilde{s}_{RGB}$ for separable pixels, $\kappa_z^{ij} > 0$. Through projecting $\tilde{E}_{UV}^{ij}$ along $\tilde{d}_{UV}^{ij}$, the cost function to be minimized is the same as equation (4.10). The solution for this optimization problem is referred to as PS in UV algorithm. Due to the presence of image noise, the PS in UV algorithm is modified to reject noise-corrupted image irradiances as outliers using studentized residuals. The modified algorithm is given as follows:

Let the residual vector $r^{ij}$ be written as:

$$r^{ij} = \left[ \tilde{E}_{UV}^{ij} \right] \tilde{d}_{UV}^{ij} - [L^{ij}] n^{ij},$$

(4.15)

where $n^{ij}$ is the scaled normal and $[L^{ij}] = [\tilde{l}_1^{ij}, \tilde{l}_2^{ij}, ..., \tilde{l}_M^{ij}]^T$. Denote $\rho^{ij} as \tilde{k}_d^{ij} \kappa_z^{ij}$, then $n^{ij} = \rho^{ij} \tilde{n}^{ij}$. The hat matrix $[H_a^{ij}]$ is then represented by:

$$[H_a^{ij}] = [L^{ij}] \left( [L^{ij}]^T [L^{ij}] \right)^{-1} [L^{ij}]^T.$$

(4.16)

Diagonal entries in $[H_a^{ij}]$ are leverages and the $k^{th}$ leverage is denoted as $h_{a,k}^{ij}$. These leverages quantify the influence that the observed $\tilde{E}_{UV}^{ij} \tilde{d}_{UV}^{ij}$ on their predicted values $[L^{ij}] n^{ij}$. 

Boren Li  Chapter 4. DRM-based Color Photometric Stereo 47
Each entry in the studentized residual vector $\tilde{r}^{i,j}$ is then approximated using:

$$\tilde{r}^{i,j}_k = \frac{r^{i,j}_k}{\sqrt{(\text{MSE})^{i,j}_k (1 - h^{i,j}_{a,k})}}, \quad (4.17)$$

where $(\text{MSE})^{i,j}_k$ represents the mean squared error of $r^{i,j}_k$. If any entry in $|\tilde{r}^{i,j}|$ exceeds $T_o$, this entry is regarded as outlier. The PS in UV repeats rejecting outliers until no more can be detected or $(\text{MSE})^{i,j}_k$ is smaller than a tolerance, $T_m$. With the general rule of thumb in detecting outliers using studentized residuals, $T_o$ and $T_m$ can be chosen as $2.5$ and $9\sigma^2_n$, where $\sigma_n$ is the standard deviation of additive white Gaussian noise that can be estimated using the method given by [78].

After PS in UV, $\bar{n}^{i,j}$ is initialized as $(\bar{n}^{i,j})_0$ and $\rho^{i,j}$ is simultaneously obtained. $\hat{k}^{i,j}_d$ is then initialized as $(\hat{k}^{i,j}_d)_0$ by computing $\rho^{i,j}/\kappa^{i,j}_z$. With the estimated $(\hat{k}^{i,j}_d)_0$ and by assuming no cast-shadows appear, the diffuse component matrix, $[\hat{E}^{i,j}_{d,RGB}]$, is rerendered as:

$$[\hat{E}^{i,j}_{d,RGB}] = \max \left( (\hat{k}^{i,j}_d)_0 [L^{i,j}] (\bar{n}^{i,j})_0, 0_N \right) (\bar{d}^{i,j}_{RGB})^T, \quad (4.18)$$

where $0_N$ is an $N \times 1$ vector with all zero entries. The specular component matrix, $[\hat{E}^{i,j}_{s,RGB}]$, is obtained by $\max \left( (\hat{f}^{i,j}_s)_0 \bar{s}^{T}_{RGB}, 0_N \right)$. With the separated color diffuse and specular components, the proposed method allows the functionality of specular removal and intrinsic image decomposition.
4.2.4 Surface Normal Refinement

Specular Parameter Initialization

If dense non-Lambertian reflectance appears at pixel \((i, j)\), which implies that more than one entries in \((f_s^{i,j})_0\) are in the specularity map, surface normal refinement is necessary since the sparse non-Lambertian reflectance assumption in the diffuse-specular separation is violated. With the estimated \((f_s^{i,j})_0\), the next objective is to initialize the specular parameters, \((k_s^{i,j})_0\) and \((\beta^{i,j})_0\) with fixed \((\hat{n}^{i,j})_0\). The specific cost function is given by:

\[
\epsilon_s^{i,j} = \sum_{k=1}^{Q} \left( (f_s^{i,j})_0 - k_s^{i,j} \left( (\hat{h}_k^{i,j})^T (\hat{n}^{i,j})_0 \right)^{\beta^{i,j}} \right)^2,
\]

(4.19)

where \(Q\) is number of pixels in the specularity map for pixel \((i, j)\) and \(Q \geq 2\). Equation (4.19) suggests a nonlinear least-squares problem, while it can be manipulated to the natural logarithm domain as a linear problem given by:

\[
\epsilon_{ls}^{i,j} = \sum_{k=1}^{Q} \left( \ln (f_s^{i,j})_0 - \ln (k_s^{i,j}) - \beta^{i,j} \ln \left( (\hat{h}_k^{i,j})^T (\hat{n}^{i,j})_0 \right) \right)^2.
\]

(4.20)

The solution for (4.20) is given as follows. Let \(\left(f_s^{i,j}\right)_0\) with \(Q\) rows consist of entries in \((f_s^{i,j})_0\) in the specularity map and \(\left[H\right]^{i,j} = [\hat{h}_1^{i,j}, \hat{h}_2^{i,j}, ..., \hat{h}_Q^{i,j}]^T\) be the matrix comprised of the corresponding unit half vectors. Then the solution of \(\ln k_s^{i,j}\) and \(\beta^{i,j}\) minimizing equation (4.20) is given by:

\[
\begin{bmatrix}
\beta^{i,j} \\
\ln k_s^{i,j}
\end{bmatrix} = \left(\left[S^{i,j}\right]^T \left[S^{i,j}\right]\right)^{-1} \left[S^{i,j}\right]^T \ln \left(f_s^{i,j}\right)_0,
\]

(4.21)

where \(\left[S^{i,j}\right] = \left[\ln \left(\left[H\right]^{i,j} (\hat{n}^{i,j})_0\right), 1\right]_Q\) and \(1_Q\) is a \(Q \times 1\) vector with all entries equal to 1. Up to this point, all parameters in the DRM have been initialized pixel-wise.
Segmentation Based on Diffuse Color

In the presence of image noise with just a small number of observations with specularities on each pixel, the initial estimates of \( (\hat{k}^i_j)_0 \) and \( (\beta^i_j)_0 \) for each pixel is sensitive to image noise, resulting in problematic parameter initializations. However, in reality, the same material is usually spread spatially in a scene among a variety of pixels. The specular parameters, \( \hat{k}^i_j \) and \( \beta^i_j \), are the same for a certain material. Therefore, based on the statistical metrics of the estimated \( \hat{k}^i_j \) and \( \beta^i_j \) for the same material, better initializations can be achieved.

Based on the above analyses, the next pertinent issue at hand is how to segment the scene to regions with each comprised of the same material. A straight-forward feature to judge the same material is the unit diffuse color, \( \bar{d}^{i,j}_{\text{RGB}} \). Based on the k-means clustering method \[79\], the scene can be segmented. For each segmented region, the medians of \( (\hat{k}^i_j)_0 \) and \( (\beta^i_j)_0 \) are adopted as the common initialization parameters.

Surface Normal Refinement

The next step is to jointly refine \( \hat{k}^i_d, \bar{k}^i_s, \beta^i_j \) and \( \bar{n}^{i,j} \) to enhance accuracy at pixels where dense non-Lambertian reflectances appear. The trust-region method with boundary constraints \[80\] is employed to iteratively refine \( \hat{k}^i_d, \bar{k}^i_s, \beta^i_j \) and \( \bar{n}^{i,j} \) by minimizing the cost function given by (4.12). \( \hat{k}^i_d, \bar{k}^i_s, \beta^i_j \) are first optimized by fixing \( \bar{n}^{i,j} \) as a constant. \( \bar{n}^{i,j} \) is then refined by fixing the other parameters. \( \hat{k}^i_d \) and \( \bar{k}^i_s \) are bounded with the deviation from \( (\hat{k}^i_d)_0 \) and \( (\bar{k}^i_s)_0 \) by ±10%. \( \beta^i_j \) is bounded by ±5% deviating from \( (\beta^i_j)_0 \). These boundary constraints are necessary to find the local minimum close to the initial guess for the purpose of increasing the robustness for the non-linear optimization and mitigating over-fitting. Note that the fixed \( \bar{d}^{i,j}_{\text{RGB}} \) will not affect the final estimation of \( \bar{n}^{i,j} \) since \( \hat{k}^i_d \) is refined in the optimization as the variation for the term, \( \hat{k}^i_d (\bar{d}^{i,j}_{\text{RGB}})^T \bar{s}_{\text{RGB}} \). In other words,
this step is not capable to refine $\bar{d}_{i;j}^{RGB}$ so that there exists a diffuse color ambiguity. This fact strengthens the necessity of using the RPCA for more accurate $d_{i;j}^{RGB}$ estimation in the diffuse-specular separation process. Another important issue being considered is the alternative cost function from (4.12) when dealing with reflectances having very strong specularities. Sometimes, the specular peaks have a few orders of magnitude higher than the diffuse component. Therefore, minimizing the cost function of (4.12) tends to overemphasize the importance of these specular peaks. Alternatively, minimizing the cost function in the log domain can provide better results. This cost function is given by:

$$
\epsilon_{i;j}^{d,s} = \sum_{k=1}^{N} \left( \log_{10} (\hat{e}_{k}^{i,j} s_{i,j}^{RGB})^T \hat{s}_{RGB} - \log_{10} \left( k_{d}^{i,j} (\hat{n}_{i,j}^{d})^T \hat{l}_{k}^{i,j} (\hat{d}_{i;j}^{RGB})^T \hat{s}_{RGB} - k_{s}^{i,j} (\hat{n}_{i,j}^{s})^T \hat{l}_{k}^{i,j} )^{\beta_{i;j}} \right) \right)^2.
$$

(4.22)

The physical meaning of using the log domain is similar to the usage of decibel watt rather than watt directly.

In this step, both the diffuse and specular components are incorporated into the cost function such that all the information encoding the surface orientations is exploited for more accurate estimations of $\bar{n}_{i;j}$. The simultaneously refined parameters, $k_{d}^{i,j}$, $k_{s}^{i,j}$ and $\beta_{i;j}$, are also valuable for more functionalities, such as digital relighting and material classification. These advantages take credit for the fact that the proposed PS method treats specularities as friends rather than foes.

### 4.2.5 Remarks

The strength of the proposed color PS method originates from the diffuse-specular separation process. With the separated diffuse components, surface orientations can be reliably initialized. The estimated unit diffuse colors, $\hat{d}_{i;j}^{RGB}$, are also valuable for the later material segmentation purpose. Without the separated specular components, non-linear optimiza-
tion is unavoidable and the cost function (4.20) can be never formulated into the log-linear form to initialize $\left(\hat{k}^i_j\right)_0$ and $\left(\beta^i_j\right)_0$. Taking advantage of the diffuse-specular separation, all the parameters in the DRM can be reliably initialized, which opens the door to treat specularities as meaningful signals to enhance the accuracy of surface orientation estimation at pixels where dense non-Lambertian reflectances appear. The information encoding in the specularities provides more potential functionalities for the proposed PS method, such as digital relighting and material classification.

4.3 Evaluations on Surface Orientation Estimation

4.3.1 Evaluations Using Synthetic Input Images

The first experiment aims to demonstrate the effectiveness of flows in the proposed color PS method and emphasize on the evaluation of the novel surface normal refinement process for dense non-Lambertian reflectances using synthetic input images. A scene with six different-colored spheres were rendered under eight illuminants using the Blinn-Phong model. Figure 4.4(a) shows the rendered image under the first illuminant. The sphere was adopted as the scene geometry since it samples uniformly of the surface orientations longitudinally and altitudinally. Figure 4.4(b) gives the normal map of the scene geometry. Different colors of the spheres were introduced for more comprehensive evaluations of the proposed PS method on surfaces with various spectral reflectances. $\bar{s}_{RGB}$ was set the same across the scene as $[0.5774, 0.5774, 0.5774]^T$. Three colors of the spheres were red, green and blue with the same chromatic angle, $\psi^i_j$, of 57.74°, while the other three spheres were yellow, cyan and magenta with $\psi^i_j$ of 35.26°. These unit diffuse colors of the spheres are given in Figure 4.4(c). The eight illuminants were configured in a circular pattern to create sufficient number of pixels
with dense non-Lambertian reflectances. Distant point source was assumed such that light directions across the scene were the same under certain illumination. The zenith angles of the light direction, $\eta_k$, were configured the same for each illuminant as $25^\circ$. The azimuth angles of the eight illuminants were uniformly distributed with $\phi_0 = 0^\circ$. $\tilde{k}^{i,j}_d$, $\tilde{k}^{i,j}_s$ and $\beta^{i,j}$ were set the same across the scene as 0.4, 0.2 and 100, respectively. Additive white Gaussian noise was introduced with standard deviation, $\sigma_n$, of 0.01. 100 repeated tests were conducted and the median values were adopted for evaluation.

Figure 4.4(d) presents the angular error for the $\tilde{d}^{i,j}_{RGB}$ estimation under $T_d = 0.02$. It shows that the proposed RPCA algorithm in estimating $\tilde{d}^{i,j}_{RGB}$ was effective in detecting strong specularities. In such regions, the angular errors were at similar levels as those without specularities. The choice of $T_d$ affects the $\tilde{d}^{i,j}_{RGB}$ estimation significantly. Figure 4.5 illustrates two examples when $T_d$ was chosen improperly. Figure 4.5(a) shows the angular error of the estimated unit diffuse colors under $T_d = 0.002$ and Figure 4.5(b) presents that under $T_d = 0.2$. When $T_d$ was chosen too small as 0.002, the RPCA algorithm was capable to detect even weaker specularities at the cost of more sensitive to image noise, especially at regions near the boundaries of the spheres where the signal to noise ratios (SNR) were weak. When $T_d$ was set too big as 0.2, the RPCA algorithm performed similar to the DPCA algorithm as used in Barsky’s PS and the error was large at pixels where specularities appeared. This observation further justified the necessity of using RPCA rather than DPCA. Another important observation was that regions with the same chromatic angle, $\psi^{i,j}$, have similar angular error of the estimated unit diffuse colors. For regions where $\psi^{i,j}$ were bigger, it was easier to identify the specularities using RPCA, resulting in smaller errors as shown by Figure 4.5(a). By contrast, if the specularities were not properly detected as shown by Figure 4.5(b), their errors were bigger than those with smaller $\psi^{i,j}$.

The efficacy of the additional surface normal refinement process was then demonstra-
Figure 4.4: (a) Rendered image under the first illuminant; (b) Normal map of the scene; (c) Unit diffuse colors of the scene; (d) Angular error of the estimated unit diffuse colors under $T_d = 0.02$.

ed by comparing the angular errors of surface orientation estimation with and without it. The other parameters of the proposed PS method were configured as: $T_c = 2^\circ$, $T_o = 2.5$, $T_m = 0.0009$. Figure 4.6(a) and (b) shows the angular error of the estimated surface orientations without and with surface normal refinement, respectively. Figure 4.6(a) is the result after PS in UV, while Figure 4.6(b) is the final result. As was shown from Figure 4.6(a), the central regions of the spheres where dense non-Lambertian reflectances appear had larger error. After the proposed surface normal refinement step, the angular error of surface orientations at the detected dense non-Lambertian reflectance region were significantly reduced as shown by 4.6(b). Errors in such regions were even smaller than those with sparse
non-Lambertian reflectance, which further enhanced the ground that making specularities as meaningful signals is beneficial for surface orientation estimation. For further quantitative evaluations, the improvement percentage of surface orientation estimation by including the surface normal refinement step is shown in Figure 4.6(c). Only regions with dense non-Lambertian reflectances were shown since the surface normal refinement was performed only in such regions. The histogram of this improvement percentage is given in Figure 4.6(d). From the histogram, 86.4% of the surface orientations converge to a better solution after the surface normal refinement process and the median improvement percentage was 34.56%. The first and third quantile of the improvement percentage were 11.85% and 57.92%, respectively. Under heavily corrupted image irradiances, the surface normal refinement process with just a few number of observations could converge to a worse solution due to the nonlinearity of the optimization problem. This deficiency can be reduced by including more observations. The performance improvements onto the six different-colored spheres were shown in Table 4.1. The improvements were obvious for all spheres, implying that the method was applicable to a wide range of surfaces with different spectral reflectances. Better estimation of $\mathbf{d}_{RGB}^{i,j}$ led to more accurate recovery of $\left( \mathbf{f}_{s}^{i,j} \right)_0$, resulting in the more reliable initialization of $\left( \mathbf{n}^{i,j} \right)_0$. Such
initializations of \((\hat{n}^{ij})_0\) ultimately affected the different improvement performances. Overall, the accuracy was mostly enhancement by including the additional surface normal refinement step. The improvement was more phenomenal at pixels where \(\hat{d}_{RGB}^{ij}\) was estimated more accurately.

In the second experiment, the performance improvement due to surface normal refinement was verified onto a wide range of reflectances. Experimental settings and parameters of the proposed method were the same as the first experiment except that different values of \(\hat{k}_d^{ij}, \hat{k}_s^{ij}\) and \(\beta^{ij}\) were applied to represent various surface reflectances when rendering
Table 4.1: Improvement on different-colored spheres

<table>
<thead>
<tr>
<th>Index</th>
<th>Color</th>
<th>$\psi_{i,j}$</th>
<th>Median error of $\vec{d}_{i,j}^{RGB}$</th>
<th>Median improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>red</td>
<td>57.74°</td>
<td>0.69°</td>
<td>31.27%</td>
</tr>
<tr>
<td>2</td>
<td>yellow</td>
<td>35.26°</td>
<td>0.63°</td>
<td>37.88%</td>
</tr>
<tr>
<td>3</td>
<td>green</td>
<td>57.74°</td>
<td>0.69°</td>
<td>31.45%</td>
</tr>
<tr>
<td>4</td>
<td>cyan</td>
<td>35.26°</td>
<td>0.63°</td>
<td>37.28%</td>
</tr>
<tr>
<td>5</td>
<td>blue</td>
<td>57.74°</td>
<td>0.69°</td>
<td>31.04%</td>
</tr>
<tr>
<td>6</td>
<td>magenta</td>
<td>35.26°</td>
<td>0.63°</td>
<td>38.43%</td>
</tr>
</tbody>
</table>

The ratio of $\kappa_{d_s}^{i,j} = \tilde{k}_d^{i,j} / \tilde{k}_s^{i,j}$ represents the relative strength between the diffuse and specular components, whereas $\beta_{i,j}$ indicates the width of the specular lobe. The mean, median, first and third quantile values of the improvement were adopted for evaluation and the angular error of $\vec{d}_{i,j}^{RGB}$ estimation was used for analyses. As shown from Table 4.2, the improvements on all the different reflectances were obvious. The improvements were larger for reflectances where $\beta_{i,j}$ and $\kappa_{d_s}^{i,j}$ were bigger. Larger value of $\beta_{i,j}$ suggesting narrower specular lobe made the $\vec{d}_{i,j}^{RGB}$ estimation more accurate since the sparse non-Lambertian reflectance assumption was valid. Similarly, larger value of $\kappa_{d_s}^{i,j}$ indicating stronger diffuse components made better $\vec{d}_{i,j}^{RGB}$ estimation due to stronger inliers. More accurate $\vec{d}_{i,j}^{RGB}$ estimation led to bigger improvement from the surface normal refinement. In summary, with the additional surface normal refinement step, the accuracy of surface orientation estimation is enhanced by around 30% by median and the improvement is more obvious for reflectance that has narrower specular lobe and stronger diffuse component.

The third experiment aims to evaluate the applicability of the proposed PS method onto a wide range of dielectric material reflectances. Thirty material BRDFs were evaluated, where the different BRDFs were provided from the MERL database \cite{69}. Experimental settings and the method parameters were exactly the same as the first experiment except that the BRDFs were replaced. Figure 4.7 shows the performance of surface orientation
Table 4.2: Evaluation of the improvement for different reflectances

<table>
<thead>
<tr>
<th>$\tilde{k}_d/\tilde{k}_s$</th>
<th>$\beta$</th>
<th>mean</th>
<th>median</th>
<th>first quantile</th>
<th>third quantile</th>
<th>median error of $\tilde{d}_{RGB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4/0.2</td>
<td>100</td>
<td>24.42%</td>
<td>34.56%</td>
<td>11.85%</td>
<td>57.92%</td>
<td>0.66°</td>
</tr>
<tr>
<td>0.4/0.4</td>
<td>100</td>
<td>22.87%</td>
<td>34.12%</td>
<td>9.96%</td>
<td>59.94%</td>
<td>0.79°</td>
</tr>
<tr>
<td>0.4/0.8</td>
<td>100</td>
<td>22.38%</td>
<td>34.09%</td>
<td>8.56%</td>
<td>60.89%</td>
<td>1.08°</td>
</tr>
<tr>
<td>0.4/0.4</td>
<td>20</td>
<td>4.81%</td>
<td>21.60%</td>
<td>−3.61%</td>
<td>49.10%</td>
<td>1.14°</td>
</tr>
<tr>
<td>0.4/0.8</td>
<td>20</td>
<td>1.88%</td>
<td>21.46%</td>
<td>−5.30%</td>
<td>49.27%</td>
<td>1.43°</td>
</tr>
<tr>
<td>0.4/0.8</td>
<td>20</td>
<td>0.11%</td>
<td>21.01%</td>
<td>−8.37%</td>
<td>50.51%</td>
<td>2.05°</td>
</tr>
</tbody>
</table>

estimation on the 30 materials using box-and-whisker plot. The red and black dot represent the mean and median value, respectively. The lower and upper bound of the blue box indicate the first and third quantile values. As is shown, the proposed method can accurately estimate $\tilde{n}^{i,j}$ for most dielectric materials within 5 degrees on average.

4.3.2 Evaluations Using Real Images

The proposed method was evaluated on six different datasets comprised of real images in the DiLigenT database [20]. These six datasets were chosen to evaluate different representative scenes in the real world and they were BUDDHA, BEAR, POT1, POT2, READING and GOBLET whose names were the same as given by the database. In terms of BRDFs, these datasets covered materials with mostly diffuse (POT1), strong and sparse specular spikes (READING), broad and soft specular lobes on uniform (BUDDHA, BEAR) and spatially-varying materials (POT2, GOBLET). In terms of surface geometry, these datasets contained smoothly curved surfaces (BEAR, GOBLET), smooth surfaces with local details (POT1, POT2), and surfaces with complicated geometry (BUDDHA, READING). The parameters in the proposed method were set as: $T_d = 0.01$, $T_c = 2^\circ$, $T_o = 2.5$ and $T_m = 0.0009$. The specular color, $s_{RGB}$, was assumed to be $[0.5774, 0.5774, 0.5774]^T$.

From the first to six column in Figure 4.8, the results on BUDDHA, BEAR, POT1,
Figure 4.7: Box-and-whisker plot for angular error of surface orientations for 30 different dielectric materials

POT2, READING and GOBLET are respectively given. Figure 4.8(a) shows the image irradiance under the first illuminant. Figure 4.8(b) demonstrates the ground truth of the normal map and Figure 4.8(c) gives the angular error of the estimated surface orientations. As can be clearly observed from the results, the angular error of the estimated surface orientations was even smaller at regions where specularities appear in the first five datasets, such as at the 'abdomen' of the BUDDHA and the 'nose' of the BEAR. This fact was due to the strength of the surface normal refinement step by treating specularities as meaningful signals rather than outliers. In the GOBLET dataset, however, deficits can be clearly observed from the specular regions, which was due to the shift of the specular color for metallic surfaces. The dichromatic reflectance model was not accurate in characterizing such surface reflectances since the wavelength and geometry exhibit inter-dependency. Therefore, the proposed PS method could be only effective for dielectric material reflectances from the above observations. Note also that the proposed method was only effective for surfaces whose diffuse color
Figure 4.8: Evaluations on surface orientation estimation using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Ground truth of the normal map; (c) Angular error of the estimated surface orientations

was distinct from the specular color as shown clearly on the 'book' in the READING dataset. Observed from Figure 4.8(c), errors most likely occur at three types of regions: 1) concave regions, such as the 'neck' in the BEAR dataset, due to inter-reflections; 2) dark regions, such as the 'cap' in the POT1 dataset, due to low SNR; 3) occluding boundaries due to both low SNR and lack of modeling of Fresnel reflection with low grazing angles.

For more comprehensive quantitative evaluations, the results from the proposed method were compared with nine other PS methods which are listed in Table 4.3. The results from the other methods were provided from the DiLigenT database. These methods were representative and have been well-recognized. Due to the data availability, [39] and [41] were not involved in the comparison, while the essence of their works in using DRM with known $s_{RGB}$ was well-inherited by the proposed method. In order to compare the efficacy of surface normal refinement, the surface orientation estimation after PS in UV and after the surface normal refinement were separately given as abbreviated by UV and DRM in Table 4.3, respectively.
Table 4.3: Comparison with other existing PS methods

<table>
<thead>
<tr>
<th>Method index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>10.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>[19]</td>
<td>[25]</td>
<td>[27]</td>
<td>[28]</td>
<td>[37]</td>
<td>[38]</td>
<td>[29]</td>
<td>[81]</td>
<td>[30]</td>
<td>UV</td>
<td>DRM</td>
</tr>
</tbody>
</table>

Figure 4.9: Comparison of surface orientation estimation on six datasets using ten different PS methods.

Figure 4.9 shows the comparison results from the ten different PS methods on the six datasets. The X-axis represents the method index and the Y-axis indicates the angular error of surface orientations. The box-and-whisker plot was adopted to demonstrate these angular errors. The red dot represents the mean value and the black dot indicates the median value. The lower and upper bound of the blue bars suggest the first and third quantile values, respectively. As was shown, the proposed method performed well comparatively in the first five datasets whose scenes were made of dielectric materials. It did not outperform the other existing methods as [38] or [30] due to the drawback of the DRM in characterizing metallic reflectances. Comparing the results of the proposed method without and with the additional
surface normal refinement step (See index 10 and 10.5), it can be clearly noticed that the angular error of surface orientations have been reduced statistically in the first five datasets. This fact further justified the significance by including this additional step.

### 4.4 Implications

![Figure 4.10: A schematic illustration for the implications of the proposed DRM-based color PS method](image)

Figure 4.10 shows a schematic illustration for the implications of the proposed DRM-based color PS method using the POT2 dataset in the DiLigenT database. After the proposed DRM-based color PS method, the normal map, albedo map \((\tilde{k}_d^{i,j}\tilde{d}_{\text{RGB}}^{i,j})\), \(\tilde{k}_s\) map and \(\beta\) map can be simultaneously obtained. The acquisition of such parameters allows more functionalities through scene synthesis, such as specular removal, intrinsic image decomposition, digital relighting, material-based segmentation, material transfer, material inference
and classification. In this section, brief discussions with preliminary results for some of these functionalities are given.

### 4.4.1 Specular Removal

![Specular Removal Images](image)

Figure 4.11: Specular removal using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Separated diffuse component; (c) Separated specular component

Specular removal is significant in many applications in computer vision and computer graphics, such as stereo vision, visual recognition and tracking, where visually consistent surface appearances are desired and the presence of specularities contaminates such appearances. Since the color diffuse and specular components are obtained through diffuse-specular separation in the proposed method, specular removal can be viewed as a by-product and is performed pixel-wise. The diffuse-specular separation is feasible owing to the known specular color with the validity of DRM. The results on the six datasets in the DiLigenT database are shown in Figure 4.11. Though these results looked promising visually, quantitative analyses were required and left of future works.
4.4.2 Intrinsic Image Decomposition

![Figure 4.12: Intrinsic image decomposition using DiLigenT database: (a) Image irradiance under the first illuminant; (b) Estimated albedo map; (c) Estimated shading map](image)

Intrinsic image decomposition calls for factorizing images into components of intrinsic material properties of the scene from illumination effects [82]. Many computer vision algorithms, such as segmentation, recognition and motion estimation are confounded by illumination effects and substantially benefited from reliable estimation of the illumination-invariant material properties for the scene. Conventionally, images are decomposed into the albedo map and the shading map, where the albedo map is invariant to illuminations. In the proposed method, the albedo map is obtained by estimating $\mathbf{d}_{RGB}^{ij}$ through RPCA and $\mathbf{k}_{d}^{ij}$ through PS. The shading map is rerendered using the known lighting and the estimated surface orientations. Figure 4.12 shows the results of the intrinsic image decomposition on the six datasets in the DiLigenT database. Visually pleasing results have been achieved. The acquired albedo map is beneficial to solve the correspondence problem in the presence of specularities. It brings a new horizon to the fusion problem of PS with other techniques, such as multiple view geometry and structure from motion [58].
4.4.3 Digital Relighting

Digital relighting refers to the functionality that is capable to online capture images of the scene under different illuminations at first and then relight the scene offline under a completely new illumination condition. The captured images are used to train the parameters in the image photometry formation model, including all the parameters in the DRM. Relighting the scene is simply to rerender an image by varying the illumination parameters and fixing the others in the DRM. It is of particularly useful to relight the scene with lighting direction the same as the viewer direction. This lighting condition is important since no cast-shadows occur that can maximize the appearance information in a single image. This lighting condition is also not physically plausible and images under such lighting condition can be only generated through digital relighting. Figure 4.13 (a) and (b) respectively shows the captured image that were excluded from the training inputs and the digital relighted image. The errors on the six datasets are shown in Figure 4.13(c) and the RMS errors are listed in Table 4.4. The error of digital relighting relates to the residual of the proposed DRM-based color
Table 4.4: RMS error of digital relighting

<table>
<thead>
<tr>
<th>Dataset</th>
<th>BUDDHA</th>
<th>BEAR</th>
<th>POT1</th>
<th>POT2</th>
<th>READING</th>
<th>GOBLET</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS error</td>
<td>0.0183</td>
<td>0.0504</td>
<td>0.0194</td>
<td>0.0258</td>
<td>0.1341</td>
<td>0.1208</td>
</tr>
</tbody>
</table>

PS method. The error can be reduced by using a more sophisticated BRDF model, such as the Cook-Torrance model, instead of the Blinn-Phong model in characterizing the surface reflectance.

Digital relighting can be also applied to classify shadows into two categories: self-shadow and cast-shadow. Self-shadows can be rerendered locally using digital relighting since they are determined only from the negativity of the surface irradiance terms ($\cos \theta_{i;j}^{i}$). Cast-shadows, however, cannot be rendered without knowing the geometry of the surface. Figure 4.14 illustrates an example for the shadow classification problem using digital relighting. Figure 4.14(a) shows the captured image irradiance and Figure 4.14(b) demonstrates the relighted image. Figure 4.14(c) indicates the error of digital relighting. Since cast-shadows cannot be determined in the relighted image, the difference between Figure 4.14(a) and (b) at cast-shadow regions was phenomenal. By setting a threshold on this difference and performing the ‘AND’ operation with the shadow map given by Figure 4.14(d) where bright pixels represent shadows, cast-shadows can be identified. The rest pixels in the shadow map are then classified as self-shadows. The shadow classification results in this example are shown in 4.14(e) where pixels in green suggest cast-shadow regions and yellow pixels indicate self-shadows. This shadow classification is significant since cast-shadows encode powerful geometric cue: if one pixel casts a shadow onto another, then the two pixels are collinear with the light direction. By employing shape-from-shadow techniques, such as [83], using the identified cast-shadows and known light positions, a sparse point cloud can be acquired and posed as a shape priori to the surface normal integration (SNI) in reconstructing the
surface geometry. Another possibility to use the cast-shadow constraint to reconstruct the surface is by adding more terms in the cost function of SNI as proposed by [36]. The proper usage of these cast-shadows remains open.

![Figure 4.14: Shadow classification using digital relighting: (a) Captured image; (b) Relighted image; (c) Error of the relighted image; (d) Detected shadows before performing PS (bright pixel represents shadow); (e) Shadow classification results (green indicates cast-shadows and yellow implies self-shadows)](image)

Material-based Segmentation

![Figure 4.15: Material-based segmentation based on extracted material-dependent features](image)

Partitioning the scene into regions that have common properties plays a crucial role in many computer vision tasks. Segmentation based on the material composition is a very
challenging problem since the appearance of materials varies significantly under different viewing and lighting conditions. Due to the uniqueness of the PS problem, the viewer and lighting are fully controllable, which makes the problem more tractable. As is shown from Figure 4.15, the material-dependent features, $\bar{d}_{R, G, B}^{i,j}$, $\bar{k}_d^{i,j}$, $\bar{k}_s^{i,j}$, and $\beta^{i,j}$ can be obtained from the proposed color PS method. These features are invariant to the surface geometry and illumination direction. Therefore, they have provided several independent features characterizing the reflectances of different materials. By applying a simple k-means clustering method [79] on the $\bar{d}_{R, G, B}$ map, the segmentation map was obtained as shown in Figure 4.15 where the same color represents the same segmented region. Incorporating the other material-dependent features can provide more variability of the segmentation.

### 4.4.4 Material Transfer

![Material transfer using BRDF database](image)

Figure 4.16: Material transfer using BRDF database
Using the material-based segmentation results, regions comprised of the same material can be easily transferred to another one using the BRDF database. An example is shown in Figure 4.16. The BRDFs used here were from the MERL database. The names of the edited scene after material transfer were the same as transferred material names in the MERL database. This material transfer operation provides a convenient way to edit the scene and the proposed method further enhanced this capability that can transfer materials from the other captured scenes. Figure 4.17 illustrates an example of this capability. The material of the BUDDHA was transferred in this case from parts of the other five scenes in the DiLigenT database. This operation is feasible owing to the simultaneously obtained material-dependent features, i.e. $\hat{d}_{RGB}^{ij}$, $\hat{k_d}$, $\hat{k_s}$, and $\beta$, by the proposed method.

Through training the BRDF database to a certain analytical BRDF model and using the captured material-dependent features in the same model, material inference and classification can be also feasible. It could open up more potential applications, such as for online
shopping where the customer cannot physically inspect the product. For this purpose, more complicated analytical BRDFs may be necessary to create a higher-dimensional material-dependent feature space. Thorough investigation of this topic is beyond the scope of the dissertation.

4.5 Conclusions and Future Works

A PS method using color images dealing with non-Lambertian reflectance has been proposed. The method formulates the imaging photometry using DRM with known specular color. It extracts surface orientations not only from the diffuse components but also specularities owing to the diffuse-specular separation. The proposed method improves the accuracy for surface orientation estimation at pixels where dense non-Lambertian reflectance appear by introducing the additional surface normal refinement step using information from specularities. The simultaneously acquired DRM parameters can be applied for more potential functionalities, such as digital relighting and material classification..

From the evaluations on the newly proposed surface normal refinement step, the results indicate that with the additional step, the accuracy is enhanced by around 30% in median and the improvement is more phenomenal where the unit diffuse colors are better estimated. From the systematic comparison on six datasets with nine other representative PS methods, the proposed method shows its descent performance on reflectances of dielectric materials and degradation on metallic surfaces due to the limitation of DRM. The result has also demonstrated that the proposed method is only feasible for surfaces whose diffuse and specular color are distinct. Under the framework of the proposed method, more functionalities, including specular removal, intrinsic image decomposition, digital relighting, material-based segmentation, material transfer, and material classification, are feasible, which were demon-
strated through preliminary results.

This chapter presented the first set of results by the proposed DRM-based color PS method, while it can be extended in a variety of ways. First, other physically-based analytical BRDFs can be employed instead of the Blinn-Phong model to better account for the Fresnel reflection. Second, an alternative method can be proposed specifically for surfaces whose diffuse and specular color are close to extend the method applicability. Last but not least, the implications discussed in Section 4.4 contain huge amounts of future works in theories and applications.
Chapter 5

Surface Normal Integration Dealing with Perspective Distortion

This chapter presents a surface normal integration (SNI) method dealing with perspective distortion. The proposed method is independent of PS, which facilitates to keep the image irradiance equation in a simple form that can be inverted using less images. The perspective SNI is performed on the image plane instead of on the target surface as did by orthographic SNI owing to the newly derived representation of surface normals. This is crucial since spatial surface points are not uniformly distributed under perspective projection, whilst image points always are. This new surface normal representation is then manipulated to the so-called log gradient space that shows strong analogy to the gradient space used in orthographic SNI. The manipulation is inspired by [55]. Having this analogy, past orthographic SNI techniques, such as [47][49], are directly applicable. Performances were evaluated via comparing the results between orthographic and perspective SNI through simulations. The proposed method opens a door to deal with perspective distortion in a simple way that is PS-independent and adaptable to most past orthographic SNI techniques.
Chapter 5. Perspective Surface Normal Integration

This chapter is organized as follows. The next section briefly reviews orthographic SNI. Section II presents the proposed method dealing with perspective distortion. Section III demonstrates the efficacy of the proposed method via simulations, while conclusions and future works are summarized in the last section.

5.1 Orthographic Surface Normal Integration

Unit surface normals, \( \tilde{n}^{i,j} \), are obtained for each pixel after PS. The objective of SNI is to integrate \( \tilde{n}^{i,j} \) to reconstruct the surface.

Let unit surface normal in \( \{C\} \) be represented as \( \tilde{n}^{i,j} = [\{C\} \tilde{n}^{i,j}_x, \{C\} \tilde{n}^{i,j}_y, \{C\} \tilde{n}^{i,j}_z]^T \). \( \{C\} \tilde{n}^{i,j} \) having two degrees of freedom can be transformed to the gradient space \([71]\) as

\[
\{C\} n^{i,j} = [\{C\} p^{i,j}, \{C\} q^{i,j}, -1]^T,
\]

where \( \{C\} p^{i,j} = -\frac{\{C\} \tilde{n}^{i,j}_x}{\{C\} \tilde{n}^{i,j}_z} = \frac{\partial \{C\} Z}{\partial \{C\} X} \) and \( \{C\} q^{i,j} = -\frac{\{C\} \tilde{n}^{i,j}_y}{\{C\} \tilde{n}^{i,j}_z} = \frac{\partial \{C\} Z}{\partial \{C\} Y} \).

If the depth relief is small compared with the average surface depth and the distance between image point and the optical axis is also small, orthographic projection can be applied to approximate perspective projection that is given by equation (3.1). The term, \( \frac{1}{\{C\} Z} \), can be then approximated by a constant as \( \frac{1}{\{C\} Z_0} \) and therefore, surface points in \( \{C\} X, \{C\} Y \) form a regular grid. The SNI process then happens on the surface directly and often tends to minimize the functional given by [46]:

\[
\epsilon = \int \int_{\Omega} \left[ \left( \frac{\partial \{C\} Z}{\partial \{C\} X} - \{C\} p \right)^2 + \left( \frac{\partial \{C\} Z}{\partial \{C\} Y} - \{C\} q \right)^2 \right] d\Omega, \tag{5.1}
\]

where \( \Omega \) is the surface domain to be integrated. Minimizing the functional leads to the Euler equation given by:

\[
\nabla^2 \{C\} Z = \{C\} p_X + \{C\} q_Y, \tag{5.2}
\]
where $\nabla^2$ denotes the Laplacian operator, $\{C\}_p X = \frac{\partial (C) p}{\partial (C) X}$ and $\{C\}_q Y = \frac{\partial (C) q}{\partial (C) Y}$. Finite difference method is then applied to approximate the Laplacian. The depth, $\{C\} Z$, can be recovered via solving a linear system of equations [44].

5.2 SNI under Perspective Projection

Perspective distortion is more obvious if assumptions for orthographic projection do not hold. The spatial distribution of the estimated $\{C\} \hat{n}^{ij}$ cannot be regarded as a regular grid considering perspective distortion and therefore, SNI cannot be appropriately performed on the target surface directly. In this section, a proper surface normal representation on the image plane is first derived, which facilitates the SNI to be performed on the image plane instead of on the target surface. SNI on the image plane is feasible since the surface corresponding image points are always uniformly distributed spatially due to the uniform distribution of the camera sensor cells. The proposed perspective SNI method based on the novel surface normal representation is then presented. The uniqueness of solution is finally discussed.

5.2.1 Surface Normal on Image Plane

Compared with the conventional 2D image coordinate system, $\{I\}$, the third dimension, referred to as the image point depth, $\{I\} Z$, is newly introduced. $\{I\} Z$ is defined as $\{I\} Z = \{C\} Z - f$ and is the deviation from the image point to its corresponding surface point along the positive optical axis direction.

Rewriting perspective projection in equation (3.1) with the definition of $\{I\} Z$, relation-
ships to derive the representation of \( \{C\} n \) in \( \{I\} \) are given by:

\[
\begin{align*}
\{I\} x &= \frac{f\{I\} X}{\{I\} Z + f} \\
\{I\} y &= \frac{f\{I\} Y}{\{I\} Z + f} \\
\{C\} Z &= \{I\} Z + f.
\end{align*}
\] (5.3)

Using such constraints, it is simple to derive \( \{C\} n \) in \( \{I\} \) as:

\[
\begin{align*}
\{C\} p &= \frac{f\{I\} Z_x}{\{I\} Z + f + \{I\} x\{I\} Z_x + \{I\} y\{I\} Z_y} \\
\{C\} q &= \frac{f\{I\} Z_y}{\{I\} Z + f + \{I\} x\{I\} Z_x + \{I\} y\{I\} Z_y},
\end{align*}
\] (5.4)

where \( \{I\} Z_x = \frac{\partial \{I\} Z}{\partial \{I\} x} \) and \( \{I\} Z_y = \frac{\partial \{I\} Z}{\partial \{I\} y} \). These two partial derivatives can be estimated using finite difference method on a uniform grid. SNI can be performed on the image plane.

What type of error the prevalently used orthographic projection assumption will bring about to the SNI? In other words, under what condition the orthographic SNI is valid? As can be observed, when \( \{I\} Z + f >> \{I\} x\{I\} Z_x + \{I\} y\{I\} Z_y \), equation (5.4) can be rewritten as:

\[
\begin{align*}
\{C\} p &\approx m_a \{I\} Z_x \\
\{C\} q &\approx m_a \{I\} Z_y,
\end{align*}
\] (5.5)

where \( m_a = \frac{f}{\{I\} Z + f} \) is the magnification ratio. Equation (5.5) is the surface normal representation under scaled orthography. Therefore, distortions are more likely to occur in orthographic SNI when:

- Image point is far away from the principal point, or
- Large surface gradient along \( \{C\} x \) or \( \{C\} y \) direction.
5.2.2 Perspective Surface Normal Integration

It is straightforward to formulate the perspective SNI problem using quadratic regularization, such as minimizing the functional given by the following form:

\[ \epsilon = \int \int_{\Omega} \left( (p(Z + f) + pyZ_y + (px - f) Z_x)^2 + (q(Z + f) + qxZ_x + (qy - f) Z_y)^2 \right) d\Omega, \]

where \( p \) and \( q \) are in \( \{C\} \), and the other variables are in \( \{I\} \) for notation simplicity. Finite difference method is applied to approximate the partial derivatives as:

\[
\begin{align*}
\frac{\partial Z^{i,j}}{\partial x} &\approx \frac{Z^{i,j+1} - Z^{i,j}}{\Delta x} \\
\frac{\partial Z^{i,j}}{\partial y} &\approx \frac{Z^{i+1,j} - Z^{i,j}}{\Delta y},
\end{align*}
\]

where \( \Delta x \) and \( \Delta y \) are the size of camera sensor cell in the \( x \) and \( y \) direction, respectively. Substituting equation (5.7) into (5.6), the discrete form of \( \epsilon \) can be manipulated. Taking the partial derivative of \( \epsilon \) with respect to each \( Z^{i,j} \) and let them equate to zero, the minimization problem can be formulated into a linear system of equations. Numerical techniques, like the conjugate gradient method, can be then applied to solve this linear system.

Though the minimization problem given by equation (5.6) is directly solvable, its complex form makes the solution computationally heavy. A more elegant form of the cost functional is, therefore, derived and is inspired by [55].

Denote:

\[
\begin{align*}
\hat{p} &= \frac{Z_x}{Z + f} = \frac{\partial \hat{Z}}{\partial x} \\
\hat{q} &= \frac{Z_y}{Z + f} = \frac{\partial \hat{Z}}{\partial y},
\end{align*}
\]

(5.8)
where $\tilde{Z} := \ln(Z + f)$. The ln operator naturally gives a bijective mapping. Since $Z$ is always positive because the scene is in front of the camera, a one-to-one mapping is guaranteed. $(x, y, \ln(Z + f))$ forms a 3D Cartesian space and is referred to as the log depth space. $(\hat{p}, \hat{q})$, in turn, forms a 2D subspace in the log depth space and is referred to as the log gradient space.

Substitute equation (5.8) into (5.4), the transformation from the log gradient space to the gradient space is derived as:

$$
\begin{align*}
    p &= \frac{f\hat{p}}{1 + x\hat{p} + y\hat{q}} \\
    q &= \frac{f\hat{q}}{1 + x\hat{p} + y\hat{q}}.
\end{align*}
$$

Note that the gradient space is in $\{C\}$ and the log gradient space is in $\{I\}$. $(\hat{p}, \hat{q})$ can be also represented in terms of $(p, q)$ deriving from equation (5.9) as:

$$
\begin{align*}
    \hat{p} &= \frac{p}{f - xp - yq} \\
    \hat{q} &= \frac{q}{f - xp - yq}.
\end{align*}
$$

Degeneracy happens when:

$$
f = xp + yq.
$$

However, from equation (5.9), it can be easily shown that (5.11) holds if and only if $f = 0$. Degeneracy cannot happen since the effective focal length can never be zero. As can be also observed, the right hand side of equation (5.10) is completely known after PS. From equation (5.8), the relationship of the depth and $(\hat{p}, \hat{q})$ has also been established. The perspective SNI
problem can be solved via minimizing the functional given by the more compact form as:

\[
\epsilon = \int \int_{\Omega} \left[ \left( \frac{\partial \hat{Z}}{\partial x} - \hat{p} \right)^2 + \left( \frac{\partial \hat{Z}}{\partial y} - \hat{q} \right)^2 \right] d\Omega.
\] (5.12)

Equation (5.12) is analogous to (5.1) and therefore, previous methods of orthographic SNI are directly applicable to perspective SNI.

In summary, the problem of perspective SNI can be efficiently solved following the steps below:

1. Estimate surface orientations, \( \hat{n} \), using PS and transform \( \hat{n} \) in \( \{C\} \) to the gradient space, \((p, q)\);

2. Transform \((p, q)\) to the log gradient space, \((\hat{p}, \hat{q})\), in \( \{I\} \) using equation (5.10);

3. Perform SNI by minimizing a functional, such as equation (5.12), and solve for \( \hat{Z} \) in the log depth space in \( \{I\} \);

4. Compute the depth map in \( \{C\} \) for each pixel corresponding surface point using \( \{C\} Z = e^{\hat{Z}} \);

5. Compute the point cloud using the reconstructed depth map with camera intrinsics.

The proposed method is completely PS-independent and the perspective distortion is treated by simply modifying the SNI process. It makes the proposed method adaptable to any past PS techniques in estimating surface orientations and enables the imaging photometry formation equation in a much simpler form represented by vectors in \( \{C\} \) rather than in \( \{I\} \), unlike methods of [55][56]. It takes full advantage of PS dealing with the complicated surface orientation and depth estimation as separate problems rather than using shape-from-shading-like methods in estimating them simultaneously.
5.2.3 Uniqueness of Solution

In general, the functional minimization problems given by equation (5.1) and (5.12) are ill-posed in the absence of suitable boundary conditions and integrability constraints. The uniqueness of solution in perspective SNI is concerned in this subsection.

Boundary conditions

The types of boundary condition to ensure the well-posedness of the minimization problem depend on the particular types of partial differential equations. There are two types of boundary conditions commonly applied, the Neumann (or natural) boundary condition and the Dirichlet (or essential) boundary condition.

The Neumann boundary condition specifies depth derivatives on the boundary in the integration domain, whilst the Dirichlet boundary condition imposes the depth directly. When prior knowledge of depth is unknown, the Neumann boundary condition is enforced when surface at the boundary is smooth. For perspective SNI, it is necessary to transform the boundary conditions to the log depth space before imposing these boundary conditions. The mixed (Robin) boundary condition can be also applied depending on the boundary information availability.

Integrability

Even with the boundary constraints, the minimization problem in the form of equation (5.12) does not specify a unique solution. In fact, adding any harmonic function on the solution results in a different solution yielding the same Euler equation given by:

$$\nabla^2 \hat{Z} = \hat{p}_x + \hat{q}_y.$$  (5.13)
This is because the above Euler equation is the divergence expression [44] where the integrability constraint, as given by equation 5.14, makes no contribution.

\[ \hat{p}_y = \hat{q}_x. \]  

(5.14)

It is easy to prove that equation (5.14) holds if and only if \( Z_{xy} = Z_{yx} \) using equation (5.8). Therefore, the integrability property can be studied directly in the log depth space.

The curl-divergence space [49], illustrated by Figure 5.1, makes the integrability problem easier to understand. Denote the curl and divergence operator in the log gradient space as \( c \) and \( d \), respectively. \( c \) and \( d \) is defined as:

\[
\begin{align*}
  c &= \frac{\partial \hat{p}}{\partial y} - \frac{\partial \hat{q}}{\partial x} = \hat{p}_y - \hat{q}_x \\
  d &= \frac{\partial \hat{p}}{\partial x} + \frac{\partial \hat{q}}{\partial y} = \hat{p}_x + \hat{q}_y.
\end{align*}
\]

(5.15)

Points in the 2D vector space are denoted as \( \mathbf{v} (\hat{p}, \hat{q}) = [c (\hat{p}, \hat{q}), d (\hat{p}, \hat{q})]^T \). Suppose \( \mathbf{v}_1 \) is a vector in the curl-divergence space for a particular pixel corresponding surface point. In fact, for this point, any vector with the same divergence, such as \( \mathbf{v}_2 \), results in the same constraint as equation (5.13). If the freedom from the curl is not restricted, the uniqueness
The integrability constraint is, therefore, introduced by minimizing the curl-square to make the resulting surface smooth and constrained. Two classes of methods were proposed. The first one is the classical method adding some penalty terms on the functional given by equation (5.12), such as [44]. The other class is to modify the reconstructed gradient field to be curl-free before minimizing the functional, such as [49]. Both classes are perfectly adaptable to the proposed method. The second class is suggested since the gradient field refinement is conducted locally prior to SNI such that the error does not propagate across the surface.

5.3 Performance Evaluation

In this section, the proposed perspective SNI methods were evaluated using three numerical experiments. The method directly minimizing equation (5.6) in the gradient space is referred to as PerSNIG. The method minimizing equation (5.12) in the log gradient space is referred to as PerSNILog. The orthographic SNI minimizing equation (5.1) is referred to as OrthSNIG. The performances of the three SNI methods were compared.

The first numerical simulation aims to demonstrate the efficacy of the proposed perspective SNI methods. A hemisphere with radius of 30 mm was adopted as the target surface geometry and modelled using triangular patches. The target surface was placed on the reference plane as illustrated in Figure 3.1. The transformation between \{W\} and \{C\} was related using \( R_W \) and \( t_W \), where \( R_W = diag(1, -1, -1) \) and \( t_W = [0, 0, 783]^T \). The image resolution was 600 × 400 and the camera intrinsic parameters were listed in Table 5.1. The surface visibility was then determined by computing ray-triangle intersection using Moller-Trumbore algorithm [84] and unit surface normals were then computed using finite difference
Table 5.1: Camera intrinsic parameters

<table>
<thead>
<tr>
<th>symbol</th>
<th>value</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f )</td>
<td>282.5</td>
<td>mm</td>
</tr>
<tr>
<td>( m_x )</td>
<td>16.67</td>
<td>pixel/mm</td>
</tr>
<tr>
<td>( m_y )</td>
<td>16.67</td>
<td>pixel/mm</td>
</tr>
<tr>
<td>( o_x )</td>
<td>300</td>
<td>pixel</td>
</tr>
<tr>
<td>( o_y )</td>
<td>200</td>
<td>pixel</td>
</tr>
</tbody>
</table>

method. The visible surface and the normal map were respectively shown in Figure 5.2(a) and (b).

Using the unit surface normals as inputs, the three SNI methods were then applied to reconstruct the surface. The pixel-wise distance error was adopted as the evaluation metric. The results are shown in Figure 5.3. Figure 5.3(a), (b) and (c) respectively show the distance error from OrthSNIG, PerSNILog and PerSNIG. The corresponding histograms of distance error from the three methods are given in Figure 5.3(d), (e) and (f), respectively. The mean distance errors were 1.19\( mm \), 0.68\( mm \) and 0.58\( mm \) from OrthSNIG, PerSNILog and PerSNIG, respectively. As is shown, the distance error from OrthSNIG was larger than either PerSNILog or PerSNIG, which justified the necessity of applying perspective SNI instead of orthographic SNI. As can be also observed, the distance error from PerSNILog was larger
than PerSNIG. This was especially true at pixels where discontinuities occurred. The larger error was because the smoothening effect of logarithm function resulted in over-smoothening the surface by minimization in the log gradient space.

The second numerical simulation aims to investigate the noise robustness of the three SNI methods. The experimental conditions were the same as the first numerical simulation except that additive white Gaussian noise was added onto the input surface normals in the gradient space. The standard deviation varied from 0 to 10% of the data amplitude with 1% in interval. Monte-Carlo simulation of 100 repeated tests were performed under each noise level. The mean distance error was used for evaluation. Figure 5.4 shows the noise sensitivity of OrthSNIG, PerSNILog and PerSNIG. OrthSNIG and PerSNIG exhibited consistent performances among the different noise levels, while PerSNILog was more sensitive to noise. The noisy discontinuous pixels contributed the most to PerSNILog’s noise sensitivity.
Figure 5.4: Noise sensitivity of OrthSNIG, PerSNILog and PerSNIG

The third numerical simulation aims to investigate the speed of the three SNI methods. The experimental conditions were the same as the first numerical simulation except that four different image resolutions were used to test the computation time. 20 repeated tests were conducted for each image resolution using a particular SNI method and the mean computation time was recorded for evaluation. The results are shown in Table 5.2. As shown from the results, the computation time of OrthSNIG and PerSNILog were similar and they were around 1.5 times faster than PerSNIG. OrthSNIG and PerSNILog, in essence, were solving the same type of minimization problem and therefore, had similar computational cost. The slower speed of PerSNIG was due to its more complicated computation in the
gradient space compared with PerSNILog computed in the log gradient space.

In summary, perspective SNI outperforms orthographic SNI in terms of accuracy. Perspective SNI in the log gradient space (PerSNILog) has similar computational cost as orthographic SNI, while PerSNILog is more sensitive to discontinuities than PerSNIG. PerSNIG could provide better accuracy than PerSNILog in the presence of discontinuities at the cost of heavier computations.

5.4 Conclusions and Future Works

This chapter has proposed a new method that generalized SNI dealing with perspective distortion. The proposed perspective SNI method is PS-independent, making the image irradiance equation being inverted during PS in a simple form. The proposed method performs SNI on the image plane instead of on the target surface unlike past methods, which ensures the integration is performed on a regular grid. The perspective SNI problem has been formulated in analogy to its orthographic counterpart, which makes the proposed method adaptable to most past SNI methods. From results of numerical simulations, perspective SNI outperformed orthographic SNI in terms of accuracy. The computational cost of perspective SNI in the log gradient space was similar to that of orthographic SNI owing to the derived analogy. The problem of discontinuity-sensitivity posed by the perspective SNI in the log gradient space was also noticed. Performing perspective SNI in the gradient space instead of in the log gradient space could provide better solution in the presence of discontinuity at the cost of more computation time.

This chapter has demonstrated the first set of results, whilst much more researches must be conducted, especially on the discontinuity issue on the proposed perspective SNI method. Investigations of proper regularization methods are relevant.
Chapter 6

Micro-scale Shape Reconstruction System Using Photometric Stereo

This chapter presents the design and calibration of the micro-scale shape reconstruction system using photometric stereo. This system is stationary and originally designed to measure pavement microtexture, while its applicability can be generalized to a wider range of surfaces. Throughout this chapter, measurement of the pavement microtexture is posed as an illustrated background for the micro-scale shape reconstruction problem. The proposed system mainly consists of a color digital single-lens reflex (DSLR) camera with a macro lens, multiple LEDs, a control circuit and a cover. The LEDs are sequentially turned on and the camera captures one image at a fixed position with one LED lit at a time. The camera and the LEDs are synchronized using the control circuit. The cover is used to block ambient light to make the lighting conditions fully controllable. Using images of the same target surface area under different illuminations, photometric stereo (PS) and surface normal integration (SNI) are sequentially performed to reconstruct the surface. The developed system has several advantages over the other existing systems. Firstly, PS with SNI achieves
pixel-resolution and since a high-resolution DSLR camera is used, very dense measurement is obtained. Secondly, because of its area-scanning nature, fast on-site data acquisition is feasible. Thirdly, the system using color PS with SNI is less sensitive to specularities and shadows compared with most optical-based methods since images under diverse lighting conditions provide more cues in detecting and exploiting information from them. Fourthly, since more parameters in the image photometry formation model other than surface orientations can be achieved, more functionalities, such as specular removal, intrinsic image decomposition and digital relighting, can be performed that are well-beyond the capability of most other methods. Last but not least, due to the simplicity of the hardware, the proposed system can be made compact at low cost.

This chapter is organized as follows. The next section formulates the problem of microtexture road profiling using color PS with SNI. Section II presents the proposed system design and Section III gives the calibration methods. Performance evaluations on individual calibration methods and the whole system are demonstrated in Section IV. Conclusions and future works are summarized in the last section.

6.1 Problem Formulation

6.1.1 Microtexture Road Profiling Requirement

Figure 6.1 is the diagram to define the microtexture road profiling problem. A road surface with area of $L_x \times L_y$ and surface deviation of $L_z$ is to be measured in the form of 3D point cloud. The spatial resolution along $X$ and $Y$ are respectively denoted as $\Delta X$ and $\Delta Y$. The vertical resolution is denoted as $\Delta Z$. Road profiling is to measure the pavement deviation from a planar surface, called the reference plane. This plane is, therefore, defined as the
Table 6.1: Requirement for microtexture road profiling system

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal range</td>
<td>$L_x$</td>
<td>$\geq$</td>
<td>20 mm</td>
</tr>
<tr>
<td>Transverse range</td>
<td>$L_y$</td>
<td>$\geq$</td>
<td>20 mm</td>
</tr>
<tr>
<td>Elevation range</td>
<td>$L_z$</td>
<td>$\geq$</td>
<td>10 mm</td>
</tr>
<tr>
<td>Longitudinal resolution</td>
<td>$\Delta X$</td>
<td>$\leq$</td>
<td>0.2 mm</td>
</tr>
<tr>
<td>Transverse resolution</td>
<td>$\Delta Y$</td>
<td>$\leq$</td>
<td>0.2 mm</td>
</tr>
<tr>
<td>Elevation resolution</td>
<td>$\Delta Z$</td>
<td>$\leq$</td>
<td>0.01 mm</td>
</tr>
</tbody>
</table>

$X-Y$ plane of the world coordinate $\{W\}$ whose origin is arbitrary on the plane. According to [85], the requirement for a microtexture road profiling system is tabulated in Table 6.1.

Figure 6.1: Microtexture road profiling problem

6.1.2 Flows of Road Profiling Using Color PS and SNI

Figure 6.2 shows the schematic diagram of the hardware components of the system and general flows of road profiling using color PS and SNI. The system contains two major hardware components, an RGB camera and $N$ directional point sources whose positions
are priorly known. A cover is additionally included to block the ambient light so that the lighting condition is fully controllable. PS-based road profiling requires two major steps, PS and SNI. Color PS is first conducted that requires the inputs of $N$ color images of the road surface illuminated under the $N$ different sources with one source lit at a time. Given these color images and the known light directions determined from the known light positions, color PS estimates unit surface normals pixel-wise. These unit surface normals are then integrated through SNI to reconstruct the 3D point cloud of the surface in the second step. Mathematically, the first step of color PS aims to estimate the unit surface normal, $\mathbf{n}_{i,j}^k$, given the color image irradiances at pixel $(i, j)$, $\{e_{i,j}^k\}_{k \in [1,N]}$, under the unit light directions, $\{l_{i,j}^k\}_{k \in [1,N]}$. In the second step of SNI, the 3D point $\{W\}_{X_{i,j}}$ is obtained by integrating the estimated normal field.

![Figure 6.2: Hardware components and general flows of road profiling using color PS and SNI](image-url)
6.1.3 Light Direction Estimation from Light Position

Estimating light direction $\vec{l}_{i,j}^k$ from the light position $\vec{S}_k$ is crucial when light sources cannot be assumed as distant source. The given method below is valid when the surface deviation is small compared with the light-surface distance, $R_{i,j}^k$.

Let the reference plane equation in $\{C\}$ be written as:

$$A_r x + B_r y + C_r z + D_r = 0,$$

where $A_r = [1, 0, 0]^T (r_{w1} \times r_{w2})$, $B_r = [0, 1, 0]^T (r_{w1} \times r_{w2})$, $C_r = [0, 0, 1]^T (r_{w1} \times r_{w2})$, $D_r = -(r_{w1} \times r_{w2}) \cdot (t_w)$. $R_w = [r_{w1}, r_{w2}, r_{w3}]$ and $t_w$ are priorly known through camera geometrical calibration. The intersection point of ray $(i,j)$ with the reference plane, $\{C\} \vec{X}_{i,j}^r = [(C) X_{i,j}^r, (C) Y_{i,j}^r, (C) Z_{i,j}^r]^T$ is computed as:

$$\begin{cases}
(C) Z_{i,j}^r &= -\frac{D_r}{A_r \frac{i-o_x}{f_{mx}} + B_r \frac{i-o_y}{f_{my}} + C_r} \\
(C) X_{i,j}^r &= \frac{j - o_x (C)}{f_{mx}} Z_r \\
(C) Y_{i,j}^r &= \frac{i - o_y (C)}{f_{my}} Z_r
\end{cases}$$

The unit light direction on pixel $(i,j)$ for the $k^{th}$ illuminant in $\{C\}$ can be then estimated using:

$$\vec{l}_{i,j}^k = \frac{(C) \vec{S}_k - (C) \vec{X}_{i,j}^r}{\| (C) \vec{S}_k - (C) \vec{X}_{i,j}^r \|}.$$
6.2 System Design

6.2.1 Design objectives

The proposed system aims to measure road surface at microtexture scale and meets all the requirements given in Table 6.1. The system should be also designed as stationary fast, as categorized by [85], whose time on lane per single measurement is less than one minute. For on-site measurements, the system requires portability and easy-operation by a single person.

6.2.2 System Overview

Figure 6.3: Hardware components of the developed system for online data capturing

Figure 6.3 shows the 3D high-resolution microtexture road profiling system for online
data capturing. The system contains three major hardware components: an RGB camera with macro-lens, a circular pattern of LEDs and a control circuit. A cover (not shown in the figure) made of black fleece fabric is also used to block ambient light and absorb second illumination reflected by the road surface to make the lighting condition fully controllable when the system operates. A high resolution RGB camera with a macro telephoto lens faces perpendicularly to the target road surface of concern. The camera captures one color image with one LED lit at a time, which is synchronized by the control circuit. These color images are saved in an SD card and later off-line processed by a computer using color PS with SNI.

6.2.3 Optimal Illumination Configuration

PS is susceptible to a number of error sources, including the model error, the calibration error, and the measurement error. The model error is determined by how accurate the image formation model can faithfully characterize the image formation process, such as the proper usage of BRDFs. The calibration error is related to the capability in acquisition of PS-isolated parameters, such as light positions. The measurement error usually refers to the corrupting influence of noise on the input images. These error sources can be studied individually through simulations and different optimal illumination configurations can be derived based on their respective criteria. To limit the scope for the optimal illumination study, the objective is to find the optimal illumination settings that are most robust to image noise. For this purpose, the following assumptions or constraints are made:

- Calibration errors are negligible;
- Distant point source is assumed;
- The sources are distributed in the circular pattern as modelled in Section 4.1.2;
• The surface reflectance can be characterized using Lambertian BRDF;

• Distant viewer is assumed such that the imaging geometry can be simplified using orthographic projection instead of perspective projection;

• Input images are in gray-scale.

Under these assumptions and from equation (3.29), the corrected image irradiance is written as:

\[ \tilde{e}^{ij} = \tilde{k}_d^{ij} \left( \mathbf{I}_k \right)^T \tilde{n}^{ij}, \]  

(6.4)

where \( \mathbf{I}_k \) is the \( k^{th} \) light direction and the same for each pixel. Note that here, \( \mathbf{I}_k \) and \( \tilde{n}^{ij} \) are in \( \{W\} \).

Consider first that three images are captured under different non-coplanar light directions, which is the basic requirement for the Lambertian-based PS [19]. These images are contaminated by additive white Gaussian noise. Incorporating image noise and stacking equation (6.4) into the matrix form, equation (6.5) is written as:

\[
\begin{align*}
\tilde{e}^{ij} + \Delta \tilde{e}^{ij} &= [L] \tilde{n}^{ij}, \\
\tilde{e}^{ij} &= [L] n^{ij},
\end{align*}
\]

(6.5)

where \( \Delta \tilde{e}^{ij} = [\Delta \tilde{e}^{ij}_1, \Delta \tilde{e}^{ij}_2, \Delta \tilde{e}^{ij}_3]^T \) is the additive image noise vector. \( n^{ij} = \tilde{k}_d^{ij} \tilde{n}^{ij} \) and \( n^{ij} = \tilde{k}_d^{ij} \tilde{n}^{ij} \) are contaminated and uncontaminated scaled normal vectors, respectively.

For the choice of optimal light configuration, the difference between the contaminated and uncontaminated unit surface normal is to be minimized and the cost function is given by:

\[ \epsilon_i = \| \tilde{n}^{ij} - \tilde{n}^{ij} \|^2 = \| [L]^{-1} \Delta \tilde{e}^{ij} \|^2 = \frac{1}{(\det [L])^2} \| \text{adj} ([L]) \Delta \tilde{e}^{ij} \|^2, \]

(6.6)
where \( \det (\cdot) \) and \( \text{adj}(\cdot) \) denote the determinant and matrix adjugate, respectively. Maximizing \( \det [L] \) can minimize \( \epsilon_i \).

**Theorem 1.** The determinant of \( [L] = [\bar{l}_1, \bar{l}_2, \bar{l}_3]^T \), where \( \{\bar{l}_k\}_{k=1,2,3} \) are unit vectors, is maximized if and only if \( [L] \) is an orthogonal matrix.

**Proof.** Applying the relationship between determinant and triple product, \( \det ([L]) \) is written as:

\[
\det ([L]) = [\bar{l}_1, \bar{l}_2, \bar{l}_3]^T = \bar{l}_1 \cdot (\bar{l}_2 \times \bar{l}_3).
\]  

(6.7)

From the definition of cross product, \( \mathbf{n}_{23} = \bar{l}_2 \times \bar{l}_3 = \|ar{l}_2\|\|ar{l}_3\| \sin \psi_{23} \bar{n}_{23} \), where \( \psi_{23} \) is the angle between \( \bar{l}_2 \) and \( \bar{l}_3 \), and \( \mathbf{n}_{23} \) and \( \bar{n}_{23} \) are scaled normal vector and unit normal vector of the plane spanned by \( \bar{l}_2 \) and \( \bar{l}_3 \), respectively. Therefore, equation (6.7) is modified as:

\[
\det ([L]) = [\bar{l}_1, \bar{l}_2, \bar{l}_3]^T = \bar{l}_1 \cdot \mathbf{n}_{23}.
\]  

(6.8)

Since \( \{\bar{l}_k\}_{k\in[1,3]} \) are unit vectors, \( \det ([L]) = \sin \psi_{23} \cos \psi_{123} \), where \( \psi_{123} \) is the angle between \( \bar{l}_1 \) and \( \bar{n}_{23} \). The maximum value of \( \det ([L]) \) is 1 and this is true if and only if \( \bar{l}_1 \perp \bar{l}_2 \perp \bar{l}_3 \). Under this condition, \( [L] \) is an orthogonal matrix.

Since the lights are configured in a circular pattern as given by equation (4.5) and (4.6), the \( k^{th} \) light direction in \( \{W\} \) is written as:

\[
\begin{align*}
\bar{l}_{x,k} &= \sin \eta \cos \phi_k, \\
\bar{l}_{y,k} &= \sin \eta \sin \phi_k, \\
\bar{l}_{z,k} &= \cos \eta,
\end{align*}
\]  

(6.9)

where \( \{\phi_k\}_{k\in[1,3]} \) are uniformly distributed in \( 360^\circ \). Therefore, the three light directions are perpendicular if and only if \( \eta = 54.74^\circ \). Under this condition, the optimal light configuration
with three lights is obtained and this setting is the most robust to image noise.

Next, the optimal light configuration with \( N \) lights is considered. Similar to the derivation of the cost function given by equation (6.6), the cost function for the \( N \)-light case is given by:

\[
\epsilon_i = \| \tilde{n}_{i;j} - \tilde{n}_{i;j} \|^2 = \| [L]^\dagger \Delta \tilde{e}_{i;j} \|^2 = \frac{1}{\det [L]^T [L]} \| \text{adj} \left( [L]^T [L] \right) [L]^T \Delta \tilde{e}_{i;j} \|^2, \tag{6.10}
\]

where \((\cdot)^\dagger\) represents the pseudo inverse. Maximizing \( \det [L]^T [L] \) can minimize the cost function.

**Theorem 2.** If the light directions are configured in the circular pattern as given by equation (6.9), the determinant of \( \det [L]^T [L] \) is maximized if and only if \( \eta = 54.74^\circ \).

**Proof.** After algebraic manipulation, \( \det [L]^T [L] \) is written as:

\[
\det [L]^T [L] = \frac{N^3}{4} \sin^4 \eta \cos^2 \eta. \tag{6.11}
\]

Compute the first derivative of \( \det [L]^T [L] \) with respect to \( \eta \) and let it be zero, \( \eta = \arctan \sqrt{2} \approx 54.74^\circ \). The second derivative at \( \eta = 54.74^\circ \) is smaller than zero. Therefore, \( \det [L]^T [L] \) is maximized at \( \eta = 54.74^\circ \). 

The next problem at hand is how the number of lights being used affects the error of surface orientation estimation. Since \( \epsilon_i \) can be decomposed using Eigen decomposition as:

\[
\epsilon_i = \Delta \tilde{e}_{i;j} [Q_i] [A_i] [Q_i]^T \Delta \tilde{e}_{i;j}, \tag{6.12}
\]
where the columns of $[Q_i]$ are the eigenvectors of $\left([L]^\dagger\right)^T[L]^\dagger$ and $[\Lambda_i]$ is a diagonal matrix containing the $N$ eigenvalues of $\left([L]^\dagger\right)^T[L]^\dagger$. Only three of these eigenvalues are nonzero and they are equal to $\frac{3}{N}$ under $\eta = 54.74^\circ$. Without losing the generality, suppose the first three eigenvalues are non-zero. Therefore, $\epsilon_i = \frac{3}{N} \sum_{k=1}^3 \| (q_k)^T \Delta \hat{e}_{i;j} \|^2$, where $q_k$ is the $k^{th}$ column of $[Q_i]$. This relationship suggests that $\epsilon_i$ is inversely proportional to $N$, which implies that the error of surface orientation estimation can be reduced through adding the number of lights.

From the above analyses, when $\eta = 54.74^\circ$, the Lambertian-based PS algorithm is the most robust to image noise. The surface orientation estimation can be enhanced through adding the number of lights. However, the shadow impacts in the previous analyses are not considered. Shadows are important outliers in the PS methods since in the shadow region, no shading information is contained theoretically without considering second illuminations. It is easy to prove that the area of cast-shadow region is minimized when the illumination direction is the same as the viewer direction, which means $\eta = 0^\circ$. Therefore, consider both the performance of image-noise robustness and cast-shadow reduction, $\eta$ should be chosen between $0^\circ$ to $54.74^\circ$. From the numerical studies presented in Section 6.4, the system is finally designed to have eight LEDs distributed in the circular pattern with $\eta = 25^\circ$.

### 6.2.4 Design of Imaging System

A high resolution ($7360 \times 4912$) DSLR camera (Nikon D810) with a macro telephoto lens (200mm f/4D) is used as the imaging system. The high resolution camera aims to provide dense measurements of the road surface, while the macro lens is used to capture fine details. Assume the optical axis is perpendicular to the road surface, the proposed system is designed to cover an area of $L_x = 100mm$ and $L_y = 66.85mm$. Since the area of the image sensor is
35.9mm × 24mm, the desired magnification ratio, \( m_a \), of the imaging system is 0.359. From the lens specs, the working distance is around 783mm, i.e. \( d_b = 783\text{mm} \) (see \( d_b \) from Figure 4.2). Under this working distance, the longitudinal and transverse resolution, \( \Delta X \) and \( \Delta Y \), are the same and equal to 13.6\( \mu \text{m} \).

In order to maintain the depth of field (DoF) beyond the required measurement elevation range, the f-number is chosen as \( f/22 \). The DoF can be estimated using:

\[
\text{DoF} = \frac{2F_a c_o (1 + m_a)}{(m_a^2 - (F_a c_o / f)^2)},
\]

(6.13)

where \( F_a \) is the f-number and \( c_o \) is the circle of confusion. With standard \( c_o \) for a full-frame image, the DoF for the proposed system is estimated to be 13.456mm. Therefore, the elevation range of the system, \( L_z \), is 13.456mm.

Having chosen the f-number as \( f/22 \) and in order to capture well-exposed images, the combination of exposure time and ISO speed requires to be determined. High ISO speed can shorten the data capturing time, while it will introduce more noise to the images. In order to balance the speed and noise level, ISO speed of 200 with exposure time of 2.5 seconds is adopted. Therefore, the measurement time on lane for the proposed system is 20 seconds. 12-bit RAW images for each color channel without Gamma correction is utilized as raw measurement data. RAW images are used since its observed image intensity is proportional to the image irradiance.

### 6.2.5 Design of Control Circuit

Figure 6.4 shows the design and developed control circuit which is used to synchronize the camera and the LEDs. A micro-controller (Arduino) and a circuit with a digital multiplexer (74HC4067) are the major components. The camera is set at ‘BULB’ mode so that when
both camera control 1 and 2 are digital high, the camera shutter opens and by contrast, if one of the camera controls is set low, the shutter closes. Therefore, the exposure time is totally controllable by the micro-controller. Note that for each cycle, the camera shutter opens 100 ms before the LED lits and closes 100 ms after the LED turns off. The design of the timing is to reduce motion blurs due to the mechanical shutter vibration.

6.2.6 System Integration and Specification

With the chosen camera working distance of $d_b = 783 mm$ and light angle $\eta$ of $25^\circ$, $r$ and $\theta$ (See Figure 4.2) are designed as $440 mm$ and $25^\circ$, respectively. The distance from the plane determined by the circular LEDs and the reference plane is, therefore, $384.22 mm$. The size of the $5 mm$ LED is not comparable with this distance and hence, the LEDs can be regarded as directional sources. The overall size of the system are then bounded by a box of $1150 mm \times 406.4 mm \times 406.4 mm$ after development and the weight is $13.6 kg$. The estimated system specifications are listed in Table 6.2. The on-site data capturing time per single measurement is 20 seconds.
Table 6.2: Estimated system specifications

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal range</td>
<td>$L_x$</td>
<td>100</td>
<td>mm</td>
</tr>
<tr>
<td>Transverse range</td>
<td>$L_y$</td>
<td>66.85</td>
<td>mm</td>
</tr>
<tr>
<td>Elevation range</td>
<td>$L_z$</td>
<td>13.46</td>
<td>mm</td>
</tr>
<tr>
<td>Longitudinal resolution</td>
<td>$\Delta X$</td>
<td>0.0136</td>
<td>mm</td>
</tr>
<tr>
<td>Transverse resolution</td>
<td>$\Delta Y$</td>
<td>0.0136</td>
<td>mm</td>
</tr>
<tr>
<td>Elevation resolution</td>
<td>$\Delta Z$</td>
<td>0.0049</td>
<td>mm</td>
</tr>
</tbody>
</table>

6.3 Calibrations

A series of calibrations require to be performed after the system is developed. Five calibrations are performed, including camera geometrical calibration, camera radiometric calibration, light position calibration, light attenuation calibration and specular color calibration. Figure 6.5 gives the necessary calibration sequence since the subsequent sequences will use the results from their preliminary ones. In this section, these calibration methods are given.

![Calibration sequence diagram](image)
6.3.1 Camera Geometrical Calibration

The major purpose of the camera geometrical calibration is to determine the camera’s intrinsics, \( M_p^K \), and the camera’s relative pose to the reference plane, \( R_w \) and \( t_w \). It also aims to correct the lens distortion to make the pin-hole camera model valid. Standardized method popularized by Zhang [86] was adopted with the usage of traditional lens model incorporating radial and tangential distortions [87]. The Caltech calibration toolbox [88] aided this calibration purpose for analyses.

6.3.2 Camera Radiometric Calibration

The function converting image intensity to image irradiance is referred to as the camera radiometric function (CRF). The process in identifying the CRF is camera radiometric calibration. RAW images are used by the proposed system and ideally, the observed image intensity is directly proportional to the image irradiance. If RAW images are not used, due to the nonlinearities introduced by most cameras, such as for the purpose of dynamic range compression, the CRF, denoted as \( F_c \), maps from image irradiance to image intensity as \( I^{i,j}_c = F_c(e^{i,j}_c) \). For color cameras, each color channel has its own radiometric response as \( \{F_c\}_{c=R,G,B} \). Here, a general CRF calibration method is proposed. This method is similar to [89] with several modifications.

The proposed method utilizes \( N_r \) images for the same scene under different exposure values, denoted as \( \{E_{v,u}\}_{u \in [1,N_r]} \). These images are first sorted in the descending order based on the exposure values that are priorly known. Assume the scene being exposed is static. Suppose the mapping from image irradiance to image intensity is nonlinear, which can be
characterized by a high-order polynomial function as:

\[ e^{i,j}_c = F^{-1}_c(I^{i,j}_c) = \sum_{n=0}^{V} c_n(I^{i,j}_c)^n, \]  

(6.14)

where \( V \) is the maximum order of polynomial in characterizing \( F_c \) and \( \{c_n\}_{n \in [0,V], c=\{R,G,B\}} \) is the corresponding polynomial coefficient for each color channel to be identified. Note that in equation (3.28), \( F_c \) was assumed to be 1 since image intensities from RAW images were used. The nonlinear mapping given here is to generalize the problem when RAW images are not available. \( F_c \) is monotonic since image irradiances and image intensities are positive correlated. Therefore, the nonlinear mapping is one-to-one and the inverse of \( F_c \) exists. Considering equation (3.24) and generalizing equation (3.28) with \( F_c \), the color image irradiance, \( \{e^{i,j}_c\}_{c=\{R,G,B\}} \), is rewritten in terms of color image intensity, \( \{I^{i,j}_c\}_{c=\{R,G,B\}} \), as:

\[ e^{i,j}_c = L^{i,j}_r \kappa^{i,j}_e E_v = F^{-1}_c(I^{i,j}_c), \]  

(6.15)

where \( \{L^{i,j}_r\}_{c=\{R,G,B\}} \) is the scene radiance, \( \kappa^{i,j}_e = \kappa_{iso} \cos^4 \theta_o^i / f^2 \) is a local constant scaling and \( E_v = \pi d^2 t_e / 4 \) is the exposure value. The exposure ratio between consecutive images is then:

\[ R^{i,j}_{u+1,u} = \frac{\sum_{n=0}^{V} c_n(I^{i,j}_{c,u+1})^n}{\sum_{n=0}^{V} c_n(I^{i,j}_{c,u})^n} = \frac{e^{i,j}_{c,u+1}}{e^{i,j}_{c,u}} = \frac{E_{v,u+1}}{E_{v,u}}, \]  

(6.16)

where \( R^{i,j}_{u+1,u} (< 1) \) is the exposure ratio. The camera radiometric calibration can be then formulated into a least square problem and the objective function is given by:

\[ \epsilon_c = \sum_{u=1}^{N_v} \sum_{i=1}^{H} \sum_{j=1}^{W} \left[ \left( \sum_{n=0}^{V} c_n(I^{i,j}_{c,u+1})^n - \frac{P^{i,j}_{u+1,u}}{L^{i,j}_{c,u}} \sum_{n=0}^{V} c_n(I^{i,j}_{c,u})^n \right)^2 \right] \rightarrow \min_{\{c_n\}_{n \in [0,V]}}, \]  

(6.17)

where \( \epsilon_c \) is the sum of squared residuals. Each residual term is scaled by \( 1/L^{i,j}_{c,u} \) to avoid overemphasizing the impacts of bright pixel residuals, which makes the cost function different
from that given in [89]. In order to constrain \( F_c^{-1}(1) = 1 \), \( c_V = 1 - \sum_{n=0}^{V-1} c_n \). By taking the partial derivative of \( \epsilon_c \) with respect to each \( c_n \) and let them be zero, \( \{c_n\}_{n \in [1, V]} \) can be derived by solving \( V \) number of linear equations. The CRF is estimated individually for each color channel.

Several issues need to be addressed by using this method. Firstly, the images under multiple exposures should cover sufficient exposure range containing both under-exposed and over-exposed images for the purpose of estimating the CRF in the full dynamic range of the camera. Secondly, dark and saturated pixels should be discarded in the optimization since they are out of the camera’s dynamic range and violate equation (6.16). Thirdly, the best way to vary the exposure values is to alter the shutter speed while maintain the aperture setting. This is because the change of aperture will also affect the blur level of the image. Last but not least, in order to maintain the monotonicity of \( F_c \), if \( I_{c,u}^{i,j} \) is not greater than \( I_{c,u+1}^{i,j} \), the term comprised of this pair of observations in equation (6.17) is discarded as outlier since they violate the constraint of \( \mathcal{R}_{u+1,u}^{i,j} < 1 \).

### 6.3.3 Light Position Calibration

Light positions are crucial parameters for PS and significantly affect the measurement accuracy. Due to imperfect manufacturing and assembly, the light positions may not be located exactly as designed. Therefore, there is necessity to calibrate the light positions after system installation. The proposed light position calibration method contains two major steps: light direction calibration and light position determination. The closest work is [90] since the same geometric cue is used, whereas the calibration object and method are different.

Figure 6.6 shows the schematic diagram to explain the light direction calibration method. A chrome steel ball fixed with a black matte support is adopted for the calibration purpose.
The radius of the ball and the distance from the bottom of the support to the bottom of the ball are known as $R_b$ and $d_s$, respectively. An image for the ball with the support is captured under the $k^{th}$ illumination. The ball is projected onto the image plane in a near-circle shape whose center locates at $\{P\}x^*_b$. $\{P\}x^*_b$ corresponds to the zenith point of the hemisphere facing the camera, $\{C\}\bar{X}_z$. The brightest spot mirroring the $k^{th}$ light appears at $\{C\}\bar{X}_{s,k}$ and projects onto the image plane at $\{P\}x^*_s$. $\{P\}x^*_s$ and $\{P\}x^*_s$ can be detected from the image. The major objective in this light direction calibration step is to derive the location of the ball center and the brightest spot under the $k^{th}$ illumination, $\{C\}\bar{X}_b$ and $\{C\}\bar{X}_{s,k}$. These two points are concerned since the $k^{th}$ light position must locate on the 3D line determined by the two points.

The ball center location, $\{C\}\bar{X}_b$, is first determined. The plane passing through the ball center and parallel to the reference plane is referred to as the equator plane. The equation for the equator plane in $\{C\}$ is first derived and given by:

$$A_e x + B_e y + C_e z + D_e = 0,$$

(6.18)
where \( A_e = A_r, B_e = B_r, C_e = C_r, D_e = -(r_{w1} \times r_{w2}) \cdot (r_{w3} (R_b + d_s) + t_w) \). Next, the ball center is determined as the intersection point between the equator plane and the ray passing both the camera center and the zenith point. The ball center, \( \{C\} \tilde{X}_b = \begin{bmatrix} \{C\} X_b, \{C\} Y_b, \{C\} Z_b \end{bmatrix}^T \), is given by:

\[
\begin{align*}
\{C\} Z_b &= \frac{-D_e}{A_e \frac{\{P\} x^*_s - o_x}{f_{m_x}} + B_e \frac{\{P\} y^*_s - o_y}{f_{m_y}} + C_e} \\
\{C\} X_b &= \frac{\{P\} x^*_s - o_x \{C\} Z_b}{f_{m_x}} \\
\{C\} Y_b &= \frac{\{P\} y^*_s - o_y \{C\} Z_b}{f_{m_y}}
\end{align*}
\]

where \( \{P\} x^*_s = \begin{bmatrix} \{P\} x^*_s, \{P\} y^*_s, 1 \end{bmatrix}^T \) is obtained from the image. The brightest spot location under the \( k^{th} \) illumination, \( \{C\} \tilde{X}_{s,k} = \begin{bmatrix} \{C\} X_{s,k}, \{C\} Y_{s,k}, \{C\} Z_{s,k} \end{bmatrix}^T \), is then determined by finding the intersection point between the ray passing through the brightest spot and the hemisphere of the ball visible by the camera. The solution is given by:

\[
\begin{align*}
\{C\} Z_{s,k} &= \frac{l_{s1} \{C\} X_b + l_{s2} \{C\} Y_b + \{C\} Z_b - \sqrt{\Delta_s}}{l_{s1}^2 + l_{s2}^2 + 1} \\
\{C\} X_{s,k} &= l_{s1} \{C\} Z_{s,k} \\
\{C\} Y_{s,k} &= l_{s2} \{C\} Z_{s,k} \\
\Delta_s &= (l_{s1} \{C\} X_b + l_{s2} \{C\} Y_b + \{C\} Z_b)^2 - (l_{s1}^2 + l_{s2}^2 + 1) \left( \{C\} X_b^2 + \{C\} Y_b^2 + \{C\} Z_b^2 - R_b^2 \right) \\
l_{s1} &= \frac{\{P\} x^{s,k}_s - o_x}{f_{m_x}} \\
l_{s2} &= \frac{\{P\} y^{s,k}_s - o_y}{f_{m_y}}
\end{align*}
\]

where \( \{P\} x^{s,k}_s = \begin{bmatrix} \{P\} x^{s,k}_s, \{P\} y^{s,k}_s, 1 \end{bmatrix}^T \) is detected from the image using the centroid of the bright spot. Having computed \( \{C\} \tilde{X}_{s,k} \) and \( \{C\} \tilde{X}_b \), a 3D line across \( \{C\} \tilde{X}_b \) and \( \{C\} \tilde{X}_{s,k} \) is determined which also passes through the \( k^{th} \) light position, \( \{C\} \tilde{S}_k \).
The second step is to determine the light positions using triangulation. By placing the ball with the support at $M$ different locations on the reference plane within the FoV and applying the first step, $M$ ball center and brightest spot locations under the $k^{th}$ illumination are obtained and denoted as $\{\{^{(C)} \bar{X}_{b,m}\}_{m \in [1,M]} \}$ and $\{\{^{(C)} \bar{X}_{s,k,m}\}_{m \in [1,M]}\}$, respectively. Figure 6.7 shows the schematic diagram for explanation of the light position determination for the case where $M = 2$. $\{\bar{s}_{k,m}\}_{m \in [1,M]}$ are unit vectors representing the 3D line directions, where $\bar{s}_{k,m} = \left(\{^{(C)} \bar{X}_{s,k,m} - \{^{(C)} \bar{X}_{b,m}\}\right) / \|\{^{(C)} \bar{X}_{s,k,m} - \{^{(C)} \bar{X}_{b,m}\}\|$. Determining the light position for the $k^{th}$ light, $\{^{(C)} \bar{S}_{k}\}$, is equivalent to find the intersection point of the $M$ 3D lines. This intersection point is defined as the point whose distances to each 3D line are minimized. The solution is given by:

$$\{^{(C)} \bar{S}_{k}\} = \left(\sum_{m=1}^{M} \left[ \bar{s}_{k,m} (\bar{s}_{k,m})^T - I_{3 \times 3}\right]\right)^{-1} \left(\sum_{m=1}^{M} \left[ \bar{s}_{k,m} (\bar{s}_{k,m})^T - I_{3 \times 3}\right] \{^{(C)} \bar{X}_{b,m}\}\right), \quad (6.21)$$

where $I_{3 \times 3}$ is an identity matrix with rank 3. The distance from the intersection point to
each 3D line, \( d_{k,m} \), is given by:

\[
d_{k,m} = \frac{\| (\{C\} \tilde{S}_k - \{C\} \tilde{X}_{h,m}) \times (\{C\} \tilde{S}_k - \{C\} \tilde{X}_{s,k,m}) \|}{\| \{C\} \tilde{X}_{h,m} - \{C\} \tilde{X}_{s,k,m} \|}.
\]

(6.22)

Ideally, \( \{d_{k,m}\}_{m=1}^M \) are near-zero and it becomes a direct metric to evaluate the performance of the light position calibration.

### 6.3.4 Light Attenuation Calibration

\( \{\kappa_{i,j}^{n,k}\}_{k=1}^N \) which encodes the light attenuations and off-axis illumination requires to be calibrated and normalized before PS. Figure 6.8 shows the schematic diagram for explaining light strength calibration. A flat white balance card with matte-finish is used for the calibration purpose. This white balance card is sufficiently big to cover the camera’s FoV and its thickness, \( d_w \), is also known. The white balance card is placed on top of the reference plane and in the center of the FoV. The top surface of the white balance card visible by the RGB camera, referred to as the top plane, is illuminated under the same lighting conditions as PS and \( N \) images are, therefore, captured. Using these \( N \) images, \( \kappa_{i,j}^{n,k} \) can be calibrated.

Assume both sides of the white balance card are flat and parallel to each other. When placing the card onto the reference plane, the top plane equation is given by:

\[
A_t x + B_t y + C_t z + D_t = 0,
\]

(6.23)

where \( A_t = A_r, B_t = B_r, C_t = C_r, D_r = - (r_{w1} \times r_{w2}) \cdot (r_{w3} d_w + t_w) \). The 3D point visible by pixel \((i,j)\) on the top plane, \( \{C\} \tilde{X}_{t}^{i,j} \) can be computed similar to equation (6.3). The unit surface normal of the plane in \( \{C\} \) equals to \( \tilde{n}_t = (r_{w1} \times r_{w2}) / \| r_{w1} \times r_{w2} \| \). For each pixel corresponding surface patch, \( \tilde{n}^{i,j} \) are the
Figure 6.8: Schematic diagram for explanation of the light strength calibration

same and $\mathbf{n}^{i,j} = \mathbf{n}_t$. Having $(C)\mathbf{X}_t^{i,j}$ and the calibrated light position, $(C)\mathbf{S}_k^{i,j}$, the unit light direction at $(C)\mathbf{X}_t^{i,j}$ under the $k^{th}$ illumination is computed by:

$$
\mathbf{l}_k^{i,j} = \frac{(C)\mathbf{S}_k^{i,j} - (C)\mathbf{X}_t^{i,j}}{\| (C)\mathbf{S}_k^{i,j} - (C)\mathbf{X}_t^{i,j} \|}.
$$

(6.24)

With the obtained $\mathbf{l}_k^{i,j}$ and $\mathbf{n}^{i,j}$, the surface irradiance is computed using:

$$
E_k^{i,j} = (\mathbf{n}^{i,j})^T \mathbf{l}_k^{i,j},
$$

(6.25)

where $E_k^{i,j}$ is the surface irradiance of pixel $(i,j)$ under the $k^{th}$ illumination.

Let the color image irradiance observed at pixel $(i,j)$ under the $k^{th}$ illumination be
$e_{i,t,k}^{i,j}$. Assume the reflectance of the top plane is Lambertian, the unit diffuse color for pixel $(i,j)$, $\bar{d}_{t,RGB}^{i,j}$, can be then estimated using the principal component of $[E_i^{i,j*}]^T [E_i^{i,j*}]$, where $[E_i^{i,j*}] = [e_{t,1}^{i,j*}, e_{t,2}^{i,j*}, ..., e_{t,N}^{i,j*}]^T$. $\kappa_{n,k}^{i,j}$ scaled by $\hat{k}_d$ under the $k^{th}$ illumination can be derived from equation (3.29) without specular component as:

$$
\hat{k}_d^{i,j} \kappa_{n,k}^{i,j} = (e_{t,k}^{i,j*})^T \bar{d}_{t,RGB}^{i,j} / E_i^{i,j*}.
$$

(6.26)

Since the roughness of the top plane cannot be ignored at microtexture scale, it is then necessary to filter out the high-frequency components. A low-pass filter, such as a moving average filter, is, therefore, applied on the scaled $\kappa_{n,k}^{i,j}$ spatially. Note that the scaling $\hat{k}_d^{i,j}$ is the same for the $N$ lights on pixel $(i,j)$ and therefore, it can be assumed to be 1 without losing the generality to use $\kappa_{n,k}^{i,j}$ representing the relative values.

Though the recovered $\kappa_{n,k}^{i,j}$ can be applied directly to normalize $e_{RGB}^{i,j}$ to $\hat{e}_{RGB}^{i,j}$, these parameters consume huge storage spaces. Therefore, the next objective is to compress $\{\kappa_{n,k}^{i,j}\}_{i \in [1,H], j \in [1,W], k \in [1,N]}$ into a few number of parameters while keeping the characteristics. Based on equation (3.32), $\kappa_{n,k}^{i,j}$ is modelled as:

$$
\kappa_{n,k}^{i,j} = \kappa_{a,k} \cos \chi_k \theta_{t,k}^{i,j} \kappa_{b,k}^{i,j},
$$

(6.27)

where $\{\kappa_{a,k}\}_{k \in [1,N]}$ and $\{\chi_k\}_{k \in [1,N]}$ are light attenuation parameters to be determined for each source. $\theta_{t,k}^{i,j} = \arccos \left( (\bar{1}_k^{i,j})^T \bar{g}_k \right)$ where the $k^{th}$ principle light direction, $\bar{g}_k$, can be determined using the unit vector from the brightest surface patch to the $k^{th}$ source. $\kappa_{b,k}^{i,j}$ is known and computed by:

$$
\kappa_{b,k}^{i,j} = \pi / 4 \left( d_{i,j} / f \right)^2 \cos^4 \theta_o^{i,j} / (P_i^{i,j})^2,
$$

(6.28)

where $\cos \theta_o^{i,j} = (\bar{v}^{i,j})^T [0, 0, -1]^T$. Transforming equation (6.27) into the logarithm domain
and after algebraic manipulation, the following equation is derived:

$$\left( \ln \cos \theta_{t,k} \right) \chi_k + \ln (\kappa_{a,k}) = \ln \kappa_{n,k}^{i;j} - \ln \kappa_{b,k}^{i;j}. \quad (6.29)$$

A linear least squares problem can be formulated to determine $\chi_k$ and $\ln (\kappa_{a,k})$. The cost function is given by:

$$\epsilon_{a,k} = \sum_{i=1}^{H} \sum_{j=1}^{W} \left[ \left( \ln \cos \theta_{t,k} \right) \chi_k + \ln (\kappa_{a,k}) - \ln \kappa_{n,k}^{i;j} + \ln \kappa_{b,k}^{i;j} \right]^2 \rightarrow \min_{\chi_k, \ln (\kappa_{a,k})}. \quad (6.30)$$

The optimal solution is derived as:

$$\begin{bmatrix} \chi_k \\ \ln \kappa_{a,k} \end{bmatrix} = \left( A_i^{i,j} \right) ^T \left( A_i^{i,j} \right) ^{-1} \left( A_i^{i,j} \right) ^T \left( c_k^{i;j} - b_k^{i;j} \right), \quad (6.31)$$

where $[A_i^{i,j}] = [a_k^{i,j}, 1_{HW}]$ and $a_k^{i,j} = [\ln \cos \theta_{t,k}^1, \ln \cos \theta_{t,k}^2, \ldots, \ln \cos \theta_{t,k}^{HW}]^T$. $c_k^{i;j}$ and $b_k^{i;j}$ are $[\ln \kappa_{n,k}^1, \ln \kappa_{n,k}^2, \ldots, \ln \kappa_{n,k}^{HW}]^T$ and $[\ln \kappa_{b,k}^1, \ln \kappa_{b,k}^2, \ldots, \ln \kappa_{b,k}^{HW}]^T$, respectively. The obtained $\chi_k$ and $\kappa_{a,k}$ are, therefore, used to characterize $\{\kappa_{n,k}^{i;j}\}_{i \in [1,H], j \in [1,W]}$.

### 6.3.5 Specular Color Calibration

Accurately estimating the specular color, $\mathbf{s}_{RGB}$, is the cornerstone to accomplish the proposed DRM-based color PS method. $\mathbf{s}_{RGB}$ is a function of the light SPD and the camera spectral sensitivity. The light SPD can be measured using a spectrometer or obtained from the manufacturing nominal values. The camera spectral sensitivity can be measured using method given in [91]. Using both the light SPD and the camera spectral sensitivity, the specular color can be computed using equation (3.30). However, neither the light SPD nor the camera spectral sensitivity is necessary. In addition, calibrating the camera’s spectral
sensitivity follows a tedious procedure requiring expensive hardware, making this method not flexible enough when an alternative camera needs to be used. Therefore, a simple fast method to calibrate specular color is presented.

A plane mirror is used for this purpose. Assume the mirror is perfect, it reflects all the wavelength in the visible spectrum and perfect interface reflection appears as characterized by equation (3.30). The experiment is set via tuning the orientation of the camera with fixed mirror and source such that the camera’s principal axis is tilted near the perfect specular angle of reflection. The camera is set with shorter exposure time to avoid sensor saturation. Excluding saturated pixels, $\mathbf{s}_{RGB}$ is computed as the principal component of $[\mathbf{E}_C]_k^T [\mathbf{E}_C]_k$, where $[\mathbf{E}_C]_k = [e_{k,1}, e_{k,2}, ..., e_{k,U}]^T$ is the matrix comprised of the remaining image irradiances under the $k^{th}$ illumination. All the sources are calibrated individually and their color consistency was checked as required by the proposed DRM-based color PS method.

6.4 Experimental Results

6.4.1 Experiments on Optimal Illumination Configuration

The first two experiments aim to verify the theoretical analyses results in Section 6.2.3. The purpose of the first experiment is to investigate the optimal zenith angle, $\eta$, that is the most robust to image noise. Additive white Gaussian noise with standard deviation of 0.05 was added. A hemisphere was adopted as the scene geometry since it can be viewed as uniform sampling of the possible surface orientations altitudinally and longitudinally. The normal map and one example of synthetic input image are shown in Figure 6.9. With the limitations from the data acquisition time and the constraints from the imaging system, eight LEDs were chosen to investigate the optimal $\eta$. These eight LEDs were distributed in the
circular pattern. The zenith angle of concern to be studied ranged from $10^\circ$ to $80^\circ$. Figure 6.10 gives the box-and-whisker plot of the error for surface orientation reconstruction using the metric given by equation (6.6). The red and black dots represent the mean and median values, respectively. The lower and upper bound of the blue bar respectively show the first and third quantile values. As is shown, the minimum mean error occurs at $\eta = 44^\circ$, while the minimum median error occurs at $\eta = 52^\circ$. The optimal $\eta$ was smaller than the theoretical value of $54.74^\circ$ since for larger $\eta$, more self-shadows occurred, making less number of valid observations used for PS. In addition, when $\eta$ was larger, more area was darker, resulting in lower SNR that finally led to bigger error. This fact also explained why the median error was smaller than the mean error. Another observation was that the error was at similar level from $\eta = 20^\circ$ to $\eta = 60^\circ$, which offered more flexibilities for the design purpose.

The second experiment aims to study how the error changes by adding more lights. The experimental condition was the same as the first experiment except that $\eta$ was fixed at $30^\circ$. The light number varied from 4 to 96. The box-and-whisker plot having the same meaning as the first experiment is shown in Figure 6.11. The green line fit the median error using $\epsilon_i = a_i/N$, where $a_i = 0.0365$. The R-square of the fitting was 0.9999, which implied
that $\epsilon_i$ was inversely proportional to $N$, as predicted and analyzed in theory. As is also shown, the error dropped sharply at the beginning when less lights were used and was nearly saturated at the end. For practical considerations, adding more lights will lead to longer data acquisition time. To balance the performance, eight lights were finally chosen for the proposed system.

The third experiment aims to investigate the optimal zenith angle for different material reflectances using the proposed DRM-based color PS method. Thirty material BRDFs, the same as used in Figure 4.7, were adopted to render the scene. The experimental settings were the same as the first experiment in Section 4.3.1 except that the zenith angle varied from $20^\circ$ to $60^\circ$. Figure 6.12 shows the box-and-whisker plot of the reconstructed angular error of surface orientations. Under each zenith angle, the mean, median, first quantile and
third quantile values for the 30 different materials were given. As is shown, the error started to increase after the zenith angle was over 30°. The mean error changed more significantly than the median error as the zenith angle was bigger since more areas were near the grazing angle and the Fresnel reflection was not modelled in the proposed method. From Figure 6.10 and 6.12, and also consider minimizing cast-shadow region, the zenith angle, \( \eta \), was finally chosen as 25° for the proposed system.
6.4.2 Evaluation on Calibration Methods

Evaluation on Camera Geometrical Calibration

This experiment aims to evaluate the camera geometrical calibration. The imaging system was first mounted on the system frame at its location of operation. Its focus was tuned at the ground plane with aperture set at minimum F-number to have shallowest DoF for more precise focus tuning. The camera geometrical calibration was then performed with the system operated horizontally rather than vertically. The imaging system, set at maximum F-number to achieve deepest DoF, captured a checker board planar pattern at 50 different poses. Using these images and by applying [86], parameters in the camera intrinsics and the lens distortion model were determined.
Figure 6.13 shows the results of the camera geometrical calibration. Figure 6.13(a) presents the relative poses between the camera and the checker board. The reprojection errors of the checker board corners are demonstrated in Figure 6.13(b). Figure 6.13(c), (d) and (e) respectively show the estimated radial, tangential and complete distortion. The reprojection errors were adopted to evaluate the performance of the camera geometrical calibration. The mean reprojection errors were respectively 2.533 and 2.529 pixels in the \( (P)x \) and \( (P)y \) directions, which implies the reprojection percent errors are 0.034% and 0.052%.
Evaluation on Camera Radiometric Calibration

This experiment aims to evaluate the proposed camera radiometric calibration method. Nine images for the same scene were captured under different exposure values by varying the shutter speed. This is simply achieved by using the bracketing (BKT) mode in Nikon cameras. These images are shown in Figure 6.14. Since these images were RAW images, the first order

![Figure 6.14: Captured images for the same scene under nine different shutter speed](image)

polynomial was used to estimate the CRF. The result is shown in Figure 6.15, which shows a perfect linear radiometric response. The residual was applied to evaluate the performance
Figure 6.15: Estimated camera radiometric response

of the estimation and defined as:

$$r_{u+1,u}^{i,j} = \left| \sum_{n=0}^{V} c_n(I_{c,u+1}^{i,j*})^n - R_{u+1,u}^{i,j} \sum_{n=0}^{V} c_n(I_{c,u}^{i,j*})^n \right|,$$

(6.32)

where $c_n$ was estimated. In order to compare the results with and without scaled by the image irradiance ($1/I_{c,n}^{i,j*}$) in equation (6.17), the histograms of residuals obtained from the two cost functions are shown in Figure 6.16. The cost function without scaled by image irradiance is referred to as the original cost function and equation (6.17) is referred to as the modified cost function. Figure 6.16(a), (b) and (c) show the histograms of residuals obtained through the original cost function for the red, green and blue channel, respectively. Figure 6.16(d), (e) and (f) demonstrate the respective histograms of residuals for the three color channels using the modified cost function. The mean residuals are listed in Table 6.3. From Figure 6.16 and Table 6.3, the residuals were significantly dropped by around 5 times by using the modified instead of the original cost function. Since 12-bit RAW images were
Figure 6.16: Histograms of residuals obtained through: (a) original cost function for the red channel; (b) original cost function for the green channel; (c) original cost function for the blue channel; (d) modified cost function for the red channel; (e) modified cost function for the green channel; (f) modified cost function for the blue channel

Table 6.3: Mean residuals during CRF estimation

<table>
<thead>
<tr>
<th>COST FUNCTION</th>
<th>RED</th>
<th>GREEN</th>
<th>BLUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>$12.0 \times 10^{-4}$</td>
<td>$6.0 \times 10^{-4}$</td>
<td>$8.1 \times 10^{-4}$</td>
</tr>
<tr>
<td>Modified</td>
<td>$2.9 \times 10^{-4}$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$1.6 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

used, the quantization error was $2.44 \times 10^{-4}$. By using the modified cost function, the mean residuals have achieved the level of quantization error and the CRF estimation was more accurate.

The linearity of the CRF was also checked through correlation, where the correlation
coefficients were 0.9933, 0.9944 and 0.9950 for the red, green and blue color channel, respectively. This fact further validated the linearity of the camera radiometric response by using RAW images and the results conformed to the estimated CRF.

Evaluation on Light Position Calibration

An experiment was conducted to evaluate the proposed light position calibration method. A G25 grade chrome steel bearing ball with $R_b = 6.35mm$ was fixed by a 3D printed support with $d_s = 3.81mm$. The ball with the support was adopted to calibrate the light positions and placed in the camera’s FoV as illustrated by Figure 6.17(a). The ball center was found by using the center of a manually labelled bounding box in the image with its edge tangential to the circle. Automatic circle detection method, such as using Hough transform, is not feasible because of the strong specularity of the ball. The brightest spot is detected by using the centroid of bright pixel locations within the bounding box. Bright pixels are determined using a threshold. Having determined the ball center and the brightest spot, the proposed light direction calibration method is performed. By capturing images for the ball at different locations in the FoV, the light positions were derived through triangulation. Theoretically, placing the ball on two different locations are sufficient to determine the light positions. In order to validate the performance, images for the ball at 15 different locations spread across the FoV were captured. The distances from the triangulated light positions to the 15 lines of light directions were used to evaluate the method performance. These distances were calculated using equation (6.22). Figure 6.17(b) shows these distances using box-and-whisker plot. The red and black dots represent the mean and median values, respectively. The lower and upper bound of the box indicate the maximum and minimum values. The mean and median distances for all the eight light positions are 2.75mm and 2.65mm, respectively. Since 5mm LEDs are used, the triangulated light positions fall mostly inside the LEDs, which have
demonstrated the validity of the proposed light position calibration method. The calibrated light positions in \( \{C\} \) was finally obtained as shown in Figure 6.17(c).

Figure 6.17: (a) Image used for light direction calibration under the first illuminant; (b) Box-and-whisker plot of distances from triangulated light positions to each 3D line (light direction); (c) The calibrated light positions in \( \{C\} \)

**Evaluation on Light Attenuation Calibration**

The first experiment aims to evaluate the proposed light attenuation calibration method. A white balance board with \( d_w \) of 4mm was used as the calibration object. The scaled \( \kappa_{n,k}^{i,j} \) was
obtained following the given procedures. When filtering the high frequency components, an average filter with window size of $150 \times 150$ was applied. $\chi_k$ and $\kappa_{a,k}$ were then achieved by solving equation (6.31). The validity of the light attenuation model and the proposed method was verified using the rerendering percent error whose histogram was shown in Figure 6.18. Table 6.4 shows the mean rerendering percent error using the proposed light attenuation model. As was shown, by using just four parameters, i.e. two in $\bar{g}_k$, $\chi_k$ and $\kappa_{a,k}$, the $k^{th}$ light attenuation can be encoded with mean accuracy of 97.38% to 98.79%. More complicated light attenuation model with more degrees of freedom may be proposed to enhance the accuracy, which is left of future work.

The next experiment aims to address the importance of light attenuation correction. It compares the surface reconstruction results of a print paper using PS with SNI with and without the correction. All the calibrations were priorly completed. Figure 6.19(a) and (b)
respectively show the reconstructed surface of the print paper with and without light attenuation correction. Comparing the two results, it was obvious that if the image irradiances were not corrected by the light attenuations, the reconstructed surface was severely distorted. The one with light attenuation correction was closer to a flat plane and the roughness of the print paper was clearly visualized, which also indirectly suggested that the light attenuation calibration was successful.

Figure 6.19: (a) Reconstructed surface with light attenuation correction; (b) Reconstructed surface without light attenuation correction

Evaluation on Specular Color Calibration

In order to evaluate the performance of the proposed specular color calibration method, ground truth must be determined at first. The nominal value of light SPD and the camera’s spectral sensitivity provided by [91] were used to compute the ground truth of $\bar{s}_{RGB}$ using equation (3.30). The light SPD and the camera’s spectral sensitivity being used are respectively given in Figure 6.20(a) and (b). $\bar{s}_{RGB}$ was computed as $\begin{bmatrix} 0.4373, 0.5577, 0.7055 \end{bmatrix}^T$. 

**Table 6.4: Mean rerendering percent error**

<table>
<thead>
<tr>
<th>Light Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean percent error [%]</td>
<td>1.28</td>
<td>1.24</td>
<td>1.41</td>
<td>1.21</td>
<td>2.62</td>
<td>1.39</td>
<td>1.72</td>
<td>2.22</td>
</tr>
</tbody>
</table>
Using the proposed specular color calibration method for each light source, the estimated $\bar{s}_{RGB}$ was deviated from the ground truth by $4.88^\circ$ and the standard deviation of the eight measurements was $0.79^\circ$. As shown from the results, $\bar{s}_{RGB}$ obtained by the two methods were similar, which proved the validity of the proposed method. The consistency of $\bar{s}_{RGB}$ further justified the applicability of the proposed DRM-based color PS method.

### 6.4.3 Evaluation of the Developed System

The first experiment aims to evaluate the accuracy of the developed system using sandpapers. Sandpapers were used since their embedded particle size of the abrasive material, referred to as the grit size, is strictly controlled when manufacturing \cite{92}. Therefore, their statistical metrics can be used as the ground truth. Two different grit of sandpapers were measured using the developed system, which were grit 80 and grit 120 manufactured by 3M. Fig. 6.21(a) and (b) demonstrate the reconstruction results of the sandpaper with grit 80 and 120 using the developed system, respectively. Figures on the left were the digital relighted RGB images of the surface, while figures on the right were the reconstructed surface height.
Figure 6.21: (a) Measurements for sandpaper of Grit 80; (b) Measurements for sandpaper of Grit 120

The height was shifted with mean-zero. The given results were zoomed to an area of $10mm \times 10mm$ to visualize details of the surfaces. Comparing results of the RGB images and the height images qualitatively, clear associations can be identified. For quantitative analyses, the results were compared with those measured by the Nanovea ST400 using chromatic confocal optical technologies [93]. Several statistical metrics were adopted for the comparison purpose, which were skewness (Ssk), kurtosis (Sku), maximum (Sz), arithmetical mean (Sa) and root mean square (Sq) for the height. Table 6.5 shows the comparison results. 'PS' and 'NV' were results by the developed system and the Nanovea profilometer, respectively. From
Table 6.5: Measurement comparison on sandpapers

<table>
<thead>
<tr>
<th>System</th>
<th>Grit</th>
<th>Ssk</th>
<th>Sku</th>
<th>Sz [micron]</th>
<th>Sa [micron]</th>
<th>Sq [micron]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS</td>
<td>80</td>
<td>0.264</td>
<td>3.06</td>
<td>499.1</td>
<td>52.96</td>
<td>66.37</td>
</tr>
<tr>
<td>NV</td>
<td>80</td>
<td>0.893</td>
<td>3.59</td>
<td>520.2</td>
<td>53.8</td>
<td>66.5</td>
</tr>
<tr>
<td>PS</td>
<td>120</td>
<td>−0.1431</td>
<td>2.98</td>
<td>367.3</td>
<td>39.32</td>
<td>42.48</td>
</tr>
<tr>
<td>NV</td>
<td>120</td>
<td>1.099</td>
<td>4.187</td>
<td>273.2</td>
<td>24.41</td>
<td>31.18</td>
</tr>
</tbody>
</table>

Table 6.6: Measurement setting comparison between PS and GFM

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>PS</th>
<th>GFM</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal range</td>
<td>$L_x$</td>
<td>100.00</td>
<td>50.00</td>
<td>mm</td>
</tr>
<tr>
<td>Transverse range</td>
<td>$L_y$</td>
<td>66.85</td>
<td>38.00</td>
<td>mm</td>
</tr>
<tr>
<td>Elevation range</td>
<td>$L_z$</td>
<td>13.46</td>
<td>10</td>
<td>mm</td>
</tr>
<tr>
<td>Longitudinal resolution</td>
<td>$\Delta X$</td>
<td>0.0136</td>
<td>0.0308</td>
<td>mm</td>
</tr>
<tr>
<td>Transverse resolution</td>
<td>$\Delta Y$</td>
<td>0.0136</td>
<td>0.0307</td>
<td>mm</td>
</tr>
<tr>
<td>Elevation resolution</td>
<td>$\Delta Z$</td>
<td>0.0049</td>
<td>0.0025</td>
<td>mm</td>
</tr>
<tr>
<td>Data acquisition rate</td>
<td>1.808</td>
<td>0.02</td>
<td>million points/second</td>
<td></td>
</tr>
</tbody>
</table>

The table, the root mean square, arithmetical mean and kurtosis values were in agreement between the two measurements, while the skewness was not. The major reason was due to the inherent measurement errors appearing in the two systems. For the developed system, the error was mostly originated from the reflectance model assumption. In this case, the specularities caused by the aluminum oxide abrasive particles were more dominant and the DRM model was insufficient to characterize such surface reflectance. Another plausible error source, but not limited to, was inter-reflection which was not modelled due to its complexity. Conservatively speaking, from the results analyses, the developed system has achieved the accuracy in the order of 10 microns.

The second experiment aims to compare the measurement results for an asphalt road sample between the developed system (PS) and the GFM MikroCAD 3D profilometer (GFM) which utilizes structured light method with fringe patterns [94]. The comparison of measurement settings between the two systems is given in Table 6.6. Four regions of an asphalt
road surface were measured and compared. Figure 6.22 to 6.25 respectively show the results for region 1 to 4. In these figures, (a) represents the estimated normal map from PS and (b) demonstrates the relighted color image. (c) and (d) give the reconstructed surface by PS and by GFM, respectively. In the elevation direction, the reconstructed surfaces were shifted to mean-zero. Only the 'same' area of the surface measured by both systems was plotted for comparison purpose. Measurement on exactly the same area at the scale of microns is difficult, considering the sample’s offset in the longitudinal and transverse direction, the existence of a rotation around z-axis, and the different sampling rate between the two systems. To compare the measurements, data from GFM was first matched to data from PS using
Figure 6.23: Asphalt surface region 2: (a) Estimated normal map from PS; (b) Digital relighted image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM iterative closest point (ICP) [95] in order to determine a 3D rigid transformation between the two point clouds. By transforming data from GFM to the same coordinate system as data from PS and interpolating data from GFM to have the same sampling points as data from PS, the corresponding z values were compared and the mean absolute errors for the four regions were 27.1\(\mu m\), 37.8\(\mu m\), 46.7\(\mu m\) and 28.1\(\mu m\), respectively. From the analyses, the mean measurement difference between the two systems was within 50\(\mu m\).

Measurements on the four regions by the two systems were then compared in the spatial frequency domain. For each measured 3D profile of an area, the data was first converted to one linear profile by concatenating columns of data end-to-end back and forth. Fast Fourier
Figure 6.24: Asphalt surface region 3: (a) Estimated normal map from PS; (b) Digital relighted image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM

Transform (FFT) in 1D was then performed on the linear profile. Figure 6.26 shows the FFT analyses comparison between PS and GFM on the four measured regions of the asphalt road sample. As shown from the figure, results from the two systems were comparable with small differences. At low frequencies, results from the two systems matched well, whilst at higher frequencies, the amplitudes of PS were systematically lower than those of GFM. This was due to the smoothening effect of the developed PS-based system by enforcing the integrability during SNI. This smoothening effect can be also clearly observed from Figure 6.22 to 6.25. Performing SNI while preserving the surface discontinuities remains open and
Figure 6.25: Asphalt surface region 4: (a) Estimated normal map from PS; (b) Relighted color image; (c) Reconstructed surface from PS; (d) Reconstructed surface from GFM

is left of future work. The small measurement differences might also originate from the different poses of the sensors relative to the target surface during measurements.

The third experiment aims to investigate the applicability of the system for on-site measurement. Figure 6.27(a) shows the on-site experiment setting of the developed system including the cover to block ambient light. Figure 6.27(b) specifies the eight different locations of experiments. Surfaces on these locations were representative to test the system’s applicability. Table 6.7 shows the types of surface on different measurement locations. Figure 6.28 and 6.29 demonstrate the results on the eight locations. (a) gives the estimated
normal map from PS and (b) shows the relighted color image of the target surface. The reconstructed surface geometry is shown in (c). Due to the lack of ground truth for on-site measurements, residual from PS and from SNI were adopted for evaluation. Residual from PS is given in (d) and residual from SNI is provided in (e). As shown from the figures, actual shapes of the target surface were qualitatively characterized. The captured color information provided more cues for recognition to rule out outliers, such as the leaf lying on the target surface on Location 5. The digital relighted image illuminating from the viewer direction

Figure 6.26: FFT analyses comparison between PS and GFM: (a) Region 1; (b) Region 2; (c) Region 3; (d) Region 4
has also minimized cast-shadows and maximized the color information. For near-Lambertian surfaces as Location 1-4, residuals from PS at boundaries of the FoV are higher than those at central region, which is due to both the limited angle of view (AoV) of the LEDs and second illumination from the cover. Residuals from SNI were mostly high where discontinuities appeared. More comprehensive evaluations were beyond the scope of the dissertation.

6.5 Conclusions and Future Works

This chapter has presented the design and calibration of an on-site microtexture road profiling system using color PS with SNI. Optimal illumination configurations were investigated through theoretical analyses and numerical experiments. The illumination was finally set in a circular pattern having eight LEDs with zenith angle of $25^\circ$. Five different calibrations for the developed system were proposed and evaluated through experiments individually.
Table 6.7: Surface types of different measurement locations

<table>
<thead>
<tr>
<th>Location index</th>
<th>Surface type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>asphalt rough</td>
</tr>
<tr>
<td>2</td>
<td>asphalt fine</td>
</tr>
<tr>
<td>3</td>
<td>concrete fine</td>
</tr>
<tr>
<td>4</td>
<td>concrete rough</td>
</tr>
<tr>
<td>5</td>
<td>concrete with leaf</td>
</tr>
<tr>
<td>6</td>
<td>gravel</td>
</tr>
<tr>
<td>7</td>
<td>metallic</td>
</tr>
<tr>
<td>8</td>
<td>plastic</td>
</tr>
</tbody>
</table>

The overall performance of the developed system was evaluated on sandpapers and an asphalt road sample through comparative studies. The results have shown that the system has achieved the accuracy in the order of $10\mu m$. Though having exhibited comparable results with existing systems, the reconstructed surface by the developed system was systematically smoothened due to the enforced integrability during SNI. The capability for on-site measurements was finally demonstrated on eight different representative surfaces. Digital relighted color image was simultaneously obtained by the developed system and it provided richer information of the target surface. More functionalities, as was shown in chapter 4, can be also incorporated. Though this system was designed for microtexture road profiling purpose, it can be generalized to solve the overall micro-scale shape reconstruction problem. In the applicability tests, results for metallic and plastic surfaces were also demonstrated, whilst more advanced PS algorithms may be required and their developments are left of future works.
Figure 6.28: Measurement results on Location 1-4: (a) Estimated normal map; (b) Relighted image; (c) Reconstructed surface; (d) Residual from PS; (e) Residual from SNI
Figure 6.29: Measurement results on Location 5-8: (a) Estimated normal map; (b) Relighted color image; (c) Reconstructed surface; (d) Residual from PS; (e) Residual from SNI
Chapter 7

Conclusions and Future Works

7.1 Conclusions

This dissertation has presented three major contributions. First, a PS method based on DRM using color images was proposed to generalize PS applicable to a wider range of surfaces with non-Lambertian reflectances. The proposed method not only estimates surface orientations from diffuse reflection but also exploits information from specularities owing to the proposed diffuse-specular separation algorithm. Using the proposed DRM-based color PS method, material-dependent features can be simultaneously obtained in addition to surface orientations, which offers much richer information in understanding the 3D scene and poses more potential functionalities, such as specular removal, intrinsic image decomposition, digital relighting, material-based segmentation, material transfer and material classification. The experimental results suggested that by incorporating the newly proposed surface normal refinement step using specularity information, the accuracy of surface orientation estimation was enhanced by around 30% in median. The results also indicated the descent performance of the proposed method on dielectric materials and degradation on metallic surfaces due to
the limitation of DRM. In addition, the applicability of the proposed method was identified that it was only applicable to surfaces whose diffuse and specular color are distinct.

The second contribution is the development of an SNI method dealing with perspective distortion. The proposed SNI method was performed on the image plane instead of on the target surface normally done by orthographic SNI owing to the newly derived representation of surface normals. This new representation was manipulated to the so-called log gradient space in analogy to the gradient space in orthographic SNI. Using the proposed SNI method, perspective distortion can be efficiently tackled with for smooth surfaces. In addition, the method was PS-independent, which can keep the image irradiance equation in a simple form during PS. Due to the derived analogy, the proposed method was capable of building on most past algorithms developed for orthographic SNI. From the experimental results, the proposed perspective SNI method provided more accurate surface reconstruction than orthographic SNI. By comparing the results from minimizing the two proposed cost functionals, performing perspective SNI in the gradient space instead of in the log gradient space could provide better solution in the presence of discontinuity at the cost of more computation time.

The third contribution is the development of a 3D micro-scale shape reconstruction system using the proposed PS with SNI method. This system was designed to measure pavement microtexture on-site, while its applicability can be generalized to a wider range of surfaces. The optimal illumination configuration was investigated. Five different calibrations regarding various aspects of the system were either newly proposed or modified from existing methods. The performance of the developed system was comprehensively evaluated. The results suggested that the system has achieved the accuracy in the order of $10\mu m$ and exhibited comparable results with existing systems. Its capability for on-site measurement was also confirmed.
7.2 Future Works

Direct future works can be explained from two aspects. In theory, the following topics may be investigated:

- Employ other physically-based BRDF instead of Blinn-Phong BRDF to better account for Fresnel reflection in DRM-based color PS;
- Propose an alternative method when diffuse and specular color are close for DRM-based color PS;
- Investigate each implication listed in the DRM-based color PS, such as exploring proper usage of classified cast shadows from digital relighting to constrain SNI;
- Investigate a perspective SNI method to preserve discontinuities;
- Formulate PS from deterministic to stochastic problem.

From the engineering aspect, the works may involve implementation of the software in a more efficient platform instead of MATLAB, applicability tests and evaluations on various surfaces, and investigate the engineering design for different machine vision applications.

For future works in the long run, the fusion problem of PS with other existing 3D reconstruction techniques is of particular interest, such as multiple-view PS. PS with focal stacking, PS with high dynamic range imaging, multi-spectral PS are also left of future works to enhance different aspects of the PS-based system. PS using machine learning techniques to deal with different surface reflectances is also interesting to be investigated. The ultimate goal of the research could be to develop a 3D scanning system that can provide very dense and accurate measurements while at the same time, encoding the scene into a series of parameters to even infer its material composition.
Bibliography


[92] Iso8486-1: Bonded abrasives - determination and designation of grain size distribution - part 1: Macrogrits f4 to f220, 1996.
