Image-Based 3D Reconstruction of Surfaces with Highly Complex Reflectance Properties

Von der Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Dortmund
genehmigte

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
eingereicht von

Malte Lenoch

Tag der mündlichen Prüfung:	09.01.2019
Hauptreferent:	Prof. Dr. rer. nat. habil. Christian Wöhler
Korreferent:	Prof. Dr. rer. nat. habil. Josef Pauli

Arbeitsgebiet Bildsignalverarbeitung Technische Universität Dortmund

Acknowledgment

First of all, I want to sincerely thank my advisor Prof. Christian Wöhler for his support, encouragement and the freedom to pursue my ideas. I have been able to explore many different fields of study, inside and outside the areas that are relevant for this thesis and learned a lot during my time as a PhD student. Furthermore, I am grateful to my second advisor Prof. Josef Pauli for his effort and interest in the topic. The time he spent to analyze this written text in depth contributed to making this thesis a little more approachable from a readers point of view.

I very much enjoyed working at the Image Analysis Group of TU Dortmund University and want to express my thanks to my colleagues Hussam Al-Behadili, Gabriele Berberich, Arne Grumpe, Daniela Rommel, Atheer Al-Tameemi and Thorsten Wilhelm for their feedback, valuable discussions on research topics, delightful breaks and the occasional laugh that we shared. Likewise I am thankful to all the colleagues of LS KT and AG DT. Many student coworkers and thesis-authors I have been working with made the department an even more more interesting place of work and the past five years a fun experience. I would like to name Pia Biebrach, Christian Dopp, Timur Güven, Patrick Klokowski, Dominik Kossmann, Christian Leismann and Natascha Mengewein and apologize to all those that I may have forgotten in this list, I wish you all the best for your personal and professional future.

Last, but certainly not least, I want to thank my family and friends that continually supported and motivated me in the past five years. Additionally, if the need arose they were able to provide the distraction I needed to gather a new perspective on challenging problems. Wibke in particular deserves my special gratitude as her encouragement, patience and confidence made finishing this thesis possible.

ABSTRACT

The camera-based acquisition of the environment has become an ordinary task in today's society as much in science as in everyday-life situations. Smartphone cameras are employed in interactive video games and augmented reality, just as industrial quality inspection, remote sensing, robotics and autonomous vehicles rely on camera sensors to analyze the outside world. One crucial aspect of the automated analysis is the retrieval of the 3D structure of unknown objects in the scene – be it for collision prevention, grabbing, or comparison to a CAD model – from the acquired image data. Reflectance-based surface reconstruction methods form a valuable part of the set of camera-based algorithms. Stereo cameras exploit geometrical optics to triangulate the 3D position of a scene point while photometric procedures require one camera only and estimate a surface gradient field based on the shading of an object. The reflectance properties of the object have to be known to achieve this which results in a chicken-and-egg problem on unknown objects since the surface shape has to be available to approximate the reflectance properties, and the reflectance properties have to be known to estimate the surface shape. This situation is circumvented on Lambertian surfaces, yet, those that are of interest in real-world applications exhibit much more complex reflectance properties for which this problem remains.

The challenge of estimating the unknown spatially varying bidirectional reflectance distribution function (BRDF) parameters of an object of approximately known shape is approached from a Bayesian perspective employing reversible jump Markov chain Monte Carlo methods to infer both, reflectance parameters and surface regions that show similar reflectance properties from sampling the posterior distributions of the data. A significant advantage compared to non-linear least squares estimates is the availability of statistical information that can directly be used to evaluate the accuracy of the inferred patches and parameters. In the evaluation of the method, the derived patches accurately separate a synthetic and a laboratory dataset into meaningful segments. The reflectance of the synthetic dataset is almost perfectly reproduced and misestimated BRDF parameters underline the necessity for a large dataset to apply statistical inference. The real-world dataset reveals the inherent problems of BRDF estimation in the presence of cast shadows and interreflections. Furthermore, a procedure that is suitable to calibrate a two-camera photometric stereo acquisition setup is examined. The calibration is based on multiple images of a diffuse spherical object that is located in corresponding images. Although the calibration object is supposed to be perfectly diffuse by design, considering a specular Phong component in addition to the Lambertian BRDF model increases the accuracy of the rendered images. The light source positions are initialized based on stereo geometry and optimized by minimizing the intensity error between measured and rendered images of the calibration object.

Ultimately, this dissertation tackles the task of image-based surface reconstruction with the contribution of two novel algorithms. The first one computes an initial approximation of the 3D shape based on the diffuse component of the reflectance and iteratively refines this rough guess with gradient fields calculated from photometric stereo assuming a combination of the BRDF models of Lambert and Blinn. The second method computes the surface gradient fields for both views of a stereo camera setup and updates the estimated depth subject to Horn's integrability constraint and a new regularization term that accounts for the disparity offset between the two matching gradient fields. Both procedures are evaluated on objects that exhibit complex reflectance properties and challenging shapes. A fringe projection 3D scanner is used for reference data and error assessment. Small details that are not visible in the coarse initial 3D data, that is supplied to the first algorithm, are recovered based on the high-quality gradient data obtained from photometric stereo. The error of the test data with respect to the reference scanner is less than 0.3 mm. In contrast to the first method that computes shape information, the stereo camera algorithm yields absolute 3D data and produces very good reconstruction results on all datasets. The proposed method even surpasses the reconstruction accuracy of the 3D scanner on a metallic dataset. This is a notable contribution, as most existing camera-based surface reconstruction methods exclusively handle diffusely reflecting objects and those that focus on non-Lambertian objects still struggle with highly specular metallic surfaces.

KURZFASSUNG

Eine kamerabasierte Erfassung der Umwelt ist heutzutage für viele wissenschaftliche wie alltägliche Anwendungen zweifellos normal. Kameras in Smartphones werden für interaktive Videospiele und Augmented Reality Anwendungen genutzt. Auch Methoden der industriellen Oberflächeninspektion, der Fernerkundung, der Robotik und des autonomen Fahrens verlassen sich auf bildgebende Verfahren, um ihre Umgebung zu analysieren. Ein grundlegender Aspekt dieser Analyse ist die Erfassung und Rekonstruktion der dreidimensionalen Struktur von unbekannten Objekten, sei es zur Kollisionsvermeidung, um einen Greifvorgang zu planen oder um Abweichungen zu einem CAD Modell zu messen. Einen wichtigen Teil dieser bildbasierten 3D Rekonstruktionsalgorithmen stellen reflektanzbasierte Verfahren dar. Im Gegensatz zu Stereo-Algorithmen, die per Triangulation die 3D Koordinaten der Szenenpunkte ermitteln, arbeiten reflektanzbasierte Verfahren nur mit einer Kameraperspektive und schätzen das Gradientenfeld einer Objektoberfläche anhand ihrer Schattierung. Dabei begegnet man einem Henne-oder-Ei-Dilemma, denn um die Oberfläche rekonstruieren zu können müssen die Reflektanzeigenschaften bekannt sein, aber um die Reflektanzeigenschaften zu berechnen muss vorher die Oberflächenstruktur (mindestens grob) bekannt sein. Unter der Annahme einer rein diffus reflektierenden Oberfläche lassen sich Gradientenfeld und Albedo zwar gleichzeitig schätzen, diese Vereinfachung ist jedoch für viele reale Anwendungen zu stark und daher insgesamt nur von untergeordneter Relevanz.

Die Schätzung einer unbekannten, lokal veränderlichen Bidirektionalen Reflektanzverteilungsfunktion (BRDF) auf einer grob bekannten Oberfläche wird in dieser Arbeit mit Methoden der Bayes'schen Statistik untersucht. Ein Reversible Jump Markov chain Monte Carlo Algorithmus wird genutzt, um gleichzeitig Parameter und Regionen gleicher Reflektanzeigenschaften aus den Messdaten abzuleiten. Ein maßgeblicher Vorteil, verglichen mit nichtlinearen Schätzungen im Sinne der kleinsten Quadrate, ist die Nutzbarkeit von statistischen Analysemethoden, um die berechneten Werte direkt einer Genauigkeitsbewertung zu unterziehen. Die in der Evaluierung ermittelten Oberflächensegmente teilen die Testobjekte akkurat in Bereiche gleicher Reflektanzeigenschaften ein. Gleichzeitig werden bei der Analyse der Ergebnisse die Notwendigkeit von großen Datensätzen für statistische Schätzungen und die inherenten Probleme von BRDF Parameterschätzungen bei Vorhandensein von Schatten und Interreflexionen deutlich gemacht. Desweiteren wird ein Algorithmus erläutert, mit dem ein Zwei-Kamera Aufnahmesystem unter Nutzung einer diffus reflektierenden Kugel für photometrisches Stereo kalibriert werden kann. Die Lichtpositionen werden mit Überlegungen der Stereogeometrie initialisiert und durch die Minimierung des Intensitätsfehlers zwischen gemessenen und gerenderten Bildern optimiert. Obwohl das Kalibrierobjekt per Definition diffus reflektieren soll, wird die Ähnlichkeit der Bilder und damit Genauigkeit der Kalibrierung durch Nutzung eines zusätzlichen Phong Spiegelterms erhöht.

Abschließend werden zwei neue Methoden vorgestellt, die sich dem Thema der bildbasierten Oberflächenrekonstruktion widmen. Im ersten Verfahren wird eine initiale 3D Struktur basierend auf der diffusen Reflektanzkomponente bestimmt und iterativ durch ein Gradientenfeld verfeinert, das unter der Annahme der BRDF von Lambert und Blinn mittels photometrischem Stereo geschätzt wird. Der zweite Algorithmus nutzt ein Stereokamerasystem, um eine grobe Schätzung der Oberfläche zu ermitteln, deren Genauigkeit unter Horns Integrabilitätsbedingungen verbessert wird. Zusätzlich wird eine neue Regularisierungsbedingung definiert, welche die Disparität zwischen beiden Kameras durch das Matching der Gradientenfelder aus beiden Perspektiven berücksichtigt. Beide Algorithmen werden auf Labordaten von Objekten ausgewertet, die lokal veränderliche Reflektanzeigenschaften und komplexe Formen aufweisen. Messungen mit einem 3D Streifenprojektionsscanner werden als Referenzdaten und zur Fehleranalyse genutzt. Artefakte und Fehler der initialen Schätzung des ersten Algorithmus können korrigiert werden und das Ergebnis weist das erhöhte Detailniveau des Gradientenfeldes auf. Die Abweichung der ermittelten Form im Vergleich zu den Referenzdaten liegt bei unter 0.3 mm. Der Stereo-Algorithmus liefert absolute Tiefeninformation und gute Ergebnisse auf allen Datensätzen. Die Rekonstruktionsgenauigkeit des Referenzscanners wird auf einer metallischen Oberfläche sogar bei weitem übertroffen. Das ist ein beachtenswerter Erfolg, da stark spiegelnde und metallische Oberflächen die meisten kamera-basierten 3D Rekonstruktionsverfahren auch heutzutage noch vor große Probleme stellen.

PUBLICATIONS

The following papers have been published during the author's employment at TU Dortmund University. Publication (1) is an extension of the author's Master thesis and (2)–(5) comprise research that is relevant in the area of this doctoral thesis.

1. Lenoch, M., Herbort, S., Grumpe, A., and Wöhler, C. (2014). Linear unmixing in BRDF reproduction and 3D shape recovery. *Proc. Int. Conf. Pattern Recognition.*
2. Lenoch, M., Mistry, H., and Wöhler, C. (2015). Photometric surface reconstruction of surfaces with non-uniform reflectance properties. In Luhmann, T. and Müller, C., editors, *Photogrammetrie, Laserscanning, Optische 3D Messtechnik - Beiträge der Oldenburger 3D Tage 2015*, pages 228–236. Wichmann.
3. Lenoch, M., Wilhelm, T., and Wöhler, C. (2016). Simultaneous surface segmentation and BRDF estimation via Bayesian methods. *Proc. Int. Conf. on Computer Vision Theory and Applications*, pages 39–48.
4. Lenoch, M. and Wöhler, C. (2016). Reflectance-based 3D shape refinement of surfaces with spatially varying BRDF properties. *9th IAPR Workshop on Pattern Recogniton in Remote Sensing*, pages 1–6.
5. Lenoch, M., Biebrach, P., Grumpe, A., and Wöhler, C. (2017). Multi-view shape from shading constrained by stereo image analysis. In Liu, M., Chen, H., and Vincze, M., editors, *Proc. Int. Conf. on Computer Vision Systems*, volume 10528 of *Lecture Notes in Computer Science*, pages 325–335. Springer International Publishing.

CONTENTS

Contents IX

1 Introduction 1

2 Related Work: Measurement and Modeling of Reflectance Functions 5
- 2.1 The Image Formation Process 5
- 2.2 Physical Modeling of Light Transport 8
- 2.3 Reflection Phenomena 10
- 2.4 BRDF Models 13
- 2.5 Data-driven Reflectance Models 18

3 Related Work: Surface Reconstruction 21
- 3.1 Shape from Shading 22
- 3.2 Photometric Stereo 25
- 3.3 Generalized Lambertian Photometric Stereo 26
- 3.4 Non-Lambertian Photometric Stereo 28
- 3.5 A Brief Outlook on the Variety of Photometric Reconstruction Methods . . . 31
- 3.6 Surface Reconstruction from Triangulation 33
- 3.7 Combining Multi-View Stereo and Shape from Shading 38

4 Data Acquisition 45
- 4.1 Acquisition Setup 45
- 4.2 Depth Data Quality Assessment 45
- 4.3 Image Data Quality Assessment 48
- 4.4 High Dynamic Range Images 49
- 4.5 Relative Change of Projected Sensor Cell Size 51

5 Contribution: Bayesian Approximation of Reflectance Functions 53
- 5.1 Bayesian BRDF Modeling 54
- 5.2 Markov Chain Monte Carlo Methods 55
- 5.3 Reversible Jump Markov Chain Monte Carlo 58
- 5.4 Application 62
- 5.5 Evaluation 68
- 5.6 Summary 73

6 Contribution: Calibrating Light Sources in a Stereo Setup 75
- 6.1 Examining the Calibration Object 75

6.2 Linear Initialization 79
6.3 Non-Linear Optimization and Evaluation 84
6.4 Summary 86

7 Contribution: Image-Based Surface Reconstruction 89
7.1 Reflectance-based 3D Shape Refinement 89
7.2 Multi-View Photometric Stereo Constrained by Stereo Image Analysis 102

8 Summary and Conclusion 127

Bibliography 131

List of Symbols i

List of Figures xi

List of Tables xiii

A Full Datasets xv
A.1 Camera Image Assessment xv
A.2 RJMCMC Datasets xix
A.3 Surface Refinement Datasets xxiv
A.4 Stereo Fusion Datasets xxviii

INTRODUCTION

1

The capabilities of the human observer of a scene regarding the interpretation of that scene are tremendous. Recognition of objects including heavily occluded ones is processed literally within the blink of an eye, as well as the estimation of the 3D shape of natural and human-made objects from different materials and very distinct reflectance properties. Of course, this is not only an act of the stereo vision system that many mammals are equipped with but also of the human brain that realizes a database search comparing the visual input to everything that one has learned so far. The necessary processing steps to achieve this immediate and adaptive understanding of the environment are very complex, and researchers have tried for decades already to equip a machine with (some of) the skills of the human visual system. However, these abilities are still far from being matched by today's computer vision systems, especially regarding the true *understanding* of a scene although efforts have been made to empower machine vision to a form of semantic image analysis (e.g., Krause et al., 2014; Fei-Fei and Perona, 2005). Nevertheless, since the human brain is fundamentally – as far as it has been understood by scientists until today – a computing engine, the human mind sets the limits of what intelligent machines can achieve or in the future maybe even outperform. Machines surpass human capacities already, yet only regarding low-level requirements (e.g., infra-red vision) and not in the processing and fusion of data.

Computer vision emerged as a field of research and study in the 1970s as a consequence of the development of the first digital image sensors. The objective to equip a computer system with the competence of a human that interprets an image or a sequence of images has by now been split up into various tasks like object recognition and classification, motion estimation, video tracking, and 3D object reconstruction, that are essential skills necessary to achieve the more abstract goal. All these tasks developed into their proper disciplines keeping researchers in computer science and engineering busy.

High-quality visual inspection of surfaces is a topic of great interest to many fields of application ranging from industrial quality inspection to medical image analysis, and this work focuses on the reconstruction of the three-dimensional shape of unknown objects. From the scope of possible procedures to measure the shape of an object, image-based algorithms have the inherent appeal not to require any physical contact with the object under inspection. Even proximity is not necessarily required if the camera lens is adequate. Both qualities are precious if physical contact is prohibitive due to e.g., high temperatures in industrial facilities or vast distances on celestial bodies. Active and passive range-scanning techniques can be distinguished, where passive methods comprise e.g., stereo vision (passive triangulation), structure-from-motion, bundle adjustment, shape from focus and shape from defocus that do not require any active projection onto the surface or specific lighting of the scene. Active procedures rely on an emitting and a receiving component and can rely on e.g., active triangulation, time-of-flight, interferometry, and shading cues. The listed techniques are only a subset of the enormous variety of existing methods that have their flaws and merits. Those that are more closely related to this work will be reviewed in more detail in the following chapters. The contribution of this thesis focuses on shape from shading, an active technique that estimates the surface gradient from shading cues obtained from photographs of the object under varying illumination.

Photometric estimation of surface gradient fields was originally introduced by Rindfleisch (1966), Horn (1970) and Woodham (1980) for the computation of lunar topography and diffusely reflecting surfaces. Diffuse surfaces are attractive since they allow computation based on a system of linear equations, yet most real-world and human-made objects do not comply with the assumption of pure diffuse reflection. Considering a wider range of reflectance properties yields two consequences. First, the reflectance characteristics need to be modeled to reproduce the measured intensities which is either done with analytical reflectance models or with large databases of real materials. Second, the adequate parameters of the reflectance models or entries from the database have to be selected and to do so an estimate of the surface structure is necessary. Hence, if there is no prior knowledge available, as it is assumed in this thesis, one faces a chicken-and-egg problem: To estimate the reflectance parameters the surface structure has to be known, to compute the surface structure the reflectance parameters have to be known. Possible approaches to resolve this situation are illustrated in this work.

Motivation

Many research in the domain of photometric surface reconstruction is still limited to diffusely reflecting surfaces and techniques have been developed to exclude surface regions that violate this assumption from the computations. This simplification ignores additional information that is available and could be exploited. Additionally, reflectance properties that are spatially varying are considered in many recent publications but in many cases refer to albedo variations only. The greater challenge is a constant albedo while the amount of specular reflection is (gradually) changing. Moreover, metallic and highly specular surfaces pose a great challenge to photometric reconstruction algorithms since the incident light is only reflected into a narrow solid angle. However, these are the surfaces that are of great interest in industrial applications like quality inspection of manufactured components. Fringe projection 3D scanners use a form of active triangulation and are the state-of-the-art regarding quality inspection in industrial contexts since the processing of the input data can be conducted in a real-time system and the results are submillimeter accurate. Photometric systems have not yet reached that level of accuracy, but precision in the range of 0.5 mm is achievable, which will be demonstrated in this thesis even on challenging surfaces. The equipment necessary to capture the input data amounts to two industrial cameras and a handful of light sources and is therefore at least an order of magnitude less expensive than a competing 3D range scanning system. All in all, this thesis tackles the topic of high-quality surface reconstruction of objects with highly complex reflectance properties including metallic surfaces solely based on camera images of these objects.

Outline

The work will be structured as follows. Chapter 2 will revise the basic principles of the image formation and the modeling of light transport and surface reflectance. The work contributed by other researchers that affect the various topics of image-based surface reconstruction that are relevant throughout this thesis will be covered in Chapter 3. The laboratory setup that is used to acquire the data that are examined during the experimental evaluation of the proposed algorithms is illustrated in Chapter 4 together with necessary steps to calibrate the measurement equipment. The contributions of this thesis will be elaborated in Chapters 5 to 7 and comprise a Bayesian approach to infer the unknown spatially varying reflectance parameters and their allocation from the input data based on a Markov chain Monte Carlo algorithm in Chapter 5. Chapter 6 contains a new procedure to calibrate the light source positions of a two-camera photometric stereo setup based on a diffuse spherical object. Finally, two new procedures to recover 3D information from camera images are presented in Chapter 7 and tested on multiple laboratory data that include an object with spatially varying reflectance properties and a metallic surface. Chapter 8 concludes the presented research and gives an outlook on potential future work.

RELATED WORK: MEASUREMENT AND MODELING OF REFLECTANCE FUNCTIONS

2

2.1 The Image Formation Process

The following section will be used to define and illustrate the conventions in describing points, vectors, and matrices in 2D and 3D space and corresponding transformations that are essential to describe the image formation process. Considerations of pros and cons of different types of camera lenses and digital sensor models are lying beyond the scope of this work, such that a geometric description of a camera image will be the main focus. The definitions will closely follow Wöhler (2013). Lowercase bold letters will denote vectors and points, as they are basically the same. However, if not specified otherwise, vectors are assumed to be of unit length, such that $\mathbf{n} = (n_{\mathrm{x}}, n_{\mathrm{y}}, n_{\mathrm{z}})^{\mathrm{T}}$, with $||\mathbf{n}||_2 = 1$, whereas points are in general not. A scalar product will be noted $\langle \mathbf{x} \cdot \mathbf{y} \rangle = \sum_i x_i \cdot y_i$ and matrices $\mathbf{R}$ of any size will be represented by bold uppercase letters. A function with an italic subscript variable indicates the partial derivative with respect to this variable or these variables, e.g., $\frac{\partial z}{\partial x} = z_x$ and $\frac{\partial}{\partial x}\frac{\partial}{\partial y} z = z_{xy}$.

2.1.1 The Pinhole Camera

Although there exists a variety of more complex camera models, e.g., thin-lens model, fisheye model (e.g., Kingslake, 1992), the pinhole camera model is commonly used and a reasonable approximation of the camera lenses used in the experimental evaluation. Therefore the following explanations will only briefly introduce the concept of a pinhole camera as depicted in Fig. 2.1. A lens is described by its *optical center* $\mathbf{c}$ and the *optical axis*, which is perpendicular to the plane defined by the camera lens and passes through the optical center. Additionally, every straight line connecting a scene point and its projection onto the image plane has to pass through the optical center. The intersection between the optical axis and the image plane is termed *principal point* $\mathbf{c}_0$, and the distance between the principal point and the optical center is the *principal distance* b. Every lens has a focal length f that is in a real camera always smaller than b, however, since the value of f approaches b if the object distance is much larger than the principal distance (Wöhler, 2013), the focal length will be used as the principal distance in the experiments.

A maximum of four coordinate systems can be necessary to describe the projection of an arbitrary scene point to the (digital or analogue) image sensor. This comprises the world coordinate system W, camera coordinate system C, image coordinate system I and sensor

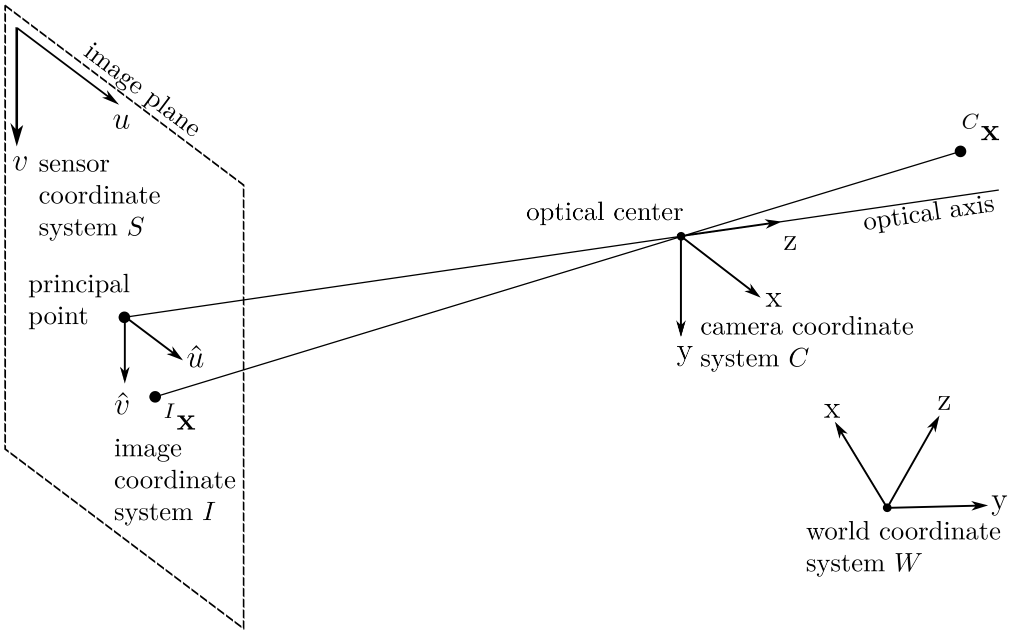


Figure 2.1. The pinhole camera model and its associated coordinate systems. A scene point defined in the camera coordinate system ${}^{C}\mathbf{x}$ is projected onto the image plane at location ${}^{I}\mathbf{x} = {}^{I}_{C}T({}^{C}\mathbf{x})$. Adapted from Wöhler (2013).

coordinate system S in this order. The 3D world coordinate system may have an arbitrary orientation and origin, such that the transformation from world to camera coordinate system ${}^{C}_{W}T$ contains a rotation and a translation. While in a single camera system the obvious choice is to align world and camera coordinate system, there exists a separate coordinate system C_i for each camera i in a multi camera system. A point in (Euclidean) 3D space, defined in the world coordinate system W, will be denoted ${}^{W}\mathbf{x}$ and is transformed into the camera coordinate system by

$$ {}^{C_i}\mathbf{x} = {}^{C_i}_{W}T({}^{W}\mathbf{x}) = \hat{\mathbf{R}}_i\,{}^{W}\mathbf{x} + \mathbf{t}_i. \tag{2.1} $$

Usually, the world coordinate system is aligned with the first camera, such that $\mathbf{R}_1 = \mathbb{I}_3$, where $\mathbb{I}_3$ is the 3×3 identity matrix, and the translation vector $\mathbf{t}_1 = (0,0,0)^{\mathrm{T}}$. In the case of a single camera system the index i is omitted.

The projection of ${}^{C}\mathbf{x} = (x,y,z)^{\mathrm{T}}$ into the point ${}^{I}\mathbf{x}$ on the image plane, connected by a straight line in accordance with the pinhole camera model, is denoted by ${}^{I}\mathbf{x} = {}^{I}_{C}T({}^{C}\mathbf{x})$. This mapping already displays the inherent loss of information by capturing a 3D scene in a 2D image, as every point on the connecting line would be projected into the same image point, and no information about the depth is preserved. The theorem of intersecting lines yields the image coordinates ${}^{I}\mathbf{x} = (\hat{u}, \hat{v})^{\mathrm{T}}$ as

$$ \begin{aligned} \frac{\hat{u}}{b} &= -\frac{x}{z} \\ \frac{\hat{v}}{b} &= -\frac{y}{z}\,. \end{aligned} \tag{2.2} $$

So far, all units, from world to camera to image coordinate system, are measured in the same metric scale (usually mm). If a digital camera sensor is used, which is the case for all experiments in this thesis, the sensor coordinates are measured in pixels and the mapping from image to sensor is an affine transformation in the most general case of pixels in the shape of parallelograms (Wöhler, 2013). Square pixels reduce the complexity to a similarity transform. The principal point is located in the image plane at ${}^{I}\mathbf{c}_0 = (\hat{c}_\mathrm{u}, \hat{c}_\mathrm{v})^\mathrm{T} = (0,0)^\mathrm{T}$ and on the sensor at ${}^{S}\mathbf{c}_0 = (c_\mathrm{u}, c_\mathrm{v})^\mathrm{T}$.

2.1.2 Camera Calibration

As the imaging process does in general not preserve lengths and angles it is adequate to depart from the use of Euclidean geometry. Perspective geometry allows for a wider variety of transformations that perfectly describe the perspective projection (therefore the name) of a 3D scene onto a 2D image plane. In two-dimensional Euclidean space a point $\mathbf{x} = (x, y)^\mathrm{T}$ corresponds to a point $\tilde{\mathbf{x}} = (X, Y, W)^\mathrm{T}$ in two-dimensional projective space $\mathbb{P}^2$. The vector that describes this point $\tilde{\mathbf{x}} = (X, Y, W)^\mathrm{T} = (\beta X, \beta Y, \beta W)^\mathrm{T}$ and as such its norm has no relevance for every $\beta \neq 0$. The Euclidean point corresponding to the projective point can be computed by normalizing the projective vector $\mathbf{x} = (X/W, Y/W)^\mathrm{T} = (x, y)^\mathrm{T}$.

The extrinsic calibration of a multi-camera system yields the orientation between each of the camera coordinate systems and the world coordinate system as given in Eq. (2.1). The camera-specific parameters, containing properties of sensor and lens, are denoted intrinsic parameters and have to be calibrated with an appropriate procedure. The camera calibration method of Heikkila and Silven (1997) in the extended implementation by Bouguet (2008) has been used in this thesis to calibrate the intrinsic parameters of the single camera setup and the intrinsic and extrinsic parameters of the stereo camera setup.

The intrinsic parameters comprise the coordinates of the principal point in the sensor plane c_u and c_v, the principal distance b scaled by the effective number of pixels per unit length k_u and k_v along the horizontal and the vertical image axis, respectively, and the pixel skew angle α (Wöhler, 2013). The camera matrix $\mathbf{A}$ then yields

$$\mathbf{A} = \begin{pmatrix} -bk_\mathrm{u} & -bk_\mathrm{u}\cot\alpha & c_\mathrm{u} \\ 0 & -bk_\mathrm{v}/\sin\alpha & c_\mathrm{v} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} c_\mathrm{u} & d_\mathrm{u}\cot\alpha & c_\mathrm{u} \\ 0 & d_\mathrm{v}/\sin\alpha & c_\mathrm{v} \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

b is sometimes termed camera constant and $(d_\mathrm{u}, d_\mathrm{v}) = (-bk_\mathrm{u}, -bk_\mathrm{v})$ is the camera constant in pixels. Most modern camera sensors employ square pixels without skew such that $k_\mathrm{u} = k_\mathrm{v}$ and $\alpha = \pi/2$ which leads to $c = d_\mathrm{u} = d_\mathrm{v}$. A simplified version of $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} c & 0 & c_\mathrm{u} \\ 0 & c & c_\mathrm{v} \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.4}$$

The 3×4 projection matrix $\mathbf{P}_i$ describes, in projective space, the transformation of a point from world to sensor coordinate system S_i of camera i and combines the camera matrix with rotation and translation, i.e., intrinsic and extrinsic calibration are fused into a single matrix.

$${}^{S_i}\tilde{\mathbf{x}} = \mathbf{P}_i \, {}^{W}\tilde{\mathbf{x}} = \left[\mathbf{M}_i | \mathbf{p}_4^{(i)}\right] {}^{W}\tilde{\mathbf{x}} = \mathbf{A}_i \left[\hat{\mathbf{R}}_i | \mathbf{t}_i\right] {}^{W}\tilde{\mathbf{x}} \tag{2.5}$$

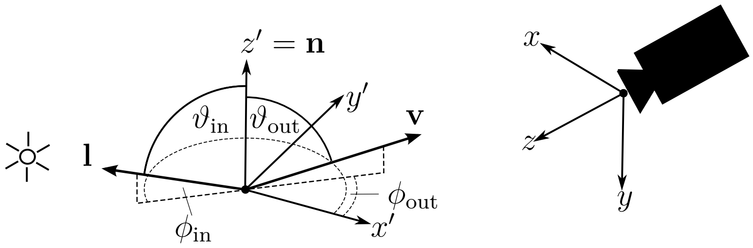


Figure 2.2. Definitions of a local coordinate system in contrary to the camera coordinate system. The surface normal direction $\mathbf{n}$ is aligned with the z-axis of the coordinate system. Incident $\mathbf{l}$ and emitting $\mathbf{v}$ directions of light are defined in terms of an infinitesimal small solid angle around their principal direction, $\omega_{\text{in}} = \{\vartheta_{\text{in}}, \phi_{\text{in}}\}$ and $\omega_{\text{out}} = \{\vartheta_{\text{out}}, \phi_{\text{out}}\}$. An often used quantity is the half-angle $\mathbf{h}$ that is the bisector of $\mathbf{l}$ and $\mathbf{v}$ according to $\mathbf{h} = \frac{\mathbf{l}+\mathbf{v}}{\|\mathbf{l}+\mathbf{v}\|_2}$.

2.2 Physical Modeling of Light Transport

2.2.1 Local Coordinate System

In addition to the viewer-specific coordinate systems described in Section 2.1.1 exists an object centered *local* coordinate system as depicted in Fig. 2.2. A spherical coordinate system with its origin at the currently considered surface point, defining a direction via elevation angle ϑ and azimuth angle ϕ, is best suited to express the transfer of energy incident from one and emitted to another direction. The surface normal direction at that point $\mathbf{n}$ is aligned with the z-axis of the local coordinate system.

2.2.2 Basics of Radiometry

It is necessary to briefly define some basic radiometric units to describe the physical properties of light transport in a scene in terms of mathematics, for a more comprehensive overview the reader is referred to the literature e.g., (Nicodemus et al., 1977; Dorsey et al., 2008). The object-centered coordinate system favors the use of spherical coordinates, and a section A of the unit sphere surrounding the origin is well-defined by the elevation ϑ and azimuthal angle ϕ. Thus, a collection of these angles $\{\vartheta, \phi\}$ uniquely defines the corresponding section of the sphere surface $\omega = \{\vartheta, \phi\}$. ω is the so-called *solid angle* (measured in steradians = sr), which corresponds to the extension of the radian to the third dimension, the surface area of the unit sphere is 4π sr.

$$\omega = \frac{A}{r^2} \qquad \text{(sr)} \tag{2.6}$$

The solid angle is given by the ratio of the size of the sphere section A to the sphere's squared radius r^2, such that the aperture viewed from the center of the sphere is independent of the absolute values of surface area and radius, and only defined by the angular section. Analogously, the surface elements $\mathrm{d}A$ spanning from ϕ to $\phi + \mathrm{d}\phi$ and ϑ to $\vartheta + \mathrm{d}\vartheta$ define the differential solid angle $\mathrm{d}\omega$. Since the width of the surface patch depends on ϑ and is given by $\sin(\vartheta)\,\mathrm{d}\phi$, thus being very small at ϑ near 0 and wide at ϑ near $\pi/2$, the differential element can be expressed as

$$\mathrm{d}\omega = \frac{\mathrm{d}A}{r^2} = \sin(\vartheta)\,\mathrm{d}\vartheta\,\mathrm{d}\phi. \tag{2.7}$$

The average radiant flux Φ measured in watts (W) that is incident on – or emitted from – a certain surface area A is termed *irradiance* E and defined as radiant flux per unit area. However, the directional dependence of the radiant flux, i.e., a certain viewing or incident direction, is of greater interest than an average value for the entire hemisphere. If the direction of emission ω_{out} is not perpendicular to the surface A, the apparent size of the area of emission shrinks equivalent to $A\cos(\vartheta_{\text{out}})$. This effect is accommodated in the *radiance* L, that is the radiant flux emitted from a surface patch A and passing in the direction of a solid angle $\mathrm{d}\omega$ measured with respect to the surface normal direction of A.

$$\mathrm{d}L(\vartheta_{\text{out}}, \phi_{\text{out}}) = \frac{\mathrm{d}\Phi(\vartheta_{\text{out}}, \phi_{\text{out}})}{\mathrm{d}A\cos(\vartheta_{\text{out}})\,\mathrm{d}\omega_{\text{out}}} \qquad (\mathrm{W\,m^{-2}\,sr^{-1}}) \tag{2.8}$$

If the appearance of an object shall be modeled and the surface shape is known the remaining question is how much light – that is the electromagnetic radiation visible to the camera[1] – is reflected from a surface subject to some incident irradiance. First introduced by Nicodemus (1965) and Nicodemus et al. (1977), the *bidirectional reflectance distribution function* (BRDF) f_{r} describes the reflection of light from a known incident direction ω_{in} into an arbitrary reflection direction ω_{out} as the ratio of radiance and irradiance for a certain surface point.

$$f_{\text{r}} = \frac{\mathrm{d}L_{\text{out}}(\vartheta_{\text{out}}, \phi_{\text{out}})}{\mathrm{d}E_{\text{in}}(\vartheta_{\text{in}}, \phi_{\text{in}})} = \frac{\mathrm{d}L_{\text{out}}(\vartheta_{\text{out}}, \phi_{\text{out}})}{L_{\text{in}}(\vartheta_{\text{in}}, \phi_{\text{in}}) \cdot \cos(\vartheta_{\text{in}}) \cdot \mathrm{d}\omega_{\text{in}}} \tag{2.9}$$

The incident irradiance per differential surface area $\mathrm{d}E_{\text{in}}(\vartheta_{\text{in}}, \phi_{\text{in}})$ can be replaced by the equivalent radiance incident on the surface scaled with the cosine of the incidence angle to again account for the change of the projected surface area. A wide variety of approaches to model the BRDF of certain materials exists throughout the research community, some of which will be discussed in Section 2.4. Assuming the strength of a point light source I_0 and a distance between light source and surface r, the light reflected in direction $\omega_{\text{out}} = (\vartheta_{\text{out}}, \phi_{\text{out}})$ incident from $\omega_{\text{in}} = (\vartheta_{\text{in}}, \phi_{\text{in}})$ can be expressed as

$$\mathbf{R} = \frac{I_0}{r^2} \cdot \underbrace{f_{\text{r}}(\vartheta_{\text{in}}, \phi_{\text{in}}, \vartheta_{\text{out}}, \phi_{\text{out}})}_{\text{BRDF}} \cdot \langle \mathbf{n} \cdot \mathbf{l} \rangle\,, \tag{2.10}$$

where $\mathbf{v}$ and $\mathbf{n}$ are the unit-length vectors pointing in the direction of the light source and the surface normal direction, respectively, as defined in Fig. 2.2. $\mathbf{R}$ is considered to be a matrix of synthetic values that are subject to certain simplifications and a chosen BRDF model, in contrast to measured intensity data $\mathbf{I}$.

2.2.3 Properties of the Bidirectional Reflectance Distribution Function

Although many researchers at the beginning of computer vision were dedicated to the specific field of computer graphics, i.e., creating artificial images of scenes based on mathematical descriptions of the object properties that appear realistic, the approaches that existed to model a BRDF have been applied to scene recovery as well. Especially in this domain not the appearance of the rendered image but the obedience to the laws of physics are of utmost

---

[1] Light oftentimes describes the portion of wavelengths of electromagnetic waves that is detectable by a human viewer. However, the luminous efficiency of a camera is in general different in that it detects a wider band of wavelengths than a human.

importance. There exists a consensus among researchers, (e.g., Lewis, 1994; Giesen, 2009; Soldado and Ureña, 2012), that three critical properties have to be fulfilled to call a BRDF model physically plausible, that is it does not violate the laws of physics while not necessarily being derived from these.

Energy Conservation

Since a BRDF describes the transfer of power from a source to a sink it has to comply with energy conservation. As all considerations in this thesis cover non-emissive surfaces only, the energy reflected into the entire hemisphere Ω can never surpass the incident energy.

$$\int_{\Omega} f_{\mathrm{r}}(\vartheta_{\mathrm{in}}, \phi_{\mathrm{in}}, \vartheta_{\mathrm{out}}, \phi_{\mathrm{out}}) \cos(\vartheta_{\mathrm{out}}) \; \mathrm{d}\omega_{\mathrm{out}} \leq 1, \forall \, \omega_{\mathrm{in}} \tag{2.11}$$

Reciprocity

The second principle is the reciprocity according to von Helmholtz (1924). The viewing and incident direction of the BRDF have to be exchangeable without influencing the function result.

$$f_{\mathrm{r}}(\vartheta_{\mathrm{in}}, \phi_{\mathrm{in}}, \vartheta_{\mathrm{out}}, \phi_{\mathrm{out}}) \overset{!}{=} f_{\mathrm{r}}(\vartheta_{\mathrm{out}}, \phi_{\mathrm{out}}, \vartheta_{\mathrm{in}}, \phi_{\mathrm{in}}) \tag{2.12}$$

Positivity

Since light can be reflected, transmitted or absorbed at the interface of two media, the BRDF value can either be a positive value, of course in accordance with Eq. (2.11), or zero in the case of transmission and absorption. However, the contribution can never be negative and as such no BRDF model shall generate negative values for any combination of angles of incidence and reflection.

$$f_{\mathrm{r}}(\vartheta_{\mathrm{in}}, \phi_{\mathrm{in}}, \vartheta_{\mathrm{out}}, \phi_{\mathrm{out}}) \geq 0 \tag{2.13}$$

A BRDF is generally a four-dimensional function $f_{\mathrm{r}}(\vartheta_{\mathrm{in}}, \phi_{\mathrm{in}}, \vartheta_{\mathrm{out}}, \phi_{\mathrm{out}})$, depending on absolute values of the angles of incidence and reflection. Anisotropic materials (e.g., fabrics, brushed aluminum) exhibit a macroscopic structure that results in a directional dependence of the reflected light on the azimuth angle (e.g., Ward, 1992; Ashikhmin and Shirley, 2000). However, the simplified case of isotropic reflection, that suits many materials and is assumed for a variety of BRDF models, reduces the dependence from absolute values of the azimuth angle to the relative angle between incident and viewing direction. The BRDF can thus be expressed as a function of three variables with the introduction of $\phi_{\mathrm{iso}} = |\phi_{\mathrm{in}} - \phi_{\mathrm{out}}|$.

Furthermore, strictly speaking, a BRDF is a function of the wavelength λ of the incident light as well. Since the experimental evaluations of this thesis rely on monochrome image sensors and narrow-band incident light, there is no need for explicitly modeling varying reflection characteristics at different wavelengths.

2.3 Reflection Phenomena

The reflection of light on a surface can cause a variety of different effects that mostly depend on the material properties. The most common, that are accounted for by the mathematical BRDF models to a different extent are discussed in this section.

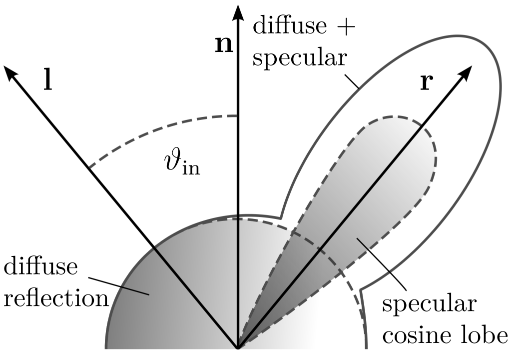


Figure 2.3. Superposition of diffuse and specular reflection component according to a cosine lobe model.

2.3.1 Diffuse and Specular Reflection

Shafer (1985) proposed the *dichromatic reflection model* according to which the two basic properties of reflectance of real surfaces are *diffuse* and *specular* reflection and the total amount of reflected light can be expressed as a superposition of both independent parts

$$L(\lambda, \vartheta_{\text{in}}, \vartheta_{\text{out}}, \phi_{\text{iso}}) = L_{\text{d}}(\lambda, \vartheta_{\text{in}}, \vartheta_{\text{out}}, \phi_{\text{iso}}) + L_{\text{s}}(\lambda, \vartheta_{\text{in}}, \vartheta_{\text{out}}, \phi_{\text{iso}}). \tag{2.14}$$

Note that as stated before, the wavelength of the light is neglected in any further considerations. Subsurface scattering of the incident light that occurs when it penetrates the surface causes the diffuse reflection component. The distribution of the reflected light is uniform in the entire hemisphere and independent of the viewing direction. Additionally, the atomic structure of the material acts as a bandpass filter on the incoming light wave and the emitted light obtains the color of the surface. Thus the diffusely reflected light determines the perceived color of an object and this effect is also denoted *body* reflection.

Since it originates from the reflection of light at the interface between the light-carrying and the object medium, the specular component is also referred to as *interface* reflection. For an ideal mirror, the reflection is a single spike directed along the direction of perfect reflection $\vartheta_{\text{in}} = \vartheta_{\text{out}}$, real specular reflection tends to produce a symmetric cone centered around ϑ_{out} that is oftentimes approximated with a cosine lobe raised to a power (e.g., Phong, 1975; Blinn, 1977; Lafortune et al., 1997; Ashikhmin and Shirley, 2000). For interface reflections, the wavelength of the reflected light equals the wavelength of the incident light. Figure 2.3 depicts the superposition based on a cosine lobe to model the specular reflection component.

2.3.2 Fresnel Effect

The effect named after Augustin-Jean Fresnel (1788 – 1827) describes the strong increase of the specular reflection component when the viewing angle becomes larger than ≈ 70°. This effect is strong on translucent materials (e.g., glass, water) and dielectrics, and negligible on metallic surfaces. Figure 2.4 illustrates this effect. The Fresnel equations explaining the physical background can be derived from Maxwell's laws of electro-magnetic waves at the interface of two smooth and homogeneous media (e.g., Band, 2006). The amount of perpendicularly polarized light that is reflected ρ_{r} or transmitted ρ_{t} at the interface of the two media is described by the following equations (e.g., Bennett, 1995, pp. 5.4–5.7). The interface is defined by incident angle ϑ_{in}, transmission angle ϑ_{t} and the refractive indices of

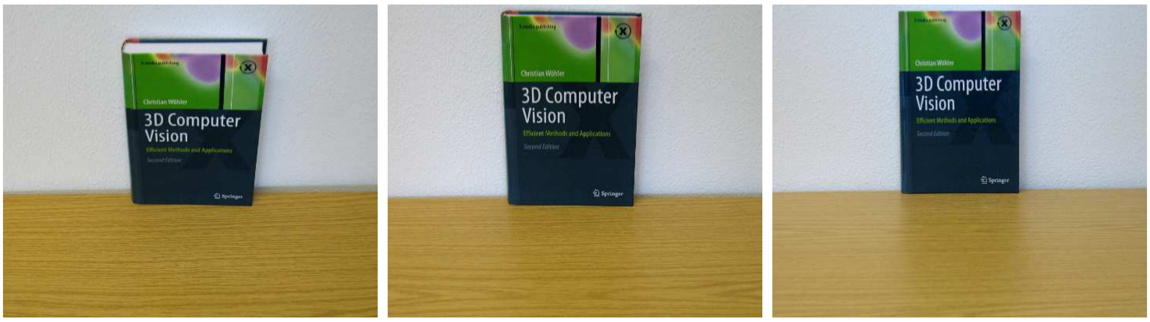


Figure 2.4. Illustration of Fresnel effect. For grazing viewing directions the diffuse component of the table fades and the intensity of the specular reflection increases. Adapted from Lafortune et al. (1997).

the media n_1 and n_2.

$$\rho_\mathrm{r} = \frac{n_1\cos(\vartheta_\mathrm{in}) - n_2\cos(\vartheta_\mathrm{t})}{n_1\cos(\vartheta_\mathrm{in}) + n_2\cos(\vartheta_\mathrm{t})} \tag{2.15}$$

$$\rho_\mathrm{t} = \frac{2n_1\cos(\vartheta_\mathrm{in})}{n_1\cos(\vartheta_\mathrm{in}) + n_2\cos(\vartheta_\mathrm{t})} \tag{2.16}$$

If passing from an optically denser to a less dense medium, i.e., if $n_1 > n_2$, the incident wave is completely reflected at the interface for $\vartheta_\mathrm{in} > \vartheta_\mathrm{crit}$. This effect enables technologies like fiber optic communication. The critical angle is derived from Snell's law and given by

$$\vartheta_\mathrm{crit} = \arcsin\left(\frac{n_1}{n_2}\right). \tag{2.17}$$

The extinction coefficient κ takes the amount of light into account that is absorbed by a medium and its use is facilitated employing a complex refractive index

$$\tilde{n} = n + j\kappa. \tag{2.18}$$

Nevertheless, κ is omitted in most cases of the modeling of Fresnel reflection for the sake of simplicity. The BRDF model by Cook and Torrance (1981) (cf. Section 2.4.5) includes the Fresnel effect, yet without using the extinction coefficient as there is only a weak angular dependence of the Fresnel term on the extinction coefficient according to the authors.

2.3.3 Backscatter, Off-specular Reflection and Diffuse Fall-off

Backscatter and *off-specular reflection* denote a specular reflection component that is not aligned with the direction of mirror-like reflection and usually occurs in addition to the dominant specular lobe. Many real-world materials exhibit these properties, which go beyond the simple cosine lobe model, up to some degree. A variety of the materials contained in the MERL database of BRDFs (Matusik et al., 2003a,b) show strong off-specular lobes towards grazing incident angles. Backscatter denotes a particular form of off-specular reflection in which a – normally less distinct – specular lobe is directed towards the light source. Obviously, it is challenging to measure this effect accurately since the light source occupies the direction of exact backscatter.

Diffuse reflection according to Lambert's law (cf. Section 2.4.1) distributes the incident light equally into the entire upper hemisphere. Nonetheless, the BRDFs of real materials tend to exhibit lower emittance values towards grazing viewing angles than predicted by the ideal diffuse model. This effect is denoted as *diffuse fall-off* or *generalized diffuse* reflection (e.g., Dorsey et al., 2008). Figure 2.5 shows examples of all three phenomena.

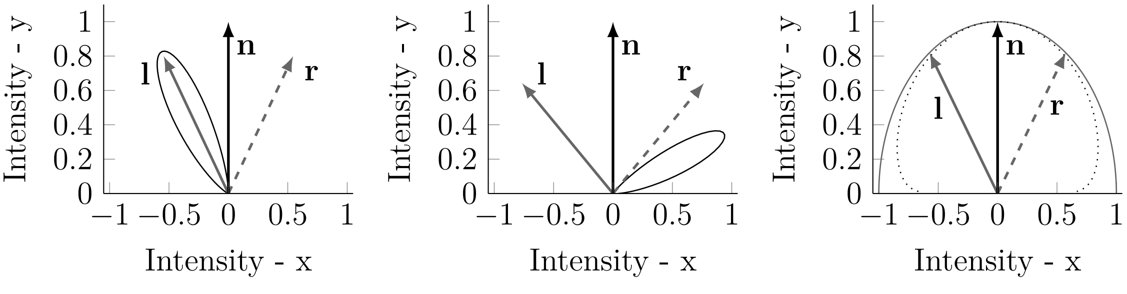
2.3.4 Direct and Indirect Illumination

A BRDF model accounts for the light that is reflected from an infinitesimal small surface patch incident from an arbitrary direction. A standard practice of estimating the shading of an entire scene is to evaluate the BRDF at every surface point subject to the incident light and the viewing direction. Both directions can be regarded as uniform for simplicity, assuming a large distance between light source and surface or between surface and camera, respectively. Computing the illumination angles requires the knowledge of the surface gradient at every point, and as such takes only the local structure of the surface into account, the result is the *direct* illumination that is reflected from every point of the surface into the camera.

However, every surface patch also reflects the incoming light, without loss of generality, into the entire hemisphere surrounding that patch and thereby possibly becomes a point light source itself and illuminates other parts of the surface. The second-hop of the light rays[2] in between objects or surface parts that can “see” each other contributes to the *indirect* or *global* illumination of the surface and massively increases the perceived realism of rendered scenes. The computation of the indirect illumination requires knowledge of not only the surface gradients but the 3D structure of the scene and the visibility between all rendered scene points, making it complex and cumbersome regarding computation time. With the focus on photorealistic rendering, some techniques have been developed that approximate the indirect illumination of a given scene. Radiosity is based on heat-transfer equations and used to compute the diffusely emitted radiation between scene points (e.g., Cohen et al., 1993). Ray tracing emits rays from the light source and follows multiple hops between surface points of the objects or follows only those rays that reach the viewer to save computational time and trace them to the source of the intensity (e.g., Pharr and Humphreys, 2010). Ray tracing is an iterative procedure since every ray that reaches a surface generates many more that have to be superimposed. However, few iterations yield satisfying results already as the intensity

---

[2] And of course all consecutive ones, although the second hop is the most important one since the intensity fades with the squared distance between source and sink.



(a) Backscatter (b) Off-specular Reflection (c) Diffuse Fall-off

Figure 2.5. Examples of different reflection phenomena. (a) Backscatter, (b) off-specular reflection, (c) Lambertian diffuse reflectance (solid line) and exemplary diffuse fall-off (dashed line). **l** is the incident light direction and **r** the direction of mirror-like reflectance.

fades quickly. Figure 2.6 shows an exemplary scene rendered under direct illumination with and without occlusion information and indirect illumination using Blender[3].

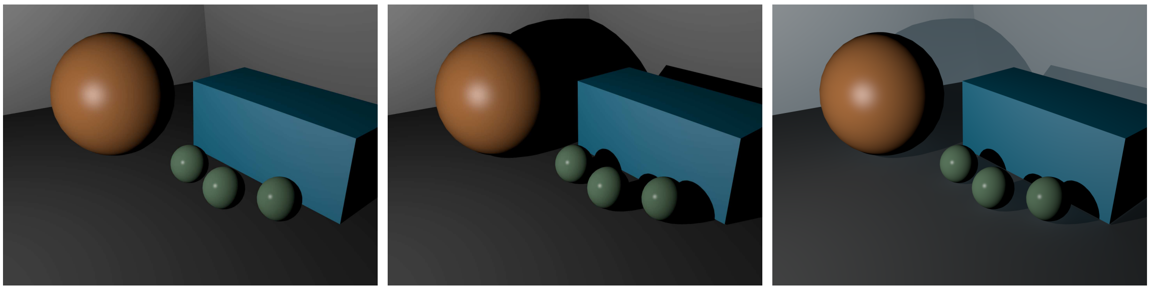

(a) (b) (c)

Figure 2.6. Basic scene rendered under a single point light source and (a) direct illumination, (b) direct illumination with occlusion information and (c) indirect illumination with occlusion information. Images rendered with Blender version 2.69.0.

2.4 BRDF Models

A possibility to describe the reflectance properties of a surface following the dichromatic reflectance model is the use of analytical BRDF models. Computing the amount of reflected light is thus reduced to the (more or less) simple function evaluations subject the object and lighting geometry and a set of model parameters that is contained in the vector $\boldsymbol{\theta}$. These models can be empirically (e.g., Blinn, 1977; Phong, 1975) as well as physically (e.g., Torrance and Sparrow, 1967; He et al., 1991) motivated. This section will cover some of the most common and for this thesis most relevant BRDF models. A brief overview of the considered effects and limitations of these BRDF models is given in Table 2.1.

Table 2.1. Limitations and features of BRDF models.

feature	Blinn	Phong	Lafortune	Cook-Torrance
diffuse reflection	yes	yes	yes	yes
specular highlight	yes	yes	yes	yes
Fresnel reflection	no	no	implicit	yes
arbitrary direction of specular reflection	no	no	yes	no

2.4.1 Lambert

First mentioned by Lambert (1760), the simplest form of a reflectance function is the assumption of perfect diffuse reflection. Hence, the incident light is completely reflected and distributed equally into the upper hemisphere. To ensure conservation of energy, the integral of the upper hemisphere has to be limited to 1, cf. Eq. (2.11), and thus according to, e.g., Horn (1977) the BRDF becomes

$$f_{\mathrm{r}}^{(\mathrm{Lambert})} = \frac{1}{\pi}. \tag{2.19}$$

---

[3] www.blender.org, version 2.69.0

Note that the solid angle of a hemisphere is 2π, however, the division of the BRDF value is only by π (e.g., Horn, 1986). This is due to the fact that the projected size of the surface is dependent on the incident angle ϑ_{in} and therefore the integral that has to be evaluated is

$$\begin{aligned}\int_{\Omega} \cos(\vartheta)\,\mathrm{d}\omega_{\mathrm{out}} = & \int_0^{2\pi}\int_0^{\pi/2} \cos(\vartheta)\sin(\vartheta)\,\mathrm{d}\vartheta\,\mathrm{d}\phi = 2\pi\int_0^{\pi/2}\cos(\vartheta)\sin(\vartheta)\,\mathrm{d}\vartheta \\ =& 2\pi\left[-\frac{1}{2}\cos^2(\vartheta)\right]\Bigg|_0^{\pi/2} = 2\pi\frac{1}{2} = \pi\end{aligned} \tag{2.20}$$

instead of the just integrating over the surface of (half) a sphere

$$\int_0^{2\pi}\int_0^{\pi/2}\sin(\vartheta)\,\mathrm{d}\vartheta\,\mathrm{d}\phi = 2\pi. \tag{2.21}$$

Now, a little more realism can be achieved if it is considered that the surface (partly) absorbs the incident light. The albedo $\rho \in [0\ 1] \subset \mathbb{R}^+$ is the ratio of incident energy and reflected energy that describes the surface reflectivity and thus the only parameter of the reflectance model. $\rho = 0$ indicates complete absorption and for $\rho = 1$ all light is reflected into the upper hemisphere. The BRDF thus becomes

$$f_{\mathrm{r}}^{(\mathrm{Lambert})} = \frac{\rho}{\pi}. \tag{2.22}$$

2.4.2 Phong

The earliest version of an empirically based cosine lobe BRDF model was developed by Phong (1975). Lambert's law is enhanced by an additional ambient light term $k_{\mathrm{a}}/\langle\mathbf{n}\cdot\mathbf{l}\rangle$ and, more importantly, the specular reflection is modeled by a weighted cosine lobe $k_{\mathrm{s}}\langle\mathbf{r}\cdot\mathbf{v}\rangle^{\gamma}/\langle\mathbf{n}\cdot\mathbf{l}\rangle$. The specular lobe is centered according to Phong along the ideal direction of mirror-like reflection $\mathbf{r} = 2\langle\mathbf{n}\cdot\mathbf{l}\rangle\,\mathbf{n} - \mathbf{l}$. The angle $\vartheta_{\mathrm{r}} = \arccos(\langle\mathbf{r}\cdot\mathbf{v}\rangle)$ between mirror and viewing direction and the specular weight k_{s} determine the intensity of the specular highlight. Note that $k_{\mathrm{s}} = 0$ when $\langle\mathbf{r}\cdot\mathbf{v}\rangle < 0$. The diffuse weight k_{d} accounts for the diffuse reflection component of the surface. All weights contain the albedo ρ as a global scaling.

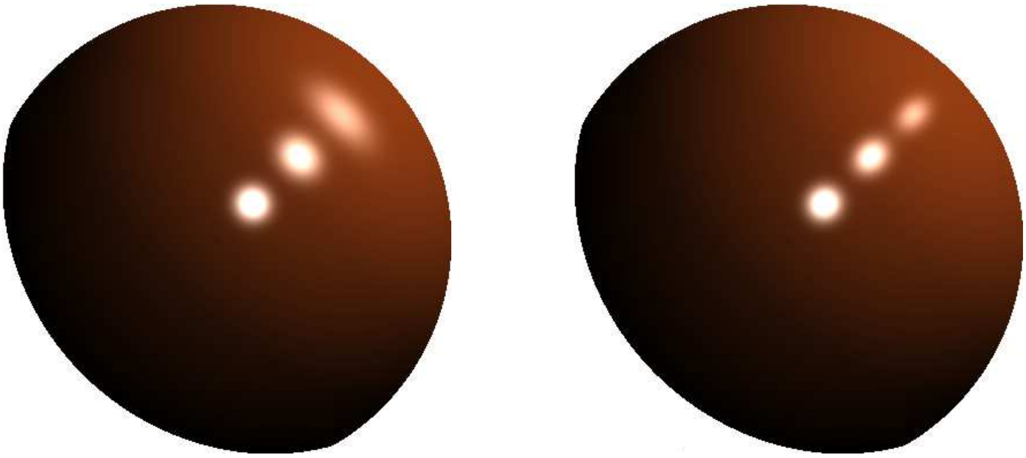
The exponent of the cosine $\gamma \in \mathbb{R}^+$ controls the width of the cosine lobe and as such the width of the specular reflection on the surface. Ranging from equal reflection over the whole hemisphere ($\gamma = 0$, i.e., ideal diffuse) to ideal mirror ($\gamma \to \infty$).

$$f_{\mathrm{r}}^{(\mathrm{Phong})} = \frac{k_{\mathrm{a}}}{\langle\mathbf{n}\cdot\mathbf{l}\rangle} + k_{\mathrm{d}} + k_{\mathrm{s}}\frac{\langle\mathbf{r}\cdot\mathbf{v}\rangle^{\gamma}}{\langle\mathbf{n}\cdot\mathbf{l}\rangle} \tag{2.23}$$

Yet, the model proposed by Phong complies neither with energy conservation nor with reciprocity. To ensure physical plausibility, a modification of Phong's BRDF was introduced by Lafortune and Willems (1994). The ambient light component is omitted and a normalization factor $N_{\mathrm{Phong}} = \frac{2+\gamma}{2\pi}$ is introduced to comply with energy conservation. The model thus becomes

$$f_{\mathrm{r}}^{(\mathrm{Plausible\ Phong})} = \frac{k_{\mathrm{d}}}{\pi} + k_{\mathrm{s}}\frac{2+\gamma}{2\pi}\langle\mathbf{r}\cdot\mathbf{v}\rangle^{\gamma} \tag{2.24}$$

with the additional constraint $k_{\mathrm{d}} + k_{\mathrm{s}} \leq 1$. $k_{\mathrm{d}}, k_{\mathrm{s}} > 0$ has to hold for both formulations of the Phong BRDF to comply with the positivity constraint in Eq. (2.13).



(a) Phong (b) Blinn

Figure 2.7. Rendered image of a sphere with (a) Phong and (b) Blinn BRDF model, both physically plausible and with three light sources. The color is arbitrarily chosen to illustrate the difference between diffuse and specular component. Note the elongation of the specular highlights in (a). Rendering parameters are: $I_0 = 1$, $k_\mathrm{d} = 0.8$, $k_\mathrm{s} = 0.2$, $\gamma_\mathrm{Phong} = 50$, $\gamma_\mathrm{Blinn} = 200$. γ_Blinn has to be $\approx 4\gamma_\mathrm{Phong}$ to achieve a similar width of the specular highlight.

2.4.3 Blinn

A variation of the Phong model was introduced by Blinn (1977) in using the half-vector $\vartheta_\mathrm{h} = (\mathbf{l}+\mathbf{v})/\left\|\mathbf{l}+\mathbf{v}\right\|_2$ to replace the vector product $\langle\mathbf{r}\cdot\mathbf{v}\rangle$ by $\langle\mathbf{h}\cdot\mathbf{n}\rangle$. $\mathbf{h}$ represents the theoretical surface normal direction to a facet that would yield mirror like reflection for given incident light and viewing directions, thus the vector product accounts for the angular distance to that ideal direction and the BRDF is

$$f_\mathrm{r}^{(\mathrm{Blinn})} = \frac{k_\mathrm{a}}{\langle\mathbf{n}\cdot\mathbf{l}\rangle} + k_\mathrm{d} + k_\mathrm{s}\frac{\langle\mathbf{h}\cdot\mathbf{n}\rangle^\gamma}{\langle\mathbf{n}\cdot\mathbf{l}\rangle}. \tag{2.25}$$

Since this is only a small modification, the original Blinn model is not compliant with energy conservation either. However, the rendered images look slightly more realistic and the more expensive computation of the reflection direction is circumvented. An additional normalization factor $N_\mathrm{Blinn} = (\gamma+2)(\gamma+4)/(8\pi(2^{-\frac{\gamma}{2}}+\gamma))$ derived by Giesen (2009) and the omission of ambient light (i.e., $k_\mathrm{a} = 0$) as proposed by Lewis (1994) are required to ensure energy conservation. Furthermore, the denominator $\langle\mathbf{n}\cdot\mathbf{l}\rangle$ has to be removed to guarantee reciprocity. Hence, the physically plausible Blinn BRDF becomes

$$f_\mathrm{r}^{(\mathrm{Plausible\ Blinn})} = \frac{k_\mathrm{d}}{\pi} + k_\mathrm{s}\frac{(\gamma+2)(\gamma+4)}{8\pi(2^{-\frac{\gamma}{2}}+\gamma)}\langle\mathbf{h}\cdot\mathbf{n}\rangle^\gamma. \tag{2.26}$$

The formulations of the BRDF models and the idea of a symmetrical cosine lobe are similar. However, they differ regarding perceived realism which is exemplified in Fig. 2.7. The specular highlight of the Phong model starts to get elongated towards grazing incident angles. This is not consistent with experimental analysis.

Table 2.2. Comparison of eigenvalue configurations for Lafortune model.

configuration	reflective behavior
$\gamma = 0$	lambertian reflectance
$-D_{\mathrm{xx}} = -D_{\mathrm{yy}} = D_{\mathrm{zz}}$	classic cosine lobe
$D_{\mathrm{xx}} \neq D_{\mathrm{yy}}$	anisotropic reflection
$-D_{\mathrm{xx}} = -D_{\mathrm{yy}} > D_{\mathrm{zz}}$	off-specular reflectance
$D_{\mathrm{xx}} = D_{\mathrm{yy}} > 0$	retro-reflection

2.4.4 Lafortune

Lafortune et al. (1997) developed a BRDF model that generalizes the classical formulation of the cosine lobe model by Phong. Its high flexibility allows for the reproduction of a broad variety of reflection phenomena (Section 2.3) including generalized diffuse reflection, backscatter, off-specular peaks and anisotropic surfaces. The model is derived from Phong's cosine lobe expression

$$f_{\mathrm{r}} = k_{\mathrm{s}} \left\langle \mathbf{r} \cdot \mathbf{v} \right\rangle^{\gamma} = k_{\mathrm{s}} \left\langle (\mathbf{l}^{\mathrm{T}} \cdot \mathbf{M}_{\mathrm{L}}) \cdot \mathbf{v} \right\rangle^{\gamma} \tag{2.27}$$

using the Householder transformation $\mathbf{M}_{\mathrm{L}} = 2\mathbf{n}\mathbf{n}^{\mathrm{T}} - \mathbb{I}_3$. This equals the Phong model for $\mathbf{M}_{\mathrm{L}} = \mathrm{diag}(-1, -1, 1)$. Since the vector product is replaced by the vector-matrix-vector multiplication, the degrees of freedom regarding the direction of the specular highlight are increased by three. Similar to Phong's definition, the cosine lobe can be adjusted in width and strength relative to the light source by means of the exponent γ and the weight k_{s}, respectively. Yet, in Phong's definition, the highlight is always centered around the perfect reflection direction. The additional freedom of the rotation of $\mathbf{l}$ about $\mathbf{M}_{\mathrm{L}}$ enables the orientation of the specular highlight towards an arbitrary direction.

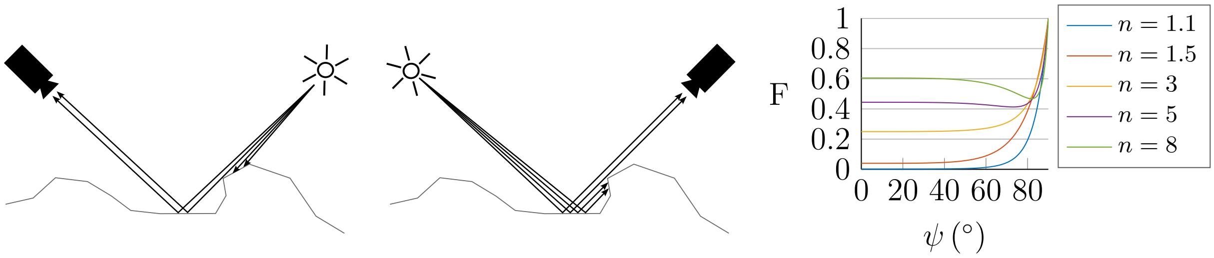
The rotation matrix $\mathbf{M}_{\mathrm{L}}$ is defined by the surface normal direction, but applying an eigenvalue decomposition (EVD) $\mathbf{M}_{\mathrm{L}} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}\mathbf{Q}$ extracts the three degrees of freedom as the eigenvalues of $\mathbf{D} = \mathrm{diag}(D_{\mathrm{xx}}, D_{\mathrm{yy}}, D_{\mathrm{zz}})^{\mathrm{T}}$. The matrix $\mathbf{Q}$ is used to transform the light and viewing direction to a new local coordinate system whose $\mathbf{x}'$- and $\mathbf{y}'$-axis are aligned with the main directions of anisotropy (if existent) and $\mathbf{z}' = \mathbf{n}$. The transformed light and viewing vectors are given by $\mathbf{l}' = \mathbf{l}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}} = (l'_{\mathrm{x}}, l'_{\mathrm{y}}, l'_{\mathrm{z}})$ and $\mathbf{v}' = \mathbf{Q}\mathbf{v} = (v'_{\mathrm{x}}, v'_{\mathrm{y}}, v'_{\mathrm{z}})$. Using all of the above, Eq. (2.27) can be rewritten as:

$$f_{\mathrm{r}}^{(\mathrm{Lafortune})} = k_{\mathrm{s}}(D_{\mathrm{xx}} l'_{\mathrm{x}} v'_{\mathrm{x}} + D_{\mathrm{yy}} l'_{\mathrm{y}} v'_{\mathrm{y}} + D_{\mathrm{zz}} l'_{\mathrm{z}} v'_{\mathrm{z}})^{\gamma} \tag{2.28}$$

As a result, the transformed illumination vectors are simply weighted by the eigenvalues of $\mathbf{M}_{\mathrm{L}}$. Table 2.2 gives a brief overview of the reflective behavior that can be assessed based on the configuration of the eigenvalues.

2.4.5 Cook-Torrance

The reflectance model described by Cook and Torrance (1981) is based on the proposal of Torrance and Sparrow (1967), which, in turn, is based on geometrical optics. Hence, in contrast to the cosine lobe models based on measurements of real-world reflectance, it is derived from physical principles directly. The surface of a given object is regarded as a collection of ideal mirror-like microfacets which contribute to the specular component if the facet's normal $\mathbf{n}'$ and $\mathbf{h}$ are aligned. The contribution of each surface point is modeled by a



(a) (b) (c)

Figure 2.8. (a) Shadowing and (b) masking on a rough surface. (c) Fresnel factor F for different refractive indices n.

probability distribution and decreases with the distance between both vectors. The roughness of the surface corresponds to the standard deviation of the distribution function, i.e., a rough surface exhibits a broad lobe, which is modeled by a wide probability distribution function, such that facets contribute to the reflectance although they are not completely aligned with the perfect reflection direction. The Cook-Torrance BRDF is given by

$$f_{\mathrm{r}}^{(\text{Cook-Torrance})} = \frac{k_{\mathrm{d}}}{\pi} + k_{\mathrm{CT}} \frac{F(n, \langle \mathbf{v} \cdot \mathbf{h} \rangle) \cdot D(m, \langle \mathbf{n} \cdot \mathbf{h} \rangle) \cdot G(\mathbf{n}, \mathbf{v}, \mathbf{h})}{\pi \langle \mathbf{n} \cdot \mathbf{v} \rangle \langle \mathbf{n} \cdot \mathbf{l} \rangle} \tag{2.29}$$

where

$$F = \frac{1}{2} \frac{(g-c)^2}{(g+c)^2} \left(1 + \frac{(c(g+c)-1)^2}{(c(g-c)+1)^2} \right) \tag{2.30}$$

$$c = \langle \mathbf{v} \cdot \mathbf{h} \rangle = \cos(\psi)$$

$$g = \sqrt{n^2 + c^2 - 1}$$

$$G = \min \left\{ 1, \frac{2 \langle \mathbf{n} \cdot \mathbf{h} \rangle \langle \mathbf{n} \cdot \mathbf{v} \rangle}{\langle \mathbf{v} \cdot \mathbf{h} \rangle}, \frac{2 \langle \mathbf{n} \cdot \mathbf{h} \rangle \langle \mathbf{n} \cdot \mathbf{l} \rangle}{\langle \mathbf{v} \cdot \mathbf{h} \rangle} \right\} \tag{2.31}$$

$$D = \frac{1}{m^2 \langle \mathbf{n} \cdot \mathbf{h} \rangle^4} \exp \left\{ -\frac{\tan^2 \vartheta_{\mathrm{h}}}{m^2} \right\}. \tag{2.32}$$

n is the refractive index and m the root mean square (RMS) of the facet slope.

The Fresnel term F is an approximation of the Fresnel equation (cf. Section 2.3.2) without consideration of the extinction coefficient κ since the authors claim only a weak angular dependence on κ (Cook and Torrance, 1981). Hence, for normal incidence ($\psi = 0 \Rightarrow c = 1 \Rightarrow g = n$) F becomes

$$F_0 = \left(\frac{n-1}{n+1} \right)^2 \tag{2.33}$$

and the approximation is exact. Figure 2.8c shows the approximated Fresnel factor in dependence of ψ for different refractive indices n.

Shadowing and masking effects on a rough surface are accounted for by the geometrical attenuation factor $G \in [0\ 1] \subset \mathbb{R}^+$. G reduces the predicted intensity with increasing incident or viewing angle since it is more likely at grazing angles that a certain amount of light does not reach the viewer as illustrated in Figs. 2.8a and 2.8b. Different probability distribution functions have been examined to model the distribution of the facet slope D.

The Beckmann-Distribution, given in Eq. (2.32) is chosen since it only depends on the facet slope m and the geometry of the given situation and does not introduce additional arbitrary constants (Cook and Torrance, 1981).

2.5 Data-driven Reflectance Models

A second popular way of reproducing the reflectance of an object in rendered images is based on high-density measurements of the material BRDF. Analytical BRDFs rely on simplifications to model the real-world behavior that is usually more complex than the model assumptions. The question is if the simplifications significantly violate or deteriorate the resulting renderings since the BRDF models depend on few parameters only and can be quickly evaluated. Data-driven BRDF reproduction, on the other hand, is a tedious procedure that has become feasible only in the last two decades with advances in computation power and memory efficiency. Storing densely sampled values of a function of up to seven variables – spatial pixel location (u, v), wavelength λ, and illumination and viewing directions $\vartheta_{\text{in}}, \phi_{\text{in}}, \vartheta_{\text{out}}, \phi_{\text{out}}$ – requires an enormous amount of storage. Furthermore, acquiring such data was complicated and cumbersome using a gonioreflectometer[4], which was the first famous device for measuring a BRDF (e.g., Erb, 1980). However, the table based reproduction of reflectance data is of less interest regarding the experiments in this thesis and therefore will be introduced only briefly.

Since the availability of cheap storage has promoted the feasibility of large databases as look-up tables for accurate BRDF measurements, researchers have advanced in presenting faster and easier acquisition techniques. Marschner et al. (1999a) proposed an acquisition device that uses a digital camera to capture 2D sensor data of cylindrical objects, in contrast to the single measurement of a gonioreflectometer per position. They apply their algorithm as well to the particular focus of the BRDF of human skin (Marschner et al., 1999b). Matusik et al. (2003b) use a similar setup to create the famous MERL database of 100 different, densely sampled materials ranging from metals to plastics and ceramics. Filip et al. (2014) employ a comparable device with two rotational arms to acquire the BRDF of anisotropic fabrics. Fabrics like silk are extraordinarily complicated to measure since their macroscopic structure creates complex anisotropic reflectance patterns.

---

[4] A spherical gantry that yields a computer controlled source and sensor that can be separately positioned to cover a wide range of illumination situations.

RELATED WORK: SURFACE RECONSTRUCTION

3

The idea of reconstructing the shape of an object from the object's appearance has been introduced and significantly influenced by Rindfleisch (1965), Horn (1970) and Woodham (1980) in the beginning of the era of computer vision and further elaborated in other publications (e.g., Horn et al., 1978; Ikeuchi and Horn, 1981; Horn and Brooks, 1986; Horn, 1990; Woodham, 1994). The following chapter will be used to shed a little light on the many variations and sub-disciplines that have emerged in the field of camera-based 3D reconstruction since these early days. Many distinct acquisition setups have been developed in conjunction with new algorithms, i.e., some facilitate the surface estimation by exploiting geometrical constraints, for example, reciprocity, others focus on specialized application scenarios. The laboratory setup used in this work was designed to put low demands on operators and equipment and therefore yields no special request regarding geometry or hemispherical coverage with incident light directions.

A fundamental and inherent flaw of all local shape-from-intensity methods, discussed already almost three decades ago by Nayar et al. (1990), is the assumption that every scene point is lit by the light source only. Every non-convex surface violates this assumption inherently since the incoming light is in general distributed from a surface point into the entire hemisphere surrounding that point. If now the connecting ray between two surface points neither is blocked nor yields angles larger than 90° with the facet normal direction, both points can be considered *visible* to one another, and the lit surface point is an additional light source to every other visible point causing so-called interreflections. Gradient-based surface estimates produce pseudo surfaces that correspond to the perceived intensity but not to the real surface since an interreflection-affected surface point is measured brighter than predicted by the BRDF models subject to the correct surface gradient (Nayar et al., 1990).

The basic statement can be extended to the assumption that every scene point *is* lit by the light source, thereby including the inaccuracies stemming from self-shadowing as well. This assertion now is valid only for flat or nearly-flat convex surfaces with little-to-no curvature. The supposition of the non-existence of both interreflections or shadowing and masking limits the applicability of such methods to real-world scenarios and industrial applications. The different portions of direct and indirect illumination that contribute to the appearance of a real-world object have been discussed already in Section 2.3.4 and illustrated in Fig. 2.6. Obviously, considering indirect illumination immensely increases the impression of realism of a rendered scene. However, a BRDF is evaluated only considering the constellation of the incident and viewing direction and computing visibility between surface points and shadowing barriers is complex and often prohibitive in reconstruction algorithms.

Up to some degree, the adverse effects of self-cast shadows and interreflections can be ignored just as the Lambertian BRDF can be a satisfactory approximation to many real surfaces. Other methods aim at identifying at least the shadow pixels to exclude them from further consideration. In any case, it is beyond question that exact knowledge of direct illumination and shadow boundaries improves the performance of every image-based surface reconstruction algorithm (e.g., Herbort et al., 2013). As the detection and elimination of both shadows and interreflections are beyond the scope of this work, this issue will be addressed only marginally.

3.1 Shape from Shading

Horn (1970) published a technique termed *shape from shading* (SfS) to tackle the inverse vision problem of reconstructing a 3D-surface from a single two-dimensional image. Each pixel $\mathbf{I}(u,v)$ yields a measurement corresponding to the irradiance of a surface point ${}^{S}\mathbf{x}(u,v)$ scaled by the camera response curve that comprises e.g., quantum efficiency, pixel size, and ADC characteristics. The definition of the term *intensity* varies among different fields of study and in this thesis it will be used to indicate the scaled radiance. Since the measurement equipment relies on CCD camera sensors that exhibit a linear relationship between incident radiant flux and the pixel value (cf. Section 4.4) this assumption is valid and the actual scaling is not of interest and neglected in all further computations. Note that this implies that the parameters of the BDRF models lose their physical meaning – if they had any – and are replaced by their corresponding scaled versions. Parameter ranges that are derived from physical constraints are affected accordingly.

Shape from shading in its original formulation is designed for Lambertian surfaces only. The "canonical setting" of the experimental setup (Herbort and Wöhler, 2011) comprises the following restrictions for the fundamental ideas of SfS to be valid:

- The object surface is strictly Lambertian;
- the surface albedo ρ is known;
- direction $\mathbf{l}$ and strength I_0 of sources of illumination are known;
- illumination distance is $\rightarrow \infty$, such that distant light sources can be assumed;
- viewing distance is $\rightarrow \infty$, such that the orthographic projection is a valid assumption;
- no shadows are cast on the surface; *and*
- no interreflections due to surface concavities exist.

Recall the definition of the BRDF and Lambert's law, Eqs. (2.10) and (2.22):

$$\mathbf{I}(u,v) = \frac{I_0}{r^2}\frac{\rho}{\pi}\left\langle \mathbf{n} \cdot \mathbf{l} \right\rangle \tag{3.1}$$

As the matrix of intensities $\mathbf{I}(u,v)$ is measured and the incident light direction $\mathbf{l}$ is calibrated and thus supposed to be known, the surface normal direction $\mathbf{n}$ and albedo ρ are the only unknowns in this equation. Unfortunately, this problem statement is ill-posed, since the normal direction yields three degrees of freedom to be determined from one intensity measurement per pixel.

$$\mathbf{I}(u,v) = \frac{I_0}{r^2}\frac{\rho}{\pi}\begin{pmatrix} l_\mathrm{x}(u,v) \\ l_\mathrm{y}(u,v) \\ l_\mathrm{z}(u,v) \end{pmatrix}^\mathrm{T}\begin{pmatrix} n_\mathrm{x}(u,v) \\ n_\mathrm{y}(u,v) \\ n_\mathrm{z}(u,v) \end{pmatrix} \tag{3.2}$$

Under the assumption that the albedo ρ is known and reformulating the normal vector in terms of the surface gradient fields, since the gradient is the first partial derivative of the surface z with respect to x and y,

$$p = \frac{\partial z}{\partial x} = z_x \quad \text{and} \quad q = \frac{\partial z}{\partial y} = z_y, \tag{3.3}$$

yields a similar vector $\mathbf{n}^* = (p, q, -1)^\mathrm{T}$ that has the same direction as $\mathbf{n}$ and can easily be normalized to unit length afterwards.

$$\mathbf{n} = \frac{1}{\sqrt{1+p^2+q^2}}\begin{pmatrix} p \\ q \\ -1 \end{pmatrix} \tag{3.4}$$

Hence, it is possible to reduce the number of unknowns by one, although still not allowing a unique solution. To make a solution feasible certain regularizations are necessary, e.g., a smoothness constraint for the surface or a partially known gradient as is the foundation of *photoclinometry* (e.g., Jankowski, 1997).

3.1.1 Variational Approaches to Shape from Shading

Horn and his coworkers extended their research regarding the extraction of shape information from image data, (e.g., Ikeuchi and Horn, 1981; Horn and Brooks, 1986; Horn, 1990), to find possible solutions to the shape-from-shading problem. Note that the measured radiance $\mathbf{I}$ and the predicted reflectance map $\mathbf{R}$ can be defined both in sensor coordinate space $\mathbf{I}(u,v)$ and 3D world coordinates $\mathbf{I}(x,y)$[1], where the measurements (or predicted reflectance values) correspond to a certain point on the object surface. This can be read as "the intensity measurement at pixel (u,v)" and "the intensity measurement of the surface point at (x,y,z)" which in turn is captured at pixel (u,v). The latter formulation provides more flexibility when multi-camera systems are used since the same surface point is captured at different sensor locations in the different cameras. Since the literature on Shape from Shading follows the definition in 3D coordinates, this will be applied here in what follows.

The principal objective is always to find a (smooth or integrable) surface $z(u,v)$ that explains $\mathbf{I}(x,y)$ utilizing a certain reflectance map $\mathbf{R}(x,y)$. This reflectance map is computed from the chosen BRDF model (or modeling technique), which in turn depends on the surface geometry. Given a calibrated acquisition environment, the remaining unknown is the surface normal vector that can also be expressed in terms of the surface gradient fields $(p(x,y), q(x,y)$ and the true surface should satisfy $\mathbf{I}(x,y) = \mathbf{R}(p(x,y), q(x,y), \mathbf{l}, \mathbf{n}, \boldsymbol{\theta})$, i.e., the measured intensity values should be as similar as possible to the predicted ones.

---

[1] Note that the dependency on z might be neglected since the resulting image only has two dimensions.

However, this is never going to be the case for multiple reasons including noise in the image acquisition process, discrete approximations of the gradient fields and non-ideal BRDF models. Now, to minimize an integral equation evaluating the squared differences between image and reflectance map is the better choice to find a surface that is very close to the true surface.

$$\iint (\mathbf{I}(x,y) - \mathbf{R}(p(x,y), q(x,y), \mathbf{l}, \mathbf{n}, \boldsymbol{\theta}))^2 \,\mathrm{d}x\,\mathrm{d}y \tag{3.5}$$

Unfortunately, there exists an infinite number of solutions that minimizes Eq. (3.5) and further constraints and regularizations are required to enable a unique solution. Horn and Brooks (1986) also noted that it appears to improve convergence to add a penalty term based on a constraint instead of imposing the constraint itself on a functional. Ikeuchi and Horn (1981) proposed to add a functional that measures the *lack of smoothness* of a surface

$$\iint (f_x^2 + f_y^2 + g_x^2 + g_y^2) \,\mathrm{d}x\,\mathrm{d}y \tag{3.6}$$

to the irradiance equation (3.5), where f_x indicates the partial derivative of f with respect to x and

$$f = \frac{2p}{1+\sqrt{1+p^2+q^2}} \quad \text{and} \quad g = \frac{2q}{1+\sqrt{1+p^2+q^2}}. \tag{3.7}$$

Nevertheless, this leads to an over-smoothing of the resulting surfaces and makes the iterative scheme even converge to an incorrect solution when initialized with the correct solution, since a correct solution does not necessarily coincide with maximum flatness as is implied by Eq. (3.6).

The constraint of *integrability*, that is the surface corresponds to an integrable gradient field $(p(x,y), q(x,y))$, is introduced by Horn and Brooks (1986). The condition

$$\frac{\partial p}{\partial y}(x,y) = \frac{\partial q}{\partial x}(x,y) \tag{3.8}$$

leads to the functional

$$\iint (z_{xy} - z_{yx}) \,\mathrm{d}x\,\mathrm{d}y \tag{3.9}$$

which does not require the explicit knowledge of the gradient fields to ensure integrability.

Horn (1990) further elaborated on the integrability constraint to fuse the recovery of height and gradient in one scheme in the case where the gradient field, in fact, is known. The partial derivatives of the surface then shall equal the gradient field to yield an integrable surface, which can be expressed as

$$\iint (z_x - p)^2 + (z_y - q)^2 \,\mathrm{d}x\,\mathrm{d}y. \tag{3.10}$$

Calculus of Variations

Consider a function $F(x, y, z(x,y), z_x(x,y), z_y(x,y))$ that accounts for the distance between a surface $z(x,y)$ and an optimal solution at a point (x,y). The goal is to minimize F in an area Ω with respect to the surface $z(x,y)$

$$\min_{z(x,y)} \iint_\Omega F(x, y, z(x,y), z_x(x,y), z_y(x,y)) \,\mathrm{d}x\,\mathrm{d}y. \tag{3.11}$$

As F is dependent on z and its partial derivatives $\frac{\partial}{\partial x}z(x,y) = z_x(x,y)$, $\frac{\partial}{\partial y}z(x,y) = z_y(x,y)$, the search has to be conducted over a space of functions and the minimization is a problem in the calculus of variations. The well known Euler-Lagrange equation (e.g., Bronstein et al., 2016) to find the extrema of F is

$$F_z - \frac{\partial}{\partial x}F_{z_x} - \frac{\partial}{\partial y}F_{z_y} = 0. \tag{3.12}$$

Applying this to Eq. (3.10), the Euler-Lagrange equation corresponding to the variational problem is

$$\Delta z = p_x + q_y \tag{3.13}$$

where $p_x = \frac{\partial p}{\partial x}$ and $q_y = \frac{\partial q}{\partial y}$ and $\Delta = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$ is the Laplace operator. Since this involves solving partial differential equations, the problem is ill-posed when there are no boundary conditions available. Luckily, the calculus of variations provides a specialized form of boundary conditions in the case where there is no additional knowledge about the boundaries of the problem statement. These so-called “natural boundary conditions” for the boundary of the region Ω are in this case

$$\mathbf{n}_\Omega^{\mathrm{T}} \begin{pmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{pmatrix} = \mathbf{n}_\Omega^{\mathrm{T}} \begin{pmatrix} p \\ q \end{pmatrix} \tag{3.14}$$

if $\mathbf{n}_\Omega^{\mathrm{T}}$ is a vector, perpendicular to the boundary.

It is possible to derive an iterative update scheme for the desired variables, say (p, q, z) in (Horn, 1990), by expanding the Euler equations and gathering the respective terms. This approach will be presented in a more elaborated way in Section 7.2 in conjunction with a new error functional that is to be minimized for a stereo camera setup. For further insight into the calculus of variations, the reader is referred to the literature (e.g., Kielhöfer, 2010; Liberzon, 2012).

3.2 Photometric Stereo

In an endeavor to overcome the strict need for the regularizations of Horn’s original shape from shading, Woodham (1980) proposed to capture $N = 3$ images $\mathbf{I}(u,v)_i$ with $i \in \{0 \ldots N-1\}$ of the same object illuminated under different conditions. In the style of binocular stereo, where there should be no change in the lighting during acquisition of both images, he termed the use of multiple light sources *photometric stereo*. Assuming a matte, convex, smooth and continuous surface, the use of a Lambertian BRDF seems reasonable and the following equation arises:

$$\mathbf{I}_i(u,v) = \frac{\rho}{\pi} \cdot \hat{I}_{0_i} \cdot (n_{\mathrm{x}} l_{\mathrm{x}_i} + n_{\mathrm{y}} l_{\mathrm{y}_i} + n_{\mathrm{z}} l_{\mathrm{z}_i}) \tag{3.15}$$

Here, $\hat{I}_{0_i}$ are the entries of a diagonal matrix $\hat{\mathbf{I}}_0$ corresponding to the intensity of each light source scaled by the inverse squared distance between light source and object. Note that n_{x} and l_{x_i} depend on their respective location (u, v). Yet, this is omitted here for the sake of readability. A more general formulation of photometric stereo favors the use of $N \geq 3$ images and thus increases robustness to outliers and image noise. From Eq. (3.15) follows a system of linear equations for every surface normal:

$$\begin{pmatrix} I_0 \\ I_1 \\ \vdots \\ I_{N-1} \end{pmatrix} = \frac{\rho}{\pi} \underbrace{\begin{pmatrix} \hat{I}_0 & 0 & \cdots & 0 \\ 0 & \hat{I}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{I}_{N-1} \end{pmatrix} \begin{pmatrix} l_{\mathrm{x}_0} & l_{\mathrm{y}_0} & l_{\mathrm{z}_0} \\ l_{\mathrm{x}_1} & l_{\mathrm{y}_1} & l_{\mathrm{z}_1} \\ \vdots & \vdots & \vdots \\ l_{\mathrm{x}_{N-1}} & l_{\mathrm{y}_{N-1}} & l_{\mathrm{z}_{N-1}} \end{pmatrix}}_{\mathbf{L}'} \mathbf{n} = \frac{\rho}{\pi} \hat{\mathbf{I}}_0 \mathbf{L} \mathbf{n} \tag{3.16}$$

Multiplying the rows of $\mathbf{L}$ with the matrix of scaled intensities $\hat{\mathbf{I}}_0$ and the inversion of the resulting matrix $\mathbf{L}'$ enables the estimation of the surface normal direction that caused the measured combination of intensities. This requires the matrix $\mathbf{L}$ to be of rank 3 and therefore the incident light directions must not be coplanar. If $N > 3$, the system is overdetermined and can be solved using a pseudo inverse matrix $\mathbf{L}^+$ (e.g., Penrose, 1955),

$$\mathbf{L}' \cdot \mathbf{L}^+ \cdot \mathbf{L}' = \mathbf{L}' \tag{3.17}$$

which solves the problem in a least squares sense. The final computation of the surface normal thus becomes

$$\mathbf{n} = \frac{\pi}{\rho} \cdot \mathbf{L}^+ \cdot \mathbf{I}. \tag{3.18}$$

If the albedo is unknown it can easily be absorbed into the surface normal vector and normalization of the vector directly yields the pixelwise albedo.

3.3 Generalized Lambertian Photometric Stereo

The confinement of Woodham's photometric stereo to diffuse reflecting surfaces under a well-calibrated acquisition setup, on the one hand, facilitates the problem of surface estimation drastically. On the other hand, of course, this limits the applicability to very few situations, mostly because Lambert's law is violated by many materials more or less severely. Furthermore, the well-calibrated setup is hard (up to impossible) to achieve for example in outdoor environments. Naturally, there has been extensive research in the computer vision community on a grand variety of approaches to loosen or circumvent this limitation. The following section aims to give a brief overview of the most relevant to the subjects of this thesis.

3.3.1 Diffuse-Enforced Reflectance

As the possibility of a linear reflectance function is a massive advantage of the Lambertian formulation, one of the approaches is to *enforce* the validity of the diffuse assumption onto the image data. Non-diffuse effects can thus be regarded as outliers to the correct model. However, this ignores possibly valuable additional information in the first place and, additionally, requires a significant amount of input data to ensure the solvability of the inverse vision problem in the face of many rejected data points.

Mallick et al. (2005) propose a data-dependent color space transformation to separate diffuse and specular reflection based on the dichromatic reflectance model (Shafer, 1985). They propose a linear coordinate transformation from RGB to SUV to maintain the photometric properties of the acquired image data while aligning the main axis of the new color space to the effective light source color. The projection of the reflectance vectors onto the plane spanned by the remaining two axis yields the diffuse component of the reflection. White objects or saturation caused by high specularities violate the central assumption of color-separability and distort the estimated surface gradient. Mukaigawa et al. (2007) eliminated the necessity for different colors of surface and light source and extended their previous work (Mukaigawa et al., 2001) on *photometric linearization*, a technique that renders ideal diffuse images from real images including specularities and shadows. The key concept is that a diffuse image under infinitely distant illumination can be composed as the linear superposition of three diffuse base images. The adequate weights are selected via random sampling (RANSAC) from the set of input images, followed by a center of gravity computation and outlier detection. The comparison of the linearized image and the acquired image allows for the classification of the input image pixels into diffuse, specular, attached shadow and cast shadow. Still, a narrow specular lobe is required to enable distinction between shading caused by surface slope and specularity, which makes these approaches inapplicable on metallic surfaces.

3.3.2 Outlier Rejection Techniques

A variety of approaches aims to handle all non-Lambertian effects of a scene directly as outliers from the diffuse model. While this has the advantage that cast shadows can be detected and treated, generally a large number of input images is required to yield stable outlier detection.

Adding a fourth incident direction is proposed by Coleman and Jain (1982) assuming that only one of the four light sources will result in a specular highlight for a particular surface point. If this is correct, the specular highlight will result in a higher magnitude of the corresponding surface normal and can thus be eliminated via thresholding. However, it is evident that these assumptions are built on a fragile basis. A similar idea has been pursued by Barsky and Petrou (2003) to cope with the appearance of shadows and specular highlights.

Motivated by the difficulties in object recognition caused by the changes in object appearance by changes of the lighting conditions, Belhumeur and Kriegman (1998) investigate the possible variety of images that represent the same object under different illumination. They find that all images of a diffuse object form a convex cone in $\mathbb{R}^n$, where n is the number of image pixels, that can be approximated by low order subspaces. Their work encouraged many researchers to explore the possibilities of matrix factorization in photometric stereo.

Sunkavalli et al. (2010) use RANSAC clustering to find subspaces of the visibility space of all light sources to exclude shadows from the estimation of Lambertian photometric stereo. Three randomly chosen surface points under three different (including uncalibrated) light directions yield an $m \times n$ intensity matrix $\mathbf{I}$ that contains m surface points illuminated by n light sources. $\mathbf{I} = \mathbf{L}^{\mathrm{T}}\mathbf{N}$, where $\mathbf{L}$ and $\mathbf{N}$ contain the light and surface normal directions, can be factorized using singular value decomposition (SVD), i.e., $\mathbf{I} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$, to approximate normals and light directions as $\hat{\mathbf{N}} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}$ and $\hat{\mathbf{L}}^{\mathrm{T}} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$, respectively. By maximizing the set of inliers that can be explained based on the estimated light direction $\hat{\mathbf{L}}_k$ corresponding to subset S_k the different visibility subspaces are estimated. Surface normals can be estimated from light directions and visibility information (up to an ambiguity, if the light directions are not calibrated a-priori).

In a similar fashion Wu et al. (2011b) use matrix factorization to find a low-rank submatrix to the intensity matrix and an error matrix that accounts for shadows and specularities. They rely on recent advances to solve the matrix rank minimization problem that is NP-hard in general. While the mathematical explanations are sound, the necessity of the procedure is not evident as they only compare their results to (and out-perform) least-squares diffuse photometric stereo. However, the works of both Sunkavalli et al. (2010) and Wu et al. (2011b) cannot be applied to the domain of non-Lambertian photometric stereo as it is used in the scope of this dissertation since the intensity matrix that is to be factorized can be formed only through the linearity of the Lambertian reflectance model.

Ikehata et al. (2014) propose a Bayesian model that decomposes the appearances of an object into a piecewise continuous linear diffuse component and a non-diffuse component that comprises all possible corruptions of the ideal diffuse surface including specularities and shadows. Their version of the dichromatic reflectance model comprises an error term that is merged with the specular component. Assuming monotonicity, the diffuse reflectance function is represented as the weighted sum of piecewise linear functions, each valid on a specific interval, thus increasing the generality of the diffuse component beyond the Lambertian model. The main limitation of their method is the necessity of a dominant diffuse basis of the reflective behavior that is not existent on metals or many natural materials.

3.4 Non-Lambertian Photometric Stereo

The previous section elaborated on recent research with the primary goal to be able to apply photometric stereo in its original form, that is based on a pure Lambertian reflectance model. To achieve this every non-diffuse effect of reflection has to be detected and eliminated. Another important approach to handling arbitrary BRDF is the use of more sophisticated BRDF models, again to minimize the difference between reflectance map and captured image. For many years developers have come up with new ideas on how to realistically render scenes based on physics and empirical studies alike (cf. Section 2.4). Initially motivated by computer graphics, the transition from creating scenes to analyzing real-world images started in the 1990s. The publications of novel BRDF models usually already contain proof of their efficiency in modeling real-world reflectance behavior, (e.g., Phong, 1975; Blinn, 1977; Cook and Torrance, 1981; He et al., 1991; Ward, 1992; Lafortune et al., 1997; Ashikhmin and Shirley, 2000). Some of their adoptions by other researchers are presented in the following parts of this section.

Georghiades (2003) uses the Torrance-Sparrow BRDF model to reconstruct the shape of human faces and shows that it is possible (in contrast to a pure Lambertian BRDF) to resolve the generalized bas-relief (GBR) ambiguity (e.g., Belhumeur et al., 1999) based on a minimum of 4 independent incident directions. While the albedo is spatially varying, the specular weight and surface roughness parameters are constant throughout the entire image.

Chung and Jia (2008) present a variation of photometric stereo to address glossy surfaces with a broad specular lobe. They aim to overcome the problem of specular component removal, which proved difficult on rough specular surfaces since the boundaries of the specular lobe are blurred. In contrast to previous work in the field of reflectance separation that is mostly based on color information (e.g., Mallick et al., 2005; Sato et al., 1997), the objects are allowed to have arbitrary texture but are restricted to relatively simple convex shapes. A Ward BRDF models the object reflectance. Diffuse weight, specular weight, and surface roughness are optimized separately while fixing the others. Their main contribution is the estimation of few initial surface normals using a shadow boundary cue. The angle between the surface normal and the incident light direction at the shadow boundary has to be zero, $\langle \mathbf{n} \cdot \mathbf{l} \rangle = 0$. Since sharp boundaries are often not to be found in real-world data, a set of candidates in the transition between bright pixels and cast shadow is considered for every surface normal. The specular parameters (weight, and cosine exponent) that are computed during the candidate normal estimation are clustered using mean shift.

3.4.1 Example-based Photometric Stereo

The early work of Silver (1980) inspired by Woodham (1977) to use images of a wooden sphere to estimate the 3D shape of other wooden objects captured under the same lighting was pursued and adapted by various researchers more than twenty years later. Hertzmann and Seitz (2005) describe and popularize the idea of *orientation consistency* that states that “under the right conditions, two points with the same surface orientation must have the same or similar appearance in an image”. This fact can be exploited when there is an object in the scene whose shape is known and reflectance properties are equal to the object under inspection which founds the new sub-genre of *example-based* photometric methods on the basis of the work by Silver (1980).

The additional requirement of a reference object of the same material in the scene (which is not always feasible) is leveled by the minimal demand for calibration of the environment, since illumination and BRDF may be unknown without hindering the reconstruction. For an object of unknown shape and its reference object, multiple images are captured from the same viewpoint and different illumination. More available images per surface point increase the stability of the solution. Since spheres yield a broad distribution of surface orientations without any ambiguities for a fixed viewing and incident direction, and their shape is described correctly regarding mathematics, they have emerged as the ideal reference object. The surface normal of the unknown object is found merely by searching for the vector of intensities on the reference object that best matches the vector on the unknown surface point $\mathbf{x}$ and assigning the reference normal direction to that point.

Apart from applying the reference object reconstruction to various materials, Hertzmann and Seitz (2005) present two extensions of the work of Silver (1980). They show that anisotropic materials can be reconstructed as well if the reference object contains a sufficient number of different local surface orientations for the same surface normal. Additionally, they consider non-uniform surfaces to be composed of a known number of multiple materials of which each is composed of a different mixture of diffuse and specular weights. The virtual materials that form the resulting reference objects are composed of a diffuse and a specular sphere and corresponding weights in the RGB color space.

Ackermann et al. (2012) extend the work of Hertzmann and Seitz (2005) by replacing the explicit need for reference objects in the scene with their concept of *scene-intrinsic reference geometry* (SIRG). Thus, they are able to apply their procedure to wide baseline stereo in general outdoor settings that for example are encountered when analyzing aerial images. They use a partial reconstruction of the scene, obtained from multi-view stereo and then transfer normals from reconstructed to unreconstructed regions. This bootstrap limits their method to objects of relatively similar reflectance properties. They capture different sets of images for photometric stereo (PS) and multi-view stereo (MVS) and register them via structure from motion. They rely on the volumetric range image processing (VRIP) developed by Curless and Levoy (1996) to merge multiple depth maps and smooth the normals with an iterative Laplacian smoothing to create the reference geometry of the scene.

Due to errors and simplifications, the directly matched normals show significant artifacts, which can be greatly reduced by transferring the average of the 50 best matching normals to the filled-in regions.

3.4.2 Non-uniform Reflectance Properties

A huge limitation on the flexibility of surface reconstruction algorithms is the assumption of uniform reflectance properties. The naive algorithmic realization considering analytic BRDF models is relatively simple: The chosen BRDF model yields a different set of parameters for every pixel or surface patch or group of pixels that are supposed to belong to the same material. However, if analytic reflectance models are used, non-uniformity greatly decreases the stability of the parameter estimation process without further measures and at the same time increases the computation time required for the model fitting[2]. Additionally, the strict allocation of patches or pixel groups to a certain set of parameters is set to fail in the presence of subtle changes in the reflectance properties or blending of two (or more) materials. A lot of research that has been conducted regarding spatially varying reflectance properties is in fact still limited to diffuse surfaces just with non-uniform albedo. Exemplary research that goes beyond this simple version of non-uniform reflectance is presented in the following.

Lensch et al. (2003) compute realistic BRDF model representations of real-world materials from 15 to 25 camera images and known 3D structure. They introduce a data structure called *lumitexel* that contains 3D position $\mathbf{x}$ and normal direction $\mathbf{n}$ for each surface point. Additionally, each lumitexel is linked to a list of reflectance samples and their corresponding incident light and viewing direction. An isotropic version of the Lafortune BRDF model is used to model the reflectance properties of the object. The 3D shape of the objects is provided by a structured light scanner or a computer tomography scan. Since spatially varying material properties are explicitly considered, the lumitexels are split into two clusters based on the largest eigenvalue of the covariance matrix of the nonlinear fitting process, that is the direction of strongest variation in the lumitexel parameter space. Then the lumitexels are redistributed to the cluster for which the intensity and color error is smallest, and the cluster BRDFs are refit. The split-recluster-fit procedure is iterated until all clusters converge. In total, up to 5 different Lafortune terms are used to produce the reflectance map and refine the initially supplied 3D information.

Many researchers lean towards the ideas of example based PS with a linear mixture of a certain number of base reflectances that is supposed to be able to explain a mixture of those materials or even unknown data. Goldman et al. (2005) constrain the spatially-varying reflectance properties of an unknown object to a pixelwise mixture of fundamental materials represented by Ward BRDFs whose parameters are uniform across the image. Thus, only the material weights η_i are used to compose a spatially varying BRDF, and since $\eta_1 + \eta_2 = 1$ this means only one unknown variable, in contrast to per-pixel sets of BRDF parameters that have been used in this thesis. The optimization alters between the global BRDF parameters and the local weights and normal directions. The normal maps are initialized with Lambertian photometric stereo, shadows and specular highlights are manually rejected for the initialization, and updated employing the mixture of Ward BRDFs under known distant lighting and constant viewing direction. The representation of the reflectance with an analytical model allows for easy editing of rendered images.

---

[2] This is, of course, a problem that almost cries out for parallelization given a high level of non-uniformity.

Alldrin et al. (2008) build on their previous work (Alldrin and Kriegman, 2007) and simultaneously estimate gradient and spatially-varying reflectance. Since analytic BRDF models pose a limitation of their applicability to a specific range of materials, they propose reflectance representation as a linearly weighted combination of basis reflectances similar to (Matusik et al., 2003a; Goldman et al., 2005; Lensch et al., 2003). However, in contrast to Lensch et al. (2003) and Goldman et al. (2005), who use the parametric BRDF models of Lafortune and Ward, respectively, to create the basis BRDF, Alldrin et al. (2008) propose a non-parametric bi-variate image formation model that consists of material weights and basis BRDFs. The algorithm requires a particular setup of (at least) a rotationally symmetric cone of light sources centered about the viewing direction to compute the surface normal estimates. From an initial estimate of the surface normals based on the symmetry of the plane spanned by the surface normal and viewing direction the algorithm iteratively optimizes BRDF, material weights, and normal estimates while keeping the other ones fixed in a procedure very similar to alternating constrained least squares (ACLS) (Lawrence et al., 2006). A-priori knowledge of the number of BRDFs is necessary to pose a stable environment for the optimization.

A dictionary-based approach for simultaneously estimating shape and reflectance of an object is derived by Hui and Sankaranarayanan (2017). To reduce the complexity of the estimation process, the authors assume that the pixelwise BRDF is a weighted combination of a small number of reference BRDFs taken from the MERL database (Matusik et al., 2003a). Inspired by Hertzmann and Seitz (2005), the reference objects used as a look-up table for the correspondence of normal directions and perceived intensity are virtual spheres rendered from measurements of the MERL database. A coarse-to-fine search is implemented to reduce computational complexity, where the sampling of the normal maps of the reference spheres increases in each step and the search is refined only in the vicinity of the previous match. The light-stage described by Einarsson et al. (2006) inspired the virtual light sources and their method explicitly focuses successfully on complex real-life scenes and includes metallic surfaces.

3.5 A Brief Outlook on the Variety of Photometric Reconstruction Methods

There exists a broad variety of algorithms derived from the ideas of Horn and Woodham that are only loosely connected to this dissertation and therefore will only be hinted at in the remainder of this short section.

3.5.1 Uncalibrated Photometric Stereo

Many researchers have tried to overcome the strict limitation for a well-defined acquisition setting (e.g., Hayakawa, 1994; Chandraker et al., 2005; Sunkavalli et al., 2010; Quéau et al., 2017). This includes simply incomplete calibration scenarios like missing source intensity values and no prior knowledge of the illumination conditions alike. The possibilities range from on-the-fly calibration of incident light to surface estimation that can be conducted up to a linear transformation only. As none of these techniques is considered in the scope of this work the interested reader is referred to the listed publications.

A general problem in uncalibrated Lambertian PS settings is that there is an equivalence class of surfaces that produce the same images given the corresponding (unknown) incident light direction. Resolving the generalized bas-relief (GBR) ambiguity is crucial to deduce the correct surface structure from the given images and there exist many different ways to do so, e.g., Hayakawa (1994) show that a constant diffuse albedo is sufficient and Chandraker et al. (2005) employ interreflections to determine a unique solution.

3.5.2 Additional Variational Approaches to Shape from Shading

Chandraker et al. (2007) use a graph cut algorithm to minimize the residual error of Lambertian photometric stereo by swapping shadow labels. They employ the Hamming distance as a smoothness operator between neighboring shadow label vectors. The computed surface normal directions are integrated using Horn's integrability constraint and an additional shadow constraint based on shadow graphs (Yu and Chang, 2005). For a given incident light direction, the surface can be parted based on shadow edges into intervals that are either always or never shadowed and thus provide inequalities that constrain the height function.

Oswald et al. (2009) propose a variational method in the exclusive domain of single view surface reconstruction without exploiting shading information. They minimize an energy function that includes a smoothness term and a data term that realizes volume inflation and maintains consistency with an a-priori computed object silhouette. The convex optimization is implemented on graphics hardware and allows close-to-realtime computation of surface estimates. However, the algorithm requires much user interaction to enable foreground and background separation and to cope with over-smoothing of discontinuities.

Quéau et al. (2017); Quéau et al. (2017) use the same alternating re-weighted least-squares (ARLS) algorithm to optimize a non-convex variational formulation of photometric stereo under the assumption of uncalibrated or inaccurately calibrated light sources. Surface information, albedo, and lighting are estimated in alternating fashion from a reasonable initialization. Quéau et al. (2017) consider Lambertian surfaces only and Quéau et al. (2017) regard specularities and non-Lambertian surfaces effects like outliers in the model space. To make this approach feasible a high number of input images is required.

3.5.3 Specialized Acquisition Setups

Capturing and reproducing the reflectance properties of human skin has received much attention in the last decade, for example in photorealistic rendering for computer-generated imagery (CGI) in motion picture post-production to create virtual persons that appear entirely realistic (e.g., Marschner et al., 1999b; Angelopoulou, 1999; Donner and Jensen, 2006; Li and Ng, 2009). Apart from reproducing the appearance, researchers, first and foremost the team led by Paul Debevec (Debevec et al., 2000; Einarsson et al., 2006; Vlasic et al., 2009), have created sophisticated and highly specialized acquisition equipment to capture the 3D structure of organic objects including the motion of performing actors.

The polarization of the reflected light changes depending on the type of reflection, i.e., subsurface scattering or interface reflection (cf. Section 2.3.1). Diffusely reflected light is (almost) unpolarized regardless of the polarization state of the incident while the polarization of the reflected light is determined by the polarization of the incident light for specular reflections. Therefore, the separation of diffuse and specular reflection component is conducted in many approaches using polarized light or linear polarization filters in front of camera and light source (e.g., Wolff, 1989; Wolff and Boult, 1991; Nayar et al., 1997; Debevec et al., 2000; Chen et al., 2007).

Ma et al. (2007) propose a circular polarization pattern such that the state of polarization is independent of the emitting light source and multiple light sources can be used simultaneously. They capture individual diffuse and specular normal maps and fuse them with low-pass filtered structured light scans to achieve high-quality depth estimates of unmoved organic materials. Additionally, the applicability of the specular normal estimation on a polished obsidian object is demonstrated by indirectly reflecting light from a source onto the object and thus creating gradient illumination. This acquisition technique was published in Peers et al. (2006). Debevec et al. (2000) further noted that diffuse normal maps lack high-frequency detail if the surface exhibits a significant amount of translucency as is typical for human skin.

Vlasic et al. (2009) focus on the capture of dynamic scenes using a sophisticated acquisition setup called *lighting dome* that was initially built by Einarsson et al. (2006). This device yields 1200 light sources and captures eight views at 240 Hz of human-size working volumes resulting in an effective frequency of 60 Hz moving meshes since optical flow across four consecutive frames is used to compensate the motion. The processing pipeline computes a normal map for each view and performs multi-view matching on the integrated normal maps. Neighboring meshes are corrected with thin-plate splines and merged with volumetric range image processing (VRIP) (Curless and Levoy, 1996). They define a set of eight illumination patterns inspired by Ma et al. (2007) that is repeatedly captured and used to estimate one set of surface normals assuming Lambertian reflectance. To avoid radiometric calibration of the light sources, a diffuse grey sphere is captured to establish the conversion from radiance to normal direction analogously to Hertzmann and Seitz (2005). The lookup key is normalized to cancel the effect of varying albedo, and the initial guess for the normal estimation of each view is based on and roughly constrained by the visual hull. That, in turn, is also combined with two heuristics (maximum of normal gradient and local integrability) to detect depth discontinuities and prevent massive distortions in the integrated surfaces.

A specialized 3D scanner that simultaneously acquires reflectance and shape of still objects is described by Holroyd et al. (2010). They use beam-splitters and mirrors to align viewing direction and light direction of a modulated sinusoidal signal and an industry camera. Two of these devices are mounted to a four-arm spherical gantry and allow the acquisition of multiple pairs of high-frequency modulated reciprocal images. They compute dense depth maps for each view and merge them into a watertight model of high accuracy.

The demands regarding time and the complicated setup of those specialized assemblies can be somewhat prohibitive to an extensive usage, thus Aittala et al. (2013) highlight the practicability of their setup and focus on the easy and mostly automated acquisition of high-resolution spatially-varying isotropic BRDFs. They confine themselves to a single viewpoint and use a planar light source (an LCD screen for example) and concentrate the BRDF sampling on the mirror direction. Of course, this introduces a lack of accuracy but still covers a majority of real-world scenarios. They further reduce the number of required images by sampling in the Fourier domain.

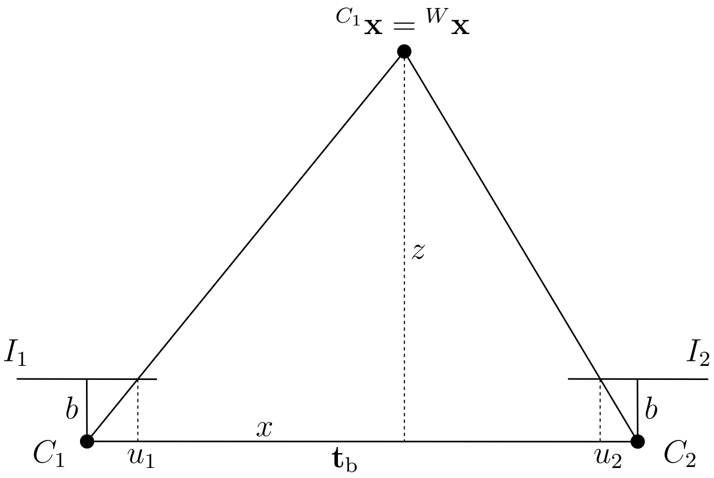


Figure 3.1. Stereo image analysis in standard geometry, the image planes I_1, I_2 are projected in front of the optical centers (C_1, C_2). The world coordinate system is aligned with the camera coordinate system of the left camera. The baseline $\mathbf{t}_\mathrm{b}$ is the translation between both cameras computed from the extrinsic calibration.

3.6 Surface Reconstruction from Triangulation

The main drawback of single camera systems is the uncertainty about a possible solution height function $z(x, y)$ that yet satisfies all constraints. An object that is twice the size and twice as far away[3], so generally speaking a scaling of the entire scene, would yield an identical image. Therefore, it is not possible to derive absolute depth information from a single image without any further knowledge, for example, the size of a reference object. Calibrated triangulation-based approaches overcome this ambiguity, as it is reduced to a scaling of the 3D model with the camera constant. The camera constant, in turn, is a manufacturing parameter and can either be derived from a data sheet or the camera calibration.

3.6.1 Stereo Image Analysis

In standard geometry the optical axes of both cameras are parallel and principal distance, image plane, and intrinsic parameters are identical for both cameras. Furthermore, the image planes are orthogonal to the optical axes and parallel to the vector connecting the optical centers, that is the stereo baseline (Horn, 1986). Although these strict requirements are usually not met in a real-world application, they can be virtually fulfilled in a processing step called *image rectification* if the intrinsic and extrinsic calibration of the stereo cameras is known. Figure 3.1 displays an image pair in standard geometry. The most advantageous property of the rectified images is that the search for image points corresponding to the same scene point is constrained to equal image rows since each pair of image rows now corresponds to a pair of matching epipolar lines. This increases algorithmic simplicity and efficiency of memory accesses alike. Since rectification can easily be achieved with a variety of existing software, rectified images will be assumed throughout the remainder of this section. In Fig. 3.1 the image planes are projected in front of the optical centers (C_1, C_2) which directly implies the inversion of the mirroring-effect of the camera.

---

[3] With the same incident irradiance and direction at every surface point.

(u_1, v_1) and (u_2, v_2) are the pixel positions in two corresponding views that have been associated with the same scene point ${}^{W}\mathbf{x}$. From simple geometrical considerations, assuming the pixel extents in millimeters $k_\mathrm{u} = k_\mathrm{v}$ to be equal and using this quantity to scale the pixel coordinates to a metric unit, and Eq. (2.2) it is obvious that

$$\begin{aligned} \frac{z}{b} &= \frac{x}{u_1/k_\mathrm{u}} \\ \frac{z}{b} &= \frac{x - \|\mathbf{t}_\mathrm{b}\|_2}{u_2/k_\mathrm{u}} \\ \frac{z}{b} &= \frac{y}{v_1/k_\mathrm{u}} = \frac{y}{v_2/k_\mathrm{u}} \end{aligned} \tag{3.19}$$

from which the 3D coordinates of ${}^{W}\mathbf{x}$ can be derived (e.g., Horn, 1986; Wöhler, 2013):

$$\begin{aligned} x &= \|\mathbf{t}_\mathrm{b}\|_2 \frac{u_1}{u_1 - u_2} \\ y &= \|\mathbf{t}_\mathrm{b}\|_2 \frac{v_1}{u_1 - u_2} \\ z &= \|\mathbf{t}_\mathrm{b}\|_2 \frac{b}{(u_1 - u_2)/k_\mathrm{u}} = \|\mathbf{t}_\mathrm{b}\|_2 \frac{b}{\delta/k_\mathrm{u}} \end{aligned} \tag{3.20}$$

$\delta = u_1 - u_2$ is the disparity, the difference between the corresponding pixel coordinates.

Hence, it is possible to obtain accurate 3D information via triangulation if the corresponding points (and thus the disparity) is known. Apparently, the real challenge in stereo vision is the correct deduction of the matching points.

Correlation-based Matching

In its traditional formulation stereo vision always relies on the principle of *photo consistency*, that is an object should appear equally bright from the two different camera views, such that it is possible to find corresponding points based on intensity cues (e.g., Seitz et al., 2006). Specular reflections and other non-diffuse effects prevent the identification of the same scene point in both images and this demand limits the suitability of stereo vision to Lambertian surfaces (cf. Section 2.4.1). However, even on surfaces that strictly satisfy these constraints the ambiguity of a single pixel intensity is too high to be identified uniquely in the second image[4]. To overcome the ambiguity, template matching with an appropriate index of similarity and window-size is a widely used solution. The discrete normalized cross-correlation (NCC) is one of these measures of similarity (e.g., Lewis, 1995).

$$\Gamma(u,v)_\mathrm{NCC} = \frac{\sum_{m,n}\left[\mathbf{I}(m,n) - \bar{I}_{u,v}\right]\left[\mathbf{T}(m-u,n-v) - \bar{t}\right]}{\left[\sum_{m,n}\left[\mathbf{I}(m,n) - \bar{I}_{u,v}\right]^2 \sum_{m,n}\left[\mathbf{T}(m-u,n-v) - \bar{t}\right]^2\right]^{1/2}} \tag{3.21}$$

where $\mathbf{I}$ is the image to be searched, $\mathbf{T}$ the template and $\bar{I}, \bar{t}$ the mean values of the image region under the template and the template, respectively. The normalization of image and

---

[4] Rectified images with known epipolar lines are assumed, such that the search for corresponding pixels is a line search.

template handles relative changes in brightness very well. Another possibility is the sum of absolute differences (SAD) (e.g., Hamzah et al., 2010):

$$\Gamma(u,v)_{\mathrm{SAD}} = \sum_{m,n} |\mathbf{I}(m,n) - \mathbf{T}(m-u, n-v)| \tag{3.22}$$

If the surface yields little texture, matching can become highly ambiguous even when template matching is used. The same problem occurs if the texture of the surface is repetitive. In these cases prior knowledge about the desired shape might be necessary, to constrain the solution space and facilitate a satisfactory reconstruction.

Feature Matching

Measures of local similarity that are based on direct comparison of image intensities, obviously deteriorate when the intensities of corresponding points do not match any longer. There are two main reasons for this to occur:

- Violations of the assumption of diffuse reflection; *and*
- affine transformations of the object with respect to the second camera caused by large baselines.

Up to some degree can correlation-based measures cope with changes in intensity and non-Lambertian reflectance. The greater challenge is to handle rotation, scaling, and transformation of object features. Local descriptors like SIFT (Lowe, 2004), SURF (Bay et al., 2006) and DAISY (Tola et al., 2010) that are computed on the intensity images and yield nonlinear feature transformations are suitable to relate object features in both images in a more robust fashion.

A brief, yet comprehensive overview of various approaches for stereo matching is given by Scharstein and Szeliski (2001); Seitz et al. (2006), another overview can be found in Wöhler (2013).

3.6.2 Semi-Global Matching

A central drawback of correspondences established by window-based block matching methods is that there tends to be a smooth transition of the disparity values at depth discontinuities (e.g., Hirschmüller, 2005). A second phenomenon that frequently occurs is *pixel-locking* (e.g., Shimizu and Okutomi, 2001; Stein et al., 2006), although there is a smooth transition between two object views some disparities are more frequent than others resulting in an artificial roughness of the resulting surface. Shimizu and Okutomi (2001) report this to be caused by the commonly used parabola fitting to increase the resolution of the disparity to sub-pixel accuracy. Stein et al. (2006) propose the use of affine transformations on the matching windows to attenuate the effect. Hirschmüller (2005) proposes a framework called semi-global matching (SGM), that will be used in the experiments of this thesis, that is a variation of scanline optimization (e.g., Scharstein and Szeliski, 2001). The search for an optimal path is not only conducted along the epipolar line but in up to sixteen directions symmetrically around the original line path. The resulting 3D cube of cost values for each possible disparity can be searched for the minimum cost which equals the most likely disparity value.

Parabola Fitting

Notwithstanding the findings of Shimizu and Okutomi (2001), fitting a parabola to the integer-valued disparity values that are the result of the SGM did not result in increased pixel-locking in the conducted experiments. On the contrary, the quality of the disparity maps is raised with the sub-pixel accuracy of the disparity estimates as the true value is in general situated between two adjacent pixels. In the case of non-rectified images, this would require a two-dimensional paraboloid.

Since a parabola is a second order polynomial, it is commonly described by $f(u) = au^2 + bu + c$. With respect to a central pixel u_0 and its left and right neighbor u_{-1} and u_1, respectively, corresponding to disparity measurements at d_{-1}, d_0 and d_1 the following system of equations defines the parameters accordingly:

$$\begin{pmatrix} d_{-1} \\ d_0 \\ d_1 \end{pmatrix} = \begin{pmatrix} u_{-1}^2 & u_{-1} & 1 \\ u_0^2 & u_0 & 1 \\ u_1^2 & u_1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{3.23}$$

Inverting matrix $\mathbf{U}$ yields the parameters corresponding to the three points

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{U}^{-1} \begin{pmatrix} d_{-1} \\ d_0 \\ d_1 \end{pmatrix} . \tag{3.24}$$

Since the points to the left and right of the minimal value are considered[5] the extremum of the fitted parabola will always be a minimum, such that $\frac{\partial f(u)}{\partial u} \overset{!}{=} 0$ is a sufficient constraint. From $\frac{\partial f(u)}{\partial u} = 2a\tilde{u} + b = 0$, for a point $\tilde{u}$ where the parabola reaches its extremum, follows $\tilde{u} = \frac{-b}{2a}$. The inverse of $\mathbf{U}$ is given by:

$$\mathbf{U}^{-1} = \frac{1}{\det \mathbf{U}} \begin{pmatrix} u_0 - u_1 & u_1 - u_{-1} & u_{-1} - u_0 \\ u_1^2 - u_0^2 & u_{-1}^2 - u_1^2 & u_0^2 - u_{-1}^2 \\ u_0^2 u_1^2 - u_0 u_1^2 & u_{-1} u_1^2 - u_{-1}^2 u_1 & u_{-1}^2 u_0 - u_0^2 u_{-1} \end{pmatrix} \tag{3.25}$$

The explicit calculation of the determinant is not necessary and from Eqs. (3.24) and (3.25) follows

$$\tilde{u} = -\frac{(u_1^2 - u_0^2)d_{-1} + (u_{-1}^2 - u_1^2)d_0 + (u_0^2 - u_{-1}^2)d_1}{2\left[(u_0 - u_1)d_{-1} + (u_1 - u_{-1})d_0 + (u_{-1} - u_0)d_1\right]}. \tag{3.26}$$

Using relative coordinates with respect to the central pixel yields $u_{-1} = -1, u_0 = 0, u_1 = 1$ and

$$\tilde{u} = -\frac{d_{-1} - d_1}{2(-d_{-1} + 2d_0 - d_1)} = \frac{d_{-1} - d_1}{2(d_{-1} - 2d_0 + d_1)} \tag{3.27}$$

which gives the minimum of the parabola at sub-pixel accuracy based on the integer values obtained from the SGM.

---

[5] The parabola fitting has to be handled differently, if the minimum value is found at the first or last element of $\mathbf{u}$.

3.6.3 Multi-View Stereo

More than two cameras views of the same scene, of course, offer more information and can increase the accuracy or the recovered object volume. A possible approach is to compute depth maps for pairs of camera views and use a stitching algorithm to fuse them into one object (e.g., Curless and Levoy, 1996). Instead of multiple calibrated cameras the consecutive pictures of a scene could also be obtained from one moving camera, for example, taken by an unmanned aerial vehicle. If the motion of the camera, that is the extrinsic parameters of each view, has to be estimated parallel to the 3D structure the problem belongs to the class of *structure from motion* (SfM). A popular and efficient method to solve the SfM setting is the *bundle adjustment* (BA) approach (e.g., Hartley and Zisserman, 2004). In conjunction with the 3D scene points, the framework can also estimate the intrinsic and extrinsic parameters of all camera views by minimizing the reprojection error between the modeled and measured image points. The nonlinear function $Q({}^{C_i}_{W}T, \mathbf{A}_i, {}^{W}\mathbf{x})$ describes the transformation of ${}^{W}\mathbf{x}$ from world to sensor coordinate system of camera i based on camera matrix $\mathbf{A}_i$ and extrinsic orientation ${}^{C_i}_{W}T$. For K image points traced along N images, the reprojection error is

$$E_{\mathrm{BA}} = \sum_{i=1}^{N} \sum_{k=1}^{K} \left\| {}^{S_i}_{I_i}T^{-1}\left(Q({}^{C_i}_{W}T, \mathbf{A}_i, {}^{W}\mathbf{x})\right) - {}^{S_i}_{I_i}T^{-1}({}^{S_i}\mathbf{x}_k) \right\|^2. \tag{3.28}$$

The transformation ${}^{S_i}_{I_i}T^{-1}$ from sensor to image coordinates is necessary to ensure that the reprojection error is measured as an Euclidean distance if the pixels are not quadratic. Minimizing this error subject to a very large number of parameters can be tedious. An efficient implementation, using sparse matrices, in C/C++ has been provided by Lourakis and Argyros (2009). Calibrated cameras obviously reduce the degrees of freedom of the nonlinear optimization and increase the stability, however, they are not necessary.

3.7 Combining Multi-View Stereo and Shape from Shading

Many works regarding multi-view stereo require the surfaces to be Lambertian which is a huge constraint on real-world data and greatly simplifies the estimation of the surface structure. However, even for Lambertian surfaces textureless regions pose a challenge to correspondence estimation procedures (cf. Section 3.6.1). Triangulation based methods enable the estimation of absolute depth information (given a fully calibrated setup) and are accurate on a large scale, yet tend to inaccuracy regarding small-scale surface details. In contrast, photometric methods yield a relative description of the surface given by its gradient field and perform best on (uniform) surfaces that have little texture. Furthermore, they provide an accurate estimation of surface details, yet small errors in the gradient fields add up during integration and lead to large-scale deviations of the obtained surfaces. Therefore, the combination of stereo vision and shading based surface reconstruction poses a perfect solution to resolve the mutual shortcomings. A new combination of stereo vision and shape from shading will be presented in Section 7.2 and the following section summarizes some of the approaches of other researchers in this field.

3.7.1 Fusing Shading Cues and Stereo Information

Fua and Leclerc (1995) present one of the first approaches on combining shading cues and stereo triangulation to compute a 3D mesh of an object. The procedure is restricted to reflectance functions that can roughly qualify as Lambertian to take advantage of viewpoint independence and simple representation. As there is only the surface albedo to estimate, they consider smoothly spatially varying parameters. The viewpoint independent intensity dramatically simplifies the search for stereo correspondences as intensity becomes a reliable feature even when using correlation-based similarity measures. Two opposing forces drive the reconstruction. One measuring the deformation from a nominal shape that is known or defined a-priori, a plane for example. The second error term depends on image-based information, that is a linear combination of multi-image intensity correlation and surface shading, whose weights are adapted based on the level of surface texture. The multi-image correlation accounts for the sum of squared differences of non-occluded intensities per surface facet, which should be equal in every image. The same fact is exploited in the shading cue to compute an estimate of the surface albedo based on the average intensity of all pixels that one facet is projected on. The variation of this albedo with respect to neighboring facets then constrains the normal directions on uniform surface areas. Substantial changes of the albedo at transitions of surface materials violate this assumption, such that only smoothly varying albedos are considered.

In contrast to many recent approaches on fusing photometric and triangulation stereo, Fua and Leclerc (1995) rely on a standard multi-view stereo setup rather than on a photometric stereo setup. They illuminate their multi-view scene with one directional light source and thus create multiple viewing directions per image point instead of multiple incident directions with a fixed camera and changing light sources. Since consistency between photometric cues is an essential condition, this introduces additional uncertainties for non-Lambertian surfaces.

A special, yet still important application scenario is tackled by Samaras et al. (2000), who build on the previous work of Fua and Leclerc (1995) and Samaras and Metaxas (1998, 1999) and apply the combination of stereo and shape from shading to the shape estimation of human faces. They rely on a deformable model as described by (Fua and Leclerc, 1995) and consider a patchwise constant diffuse albedo and handle specularities as outliers of the diffuse lighting model. This is, however, only feasible for the relatively wide specular component - in contrast to a sharp spike - that is typical of human skin. Disparity maps are generated for each image pair of a video sequence and the 3D points are fit to the face model. Using this as the initial solution, the result is refined using shape from shading in regions of little texture and low curvature that are identified based on the self-consistency method and the *minimum description length* (MDL) score (Leclerc et al., 2000). Leclerc et al. (2000) propose to assess the accuracy of multi-view stereo correspondences with their self-consistency method. The key idea is similar to the generic viewpoint constraint (e.g., Horn, 1986; Freeman, 1996), yet evaluated on real data. Assuming an established point correspondence, the camera matrices are slightly manipulated to change the reprojection and therefore the triangulated scene point. If the distance between the 3D points obtained from manipulated matches over a series of camera positions is small, the scene point qualifies as self-consistent. However, this criterion can only be applied to multi-view stereo, i.e., many more than two views.

Zickler et al. (2002) loosen the demand of traditional stereo of viewpoint-independent object radiance – strictly Lambertian surfaces – by exploiting the BRDF property of Helmholtz reciprocity (cf. Section 2.2.3). Since viewpoint and incident light are interchangeable without affecting the BRDF value, it is possible to generate a pair of two reciprocal images that define the surface normal at a point ${}^{W}\mathbf{x}$ independent of the (unknown) BRDF at this point. This enables the simultaneous estimation of surface normal based on shading and depth based on multi-view correspondences on surfaces of arbitrary reflectance properties. The accuracy of the reconstruction is similar to that feasible with photometric stereo but does not require the knowledge or estimation of the BRDF beforehand. They present a simple rotational setup that generates many reciprocal image pairs per object and demonstrate the ability to recover the surface shape of objects with little interreflection since that might hinder the reciprocity assumption.

A more general approach to combine multiple sources of information is presented by Zhang et al. (2003) coupling surface normals, surface position, object and camera motion, and illumination in one minimization problem. Traditional vision problem statements like photometric stereo, structure from motion or multi-view stereo can be seen as specific cases of the presented method. However, to simplify and enable the solution of this high-dimensional problem the assumption of Lambertian reflectance without the presence of occlusions, shadows or interreflections is a crucial assumption. Furthermore, the authors claim a pixelwise objective function to be unstable and assume a window of constant flow and normal direction around each pixel. This diminishes the advantage of photometric stereo to compute per-pixel normal estimates in contrast to window-based stereo matching algorithms. The authors provide a mathematically elegant way to estimate the Hessian matrices for computing camera motion and illumination or shape and normal estimation respectively and as such a means to assess the uncertainties of the estimates. The solution is computed from sparse structure-from-motion on manually selected features and then iteratively solved for the shape and the illumination while keeping the other unknowns fixed.

Nehab et al. (2005) combine depth and gradient information obtained from a triangulation scanner and photometric stereo. They assess the reliability of each of the sources and conclude that the triangulation scanner yields randomly distributed high-frequency noise that can distort the computed normals up to 30°. On the other hand, the normal maps measured from photometric stereo techniques exhibit low-frequency errors including systematic biases caused by simplifications in BRDF and lighting models, interreflections, and shadows. Integrating these normals introduces large errors in the depth information. The low-frequency component in the photometric normal field is replaced by that from the depth data after smoothing both normal fields with a Gaussian kernel. Each surface point $\tilde{\mathbf{x}}(x, y)$ is expressed in projective coordinates only depending on the depth function $z(x, y)$

$$\tilde{\mathbf{x}}(x, y) = \left[-\frac{u}{\hat{f}_{\mathrm{u}}} z(x, y), -\frac{v}{\hat{f}_{\mathrm{v}}} z(x, y), z(x, y)\right]^{\mathrm{T}}. \tag{3.29}$$

$\hat{f}_{\mathrm{u}}$ and $\hat{f}_{\mathrm{v}}$ are the focal lengths in pixels. They introduce a position error that is the distance between (3.29) and the measured position and a normal error that relates the tangents to the optimized surface to the measured normals. Ideally, the angle between the tangent and normal direction should be 90°. The formulation of the normal error yields the advantage of not introducing a non-linear problem statement. The method is evaluated on a number of diffuse surfaces and it is shown that the resulting depth maps follow a ground truth closer than both data sources individually.

Rohith et al. (2013) resolve the stereo ambiguity problem of finding corresponding scene points in textureless regions by generating paths between sparse disparity estimates. The interpolation is constrained by shape from shading cues deeming their procedure *gradient constrained interpolation* (GCI). For diffuse surfaces the change in luminance $L_\mathrm{v}(\mathbf{x})$ and disparity $\delta(\mathbf{x})$ of a surface point $\mathbf{x}(x, y)$ is related by an arbitrary scale factor that depends on surface albedo and scene distance, if the camera's optical axis is aligned with the incident light direction. This fact is employed to compute minimum cost paths between known stereo correspondences that result in dense stereo according to

$$\operatorname*{arg\,min}_{\delta(\mathbf{x})} \sum_{\mathbf{x} \in Z^2} | \, \|\nabla \delta(\mathbf{x})\| - \alpha_\mathrm{E} g(\mathbf{x})|^2. \tag{3.30}$$

Where α_E is the scaling factor and $g(p)$ is supposed to follow the Eikonal equation

$$g(\mathbf{x}) = \frac{1}{\sqrt{1/L_\mathrm{v}(\mathbf{x})^2 - 1}}. \tag{3.31}$$

To speed up computation efficient paths are created based on geodesic maps with the metric $|\nabla \mathbf{I}(\mathbf{x})|^2$. ∇ is the gradient operator. The manifolds are computed from *fast marching method* (FMM) (Kimmel and Sethian, 2001).

Wu et al. (2011a) combine the strengths of multi-view stereo and shape from shading under arbitrary, constant lighting. They compute an initial geometry guess from existing MVS approaches and refine the result based on the shading information of each object vertex visible in multiple cameras. To keep the problem tractable, they assume constant albedo and Lambertian reflectance and estimate the unknown lighting as an environment map based on low-order spherical harmonics (SH) on the MVS depth estimate. As many illumination effects, e.g., interreflections, radiometric calibration errors, are neglected, rather than directly comparing intensities, the gradients of image data and predicted Lambertian model are compared throughout up to 30 views of non-convex real-world objects. The results demonstrate that the approach is apparently able to refine the high-frequency shape component of diffusely reflecting objects.

Similar to Wu et al. (2011a), Langguth et al. (2016) use third-order spherical harmonics to approximate the unknown lighting on an initial coarse stereo surface model and again assume the surface to be Lambertian. Based on Edwin Land's Retinex theory[6] (e.g., McCann, 2016), they assign the intensity gradient of a surface point to a different cause, that is either change in albedo (large gradient) or changes in surface normals (small gradient). They form an energy function that combines the reprojection error of the intensity gradient in neighboring views, and a shading error also based on the gradient of the acquired image and predicted model. Operating in the log-space, they create an albedo independent formulation of the shading error making use of the Retinex theory. Surface patches are represented with bi-cubic patches and the optimization of the energy function is achieved via linearization. The algorithm is applied to the various data sets of the Middlebury database[7] and additional out-door datasets yielding visually appealing results. Obvious limitations of the method are objects whose albedo changes gradually (violating the Retinex theory) and non-Lambertian reflectance properties.

Zhou et al. (2013) present a multi-view setup to compute the 3D shape of objects of spatially varying isotropic materials comparable to Goldman et al. (2005) (cf. Section 3.4.2).

---

[6] Retinex is a coinage containing retina and cortex and refers to the human perception of color and color constancy.

[7] http://vision.middlebury.edu/stereo/ (Scharstein and Szeliski, 2001).

The simplest version of their setup requires a camera and a hand-held light source only, thus lowering the demands of Einarsson et al. (2006) and Holroyd et al. (2010) drastically. They acquire multiple views of an object situated on a turntable and use structure-from-motion to compute an initial sparse set of 3D points. The turn-table yields a checkerboard pattern to ensure sufficient information for image registration and they employ the technique proposed by Alldrin and Kriegman (2007) to compute pixelwise surface normals from rotationally symmetric lighting, by finding the symmetry axis in the illumination profile of each pixel. They handle outliers due to global illumination effects by fitting a truncated Fourier series to the data. If the direction of the surface gradient is known, it is implicitly clear that the surface height must remain constant perpendicular to that direction. Therefore it is feasible to trace an *iso-depth contour* of constant height across the surface (Alldrin and Kriegman, 2007). The sparsely known 3D points are now propagated along these contours in each of the views, generating new 3D points that are propagated and so forth. The points are sorted according to the confidence of the associated iso-depth contours and the projection to different viewpoints is checked for visibility. Poisson surface reconstruction is applied to the resulting surface normals and 3D points. The known shape is then used to estimate the reflectance functions from basic isotropic BRDFs. The basis BRDFs and mixing weights are estimated according to Lawrence et al. (2006) assuming a linear mixing. The more sophisticated acquisition setup contains 72 LEDs in two concentric circles around the camera. The mean depth error is about 0.5 mm and the relative reflectance *root mean squared error* (RMSE) is 9%. As for every reflectance based photometric approach, interreflection poses a strong limitation of the method.

Ackermann et al. (2014) present an extension of the work of Hertzmann and Seitz (2003, 2005) (cf. Section 3.4.1) with a multi-view photometric stereo setup where they explicitly re-introduce the example object into the scene. In contrast to Vlasic et al. (2009) they do not need the visual hull (nor other sources of information) for boot-strapping and focus on objects with little-to-no texture and uniform non-Lambertian BRDF lit by a distant point light source. The distant light source is a necessary constraint to allow the idea of example based normal inference. Similar to Hertzmann and Seitz (2005) they rely on orientation consistency between object and spherical example to compute normal maps without the need for calibrating camera or lighting environment. Similar to bundle adjustment they project the 3D position of a pixel in all other viewing positions to obtain a unique vector of intensities. The matching error function, solely based on the intensity, yields a wide low-error basin and is therefore insufficient to infer an accurate 3D point. The authors extend the intensity error with the error of the surface normals estimated at each corresponding pixel. If the 3D point does not lie on the true surface, the normal direction varies severely for the different views. To achieve a single optimization process, they define an energy function that contains the intensity error and a deviation from the discretely sampled sphere normals. A third term that couples normals and depth data by computing the intersection of a local plane, centered at the current depth estimate oriented according to the current normal, and the neighboring normal directions. The distance should be small, which effectively reduces depth discontinuities. The range of the initial depth search has to be initialized manually. They evaluate their algorithm on up to forty views of relatively simple shapes, comprising a spray painted plastic bottle, spheres, and glossy owl figurine. The datasets yield few concave regions and thus little interreflection. The error to ground truth data is less than 2.5 mm at a camera distance of 2 m. The authors claim to have faced “challenging, unknown reflectance properties” which appears to be an exaggeration of the encountered challenges. However, the fact that they do not rely on so-called proxy geometry, that is an initial guess of the true structure, is an advantage.

3.7.2 Variational Approaches to Multi-View Shape from Shading

Few researchers have employed variational approaches loosely related to Section 7.2 to solve object shape retrieval from image information in very different approaches. Semerjian (2014) presents a variational framework to combine multi-view photo consistency with a smoothness function on an arbitrarily scalable grid defined by finite elements. The photo consistency is based on the image gradient of 3×3 patches surrounding a central image point and the Jacobian that accounts for the coordinate transformation of the image gradient between consistent camera positions. The smoothness term is computed as the divergence of the surface normals against image coordinates in contrast to the usual approach to compute surface normals against 3D space. A coarse-to-fine minimization is implemented that has the inherent drawback of finite elements that the grid spacing has to be coarser than the pixel spacing to keep the degrees of freedom of the differential equations tractable and smaller than the number of pixels (i.e., measurements). The algorithm is evaluated on aerial photographs of mountains with ≈ 10 images per surface and the Dino dataset from the Middlebury benchmark achieving visually appealing results.

Maurer et al. (2018) present a variational approach to the combination of shading and stereo cues without the need for an initial surface mesh. They minimize an energy function that contains a stereo and a shading term alongside strong regularizations on depth, illumination, and albedo. The incident illumination is assumed to be constant or slowly varying and the depth is subject to a second-order smoothness term. Regarding the albedo regularization, they exploit the color image information to separate albedo from geometry based on similar rg-chromaticity. The results outperform competing methods including the proposal by Wu et al. (2011a). Although the albedo is not considered as constant, the algorithm is still limited to Lambertian surfaces.

DATA ACQUISITION

4

4.1 Acquisition Setup

All data that have been used in the experimental evaluation of Chapters 5 to 7 have been acquired with the setup depicted in Fig. 4.1. The first version of the setup was mentioned by Herbort et al. (2011). It has later on been extended and improved during the Bachelor thesis Lenoch (2011) and published in Lenoch et al. (2012). The new setup has been used from the on (e.g., Herbort and Wöhler, 2012). The exemplary scene contains the diffuse calibration object used in the light source calibration described in Chapter 6. The equipment is enclosed in a black box to avert the distortion of the measurement process by ambient light. In detail the acquisition setup contains:

1. 18 LED light sources arranged in 3 arcs, Seoul P4 Power LEDs, $\lambda = 525\,\mathrm{nm}$;
2. Micos DT-65N and PRS-110 rotatable object stage;
3. Lumenera Lu165M stereo cameras, CCD sensor, resolution 1392 × 1040 pixels; *and*
4. industrial quality fringe projection scanner ViALUX zSnapper Vario with AVT-Pike F421, CCD sensor, resolution 2048 × 2048 pixels.

4.2 Depth Data Quality Assessment

As indicated above, the ViALUX zSnapper Vario fringe projection scanner is used to acquire reference 3D data of the objects under inspection in the experimental analysis where image data only is used and for initial 3D structure in other experiments. Especially for the first use – reference data – it is of interest to assess the accuracy of the depth data in terms of repetition constancy. There is little to be found regarding an error analysis in the zSnapper manual apart from the reference to a numerical test of the calibration between camera and projector that is conducted on the calibration gauge and compares estimated and stored values. It is mentioned that "[v]alues of standard deviation below 0.1 mm indicate that a new calibration is not necessary" (ViALUX Messtechnik + Bildverarbeitung GmbH, 2010, p 42). Hence, an assessment and evaluation of the accuracy of the depth data is briefly conducted in the remainder of this section to motivate the data acquisition process in the employed form.

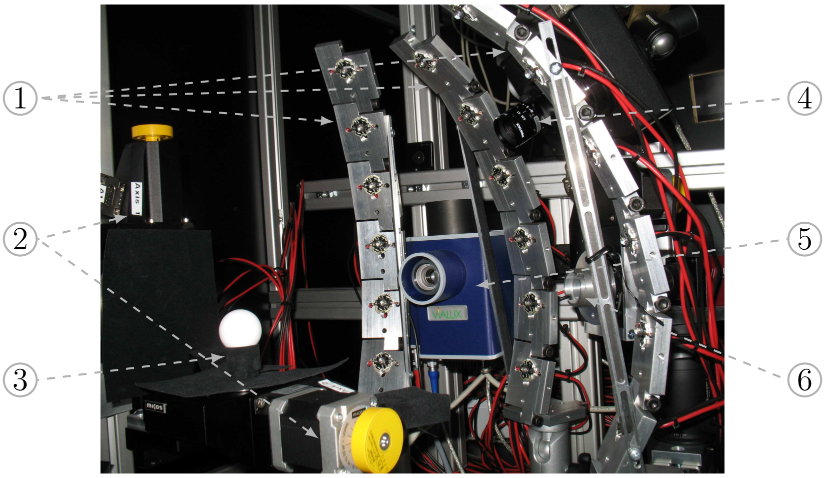


Figure 4.1. Acquisition setup consisting of 18 LED light sources arranged in 3 arcs (1), rotation stages (2), object stage (3), stereo cameras (4), fringe projection scanner composed of projection unit (5) and camera (6). The diameter of the diffuse sphere on the object stage is 30 mm.

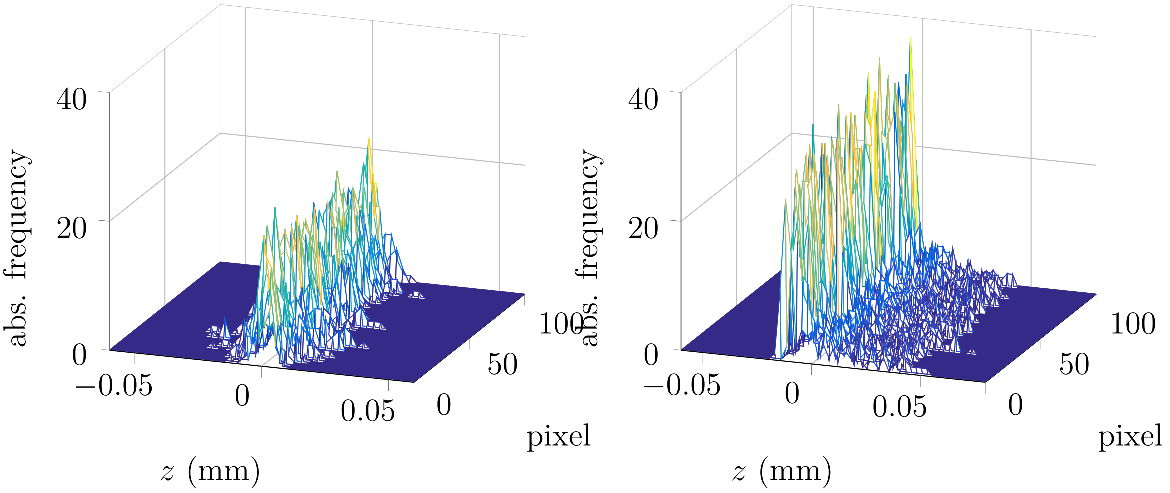
Fig. 4.2 displays the distribution of depth values of 100 randomly drawn pixels across 100 measurements reduced by their respective mean value. The Kolmogorov-Smirnov test for normality rejects the hypothesis that the pixelwise measurements are drawn from a normal distribution (e.g., Sachs and Hedderich, 2006). The rejection coincides with the observation that Fig. 4.2b yields a non-symmetrical distribution of the measurements. However, this is less evident in Fig. 4.2a. The asymmetry of the measurements of the Plaster dataset is surprising and will be addressed in more detail later. The measurements of the Leaf dataset are, although not normal distributed, symmetrically centered around their mean value that yields a high absolute frequency[1]. This suffices to deduce that it is valid to use the mean computed from multiple measurements as the final value, thereby reducing the measurement noise.

The standard deviation computed for the x, y and z component of the depth data of all pixels of the Leaf dataset is given in Fig. 4.3. x, y and z component are displayed from left to right with different scaling of the z component. x and y component display a strong dependency of the measurement variation on the spatial location of the pixel. The lines of horizontal and vertical origin that are the center of the image exhibit a minimum of variation. Since the object did not cover the entire measurement volume the images have been cropped, which is why the "middle axis" is not extremely apparent in the first image. Apart from this feature, the error in z direction is exceptionally high near invalid measurements but otherwise not outstanding with maximum values about 10 µm and an average of 4.6 µm, a comprehensive set of values is given in Table 4.1. Figure 4.4 yields similar findings for the Plaster dataset since the object is larger the cropping of the images did not conceal the symmetric dependence of the errors of the x and y component. The striking aspect is that a wave pattern is visible – especially for the z component – across the surface that is related to the acquisition procedure of the scanner [2]. This texture is most probably caused by the

---

[1] Since the distribution is based on 100 measurements, the absolute distribution equals the relative distribution given in percent.

[2] Modulated sinusoidal patterns are projected onto the surface, and their distortion indicates the depth value at each point on the surface.



(a) Leaf (b) Plaster

Figure 4.2. Distribution of zero-mean depth measurements of 100 randomly drawn pixels based on 100 measurements each. The measurements of the Leaf data are as expected distributed symmetrically around their mean value. The Plaster data in contrast yields an offset between mean value and mode probably caused by the interaction of the projected light and the dielectric surface material, since the projection patterns are visible in the standard deviation of the measurements as depicted in Fig. 4.4.

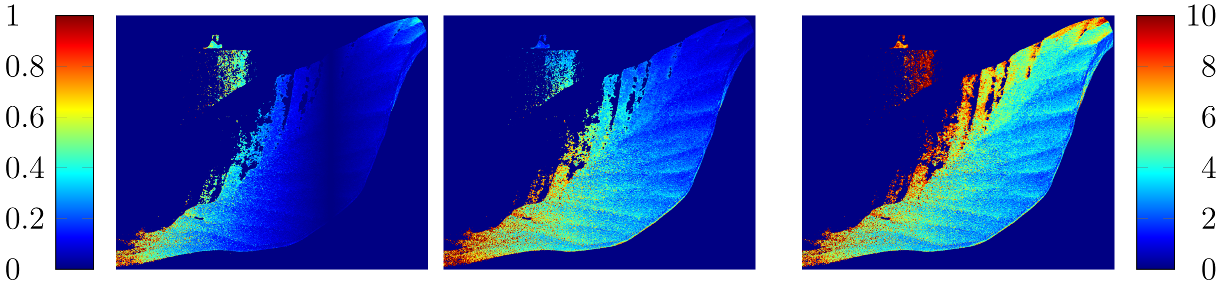


Figure 4.3. Standard deviation of depth measurements on Leaf data set based on 100 measurements given in µm. x, y and z component from left to right, z component with different scaling.

surface material of the object (lacquered plaster) being dielectric in contrast to the metallic surface of the Leaf dataset. The visibility of the sinusoidal patterns as an unwanted noise effect furthermore could explain the asymmetry of the distribution of the measurements as depicted in Fig. 4.2(b).

Nevertheless is the analysis of the scanning procedure not primarily in the scope of this thesis and these effects are noted but not accounted for during the experiments. Especially as the difference between mean and mode for the Plaster dataset (and possibly all dielectrics) is in the range of 10 µm and below, at a mean distance to the object of ≈ 250 mm, and the scanner data is used as initialization or large scale reference only. Still, this shift should be considered as another indicator that image-based 3D reconstruction methods like those that are evaluated and proposed in this work are of high value.

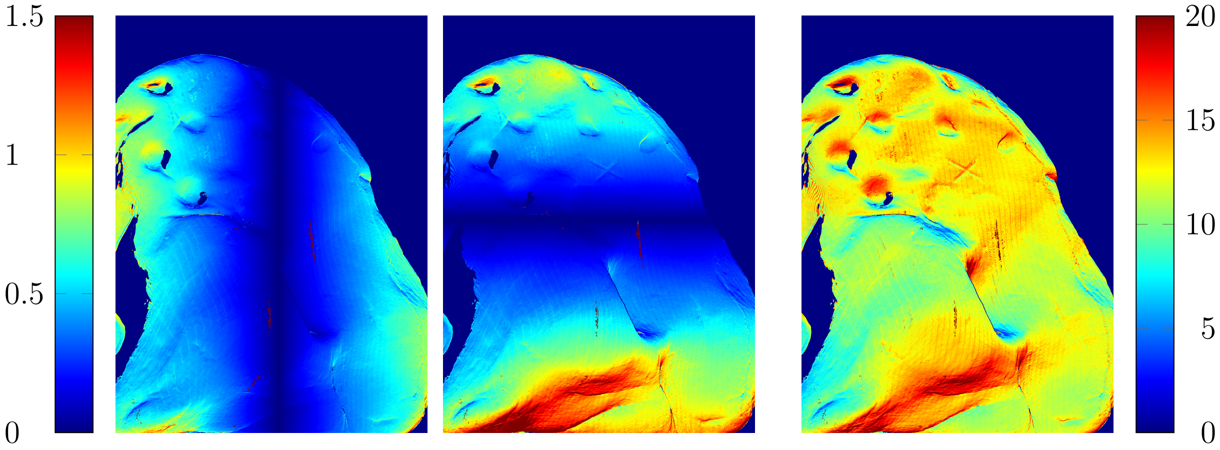


Figure 4.4. Standard deviation of depth measurements on Plaster data set based on 100 measurements given in µm. x, y and z component from left to right, z component with different scaling.

Table 4.1. Average values of standard deviation per pixelwise measurement on Leaf and Plaster datasets, the mean distance to the object is $\approx 250\,\mathrm{mm}$. The magnitudes and proportions are similar while the details are revealed in Figs. 4.2 to 4.4.

	Leaf	Plaster
$\bar{\sigma}_{\mathrm{x}}$ (µm)	0.149	0.3889
$\bar{\sigma}_{\mathrm{y}}$ (µm)	0.325	0.5643
$\bar{\sigma}_{\mathrm{z}}$ (µm)	4.632	13.7050

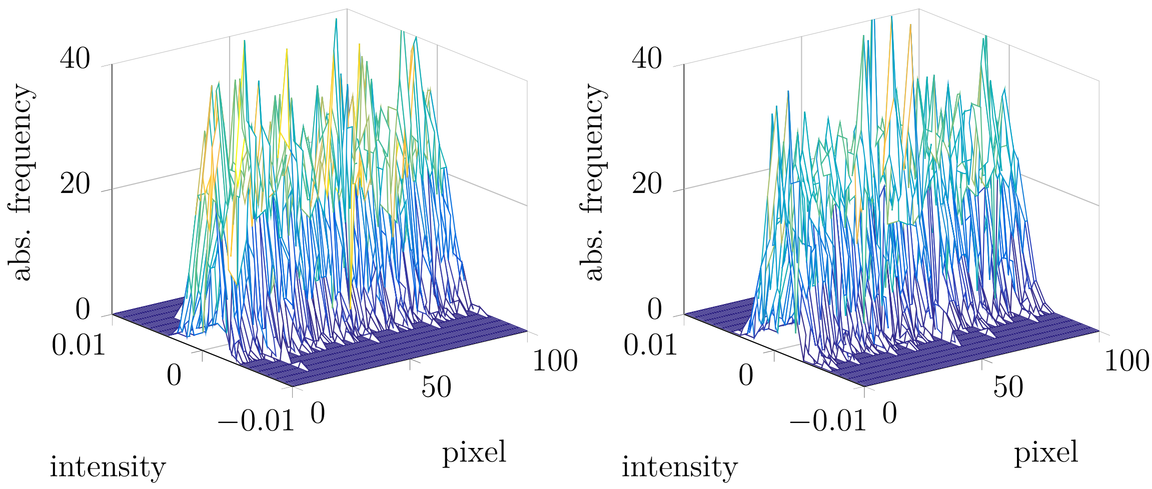
4.3 Image Data Quality Assessment

The image data is examined regarding its noise distribution similar to the assessment of the depth data. Since the acquired images are used as measurements of object reflectance, the highest possible degree of accuracy is of utmost importance in the domain of image based 3D reconstruction. Digital image sensors use the photoelectric effect to count incoming photons in a certain time interval. These events per unit time follow a Poisson distribution

$$P_{\mathrm{p}}(k) = \frac{{\lambda_{\mathrm{p}}}^{k}}{k!}\,\mathrm{e}^{-\lambda_{\mathrm{p}}} \tag{4.1}$$

where e is Euler's number and λ_{p} is the mean rate of expected events per unit time. For $k \geq 30$ the Poisson distribution converges against a Gaussian distribution, such that surface areas that are directly illuminated should satisfy the simplified assumption. As it is evident in Fig. 4.5 the distribution of the measurements is symmetrical and tightly centered around the mean value, however, the measurements in Fig. 4.5a pass the Kolmogorov-Smirnov test for normality, whereas those in Fig. 4.5b fail to do so.

The image of mean pixel values and the corresponding map of pixels that fail to pass the test for normality (indicated in white) given in Fig. 4.6 suggest a possible cause: Almost all pixels that are not normally distributed belong to surface areas of self-cast or attached shadow. Unfortunately, some areas that exhibit self-cast shadows but are relighted from interreflections (upper cavities) still follow a normal distribution. This is probably due to the indirect illumination caused by the surrounding diffuse pixels that follow themselves



(a) Non-shadowed Pixels

(b) Shadowed Pixels

Figure 4.5. Distribution of zero-mean intensity measurements of 100 randomly drawn pixels based on 100 measurements each.

a normal distribution. After all, this is an interesting observation regarding the detection of shadows in monochrome images, and this fact could be used to exclude vast areas of cast shadows from further considerations in the photometric analysis. Nonetheless, a large number of images would be necessary to provide a suitable sample size for the normality test. The full set of images of the Plaster dataset and their corresponding normality-test-maps can be found in Appendix A.1. As there exist many statistical tests and their significance always depends on a specified level, comprehensive study is necessary to determine if this observation is useful.

In any case, both shadow and non-shadow pixels are distributed symmetrically, and this motivates the acquisition of multiple images per exposure time and computing their mean value. For all experimental analysis, a mean value of 16 images per exposure time is used. Additionally, each image is compensated for a corresponding dark frame that is captured at the same exposure time. Dark frames contain fixed-pattern noise caused by dark current, a typical kind of image sensor noise.

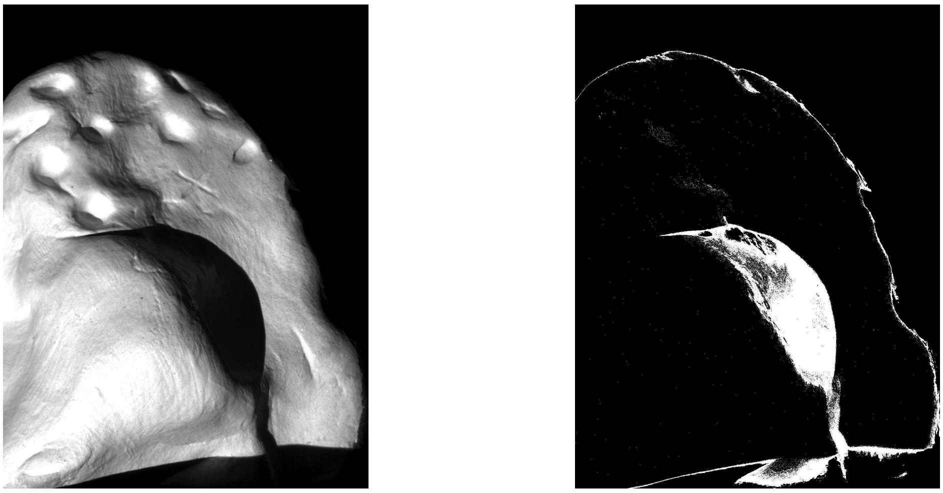
4.4 High Dynamic Range Images

The luminous intensity in nature can vary over a range of about 160 dB from distant stars emitting $\approx 0.001\,\mathrm{cd\,m^{-2}}$ to direct sunlight at $\approx 100\,000\,\mathrm{cd\,m^{-2}}$ (e.g., Schulz et al., 2007). To this date no available image sensor is capable of capturing this dynamic range in one single image[3], which results in either under- / or overexposed regions in a camera image of a high dynamic range scene. Typical dynamic ranges of image sensors are:

- consumer grade CMOS $\approx 54\,\mathrm{dB}$
- consumer grade CCD $\approx 66\,\mathrm{dB}$

---

[3] There exist high-end camera sensors that use specialized read-out techniques to drastically increase their dynamic range. However, these will not be regarded here and are still not capable to cover 160 dB.



(a) (b)

Figure 4.6. Mean camera image (a) and map of failed normality test (b) for LED 1.3. The pixels that fail to pass the Kolmogorov-Smirnov test for normality are almost exclusively located in areas of cast shadow.

Table 4.2. Validity of linear regression.

camera	RMSE	correlation coefficient
AVT-Pike F421	0.2440 %	0.999 747
Lu165M (left)	0.2853 %	0.999 907
Lu165M (right)	0.3037 %	0.999 903

- high-end CCD ≈ 80 dB

An intuitive approach to capturing the full dynamic range of a scene, even with a consumer grade image sensor, is the acquisition of a series of images taken at different exposure times (e.g., Debevec and Malik, 1997) and to compute a virtual high dynamic range (HDR) image from the series. To account for under- and overexposure, gray values lower than 2 % and larger than 98 % of the pixel range are excluded from the consideration, since they most likely contain noise or saturated pixels, respectively. The expenditure is reduced to a simple linear regression if the camera response curve can reasonably be assumed to bear a linear dependence between exposure time and brightness (pixel value). Figure 4.7 shows the exemplary response curves for a single pixel subject to exposure from 1 ms to 1000 ms for the 3 cameras that are used in the experiments. Since the cameras exhibit a different bit depth, the pixel values have been normalized to the theoretical maximum. The AVT Pike F-421 is used in 8-bit mode and both Lu165M are operated in 16-bit mode. The dynamic range of the Lu165M is given as 66 dB according to its data sheet (Lumenera Coorporation, 2008), the corresponding value for the AVT Pike F-421 was not available.

All response curves are highly linear, yielding correlation coefficients above 0.99 and errors around 0.3 % and below, detailed values are given in Table 4.2. Hence, as the assumption of linearity is validated, series of 11 images are captured of all objects and the HDR images

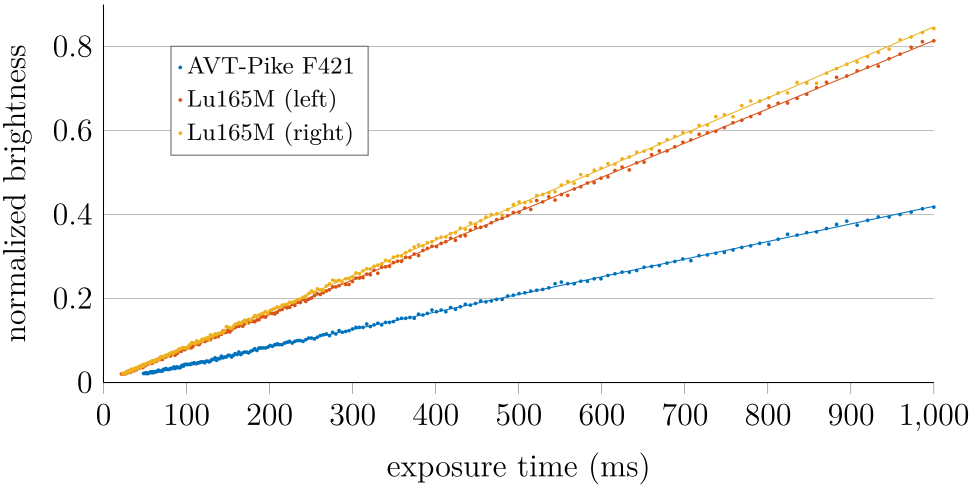


Figure 4.7. Camera response curves for the relevant interval of exposure times on one arbitrary pixel. The brightness has been normalized to the possible maxima of the 8-bit (AVT Pike-F421) and 16-bit (Lumenera Lu165M) resolution. The linearity of the camera response with respect to the exposure time is compelling. The errors of the linear regression are around 0.3 % and below for all cameras.

are computed as the linear regression according to

$$\mathbf{I}_{\text{HDR}}(u,v) = \frac{\sum_k (\mathbf{I}(u,v,k) - \bar{\mathbf{I}}(u,v))(\mathbf{e}(k) - \bar{e})}{\sum_k (\mathbf{e}(k) - \bar{e})^2} = \frac{\text{Cov}(\mathbf{I},\mathbf{e})}{\text{Var}(\mathbf{e})}, \tag{4.2}$$

where $\mathbf{I}(u,v,k)$ is the image at exposure time $\mathbf{e}(k)$ and $\bar{\mathbf{I}}$ and $\bar{e}$ are the mean values of the series, respectively. Note that HDR images will be used throughout the entire experimental evaluation without explicitly stating this again for the sake of readability.

4.5 Relative Change of Projected Sensor Cell Size

The extent of a single image pixel on the sensor is a fixed property of the camera. The AVT Pike F-421 exhibits a pixel size of $k_{\text{u}} = k_{\text{v}} = 7.4\,\mu\text{m/px}$. Facing the goal of deriving real-world depth information purely from the camera data, it is crucial to keep in mind, that the pixel size on a distant object is subject to a perspective projection. The theorem of intersecting lines, employing principal distance b and k_{u}, leads to

$$\hat{k}_{\text{u}}(z) = \frac{k_{\text{u}}}{b} z. \tag{4.3}$$

The actual surface area covered by one pixel $\hat{k}_{\text{u}}$ changes linearly depending on the distance from the camera lens, such that the spacing of the grid defining the 3D object space is (possibly) non-uniform. The focal lengths have been estimated during the calibration of the zSnapper Vario and the stereo camera calibration, respectively. Figure 4.8 depicts the change of the projected sensor cell size on a distant surface. The values of left and right stereo camera almost coincide since their estimated focal lengths are very similar. The values of k_{u} and $\bar{z}$ for each camera are given in Table 4.3, in conjunction with the minimum and maximum projected pixel size for the designated measurement volume. Note that image-based 3D reconstruction methods do not have a specific measurement volume since everything that

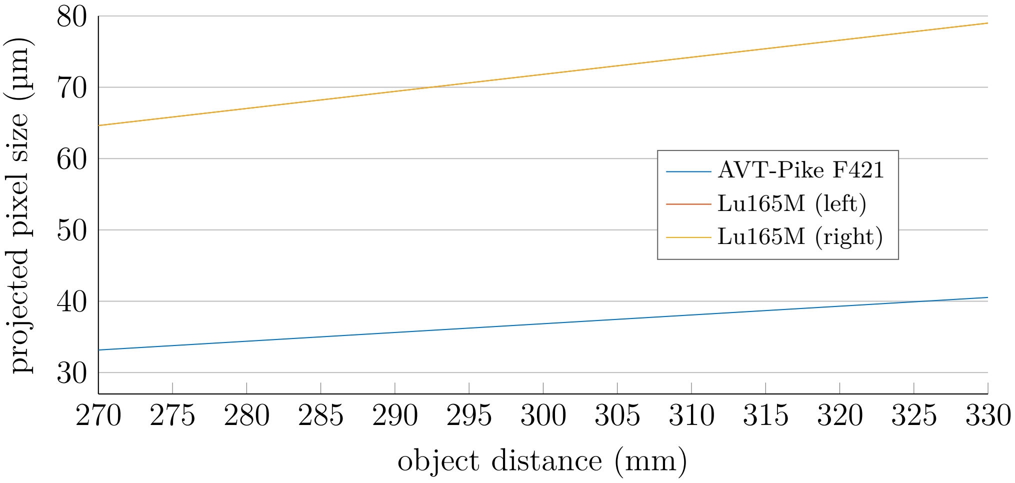


Figure 4.8. Changes of effective projected pixel size on a distant surface in dependence of object distance.

is captured in the image can be processed. However, as the zSnapper Vario is used for reference depth data, its measurement volume is used for all experiments. The differences of the projected pixel sizes between the minimum and maximum possible depth are minor – about 7 µm for the AVT-Pike F421 and 14 µm for the Lumenera Lu165M – hence, this effect will be neglected in the course of this thesis.

Table 4.3. Sensor cell extent k_u, focal length $\bar{\mathbf{l}}$ and exemplary values of projected pixel size $\hat{k}_\mathrm{u}$ for minimum and maximum distance of designated measurement volume for different cameras.

camera	k_u (µm)	$\bar{\mathbf{l}}$ (mm)	$\hat{k}_\mathrm{u}(z = 270\,\mathrm{mm})$ (µm)	$\hat{k}_\mathrm{u}(z = 330\,\mathrm{mm})$ (µm)
AVT-Pike F421	7.40	60.2500	33.1618	40.5311
Lu165M (left)	6.45	26.9456	64.6302	78.9925
Lu165M (right)	6.45	26.9448	64.6321	78.9948

CONTRIBUTION: BAYESIAN APPROXIMATION OF REFLECTANCE FUNCTIONS

5

Photometric gradient estimation relies on the minimization of a target function that compares captured images and synthetic reflectance maps. Therefore, a profound knowledge of the reflectance properties of the object surface is required to compute high-quality gradient fields. As detailed in Chapter 3 there exist a variety of techniques to reproduce the appearance of an object, however, the most important for this thesis is the use of parametric BRDF model functions. This fact comprises two implications: The BRDF model has to be selected adequately for the given material and the set of model parameters has to fit the reflectance properties of the object. This is particularly complicated to fulfill since many human-made objects or natural materials yield reflectance properties that are spatially varying across the surface, and the assumption of a constant albedo imposes a considerable limitation. Especially the properties of human-made objects tend not to change gradually, but abruptly from one region to the other, e.g., for painted objects of different color or mixtures of different materials. Consequently, it is vitally important to infer those areas of a surface that are united by identical reflectance properties and determine the BRDF parameters that best represent these areas.

This section provides a novel and mathematically elegant procedure that employs methods from Bayesian statistics to segment the surface of an arbitrary object into a set of meaningful regions implicitly based on their reflection characteristics. The borders of the surface patches and the posterior distributions of the corresponding BRDF parameters are directly computed from the input data, using prior distributions of each parameter to reflect boundaries and existing information. In contrast to commonly applied gradient descent procedures that only handle point estimates, the uncertainty of the estimate is known since every parameter is handled as a stochastic distribution. Furthermore, a drawback of confidence analysis in classical statistic – that is based on asymptotic theory and the assumption of large datasets – is circumvented (e.g., King et al., 2010). Combining Bayesian inference or statistics in general and reflectance estimation has had little attention in the past. A Gaussian mixture model to represent varying reflectance in the frequency domain is proposed by Aittala et al. (2013). They improve their initial estimate with a maximum a posterior estimate. Louw and Nicolls (2010) employ a Markov random field (MRF) that is estimated with a population Markov chain Monte Carlo algorithm (Laskey and Myers, 2003) to imitate the reflectance behavior of a surface.

The contribution described in this chapter has been previously published in Lenoch et al. (2016).

5.1 Bayesian BRDF Modeling

A commonly used approach to finding the unknown parameters of a model is an estimation that minimizes the least-squares error between measurement and model response. This estimation may be a simple linear regression or a non-linear iterative optimization, the critical thing to note is that, in either case, the solution is a point estimate. In a non-linear optimization, this might even be a local optimum. For linear regression, it is possible to derive confidence bands from the point estimates assuming normality or enormous sample sizes. However, especially the problem of estimating the parameters of a reflectance function is in general highly non-linear, and it is difficult to answer questions like

- "what is the probability that the parameter estimates are correct?";
- "how many other parameter estimates result in similar errors?"; *and*
- "how certain is the result that was computed?"

because there exist no methods to directly derive answers to those questions from the result of the point estimate. The Bayesian approach to solving this is to find an answer to the question "how probable are the estimated parameters given the measured data?" which directly relates to the conditional probability $P(\boldsymbol{\theta}|\mathbf{I})$ subject to model parameters $\boldsymbol{\theta}$ and the input data $\mathbf{I}$. The result will be a probability distribution of the individual parameters that implicitly allow for an answer to the previously stated questions. The details and relations of applying Bayesian statistics to BRDF estimation will be elaborated in this chapter.

Although being a relatively simple BRDF model, the physically plausible model according to Blinn (1977) (cf. Section 2.4.3) is chosen for the reflectance modeling since larger segments of complex reflectance can be approximated by smaller segments that yield different sets of parameters. Note that it is not required for the segmented patches to coincide with physical materials, as subtle changes can occur even within the same material and require changing BRDF parameters to achieve accurate modeling.

Concerning Bayesian statistics, the BRDF model provides the mean of a normally distributed likelihood function. The residuals of the BRDF model fit are assumed to follow a Gaussian distribution with an unknown variance ${\sigma_\mathrm{M}}^2$. Since ${\sigma_\mathrm{M}}^2$ is unknown, it is estimated as a part of the procedure, and hence it is feasible to assess the validity of the chosen model and its accuracy of fit. Recalling the definition of Blinn's BRDF Eq. (2.26), this leads to

$$\begin{aligned} P(\mathbf{I}|\boldsymbol{\theta}) &\sim \mathcal{N}(\mu, {\sigma_\mathrm{M}}^2) \\ \mu &= \frac{I_0}{r^2}\left(\frac{k_\mathrm{d}}{\pi} + k_\mathrm{s}\frac{(\gamma+2)(\gamma+4)}{8\pi(2^{-\frac{\gamma}{2}}+\gamma)}\,\langle \mathbf{h}\cdot\mathbf{n}\rangle^\gamma\right)\langle \mathbf{l}\cdot\mathbf{n}\rangle \\ 1 &= k_\mathrm{d} + k_\mathrm{s}. \end{aligned} \tag{5.1}$$

A simplification is introduced here with $1 = k_\mathrm{d} + k_\mathrm{s}$ that is considered valid, since the overall brightness can still be scaled by I_0. Thus the likelihood depends on four parameters: Light source intensity I_0, diffuse and specular weight k_d (since $k_\mathrm{s} = 1 - k_\mathrm{d}$), cosine lobe exponent γ and variance σ^2. Each parameter requires a prior function to define the expected distribution of the parameter before any input data are evaluated. In many cases normal distributions centered about a target value are a suitable choice. These are chosen here as well and the priors of the four parameters are given as

$$\begin{aligned} P(I_0) &\sim \mathcal{N}(\tilde{\mu}_{\mathrm{I_0}}, \tilde{\sigma}^2_{\mathrm{I_0}}) \\ P(k_\mathrm{d}) &\sim \mathcal{N}(\tilde{\mu}_{\mathrm{k_d}}, \tilde{\sigma}^2_{\mathrm{k_d}}) \\ P(\gamma) &\sim \mathcal{N}(\tilde{\mu}_{\gamma}, \tilde{\sigma}^2_{\gamma}) \\ P(\sigma) &\sim \mathcal{N}(\tilde{\mu}_{\sigma}, \tilde{\sigma}^2_{\sigma}). \end{aligned} \tag{5.2}$$

At this point, it is possible and useful to introduce any available knowledge about the possible range of the parameters into the prior distribution. In addition to parameter ranges defined by physical constraints – for example diffuse and specular weight factors or the index of refraction – this can be applied to parameters that have been coarsely estimated in a previous step. The light source intensities are calibrated, yet may be prone to simplification errors. Here, the previously estimated value can easily be used as the mean value $\tilde{\mu}_{\mathrm{I_0}}$ of the prior distribution, in combination with a relatively small variance $\tilde{\sigma}^2_{\mathrm{I_0}}$. If there is no prior knowledge available or only at a high level of uncertainty, a large prior variance can be used to reduce the actual influence of the prior distribution on the parameter, since its probability is (almost) equal in the entire parameter space. The latter case is termed an *uninformative prior* in a Bayesian framework. Replacing the normal distributions Eq. (5.2) with uniform distributions would completely nullify the effect of the prior distribution and can thus be a valid choice if the aim is not to bias the estimation. However, appreciating existing information is much more plausible, and therefore the Gaussian distributions are chosen. Of course, all kinds of probability functions could be chosen to model the distribution of the parameters, but using a Gaussian distribution is always a suitable choice if the superposition of different sources of uncertainty – both modeling errors and acquisition noise – can be expected. Furthermore, the impact of the prior distribution on the parameter estimation becomes small when the amount of data is sufficiently high.

Concluding from Bayes' law the probability distribution of the set $\boldsymbol{\theta}$ of model parameters corresponding to the input intensity data $\mathbf{I}$ and the a-priori probability $P(\boldsymbol{\theta})$ can now be computed.

$$P(\boldsymbol{\theta}|\mathbf{I}) = \frac{P(\mathbf{I}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})}{P(\mathbf{I})} \tag{5.3}$$

It is difficult to estimate the normalization factor $P(\mathbf{I})$ and, therefore, usually omitted since it is constant as long as the input data does not change. Hence, the posterior-density can be expressed as

$$P(\boldsymbol{\theta}|\mathbf{I}) \propto P(\mathbf{I}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta}) \tag{5.4}$$

Consequently, the only information required to compute the probability distribution[1] of the set of model parameters for the given input data are the distribution of likelihood and prior, both of which have already been derived.

5.2 Markov Chain Monte Carlo Methods

Notwithstanding the choice of the prior functions – introducing available information or completely uninformative – the exact posterior distribution $P(\boldsymbol{\theta}|\mathbf{I})$ remains unknown and

---

[1] More accurately, a function that is proportional to the unknown probability distribution.

has to be estimated. *Markov chain Monte Carlo* (MCMC) methods as e.g., introduced by Ntzoufras (2011) are an elegant way to sample an unknown distribution from existing input data. To achieve this, a Markov chain is constructed whose stationary distribution corresponds to the unknown target distribution. Hence, after convergence, the consecutive states of the Markov chain can be regarded as samples from the unknown posterior.

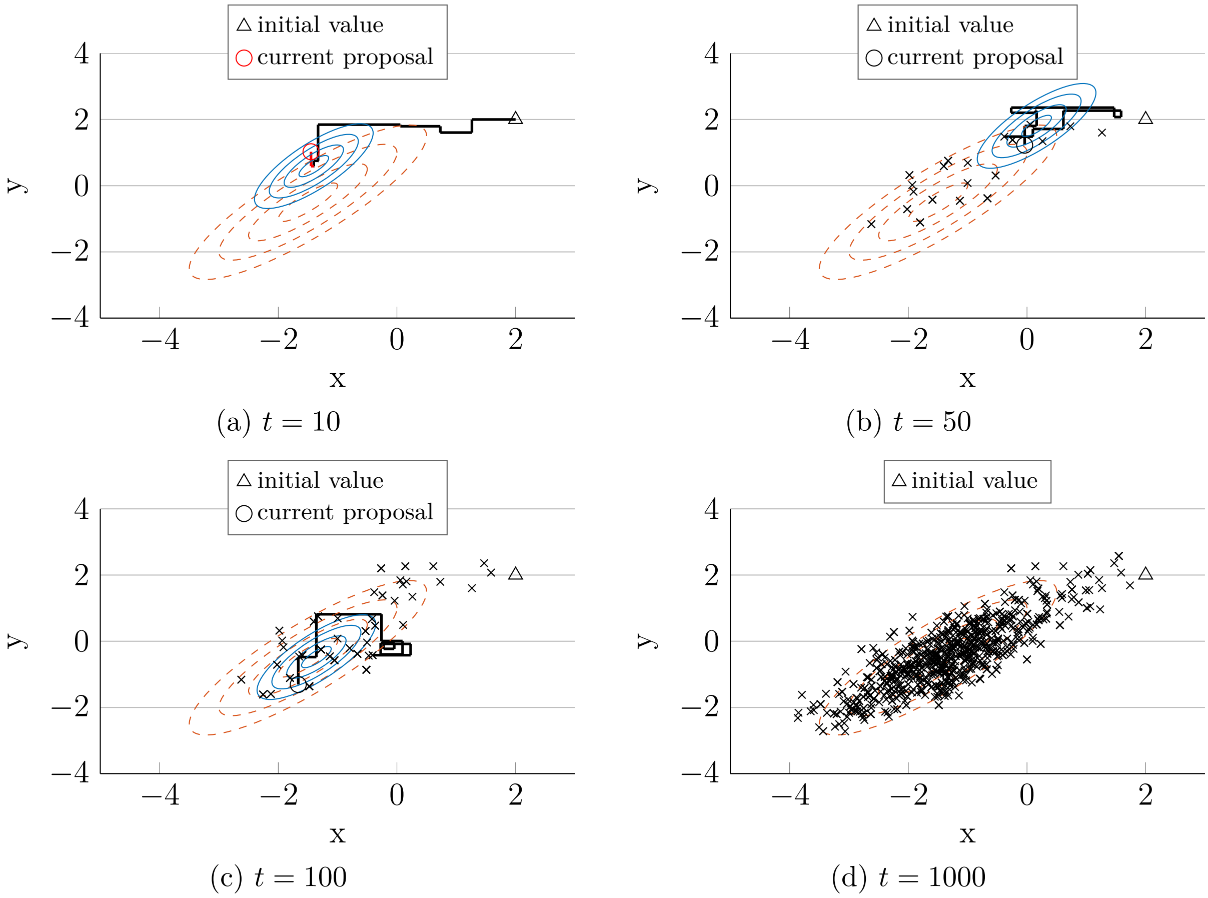
More precisely, the *Metropolis-Hastings* (MH) algorithm introduced by Metropolis et al. (1953) and Hastings (1970) is applied here. Accordingly, the constructed Markov chain proposes in each step new values for the BRDF parameters and the uncertainty of the model. The likelihood of the proposed values is assessed evaluating Eq. (5.1) subject to the acquired image data. If the likelihood of the image data with the given parameters increases, they are probably accepted as new values of the chain. Otherwise, they are probably discarded. The explicitly mentioned term *probably* refers to the acceptance probability ε that is compared to a random number ζ, such that there is a possibility to decline a new value even if the likelihood increases and vice-versa. Starting values $\boldsymbol{\theta}^{(0)} = ({I_0}^{(0)}, {k_\mathrm{d}}^{(0)}, m^{(0)}, \sigma^{(0)})$ of the Markov chain have to be set by the user and can be used to include additional information even for uninformative priors. An example of an MCMC parameter sampling is given in Fig. 5.1. The target distribution and the proposal distribution are illustrated by the dashed orange and solid blue contour, respectively. Here, the parameter space is two-dimensional and from the initial value (denoted by the triangle symbol) new values are proposed, and accepted or discarded. Note that in Fig. 5.1a the proposed value is marked in red as it is not accepted as a valid sample of the posterior distribution. In Fig. 5.1d $t = 1000$ iterations have been evaluated and the accepted samples already cover the target distribution very well.

The number of iterations of the algorithm T is set to a fixed value. To enable the generation of new states, a proposal distribution $P(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$ has to be specified to propose new values based on the last valid parameter value and the proposal variance $\hat{\sigma}^2$. The proposal variance controls the speed of the convergence and the variety of proposed values if the current estimate is already close to the real value and has to be tuned manually.

$$\begin{aligned} P({I_0}^{(t)}|{I_0}^{(t-1)}) &\sim \mathcal{N}({I_0}^{(t-1)}, (\hat{\sigma}^2_{\mathrm{I_0}}) \\ P({k_\mathrm{d}}^{(t)}|{k_\mathrm{d}}^{(t-1)}) &\sim \mathcal{N}({k_\mathrm{d}}^{(t-1)}, (\hat{\sigma}^2_{\mathrm{k_d}}) \\ P(\gamma^{(t)}|\gamma^{(t-1)}) &\sim \mathcal{N}(\gamma^{(t-1)}, (\hat{\sigma}^2_{\gamma}) \\ P(\sigma^{(t)}|\sigma^{(t-1)}) &\sim \mathcal{N}(\sigma^{(t-1)}, (\hat{\sigma}^2_{\sigma}) \end{aligned} \tag{5.5}$$

The acceptance probability ε is computed from the change of likelihood and prior and compared to the random number ζ to accept or discard the new step of the Markov chain. If there is no interdependence between the parameters – as it is assumed here – a *componentwise* MH can be applied and each parameter is proposed and accepted individually (Ntzoufras, 2011). Hence, the change of the posterior estimate simplifies as the priors of all other parameters are canceled. Given here exemplarily for the new estimate ${I_0}^{(t)}$.

$$\begin{aligned} &\frac{P(\mathbf{I}|{I_0}^{(t)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t)})P({k_\mathrm{d}}^{(t-1)})P(\gamma^{(t-1)})P(\sigma^{(t-1)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t-1)})P({k_\mathrm{d}}^{(t-1)})P(\gamma^{(t-1)})P(\sigma^{(t-1)})} \\ =&\frac{P(\mathbf{I}|{I_0}^{(t)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t-1)})} \end{aligned} \tag{5.6}$$



(a) $t = 10$

(b) $t = 50$

(c) $t = 100$

(d) $t = 1000$

Figure 5.1. Example of a Metropolis-Hastings parameter sampling in a two-dimensional parameter space. The unknown target distribution is given by the orange dashed contour and the blue contour displays the proposal distribution that determines the newly proposed values based on the last valid sample. The last 10 iterations are illustrated by the solid black line. Note that the proposal in a) is discarded and that the proposed values converge fast to the target distribution. After $t = 1000$ the unknown distribution is already accurately sampled by the Metropolis-Hastings algorithm.

Applying this to all parameter estimates yields their corresponding acceptance probabilities

$$\begin{aligned}
\varepsilon_{I_0} &= \min\left(1, \frac{P(\mathbf{I}|{I_0}^{(t)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({I_0}^{(t-1)})}\right)\\
\varepsilon_{k_\mathrm{d}} &= \min\left(1, \frac{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({k_\mathrm{d}}^{(t)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P({k_\mathrm{d}}^{(t-1)})}\right)\\
\varepsilon_{\gamma} &= \min\left(1, \frac{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t)}, \sigma^{(t-1)})P(\gamma^{(t)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P(\gamma^{(t-1)})}\right)\\
\varepsilon_{\sigma} &= \min\left(1, \frac{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t)})P(\sigma^{(t)})}{P(\mathbf{I}|{I_0}^{(t-1)}, {k_\mathrm{d}}^{(t-1)}, \gamma^{(t-1)}, \sigma^{(t-1)})P(\sigma^{(t-1)})}\right).
\end{aligned} \tag{5.7}$$

The componentwise MH yields furthermore the general advantage that simple proposal distributions can be applied. Although possibly beneficial with respect to the convergence time of the chain, the definition of a combined proposal distribution requires careful design and is oftentimes too tedious to be applied. Computing the likelihoods of current and proposed parameters in the log-space, thus computing the so-called log-likelihood, facilitates the equations and furthermore circumvents explicitly evaluating the exponential function. If exponential distribution functions are used, e.g., Gaussian distributions, these evaluations are very costly in terms of computation time.

Depending on the chosen initial values of the Markov chain it takes a certain amount of time, i.e., iterations, for the chain to converge to the unknown posterior it is supposed to approximate. This period is called *burn-in* phase and has to be estimated by the user since, obviously, only values of the chain *after* convergence sample the posterior. However, no universal criterion determines convergence, although (individual) parameters that change little throughout the various proposal iterations are a strong indicator. If multiple Markov chains are computed in parallel from different seeds, so as to prevent bias from badly chosen starting values or rather increase variety, a possible criterion for convergence is the *potential scale reduction factor* (PSCF) proposed by Gelman and Rubin (1992) and Gelman (1995). At $T \to \infty$ all sequences have converged to the target distribution and thus yield the same variance, whereas the variance between the sequences is large at the state of non-convergence and thus the parameter space requires further exploration. However, since parallel sequences have not been applied due to limits in computational time PSCF cannot be applied. The Markov chain has been shown to converge to a stationary distribution (Gilks et al., 1996, p.46) if the following properties are met:

- irreducibility
- positive recurrence
- aperiodicity

Once converged to the unknown posterior, all new steps of the Markov chain will generate values following the target distribution and can be regarded as samples from the posterior. Thus, the samples can be used to infer statistics since the adaption of the BRDF model parameters to unknown data immediately yields a confidence region for each parameter. Additionally, the variance of the likelihood σ^2 is a measure of the accuracy of the model, such that not only the model parameters can be assessed regarding their estimation certainty, but the validity of the chosen model as well.

5.3 Reversible Jump Markov Chain Monte Carlo

So far it is feasible to adapt the parameters of a BRDF model to unknown intensity data, yet the segmentation of the surface has not been addressed. Image segmentation is usually performed in a generative way (e.g., using Gaussian Mixture Models (GMM)). Here instead a discriminative approach is used, and direct inference of the borders of the surface patches from the data is achieved. Effectively, a change in surface structure or material enforces a change of the model parameters to reproduce the received impression accurately, and the task is to find the spots where another set of parameters has to be used.

Green (1995) described so-called *reversible jump Markov chain Monte Carlo* (RJMCMC) methods that can simultaneously explore the parameter and model space. In terms of a Markov chain, this means that there is a certain probability in each iteration of creating, deleting or modifying a *change point* τ that separates two sets of model parameters. By finding the correct change points, unknown data can be divided into meaningful segments. Among the contributions of this thesis is a proposal on how to extend the concept of change points to the two-dimensional application of image segmentation. This extension comprises the following new properties:

- change points are possible both in u and v direction (referring to sensor coordinates);
- a change point becomes a vector of change points since it will affect the entire row or column of image pixels;
- each pixel coordinate of the change point vector can be moved individually; *and*
- surface segments are constructed from the regions where the same change points in u and v direction coincide.

Additionally, since segment borders can cover less than an entire image row or column, change points are also allowed to lie outside of the image coordinate frame. The concept of change points and their possible movement within an image is illustrated in Fig. 5.2. Change points are proposed for an entire row or column of the image and initialized with the same random value. However, the change points can be moved individually for each pixel by the RJMCMC algorithm. Three exemplary situations are given in Fig. 5.2. The black arrows indicate the movement of the change points with respect to their initial values. Fig. 5.2a depicts the simple example of one *horizontal vector* of change points that can be moved *vertically* which will result in a maximum of two regions. Since the change points are allowed to lie outside the image frame, which is indicated by the dashed line, it is possible that the regions cover an image row or column only partially. Fig. 5.2b illustrates an additional change point in v direction and a possible solution to the image segmentation problem. Fig. 5.2c displays a more complex example of a vector of change points in both v and u direction. Note how region 3 disappears in the final result since there is no overlap any more of the image region *right* of the green vector and *above* the blue vector. Region 3 has literally been pushed out of the image frame. This feature allows for a great flexibility in the segmentation.

A possible maximum of change points in each direction has to be defined by the user. An indicator matrix is constructed based on these maxima to assign an index to each region based on the combinations of change points. For example, $\tau_{\max}(u) = \tau_{\max}(v) = 3$ yield the following assignment:

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ (3,1) & (3,2) & (3,3) \end{pmatrix} \hat{=} \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \tag{5.8}$$

Each change point itself has a statistical distribution, such that the data of the Markov chain can be used to compute confidence levels of the positions of the segment borders as well. This is a noteworthy advantage compared to other segmentation procedures since the algorithm yields a measure for the quality of the segmentation. The change points are assumed to follow a uniform distribution

$$\tau \sim \mathcal{U}(-1, \mathrm{imagesize} + 1) \tag{5.9}$$

since they are equally probable on the entire surface. Additionally, the dimension are increased with respect to the number of image pixels as the change points are allowed to lie outside the image plane.

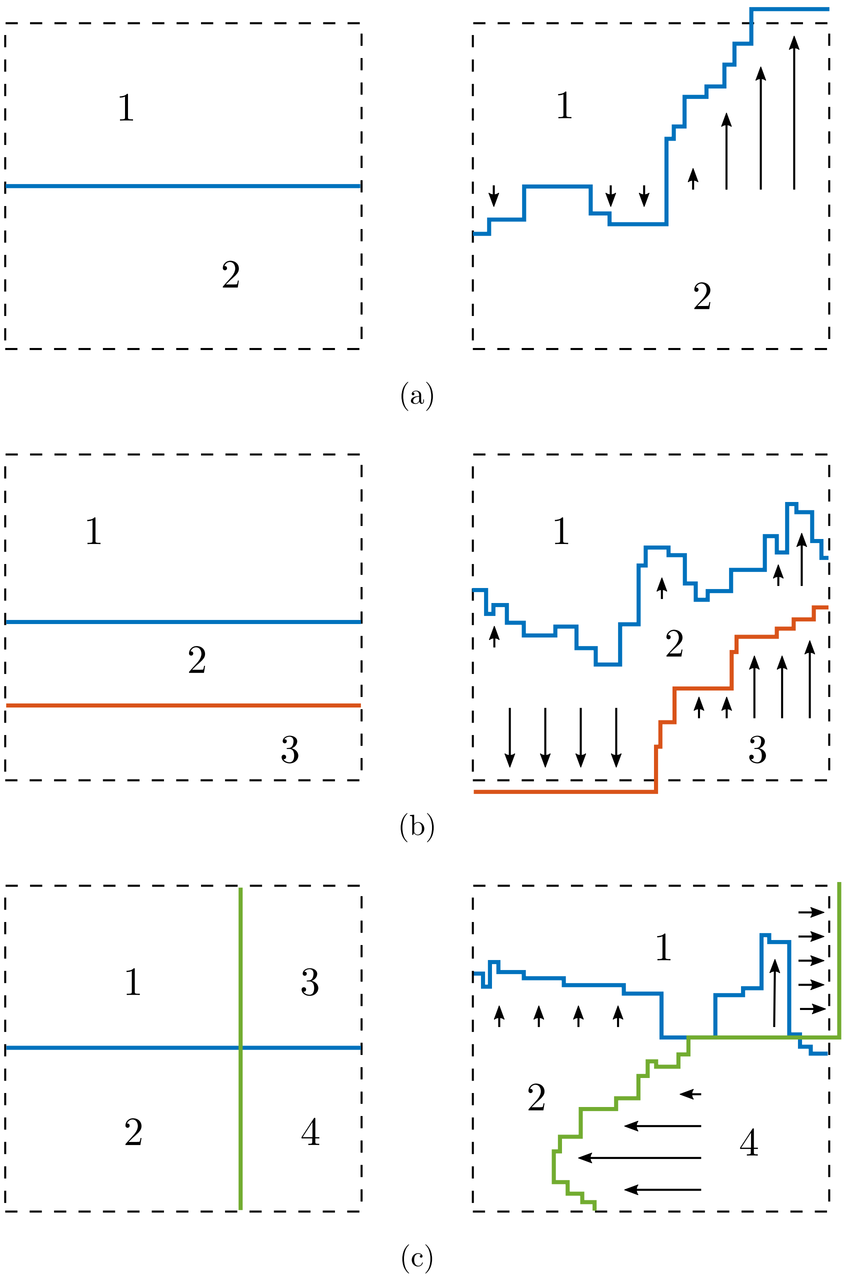


Figure 5.2. Illustration of the concept of change points in two dimensions. The initial situation of a) one horizontal vector, b) two horizontal vectors and c) one vertical and one horizontal vector of change points is given in the left column. The numbers indicate the associated image regions and the positions of the change points after convergence are given in the right column. Note that change points can lie outside the image frame (dashed line) and therefore create patches that do not cover an entire image row or column. Note further how region 3 disappears in the third example since the remaining vertical change points belongs to the lower image region 4. The black arrows indicate the movement of the change points with respect to their initial position during convergence.

5.3.1 Changing the Dimensionality of the Model

Creating or deleting a change point – which here always means a vector of change points – is in terms of reversible jump Markov chain Monte Carlo a change in dimensionality. The model of the surface (segmentation) changes drastically as the association of the image pixels to a certain region or cluster can change completely if a change point is added or removed. Consider for example Fig. 5.2c and remove the horizontal change point (green vector), the resulting segmentation would be rather similar to Fig. 5.2a. The following section covers the implications of a change in dimension on two-dimensional data.

The possibility to create or delete change points lets the needs arise for a procedure to split and merge parameter sets $\boldsymbol{\theta}_i$. Multiple options have been considered. A new region could be initialized with the initial parameters of the chain or the parameters of a randomly chosen neighboring region. Additionally, it could prevent getting stuck in local maxima of the posterior to merely draw random parameters from the prior distributions. Yet all of these options make it impossible to revert the step without storing a large amount of data. A more suitable solution is to have a bijective transformation $\mathbf{M}_\mathrm{D}$ which is independent of the actual number of change points.

Say the dimensionality, i.e., the number of existing surface patches, is increased from two to three, the following procedure can be (and is) used to derive the new sets of BRDF parameters $(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, \boldsymbol{\theta}_3^{(t)})$ from the currently established sets of parameters $(\boldsymbol{\theta}_1^{(t-1)}, \boldsymbol{\theta}_2^{(t-1)})$. A random number ζ is drawn from the distribution $q(\zeta) = \mathcal{N}(0,1)$ and symmetrically added and subtracted from the current estimate of the BRDF parameters. Inspired by Pascal's triangle the coefficients $P_\mathrm{n}(k) = (-1)^k \binom{n}{k}, k = 0, \ldots, n$ of the last column of $\mathbf{M}_\mathrm{D}$ create a unique transformation when "jumping" from dimension $n{-}1$ to n. For the step to $n = 3$ and former parameters $\boldsymbol{\theta}_1^{(t-1)}$ and $\boldsymbol{\theta}_2^{(t-1)}$ this means

$$\begin{pmatrix} \boldsymbol{\theta}_1^{(t)} \\ \boldsymbol{\theta}_2^{(t)} \\ \boldsymbol{\theta}_3^{(t)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0.5 & 0.5 & 2 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_1^{(t-1)} \\ \boldsymbol{\theta}_2^{(t-1)} \\ \zeta \end{pmatrix}. \tag{5.10}$$

Note that the transformation has to be applied componentwise to $\boldsymbol{\theta}_i$, thus the random number ζ can be scaled to match the dimension of the parameter. The inverse of $\mathbf{M}_\mathrm{D}$ is defined and the Markov chain can be reset to its state in the lower dimension if necessary. This is regarded to be helpful to reduce computational time since the chain can continue from its former state, possibly nearer to convergence.

A change in dimension is only a proposal, and it is accepted with an acceptance probability ε_D similar to the proposition of new BRDF parameters. The dimension is represented by a model number $\tilde{m}$ that is determined from the indicator matrix. For the sake of readability the iteration index t will be omitted in what follows, and an apostrophe is indicating the proposed value.

$$\varepsilon_\mathrm{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left(1, \frac{P(\boldsymbol{\theta}', \tilde{m}'|\mathbf{I}) P(\tilde{m}|\tilde{m}') q(\zeta')}{P(\boldsymbol{\theta}, \tilde{m}|\mathbf{I}) P(\tilde{m}'|\tilde{m}) q(\zeta)} \left| \frac{\partial \boldsymbol{\theta}'}{\partial(\boldsymbol{\theta}, \zeta)} \right| \right) \tag{5.11}$$

$P(\tilde{m}'|\tilde{m}) = P(\tilde{m}|\tilde{m}')$ is the probability to change from one model to the other, i.e., add or delete a change point respectively and here assumed to be equal. $\left|\frac{\partial \boldsymbol{\theta}'}{\partial(\boldsymbol{\theta}, \zeta)}\right|$ is the Jacobian of the partial derivatives, necessary since what is happening here is basically a coordinate transformation. Yet, the Jacobian is only a scaling factor, which may be omitted and thus ε_{D} simplifies to

$$\varepsilon_{\mathrm{D}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left(1, \frac{P(\boldsymbol{\theta}', \tilde{m}'|\mathbf{I})q(\zeta')}{P(\boldsymbol{\theta}, \tilde{m}|\mathbf{I})q(\zeta)}\right). \tag{5.12}$$

The acceptance of the inverse move is simply calculated as

$$\varepsilon_{\mathrm{D}}{}^{(-1)}(\boldsymbol{\theta}', \boldsymbol{\theta}) = \min\left(1, \frac{P(\boldsymbol{\theta}, m|\mathbf{I})q(\zeta)}{P(\boldsymbol{\theta}', m'|\mathbf{I})q(\zeta')}\right) \tag{5.13}$$

After the predefined number of T iterations, it has to be decided which combination of change points is the most probable one. The integration in the model space is turned into a summation in the discrete case, and the model with the most iterations is accepted as the best solution (e.g., King et al., 2010). However, comparing the “time spent” in different models yields information about the certainty of the solution.

5.4 Application

The proposed method is applied to both synthetic and laboratory data. The synthetic data consist of 12 images of 121 × 121 pixels of the surface of a sphere that are rendered with the Lambert-Blinn BRDF model under changing illumination and varying parameter sets. The surface is divided into four segments, and the BRDF parameters of the individual patches are given in Table 5.3 (colored values). Gaussian noise is added to the image and depth data to create more realistic conditions according to $\sigma_{\text{image}} = 0.5$ and $\sigma_{\text{depth}} = 0.001$. The laboratory dataset contains both, 18 HDR images and pixel-synchronous depth data of a painted plaster object that have been acquired with the setup described in Chapter 4. The illumination environment is calibrated according to Lenoch et al. (2012).

5.4.1 Synthetic Data

The algorithm is evaluated first on the synthetic dataset with $T = 20\,000$ iterations. Figure 5.3a shows one of the noise-afflicted input images and Fig. 5.3b the resulting reflectance map. The remaining images of the artificial dataset and their corresponding reflectance maps can be found in Fig. A.2. The reflectance maps exhibit a compelling similarity to the input image data disregarding the additional noise, and the root mean squared error $E_{\text{RMSE}} = 0.5197$ which directly reflects the additional noise on the input data σ_{image}. The regions will be referred to by their numbers indicated in Fig. 5.3c for the following analysis. The algorithm misplaces only very few change points separating the two lower regions. A second non-ideal behavior can be noted in the upper right region, as the surface patch is over-segmented into three different regions. However, the computed reflectance maps show neither the patch borders nor a significant deviation from the input data. The details of this effect will be assessed later.

The parameters of the algorithm consisting of the variance $\hat{\sigma}^2$ of the proposal distribution, the prior parameters and the starting values of the Markov chain $\boldsymbol{\theta}_i^{(0)}$ are stated in Table 5.2. The starting values have purposely been selected far from the true values of the target distributions to pose a challenging situation. Note that the priors have little influence on the parameter inference if the pool of data is sufficiently large since the prior is only one probability value that is outweighed by one value per data point. The histogram of visited

Table 5.2. Initial values $\boldsymbol{\theta}_i^{(0)}$ of Markov chain, variance of proposal distribution $\hat{\sigma}_i^2$, prior parameters $\tilde{\mu}$ and $\tilde{\sigma}^2$ on the synthetic dataset.

	$\boldsymbol{\theta}_i^{(0)}$	$\hat{\sigma}_i^2$	$\tilde{\mu}$	$\tilde{\sigma}^2$
I_0	1	1	80	100
k_d	0.5	0.1	0.5	1
γ	25	1	25	100
σ^2	10	0.1	0.5	0.1

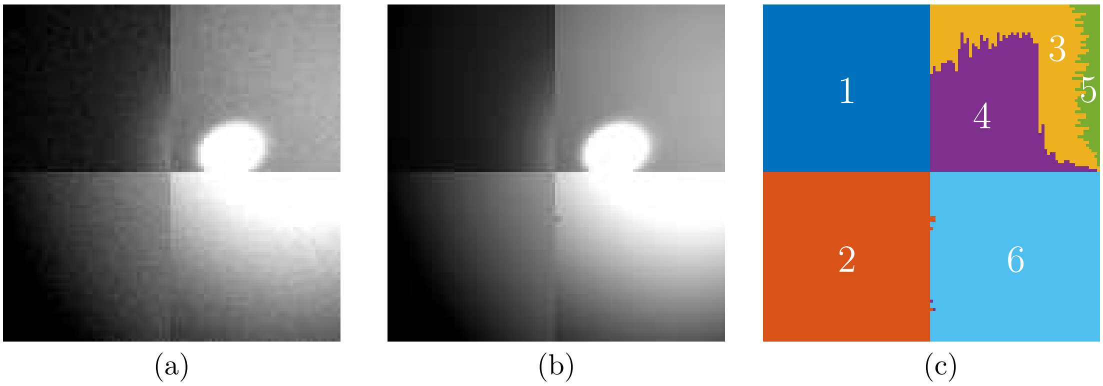


Figure 5.3. (a) Exemplary input image of synthetic dataset and (b) reflectance map computed from the algorithm. Same grey value scaling in both images. (c) The separation of the regions is almost perfect, however the surface is over-segmented in the upper right region.

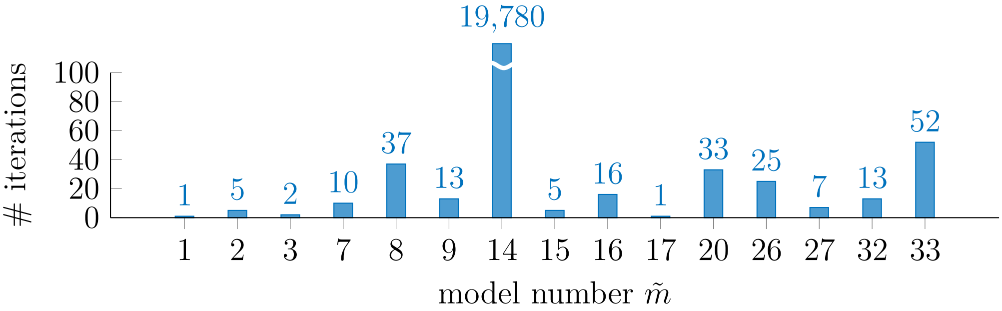


Figure 5.4. Histogram of visited models in model space of synthetic dataset. Counting indicated that model 14 is the most likely one. Note that not all models have been proposed or accepted by the algorithm.

models is depicted in Fig. 5.4. The histogram counts indicate strongly that model $\tilde{m} = 14$ is the best choice. This model equals one change point in v and two in u direction. Of course, as already stated the surface is over-segmented, and the ideal solution would have been one change point in each direction. The likelihood of this solution is extremely high since all other models have been rejected. The certainty of the algorithm based on the count of evaluations is about 98.9 %.

The sampled a-posteriori distribution of the set $\boldsymbol{\theta}$ of model parameters is shown along with the iterations of the Markov chain in Fig. 5.5. Convergence of k_d is slow on region 6 and γ appears not to converge to a solution at all on regions 3 and 5 which is apparent

Table 5.3. Maximum a-posteriori estimate of parameters and 1$-\sigma$ uncertainty for all surface segments after a burn-in phase of 20 %. The uncertainty is given with respect to the last corresponding digits. The green values have been used to render the synthetic images and are therefore the target values of the optimization. The orange and yellow values indicate ground truth values that exhibit large discrepancies with respect to the estimates. The orange values occurred on unnecessary surface patches.

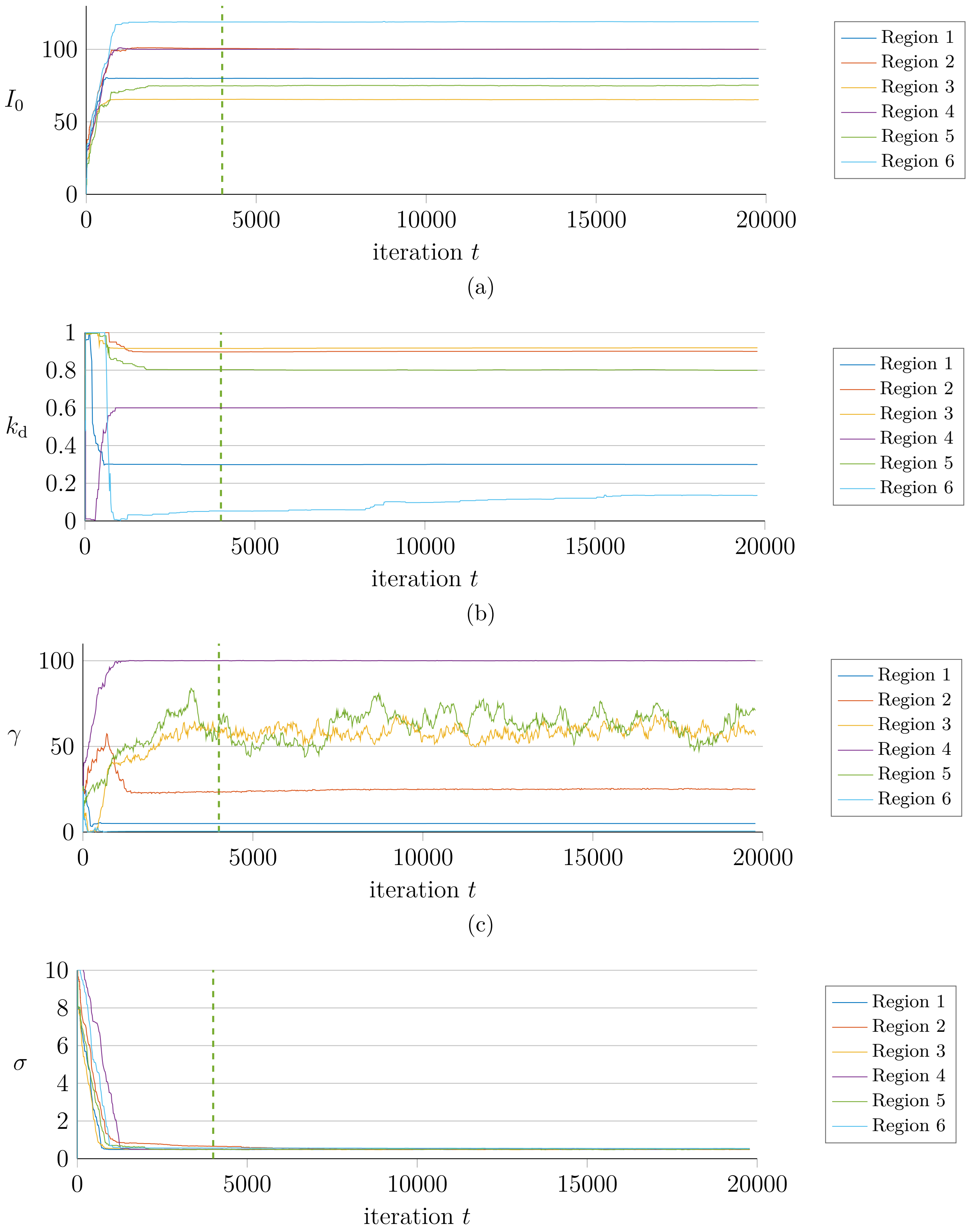
region	I_0		k_d		γ		σ	
1	79.9884(272)	80	0.2999(6)	0.3	4.9992(100)	5	0.4900(20)	0.5
2	100.1142(1666)	100	0.8995(11)	0.9	24.797(474)	25	0.5203(402)	0.5
3	65.3714(771)	100	0.9176(10)	0.6	58.47(349)	100	0.4901(56)	0.5
4	100.0444(184)	100	0.6002(1)	0.6	100.0445(490)	100	0.5018(35)	0.5
5	74.9820(1395)	100	0.8012(10)	0.6	62.60(794)	100	0.5001(99)	0.5
6	118.9728(889)	120	0.1018(313)	0.4	0.5650(256)	1	0.5538(84)	0.5

by the constantly changing proposed and accepted values throughout all iterations. Since convergence is fast for all remaining parameters on all other regions, a burn-in phase of 20 %, marked by the green dashed line in Fig. 5.5, is assumed sufficient. The inferred reflectance parameters and their uncertainties are listed in Table 5.3 together with the true values that have actually been used to generate the synthetic test data (colored values). The regions 1 and 2 that are nearly perfectly segmented exhibit mean values of the reflectance parameters that include the ground truth in their very narrow confidence tube. Region 6 is very well segmented as well. However, the parameters k_d and γ differ notably from the target values. The upper right region that is segmented into the regions 3, 4 and 5 exhibits ambivalent parameter estimates. Region 4 matches the target values similar to regions 1 and 2 with a very narrow tube of confidence, while regions 3 and 5 again differ strongly from the ground truth. Nonetheless, the rendered images equal the input data on the upper right region.

Error Analysis

The over-segmentation on the upper right region is interesting since the BRDF parameters differ notably from one another while the resulting rendered image is very similar. Examining the input data reveals that regions 3 and 5 exclusively contain the diffuse component of the reflection and thus the effective diffuse component $\hat{L}_\mathrm{d} = I_0 k_\mathrm{d}$ is essentially equal. A lack of salient data always deteriorates the estimation of the specular component as is very well visible in the high uncertainty in Table 5.3 and Fig. 5.5c, and the remaining effective diffuse component is equal for all sets of parameters as depicted in Table 5.4. Region 6 contains both, specular and diffuse reflection component at a sufficient amount, and still, the parameters do not converge to the target values although the likelihood of the rendering is very high. Figure 5.6 depicts the perceived intensity subject to the ground truth parameters and the estimated values, $\cos(\vartheta_\mathrm{in})$ is fixed as it provides a global scaling. It is evident that the similarity between both sets of parameters is very high for $\vartheta_\mathrm{h} < 70°$, which explains the convergence issues and points again to a lack of data. The estimated parameters do not equal the ground truth but are very likely to explain the data.

Hence, in both cases of falsely estimated parameters answers have been found to the Bayesian question of the existence of other sets of parameters that plausibly explain the data, i.e., yield a similar likelihood compared to the true target values. This underlines the



(a)

(b)

(c)

(d)

Figure 5.5. States of Markov chain for different parameters. The green dashed line indicates the end of the burn-in phase after 20 % of the total iterations. All parameters appear to have converged to their individual target distribution with the exclusion of γ on regions three and five.

Table 5.4. Similarity of effective diffuse component $\hat{L}_\mathrm{d}$ for different parameter sets.

region	I_0	k_d	$\hat{L}_\mathrm{d}$
3	65.3714	0.9176	59.9825
4	100.0444	0.6002	60.0451
5	74.9820	0.8012	60.0776

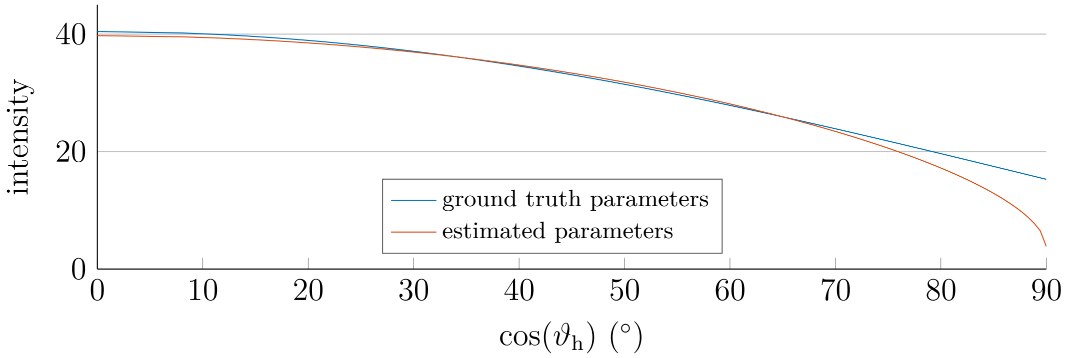


Figure 5.6. Similarity of the intensity value for the estimated parameters of region 6 and ground truth values. $\cos(\vartheta_\mathrm{in}) = 1$ is fixed, as it provides only a scaling of the BRDF value. The similarity is very high for $\vartheta_\mathrm{h} < 70°$ and thus the estimated parameters are very likely to explain the data.

Table 5.5. Initial values $\boldsymbol{\theta}_i^{(0)}$ of Markov chain, variance of proposal distribution $\hat{\sigma}_i^2$, prior parameters $\tilde{\mu}$ and $\tilde{\sigma}^2$ on the laboratory dataset.

	$\boldsymbol{\theta}_i^{(0)}$	$\hat{\sigma}_i^2$	$\tilde{\mu}$	$\tilde{\sigma}^2$
I_0	1	1	1	1
k_d	0.5	0.1	0.5	1
γ	25	1	25	100
σ^2	10	0.1	0.1	100

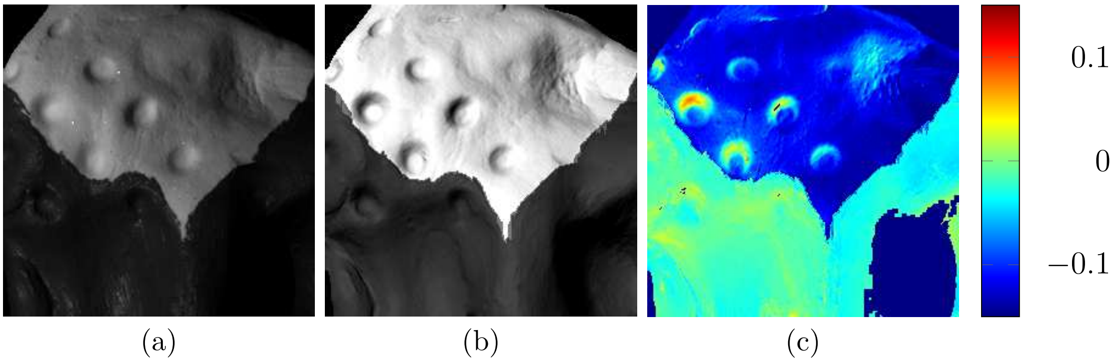
importance of a large dataset to infer the model parameters from the data reliably. However, the synthetic data should be regarded as a proof-of-concept, and applying the method to real-world data will be a more valuable assessment.

5.4.2 Laboratory Data

The test object of the laboratory dataset is made of shaped plaster that has been divided into three patches, two of which are painted with green and blue acrylic paint, respectively. The design intended to obtain a geometrically complex shape and at the same time multiple reflection characteristics at different surface areas. However, due to the image acquisition process with a monochrome camera, both painted regions appear very similar even to a human viewer. Hence, two regions can effectively be separated. One is bright white and exhibits mostly diffuse reflection while the other is dark grey and shows increased specular reflection that is caused by the shiny paint. Table 5.5 displays the initial values of the Markov chain $\boldsymbol{\theta}_i^{(0)}$, the variance $\hat{\sigma}^2$ of the proposal distribution and the prior parameters.

The object is depicted in Fig. 5.7a and Fig. 5.7b displays the corresponding reflectance map as a result of the proposed algorithm. It is based on the the estimated parameters and the surface segmentation. A pixelwise error map is depicted in Fig. 5.7c. For the sake of clarity, only one image pair of the entire laboratory dataset is given here, the remaining images and reflectance maps can be found in Fig. A.3. The gray value scaling of both images is identical, and it is obvious that the upper area of the reflectance map is too bright compared to the input image and the specular component of the lower region is too small while the segmentation of both regions is well-achieved. The largest deviations can be spotted at the dents in the upper region where the interreflections on the real dataset deteriorate the performance of the BRDF model. Hence, the overall brightness of the upper region is too high compared to the input data, and the intensity in the shadowed parts of the dents is still too low. Considering the likelihood of the measurements as normally distributed around the target values, this is well explained by a certain variance. However, there is a certain lack in the accuracy of the reproduction of the input data.

Although a segmentation of the surface based merely on intensity might be considered suitable, it has to be emphasized again that it, in contrast, is based on the likelihood of the BRDF parameters to fit a designated surface patch best. In a segmentation that employs brightness as the only feature, the darker parts of the dents in the upper region could be very well assigned to the lower area. Furthermore, the slightly glossy specular component of the lower region yields an intensity which is in turn close to the bright areas of the upper region. This demonstrates one notable advantage of the reflectance-based surface segmentation: A shiny surface is not separated into different patches due to the fact that the image intensity varies significantly depending on the viewing angle ϑ_{out}, this property is elaborated further in Section 5.5.



(a) (b) (c)

Figure 5.7. (a) One of the captured images, (b) reflectance map of segmented surface with computed parameters. Same scaling of grey values in both images. (c) Pixelwise difference of intensity values.

The procedure suggests that a segmentation of the surface in three segments is the most probable choice, although this contradicts the intuitive segmentation of upper and lower region. Figure 5.8 indicates that the model $\tilde{m} = 3$ – two vectors of change points in v direction – was visited $T_3 = 19\,565$ times and thus has a probability of $97.825\,\%$ to be the best choice. The third region covers the small area of transition between upper and lower region, where separation is not distinctive as indicated in Fig. 5.10. The proposal of changepoints as a vector of change points, i.e., a line, makes the segmentation of small encapsulated areas improbable and as a consequence, the transition region can neither be

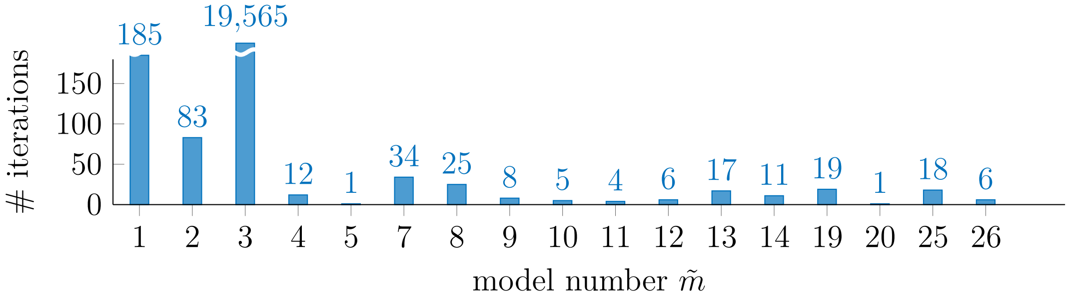


Figure 5.8. Histogram of visited models in model space on laboratory dataset. Counting indicated that model $\tilde{m} = 3$ is the most likely. Note that not all models have been proposed or accepted by the algorithm.

Table 5.6. Maximum a-posteriori estimate of parameters and $1-\sigma$ deviation for upper and lower region after burn-in phase of 20 %.

region	I_0	k_d	γ	σ
1	0.9276(8)	0.9989(28)	0.4966(9146)	43.8229(408) $\times 10^{-3}$
2	0.4410(25)	0.9979(96)	21.7542(147682)	39.1242(3823) $\times 10^{-3}$
3	0.2417(4)	0.9621(6)	31.1956(5846)	14.3936(402) $\times 10^{-3}$

added to the upper or lower area. Hence, the estimated BRDF parameters are a mixture to satisfy both reflectance properties up to a certain limit. The estimated parameters of the three BRDFs and their $1-\sigma$ confidence levels are listed in Table 5.6. While the diffuse component is almost equal in all three patches, the intensity I_0 models the changing albedo of the surface. The estimated noise σ^2 of the likelihood is very small and in the range of the differences that are visible between input images and reflectance maps. The cosine-lobe parameter γ shows a high uncertainty that is caused by the large diffuse component on regions 1 and 3. Since $k_\mathrm{s} = 1 - k_\mathrm{d} \Rightarrow k_\mathrm{s} \approx 0.01 \ldots 0.04$ and the cosine lobe has very little effect on the final reflectance map. As region 2 covers both upper and lower region in the transition area, the uncertainty of γ is even more profound. The larger exponent in the lower region is consistent with the impression of a shinier surface, although the specular component in the reflectance map is too weak.

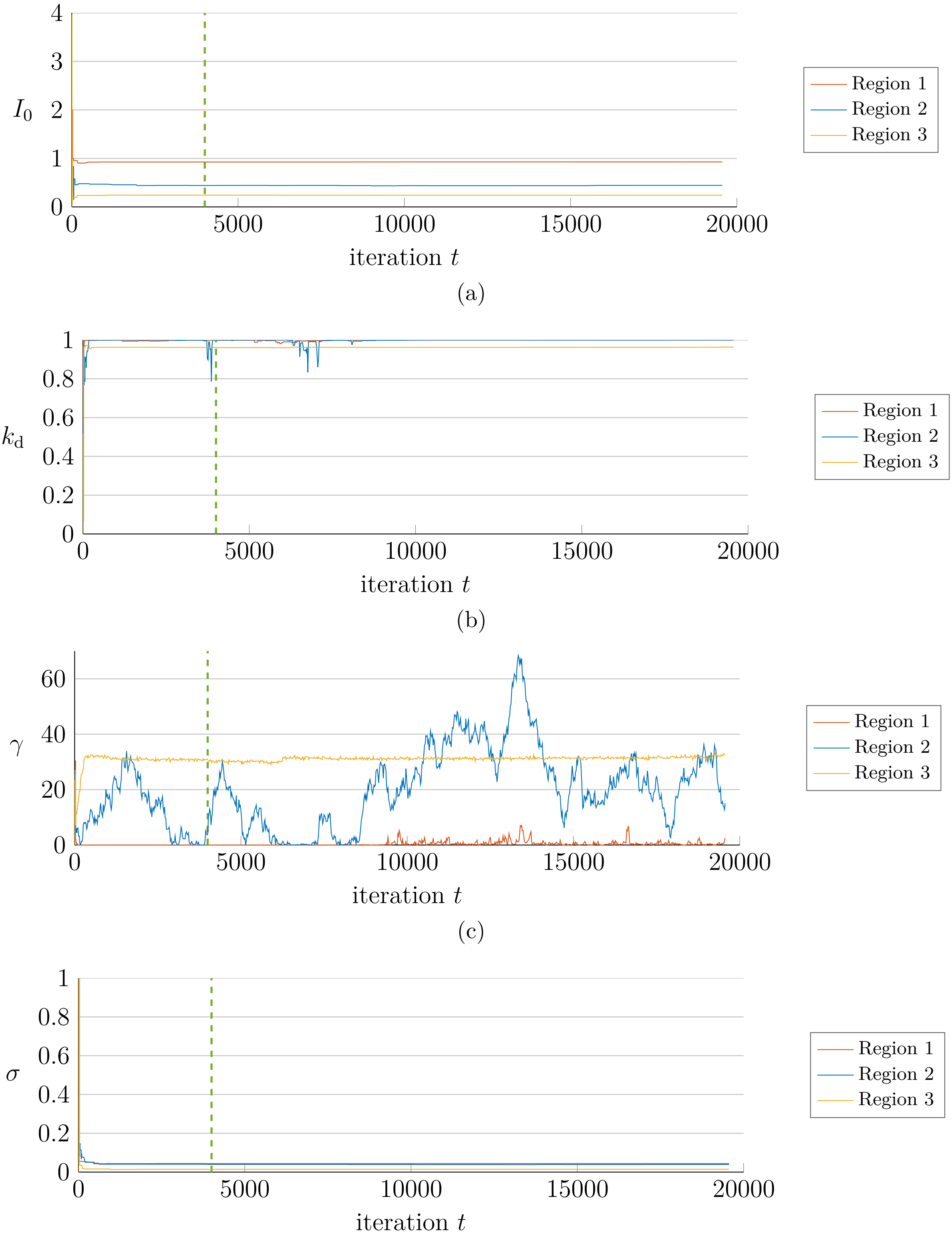
The progression of the Markov chain of $\boldsymbol{\theta}$ on all regions is depicted in Fig. 5.9. The dotted green line marks the end of the burn-in phase after 20 % of the iterations. It is evident that all parameters except γ have converged fast to an almost stationary value. Especially on region 2, γ suffers from the unclear segmentation and the contribution of upper and lower region to the considered input data. However, as already mentioned, the small specular component makes its estimation difficult on all regions.

5.5 Evaluation

To assess the prospects of the presented procedure, the results of the surface segmentation are contrasted with commonly used image segmentation algorithms. These are, in particular, k-means, (e.g., Seber, 1984; Spath, 1985), Gaussian mixture models[2] (McLachlan and

---

[2] The Matlab 2015a implementations of k-means and GMM have been used.



(a)

(b)

(c)

(d)

Figure 5.9. States of Markov chain of different parameters on final regions. Most parameters converge fast to an almost stationary value, γ shows a higher uncertainty on region 2. The green dashed line indicates the end of the assumed burn-in phase after 20% of the total iterations.

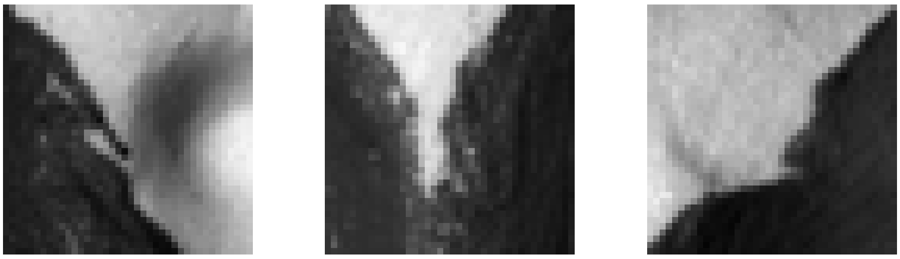


Figure 5.10. Exemplary regions of uncertain affiliation on laboratory data. The contrast has been adjusted to increase clarity.

Peel, 2000), multichannel k-means (Pichler and Hosticka, 1995) and normalized graph cuts (NCuts)[3] (Shi and Malik, 2000). Since the varying intensities and specular highlights of the synthetic dataset are the greater challenges, the cluster algorithms are applied to the synthetic images. Note, however, that it is questionable to contrast the presented algorithm with these other surface segmentation methods since the similarity criterion used to dissect the surface is neither identical nor easily comparable. The proposed RJMCMC method maximizes the average similarity of a measured grey value image and the rendering based on the estimated parameters, while the other methods use image grey values only. The NCuts algorithm is not designed to handle multispectral input data without modification, and thus the image data had to be pre-processed. However, to the author's knowledge, no equivalent method achieves image segmentation and BRDF computation by similar means as the one presented in this chapter.

k-means Multichannel image data can be exploited in a k-means clustering when treating each of the P images as a separate feature per pixel, such that there are $N \times M$ samples (assuming $\mathbf{I}_{N\times M}$) each yielding P features. Each feature, i.e., image, is standardized to exhibit a standard deviation $\sigma = 1$ and a mean $\mu = 0$. Since k-means requires that the user provides the desired number of clusters, different $k \in [4, 8]$ are evaluated. The k-means segmentation of the synthetic dataset is displayed in Fig. 5.11. The segmentation is erroneous for every considered number of clusters. The cause is obvious: Since k-means operates on grey values, the specular highlights of the surface determine the dominant center of the clusters. Only the transition between the upper clusters is detected for $k \geq 6$. The multichannel k-means algorithm as described by Pichler and Hosticka (1995) did not converge to a solution on the given data.

Gaussian mixture model Similar to the k-means algorithm the multiple intensities per image pixel are supplied as standardized features to be fitted with a Gaussian Mixture Model. The result of the fitting is depicted in Fig. 5.12. The GMM handles the data very well and a segmentation expecting 4 clusters almost yields perfect results. The border between the lower patches is slightly distorted. Adding one cluster leads to the isolation of multiple specular highlights in the upper right patch and additional Gaussian modes sub-divide the correct patches or lead to incorrect clusters based on the large central intensities. While the key approach to segmentation of k-means and GMM is similar, the additional degrees of freedom of a GMM regarding the shape and size of each cluster kernel offer the flexibility that is helpful to address the given problem.

---

[3] The existing implementation provided at http://www.cis.upenn.edu/~jshi/software/ was used, no version number available but last updated January 22nd, 2010.


(a) $k = 4$ (b) $k = 5$ (c) $k = 6$ (d) $k = 7$ (e) $k = 8$

Figure 5.11. k-means segmentation of synthetic data, the colors code different surface segments. The segmentation fails for every considered setting.


(a) $k = 4$ (b) $k = 5$ (c) $k = 6$ (d) $k = 7$ (e) $k = 8$

Figure 5.12. Gaussian mixture model segmentation of synthetic data, the colors code different surface segments. k is the number of Gaussian distributions fit to the data. $k = 4$ yields a suitable segmentation, yet the lower segment border is frayed. Adding additional clusters leads to the isolation of multiple specular highlights and the sub-division of correctly determined patches.


(a) Input (b) $k = 4$ (c) $k = 5$ (d) $k = 6$ (e) $k = 7$

Figure 5.13. (a) Input image taken from the synthetic dataset. (b-e) Normalized cuts segmentation of synthetic data, the colors code different surface segments.

normalized cuts Since the normalized cuts algorithm is not designed to handle multispectral image data, three simple forms of pre-processing have been applied to the data. Only one image from the dataset has been selected for the segmentation (Fig. 5.13a), a median image (Fig. 5.14a) and a mean image (Fig. 5.15a) are computed from all input images and supplied to the segmentation. The latter approaches are intended to reduce the influence of specular highlights. Various possibilities of the number of clusters k are evaluated.

The segmentation based on one randomly chosen input image illustrated in Fig. 5.13 fails for all considered numbers of clusters to produce a reasonable result. Setting $k = 5$ yields the best result separating all regions reasonably from one another, but the wide specular lobe in the upper left region prevents correct segment borders. Increasing the number of clusters leads to false segments separating the regions of distinct intensity differences instead

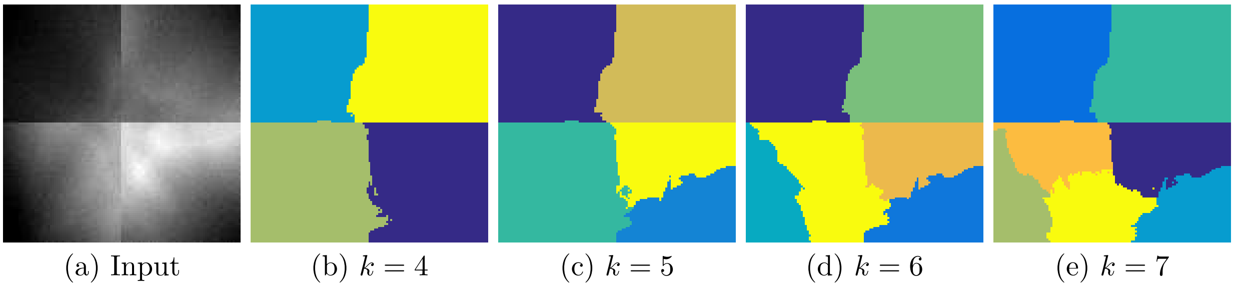


Figure 5.14. (a) Input median image of synthetic dataset. (b-e) Normalized cuts segmentation of median image, the colors code different surface segments. Reasonable results are only achieved for $k \in \{4, 5\}$.

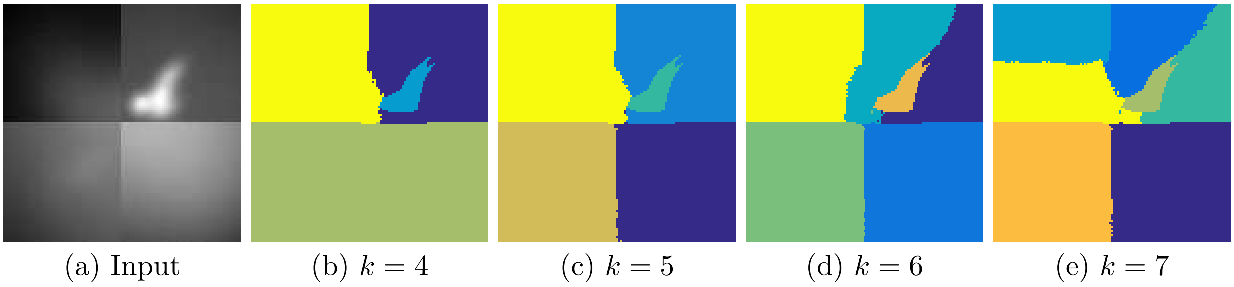


Figure 5.15. (a) Input mean image of synthetic dataset. (b-e) Normalized cuts segmentation of mean image of synthetic data, the colors code different surface segments. The dominance of the specular highlights deteriorates the segmentation. A good approximation is found for $k = 5$.

of subdividing the correct segments into smaller patches. The segmentation of the median image based on $k \in \{4, 5\}$ clusters yields the best result as is depicted in Fig. 5.14. The four patches are coarsely detected and subdivided when the number of patches surpasses the maximum of existing ones. While the segment borders are visible in the median image, the mean image in Fig. 5.15 provides the best composition to identify transitions. However, the dominance of the specular highlights in the upper right region leads to the identification of a single cluster that comprises these highlights. Thus $k = 5$ yields a proper segmentation where the separation of the regions is much clearer than in the other methods.

Note that the above mentioned segmentation procedures are superior to the proposed method in computational time. However, besides providing an elegant solution to many problems of data-based inference, the general benefit of the Bayesian approach is the inherent knowledge of the uncertainty of the computed results. If a traditional image segmentation is intended to be used in the task of reflectance based surface reconstruction, the reflectance parameters of the segmented patches have to be computed separately. The presented algorithm yields a combination of parameters and surface segments, including their uncertainty, as it is based on the reflectance estimation directly. Additionally, the methods used for comparison require the number of desired clusters as user input, which is not necessary in the proposed approach since it achieves unsupervised surface segmentation. Still, the considerable amount of processing time is a drawback and, additionally, a (coarse) estimate of the surface topology needs to be provided for the algorithm since its similarity index is based on the rendering of the image data which requires the surface normal directions to be available.

5.6 Summary

In this chapter, it was shown that Bayesian inference in its form of reversible jump MCMC algorithms can be applied to simultaneous surface segmentation and BRDF estimation. The benefit of using methods of statistical inference is the implicit knowledge of the certainty of the computed parameters and the foundation of the results on a large amount of data computed from the Monte Carlo approach of sampling the target distributions. The combination of Bayesian statistics and surface segmentation or BRDF estimation is new in this form and might proof useful in future applications. A severe drawback is the high processing time, and the linearity of the mechanism makes parallelization of a single Markov chain a complicated endeavor[4]. On an Intel Core i7-3770 machine, the processing of 20 000 states of the Markov chain on the synthetic dataset of 12 images of 121 × 121 pixels takes about 14 hours. After all, there exist approaches to compute RJMCMC problems in a parallel framework (e.g., Ye et al., 2009) and the proposed procedure would probably benefit significantly from such a realization.

However, since this proof-of-concept has conquered unknown territory that is additionally straying from the main topic of image-based 3D reconstruction, these experiments shall be considered as a first step. The experimental evaluation has demonstrated that the amount of available data has to be high enough to draw the correct conclusions. Especially on the synthetic dataset, the lack of data covering the specular component on the upper right region led to over-segmentation and falsely estimated parameters. However, that did not affect the quality of the reflectance maps, as the existing data was reproduced accurately. As such, the algorithm worked as intended and found a solution to reproduce the input data. It is just not the only valid solution. Still, that a data-dependent procedure requires a sufficient amount of data to converge to a meaningful result is an obvious observation.

There is a considerable amount of possible future work that can be deduced from the proposed algorithm. In any case, it is considered helpful if the BRDF – the core of the likelihood function – took the Fresnel effect into account. Using a more sophisticated reflectance model, like the Cook-Torrance BRDF (cf. Section 2.4.5) or a more general one like the Lafortune BRDF model (cf. Section 2.4.4) could increase the performance of the procedure on more complex laboratory data. Secondly, the necessity to provide the surface normals to the algorithm is as much an obstacle to its applicability as a limitation on the algorithm since the accuracy of the 3D data may as well be unknown and falsely estimated surface normals directly influence the complete procedure. Hence, it would be a self-evident move to include the estimation of the surface normals into the Markov chain. This, however, signifies a considerable increase of the overall uncertainty as it introduces two additional parameters per pixel, but the gain would be impressive. Thirdly, the proposal of new change points in the existing implementation is randomized in the form of vectors and therefore favors lines, which makes it difficult to segment circle-like surface patches since those would have to be based on lines that converge from opposing sides of the circle. This ultimately led to the generation of the third and unnecessary transition region on the laboratory dataset. Consequently, a different concept of change point proposition adapted from prior knowledge of the type of surface segments could enhance the performance of the procedure.

---

[4] Computing individual Markov chains in parallel is of course a very simple task.

CONTRIBUTION: CALIBRATING LIGHT SOURCES IN A STEREO SETUP

6

The accurate calibration of the light source position and strength is of utmost importance to compute high-fidelity normal fields from photometric data and multi-view photometric stereo algorithms. A widely applied approach to calibrate an acquisition setup is to mirror the reflection of the light source on a specular sphere (e.g., Lensch et al., 2003). Placing multiple spheres in the scene allows for the estimation of the light position by intersecting the different incident directions. However, since the incident direction is derived from a single measurement only, this is vulnerable to non-perfect detection of the specular highlight which is influenced by image saturation and broadening of the highlight alike. Another drawback of the specular sphere is that the intensity of the light source cannot be determined since the specific reflection characteristics of the sphere are unknown. Using a diffusely reflecting sphere resolves most of these issues, as has been shown by Lenoch et al. (2012), since the incident direction is estimated from $> 2 \times 10^5$ intensity measurement on the sphere surface and using multiple spheres again allows for the estimation of the light source location. Zhou and Kambhamettu (2002) already used a pair of stereo cameras and a sphere that exhibits both diffuse and specular reflection to estimate light source position and intensity, yet they still used the specular highlight to determine the light direction which is prone to errors. The contribution of this chapter consists in an algorithm that is capable of robustly estimating light source position and intensity in a stereo camera setup based on multiple images of a diffuse spherical object that is designed to have an albedo of $\rho = 0.99$.

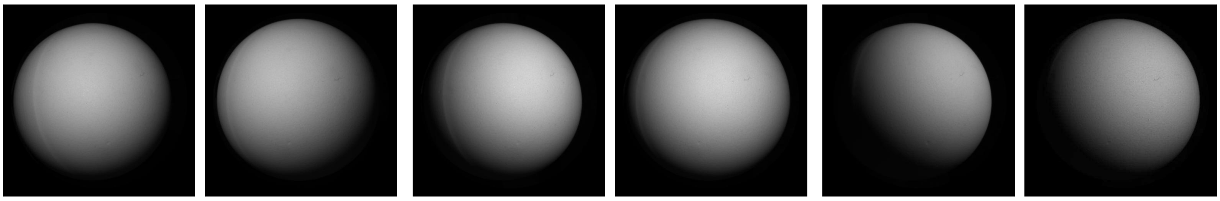
The content of the following chapter has been very briefly introduced in Lenoch et al. (2017) and builds on the prior work of Lenoch et al. (2012).

6.1 Examining the Calibration Object

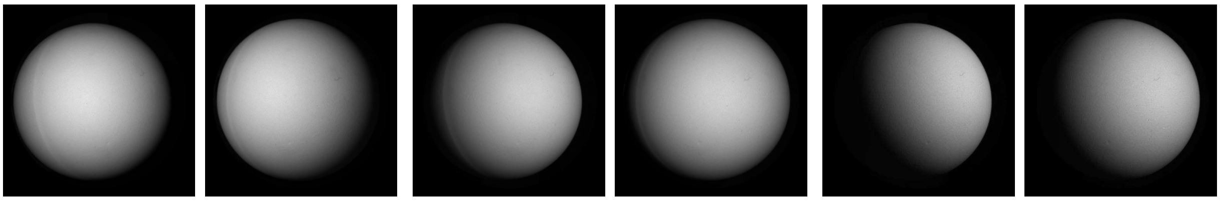
The diffusely reflecting sphere[1] that is used to calibrate the light source positions and intensities for the photometric reconstructions in this thesis has an albedo $\rho = 0.99$ according to the manufacturer and is supposed to be purely Lambertian. The complete set of input images obtained from both corresponding camera views (after rectification) is depicted in Fig. 6.1. The images have been cropped to the area containing the sphere. The images are ordered in correspondence with the 3 arcs sustaining the LEDs.

---

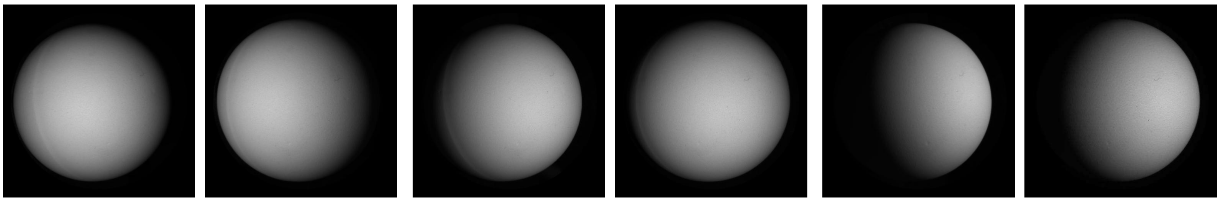
[1] manufactured by OptoPolymer, Munich.



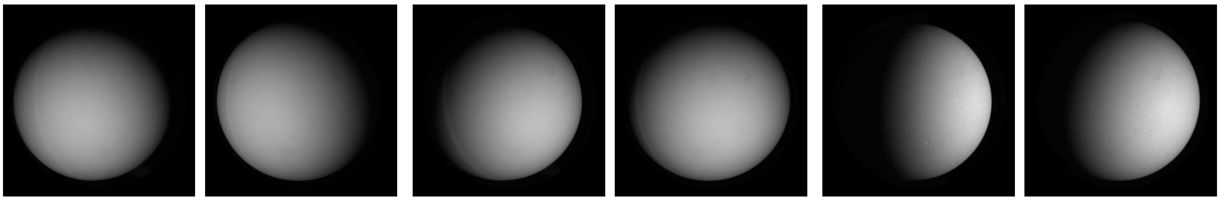
(a) LED 1.1 (b) LED 2.1 (c) LED 3.1



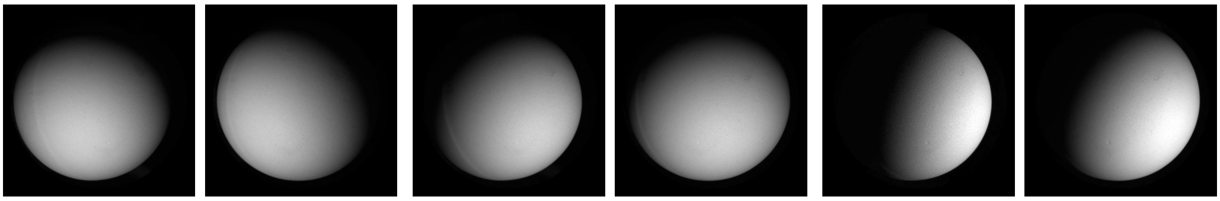
(d) LED 1.2 (e) LED 2.2 (f) LED 3.2



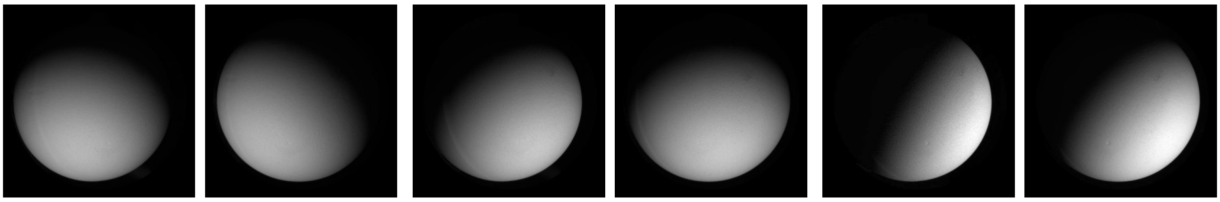
(g) LED 1.3 (h) LED 2.3 (i) LED 3.3



(j) LED 1.4 (k) LED 2.4 (l) LED 3.4



(m) LED 1.5 (n) LED 2.5 (o) LED 3.5



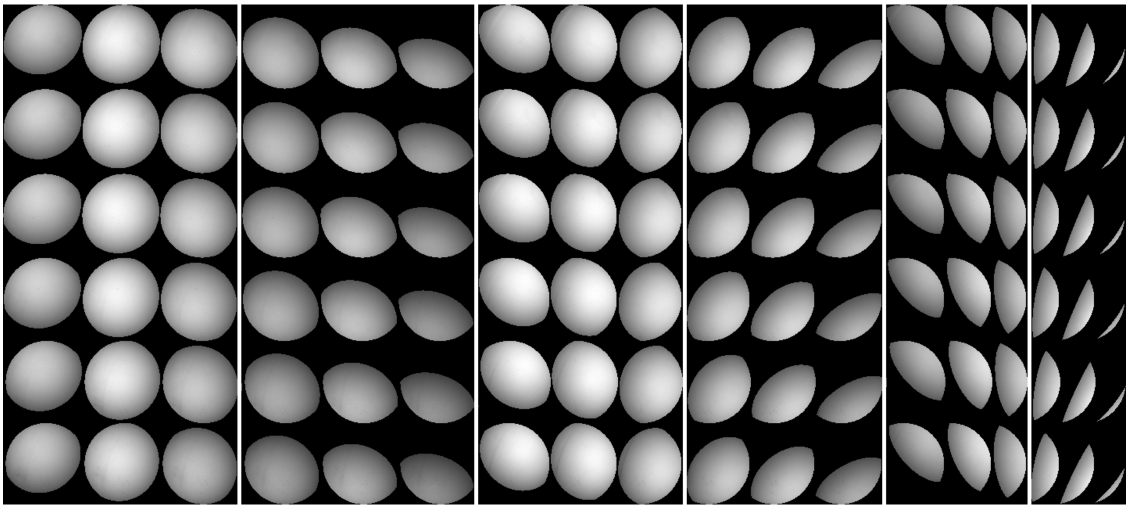
(p) LED 1.6 (q) LED 2.6 (r) LED 3.6

Figure 6.1. Calibration sphere lit from different light positions. Each sphere is displayed as seen by left and right stereo camera from the camera's perspective.

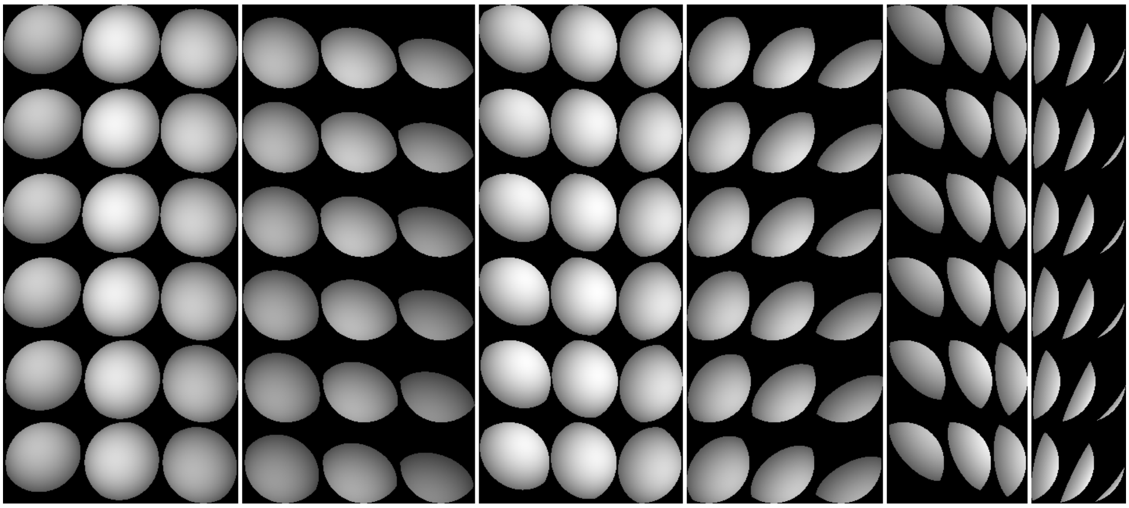
As a contradiction to the promoted Lambertian surface, a comparison between acquired images and the spheres rendered according to Lambert's law reveals systematic differences as can be seen in Fig. 6.2. The images yield a small but noticeable highlight at what appears to be the direction of mirror-like reflection. This deviation indicates that a reflectance model that accounts for specular reflection could compensate this systematic error. The physically plausible version of Phong's BRDF was selected since it can provide a subtle and wide specular lobe while being computationally very effective. The results depicted in Fig. 6.2c are visually more compelling. The parameters of the Phong BRDF have been estimated as $\boldsymbol{\theta}_{\mathrm{s}} = (0.95, 0.04, 1.42)$ (cf. Eq. (2.24)). A collection of images of the calibration sphere is depicted in Fig. 6.2 along with renderings with a pure Lambertian BRDF and Phong's BRDF. The relative pixelwise error $\mathbf{E}(u, v) = (\mathbf{I}(u, v) - \mathbf{R}(u, v))/\mathbf{I}(u, v)$ is illustrated in Fig. 6.3 and clearly shows that the diffuse assumption does not hold in its entirety.

Furthermore, many diffusely reflecting objects are known to exhibit *diffuse fall-off* (e.g., Wolff, 1996), describing the fact that the reflectance is not constant as depicted by Lambert but fading towards grazing angles (cf. Section 2.3.3). Since this would further violate the assumptions of the BRDF models and introduce additional errors in the calibration process, the reflectance properties of the diffuse sphere have been examined in greater detail. The average radiance emitted from the surface into the entire hemisphere measured for different incidence angles is shown in Fig. 6.4. The lighting geometry has been estimated based on the zSnapper depth scanner and the stereo triangulation data, respectively. The scaling differs as the data were acquired with different cameras (8- and 16-bit resolution, cf. Section 4.1). Note that the stereo rig and multiple sphere positions yield about 12 million data points to estimate the mean radiance in contrast to 5 million measurements for the depth scanner. This difference is a possible reason for the noisier appearance of the mean radiance in Fig. 6.4a. The amount of reflected light is a material property and ought to be identical for both sources of lighting geometry since the material is identical. Yet, this is the case only up to an incidence angle of $\vartheta_{\mathrm{in}} < 65°$. However, both sources of lighting geometry have their individual flaws. The fringe projection scanner exhibits high-frequency noise, and the stereo setup relies on the mathematical description of a perfect sphere to estimate the surface normals, an assumption that does not hold in practice. Additionally, the precise estimation of grazing angles is difficult, since the relative projected facet size decreases with the cosine of this angle. Moreover, small changes in the argument of the cosine lead to significant changes of the angle, making the estimation even more unstable.

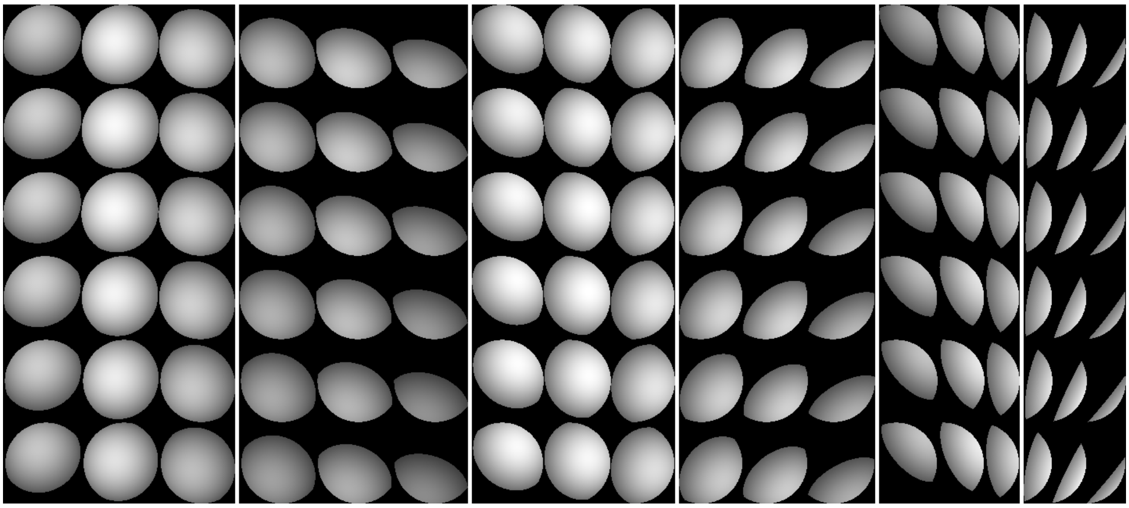
If the Lambertian statement were correct, the radiance would be a constant according to ρ/π for any uniform albedo ρ disregarding both viewing and incidence direction. However, the diffuse fall-off is visible and additionally, a second non-Lambertian effect appears to deteriorate the performance of simple BRDF models. The sharply increasing amount of reflected light towards grazing incident angles is well-known as the Fresnel effect (cf. Section 2.3.2). Since these effects are complicated to model and even more difficult to measure accurately, the range of considered incidence and emittance angles is limited to $(\vartheta_{\mathrm{in}}, \vartheta_{\mathrm{out}}) < 60°$ thereby reducing the influence of the diffuse fall-off to a level that is negligible while keeping a valid and sufficiently high amount of input data. Others researchers have stated similar results, e.g., Marschner et al. (1999b) use a range of $\vartheta_{\mathrm{in}} < 75°$ and $\vartheta_{\mathrm{out}} < 80°$ to measure and reproduce the appearance of objects and human skin.



(a) Input Image

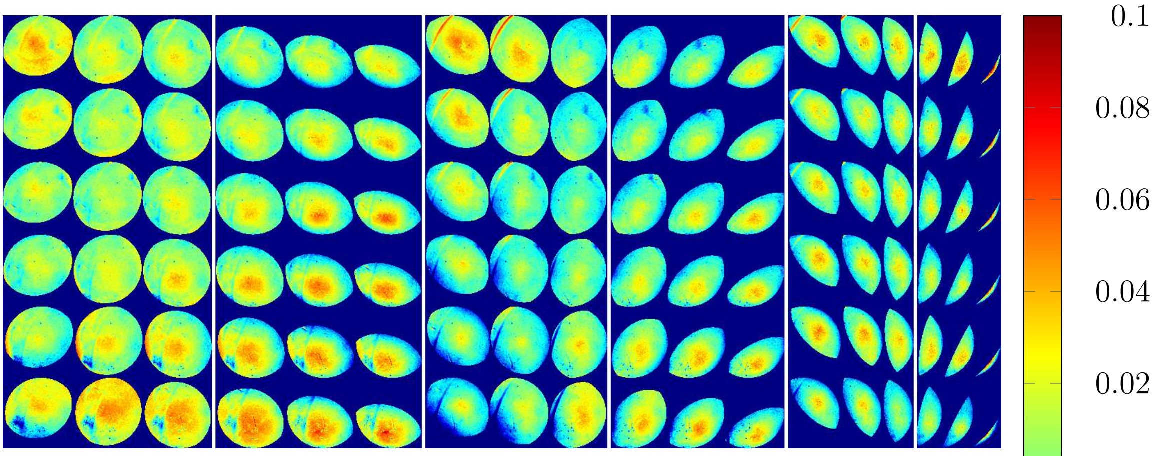


(b) Lambertian BRDF

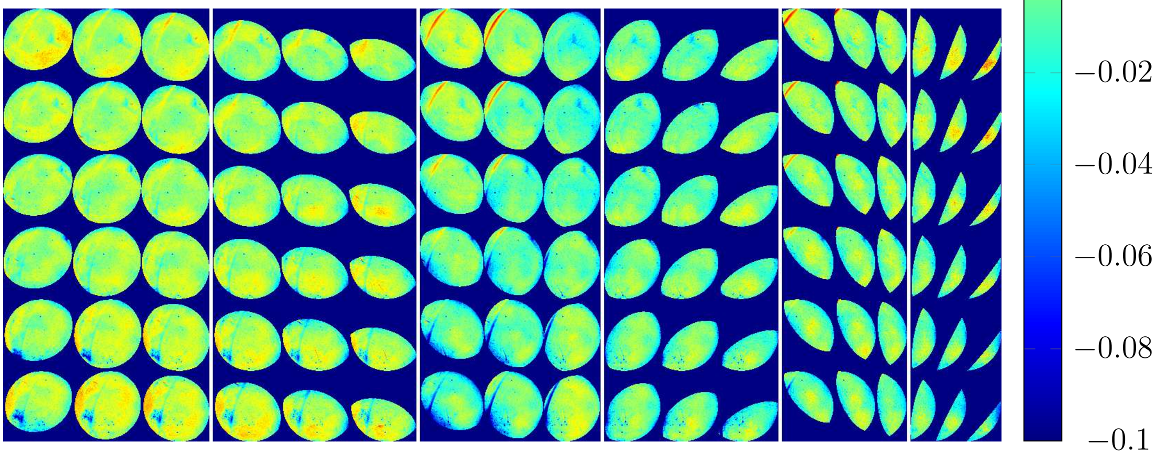


(c) Phong BRDF

Figure 6.2. Comparison of acquired images and renderings of diffuse calibration sphere for different light sources (arranged in a 3 × 6 grid) and sphere positions (block-wise) after optimization. Angles $\vartheta_{\text{in}}, \vartheta_{\text{out}} > 60^\circ$ have been excluded from the images.



(a) Lambertian BRDF



(b) Phong BRDF

Figure 6.3. Pixelwise relative error of input data and (a) Lambertian and (b) Phong BRDF model. The images comprise 18 light source positions (arranged in a 3×6 grid) and 6 sphere positions (block-wise), angles $(\vartheta_{\mathrm{in}}, \vartheta_{\mathrm{out}}) > 60°$ have been excluded from the consideration and marked invalid. The Phong BRDF clearly produces a better estimation of the reflectance data.

6.2 Linear Initialization

The calibration of the light source positions is conducted in four steps and is based on multiple images of the diffuse sphere[2], each image is illuminated by the light source to calibrate. A single image per light source would suffice to calibrate the mean incident light direction for this particular object position. However, since the position of the object varies and, more importantly, local incident light directions are preferred over one mean direction, the position of the light source in the camera coordinate system is the desired quantity. Computing the mean incident direction for one light source and multiple positions of the sphere results (theoretically) in multiple vectors that intersect at one point: the light source position. Actually, for numerical reasons and modeling errors, the lines are skewed, and a point has to be found that is the most likely position of the light source. The possibility to compute actual positions rather than incident directions based on multiple sphere positions

---

[2] Although it was shown in Section 6.1 that the sphere is not perfectly diffuse according to the Lambert law, it will still be addressed as the “diffuse sphere”.

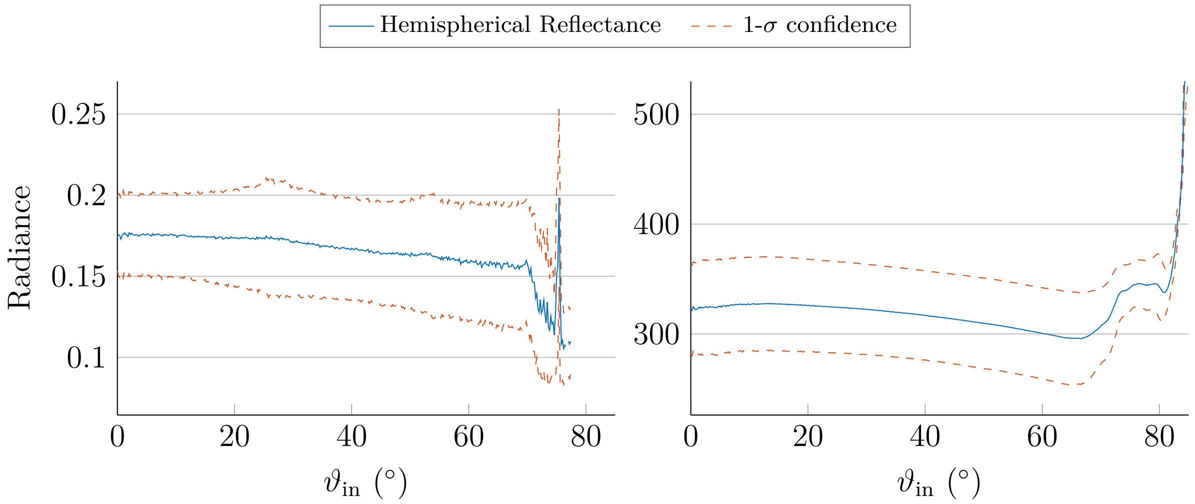


Figure 6.4. Material properties of diffuse sphere analyzed based on the mean radiance emitted at different incidence angles. The radiance would be constant and independent of ϑ_{in} if the surface was truly Lambertian. The scaling differs due to the acquisition with two different image sensors, i.e., the AVT-Pike F421 (left) and the Lumenera Lu165M (right).

has already been introduced in Lenoch et al. (2012), but calibrating the light sources for a stereo camera rig offers more flexibility concerning the optimization of the initial estimates as will be elaborated in this chapter.

6.2.1 Position of the Spheres in 3D Space

The calibration procedure depicted here is performed for the 18 LEDs shown in Fig. 4.1 and 6 positions of the diffuse sphere. Based on the insights of Section 6.1 only measurements that comply to $(\vartheta_{\mathrm{in}}, \vartheta_{\mathrm{out}}) < 60°$ are considered and the images of the valid surface points and corresponding reflectance maps are given in Fig. 6.2. Since no 3D data is available, the first step is to compute the position of the spheres in the world coordinate system. According to the projection equation (2.5) a point ${}^{W}\tilde{\mathbf{x}} = (x, y, z, 1)^{\mathrm{T}} = (x, y, z)^{\mathrm{T}} = {}^{W}\mathbf{x}$ in the world coordinate system is projected on the corresponding point ${}^{S}\tilde{\mathbf{x}} = (u, v, 1)^{\mathrm{T}} = (u, v)^{\mathrm{T}} = {}^{S}\mathbf{x}$ on the image sensor as

$$ {}^{S}\tilde{\mathbf{x}} = \mathbf{P}\,{}^{W}\tilde{\mathbf{x}} = [\mathbf{M}|\mathbf{p}_4]\,{}^{W}\tilde{\mathbf{x}}. \tag{6.1} $$

As the projection is from 3D to 2D space, every point on the line through camera center and point on the sensor is projected into the same point ${}^{S}\tilde{\mathbf{x}}$ and thus a direct inversion of Eq. (6.1) is not possible. Every point on the line

$$ g : {}^{W}\tilde{\mathbf{x}}(\nu) = \nu \mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}} - \mathbf{M}^{-1}\mathbf{p}_4 \tag{6.2} $$

is a possible candidate and the scaling factor ν needs to be estimated based on additional information to revert the projection and identify the correct points in 3D space. Sphere center and radius in sensor coordinates $(u, v)^{\mathrm{T}}$ have been estimated with thresholding, morphological operations and a linear circle fit. Since the radius of the sphere $r_{\mathrm{s}} = 15\,\mathrm{mm}$ is a manufacturing constant, it can be assumed that the intersections of a straight line through the sphere center and the opposing sides of the hull correspond to the diameter of the sphere and therefore has the value $2r_{\mathrm{s}}$. If ${}^{S}\tilde{\mathbf{x}}_0 = (u_0, v_0, 1)^{\mathrm{T}}$ is the sphere center, $\hat{r}_{\mathrm{s}}$ the radius in pixels and ${}^{S}\tilde{\mathbf{x}}_1$ and ${}^{S}\tilde{\mathbf{x}}_2$ are the intersections with the hull, they can be expressed as

$$ {}^{S}\tilde{\mathbf{x}}_1 = {}^{S}\tilde{\mathbf{x}}_0 - \hat{r}_\mathrm{s} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad {}^{S}\tilde{\mathbf{x}}_2 = {}^{S}\tilde{\mathbf{x}}_0 + \hat{r}_\mathrm{s} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{6.3} $$

At the correct projection distance ν both points must have the distance $2r_\mathrm{s}$.

$$ \begin{aligned} 4{r_\mathrm{s}}^2 &= \| {}^{W}\tilde{\mathbf{x}}_1 - {}^{W}\tilde{\mathbf{x}}_2 \|^2 = \| \nu \mathbf{M}^{-1}\, {}^{S}\tilde{\mathbf{x}}_1 - \mathbf{M}^{-1}\mathbf{p}_4 - \nu \mathbf{M}^{-1}\, {}^{S}\tilde{\mathbf{x}}_2 + \mathbf{M}^{-1}\mathbf{p}_4 \|^2 \\ &= \| \nu \mathbf{M}^{-1} ({}^{S}\tilde{\mathbf{x}}_1 - {}^{S}\tilde{\mathbf{x}}_2) \|^2 = \nu^2 \| \mathbf{M}^{-1} ({}^{S}\tilde{\mathbf{x}}_1 - {}^{S}\tilde{\mathbf{x}}_2) \|^2 \end{aligned} \tag{6.4} $$

This is a quadratic equation that yields two solutions for ν, yet the positive one is the correct solution since the object would be behind the camera otherwise.

$$ \nu = \pm \sqrt{\frac{4{r_\mathrm{s}}^2}{\| \mathbf{M}^{-1} ({}^{S}\tilde{\mathbf{x}}_1 - {}^{S}\tilde{\mathbf{x}}_2) \|^2}} = \pm \frac{2r_\mathrm{s}}{\| \mathbf{M}^{-1} ({}^{S}\tilde{\mathbf{x}}_1 - {}^{S}\tilde{\mathbf{x}}_2) \|} \tag{6.5} $$

The center of the sphere in 3D coordinates can finally either be computed from the projection of the sphere center in sensor coordinates or from the average of the projections of the intersections with the sphere hull.

$$ {}^{W}\tilde{\mathbf{x}}_0 = 0.5({}^{W}\tilde{\mathbf{x}}_1(\nu) + {}^{W}\tilde{\mathbf{x}}_2(\nu)) \overset{!}{=} \nu \mathbf{M}^{-1}\, {}^{S}\tilde{\mathbf{x}}_0 - \mathbf{M}^{-1}\mathbf{p}_4 \tag{6.6} $$

6.2.2 Incident Light Direction

Based on the initial guess of the sphere parameters, the directions of the incident light $\mathbf{l}_i$ can efficiently be computed from the measured intensities $\tilde{I}_n$ and the corresponding surface normal directions $\mathbf{n}_n$ under the assumption that the reflectance is perfectly diffuse. Now, it is already clear that this assumption is not correct, but it is sufficient to compute a coarse estimate. Additionally, the specular highlights visible in Fig. 6.2a are symmetrical and thus probably nullify the distortions of one mean incidence direction. The image intensities can be computed up to the scaling factor η based on Lambert's law

$$ \begin{pmatrix} \tilde{I}_1 \\ \vdots \\ \tilde{I}_N \end{pmatrix} = \eta \begin{pmatrix} \mathbf{n}_1^\mathrm{T} \\ \vdots \\ \mathbf{n}_N^\mathrm{T} \end{pmatrix} \mathbf{l}, \tag{6.7} $$

and inverting Eq. (6.7) based on N surface normal directions and their corresponding intensity measurements yields the scaled mean incidence direction. This is, of course, very similar to the concept of photometric stereo. Therefore the task is to derive the surface normal directions of all visible points on the sphere surface.

Reprojection of all sensor coordinates that are identified as lying on the surface of the sphere is feasible with Eq. (6.2). Again, the scaling factor ν has to be determined to identify the points in 3D space correctly. Concluding that a point ${}^{W}\tilde{\mathbf{x}}_i = (x_i, y_i, z_i, 1)^\mathrm{T}$ on the surface of the sphere – and consequently all points on the surface of the sphere – have to have the distance r_s to the center of the sphere yields

$$ (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 = {r_\mathrm{s}}^2 = \| {}^{W}\tilde{\mathbf{x}}_i - {}^{W}\tilde{\mathbf{x}}_0 \|^2. \tag{6.8} $$

Substituting ${}^{W}\tilde{\mathbf{x}}_i = \nu\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i - \mathbf{M}^{-1}\mathbf{p}_4$ according to Eq. (6.2) leads to

$$\begin{aligned} {r_\mathrm{s}}^2 &= \|\nu_i\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i - \mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0\|^2 \\ &= \nu_i^2\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2 - 2\nu_i(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0) + \|\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0\|^2, \end{aligned} \tag{6.9}$$

which again is a quadratic equation for ν_i

$$\nu_i^2 - 2\nu_i\frac{(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0)}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2} + \frac{\|\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0\|^2 - {r_\mathrm{s}}^2}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2} = 0 \tag{6.10}$$

and leads to the ambiguous solution for each ν_i

$$\begin{aligned} \nu_i =& \frac{(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0)}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2} \\ &\pm\sqrt{\left(\frac{(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0)}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2}\right)^2 - \frac{\|\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0\|^2 - {r_\mathrm{s}}^2}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2}}. \end{aligned} \tag{6.11}$$

However, the only reasonable solution is

$$\begin{aligned} \nu_i =& \frac{(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0)}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2} \\ &-\sqrt{\left(\frac{(\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i)^\mathrm{T}(\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0)}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2}\right)^2 - \frac{\|\mathbf{M}^{-1}\mathbf{p}_4 - {}^{W}\tilde{\mathbf{x}}_0\|^2 - {r_\mathrm{s}}^2}{\|\mathbf{M}^{-1}\,{}^{S}\tilde{\mathbf{x}}_i\|^2}} \end{aligned} \tag{6.12}$$

as this represents the smaller Euclidean distance to the camera and thus the visible side of the sphere. The other solution for ν_i corresponds to the points on the back of the sphere that, naturally, also have the distance r_s to the center and lie on the same straight line through the camera center.

The normal vector corresponding to each surface point ${}^{W}\tilde{\mathbf{x}}_i$ is thus simply the normalized connecting line between surface point and sphere center

$$\mathbf{n}_i = \frac{{}^{W}\tilde{\mathbf{x}}_i(\nu_i) - {}^{W}\tilde{\mathbf{x}}_0}{\|\,{}^{W}\tilde{\mathbf{x}}_i(\nu_i) - {}^{W}\tilde{\mathbf{x}}_0\|} \tag{6.13}$$

and the photometric stereo equation can be solved to find the mean incidence light direction.

6.2.3 Light Source Position

As non-distant light sources are assumed, not only the direction but the actual position of each light source is of interest. If the position of a light source is known an individual incident light direction per pixel can be computed, which enables more accurate photometric normal estimations. Theoretically, the light directions corresponding to the same light source and different sphere positions should intersect in one point. However, due to measurement noise and numerical uncertainties, one instead faces a problem of skew lines that is solved in a linear least squares fashion. Since projections from sensor to world coordinate system are no longer necessary, the following computations can be conducted in Euclidean space and furthermore all computations are executed with reference to the world coordinate system.

Hence, the notation ${}^{W}\mathbf{x}_i$ for a point in the world coordinate system will be omitted in what follows for the sake of readability, such that ${}^{W}\mathbf{x}_i = \mathbf{x}_i$.

Each light position[3] $\mathbf{s} = (s_\mathrm{x}, s_\mathrm{y}, s_\mathrm{z})^\mathrm{T}$ is subject to the following minimization problem

$$\underset{\eta_i,\mathbf{s}}{\arg\min} E = \sum_i \frac{1}{2} \|\mathbf{s} - \mathbf{x}_0^{(i)} - \eta_i \mathbf{l}_i\|^2. \tag{6.14}$$

where $\mathbf{x}_0^{(i)} = (x_0^{(i)}, y_0^{(i)}, z_0^{(i)})^\mathrm{T}$ is the center of the i-th sphere position, $\mathbf{l}_i = (l_{\mathrm{x}_i}, l_{\mathrm{y}_i}, l_{\mathrm{z}_i})^\mathrm{T}$ the corresponding incident light direction and η_i a scaling constant referring to the distance from the surface to the light source. Considering J light directions per light position at once – corresponding to J sphere positions – this can be transferred into a matrix-vector multiplication

$$\underbrace{\begin{pmatrix} -\mathbf{l}_1 & 0 & \cdots & 0 & \mathbb{I}_3 \\ 0 & -\mathbf{l}_2 & & \vdots & \vdots \\ \vdots & & \ddots & & \\ 0 & \cdots & & -\mathbf{l}_J & \mathbb{I}_3 \end{pmatrix}}_{A} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_J \\ \mathbf{s} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{x}_0^{(1)} \\ \mathbf{x}_0^{(2)} \\ \vdots \\ \mathbf{x}_0^{(J)} \end{pmatrix}}_{b} \tag{6.15}$$

The partial derivatives yield

$$\nabla E = \left(\frac{\partial E}{\partial \eta_1}, \cdots, \frac{\partial E}{\partial \eta_J}, \frac{\partial E}{\partial s_\mathrm{x}}, \frac{\partial E}{\partial s_\mathrm{y}}, \frac{\partial E}{\partial s_\mathrm{z}} \right)^\mathrm{T} \tag{6.16}$$

where each term consists of

$$\begin{aligned} \frac{\partial E}{\partial \eta_i} &= \left[(x_0^{(i)} + \eta_i l_{\mathrm{x}_i} - s_\mathrm{x}) l_{\mathrm{x}_i} + (y_0^{(i)} + \eta_i l_{\mathrm{y}_i} - s_\mathrm{y}) l_{\mathrm{y}_i} + (z_0^{(i)} + \eta_i l_{\mathrm{z}_i} - s_\mathrm{z}) l_{\mathrm{z}_i} \right] \\ \frac{\partial E}{\partial s_\mathrm{x}} &= \sum_i (s_\mathrm{x} - x_0^{(i)} - \eta_i l_{\mathrm{x}_i}) \\ \frac{\partial E}{\partial s_\mathrm{y}} &= \sum_i (s_\mathrm{y} - y_0^{(i)} - \eta_i l_{\mathrm{y}_i}) \\ \frac{\partial E}{\partial s_\mathrm{z}} &= \sum_i (s_\mathrm{z} - z_0^{(i)} - \eta_i l_{\mathrm{z}_i}). \end{aligned} \tag{6.17}$$

The solution to this system of equations is found by solving

$$\nabla E = \underbrace{\begin{pmatrix} \|\mathbf{l}_1\|^2 & \cdots & 0 & -\mathbf{l}_1^\mathrm{T} \\ \vdots & \ddots & & \vdots \\ 0 & & \|\mathbf{l}_J\|^2 & -\mathbf{l}_J^\mathrm{T} \\ & & & \\ -\mathbf{l}_1 & \cdots & -\mathbf{l}_J & J\mathbb{I}_3 \\ & & & \end{pmatrix}}_{A^\mathrm{T}A} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_J \\ s_\mathrm{x} \\ s_\mathrm{y} \\ s_\mathrm{z} \end{pmatrix} \overset{!}{=} \underbrace{\begin{pmatrix} -\mathbf{l}_1^\mathrm{T}\mathbf{x}_1 \\ \vdots \\ -\mathbf{l}_J{}^\mathrm{T}\mathbf{x}_J \\ \\ \sum_i \mathbf{x}_i \\ \\ \end{pmatrix}}_{A^\mathrm{T}\mathbf{b}}. \tag{6.18}$$

---

[3] Note the difference between the incident light direction $\mathbf{l}$ that changes per surface point and the absolute light source position $\mathbf{s}$ in the camera coordinate system.

6.3 Non-Linear Optimization and Evaluation

The coordinates obtained from the geometric considerations in the previous sections provide a guess of the 3D positions of the calibration spheres and the light source positions. Solving Eq. (6.7) with the mean incident direction for the unknown scaling factor contributes an initial value of the light source strengths. All these quantities are introduced into a global gradient descent optimization framework where the error between rendered images $\mathbf{R}$ and measured images $\mathbf{I}$ is minimized for all camera positions. Incident and viewing angles $\vartheta_{\text{in}}, \vartheta_{\text{out}} > 60°$ are excluded from all considerations, in accordance with the experimental analysis in Section 6.1 which revealed that the assumption of a diffusely reflecting object is untenable for grazing angles as Fresnel effect and diffuse fall-off dominate the reflectance behavior. As already mentioned in Section 6.1 and depicted in Fig. 6.3, the image reproduction quality increased when considering the physically plausible Phong BRDF model in addition to the Lambertian term. The parameters that have been used are $\boldsymbol{\theta}_{\text{s}} = (0.95, 0.04, 1.42)$ for diffuse weight, specular weight and specular exponent, respectively. These BRDF parameters have been estimated by including them in the optimization scheme. The final light position calibration was conducted with fixed parameters.

Table 6.2 displays the results of the non-linear optimization for different configurations comprising both the Lambert and Phong BRDF model and a constrained and an unconstrained solver. The constrained solver relied on the Matlab implementation of the interior point algorithm described by Byrd et al. (2000) and Waltz et al. (2006). The unconstrained solver is the Matlab implementation of a subspace trust-region method that is based on the interior-reflective Newton method described by Coleman and Li (1994, 1996). For the constrained parameters the initial values are assumed to have a certain degree of accuracy, and therefore the final values are confined to bounding boxes around the initial values. The sphere centers and light positions are constrained to cubes of $4\,\text{cm}^3$ and $10\,\text{cm}^3$, respectively, the light source intensities in the range of 3×10^7 are scaled to prevent numerical instabilities and limited to a range of 0.01 to 100. The final values of light source and sphere position are given as Euclidean distances and the light source intensity as a scaling factor with respect to their initial guesses.

Although intuitively the unconstrained solver should provide more freedom for exploring the parameter space it gets stuck in a local minimum indicated by the low number of iterations (6 and 10) and the almost negligible change of the parameters compared to the initial values. The constrained solver achieves to reduce the modeling error roughly by one order of magnitude when compared to the unconstrained solver, consuming considerably more computational time (595 and 452 iterations). Additionally, the modeling error is for both solvers a factor of ≈ 1.5 to 2 lower using the Phong BRDF model, underlining the importance of the non-Lambertian modeling of the calibration object. Since no ground truth is available only a qualitative assessment of the calibration results is feasible. Figure 6.5 contains the 3D view of light source positions and calibration object after the optimization. The camera is situated at the origin of the coordinate system, and the view is intentionally approximately the same as in Fig. 4.1. The sphere positions are almost equal for both approaches, and the arrangement of the light sources follows the shape of the arcs they are mounted on, yet more closely for the Phong BRDF. Especially LED 3.6 on the third arc calibrated with the Lambertian BRDF seems off and the curvature of the arcs appears more plausible for the Phong model as well. However, since the light sources are modeled as point light source which implies simplifications of the real emission characteristics according to the technical data

Table 6.2. Change of results relative to initial values after non-linear optimization of the light source positions $\mathbf{s}$, sphere positions $\mathbf{x}_0$ and light source intensity I_0. The light source intensity is given as scaling factor with regard to the initial value. The unconstrained solver fails to converge to a solution, indicated by the low number of iterations and the large error.

BRDF	**Lambert**				**Phong**			
config	unconstrained		constrained		unconstrained		constrained	
iterations	6		595		10		452	
E_{final}	253.908		30.697		170.346		15.612	
	$\mathbf{s}/\mathbf{x}_0$ (mm)	I_0 (rel)	$\mathbf{s}/\mathbf{x}_0$ (mm)	I_0 (rel)	$\mathbf{s}/\mathbf{x}_0$ (mm)	I_0 (rel)	$\mathbf{s}/\mathbf{x}_0$ (mm)	I_0 (rel)
LED 1.1	0.0014	0.9858	41.556	0.6131	0.0106	0.9928	11.575	0.9422
LED 1.2	0.0021	0.9818	40.127	0.6275	0.0164	0.9999	9.0488	0.9539
LED 1.3	0.0014	0.9899	45.182	0.6227	0.0090	1.0172	20.796	0.8331
LED 1.4	0.0011	0.9889	10.388	0.9173	0.0096	1.0172	18.356	1.2209
LED 1.5	0.0013	0.9813	15.731	1.0510	0.0096	0.9876	29.795	1.2757
LED 1.6	0.0012	0.9180	31.856	1.1641	0.0060	0.9519	45.885	1.3964
LED 2.1	0.0015	0.9186	10.011	1.0068	0.0122	0.9874	32.628	1.4337
LED 2.2	0.0002	0.9871	38.611	0.6293	0.0009	1.0138	10.061	0.9709
LED 2.3	0.0003	0.9863	47.527	0.5536	0.0015	0.9772	14.684	0.8875
LED 2.4	0.0006	0.9917	55.463	0.5183	0.0049	0.9882	21.844	0.8114
LED 2.5	0.0007	0.9854	33.872	0.6836	0.0057	0.9818	16.534	0.8788
LED 2.6	0.0004	0.9622	37.366	0.6694	0.0032	0.9792	20.690	0.8535
LED 3.1	0.0005	0.9873	33.653	1.2711	0.0041	0.9816	45.717	1.4041
LED 3.2	0.0005	0.9842	28.337	1.1693	0.0033	0.9690	37.855	1.2809
LED 3.3	0.0004	0.9733	20.780	0.8861	0.0030	0.9793	27.187	1.0703
LED 3.4	0.0004	0.9682	25.316	0.8119	0.0007	1.0006	26.503	0.9640
LED 3.5	0.0004	0.9670	38.637	0.6395	0.0017	0.9942	26.407	0.8599
LED 3.6	0.0003	0.9739	34.660	1.3856	0.0016	0.8573	24.292	0.8846
Sphere 1	0.0387		1.8510		0.2342		1.9288	
Sphere 2	0.0257		2.7704		0.2934		2.5564	
Sphere 3	0.0208		2.8048		0.1510		2.5281	
Sphere 4	0.0074		1.0256		0.0682		0.8226	
Sphere 5	0.0585		3.4928		0.3857		3.5418	
Sphere 6	0.0603		3.8042		0.4066		3.8602	

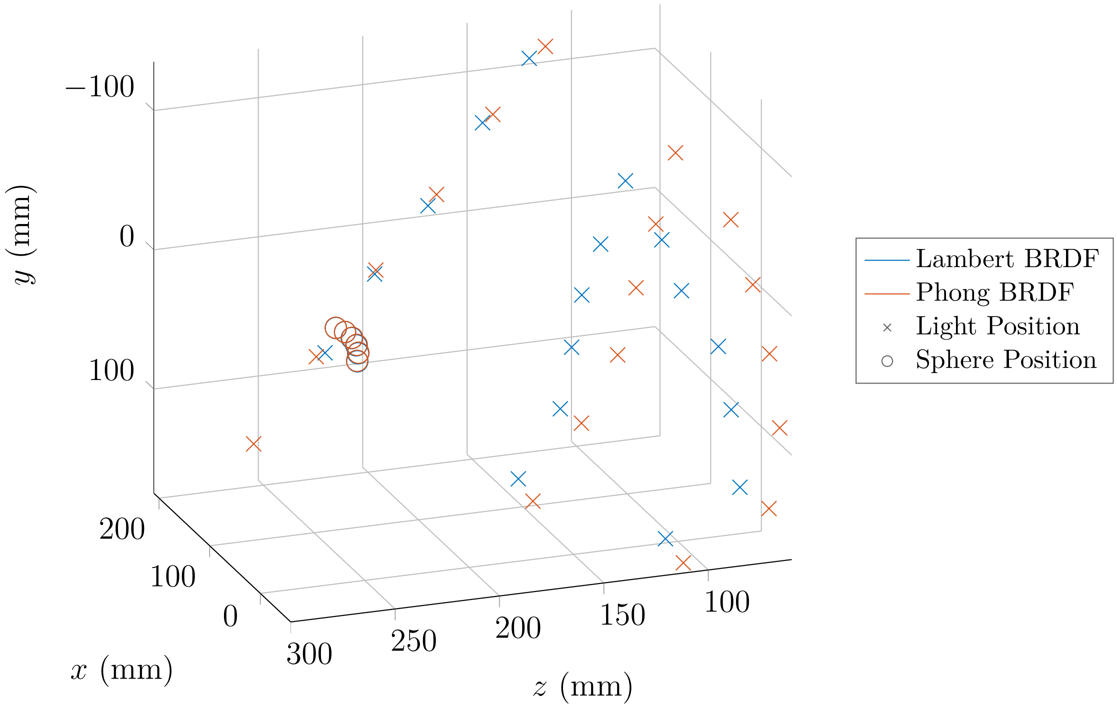


Figure 6.5. Position of light sources and spheres after calibration. Consistently with the true setup, the light sources are arranged in three arcs. The calibration based on the Phong BRDF appears more reasonable.

sheet (Seoul Semiconductor, 2013, p.26), the calibration result yields "virtual" (effective) positions of the light sources that work best to accurately reproduce the measured images under the assumption of point light sources.

Introducing additional constraints derived from prior information into the minimization procedure, such as a minimum and maximum Euclidean distance between adjacent light sources, led to no meaningful results. Furthermore, using a Monte Carlo simulation with slightly modified input values to assess the repeatability of the calibration was prohibitive due to the long execution times of about 50 hours for the constrained Phong model on full resolution images (1392 × 1040, yet the sphere covers only about a sixth part of the image). However, since the initial light source direction is based on the inversion of the Lambertian BRDF formulation, each estimate is based on $\approx 200\,000$ values contributed from a sphere surface. The entire optimization is relying on about 43 000 000 measurements[4]. A calibration of the acquisition setting has to be executed once or at maximum once before every period of acquisition and as such the execution time does not prevent its application.

6.4 Summary

This chapter provided a method to calibrated the light source positions of a measurement setup for the specific task of two-camera photometric stereo. The procedure is easily extendable to additional (calibrated) camera positions and, of course, the calibration can be used for common photometric stereo applications as well. However, since the stereo cameras are

---

[4] About 200 000 pixels on the sphere surface lit by 18 light sources at 6 positions that are captured by 2 cameras.

necessary for the calibration procedure, it is consistent to use the information obtained from the second camera and apply the algorithm illustrated in Section 7.2. An assessment of the accuracy of the procedure is difficult as there is no ground truth of the positions available. Additionally, since the model assumption of point light sources is a simplification of the real emission characteristics, the optimization renders the light sources at virtual positions. According to the similarity between photographs and rendered spheres the estimated light positions are of high accuracy and form a valid foundation for photometric gradient estimation. Moreover, it has been shown already by Lenoch et al. (2012) that the calibration based on a diffuse sphere yields increased robustness in comparison to incident directions estimated from (multiple) specular spheres.

CONTRIBUTION: IMAGE-BASED SURFACE RECONSTRUCTION

7

Photometric approaches to 3D structure retrieval have the advantage of requiring a simple setup, consisting of a camera and various light sources, that is neither complicated to set up nor prohibitively expensive and therefore open to a wide range of application scenarios and users. Structured light and laser scanning systems are the current state-of-the-art in industrial applications like quality inspection since they provide both, high accuracy and fast computation. However, active triangulation techniques are costly and suffer from sparse depth data on surfaces that exhibit strong discontinuities since the projected pattern is partially blocked. Moreover, narrow-band sources of emission are vulnerable to speckle noise on surfaces whose roughness is similar to their wavelength. This sensitivity deteriorates the quality of the 3D reconstructions and results in an artificial roughness of the estimated surfaces. In this chapter, two different procedures will be detailed that achieve accurate image-based surface estimations of objects that feature complex reflectance properties and challenging shapes.

The content of this chapter has been adapted and/or adopted from Lenoch and Wöhler (2016) and Lenoch et al. (2017).

7.1 Reflectance-based 3D Shape Refinement

While reflectance-based 3D reconstruction methods have a number of advantages that are stated in the introduction, they have the inherent disadvantage to recover gradient information, i.e., a relative description of the surface, only. The estimated gradient field can be integrated to create a coherent surface, but due to the unknown integration coefficient, that is the scaling, the real-world distance remains unknown. Furthermore, small errors in the estimated surface normals add up during the integration causing the final surface to deviate at a larger scale from the true solution. Still, the high-frequency component tends to be very accurate. Therefore it is necessary to constrain the integration of the gradient (e.g., Ikeuchi and Horn, 1981; Horn and Brooks, 1986). A coarse estimate of the 3D structure provides valuable information to resolve the scaling uncertainty and prevent large-scale deviations as shown by Grumpe et al. (2011). This section introduces an algorithm that is capable of determining the spatially varying BRDF (SVBRDF) of an arbitrary object and refining an initially estimated 3D surface iteratively using the gradient fields computed from photometric stereo. Here, the initial surface is computed from the diffuse component of the

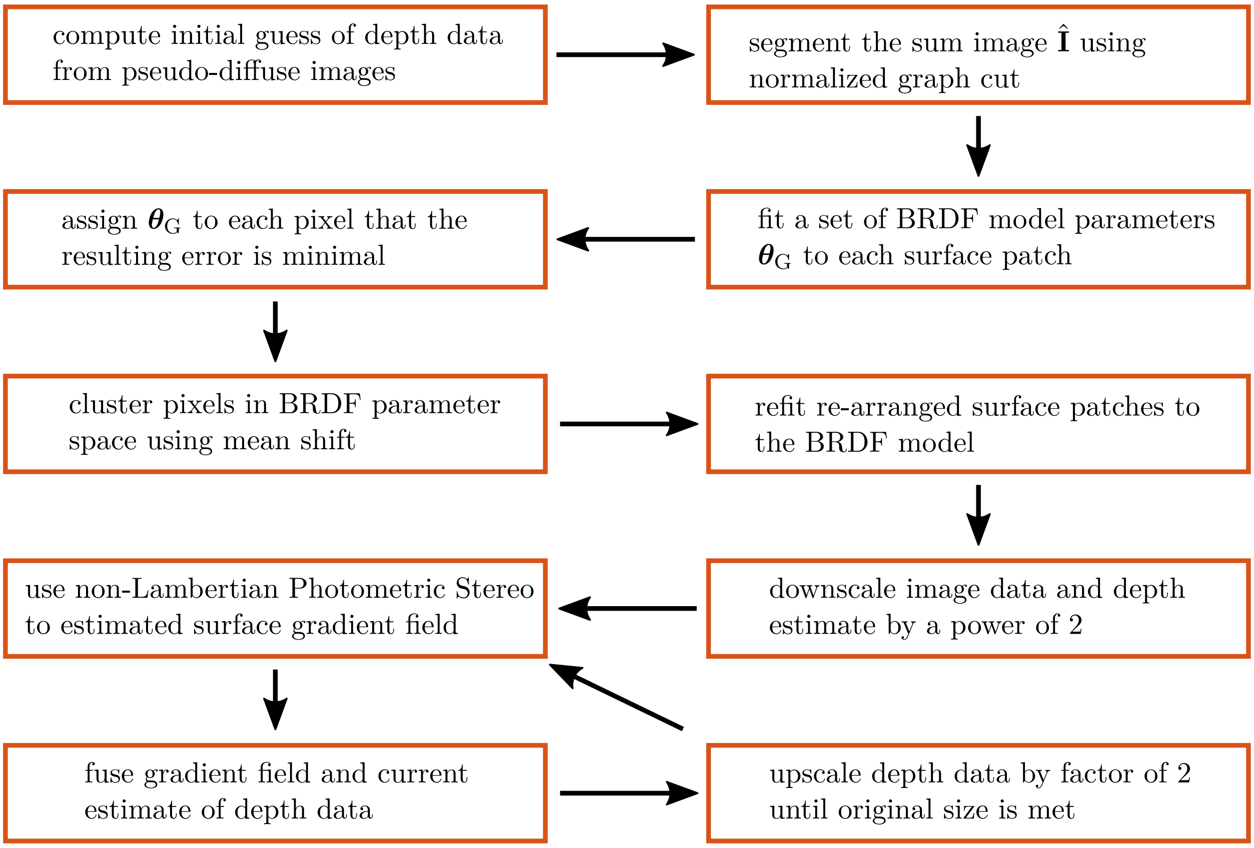


Figure 7.1. Flow chart of reflectance based refinement algorithm.

reflectance data, but the source of the 3D data has no importance and could, for example, be classical stereo as in Section 7.2 or shape-from-motion (e.g., Zhou et al., 2013).

7.1.1 Algorithm Overview

The proposed procedure to compute per-pixel BRDF model parameters and refine an initially estimated 3D structure based on the surface gradient consists of several steps that will be reviewed in the course of this section in detail. The most important ones are given in Fig. 7.1.

In contrast to the Bayesian segmentation presented in Chapter 5, this procedure computes the image segmentation based on the input intensity images. This is computationally more efficient and does not require accurate 3D information to render the reflectance maps. The algorithm can be divided into two large sections. The pre-processing of the data is illustrated in the following.

Initial pseudo-diffuse surface

Accurate knowledge of the surface is in general crucial to fit a spatially varying BRDF to a surface. It is assumed that no depth data are available to the proposed procedure and therefore a good initial guess is necessary. Inspired by Han and Shen (2015), virtual images that contain only the diffuse reflection component of the test objects are employed to solve the classical photometric stereo (cf. Section 3.2). A fundamental assumption is that the perceived intensity changes little from one light source to the next if the diffuse component of the surface contributes the majority of the reflected light. Whereas the changes of the specular component can be drastic when the incident direction changes by some degrees (cf. Fig. 2.3). Thus, the gradient of the intensity data is supposed to shed light on the principal components that contribute to the radiance. If the gradient is small, the intensity originates probably from the diffuse component. Unlike Han and Shen (2015) the laboratory data provide relatively sparse data covering only about the eighth part of the upper hemisphere with the 18 light positions.

Those pixels that are determined as non-specular according to the low intensity gradient are fused into individual pseudo-diffuse images. To these images the original Photometric Stereo can be applied (since the surface is Lambertian) to compute a least-squares estimate of the surface gradient field. The resulting gradient field can then be integrated, and the resulting surface is displaced by an approximate distance between camera and object to provide an initial guess of the scene. The number of pseudo-diffuse images that is used here is a parameter that can be set by the user. To support a robust estimate of the surface gradient while still excluding the specular pixels, a set of $N_{\text{diff}} = 8$ images is supposed to be suitable. Of course, it is feasible to replace this by depth information computed from other sources, such as stereo vision (cf. Section 7.2) or a low-resolution (and thus low-cost) 3D scanning system. The following algorithm is designed as well to refine an initially supplied surface regarding noise reduction while conserving high-frequency details.

Initial Segmentation and Patchwise BRDF Fitting

The sum image $\hat{\mathbf{I}} = \sum_n \mathbf{I}(u, v, n)$ is computed as the pixelwise sum of N input images, where invalid or missing intensities are treated as zeros. The number of measurements is not considered since it is explicitly not intended to compute average values. $\hat{\mathbf{I}}_{\text{seg}}$ is segmented using the Normalized Cuts algorithm (Shi and Malik, 2000) (the same implementation that was used in the evaluation in Section 5.5). The number of image segments G_{s} has to be chosen by the user and is set here to $G_{\text{s}} = 10$ as this should provide enough flexibility while not resulting in strong oversegmentation.

Based on the initial segmentation, the parameters of the physically plausible Blinn BRDF model (Eq. (2.26)) are computed for each of the supplied surface patches to minimize the intensity error between measured image and reflectance map $\mathbf{R}$. The parameter search is based on the non-linear optimization problem

$$\underbrace{\arg\min}_{\boldsymbol{\theta}_{\text{G}}} \sum^{U_{\text{g}}, V_{\text{g}}, N} \Big[\mathbf{I}(u, v, n) - \mathbf{R}(u, v, k, \boldsymbol{\theta}_{\text{G}}, \mathbf{n})\Big]^2 \tag{7.1}$$

with u, v denoting pixel coordinates and $U_{\text{g}}, V_{\text{g}}$ the set of image pixels that belong to surface patch s. The solution is found using the Levenberg-Marquardt algorithm (e.g., Moré, 1978). It is supposed that a total image size of $U = V = 600$ and $N = 18$ images supplies sufficient information to fit the model parameters accurately to each patch.

Re-assignment of Surface Pixels

Since the initial segmentation was based on monochrome image data only, it cannot be expected that the segmented patches necessarily represent areas of similar reflectance properties. To ensure that, the error between image and reflectance is computed for each pixel and every set of parameters $\boldsymbol{\theta}_{\text{G}}$ and the corresponding pixels are assigned to the patch for which they result in the lowest intensity error.

Mean Shift Clustering

The pixelwise BRDF parameters are normalized according to the z-score, such that they exhibit a standard deviation $\sigma = 1$ and a mean $\mu = 0$ (e.g., Marsland, 2009). Consecutively, the normalized parameters are clustered using the mean shift algorithm (Comaniciu and Meer, 2002) with kernel width $\sigma_{\mathrm{MS}} = 0.5$ and minimal distance between two clusters $d_{\min} = 10^{-6}$. Patches of very similar reflectance properties will thus be merged.

The absolute intensity error $\sum_N |\mathbf{I}(u,v,n) - \mathbf{R}(u,v,n)|$ is computed and checked for every pixel that exceeds a threshold if there is another set of parameters that would yield a lower error. By doing this, it is again ensured that the clusters are similar in their reflectance properties.

Refit of Surface Patches

As the assignment of the different patches might have changed in the previous step, the BRDF parameters of the patches are re-computed according to Eq. (7.1). After this preprocessing of the data, the initial 3D information is refined in a pyramid scheme, i.e., the given data are sub-sampled, and its resolution is iteratively increased by a factor of 2 until the original size is met. The subsampling factors are $d \in \{8,4,2,1\}$.

Photometric Stereo

Inspired by Woodham's photometric stereo (Woodham, 1980) the surface normals $\mathbf{n}_{\mathrm{PS}}(u,v)$ of each image pixel are computed based on the given BRDF model and the input images. The idea is similar to Eq. (7.1) and reverses the search: The BRDF parameters are known, and the surface normal is found that yields the lowest intensity error.

$$\underbrace{\arg\min}_{\mathbf{n}_{\mathrm{PS}}(u,v)} \sum^{U_{\mathrm{g}},V_{\mathrm{g}},N} \Big[\mathbf{I}(u,v,n) - \mathbf{R}(u,v,n,\boldsymbol{\theta}_{\mathrm{G}},\mathbf{n}_{\mathrm{PS}})\Big]^2 \tag{7.2}$$

Fusion of Gradient and Depth

The normals computed from the photometric stereo $\mathbf{n}_{\mathrm{PS}}$ are combined with those derived from the current depth estimation $\mathbf{n}_{\mathrm{depth}}$ according to

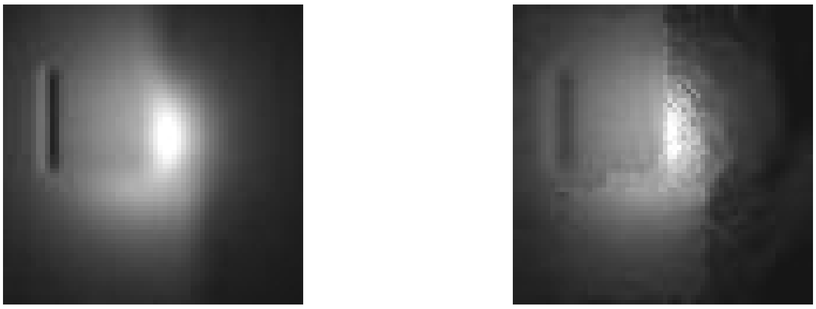
$$\mathbf{n}(u,v) = \mathbf{W}(u,v) \cdot \mathbf{n}_{\mathrm{depth}}(u,v) + (1 - \mathbf{W}(u,v)) \cdot \mathbf{n}_{\mathrm{PS}}(u,v)\,. \tag{7.3}$$

Based on the pixelwise root mean squared error $\mathbf{E}_{\mathrm{RMSE}}$ of measurements and rendered intensities after the final surface normal estimation, the weight matrix is $\mathbf{W} = \mathbf{E}_{\mathrm{RMSE}} / \max(\mathbf{E}_{\mathrm{RMSE}})$. The problem of integrating surface normals to a coherent surface $z(u,v)$ bound to approximate depth information $\hat{z}(u,v)$ can be regarded as an extension to the Shape from Shading formulation given by Horn (1989) and has been solved by Grumpe and Wöhler (2014) using calculus of variations (cf. Section 3.1.1). Accordingly, the integrability error term

$$E_{\mathrm{int}} = \sum^{U,V} \frac{1}{2}(z_x(u,v) - p(u,v))^2 + \frac{1}{2}(z_y(u,v) - q(u,v))^2 \tag{7.4}$$

with the partial derivatives $z_x = \frac{\partial x}{\partial z}$ and $z_y = \frac{\partial z}{\partial y}$ and the estimated surface gradients p and q is extended by an absolute depth error term

$$E_{\mathrm{absDepth}} = \sum^{U,V} \frac{1}{2}(g_{\sigma_{\mathrm{absDepth}}} * z(u,v) - g_{\sigma_{\mathrm{absDepth}}} * \hat{z}(u,v))\,. \tag{7.5}$$



(a) (b)

Figure 7.2. (a) Input image of synthetic dataset and (b) reflectance map rendered on refined depth data and optimized parameters. Same gray value scaling in both images.

Here $g_{\sigma_{\text{absDepth}}}$ denotes a σ_{absDepth} pixels wide Gaussian low-pass filter that is applied using a convolution ($*$). The existing depth data in the current iteration of the pyramid scheme are regarded as valid on a large scale but possibly erroneous on a small scale. Hence, a low-pass filtered version of the current depth information should be equal to the known reference. The accuracy on a small scale is provided by the surface gradient since the image data yield small details that are only partially (or not at all) visible in the depth data.

7.1.2 Evaluation on Synthetic Data

First, the proposed algorithm is tested on a small synthetic dataset that consists of the rendering of a sphere with a scratch to simulate a production error under spatially varying reflectance properties. The dataset comprises 18 images the are rendered in accordance with the laboratory setup, that is the object is placed at the approximate distance of 250 mm and the calibrated light source positions are used. The varying parameters consist of 3 initial sets of parameters assigned to surface regions with a soft transition between the regions. Noise is added to a low-pass filtered version of the depth data that are used in the rendering process to simulate a non-ideal depth estimation. One of the input images alongside the corresponding reflectance map based on the estimated varying parameters is given in Fig. 7.2. The reflectance map is rendered based on the refined depth data.

The estimated parameters of the Blinn BRDF model are shown in Fig. 7.3. The upper row displays the ground truth used to render the synthetic data, the lower row is the result of the algorithm, with the same colormap in corresponding images. Despite the noisy input data of both, image and depth data, the estimated BRDF parameters are consistent with the true values in the central image regions. The outer parts of the image reveal a deviation from the ground truth possibly caused by the increasing slope of the sphere surface. Small changes in the estimated surface normals cause high variation of the amount of reflected light towards grazing angles according to the variations of the cosine.

Fig. 7.4 displays a shading of the depth data, where a uniform BRDF that does not correspond to the true reflectance properties has been used for rendering. A simple low-pass filtering (Fig. 7.4d) reduces the noise at the cost of high-frequency image detail. Here, the scratch in the sphere is hardly visible. The proposed method Fig. 7.4c results in a good reconstruction of the ground truth data without diminishing the appearance of the scratch on the surface. However, a distortion of the surface in the upper right corner is visible and presumably caused by a mixture of grazing incident light that results in low intensities and

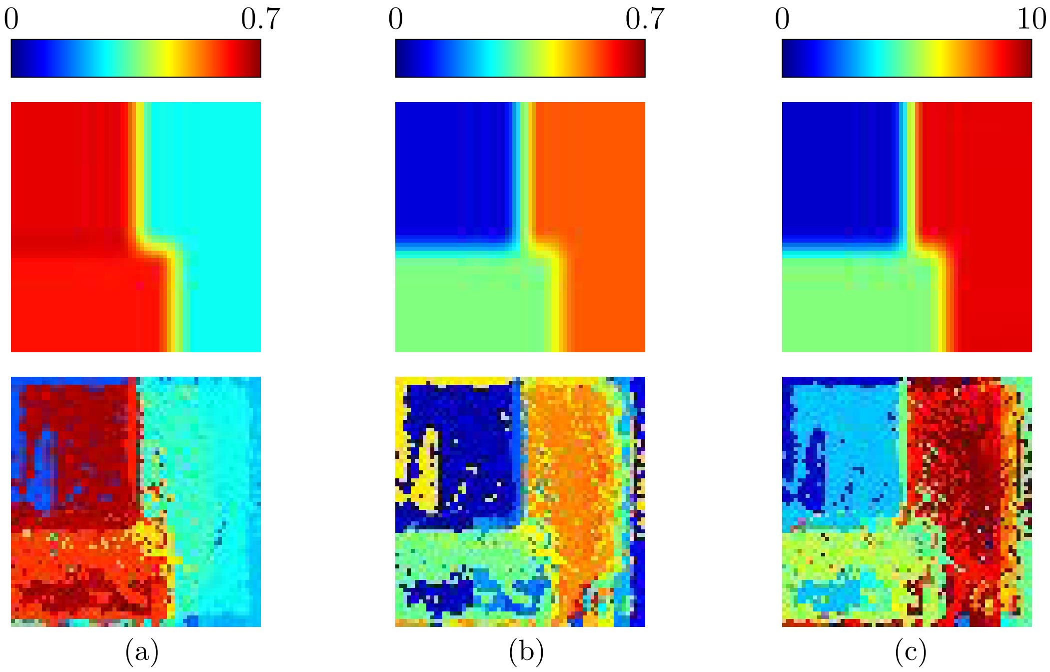


Figure 7.3. BRDF parameter maps of the synthetic dataset. Ground truth (upper row) and parameters after reassignment of every pixel to the cluster with the lowest error (lower row). (a) Diffuse weight, (b) specular weight, (c) specular lobe exponent.

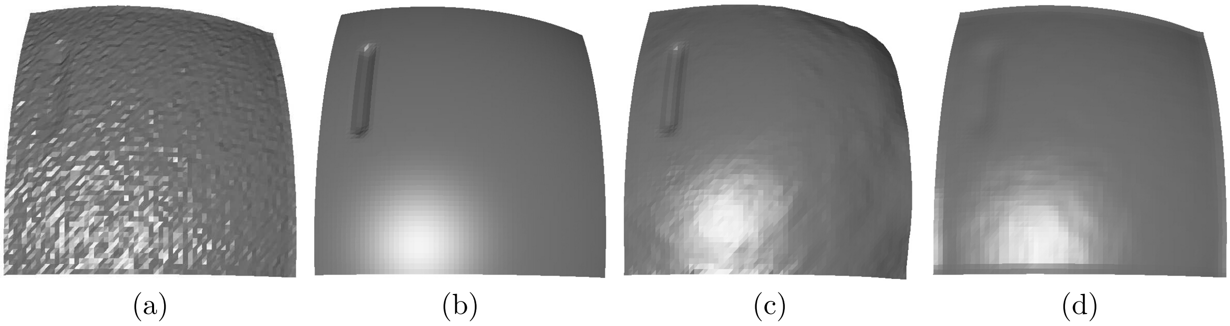


Figure 7.4. Uniform shading of depth data. (a) Noisy input image, (b) noise-free ground truth, (c) reconstructed depth based on surface normals, and (d) low-pass filtered input data. The low-passing cancels the noise but as well softens out the scratch in the surface. The upper right corner of the reconstruction is slightly distorted.

incorrectly estimated parameters. Anyhow, the experiments on synthetic data are regarded as a proof of concept, and the evaluation is continued on laboratory data acquired in a real-world environment.

7.1.3 Evaluation on Laboratory Data

The laboratory data set that was used to test the properties of the algorithm outlined in the previous section consists of images of a Styrofoam egg and a shaped plaster mass, similar to the object used in Section 5.4.2, displayed in Fig. 7.5. These images are for comprehension only and have been acquired with a camera different from those of the acquisition setting. Due to the ambient lighting, the spatially varying reflectance is more apparent in Fig. 7.7.

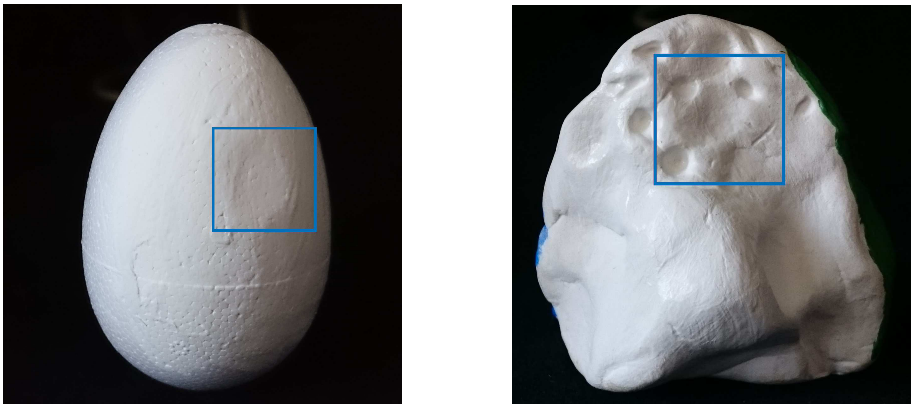


Figure 7.5. RGB images of the objects used to test the proposed algorithm, relevant area of 600×600 pixels is inside the blue box. Note: The images are displayed for comprehension only and have been captured with a consumer-grade digital camera different from the acquisition setup. Height of both objects is $\approx 7\,\mathrm{cm}$.

Both objects show inherent diffuse reflection and have been partially painted with white acrylic paint to smoothen the surface structure and satin varnish to increase the glossiness. As a result, data of objects of a similar color can be acquired that exhibit spatially varying reflectance properties at the same time. It was chosen not to change the albedo of the surface significantly since i.e., orange or blue paint would result in low image gray values that can be distinguished easily from the rest of the surface. The white ground coat with an additional specular component poses a more challenging problem.

The results of the pseudo-diffuse estimate of the surface structure are depicted in Fig. 7.6. The pseudo-diffuse images are exemplarily represented by the example in Fig. 7.6a. However, they are not very meaningful to a human viewer since the pixels are collected from various images according to the lowest intensity gradient which is not necessarily consistent for neighboring pixels. Still, as long as the light directions are ordered accordingly, the results are reasonable. Both surfaces shown in Fig. 7.6 reveal irregularities and artifacts that are not visible in the images depicted in Fig. 7.7. However, this is tolerable to a certain extent, and they provide a good initial estimate of the real depth data that will be refined iteratively in the following steps of the procedure. The final results are shown in Fig. 7.9.

Figure 7.7 depicts four of the eighteen available input images for both data sets and their corresponding reflectance maps rendered on the depth data computed by the proposed method. The full sets can be found in Figs. A.4 and A.5 in Appendix A.3. The arbitrary change in the reflection properties from diffuse to specular is obvious for both data sets. The images of the egg dataset show a hard transition from diffuse to highly specular reflection. The reproduced images contain a visibly increased specular component on the right side that is yet less strong than in the input images. The albedo is estimated well, and the diffuse component is very similar which is evident as the overall brightness in the corresponding images is very similar. The reflectance maps of the plaster dataset show an increased specular component for the upper left region of the object which coincides with the image data, yet again, the strength of the specular lobe is too low, and as such, the shininess of the object is not reproduced accurately.

The BRDF parameter maps are depicted in Fig. 7.8. The *full width at half maximum* (FWHM) of the cosine lobe $\chi = 2\arccos(0.5^{1/\gamma})$ is chosen instead of the lobe exponent γ since it is easier to derive a physical meaning from it. A narrow cosine lobe, i.e., a small

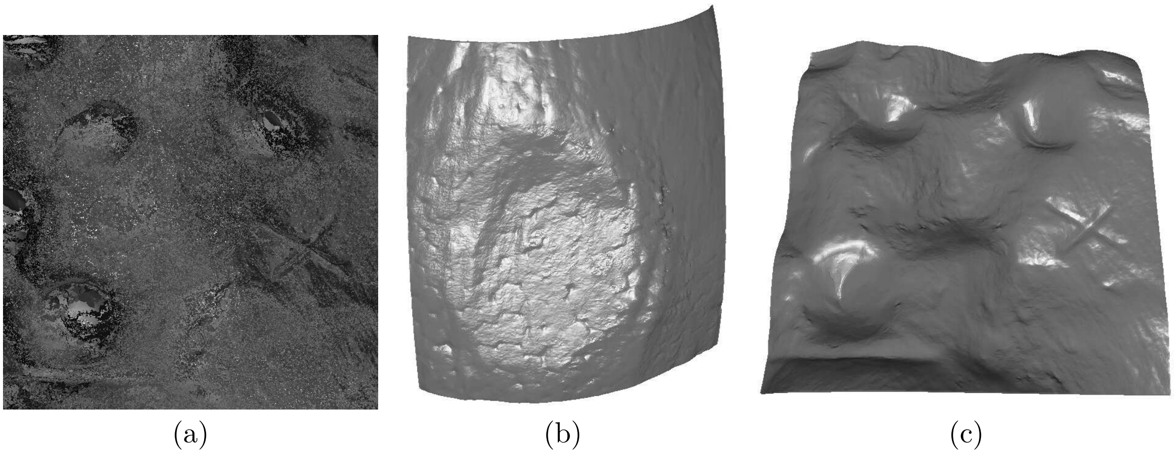
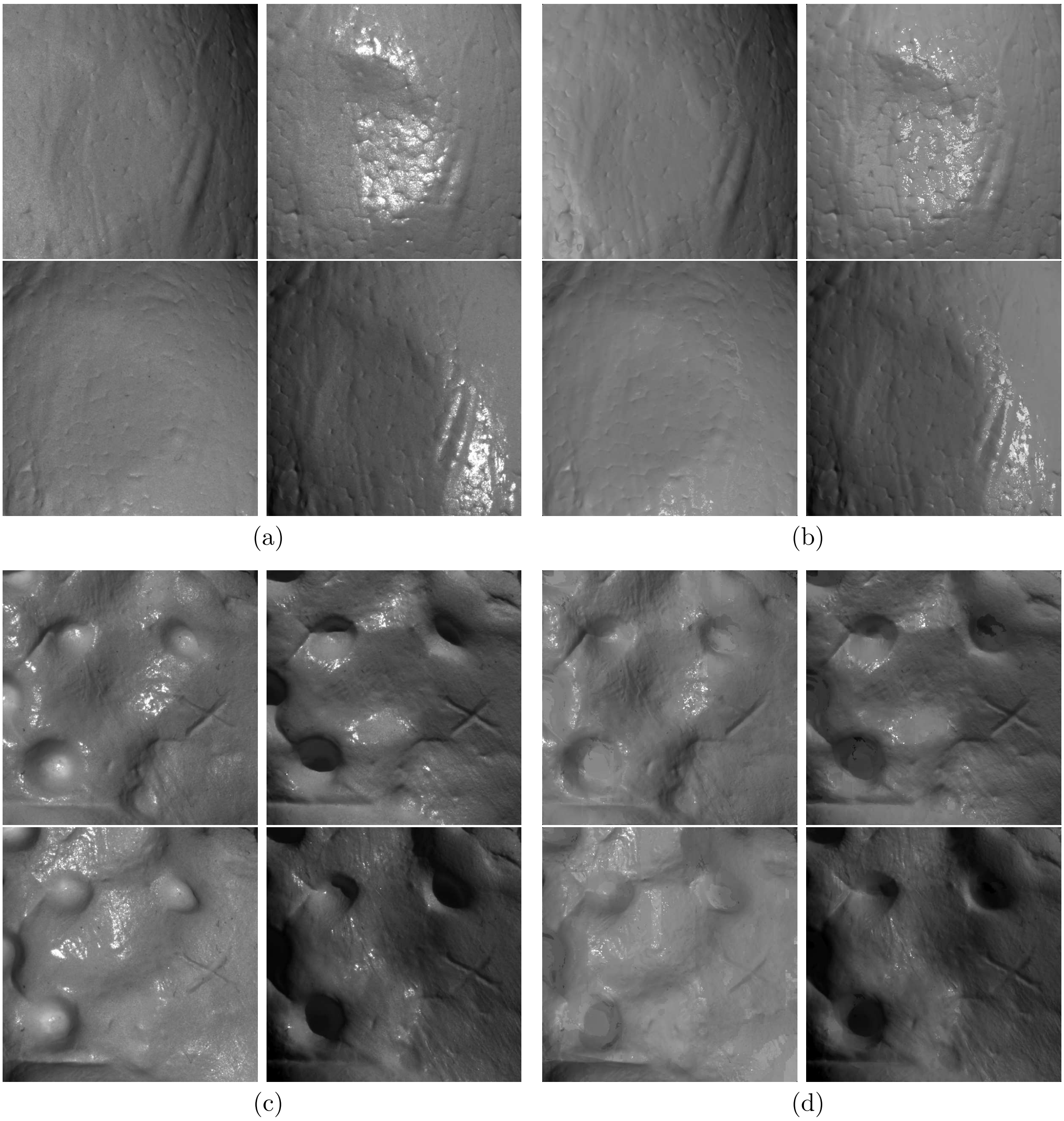

(a) (b) (c)

Figure 7.6. (a) Exemplary pseudo diffuse image. (b,c) Uniformly shaded depth data based on pseudo diffuse photometric stereo estimate of Styrofoam egg and plaster mass, respectively.

FWHM, indicates a higher amount of specular reflection and equals a high exponent. The parameter maps of the egg dataset exhibit little change along the surface with a high diffuse and very low specular weight. The maxima of the specular weight cover the shiny right side of the image. The plaster dataset features a higher variance of the diffuse weight and the specular exponent. The diffuse component exhibits its maximum in the central region and decreases to the left and right edges, and the specular lobe widens, which corresponds to a less glossy surface. The changes in the diffuse component do not agree with a uniform surface, however, since the scaling of the parameters is rather narrow, the changes have only a subtle influence on the resulting reflectance maps. There is no apparent reason for the rectangular structures that are visible in all parameter maps of the plaster dataset. These might be artifacts resulting from the parameter fitting process that weakly constrains the parameters by taking into account the surrounding surface region. Generally speaking, the obtained BRDF parameter maps might be used to derive basic material properties from unknown objects. However, in this case, the conclusions that may be drawn from the parameters are vague. All these findings support the impression of Fig. 7.7.

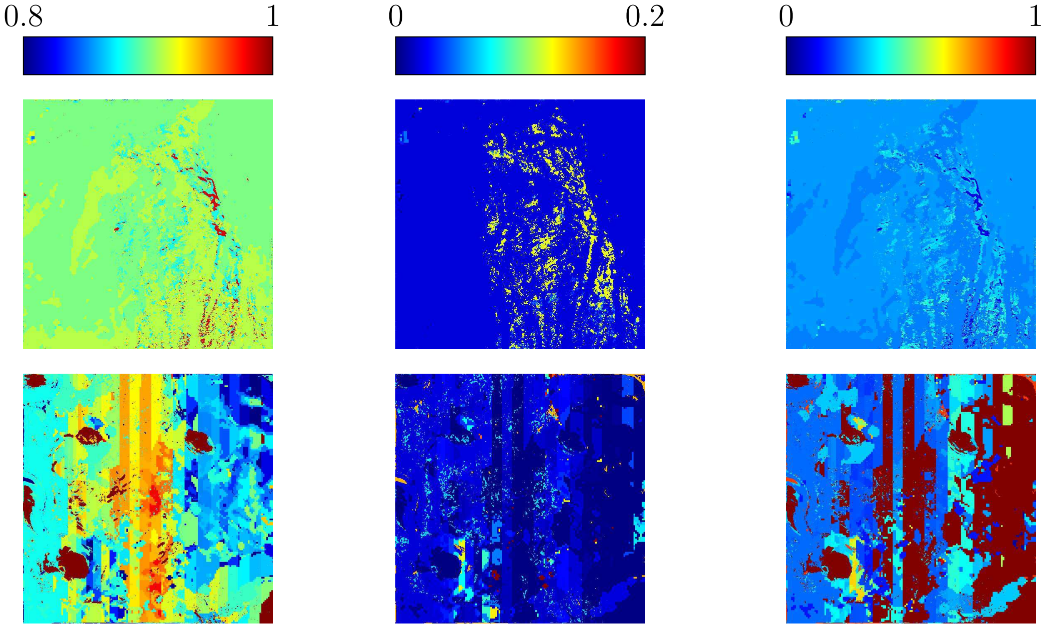
The more important part of the proposed procedure focuses on the reconstruction of the unknown surfaces. The gradient fields of the surfaces are computed from the estimated parameter maps and fused with an initial estimate of the depth information. The full view of both surfaces is depicted in Fig. 7.9. Note here that the high-frequency information that makes the scanner data appear *sharper* is noise that displays features that do not exist. A close up view of the 'X' scratch on the plaster surface is depicted in Fig. 7.10. The image in Fig. 7.10a was captured with a different camera to achieve a high-resolution zoom. The image suggests that the smoother structure of the reconstructed depth in Fig. 7.10c is reasonable. The sharp edges of the 3D scanner data are not visible in the image, and the refined surface contains significantly less noise while preserving the existing structure of the object, since the gradient information contains all visible details.

A more quantitative assessment of the refined surfaces can be conducted based on height profiles, i.e., horizontal cuts through the surfaces, two of which are displayed in Fig. 7.11. Note that the noise in the ground truth data is less apparent due to the scaling. Furthermore, the mean of the depth data has been subtracted from the surface to remove the offset between



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

Figure 7.7. Test objects with spatially varying reflectance properties lit by different light sources (4 out of 18 images). White diffuse surfaces partially painted with acrylic white paint and/or satin varnish. (a,c) Input images and (b,d) corresponding reflectance maps based on varying BRDF parameters and refined depth data. (a,b) Styrofoam egg and (c,d) plaster mass. Image resolution is 600 × 600 pixels, same gray value scaling in corresponding images.



(a) (b) (c)

Figure 7.8. Parameter maps of the Styrofoam egg (upper) and plaster object (lower) after reassignment of every pixel to the cluster with the lowest error. (a) Diffuse weight, (b) specular weight and (c) FWHM of cosine lobe of the Blinn BRDF.

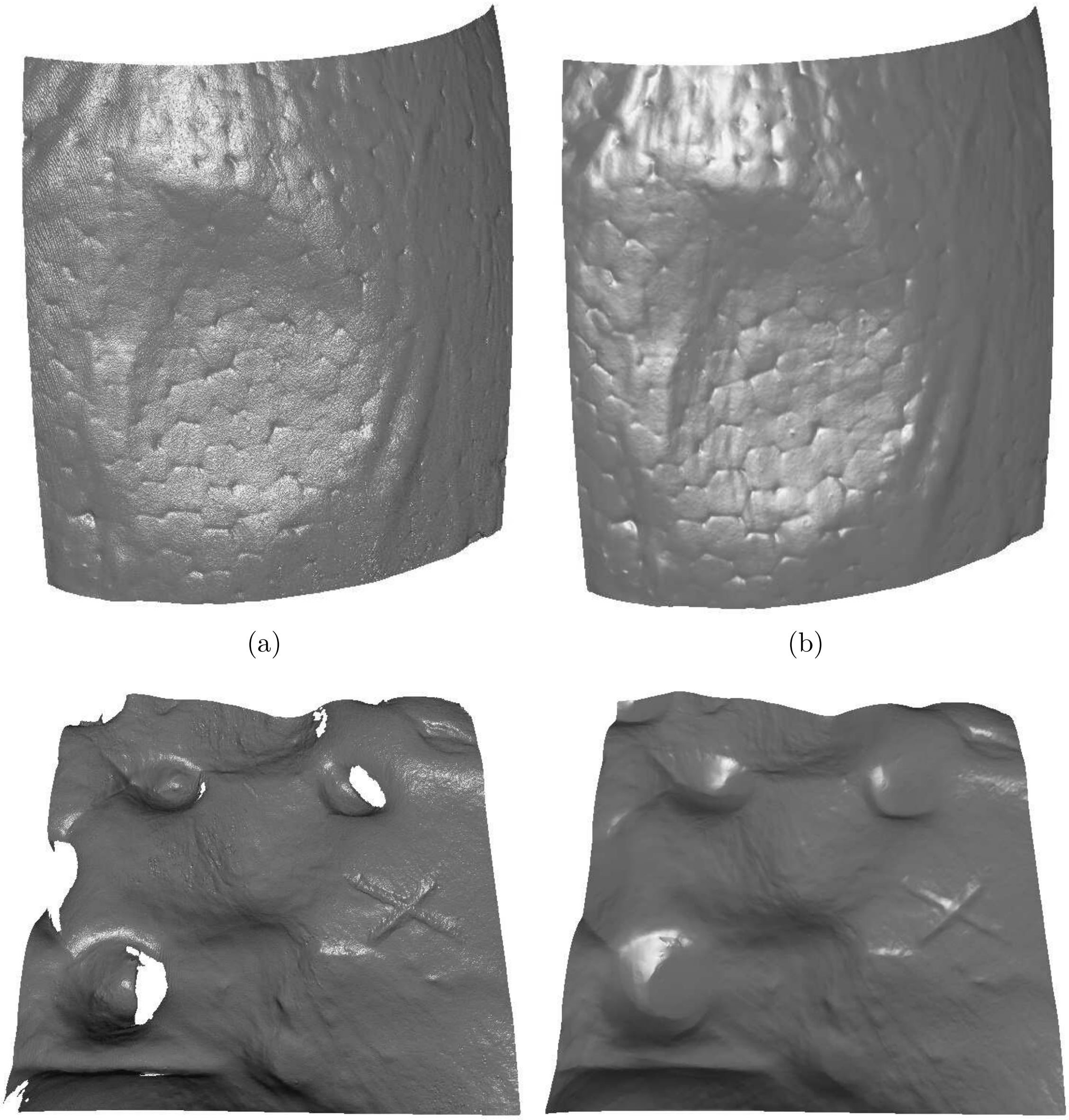
Table 7.2. RMSE in mm between zero-mean scanner data and pseudo-diffuse depth and results of the proposed algorithm, respectively.

approach	egg	plaster
initial estimate	0.469	0.582
proposed method	0.179	0.285

real and estimated distance of camera and object. The visible deviation of refined data and input in Fig. 7.11 (lower profile) is caused by interreflections that occur in concave areas and cannot be captured by BRDF models as those account for direct illumination only. Still, the results are convincing. The RMSE of initial depth estimate and the final result of the algorithm with respect to the zero mean scanner data are given in mm in Table 7.2. The presented algorithm is capable of refining the initial surface estimate and reducing the RMSE by a factor of 2. Consequently, even a complex structure like the plaster mass that exhibits cavities and cuts deviates from the ground truth by less than 0.3 mm, which is similar in accuracy to Zhou et al. (2013) at a less extensive acquisition setup. Zhou et al. (2013) require a symmetric lighting pattern to compute and trace iso-depth contours on the object and use a ring-light system that contains 72 LEDs.

7.1.4 Summary

In this section, a new method has been proposed and evaluated to compute depth data and spatially varying reflectance parameters of an arbitrary and unknown surface. The information obtained from gray value images is used to refine the initially estimated 3D



(a) (b)

(c) (d)

Figure 7.9. Uniform shading of zSnapper reference depth data (a,c) and the final result of the proposed method (b,d). The absolute distance has to be guessed and is unknown if the algorithm is initialized with the pseudo-diffuse surface. The proposed approach results in smooth surfaces that still contain valuable details. The scanner data contain a significant amount of high-frequency noise. The smoothness of the shaded surfaces in (b) and (d) is not to be mistaken for blur.

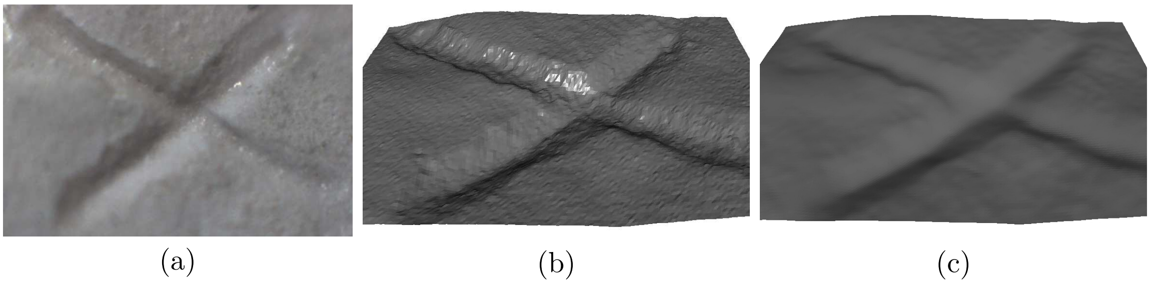


Figure 7.10. Close up view of 'X' scratch in plaster surface. (a) Image obtained with different lens and camera to achieve high-resolution zoom of the scratch mark, (b) shaded 3D scanner data, (c) shaded reconstruction. The edges and cracks inside the scratch that appear in the scanner data are not visible in the zoomed image. The smoother reconstruction of the depth is therefore probably closer to the real surface.

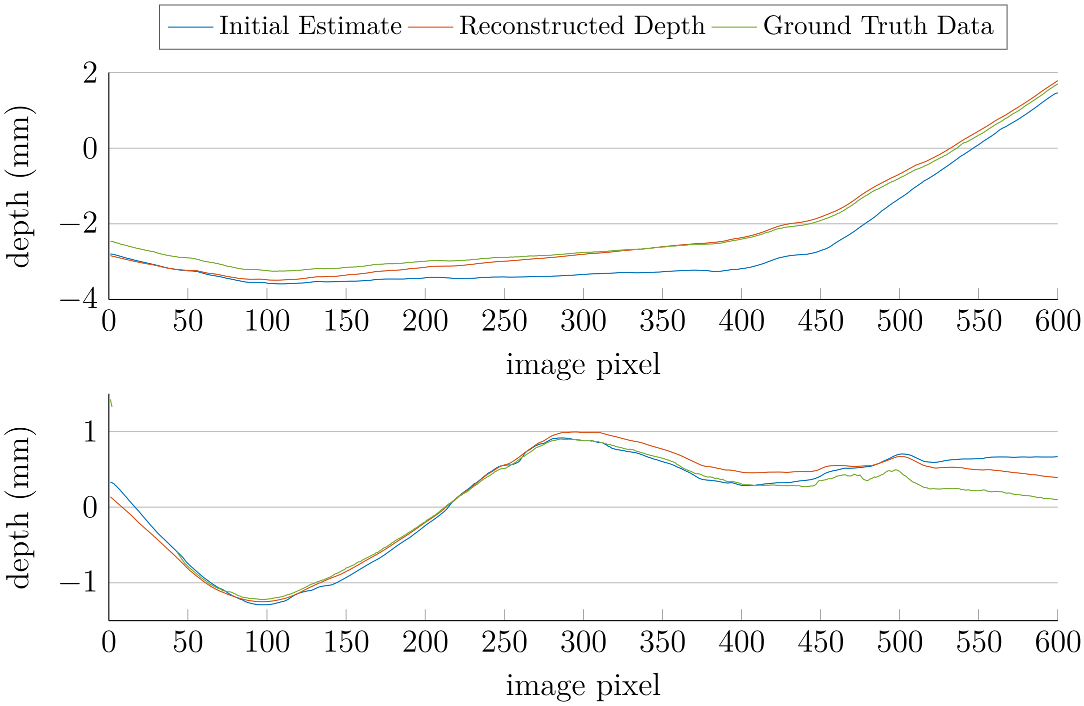


Figure 7.11. Horizontal depth profiles. White egg (upper diagram) and white plaster (lower diagram). The mean of all data sets has been subtracted to facilitate comparability. Note that the noise in the input data is less apparent than in Fig. 7.9 due to the scaling.

data. The results show that the procedure can be applied successfully to real-world objects. The resulting surface is less noisy than the 3D range scanner data while it preserves fine structures of the object under inspection. The accuracy with respect to the ground truth is high with an RMSE of less than 0.3 mm and the necessary measurement equipment amounts to a camera and some calibrated light sources.

The significance of the BRDF parameter maps to a human viewer is less evident than expected for the examined objects. The gradient field computed from those parameters is plausible and refines the initial surfaces to an accurate solution. This approach might be extended to physically based reflection functions to possibly derive physical properties to be interpreted directly by a human observer. Additionally, highly specular metallic surfaces pose a limitation of the proposed method, because the diffuse component of the reflection is minimal and the initial estimate of the surface structure fails to provide a suitable initialization. Hence, the BRDF parameter estimation results in erroneous parameters and the entire procedure diverges. This drawback could be overcome by using another source for the initial depth data, especially the features used for stereo matching that are discussed in Section 7.2 work very well on metallic objects and can provide a suitable initialization. Still, BRDF estimation on highly specular surfaces proved to be a difficult task in general.

7.2 Multi-View Photometric Stereo Constrained by Stereo Image Analysis

One of the main drawbacks of the method depicted in Section 7.1 is the scalability of the resulting surface gradients. There are no means to derive absolute depth data from a single camera image without some metric reference in the scene. In contrast, a calibrated stereo setup provides the necessary information to estimate the 3D coordinates of all scene points that are visible (and correctly located) in both images since the known distance between both cameras is the reference value. Finding corresponding points, however, is challenging and oftentimes relies on window-based features which favor pixel-locking and thus neglect details in the surface structure. Photometric methods, on the other hand, acquire gradient information locally at high precision. Yet, small errors in the estimation add up during the integration that is necessary to compute a coherent surface, resulting in depth maps that are distorted. Additionally, non-Lambertian reflectance functions impede or prohibit successful stereo matching, while many photometric algorithms have been designed to explicitly handle various types of BRDFs. Consequently, combining shape from shading and two-source stereo is an obvious choice since both procedures exhibit complementary shortcomings. Coarse traditional stereo depth data is refined with gradient information obtained from photometric stereo.

7.2.1 Motivation

A fully integrated approach that defines the estimation of 3D scene structure as the solution to a variational problem subject to constraints derived from stereo imaging is the main contribution of this section. There has been extensive research in the fields of multi-view stereo (MVS) (cf. Section 3.6) and shape from shading (SfS) (cf. Section 3.1) in the last decades. However, on many occasions, both fields of research that ultimately aim for the same goal have been treated individually. Some researchers have put an effort in the fusion of both approaches similar to the proposed idea. Wu et al. (2011a) improve MVS meshes of Lambertian surfaces based on shading cues under arbitrary lighting. Their limitation of a constant albedo has been overcome by Maurer et al. (2018) and Langguth et al. (2016) who can handle abruptly changing albedos. Langguth et al. (2016) use spherical harmonics similar to Wu et al. (2011a) to estimate the lighting setup and minimize an energy function that is independent of the surface albedo. Maurer et al. (2018) use a variational framework that does not need initialization and combine stereo and shading subject to strong regularizations. Still, these – and to the author's best knowledge all other – contributions in this field of study are limited to diffuse surfaces and additionally fail if there are gradual changes of the surface albedo since this is mistaken for changes of the surface gradient.

In contrast, the proposed method uses analytical BRDF models to handle complex uniform or spatially varying reflectance functions including metallic objects and provides accurate 3D models of the surfaces under inspection. An iterative update equation is derived from the variational problem formulation with the aim of matching the gradient fields of left and right camera view while simultaneously estimating a surface that is integrable with respect to this gradient field. The matching of the gradient fields provides a disparity map that constrains the depth estimate that itself is subject to the integrability. As *multi-view photometric stereo* method, the procedure requires a specific measurement setup that contains a calibrated stereo camera rig and a suitable number of light sources with calibrated incident

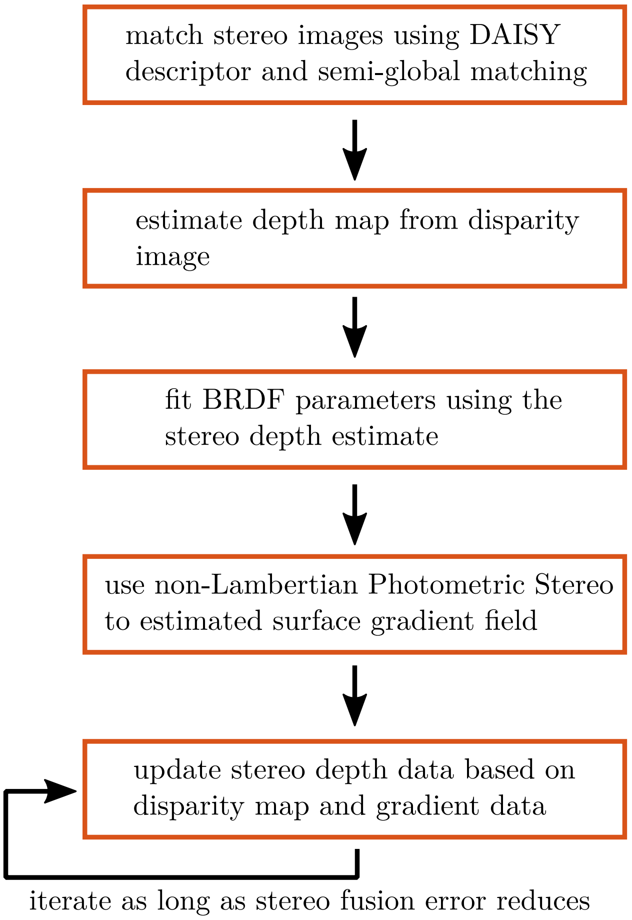


Figure 7.12. Flow chart of stereo fusion algorithm.

direction and strength. Both intrinsic and extrinsic parameters of the stereo cameras have been calibrated with Bouguet's Matlab toolbox (Bouguet, 2008), and the light sources have been calibrated according to the algorithm described in Chapter 6 prior to all computations.

7.2.2 Preliminary Computations

The principal novelty of this method is the combination of Horn (1990)'s shape from shading with the disparity map corresponding to a stereo camera setup based on a fully integrated approach. The initial disparity map is estimated by matching the DAISY descriptor (Tola et al., 2008, 2010) computed on each pair of images based on the minimal sum of squared differences. The stereo matching is refined with semi-global matching (SGM) (Hirschmüller, 2005) using the DAISY descriptor again as a measure of similarity. A left-right consistency check is applied to ensure valid disparities by eliminating every disparity value whose estimate from left or right camera differs by more than 5 pixels. Since a disparity map can be obtained for each pair of images, the final disparity is computed from the median value across all 18 images if there are at least 5 valid disparities. Note in Fig. 7.27 that this combination enables robust stereo estimation even on metallic surfaces and those that exhibit spatially varying non-Lambertian reflectance properties.

The second prerequisite for the procedure is a known pair of gradient fields $(p_\mathrm{L}, q_\mathrm{L})$, $(p_\mathrm{R}, q_\mathrm{R})$ for both, left and right camera view of the stereo rig. Photometric stereo based on the BRDF models by Blinn (Section 2.4.3) or Cook-Torrance (Section 2.4.5) is applied to both individual camera views according to the non-linear optimization problem described in Eq. (7.2). The BRDF parameters are estimated on the depth data estimated from the stereo correspondence between the cameras, yet every other source of depth data would be possible as well. If the parameters are estimated per-pixel, a 9×9 mask of Gaussian weights $\mathcal{N}(0, 4)$ with a variance $\sigma_{\text{mask}} = 4$, where the central pixel is the currently fit BRDF parameter vector, is applied to account for the measurements of surrounding pixels and reduce the influence of measurement uncertainties. Since 18 light sources provide a relatively sparse dataset per pixel this is considered necessary and could be omitted if the data were acquired from a wider range of measurements. Subsequently, the estimated parameters are used to compute the gradient fields. The complete algorithm is depicted in Fig. 7.12

7.2.3 Formulating the Variational Problem

The variational shape from shading method described by Horn (1990) solves the problem of finding the surface $z(x, y)$ that minimizes a function F dependent on z and two of its partial derivatives $z_x = \frac{\partial z}{\partial x} = p$, $z_y = \frac{\partial z}{\partial y} = q$ in a least squares sense, yielding

$$\min_z \iint F(z, z_x, z_y)\,\mathrm{d}x\,\mathrm{d}y\,. \tag{7.6}$$

The Euler-Lagrange equation used to solve this variational problem is

$$F_z - \frac{\partial}{\partial x} F_{z_x} - \frac{\partial}{\partial y} F_{z_y} = 0\,. \tag{7.7}$$

The complete error function presented in this section is composed of two parts. First, there is the integrability term that was already defined by Horn and Brooks (1986)

$$E_{\text{int}} = \iint \frac{1}{2}\Big[(z_x - p)^2 + (z_y - q)^2\Big]\,\mathrm{d}x\,\mathrm{d}y\,, \tag{7.8}$$

considering the observed gradient field (p, q). The second part and the contribution of this section is the disparity error comparing the gradient fields computed from two scene views based on the disparity δ defined by the depth estimate z

$$E_\delta = \iint \frac{1}{2}\Big[(p_{\text{R}}(\delta(z)) - p_{\text{L}})^2 + (q_{\text{R}}(\delta(z)) - q_{\text{L}})^2\Big]\,\mathrm{d}x\,\mathrm{d}y\,. \tag{7.9}$$

Consequently, the entire error function can be expressed by

$$\begin{aligned} E_{\text{SF}} = \iint \Big(&\frac{1}{2}\Big[(z_x - p_{\text{L}})^2 + (z_y - q_{\text{L}})^2\Big] \\ &+ \frac{\gamma_\delta}{2}\Big[(p_{\text{R}}(\delta(z)) - p_{\text{L}})^2 + (q_{\text{R}}(\delta(z)) - q_{\text{L}})^2\Big]\Big)\,\mathrm{d}x\,\mathrm{d}y\,. \end{aligned} \tag{7.10}$$

$\gamma_\delta \in [0, \infty]$ is a weighting factor to level the contribution of the individual error terms, such that for $\gamma_\delta = 0$ the Stereo Fusion error E_{SF} equals Horn's integrability error term. Inserting Eq. (7.10) into the Euler equation (7.7), where $F = E_{\text{SF}}$, separates both error terms again, because E_{int} is independent of z and E_δ is independent of z_x and z_y.

Disparity Error Term

The following derivations will refer to pixelwise estimates in the discrete domain and therefore skip the integration component of the equation. However, the pixel indices (u, v) will be omitted as well for the sake of readability if not necessary for comprehension. The first part of the Euler equation is given by the disparity error term Eq. (7.9):

$$F_z = \frac{\partial}{\partial z}\gamma_\delta E_\delta = \gamma_\delta \left(p_\mathrm{R}(\delta(z)) - p_\mathrm{L}\right)\frac{\partial p_\mathrm{R}}{\partial \delta}\frac{\partial \delta}{\partial z} + \gamma_\delta \left(q_\mathrm{R}(\delta(z)) - q_\mathrm{L}\right)\frac{\partial q_\mathrm{R}}{\partial \delta}\frac{\partial \delta}{\partial z}\,. \tag{7.11}$$

Disparity and depth are related through camera focal length f[1], base length $t_\mathrm{b} = \|\mathbf{t}_\mathrm{b}\|_2$ (cf. Section 3.6, Eq. (3.20)) of the stereo setup and horizontal extent of a sensor pixel k_u according to

$$\delta = \frac{f t_\mathrm{b} k_\mathrm{u}}{z} \tag{7.12}$$

and the disparity can thus be regarded as a function in z such that the discrete gradient $\Delta_\mathrm{z}\delta(z)$ exists.

$$\Delta_\mathrm{z}\delta(z) = \frac{\partial \delta}{\partial z} = -\frac{f t_\mathrm{b}}{z^2} \tag{7.13}$$

The gradients $(p_\mathrm{L}, q_\mathrm{L})$ and $(p_\mathrm{R}, q_\mathrm{R})$, respectively are known at discrete pixel values (u, v) only and have to be interpolated to match the re-projection onto the continuous image coordinates $(\hat{u}, \hat{v})$ of the current depth estimate $\hat{z}(x, y)$ (in contrast to the true, unknown depth $z(x, y)$). Reprojecting the depth value from 3D space onto the sensor will result in continuous coordinates on the image plane. These have to be converted to pixel coordinates by subtracting the principal point $\mathbf{c}_0 = (\hat{c}_\mathrm{u}, \hat{c}_\mathrm{v})$ and rounding to the nearest smaller or equal integer value. To prevent loss of information, the values of the (theoretically) continuous sensor pixels are interpolated based on the slope values m_p and m_q. Linear interpolation is assumed to be sufficiently accurate and thus

$$\begin{gathered} p_\mathrm{R}(\hat{u} - \hat{c}_\mathrm{u} - \delta, \hat{v} - \hat{c}_\mathrm{v}) = p_\mathrm{R}(u, v) + m_\mathrm{p}(\hat{u} - \hat{c}_\mathrm{u} - \delta - u, v), \\ u = \lfloor \hat{u} - \hat{c}_\mathrm{u} - \delta \rfloor\,. \end{gathered} \tag{7.14}$$

Note that rectified images in a calibrated stereo setup yield $\hat{v} - \hat{c}_\mathrm{v} = v$ disregarding the disparity δ. Of course, the same concept applies to q_R. The slope of the linear interpolation function is given by

$$m_\mathrm{p} = \frac{\partial p_\mathrm{R}}{\partial \delta}, \qquad m_\mathrm{q} = \frac{\partial q_\mathrm{R}}{\partial \delta} \tag{7.15}$$

and can be replaced in Eq. (7.11) yielding the final form of the disparity error based on the current depth estimate $\hat{z}(x, y)$:

$$F_z = -\gamma_\delta \frac{f t_\mathrm{b}}{z^2}\Big[\Big(p_\mathrm{R}(\delta(\hat{z})) - p_\mathrm{L}\Big)m_\mathrm{p} + \Big(q_\mathrm{R}(\delta(\hat{z})) - q_\mathrm{L}\Big)m_\mathrm{q}\Big]\,. \tag{7.16}$$

Integrability Error Term

The integrability error term Eq. (7.8) refers to the reference camera view only and therefore the indices of left and right camera are omitted for improved readability. Its partial

---

[1] As mentioned earlier, the focal length will be used as an approximation of the principal distance, i.e., $b = f$.

	p_u	
p_l	p_c	p_r
	p_d	

Figure 7.13. Mask to compute discrete gradients at center pixel p_c based on upper p_u, lower p_d, left p_l and right p_r pixel.

derivatives are given by

$$\begin{aligned} F_{z_x} &= \frac{\partial}{\partial x}(z_x - p) = \frac{\partial^2 z}{\partial x^2} - \frac{\partial p}{\partial x}\,, \\ F_{z_y} &= \frac{\partial}{\partial y}(z_y - q) = \frac{\partial^2 z}{\partial y^2} - \frac{\partial q}{\partial y}\,. \end{aligned} \tag{7.17}$$

Using the linear interpolation described in the previous section, it is suitable to employ sensor coordinates to index the depth estimates when operating in the discrete domain $\hat{z}(u, v)$. The first and second derivative can be approximated by finite differences from the current estimate of the depth data $\hat{z}$. The mask displayed in Fig. 7.13 denotes the necessary information with respect to a center pixel p_c based on upper p_u, lower p_L, left p_l, and right p_r pixel. For example if $p_\mathrm{c} = p(u, v)$ then $p_\mathrm{r} = p(u + 1, v)$.

Let the spatial extent of a pixel, i.e., the distance between two neighboring depth measurements, be given by k_u and k_v, respectively. The symmetric discrete derivatives in Eq. (7.17) can thus be substituted as follows (the relation for q is implicit):

$$\frac{\partial p}{\partial x} \approx \frac{p_\mathrm{r} - p_\mathrm{l}}{2k_\mathrm{u}}\,, \tag{7.18}$$

$$\frac{\partial^2 z}{\partial x^2} \approx \frac{z_\mathrm{r} - 2z_\mathrm{c} + z_\mathrm{l}}{{k_\mathrm{u}}^2}\,. \tag{7.19}$$

Note again that the derivative of the gradient is computed with respect to the camera coordinate system, i.e., in metric coordinates, while the value of the gradient is indexed in sensor coordinates. Thus, the statement $p_x(u, v)$ might appear curious but is a meaningful one that describes the gradient of the horizontal surface gradient corresponding to the intensity and depth values at sensor pixel (u, v). Introducing this relation into Eq. (7.17), the integrability error based on the current depth estimate $z_\mathrm{c} = \hat{z}(u, v)$ can be rewritten as

$$\begin{aligned} -\frac{\partial}{\partial x}F_{z_x} - \frac{\partial}{\partial y}F_{z_y} =& \frac{-z_\mathrm{r} + 2z_\mathrm{c} - z_\mathrm{l}}{{k_\mathrm{u}}^2} + \frac{p_\mathrm{r} - p_\mathrm{l}}{2k_\mathrm{u}} \\ &+ \frac{-z_\mathrm{d} + 2z_\mathrm{c} - z_\mathrm{u}}{{k_\mathrm{v}}^2} + \frac{q_\mathrm{d} - q_\mathrm{u}}{2k_\mathrm{v}}\,. \end{aligned} \tag{7.20}$$

Isolating z_c with the abbreviations $p_x = \frac{p_\mathrm{r} - p_\mathrm{l}}{2k_\mathrm{u}}$ and $q_y = \frac{q_\mathrm{d} - q_\mathrm{u}}{2k_\mathrm{v}}$ yields

$$2z_\mathrm{c}\left(\frac{{k_\mathrm{u}}^2 + {k_\mathrm{v}}^2}{{k_\mathrm{u}}^2{k_\mathrm{v}}^2}\right) - \frac{z_\mathrm{r} + z_\mathrm{l}}{{k_\mathrm{u}}^2} - \frac{z_\mathrm{d} + z_\mathrm{u}}{{k_\mathrm{v}}^2} + p_x + q_y = 0\,. \tag{7.21}$$

Stereo Fusion Error

Finally, all these findings can be inserted into the Euler equation (7.7). The result is the Stereo Fusion error term that leads to the iterative update scheme.

$$\begin{aligned} &2z_\mathrm{c}\left(\frac{{k_\mathrm{u}}^2+{k_\mathrm{v}}^2}{{k_\mathrm{u}}^2{k_\mathrm{v}}^2}\right)-\frac{z_\mathrm{r}+z_\mathrm{l}}{{k_\mathrm{u}}^2}-\frac{z_\mathrm{d}+z_\mathrm{u}}{{k_\mathrm{v}}^2}+p_x+q_y \\ &-\gamma_\delta\frac{ft_\mathrm{b}}{z_\mathrm{c}^2}\left[\Big(p_\mathrm{R}(\delta(\hat{z}))-p_\mathrm{L}\Big)m_\mathrm{p}+\Big(q_\mathrm{R}(\delta(\hat{z}))-q_\mathrm{L}\Big)m_\mathrm{q}\right]=0 \end{aligned} \tag{7.22}$$

Isolating the highest power of z_c yields the cubic equation

$$\begin{aligned} &z_\mathrm{c}^3+z_\mathrm{c}^2\frac{{k_\mathrm{u}}^2{k_\mathrm{v}}^2(p_x+q_y)-{k_\mathrm{v}}^2(z_\mathrm{r}+z_\mathrm{l})-{k_\mathrm{u}}^2(z_\mathrm{d}+z_\mathrm{u})}{2({k_\mathrm{u}}^2+{k_\mathrm{v}}^2)} \\ &-\frac{\gamma_\delta ft_\mathrm{b}\left[\Big(p_\mathrm{R}(\delta(\hat{z}))-p_\mathrm{L}\Big)m_\mathrm{p}+\Big(q_\mathrm{R}(\delta(\hat{z}))-q_\mathrm{L}\Big)m_\mathrm{q}\right]{k_\mathrm{u}}^2{k_\mathrm{v}}^2}{2({k_\mathrm{u}}^2+{k_\mathrm{v}}^2)}=0\,. \end{aligned} \tag{7.23}$$

The solution to a cubic equation of the form

$$a_3z^3+a_2z^2+a_1z+a_0=0 \tag{7.24}$$

can be found in the literature (e.g., Bronstein et al., 2016). The polynomial coefficients

$$\begin{aligned} a_3 &=1 \\ a_2 &=\frac{{k_\mathrm{u}}^2{k_\mathrm{v}}^2(p_x+q_y)-{k_\mathrm{v}}^2(z_\mathrm{r}+z_\mathrm{l})-{k_\mathrm{u}}^2(z_\mathrm{d}+z_\mathrm{u})}{2({k_\mathrm{u}}^2+{k_\mathrm{v}}^2)} \\ a_1 &=0 \\ a_0 &=-\frac{\gamma_\delta ft_\mathrm{b}\left[\Big(p_\mathrm{R}(\delta(\hat{z}))-p_\mathrm{L}\Big)m_\mathrm{p}+\Big(q_\mathrm{R}(\delta(\hat{z}))-q_\mathrm{L}\Big)m_\mathrm{q}\right]{k_\mathrm{u}}^2{k_\mathrm{v}}^2}{2({k_\mathrm{u}}^2+{k_\mathrm{v}}^2)} \end{aligned} \tag{7.25}$$

and the abbreviations

$$\begin{aligned} A_0 &=\frac{1}{9}\Big(3a_1-a_2^2\Big) \\ A_1 &=\frac{1}{54}\Big(9a_2a_1-27a_0-2a_2^3\Big) \\ A_2 &=A_0^3+A_1^2 \\ A_3 &=\Big(A_1+\sqrt{A_2}\Big)^{\frac{1}{3}} \end{aligned} \tag{7.26}$$

lead to the three possible solutions

$$\begin{aligned} z_1 &=-\frac{a_2}{3}+A_3-\frac{A_0}{A_3} \\ z_{2,3} &=-\frac{a_2}{3}+\frac{A_0}{2A_3}-\frac{A_3}{2}\pm j\sqrt{3}\left(\frac{A_0}{A_3}+A_3\right)\,, \end{aligned} \tag{7.27}$$

here j is the imaginary unit and $j^2=-1$. Since the solution is one of the three (possibly) complex roots of the function, a valid solution has to be chosen to continue the next iteration with. Here the real solution to $\hat{z}(u,v)$ that is closest in terms of Euclidean distance to the current estimate $\hat{z}(u,v)$ is accepted.

7.2.4 Evaluation

The presented procedure is evaluated on four laboratory datasets. That means 18 images per camera of the same object illuminated from different known directions are the only input data necessary. Exemplary images of the datasets are depicted in Fig. 7.14 and comprise two objects manufactured by selective laser sintering, a manually formed plaster mass partially painted with satin varnish (already used in Section 7.1.3) and a leaf-shaped piece of metal. The full datasets can be found in Appendix A.4. The 3D printed objects (*Ganesh* and *Ram*) exhibit an almost Lambertian reflectance behavior that is relatively easy to model, yet the shape is rather complicated yielding depth discontinuities and occlusions that result in many cast shadows under directional lighting and as such are difficult to be handled by image-based methods. The *Plaster* dataset provides spatially varying reflectance properties on a very similar albedo. The *Leaf* dataset exhibits a comparatively simple shape, but the metallic surface poses an obstacle to accurate reconstruction by both, classical stereo vision and laser scanning methods. As a matter of fact, the proposed method outperforms the visual quality of the 3D scanner that is supposed to provide reference data during the experimental analysis.

Reference 3D data of all objects have been acquired in the laboratory, to reduce speckle noise a Gaussian filter $\mathcal{N}(0,2)$ of size 13×13 pixels has been applied to the scanner data. Pixelwise comparison of the depth measurements are preferable, yet not possible without further processing since the stereo fusion and fringe projection scanner data have neither the same grid nor the same viewing direction of the object. Hence, the different views are matched with the *iterative closest points* (ICP) algorithm (Zhang, 1994) and the grid size is adjusted by interpolating from higher to lower resolution, which here means that the 3D scanner data is interpolated onto the grid of the stereo camera depth data.

Determining the Disparity Error Weight

Since the disparity error weight γ_δ has an impact on the optimization problem – between solely relying on integrability and completely eliminating integrability – a suitable operating point has to be identified. The iterative optimization of the disparity map has been evaluated on all datasets including three distinct setups of the Ram dataset for values of $\gamma_\delta \in [0, 1000]$. The resulting surface estimate is compared to a low-pass filtered version of the 3D scanner data to reduce the influence of high-frequency noise. The RMSE between optimization and reference data as a function of the disparity error weight is given in Fig. 7.15. It is evident that all errors increase significantly for values of $\gamma_\delta > 100$, yet for lower values, a clear minimum is hard to identify. Since the Ram exhibits the highest amount of self-cast shadows that hinder the initial estimation of adequate BRDF parameters and consequently impede a reliable estimation of surface normals, two simplifying conditions are introduced into the algorithm. The first one is the assumption of a uniform BRDF, thus facilitating the parameter estimation, and the second one is the application of an image-wise shadow mask to mark all areas of cast shadow and exclude them from any consideration. Especially the Ram dataset subject to both conditions (yellow line) hints to $\gamma_\delta \approx 7$ to 10 as a reasonable choice. Now that this also suits all other datasets, while still maintaining a notable contribution of the newly introduced disparity error term, $\gamma_\delta = 10$ is used for all experiments.

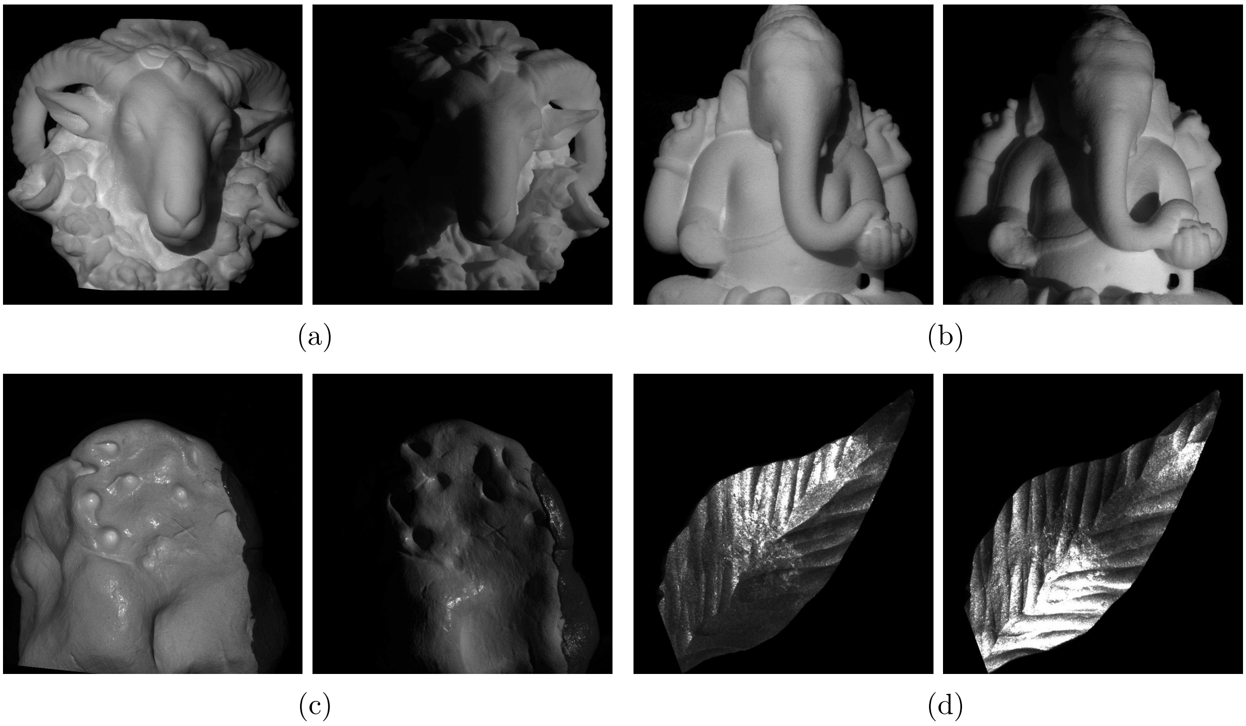


Figure 7.14. Exemplary images of the datasets used to evaluate the stereo fusion algorithm. (a) Ram, (b) Ganesh, (c) Plaster and (d) Leaf. The images are captured with the left camera under varying incident light to demonstrate object diversity. Same gray value scaling in images of the same object.

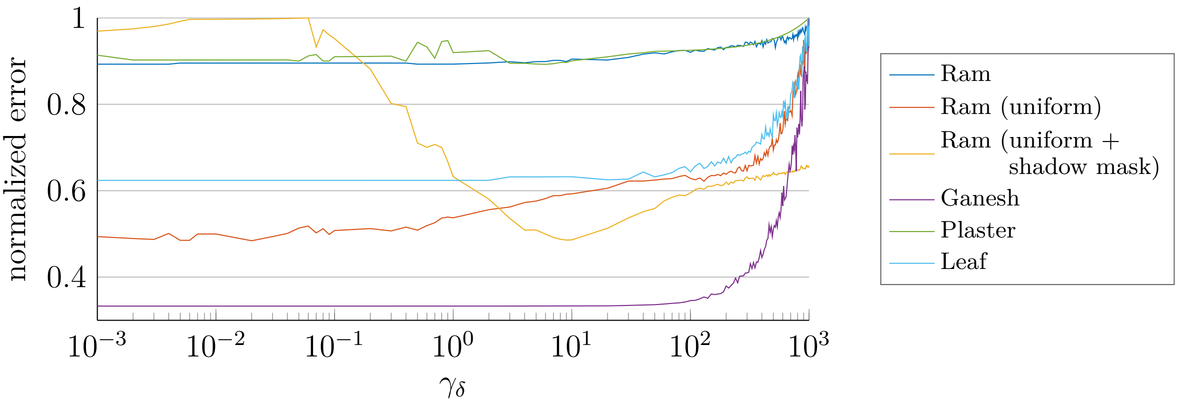


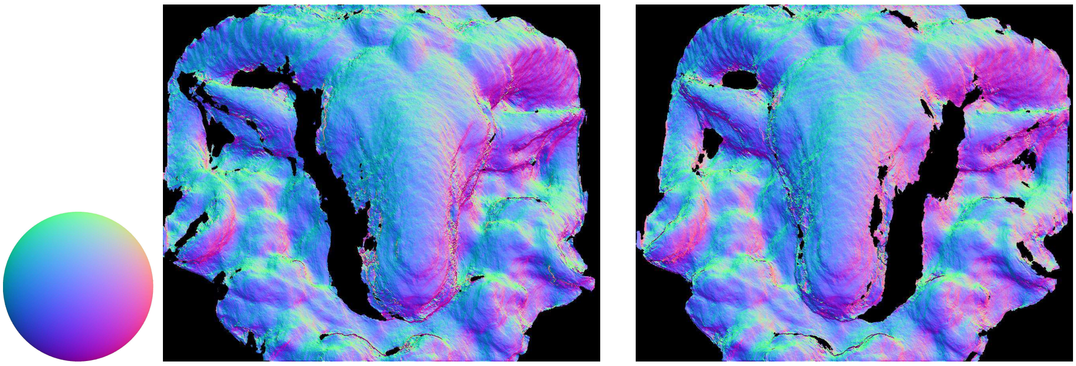
Figure 7.15. RMSE with respect to reference data for different values of γ_δ. All vectors have been normalized to their individual maximum to improve visibility. Especially the yellow Ram dataset, subject to a uniform BRDF and prior knowledge about cast shadows, hints to a good value of $\gamma_\delta \approx 7$ to 10.

Ram

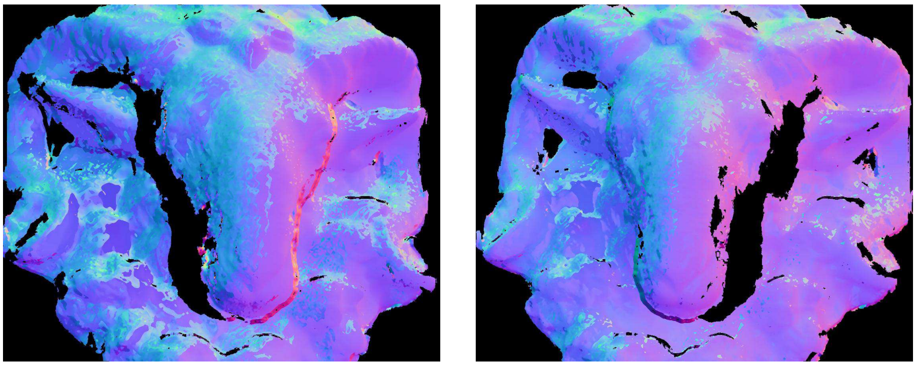
The Ram dataset has been evaluated under different settings that ultimately lead to the conclusion that – as expected – the best result is obtained by applying the shadow mask and using a uniform set of reflectance parameters. Figure 7.16 depicts the gradient fields estimated with PS for both, left and right stereo camera. The similarity of the corresponding gradient fields is a necessary precondition for the successful optimization of the depth map, and the possible level of detail is limited by the gradient as well. The surface normals computed from the stereo disparity maps in the first row of Fig. 7.16 already underline the necessity of the photometric data since the stereo gradients exhibit an artificial fine scale roughness. The pixelwise BRDF parameters lead to falsely estimated gradients especially in the areas affected by shadows, such that the uniform BRDF increases the stability of the normal estimates, and furthermore, normals are estimated for all parts of the surface that yield valid intensity measurements. Applying a-priori knowledge of cast shadows further improves the normal estimation, which of course is expected behavior. The shadow maps have been manually acquired per image to assess the prospects of the presented algorithm without the need for an evaluation of different shadow detection methods. However, there exist a variety of methods to detect shadows (e.g., Finlayson et al., 2002; Barsky and Petrou, 2003; Mukaigawa et al., 2007) and remove them (e.g., Xu et al., 2006; Khan et al., 2016) based on different techniques. Shadow removal is not of interest since there is no guarantee that the interpolated values have any physical meaning, yet the detection and hence exclusion of shadows always improves the performance of photometric applications.

Figure 7.17 displays the pixelwise depth differences between the 3D scanner and the Stereo Fusion under the above mentioned different settings. Contrary to the normal fields in Fig. 7.16, the differences between the results are small, indicating a strong dependence on the initialization. Still, a significant area of the surface cannot be evaluated since there are just no reference data. Thus the evaluation has to be (partially) qualitative again. Uniform shadings of two different perspectives corresponding to each input-data-setting, the stereo initialization, and the reference data are depicted in Fig. 7.18.

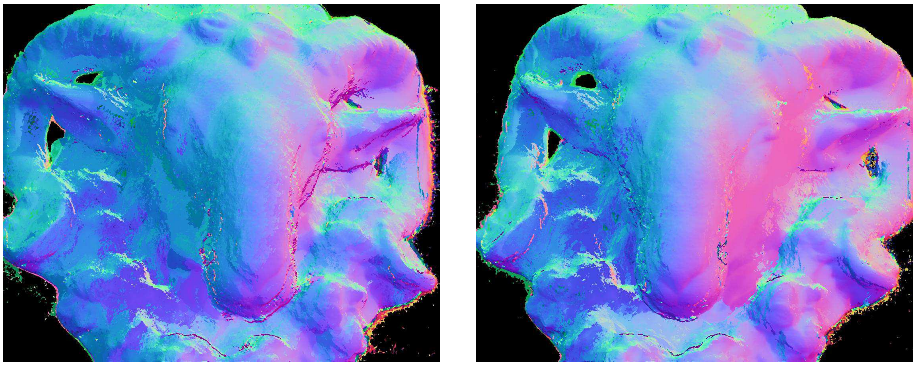
The 3D scanner reaches the highest level of fine-scale detail at the cost of the smallest surface area covered. The focus region of the scanner is already small, and the complex shape further reduces the visibility of the projection pattern as the object partially blocks it. The surface estimate obtained from the stereo matching already provides a reasonable structure of the Ram, yet it yields an artificial surface roughness and a double contour in the snout area. The surface roughness and the visibility of the double contour are reduced with the pixelwise BRDF data but still visible. A further refinement is reached with the uniform BRDF gradient data, yet additional spikes in the surrounding area reduce the quality of the reconstructions. The smoothest surface is computed from the highest quality gradient fields based on uniform BRDF and shadow masking. The headdress and the texture of the horns are well depicted while smoothing out the stereo noise. However, the estimated surface appears to exhibit over-smoothing in the flower wreath surrounding the head and a distortion in the left snout area. The flower wreath exhibits many small cavities that are severely affected by interreflections. Although the gradient field Fig. 7.16d appears to yield many fine details, the curvature in the flower wreath is too small and thus enables the over-smoothing. The discontinuity at the right-hand side of the head poses a problem to all reconstruction procedures. The side view in Fig. 7.18 illustrates that the absolute dimensions of the ram head are very similar for the stereo data and the 3D scanner, and while smoothing the noise and emphasizing details, the large scale is still correct for the stereo fusion results.



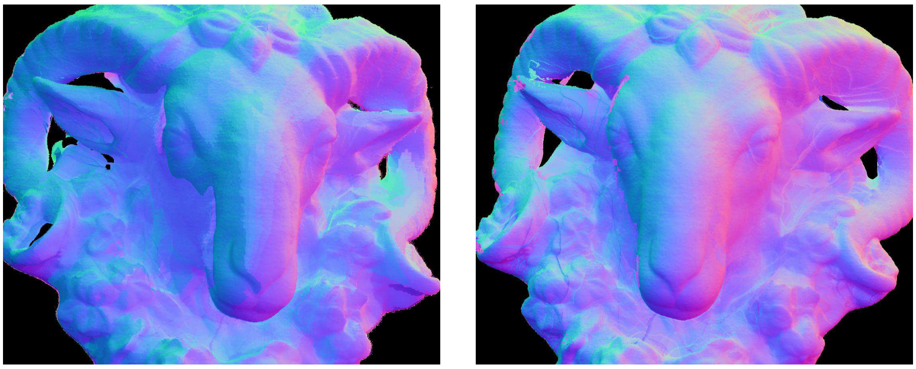
(a) Stereo



(b) Photometric Stereo

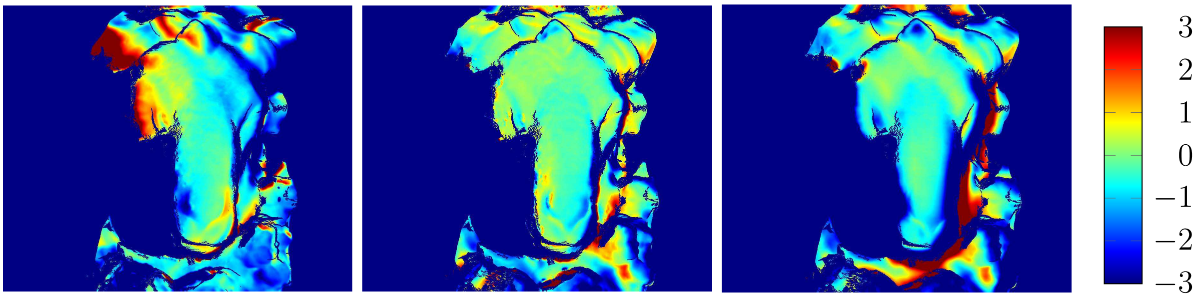


(c) Photometric Stereo (uniform BRDF)



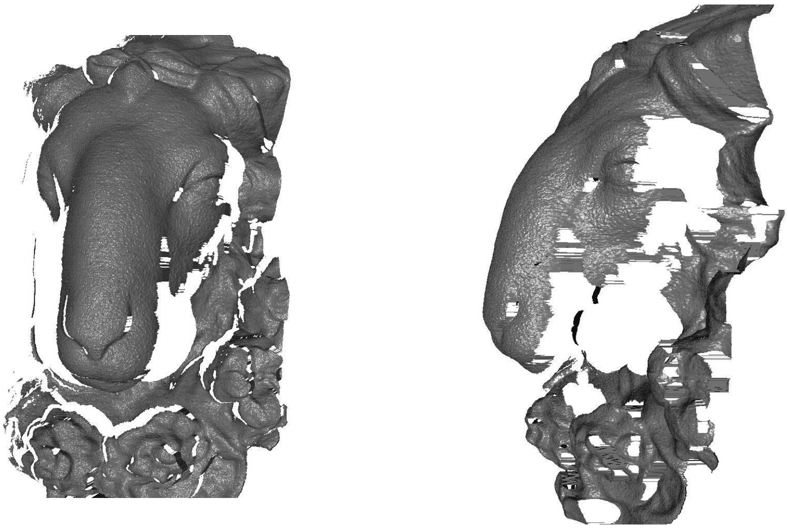
(d) Photometric Stereo (uniform BRDF + shadow mask)

Figure 7.16. Normals maps of the Ram dataset from different sources for left and right camera view. Reference sphere is given for comprehension only.

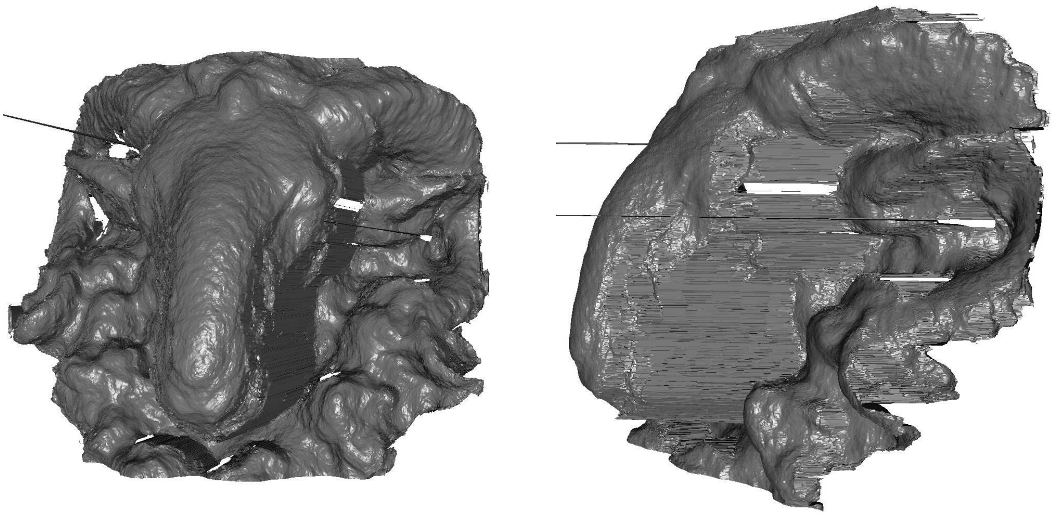


(a) $E_{\mathrm{RMSE}} = 2.7862\,\mathrm{mm}$ (b) $E_{\mathrm{RMSE}} = 2.6159\,\mathrm{mm}$ (c) $E_{\mathrm{RMSE}} = 2.2474\,\mathrm{mm}$

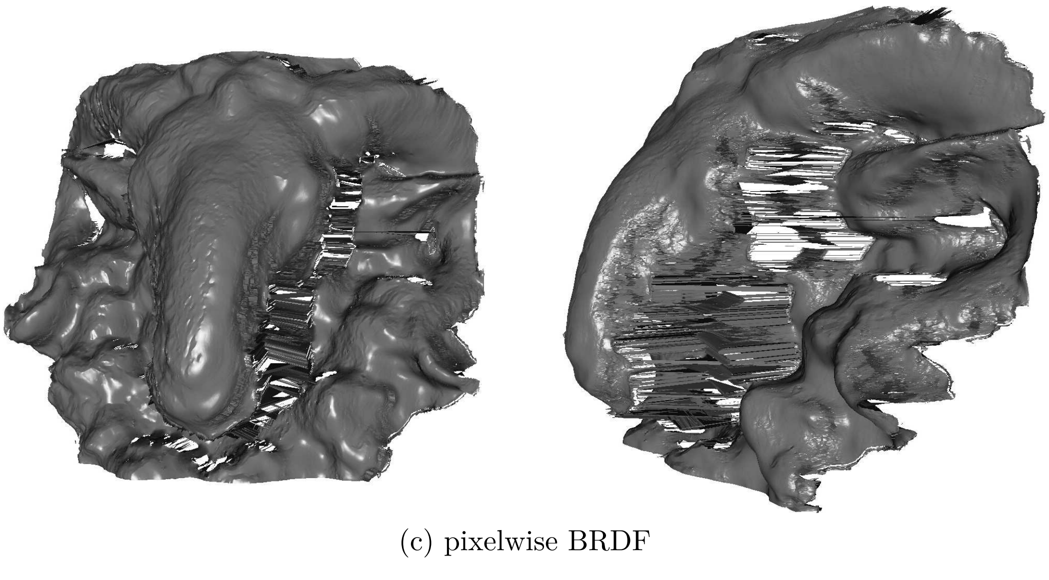
Figure 7.17. Pixelwise depth differences of the Ram dataset to 3D scanner. (a) PS with pixelwise BRDF, (b) PS with uniform BRDF, (c) PS with uniform BRDF and shadow masking. The RMSEs are quite similar, but there are no reference data for a lot of surface points. Scaling in millimeters.



(a) 3D Scanner

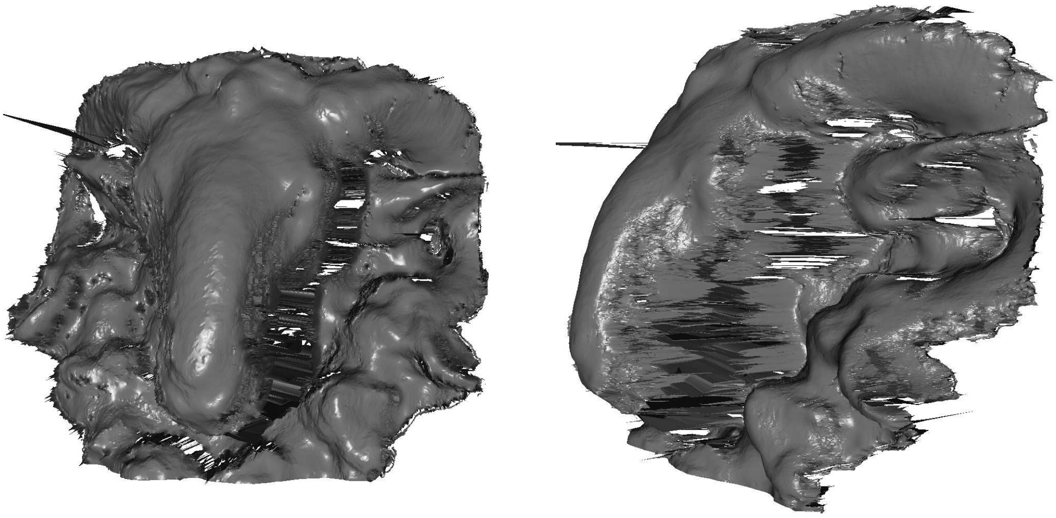


(b) Stereo Depth

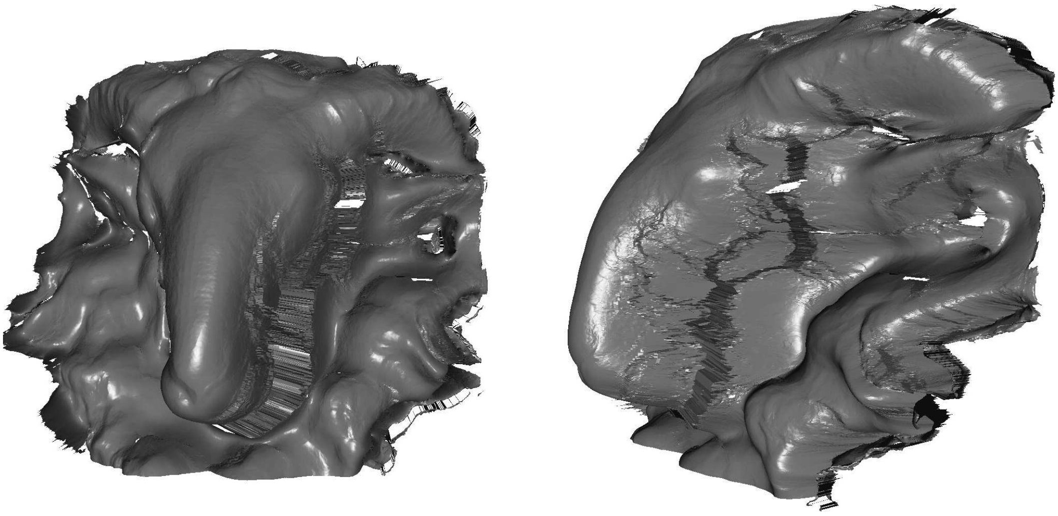


(c) pixelwise BRDF

Figure 7.18. Uniform shadings of the Ram surface.



(d) uniform BRDF



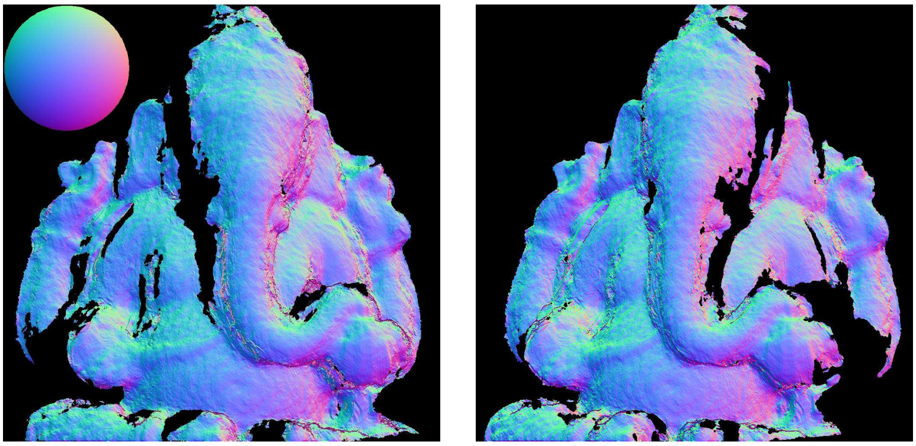
(e) uniform BRDF + shadow masking

Figure 7.18 (*cont.*). Uniform shadings of the Ram surface.

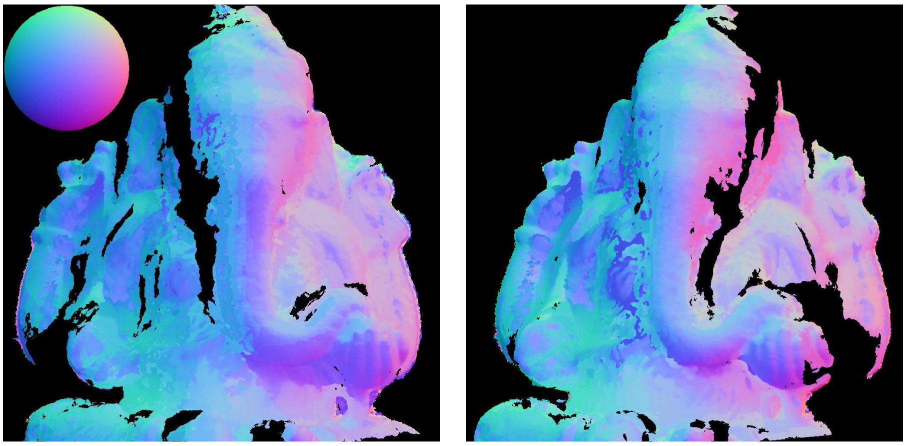
Ganesh

The Ganesh dataset exhibits reflection properties similar to those of the Ram dataset, an almost diffuse reflectance component, and a uniform BRDF, while the level of fine-grain details is lower especially concerning the flower wreath surrounding the ram head. At the same time, depth discontinuities and self-shadowing pose a difficult setting for the application of photometric surface reconstruction algorithms. The photometric normal maps computed from the stereo matching and the photometric stereo are depicted in Fig. 7.19. Typical for the stereo correspondence the surface structure is well captured but exhibits the artificial roughness that is caused by pixel-locking in the matching process. The photometric gradient fields, on the other hand, are smooth and yield an increased level of detail, above all in the region of the left hand holding Ganesh's trunk. The left side of the image yields a large area of little detail, similar to Fig. 7.16b, such that this is most likely caused by the high amount of self-cast shadows in the input image sequence and, in consequence, the pixelwise BRDF estimation that lacks valid data in the shadowed regions. Hence, the photometric normals do not reach the possible degree of high-frequency details.

Obviously, the quality of the gradient fields that are supposed to provide additional information to refine the stereo estimate has a substantial impact on the resulting surface



(a) Stereo



(b) Photometric Stereo

Figure 7.19. Normal maps of the Ganesh dataset. Reference sphere is given for comprehension only.

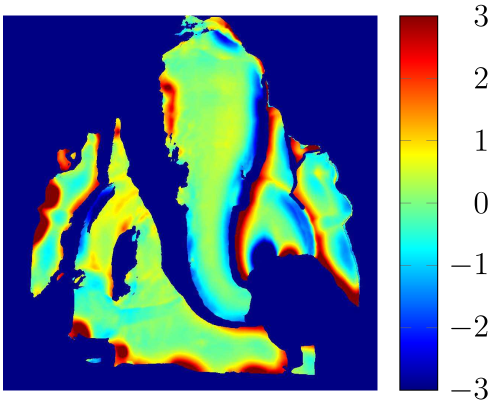
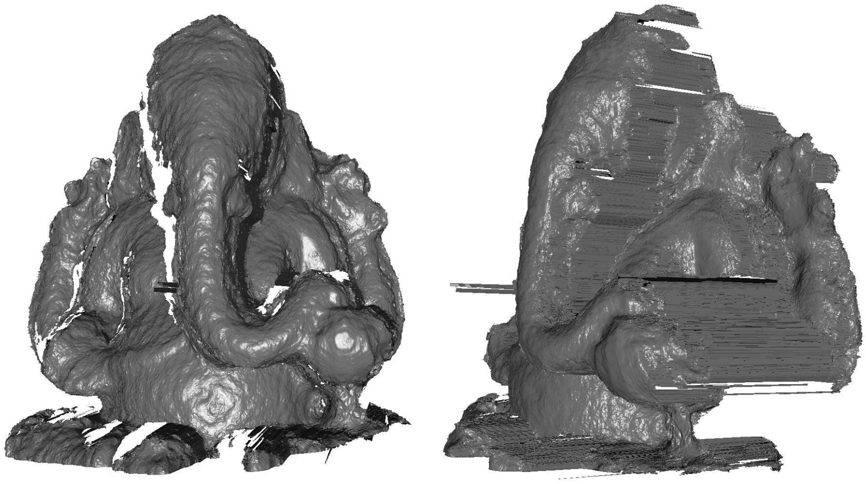


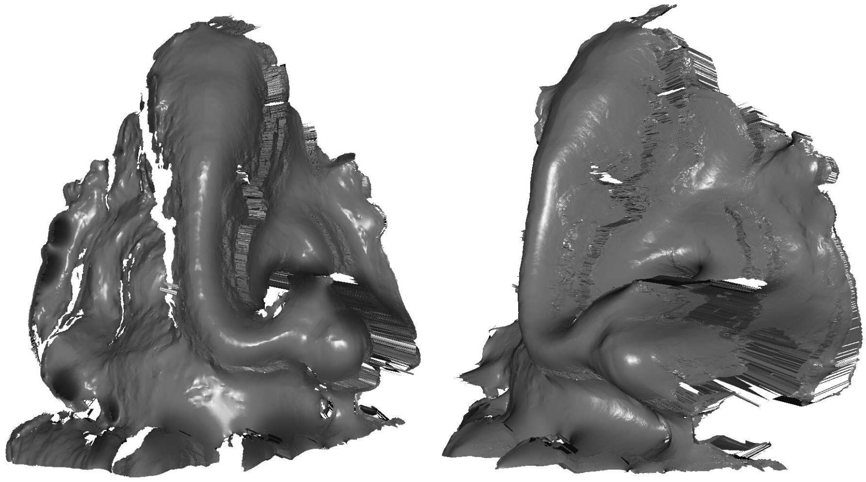
Figure 7.20. Pixelwise depth differences of the Ganesh dataset, $E_{\mathrm{RMSE}} = 1.2414\,\mathrm{mm}$. Scaling in millimeters.

estimate. The pixelwise differences to the reference data is displayed in Fig. 7.20. The falsely estimated gradient fields that are constraints to the differential equations of the surface curvature lead to divergence of the edges which is evident as well in Fig. 7.21c. The central regions of the figurine differ little from the reference data and the diverging edges still only lead to an RMSE of $E_{\mathrm{RMSE}} = 1.2414\,\mathrm{mm}$. The uniform shadings of the surface depicted in Fig. 7.21 provide an easier qualitative assessment of the resulting surface. The estimated 3D structure yields, apart from edges on the left side, a valid estimate, yet many small details of the original figurine are not captured in the stereo fusion reconstruction. Contrasting the image-based methods to the 3D scanner shows again that the 3D scanner does not capture a large part of the left side of the figurine due to shadowing of the projection pattern. This can be compensated by stitching multiple depth shots obtained for example on a rotation stage, but the stereo methods capture depth data for all image points that are visible to the reference camera.

(a) 3D Scanner



(b) Stereo Depth



(c) Stereo Fusion

Figure 7.21. Uniform shadings of the Ganesh surface.

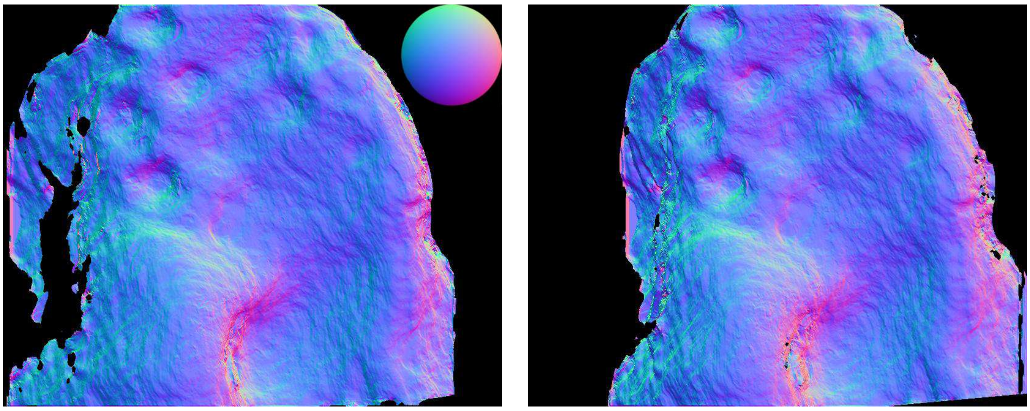
Plaster

The Plaster dataset has already been used in the evaluation in Section 7.1.3 and only the main properties are recalled here. The object exhibits small cavities and concave areas that promote interreflections, and while the albedo is constant, the specular component changes across the surface. Since the reflectance properties are spatially varying, a pixelwise BRDF estimation is required. The normal maps derived from stereo and photometric stereo are displayed in Fig. 7.22 and illustrate properties similar to the previous datasets. The stereo matching produces an artificial roughness, and the Photometric Stereo gradient fields show a high level of fine details. Note especially the 'X' scratch mark. The slope on the left-hand side of the object is the part most influenced by cast shadows[2] and therefore again lacks accurate gradient estimates. The pixelwise comparison to the reference data shown in Fig. 7.23 demonstrates the effectiveness of the refined stereo estimate with an $E_{\mathrm{RMSE}} = 0.4139\,\mathrm{mm}$. The highest deviations are found inside the small upper and larger lower cavities since the interreflection violates the assumption of direct illumination that is a foundation to photometric stereo.

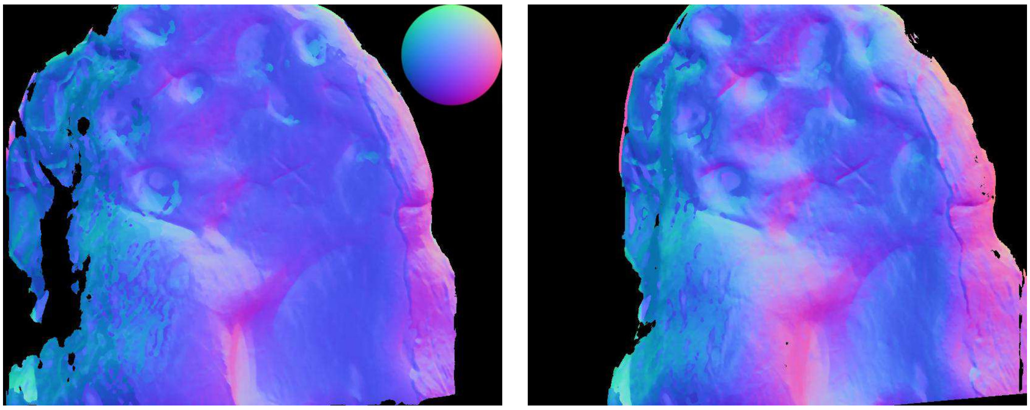
The uniform shadings of the reconstructed surface presented in Fig. 7.24 show that the stereo fusion reveals small details like the 'X' scratch mark (enlarged in the blue box) that

---

[2] Two-thirds of the light sources are arranged to the right of both cameras.



(a) Stereo



(b) Photometric Stereo

Figure 7.22. Normal maps of the Plaster dataset. Reference sphere is given for comprehension only.

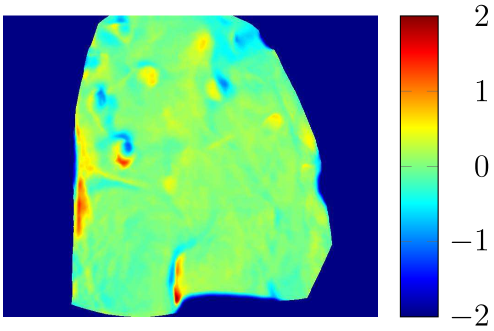
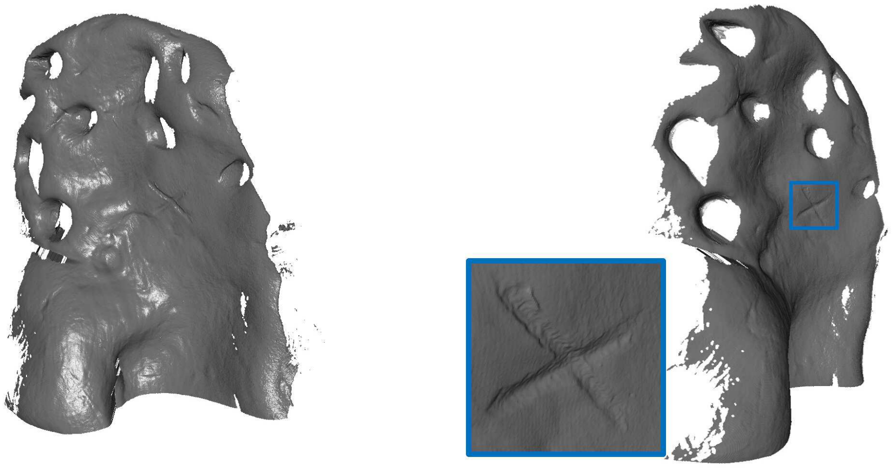
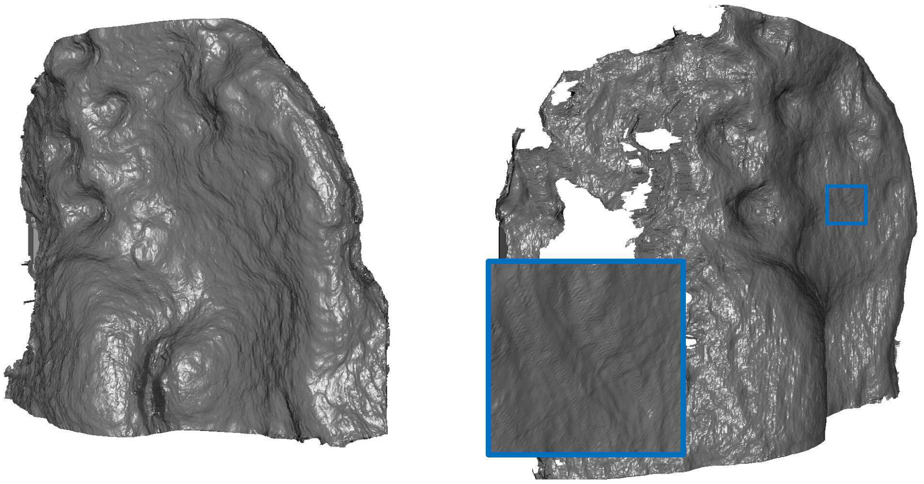


Figure 7.23. Pixelwise depth differences of the Plaster dataset, $E_{\mathrm{RMSE}} = 0.4139\,\mathrm{mm}$. Scaling in millimeters.

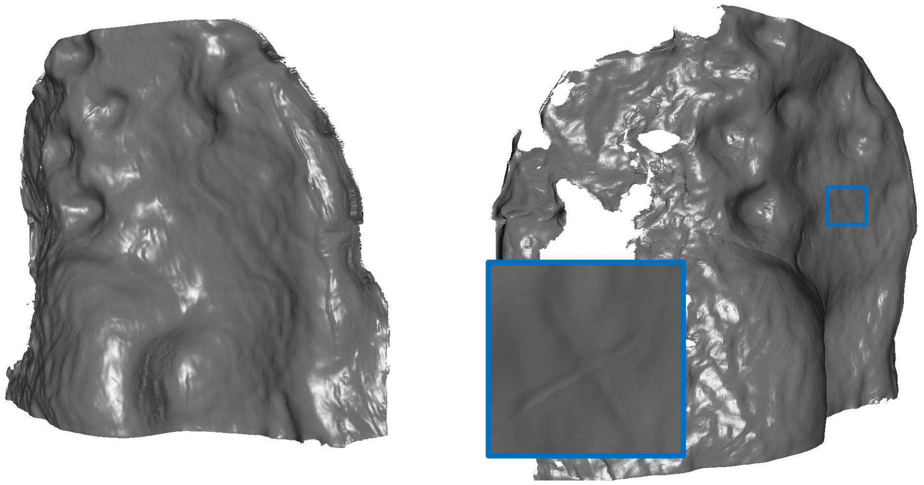
are not visible in the stereo depth data while maintaining a smoother surface. The roughness of the stereo data caused by pixel-locking is not completely removed and most visible on the left slope, exactly where the gradient data is the least accurate. Since both gradient fields exhibit the same structure on the left slope, the matching of the gradient fields is feasible. However, the integrability error that is the second constraint requires a high-quality gradient field to result in a good refinement of the initial surface. Although the 3D scanner achieves the highest level of detail, it only produces sparse depth data where almost all cavities are "holes".



(a) 3D Scanner



(b) Stereo Depth



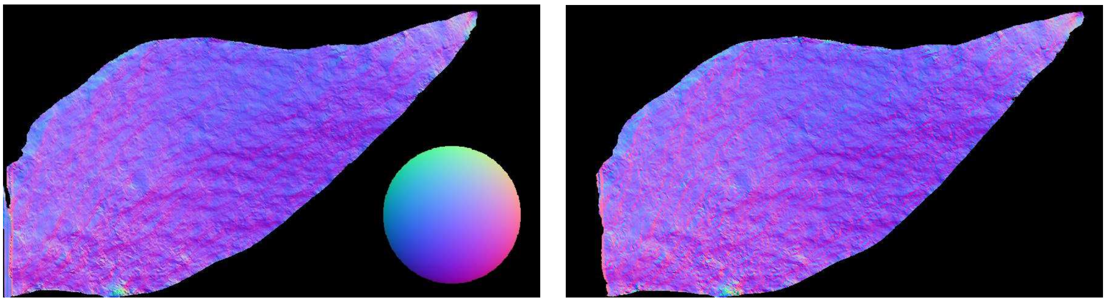
(c) Stereo Fusion

Figure 7.24. Uniform shadings of the Plaster surface. The 'X' scratch mark is not visible in the stereo depth data but maintained by the stereo fusion while smoothing the pixel-locking of the stereo correspondences.

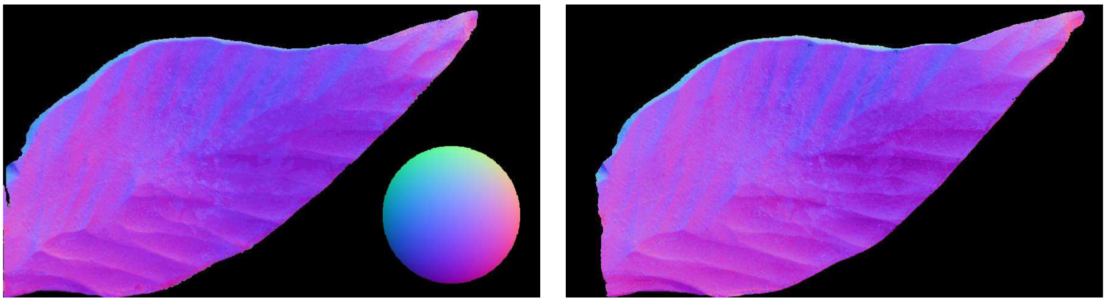
Leaf

The Leaf dataset is of special interest as it contains a metallic surface that inherently poses a difficult setting for the acquisition of accurate 3D information through various measurement procedures. Highly specular surfaces reflect (almost) all incoming light in the direction of mirror-like reflectance and make appearance capture tedious. Speckle noise and spurious reflection on shiny surfaces deteriorate the quality of the estimated 3D surface for scanning systems using lasers or other coherent sources of emission (e.g., Amir and Thörnberg, 2017). Although the light sources used in the photometric stereo step have a narrow bandwidth, it is not necessary to accurately detect small variations of a projection pattern. To facilitate the reflectance parameter estimation process on the challenging metallic surface a uniform BRDF is assumed. The Cook-Torrance BRDF and its modeling of surfaces as a collection of microfacets suit metallic objects very well and it is therefore preferred over the Blinn BRDF that was used for all remaining datasets. The normal maps in Fig. 7.25 illustrate very well that the stereo matching using the DAISY descriptor already achieves a reasonable depth estimate even on a highly non-Lambertian surface. However, it is also evident that the normal maps exhibit artificial roughness and lack all details of the fine leaf structure, whereas the PS normal fields contain all those details.

The pixelwise differences depicted in Fig. 7.26 show a high similarity between Stereo Fusion estimate and reference data with a mean error $E_{\mathrm{RMSE}} = 0.5073\,\mathrm{mm}$. The left edge of the leaf exhibits the highest deviation, and the cause is obvious from the uniform shadings displayed in Fig. 7.27. A presumably wrong correspondence between the stereo images leads to a discontinuity in the depth estimate of the lower left edge of the leaf. This error is smoothed by the proposed procedure, yet not fully corrected. On the contrary, the smoothing affects additional areas of the surface that have been correctly estimated from the stereo



(a) Stereo



(b) Photometric Stereo

Figure 7.25. Normal maps of the Leaf dataset. Reference sphere is given for comprehension only.

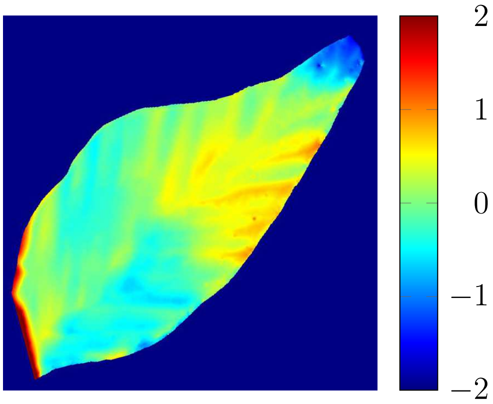
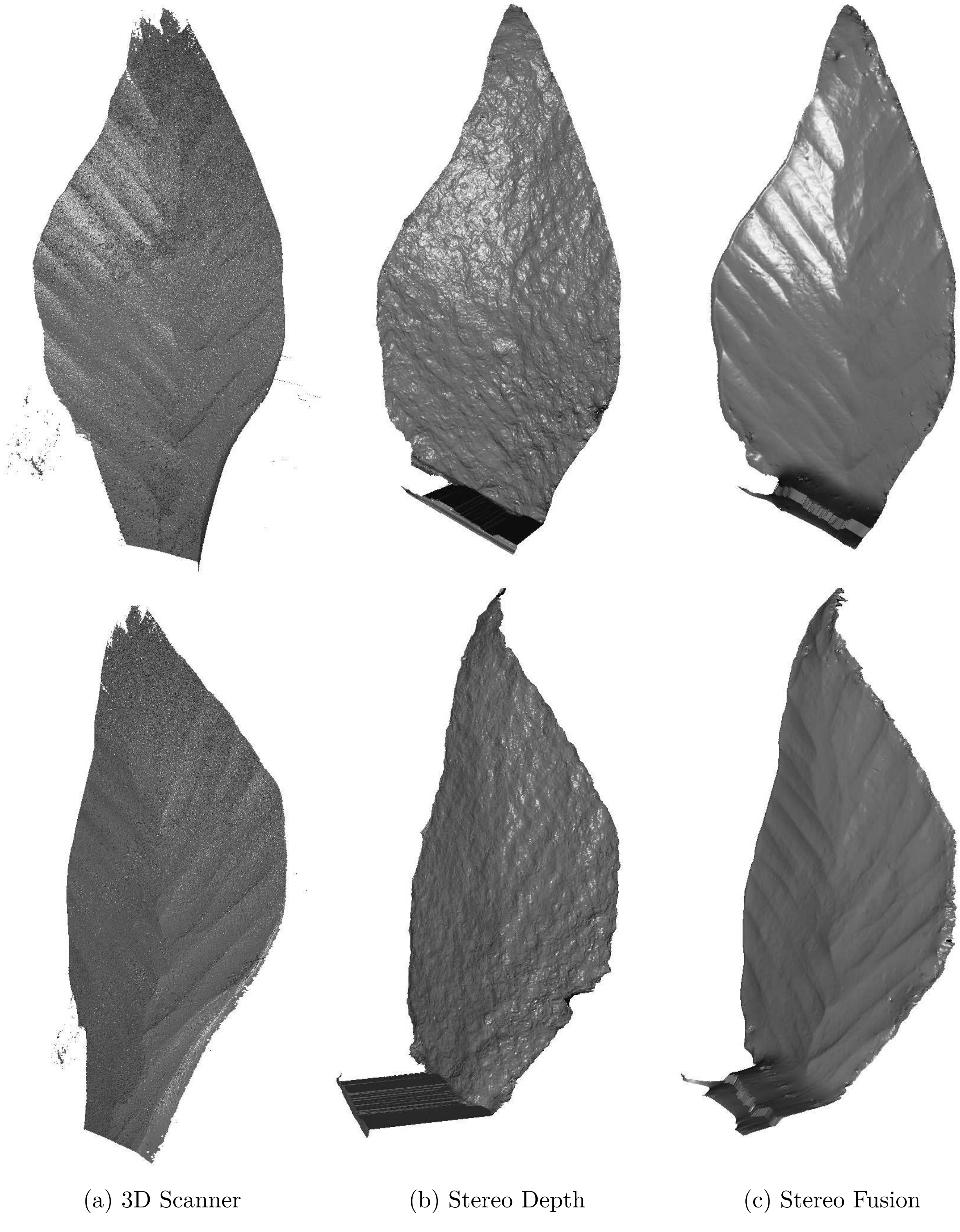


Figure 7.26. Pixelwise depth differences of the Leaf dataset, $E_{\mathrm{RMSE}} = 0.5073\,\mathrm{mm}$. Scaling in millimeters. The high error on the left lower edge is caused by a false initialization that prevented convergence in this area.

matching. The dependence of the algorithm on a proper initialization is a drawback, yet the stereo fusion algorithm very well reconstructs the remaining parts of the surface. The surface in Fig. 7.27c possesses the fine structure of the metallic leaf while being smooth whereas the reference scan is extremely noisy. In short, the source of reference data was outperformed by the proposed method.

The level of noise on the 3D scanner data is so high that only a low-pass filtered version of the shape can be considered as a valid reference. Although the stereo disparity map contains only the coarse shape of the leaf with a very bumpy structure, the ramifications of the leaf are completely visible in the Stereo Fusion results. Many image-based surface reconstruction methods do not even consider non-Lambertian objects and the proposed method surpasses the reconstruction accuracy of the reference 3D scanner on this highly specular cast iron object.



(a) 3D Scanner (b) Stereo Depth (c) Stereo Fusion

Figure 7.27. Uniform shadings of the Leaf surface. The proposed method outperforms the reference 3D scanner and yields an accurate and smooth reconstruction of the leaf structure.

7.2.5 Summary

A new approach to fuse stereo image analysis and photometric stereo has been described in this section. The differential equations building on Horn's integrability term are converted into an iterative update scheme that refines the initial depth estimation derived from the stereo correspondence by adjusting the initial disparity map to match the gradient fields obtained from both cameras. Figure 7.28 displays the cumulative distribution functions of the absolute pixelwise errors between final result and reference data for all datasets. Especially the Leaf and Plaster dataset exhibit high similarity to the 3D scanner with more than 96 % of the data points closer than 1 mm to the reference. The Ram and Ganesh dataset demonstrate that a good estimate of the gradient field is crucial to achieve a significant refinement of the initial guess. The high amount of self-cast shadows caused by the large discontinuities of the depth data leads to rough estimates of the gradient fields and the photometric stereo benefits from a shadow detection pre-processing step. This appears to be contradicted by Fig. 7.28 regarding the Ram dataset, since the reconstruction based on the uniform BRDF but without shadow masking (orange) yields a cumulative distribution function (CDF) that is closer to the optimum. However, the reference data of the 3D scanner only covers a part of the reconstructed surface. The over-smoothing and distortion of the snout area that are visible in Fig. 7.18e are accounted for while the spikes in Fig. 7.18d are not.

Shadowing effects are of minor importance on the Plaster and Leaf dataset, since their depth varies somewhat on a smaller scale and, more important, continuously without large steps. The spatially varying reflectance properties of the Plaster require a robust parameter estimation to compute correct surface normal directions successfully. This is achieved except for the lower left slope where shadowing affects the input data. The scratch mark is nonetheless reconstructed by the stereo fusion although it is not visible in the initial stereo estimate but in the PS gradient fields only. In contrast to the reference scanning system, the image-based methods do achieve a dense estimation of the 3D data. The proposed algorithm outperforms the reference 3D scanning system on the Leaf dataset whose metallic surface poses an obstacle for accurate reconstruction of most active range scanning techniques. The surface reconstructed by the 3D scanner yields a high level of noise whereas the proposed method achieves a detailed and smooth reconstruction of the metallic object.

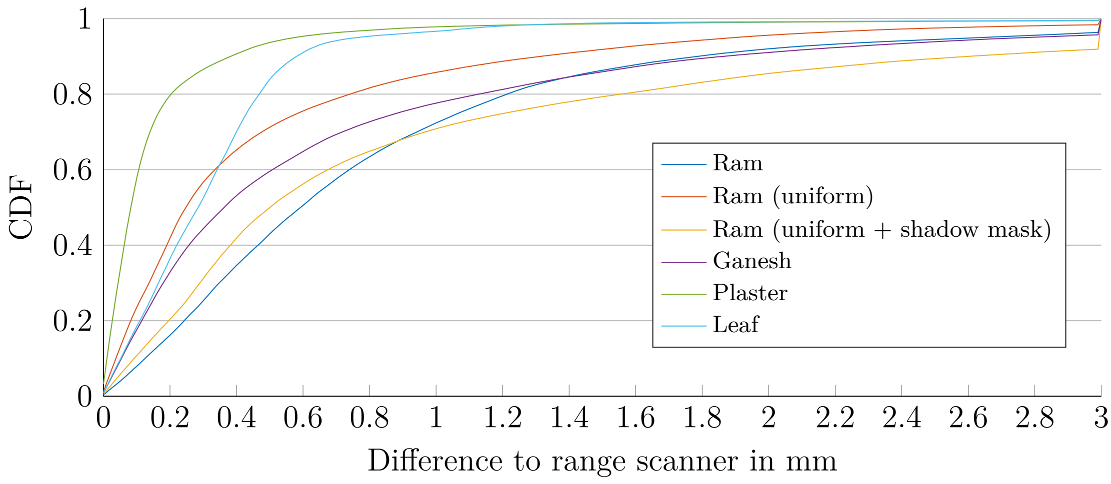


Figure 7.28. Cumulative distribution of pixelwise errors of all datasets.

Limitations

The optimization relies on the refinement of a disparity map based on corresponding gradient fields. Thus, a good estimate of the gradient field is essential to recover a meaningful disparity map. Furthermore, the iterative update scheme needs to be initialized at a depth value “close to the true value” in order not to diverge. In the case of the Leaf dataset in Fig. 7.27 it is evident that the incorrect initialization biases the lower left edge of the reconstruction and the algorithm is not able to converge to a reasonable solution. In fact, the smoothing effect of the integrability error affects surrounding parts of the surface since the step in the depth data is extremely large. A second limitation is apparent in Fig. 7.21 showing the Ganesh surface where the boundaries of the surface that are provided by the gradient data are incorrect, and thus the edges diverge. However, edges are boundary conditions to the differential equations and as such, in general, vulnerable to divergence if not constrained correctly.

SUMMARY AND CONCLUSION

8

Various research topics regarding the goal of high-accuracy image-based 3D surface reconstruction have been tackled throughout the previous pages. Chapter 5 illustrated the usability of Bayesian inference to estimate unknown BRDF model parameters on a surface that yields patchwise arbitrary parameters. The introduction of a new two-dimensional reversible jump procedure allows for the exploration of the model space and the derivation of the correct surface patches if there is a sufficient number of versatile (regarding illumination conditions) input data available. A huge advantage of the Monte Carlo approach is the possibility to directly infer statistics from the data, giving insights into confidence levels without the need for additional model assumptions. On the other hand, multiple evaluations of the model subject to many different parameter constellations require a considerable amount of processing time, especially since the linearity of the Markov chain makes parallelization of a single RJMCMC run tedious. However, research has been conducted regarding the possibilities of efficient parallelization and improved convergence of the model selection (e.g., Ye et al., 2009; Farr et al., 2015) and as such there is room for further exploration of the proof-of-concept that was presented in this thesis. Computing multiple Markov chains in parallel is of course very well manageable and augments the data base that can be sampled from. Yet, every chain needs to converge on its own to the unknown posterior distribution and this still takes a certain amount of time.

Building on previous work that has been undertaken in the author's Bachelor thesis at the Image Analysis Group at TU Dortmund University, a new procedure to calibrate light source positions for a two-camera photometric stereo is presented in Chapter 6. A spherical object that is considered as ideal diffuse at first is located at multiple positions in a calibrated stereo camera rig and the 3D coordinates are used to invert Lambert's law and derive a unique incident light direction per sphere position. These directions that should ideally intersect in one point (the exact light position) are instead a bundle of skewed lines, and the point in 3D space that is closest to all these lines is found from a linear system of equations. From these points in the world coordinate system, a global optimization is initialized that minimizes the difference between measured images and rendered spheres subject to the Phong BRDF model. Considering a small specular lobe increased the similarity of renderings and images. Using a (mostly) diffuse sphere in contrast to highly specular spheres that are commonly used for light calibration yields the advantages that the dataset from which the positions are derived is significantly more substantial compared to a few pixels that mark the specular highlight. Additionally, the light source strength can be estimated since it is part of the BRDF model and the optimization.

Considering now a fully calibrated laboratory, Chapter 7 finally provides two procedures that contribute to the task of retrieving 3D information from image data. An initial estimate of the depth data obtained from a pseudo-diffuse surface is used in Section 7.1 to estimate the spatially varying parameters of an analytical BRDF model. A pyramid subsampling scheme then iteratively refines this initial estimate of the depth data with the photometric stereo gradient fields obtained from the input images. The algorithm has been evaluated on synthetic and laboratory data and produced compelling reconstructions of the surfaces. Fine details have been preserved and reconstructed in the surfaces while exhibiting a less noisy appearance compared to the reference depth scanner. However, since absolute depth information cannot be obtained from a single camera, the 3D information is relative and scalable.

The stereo fusion algorithm illustrated in Section 7.2 extends the integrability regularization to solve shape from shading (Horn, 1990) with a disparity constraint that accounts for the similarity between the gradient fields obtained from left and right stereo camera view subject to an initial disparity map. Since depth and disparity are inversely proportional, an update of the depth data changes the disparity and vice-versa. Although the optimization requires an initial guess that is relatively close to the correct surface, the procedure achieves a refinement of the stereo depth data that maintains details that are visible in the gradient fields while reducing the artificial roughness of the stereo estimate. The quality of the final result is of course constrained by the level of detail contained in the photometric stereo gradient field. Using only 18 light sources and two cameras, the stereo fusion can compute 3D data that are dense and come close on absolute scale to the reference range scanning system. The small-scale accuracy of the 3D scanner is not matched on the Plaster, Ganesh, and Ram datasets but in contrast to the 3D scanner, depth information can be estimated for every point of the surface that is visible in the left camera. The metallic dataset poses a challenge to the reference system that produces noise-afflicted data while the proposed method yields a smooth and detailed reconstruction of the leaf. The camera-based reconstruction outperforms the competing fringe projection scanner considerably.

Future Work

In the description of the the camera calibration in Section 4.3 it was noticed that, in a series of repeatedly acquired images, the statistical distributions of image pixels that are shadowed and those that are directly lighted differ such that only the latter pass the Kolmogorov-Smirnov test for normality. This property appears interesting and may be useful for the detection of shadowed image points in single gray value images. Shadow detection has been conducted based on RANSAC and outlier rejection in multiple images (e.g., Mukaigawa et al., 2007), or using color and texture cues (e.g., Xu et al., 2006; McFeely et al., 2012), yet single gray value images still pose a situation that is difficult to handle.

A second possible aspect of continuing the presented work is the parallelization of the RJMCMC procedure. The execution on a dataset of 12 synthetic images of 121 × 121 pixels with 20 000 states of the Markov chain takes about 14 hours on an Intel Core i7-3770 machine. This computational time naturally prohibits application in any productive operation. Admittedly, there has no effort been made regarding speed-up of the code and using C++ instead of Matlab might already provide a significant improvement. Still, parallelization of the Markov chain Monte Carlo algorithm is a crucial requisite if this powerful technique should be used in industrial applications.

Concerning the stereo fusion algorithm, there is still the need for a comprehensive analysis of the convergence properties of the algorithm. As has been illustrated on the Leaf dataset, the algorithm did not converge to the correct solution after being falsely initialized with the stereo map. Additionally, the gradient fields of the Ganesh and Ram datasets provided more detailed information than has been visible in the final result of the algorithm. Thus, the algorithm probably terminated in a local minimum. Furthermore, the algorithm as it is depicted in this thesis relies exclusively on the left camera view to compute the integrability error. Since depth data and gradient for the right camera are available, augmenting the integrability error to the second camera might increase stability in the case of falsely estimated gradients for only one of the two camera views. Another possibility would be to introduce a combined integrability error that again considers the gradient of the right camera view and the disparity offset defined by the current depth estimate. However, the latter option would probably only have a small effect on the convergence since the transformation between left and right camera is already covered by in the disparity error term.

BIBLIOGRAPHY

Ackermann, J., Langguth, F., Fuhrmann, S., Kuijper, A., and Goesele, M. (2014). Multi-view photometric stereo by example. In *Proc. Int. Conf. on 3D Vision*, pages 259–266.

Ackermann, J., Ritz, M., Stork, A., and Goesele, M. (2012). Removing the example from example-based photometric stereo. In Kutulakos, K. N., editor, *Trends and Topics in Computer Vision*, pages 197–210. Springer.

Aittala, M., Weyrich, T., and Lehtinen, J. (2013). Practical SVBRDF capture in the frequency domain. *ACM Transactions on Graphics*, 32(4):110:1–110:12.

Alldrin, N., Zickler, T., and Kriegman, D. (2008). Photometric stereo with non-parametric and spatially-varying reflectance. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Alldrin, N. G. and Kriegman, D. J. (2007). Toward reconstructing surfaces with arbitrary isotropic reflectance : A stratified photometric stereo approach. In *2007 IEEE International Conference on Computer Vision*, pages 1–8.

Amir, Y. M. and Thörnberg, B. (2017). High precision laser scanning of metallic surfaces. *International Journal of Optics*, 2017:1–13.

Angelopoulou, E. (1999). The reflectance spectrum of human skin. Technical report, University of Pennsylvania Department of Computer and Information Science Technical.

Ashikhmin, M. and Shirley, P. (2000). An anisotropic phong light reflection model. *Journal of Graphics Tools*, 2:25–32.

Band, Y. B. (2006). *Light and Matter.* John Wiley & Sons Inc.

Barsky, S. and Petrou, M. (2003). The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252.

Bay, H., Tuytelaars, T., and Gool, L. V. (2006). SURF: Speeded up robust features. In *Proc. European Conf. on Computer Vision*, volume 3951 of *Lecture Notes in Computer Science*, pages 404–417.

Belhumeur, P. N. and Kriegman, D. J. (1998). What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 3:1–16.

Belhumeur, P. N., Kriegman, D. J., and Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):1040–1046.

Bennett, J. M. (1995). Polarization. In Bass, M., van Stryland, E. W., Williams, D. R., and Wolfe, W. L., editors, *Handbook of Optics Volume 1: Fundamentals, Techniques, & Design*. McGraw-Hill, Inc., 2nd edition.

Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. *ACM SIGGRAPH*, 11(2):192–198.

Bouguet, J.-Y. (2008). *Camera Calibration Toolbox for Matlab,* online *http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.*

Bronstein, I. N., Semendjajew, K., Musiol, G., and Mühlig, H. (2016). *Taschenbuch der Mathematik*, volume 10. Europa-Lehrmittel.

Byrd, R. H., Gilbert, J. C., and Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185.

Chandraker, M., Agarwal, S., and Kriegman, D. (2007). ShadowCuts: Photometric Stereo with Shadows. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Chandraker, M. K., Kahl, F., and Kriegman, D. (2005). Reflections on the generalized bas-relief ambiguity. *Conference on Computer Vision and Pattern Recognition*, 1:788–795.

Chen, T., Lensch, H. P. A., Fuchs, C., and Seidel, H. P. (2007). Polarization and phase-shifting for 3d scanning of translucent objects. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Chung, H.-S. and Jia, J. (2008). Efficient photometric stereo on glossy surfaces with wide specular lobes. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Cohen, M. F., Wallace, J., and Hanrahan, P. (1993). *Radiosity and Realistic Image Synthesis.* Academic Press Professional, Inc.

Coleman, Jr., E. N. and Jain, R. (1982). Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18:309–328.

Coleman, T. and Li, Y. (1994). On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds. *Mathematical Programming*, 67(2):189–224.

Coleman, T. and Li, Y. (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6:418–445.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Cook, R. L. and Torrance, K. E. (1981). A reflectance model for computer graphics. *SIGGRAPH Computer Graphics*, 15(3):307–316.

Curless, B. and Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proc. Annual Conf. on Computer Graphics and Interactive Techniques*, pages 303–312. ACM Press.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. (2000). Acquiring the reflectance field of a human face. In *Proc. Annual Conf. on Computer Graphics and Interactive Techniques*, pages 145–156. ACM Press.

Debevec, P. E. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proc. Annual Conf. on Computer Graphics and Interactive Techniques*, pages 369–378. ACM Press.

Donner, C. and Jensen, H. W. (2006). A spectral BSSRDF for shading human skin. In *Proc. Eurographics Conf. on Rendering Techniques*, EGSR '06, pages 409–417. Eurographics Association.

Dorsey, J., Rushmeier, H., and Sillion, F. (2008). *Digital Modeling of Material Appearance.* Morgan Kaufmann.

Einarsson, P., Chabert, C.-F., Jones, A., Ma, W.-C., Lamond, B., Hawkins, T., Bolas, M., Sylwan, S., and Debevec, P. (2006). Relighting human locomotion with flowed reflectance fields. In *Proc. Eurographics Conf. on Rendering Techniques*, EGSR '06, pages 183–194. Eurographics Association.

Erb, W. (1980). Computer-controlled gonioreflectometer for the measurement of spectral reflection characteristics. *Applied Optics*, 19(22):3789–3794.

Farr, W. M., Mandel, I., and Stevens, D. (2015). An efficient interpolation technique for jump proposals in reversible-jump markov chain monte carlo calculations. *Royal Society Open Science*, 2:1–12.

Fei-Fei, L. and Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 524–531 vol. 2. IEEE Computer Society.

Filip, J., Vávra, R., and Havlíček, M. (2014). Effective acquisition of dense anisotropic brdf. In *Proc. Int. Conf. on Pattern Recognition*, pages 2047–2052.

Finlayson, G. D., Hordley, S. D., and Drew, M. S. (2002). Removing shadows from images. In *Proc. European Conf. on Computer Vision – Part IV*, ECCV '02, pages 823–836. Springer.

Freeman, W. T. (1996). Exploiting the generic viewpoint assumption. *International Journal of Computer Vision*, 20(3):243–261.

Fua, P. and Leclerc, Y. G. (1995). Object-centred surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56.

Gelman, A. (1995). Inference and monitoring convergence. In Gilks, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, pages 131–143. Chapman & Hall.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Sciences*, 7(4):457–472.

Georghiades, A. S. (2003). Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. IEEE Int. Conf. on Computer Vision*, volume 2, pages 816–823. IEEE Computer Society.

Giesen, F. (2009). Phong and Blinn-Phong Normalization Factors. online *http://www.farbrausch.de/~fg/stuff/phong.pdf* accessed 25.10.2013, pages 1–2.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Goldman, D. B., Curless, B., Hertzmann, A., and Seitz, S. (2005). Shape and spatially-varying BRDFs from photometric stereo. *Proccedings of the 10th International Conference on Computer Vision*, pages 341–348.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Grumpe, A., Herbort, S., and Wöhler, C. (2011). 3D reconstruction of non-lambertian surfaces with non-uniform reflectance parameters by fusion of photometrically estimated surface normal data with active range scanner data. In Luhmann, T. and Müller, C., editors, *Photogrammetrie, Laserscanning, Optische 3D Messtechnik - Beiträge der Oldenburger 3D Tage 2011*, volume 10, pages 54–61.

Grumpe, A. and Wöhler, C. (2014). Recovery of elevation from estimated gradient fields constrained by digital elevation maps of lower lateral resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, 94:37–54.

Guo, R., Dai, Q., and Hoiem, D. (2011). Single-image shadow detection and removal using paired regions. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2033–2040. IEEE Computer Society.

Hamzah, R. A., Rahim, R. A., and Noh, Z. M. (2010). Sum of absolute differences algorithm in stereo correspondence problem for stereo matching in computer vision application. *Proc. Int. Conf. on Computer Science and Information Technology*, 1:652–657.

Han, T.-Q. and Shen, H.-L. (2015). Photometric stereo for general BRDFs via reflection sparsity modeling. *IEEE Transactions on Image Processing*, 24(12):4888–4903.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109.

Hayakawa, H. (1994). Photometric stereo under a light source with arbitrary motion. *Journal of Optical Society of America A*, 11:3079–3089.

He, X. D., Torrance, K. E., Sillion, F. X., and Greenberg, D. P. (1991). A comprehensive physical model for light reflection. *SIGGRAPH Computer Graphics*, 25(4):175–186.

Heikkila, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1106–1112. IEEE Computer Society.

Herbort, S., Grumpe, A., and Wöhler, C. (2011). Reconstruction of non-lambertian surfaces by fusion of shape from shading and active range scanning. *International Conference on Image Processing*, pages 1–4.

Herbort, S., Schugk, D., and Wöhler, C. (2013). 3D reconstruction of interreflection-affected surface concavities using photometric stereo. *Proc. Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 1–7.

Herbort, S. and Wöhler, C. (2011). An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3):1–17.

Herbort, S. and Wöhler, C. (2012). Self-consistent 3D surface reconstruction and reflectance model estimation of metallic surfaces. *Proc. Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 1–8.

Hernandez, C., Vogiatzis, G., and Cipolla, R. (2008). Shadows in three-source photometric stereo. In *Proc. European Conference on Computer Vision*, pages 290–303.

Hertzmann, A. and Seitz, S. M. (2003). Shape and materials by example: A photometric stereo approach. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 533–540. IEEE Computer Society.

Hertzmann, A. and Seitz, S. M. (2005). Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264.

Hiary, H., Zaghloul, R., and Al-Zoubi, M. B. (2018). Single-image shadow detection using quaternion cues. *The Computer Journal*, 61(3):459–468.

Hirschmüller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 807–814. IEEE Computer Society.

Holroyd, M., Lawrence, J., and Zickler, T. (2010). A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics*, 29(4):99:1–99:12.

Horn, B. K. P. (1970). Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report 232, Messachusets Institute of Technology.

Horn, B. K. P. (1977). Understanding image intensities. *Artificial Intelligence*, 11(2):201–231.

Horn, B. K. P. (1986). *Robot Vision*. MIT Electrical Engineering and Computer Science. The MIT Press.

Horn, B. K. P. (1989). Height and gradient from shading. Technical Report 1105A, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Horn, B. K. P. (1990). Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75.

Horn, B. K. P. and Brooks, M. J. (1986). The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 33:174–208.

Horn, B. K. P., Woodham, R. J., and Silver, W. M. (1978). Determining shape and reflectance using multiple images. Technical Report 490, Messachusets Institute of Technology.

Hui, Z. and Sankaranarayanan, A. C. (2017). Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2060–2073.

Ikehata, S., Wipf, D., Matsushita, Y., and Aizawa, K. (2014). Photometric stereo using sparse bayesian regression for general diffuse surfaces. *Transactions on Pattern Analysis and Machine Intelligence*, 36(9):1816–1831.

Ikeuchi, K. and Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184.

Jankowski, D. G. (1997). Photoclinometry. In Shirley, J. and Fairbridge, R. W., editors, *Encyclopedia of Planetary Science*, pages 576–576. Springer.

Khan, S. H., Bennamoun, M., Sohel, F., and Togneri, R. (2016). Automatic shadow detection and removal from a single image. *Transactions on Pattern Analysis and Machine Intelligence*, 38(3):431–446.

Kielhöfer, H. (2010). *Variationsrechnung*. Vieweg + Teubner, 1st edition.

Kimmel, R. and Sethian, J. A. (2001). Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14:237–244.

King, R., Morgan, B. J., Gimenez, O., and Brooks, S. P. (2010). *Bayesian Analysis for Population Ecology*. CRC Press.

Kingslake, R. (1992). *Optics in Photography*, volume 6. SPIE Publications.

Krause, J., Gebru, T., Deng, J., Li, L. J., and Fei-Fei, L. (2014). Learning features and parts for fine-grained recognition. In *Proc. Int. Conf. on Pattern Recognition*, pages 26–33.

Lafortune, E. P. and Willems, Y. D. (1994). Using the modifed phong reflectance model for physically based rendering. Technical report, Department of Computing Science, K. U. Leuven.

Lafortune, E. P. F., Foo, S.-C., Torrance, K. E., and Greenberg, D. P. (1997). Non-linear approximation of reflectance functions. In *Proc. Conf. on Computer Graphics and Interactive Techniques*, pages 117–126. ACM Press.

Lambert, J.-H. (1760). *Photometria, sive de mensura et gradibus luminis, colorum et umbrae*. Vidae Eberhardi Klett.

Langguth, F., Sunkavalli, K., Hadap, S., and Goesele, M. (2016). Shading-aware multi-view stereo. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Proc. European Conf. on Computer Vision*, pages 469–485. Springer International Publishing.

Laskey, K. and Myers, J. (2003). Population markov chain monte carlo. *Machine Learning*, 50:175–196.

Lawrence, J., Ben-Artzi, A., DeCoro, C., Matusik, W., Pfister, H., Ramamoorthi, R., and Rusinkiewicz, S. (2006). Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics*, 25(3):735–745.

Leclerc, Y. G., Luong, Q.-T., and Fua, P. (2000). Characterizing the performance of multiple-image point-correspondence algorithms using self-consistency. In *Proc. Int. Workshop on Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 37–52. Springer.

Lenoch, M. (2011). Analyse der Genauigkeit gängiger Beleuchtungs-Kalibrierverfahren und Untersuchung der Auswirkung von Kalibrierungsungenauigkeiten auf 3D Rekonstruktionsalgorithmen. *Bachelor Thesis, TU Dortmund University.*

Lenoch, M., Biebrach, P., Grumpe, A., and Wöhler, C. (2017). Multi-view shape from shading constrained by stereo image analysis. In Liu, M., Chen, H., and Vincze, M., editors, *Proc. Int. Conf. on Computer Vision Systems*, volume 10528 of *Lecture Notes in Computer Science*, pages 325–335. Springer International Publishing.

Lenoch, M., Herbort, S., Grumpe, A., and Wöhler, C. (2014). Linear unmixing in BRDF reproduction and 3D shape recovery. *Proc. Int. Conf. Pattern Recognition.*

Lenoch, M., Herbort, S., and Wöhler, C. (2012). Robust and accurate light source calibration using a diffuse spherical calibration object. In Luhmann, T. and Müller, C., editors, *Photogrammetrie, Laserscanning, Optische 3D Messtechnik - Beiträge der Oldenburger 3D Tage 2012*, pages 212–219. Wichmann.

Lenoch, M., Mistry, H., and Wöhler, C. (2015). Photometric surface reconstruction of surfaces with non-uniform reflectance properties. In Luhmann, T. and Müller, C., editors, *Photogrammetrie, Laserscanning, Optische 3D Messtechnik - Beiträge der Oldenburger 3D Tage 2015*, pages 228–236. Wichmann.

Lenoch, M. and Wöhler, C. (2016). Reflectance-based 3D shape refinement of surfaces with spatially varying BRDF properties. *9th IAPR Workshop on Pattern Recogniton in Remote Sensing*, pages 1–6.

Lenoch, M., Wilhelm, T., and Wöhler, C. (2016). Simultaneous surface segmentation and BRDF estimation via Bayesian methods. *Proc. Int. Conf. on Computer Vision Theory and Applications*, pages 39–48.

Lensch, H. P. A., Kautz, J., Goesele, M., Heidrich, W., and Seidel, H.-P. (2003). Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*, 22(2):234–257.

Lewis, J. P. (1995). Fast normalized cross-correlation. Technical report, Industrial Light & Magic.

Lewis, R. R. (1994). Making shaders more physically plausible. *Fourth Eurographics Workshop on Rendering*, pages 47–62.

Li, L. and Ng, C. S.-L. (2009). Rendering human skin using a multi-layer reflection model. *International Journal of Mathematics and Computers in Simulation*, 3:44–53.

Liberzon, D. (2012). *Calculus of Variations and Optimal Control Theory.* Princeton Univ. Press.

Lourakis, M. I. A. and Argyros, A. A. (2009). Sba: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, 36(1):2:1–2:30.

Louw, M. and Nicolls, F. (2010). Surface classification via BRDF parameters, using population monte carlo for mrf parameter estimation. *Proc. IASTED Int. Conf. on Computer Graphics and Imaging*, 11:145–154.

Lowe, D. G. (2004). Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image. *US Patent 6711293.* Washington, DC: U.S. Patent and Trademark Office.

Lumenera Coorporation (2008). *Lu160/Lu165 data sheet.*

Ma, W.-C., Hawkins, T., Peers, P., Chabert, C.-F., Weiss, M., and Debevec, P. (2007). Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proc. 18th Eurographics Conf. on Rendering Techniques*, pages 183–194. Eurographics Association.

Mallick, S. P., Zickler, T., Kriegman, D. J., and Belhumeur, P. N. (2005). Beyond lambert: Reconstructing specular surfaces using color. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 619–626. IEEE Computer Society.

Marschner, S., Lafortune, E. P. F., Westin, S. H., Torrance, K. E., and Greenberg, D. P. (1999a). Image-based BRDF measurement. *Applied Optics*, 39(16).

Marschner, S., Westin, S. H., Lafortune, E. P. F., Torrance, K. E., and Greenberg, D. P. (1999b). Image-based BRDF measurement including human skin. In *Proc. of 10th Eurographics Workshop on Rendering*, pages 139–152.

Marsland, S. (2009). *Machine Learning: An Algorithmic Perspective.* Chapman & Hall/CRC Machine.

Martin, D. R., Fowlkes, C. C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549.

Matusik, W., Pfister, H., Brand, M., and McMillan, L. (2003a). A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769.

Matusik, W., Pfister, H., Brand, M., and McMillan, L. (2003b). Efficient isotropic BRDF measurement. *ACM Eurographics Workshop on Rendering*, 44:241–247.

Maurer, D., Ju, Y. C., Breuß, M., and Bruhn, A. (2018). Combining shape from shading and stereo: A joint variational method for estimating depth, illumination and albedo. *International Journal of Computer Vision.*

McCann, J. (2016). Retinex theory. In *Encyclopedia of Color Science and Technology*, pages 1118–1125. Springer New York, New York, NY.

McFeely, R., Glavin, M., and Jones, E. (2012). Shadow identification for digital imagery using colour and texture cues. *IET Image Processing*, 6(2):148–159.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* John Wiley & Sons, Inc.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Moré, J. J. (1978). The levenberg-marquardt algorithm: Implementation and theory. In Watson, G., editor, *Lecture Notes in Mathematics - Numerical Analysis*, volume 630, pages 105–116. Springer, Berlin, Heidelberg.

Mukaigawa, Y., Ishii, Y., and Shakunaga, T. (2007). Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 24(10):3326–3334.

Mukaigawa, Y., Miyaki, H., Mihashi, S., and Shakunaga, T. (2001). Photometric image-based rendering for image generation in arbitrary illumination. *Proc. Int. Conf. on Computer Vision*, 2:652–659.

Nayar, S. K., Fang, X.-S., and Boult, T. (1997). Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186.

Nayar, S. K., Ikeuchi, K., and Kanade, T. (1990). Shape from interreflections. Technical Report CMU-RI-TR-90-14, Carnegie-Mellon University of Pittsburgh, PA, Robotics Institute.

Nehab, D., Rusinkiewicz, S., Davis, J., and Ramamoorthi, R. (2005). Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24(3):536–543.

Nicodemus, F., Richmond, J., Hsia, J., Ginsberg, I., and Limperis, T. (1977). Geometrical considerations and nomenclature for reflectance. Technical report, U.S. Department of Commerce, National Bureau of Standards.

Nicodemus, F. E. (1965). Directional reflectance and emissivity of an opaque surface. *Applied Optics*, 4(7):767–775.

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS.* John Wiley & Sons.

Oswald, M. R., Töppe, E., Kolev, K., and Cremers, D. (2009). Non-parametric single view reconstruction of curved objects using convex optimization. In *Proc. 31st DAGM Symposium on Pattern Recognition*, pages 171–180. Springer.

Peers, P., Hawkins, T., and Debevec, P. (2006). A reflective light stage. Technical Report ICT-TR-04.2006, USC Institute for Creative Technologies.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406—-413.

Pharr, M. and Humphreys, G. (2010). *Physically Based Rendering - From Theory to Implementation.* Morgan Kaufmann (Elsvier).

Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317.

Pichler, O. and Hosticka, A. T. B. J. (1995). A multichannel algorithm for image segmentation with iterative feedback. In *Proc. Inf. Conf. on Image Processing and Its Application*, pages 510–513.

Quéau, Y., Wu, T., Lauze, F., Durou, J.-D., and Cremers, D. (2017). A non-convex variational approach to photometric stereo under inaccurate lighting. In *Proc. IEEE Conf. on Computer Vision Pattern Recognition*, pages 350–359.

Quéau, Y., Wu, T., and Cremers, D. (2017). Semi-calibrated near-light photometric stereo. In Lauze, F., Dong, Y., and Dahl, A. B., editors, *Proc. Inf. Conf. on Scale Space and Variational Methods in Computer Vision*, pages 656–668. Springer.

Rindfleisch, T. (1965). A photometric method for deriving lunar topography information. Technical report, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

Rindfleisch, T. (1966). Photometric method for lunar topography. *Photometric Engineering*, 32(2):262–277.

Rohith, M., Sorensen, S., Rhein, S., and Kambhamettu, C. (2013). Shape from stereo and shading by gradient constrained interpolation. In *Proc. IEEE Int. Conf. on Image Processing*, pages 2232–2236.

Sachs, L. and Hedderich, J. (2006). *Statistik: Angewandte Statistik.* Springer, Berlin / Heidelberg, 12 edition.

Samaras, D. and Metaxas, D. (1998). Incorporating illumination constraints in deformable models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 322–329.

Samaras, D. and Metaxas, D. (1999). Coupled lighting direction and shape estimation from single images. In *Proc. IEEE Int. Conf. Computer Vision*, pages 868–874.

Samaras, D., Metaxas, D., Fua, P., and Leclerc, Y. G. (2000). Variable albedo surface reconstruction from stereo and shape from shading. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 480–487.

Sato, Y., Wheeler, M. D., and Ikeuchi, K. (1997). Object shape and reflectance modeling from observation. In *Proc. 24th Annual Conf. on Computer Graphics and Interactive Techniques*, pages 379–387. ACM Press.

Scharstein, D. and Szeliski, R. (2001). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42.

Schulz, S., Grimm, M., and Grigat, R.-R. (2007). Using brightness histogram to perform optimum auto exposure. *WSEAS Transactions on Systems and Control*, 2(2):93–100.

Seber, G. A. F. (1984). *Multivariate Observations.* John Wiley & Sons, Inc.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 519–528. IEEE Computer Society.

Semerjian, B. (2014). A new variational framework for multiview surface reconstruction. In *Proc. European Conf. on Computer Vision*, volume 8694 of *Lecture Notes in Computer Science*, pages 719–734.

Seoul Semiconductor (2013). *Z-Power LED data sheet, Rev. 11.*

Shafer, S. A. (1985). Using color to separate reflection components. In Healey, G. E., Shafer, S. A., and Wolff, L. B., editors, *Color - Research and Application*, pages 43–51. Jones and Bartlett Publishers, Inc.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 8:888–905.

Shimizu, M. and Okutomi, M. (2001). Precise sub-pixel estimation on area-based matching. In *Proc. IEEE Int. Conf. on Computer Vision*, volume 1, pages 90–97.

Silver, W. M. (1980). Determining shape and reflectance using multiple images. Master's thesis, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory (M.I.T. A.I. Laboratory).

Soldado, R. and Ureña, C. (2012). An overview of BRDF models. Technical report, Dept. Lenguajes y Sistemas Informáticos, University of Granada, Granada, Spain.

Spath, H. (1985). *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples.* Halsted Press.

Stein, A. N., Huertas, A., and Matthies, L. (2006). Attenuating stereo pixel-locking via affine window adaptation. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 914–921.

Sunkavalli, K., Zickler, T., and Pfister, H. (2010). Visibility subspaces: Uncalibrated photometric stereo with shadows. In *Proc. European Conf. on Computer Vision*, volume 6312 of *Lecture Notes in Computer Science*, pages 251–264. Springer.

Tola, E., Lepetit, V., and Fua, P. (2008). A fast local descriptor for dense matching. In *Proc. Conf. on Computer Vision and Pattern Recognition.* IEEE Computer Society.

Tola, E., Lepetit, V., and Fua, P. (2010). DAISY: An efficient dense descriptor applied to wide baseline stereo. *Transactions on Pattern Analysis and Machine Intelligence*, 32(5):815–830.

Torrance, K. E. and Sparrow, E. M. (1967). Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America A*, 57(9):1105–1114.

ViALUX Messtechnik + Bildverarbeitung GmbH (2010). *zSnapper Manual, Version 3.0.*

Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., and Matusik, W. (2009). Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics*, 28(5):174:1–174:11.

von Helmholtz, H. (1924). *Handbuch der Physiologischen Optik.* Optical Society of America.

Waltz, R., Morales, J., Nocedal, J., and Orban, D. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3):391–408.

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *SIGGRAPH Computer Graphics*, 26(2):265–272.

Wöhler, C. (2013). *3D Computer Vision - Efficient Methods and Applications*. Springer London, 2nd edition.

Wolff, L. B. (1989). Using polarization to separate reflection components. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 363–369. IEEE Computer Society.

Wolff, L. B. (1996). Generalizing lambert's law for smooth surfaces. In Buxton, B. and Cipolla, R., editors, *Proc. European Conf. on Computer Vision*, Lecture Notes in Computer Science, pages 40–53. Springer Berlin Heidelberg.

Wolff, L. B. and Boult, T. E. (1991). Constraining object features using a polarization reflectance model. *Transactions on Pattern Analysis and Machine Intelligence*, 13(7):635–657.

Woodham, R. J. (1977). *Reflectance map techniques for analyzing surface defects in metal castings*. PhD thesis.

Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144.

Woodham, R. J. (1994). Gradient and curvature from the photometric-stereo method, including local confidence estimation. *Journal of the Optical Society of America A*, 11(11):3050–3068.

Wu, C., Wilburn, B., Matsushita, Y., and Theobalt, C. (2011a). High-quality shape from multi-view stereo and shading under general illumination. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 969–976. IEEE Computer Society.

Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., and Ma, Y. (2011b). Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. Asian Conf. on Computer Vision*, volume 6494 of *Lecture Notes in Computer Science*, pages 703–717.

Xu, L., Qi, F., and Jiang, R. (2006). Shadow removal from a single image. In *Proc. Int. Conf. on Intelligent Systems Design and Applications*, volume 2, pages 1049–1054.

Ye, J., Wallace, A., and Thompson, J. (2009). Parallel markov chain monte carlo computation for varying-dimension signal analysis. In *Proc. European Signal Processing Conference*, pages 2673–2677.

Yu, Y. and Chang, J. T. (2005). Shadow graphs and 3d texture reconstruction. *International Journal of Computer Vision*, 62(1):35–60.

Zhang, L., Curless, B., Hertzmann, A., and Seitz, S. M. (2003). Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *International Conference on Computer Vision*, 1:618–626.

Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152.

Zhou, W. and Kambhamettu, C. (2002). Estimation of illuminant direction and intensity of multiple light sources. In *Proc. European Conf. on Computer Vision*, volume 1, pages 206–220.

Zhou, Z., Wu, Z., and Tan, P. (2013). Multi-view photometric stereo with spatially varying isotropic materials. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 1482–1489. IEEE Computer Society.

Zhu, J., Samuel, K. G. G., Masood, S. Z., and Tappen, M. F. (2010). Learning to recognize shadows in monochromatic natural images. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 223–230. IEEE Computer Society.

Zickler, T., Belhumeur, P. N., and Kriegman, D. J. (2002). Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *Proc. European Conf. on Computer Vision*, volume 3, pages 869–884.

List of Symbols

A	surface area in metric units
$\mathbf{A}$	camera matrix containing the intrinsic camera parameters
b	principal distance between principal point and optical center
c	camera constant in pixel units
c_u	horizontal coordinate of principal point in sensor coordinate system
$\hat{c}_\mathrm{u}$	horizontal coordinate of principal point in image coordinate system
c_v	vertical coordinate of principal point in sensor coordinate system
$\hat{c}_\mathrm{v}$	vertical coordinate of principal point in image coordinate system
$\mathbf{c}$	optical center of a lens
$\mathbf{c}_0$	principal point of a lens
C	camera coordinate system
d	sub sampling factor in image resizing
d_u	horizontal component of camera constant in pixel units
d_v	vertical component of camera constant in pixel units
d_min	minimal distance between two clusters
D	distribution of the microfacets in the Cook-Torrance BRDF model
D_xx	eigenvalue corresponding to x-component of Lafortune coordinate transformation
D_yy	eigenvalue corresponding to y-component of Lafortune coordinate transformation
D_zz	eigenvalue corresponding to z-component of Lafortune coordinate transformation
$\mathbf{D}$	square matrix of eigenvalues corresponding to eigenvectors in $\mathbf{Q}$
e	Euler’s number

$\mathbf{e}$	vector of different exposure times in images acquisition process
E	irradiance is the radiant flux incident on (or emitted from) a surface area
E_{absDepth}	absolute depth error as defined by Grumpe and Wöhler (2014)
E_{int}	integrability error as defined by Horn (1989) and detailed in Section 7.2.3
E_{SF}	Stereo Fusion error as defined in Section 7.2.3
E_{δ}	disparity error as defined in Section 7.2.3
$\mathbf{E}$	pixelwise normalized error map between measured values and reflectance model
$\mathbf{E}_{\mathrm{RMSE}}$	pixelwise RMSE error map between measured values and reflectance model
f	focal length of a real lens
$\hat{f}$	focal length of a real lens in pixel units
f_{r}	Bidirectional Reflectance Distribution Function
F	Fresnel term of the Cook-Torrance BRDF model
F_0	Fresnel term of the Cook-Torrance BRDF model at normal incidence
G	geometric attenuation factor of the Cook-Torrance BRDF model
G_{s}	set of surface segments
$\mathbf{h}$	half angle vector between incident and viewing direction
I	image coordinate system
$\tilde{I}$	single intensity measurement
I_0	strength of a point light source
$\hat{I}_0$	strength of a point light source scaled by the inverse squared distance to the object
$\mathbf{I}$	matrix of measured intensity values that compose the digital image
$\mathbf{I}_{\mathrm{HDR}}$	HDR image composed of different images captured at different exposure times
$\bar{\mathbf{I}}$	image of pixelwise mean values computed across a series
$\hat{\mathbf{I}}$	image of pixelwise sum values computed for a series of images
$\hat{\mathbf{I}}_0$	diagonal matrix containing the scaled strength of different light sources
$\mathbb{I}$	identity matrix
j	imaginary unit
J	number of sphere positions used to compute a light source position

k_a	weight of ambient light component of a BRDF model
k_CT	weight of Cook-Torrance BRDF term
k_d	weight of diffuse reflection component of a BRDF model
k_s	weight of specular reflection component of a BRDF model
k_u	horizontal extent of a camera sensor pixel per unit length
$\hat{k}_\mathrm{u}$	horizontal extent of a camera sensor pixel per unit length projected on the object surface
k_v	vertical extent of a camera sensor pixel per unit length
$\hat{k}_\mathrm{v}$	vertical extent of a camera sensor pixel per unit length projected on the object surface
K	number of points in an image to be used for a calculation
l_x	x component of incident light direction
l'_x	x component of incident light direction after Lafortune coordinate transformation
l_y	y component of incident light direction
l'_y	y component of incident light direction after Lafortune coordinate transformation
l_z	z component of incident light direction
l'_z	z component of incident light direction after Lafortune coordinate transformation
$\mathbf{l}$	unit vector from surface point to light source
$\mathbf{l}'$	unit vector from surface point to light source after Lafortune coordinate transformation
L	radiance is the radiant flux emitted from a certain surface area in a solid angle
L_d	diffuse component of the radiance
$\hat{L}_\mathrm{d}$	effective diffuse component of the radiance composed of incident light strength and diffuse weight
L_s	specular component of the radiance
L_v	luminance is the luminous intensity of light travelling per unit area in a certain direction
$\mathbf{L}$	$N \times 3$ matrix of incident light directions per surface point
$\mathbf{L}^+$	pseudo inverse of a non-square matrix
m	root mean square of microfacet slope in Cook-Torrance BRDF model
$\tilde{m}$	model number of current iteration in reversible jump Markov chain Monte Carlo algorithm
m_p	slope of the horizontal image gradient
m_q	slope of the vertical image gradient

$\mathbf{M}$	rotation matrix scaled by the camera matrix
$\mathbf{M}_{\mathrm{D}}$	transformation matrix in reversible jump Markov chain Monte Carlo algorithm to change between dimensions
$\mathbf{M}_{\mathrm{L}}$	coordinate transformation matrix in Lafortune BRDF model
n_{x}	x component of surface normal direction
n_{y}	y component of surface normal direction
n_{z}	z component of surface normal direction
n	refractive index of a medium at an interface of two media a subscript i identifies the corresponding index
$\tilde{n}$	complex refractive index of a medium, comprising refractive index and extinction coefficient
$\mathbf{n}$	surface normal direction
$\mathbf{n}'$	normal direction of microfacet in Cook-Torrance BRDF model
N	number of images to be used in Photometric Stereo
N_{diff}	number of images to be used in pseudo diffuse Photometric Stereo
N_{Blinn}	normalization factor of Blinn BRDF model to comply with energy conservation
N_{Phong}	normalization factor of Phong BRDF model to comply with energy conservation
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution defined by mean μ and standard deviation σ
p	horizontal surface gradient
p_{c}	center pixel of mask to compute discrete gradient
p_{d}	lower pixel of discrete gradient mask with respect to center pixel p_{c}
p_{l}	left pixel of discrete gradient mask with respect to center pixel p_{c}
p_{r}	right pixel of discrete gradient mask with respect to center pixel p_{c}
p_{u}	upper pixel of discrete gradient mask with respect to center pixel p_{c}
p_x	derivative of the horizontal surface gradient with respect to x coordinate
p_{L}	horizontal surface gradient as seen by left stereo camera
p_{R}	horizontal surface gradient as seen by right stereo camera
$\mathbf{p}_4$	translation vector between two camera coordinate systems scaled by camera matrix
$P(\boldsymbol{\theta})$	a-priori probability distribution of BRDF parameters

$P(\mathbf{I})$	a-priori probability distribution of measured image data
$P(\boldsymbol{\theta}\|\mathbf{I})$	conditional probability of BRDF parameters given the measured image data – the posterior distribution
$P(\mathbf{I}\|\boldsymbol{\theta})$	conditional probability of measured image data given the current BRDF parameters – the likelihood
$P_{\mathrm{n}}(k)$	k-th coefficient in n-th row of Pascal's triangle
P_{p}	Poisson distribution of events per unit time
$\mathbf{P}$	projection matrix describing the transformation of a point in the scene onto the sensor in projective coordinates
$\mathbb{P}^i$	i-dimensional projective space
q	vertical surface gradient
q_y	derivative of the vertical surface gradient with respect to y coordinate
q_{L}	vertical surface gradient as seen by left stereo camera
q_{R}	vertical surface gradient as seen by right stereo camera
$q(\zeta)$	probability distribution of random number ζ
$q(\zeta')$	probability distribution of random number ζ' for proposed consecutive step of Markov chain
$\mathbf{Q}$	square matrix of eigenvectors corresponding to eigenvalues in $\mathbf{D}$
r	metric distance between light source position and surface segment
r_{s}	real-world radius of the calibration sphere in millimeters
$\hat{r}_{\mathrm{s}}$	radius of the calibration sphere on the images sensor in pixels
$\mathbf{r}$	unit vector pointing in the direction of ideal mirror like reflection
$\mathbf{R}$	matrix of synthetic pixel values based on a reflectance modeling technique and parameters of the laboratory setup
$\hat{\mathbf{R}}_i$	rotation matrix describing the rotation from i-th camera coordinate system to the world coordinate system
s_{x}	x component of light source position
s_{y}	y component of light source position
s_{z}	z component of light source position
$\mathbf{s}$	light source position in world coordinate system
S	sensor coordinate system
t	iteration index of Markov chain
t_{b}	metric absolute value of the baseline of a stereo camera system

$\mathbf{t}$	translation vector between two camera coordinate systems
$\mathbf{t}_\mathrm{b}$	baseline vector of a stereo camera system
T	number of evaluated states of the Markov chain
${}^{I}_{C}T$	transformation from image to camera coordinate system
${}^{C}_{W}T$	transformation from world to camera coordinate system
${}^{S}_{W}T$	transformation from world to sensor coordinate system
$\mathbf{T}$	image template to be used in correlation-based matching
u	horizontal sensor coordinate
$\hat{u}$	horizontal image coordinate
$\mathcal{U}(u_\mathrm{min}, u_\mathrm{max})$	uniform distribution ranging from u_min to u_max
U_g	set of horizontal image pixels that belong to surface patch g
v	vertical sensor coordinate
$\hat{v}$	vertical image coordinate
v_x	x component of viewing direction
v'_x	x component of viewing direction after Lafortune coordinate transformation
v_y	y component of viewing direction
v'_y	y component of viewing direction after Lafortune coordinate transformation
v_z	z component of viewing direction
v'_z	z component of viewing direction after Lafortune coordinate transformation
$\mathbf{v}$	unit vector from surface point to viewer
$\mathbf{v}'$	unit vector from surface point to viewer after Lafortune coordinate transformation
V_g	set of vertical image pixels that belong to surface patch g
W	world coordinate system
$\mathbf{W}$	matrix of weighting factors
$\mathbf{x}$	point in euclidean space
$\tilde{\mathbf{x}}$	point in projective space
$\mathbf{x}'$	x axis of local coordinate system after Lafortune coordinate transformation
$\mathbf{y}'$	y axis of local coordinate system after Lafortune coordinate transformation
$z(u, v)$	depth value corresponding to sensor pixels
$\hat{z}(u, v)$	approximate depth value used as constraint or initial value to next iteration

$z(x,y)$	depth value corresponding to 3D world coordinates
$\bar{z}$	mean distance between object and origin of coordinate system
z_x	partial derivative of depth map with respect to x coordinate
z_y	partial derivative of depth map with respect to y coordinate
$\mathbf{z}'$	z axis of local coordinate system after Lafortune coordinate transformation
α	skew angle between horizontal and vertical axis of camera sensor pixels
β	scaling of coordinates in projective space
γ	exponent of the cosine lobe of a BRDF model
γ_δ	weight of the disparity error in the Stereo Fusion term
$\Gamma(u,v)_{\mathrm{NCC}}$	normalized cross correlation coefficient at pixel (u,v)
$\Gamma(u,v)_{\mathrm{SAD}}$	sum of absolute differences at pixel (u,v)
δ	disparity between pixel coordinates corresponding to the same 3D scene point
Δ	Laplace operator
ε	probability per BRDF parameter to accept newly proposed values in Markov chain Monte Carlo algorithm
ε_{D}	probability to accept a change in dimension during reversible jump Markov chain Monte Carlo algorithm
ζ	random number drawn from distribution $q(\zeta)$
η	scaling uncertainty when computing Lambert's law with non-unit vectors
ϑ	elevation angle
$\vartheta_{\mathrm{crit}}$	critical angle for which larger incident angles lead to complete reflection of the light wave
ϑ_{h}	elevation angle defining the half vector
ϑ_{in}	elevation angle defining the incident light direction
ϑ_{out}	elevation angle defining the viewing direction
ϑ_{r}	elevation angle defining the direction of mirror-like reflection
ϑ_{t}	transmission angle at the interface of two media
$\boldsymbol{\theta}$	vector of parameters of a certain BRDF model
$\boldsymbol{\theta}'$	proposed BRDF parameters in following step of Markov chain
$\boldsymbol{\theta}_{\mathrm{s}}$	vector of Phong BRDF parameters of the diffuse calibration sphere
$\boldsymbol{\theta}_{\mathrm{G}}$	vector of Blinn BRDF parameters of surface segment g

κ	extinction coefficient accounting for the amount of absorbed light at an interface of two media
λ	wavelength of radiation
λ_{p}	rate of mean events per unit time of a Poisson distribution
μ	mean value of a random variable
$\tilde{\mu}$	estimated mean value of prior distribution for a certain BRDF model parameter
$\hat{\mu}$	estimated mean value of proposal distribution for a certain BRDF model parameter
ν	scaling of the reprojection of a point from sensor to world coordinates
ρ	albedo of a surface or material that describes the overall amount of reflected light
ρ_{r}	amount of reflected light at the interface of two media
ρ_{t}	amount of transmitted light at the interface of two media
σ	standard deviation of a random variable
σ_{M}	standard deviation of normally distributed measurements around the value predicted by a BRDF model
σ_{MS}	standard deviation of mean shift kernel
σ_{image}	standard deviation of additive image noise for synthetic data
σ_{depth}	standard deviation of additive noise for synthetic depth data
$\sigma_{\mathrm{absDepth}}$	width in pixels of a Gaussian low-pass filter
σ_{mask}	width in pixels of a weighted BRDF fitting mask
$\bar{\sigma}$	mean value across multiple standard deviations of different test series of the variable indicated by a subscript
$\tilde{\sigma}$	estimated standard deviation of prior distribution for a certain BRDF model parameter
$\hat{\sigma}$	estimated standard deviation of proposal distribution for a certain BRDF model parameter
τ	change point that distinguished two sets of BRDF model parameters
τ_{max}	predefined maximum of allowed change points per image direction
$\boldsymbol{\tau}$	vector of change points in horizontal or vertical direction
ϕ	azimuth angle
ϕ_{h}	azimuth angle defining the half vector
ϕ_{in}	azimuth angle defining the incident light direction

ϕ_{iso}	isotropic azimuth angle stating only the relative difference between ϕ_{in} and ϕ_{out}
ϕ_{out}	azimuth angle defining the viewing direction
ϕ_{r}	azimuth angle defining the direction of mirror-like reflection
Φ	radiant energy emitted, reflected or transmitted per unit time
χ	full width at half maximum of specular cosine lobe
ψ	angle between viewing direction and half vector
ω	solid angle defining a section of a sphere
ω_{in}	solid angle describing an arbitrary direction of incidence
ω_{out}	solid angle describing an arbitrary direction of emission
Ω	solid angle covering the upper hemisphere of a surface

LIST OF FIGURES

2.1 The pinhole camera model and its associated coordinate systems. 6
2.2 Definitions of a local coordinate system in contrast to the camera coordinate system. 8
2.3 Superposition of diffuse and specular reflection component according to a cosine lobe model. 11
2.4 Illustration of Fresnel effect. 11
2.5 Examples of different reflection phenomena. 13
2.6 Basic scene rendered with varying complexity of illumination. 14
2.7 Rendered images to display difference in Phong's and Blinn's BRDF model. . . 16
2.8 Shadowing and masking on a rough surfaces. 18

3.1 Stereo image analysis in standard geometry 34

4.1 Acquisition setup 46
4.2 Distribution of depth measurements around mean value. 47
4.3 Standard deviation of depth measurements on Leaf data set. 47
4.4 Standard deviation of depth measurements on Plaster data set. 48
4.5 Distribution of intensity measurements around mean value. 49
4.6 Mean camera image and map of normality test. 50
4.7 Camera response curves for relevant interval of exposure times. 51
4.8 Changes of effective pixel size on distant surface. 52

5.1 Example of a Metropolis-Hastings parameter sampling. 57
5.2 Illustration of the concept of change points in two dimensions. 60
5.3 Exemplary input image and reflectance map of synthetic dataset. 63
5.4 Histogram of visited models in model space of synthetic dataset. 63
5.5 States of Markov chain for different parameters. 65
5.6 Similarity of the intensity value for the estimated parameters of region 6 and ground truth values 66
5.7 Exemplary image and reflectance map of laboratory dataset. 67
5.8 Histogram of visited models in model space on laboratory data. 68
5.9 Progress of Markov chain on laboratory data. 69
5.10 Exemplary regions of uncertain affiliation on laboratory data. 70
5.11 k-means segmentation of synthetic data. 70
5.12 Gaussian mixture model segmentation of synthetic data. 71
5.13 Normalized cut segmentation of synthetic data. 71
5.14 Normalized cut segmentation of median image of synthetic data. 72

5.15 Normalized cut segmentation of mean image of synthetic data. 72

6.1 Calibration sphere lit from different light positions. 76
6.2 Comparison of acquired images and renderings of diffuse calibration sphere. . . . 78
6.3 Pixelwise relative error of Lambertian and Phong BRDF model. 79
6.4 Mean radiance reflected from diffuse sphere. 80
6.5 Position of light sources and spheres after calibration. 86

7.1 Flow chart of reflectance based refinement algorithm. 90
7.2 Input image and reflectance map of synthetic dataset. 93
7.3 BRDF parameter maps of the synthetic dataset. 94
7.4 Uniform shading of depth data of synthetic dataset. 94
7.5 RGB images of the objects used to test the proposed algorithm. 95
7.6 Uniformly shaded depth data based on pseudo diffuse photometric stereo estimate. 96
7.7 Test objects with spatially varying reflectance properties. 97
7.8 Parameter maps of the Styrofoam egg and plaster object. 98
7.9 Uniform shading of ground truth depth data from zSnapper and final result of the proposed method. 99
7.10 Close up view of ‘X’ scratch in plaster surface. 100
7.11 Horizontal depth profiles of laboratory data. 100
7.12 Flow chart of stereo fusion algorithm. 103
7.13 Mask to compute discrete gradients. 106
7.14 Datasets used to evaluate the Stereo Fusion algorithm. 109
7.15 RMSE with respect to reference data for different values of γ_δ. 109
7.16 Normal maps of the Ram dataset. 111
7.17 Pixelwise depth differences of the Ram dataset to 3D scanner. 112
7.18 Uniform shadings of the Ram surface. 113
7.18 Uniform shadings of the Ram surface. 114
7.19 Normal maps of the Ganesh dataset. 115
7.20 Pixelwise depth differences of the Ganesh dataset. 116
7.21 Uniform shadings of the Ganesh surface. 117
7.22 Normal maps of the Plaster dataset. 118
7.23 Pixelwise depth differences of the Plaster dataset. 119
7.24 Uniform shadings of the Plaster surface. 120
7.25 Normal maps of the Leaf dataset. 121
7.26 Pixelwise depth differences of the Leaf dataset. 122
7.27 Uniform shadings of the Leaf surface. 123
7.28 Cumulative distribution of pixelwise errors of all datasets. 124

A.1 Mean camera images and maps of normality test. xvi
A.2 Input images and corresponding reflectance maps of synthetic dataset. xix
A.3 Input images and reflectance maps of laboratory dataset for RJMCMC evaluation. xxi
A.4 Input images and corresponding reflectance maps of the white egg dataset. . . . xxiv
A.5 Input images and corresponding reflectance maps of the white plaster dataset. . xxvi
A.6 Left and right stereo input images of the ram dataset. xxviii
A.7 Left and right stereo input images of the ganesh dataset. xxx
A.8 Left and right stereo input images of the plaster dataset. xxxii
A.9 Left and right stereo input images of the leaf dataset. xxxiv

List of Tables

2.1 Limitations and features of BRDF models. 14
2.2 Comparison of eigenvalue configurations for Lafortune model. 17

4.1 Average values of standard deviation per pixelwise measurement on Leaf and Plaster datasets. 48
4.2 Validity of linear regression. 50
4.3 Examplary values of projected pixel size. 52

5.2 Initial values and settings of Markov chain of synthetic dataset. 63
5.3 Maximum a-posteriori estimate of parameters and $1-\sigma$ deviation after burn-in phase. 64
5.4 Similarity of effective diffuse component for different parameter sets. 66
5.5 Initial values and settings of Markov chain of synthetic dataset. 66
5.6 Maximum a-posteriori estimate of parameters and $1-\sigma$ deviation for upper and lower region after burn-in phase of 20 %. 68

6.2 Results of the non-linear optimization of the light source positions. 85

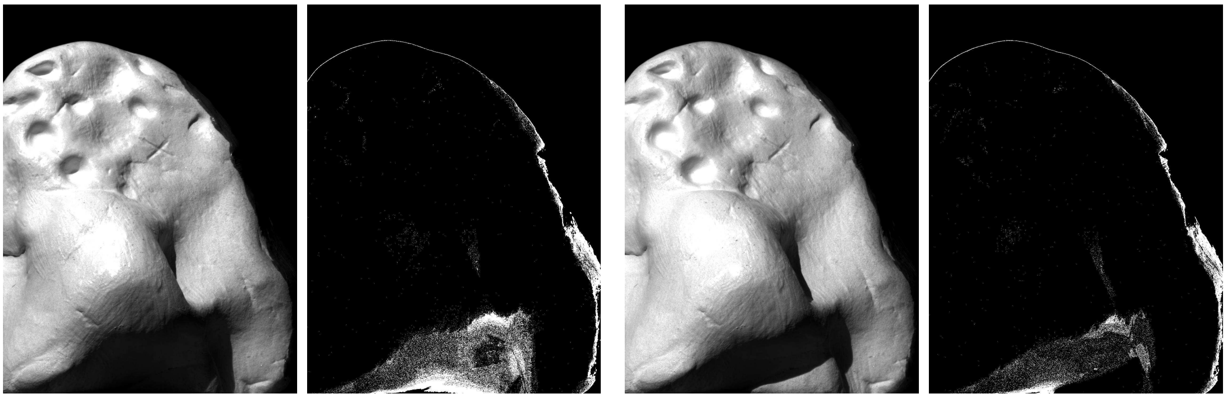
7.2 RMSE between zero-mean scanner data and proposed algorithm. 98

FULL DATASETS A

This appendix contains the full datasets that have been acquired to evaluate the proposed algorithms. In the chapters that elaborate on the experimental evaluation, only a subset of the images has been displayed to facilitate readability.

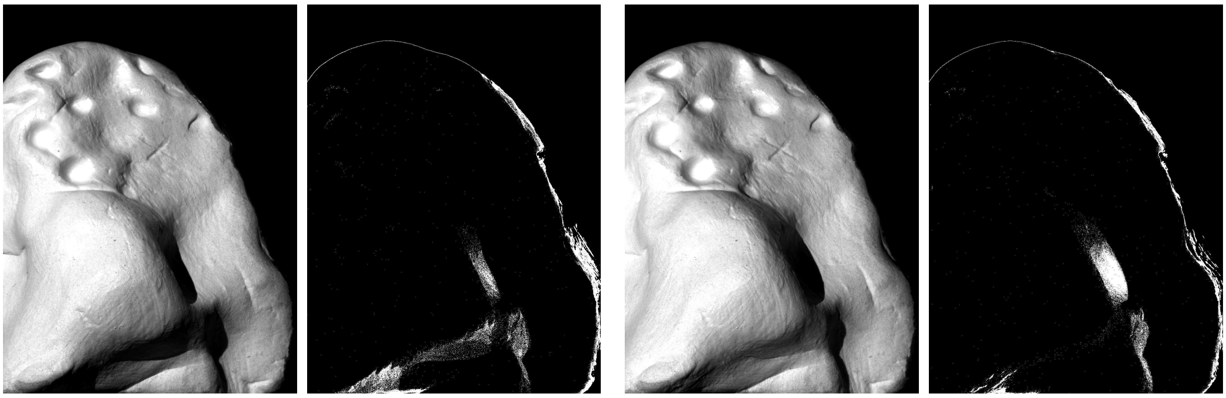
A.1 Camera Image Assessment

The full set of images of the Plaster dataset for which 100 measurement cycles have been acquired is displayed in this section in Fig. A.1. The Kolmogorov-Smirnov test for normality was applied to test the distribution of each pixel and those that failed to pass the test are located almost exclusively in areas of self-cast or attached shadow. Although not all shadowed pixels fail to pass the KS-test, this is an interesting information considering the detection of shadow pixels in monochrome images, since many shadow removal techniques rely on color space transformations (e.g., Finlayson et al., 2002), RANSAC and outlier rejection in multiple images (e.g., Mukaigawa et al., 2007), exploiting integrability of the gradient field (e.g., Hernandez et al., 2008), or combining color and texture cues (e.g., Xu et al., 2006; McFeely et al., 2012). However, single gray value images still pose a situation that is difficult to handle. Guo et al. (2011) use graph cut inference to find image regions corresponding to the same materials under similar or different illumination conditions. They use the CIELAB color space and rely on features derived by Martin et al. (2004) to detect image boundaries. The corresponding illumination pairs are used to remove the shadow. Zhu et al. (2010) apply the same features comprising statistical information and texture and intensity cues to monochromatic images. Hand-labeled ground truth data is used to train a conditional random field and detect shadow labels. In contrast to hand-crafted shadow variant and invariant features, Khan et al. (2016) train multiple convolutional neural networks (CNNs) in a supervised fashion to identify shadow maps on outdoor color scenes. They require Lambertian reflection and corresponding color distributions between shadow and non-shadow regions to remove the shadows with a Bayesian shadow generation model accurately. Hiary et al. (2018) employ quaternions to represent color images and realize custom color space transformations. The threshold to separate shadows in a bi-modal histogram is computed from the ratio of two transformed color channels. All these methods are only a brief peek at the variety of shadow detection and removal algorithms, yet they illustrate that researchers are still trying to find a robust and accurate scheme to detect shadows, especially in monochromatic single-image scenarios.



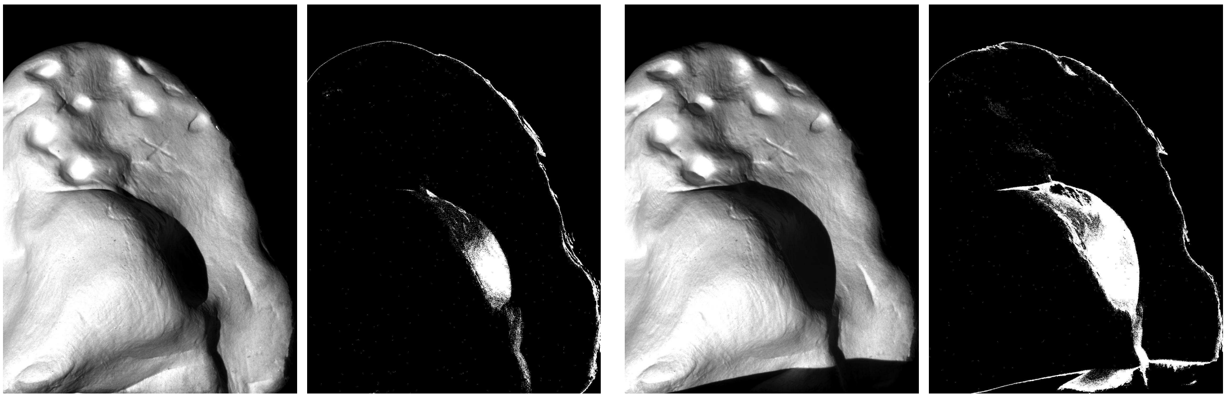
(a) LED 1.1

(b) LED 1.2



(c) LED 1.3

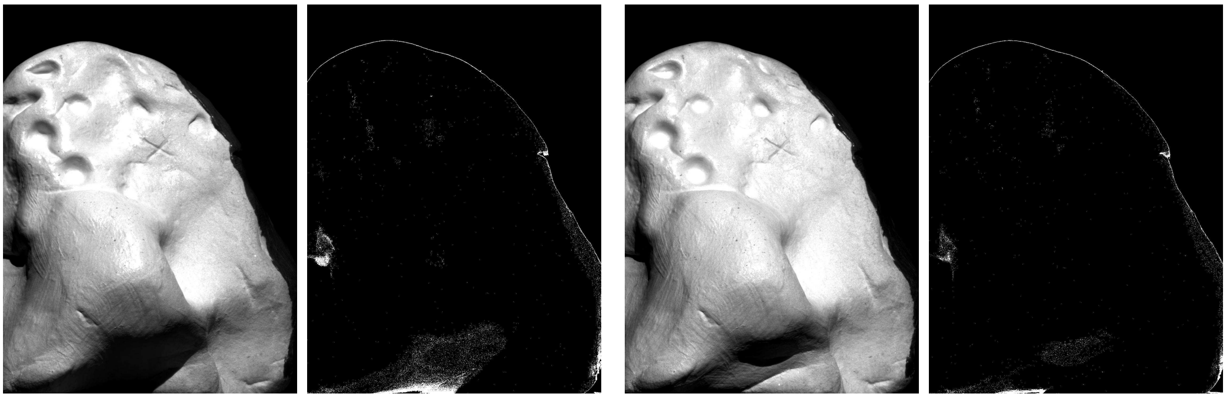
(d) LED 1.4



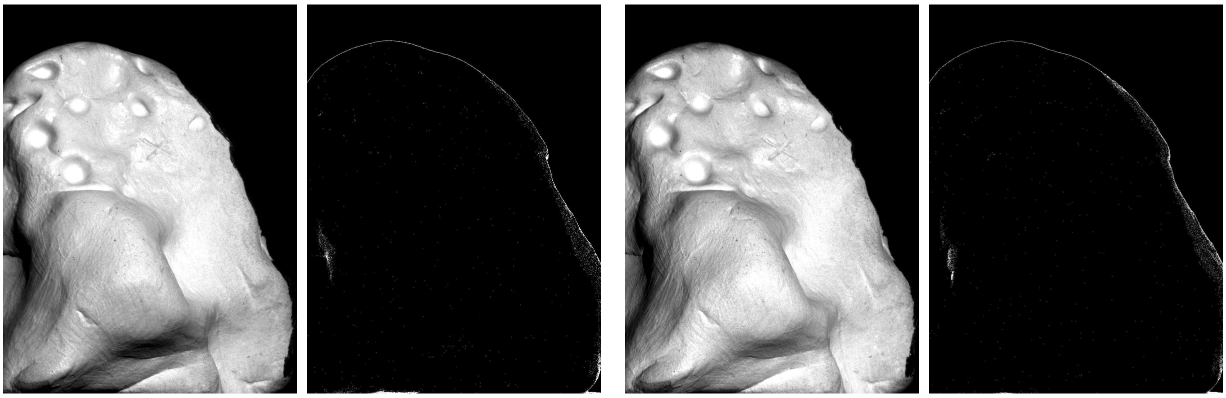
(e) LED 1.5

(f) LED 1.6

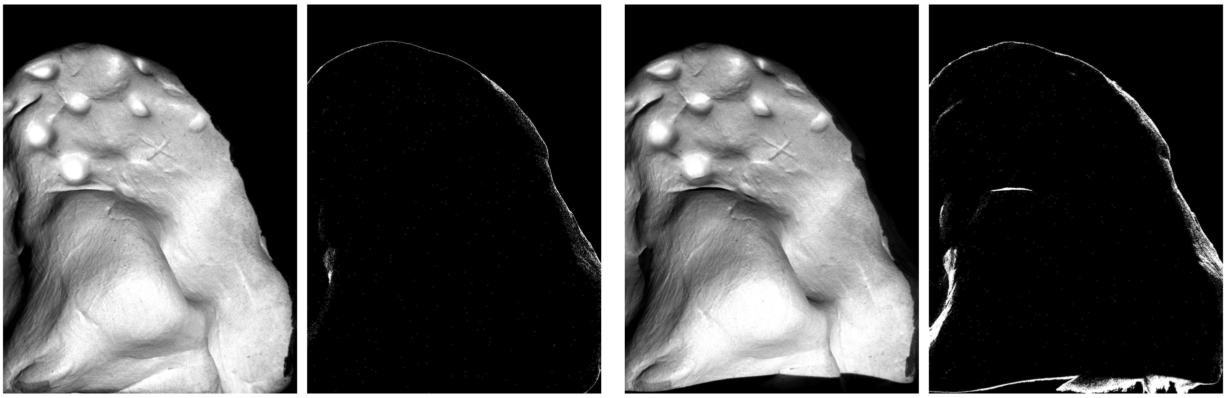
Figure A.1. Mean camera images and maps of failed normality test for full set of Plaster data. The Kolmogorov-Smirnov test was used on 100 samples of each pixel. (*cont.*)



(g) LED 2.1 (h) LED 2.2

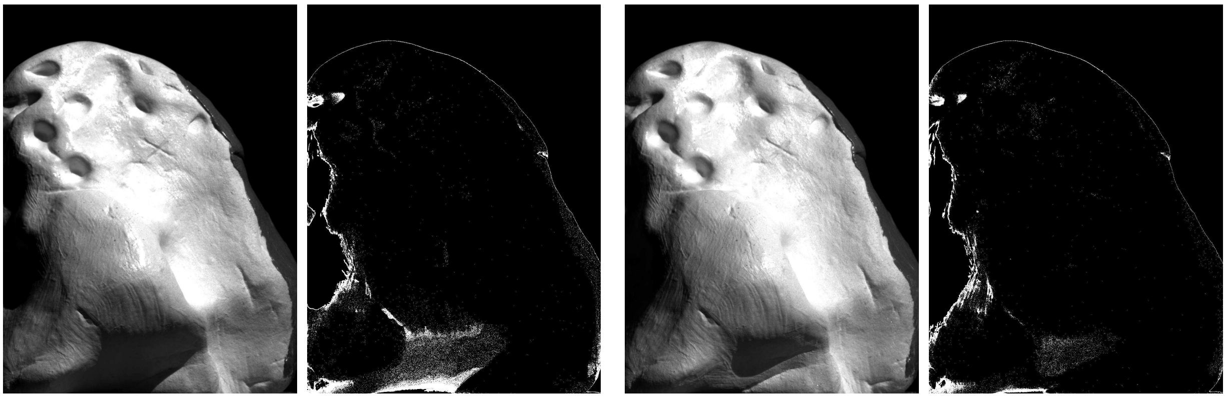


(i) LED 2.3 (j) LED 2.4



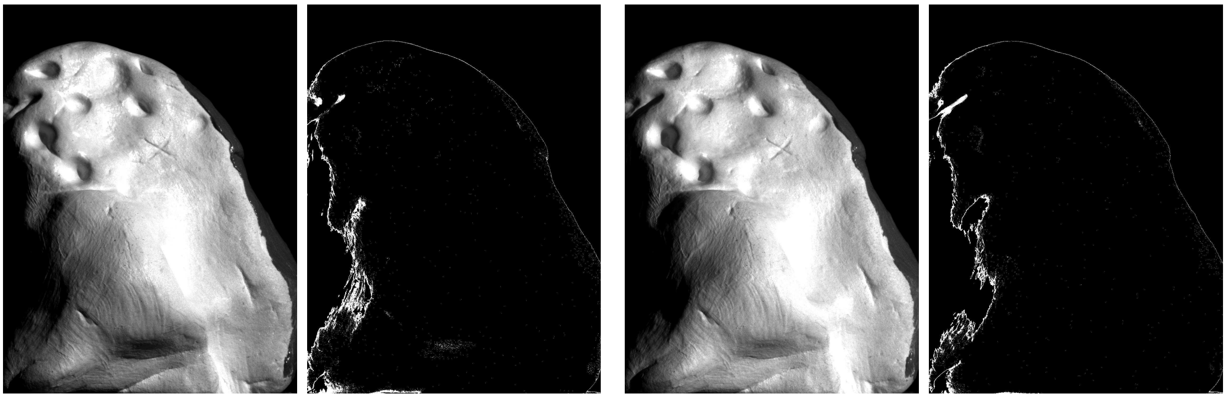
(k) LED 2.5 (l) LED 2.6

Figure A.1 (*cont.*). Mean camera images and maps of failed normality test for full set of Plaster data. The Kolmogorov-Smirnov test was used on 100 samples of each pixel.



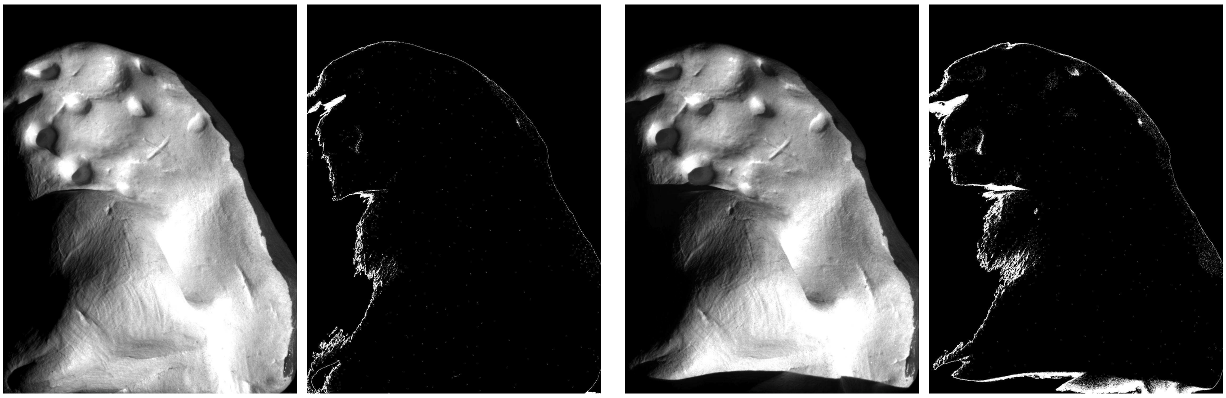
(m) LED 3.1

(n) LED 3.2



(o) LED 3.3

(p) LED 3.4



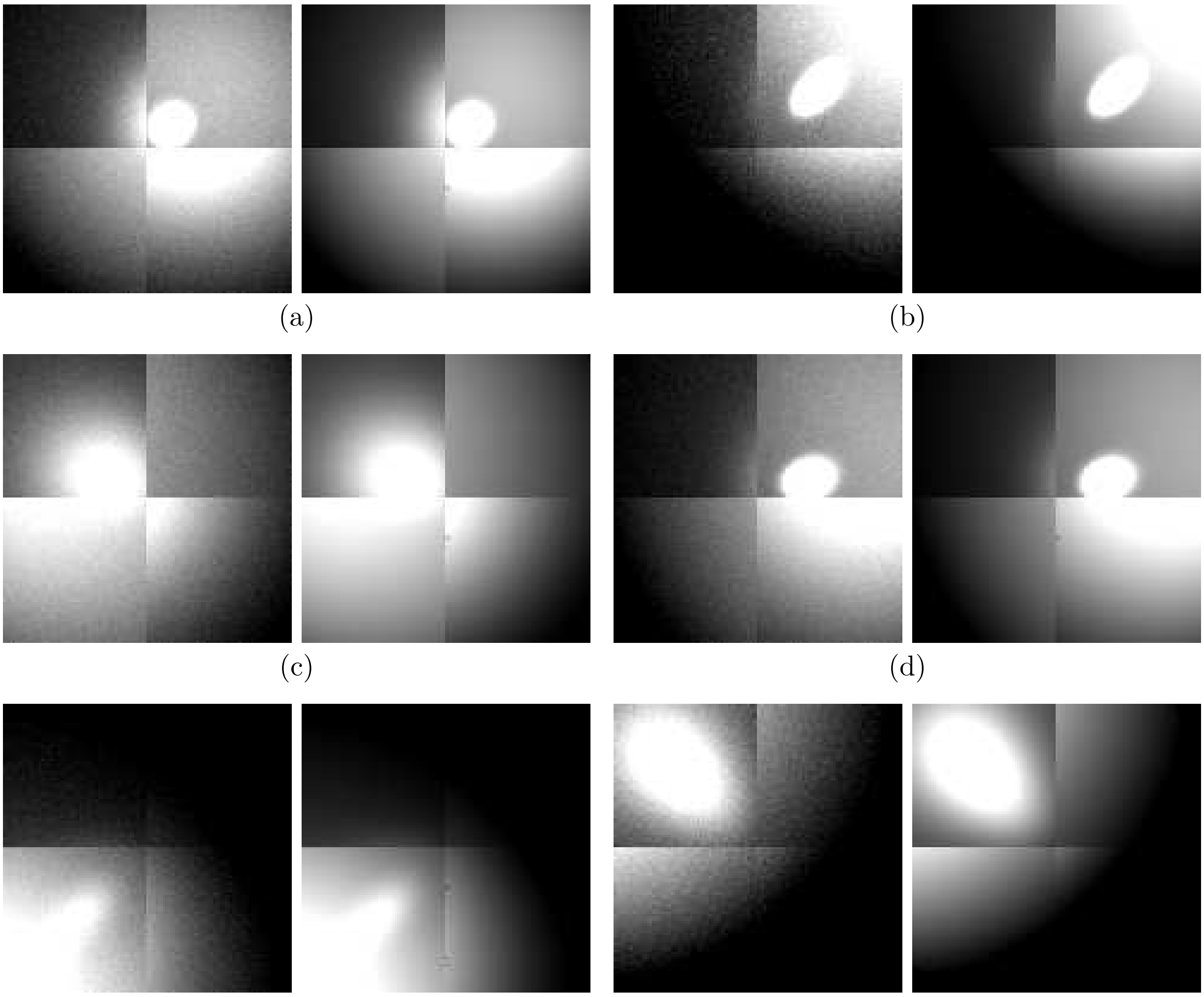
(q) LED 3.5

(r) LED 3.6

Figure A.1 (*cont.*). Mean camera images and maps of failed normality test for full set of Plaster data. The Kolmogorov-Smirnov test was used on 100 samples of each pixel.

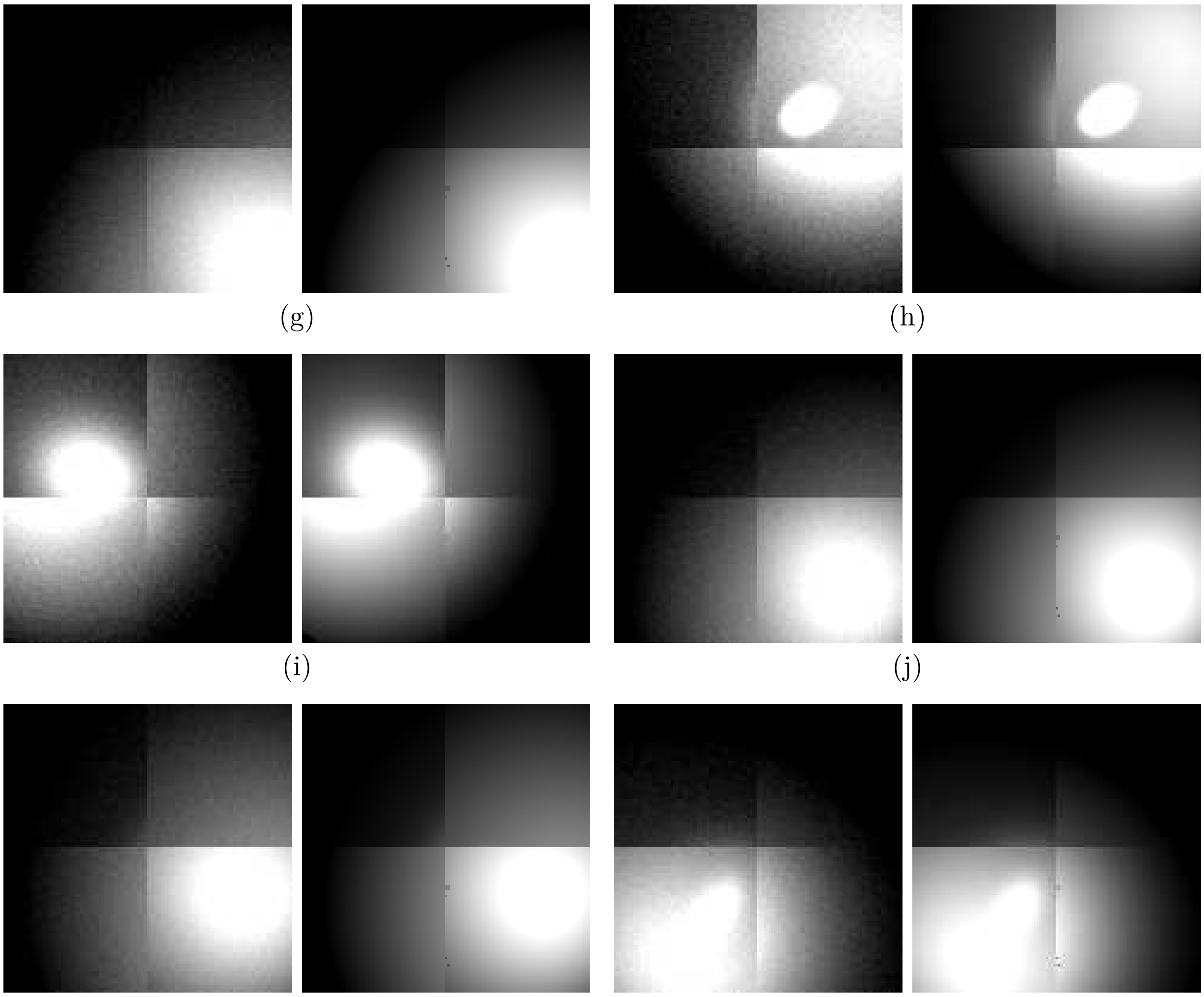
A.2 RJMCMC Datasets

Figure A.2 displays the complete input data and reflectance maps corresponding to the synthetic dataset used in the evaluation of the reversible jump Markov chain Monte Carlo analysis described in Chapter 5. The gray value scaling is identical in corresponding images. The images are reproduced at a very high quality without the additional white Gaussian noise. The input images and reflectance maps of the laboratory dataset are depicted in Fig. A.3.



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

Figure A.2. Noise afflicted input images and corresponding reflectance maps of synthetic dataset for RJMCMC evaluation. Same scaling in all images.

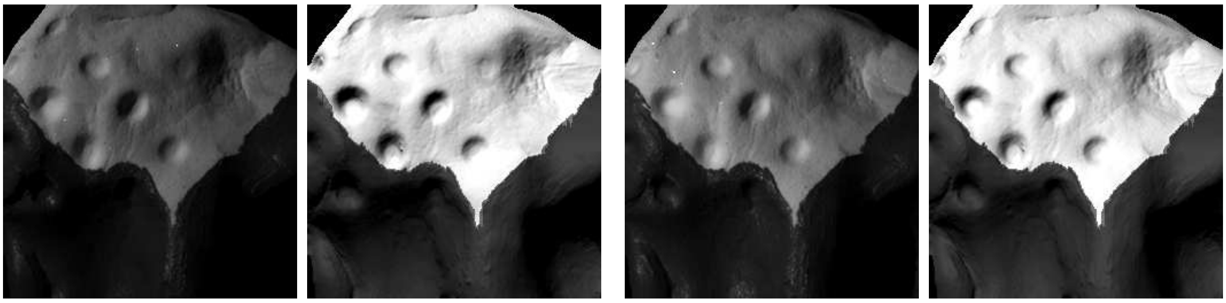


(g) (h)

(i) (j)

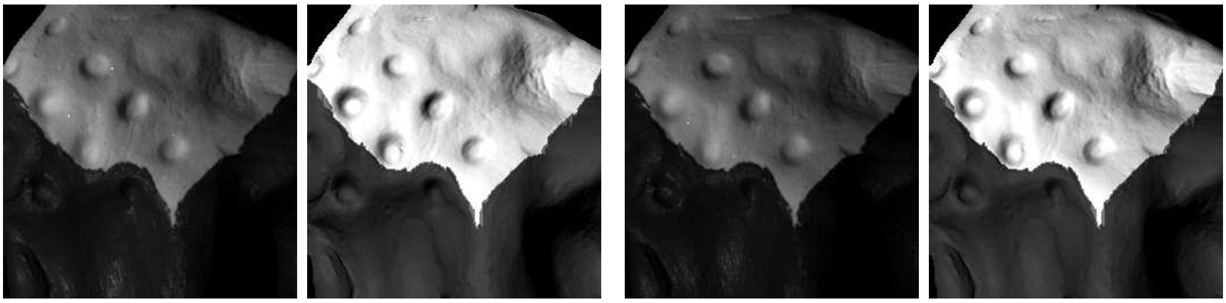
(k) (l)

Figure A.2 (*cont.*). Noise afflicted input images and corresponding reflectance maps of synthetic dataset for RJMCMC evaluation. Same scaling in all images.



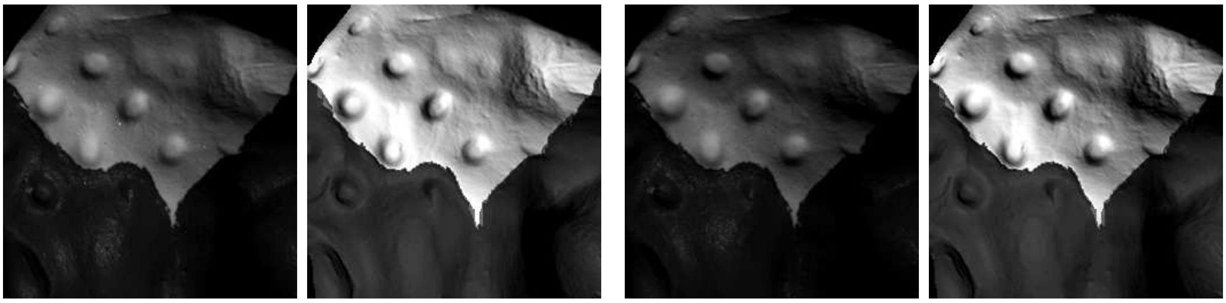
(a) LED 1.1

(b) LED 1.2



(c) LED 1.3

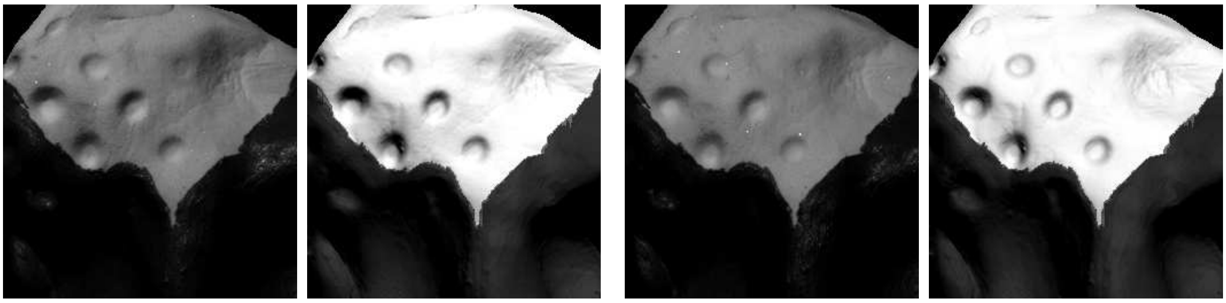
(d) LED 1.4



(e) LED 1.5

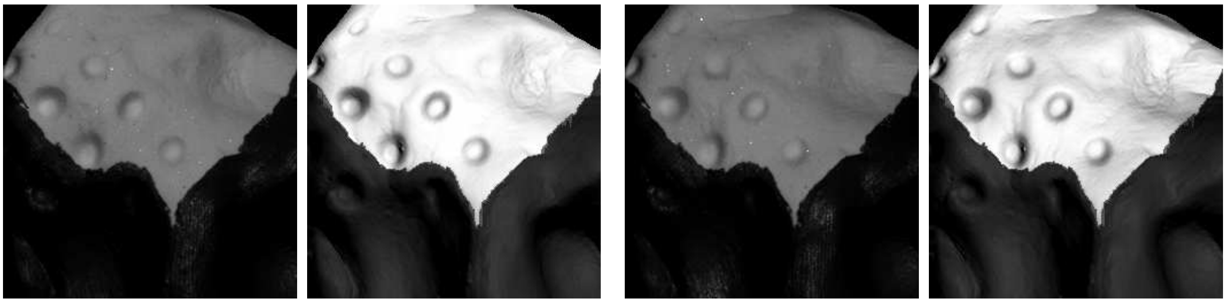
(f) LED 1.6

Figure A.3. Input images and reflectance maps of laboratory dataset for RJMCMC evaluation. Gray value scaling from 0.0229 to 0.2951 in all images.



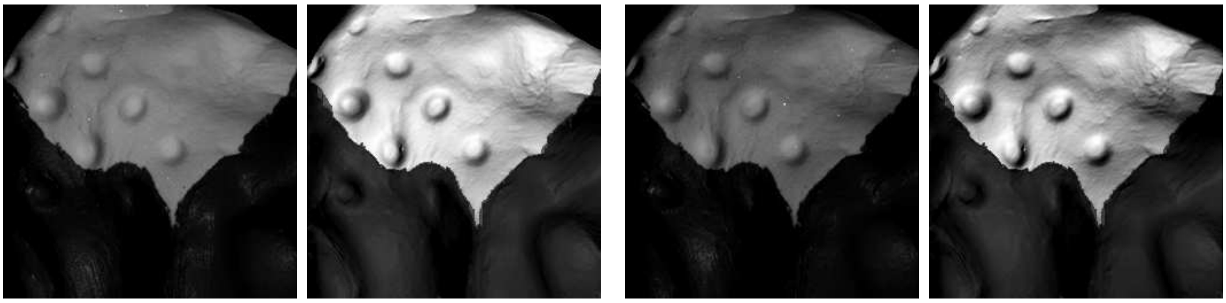
(g) LED 2.1

(h) LED 2.2



(i) LED 2.3

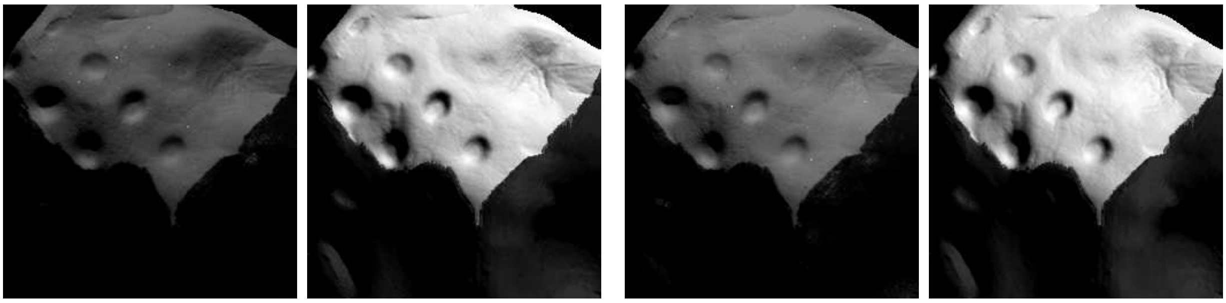
(j) LED 2.4



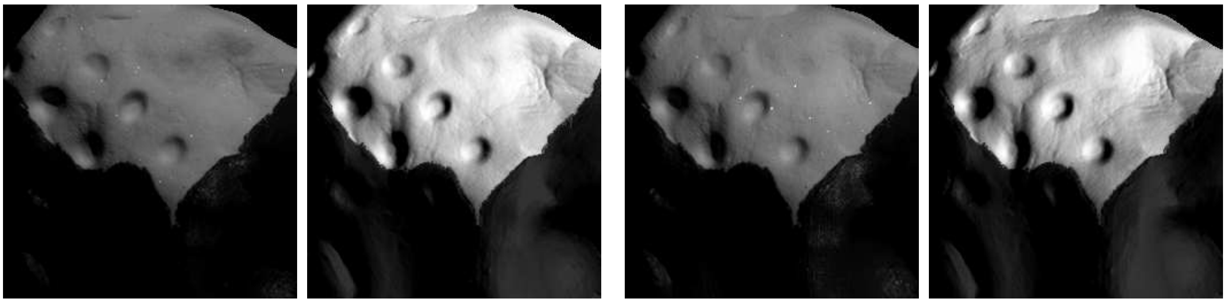
(k) LED 2.5

(l) LED 2.6

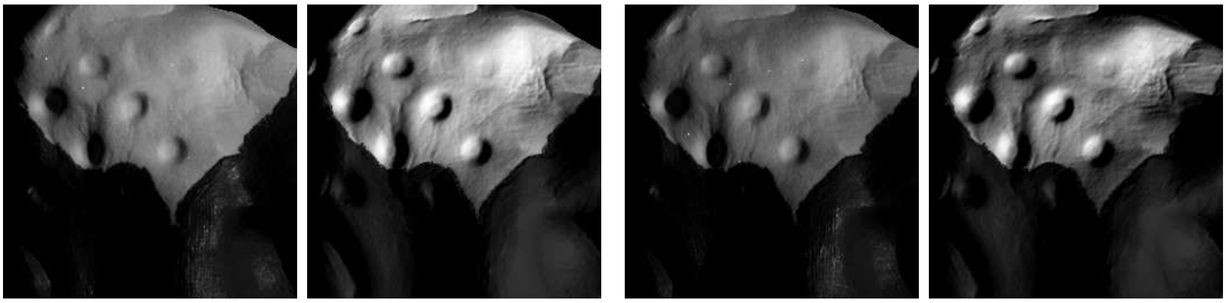
Figure A.3 (*cont.*). Input images and reflectance maps of laboratory dataset for RJMCMC evaluation. Gray value scaling from 0.0229 to 0.2951 in all images.



(m) LED 3.1 (n) LED 3.2



(o) LED 3.3 (p) LED 3.4

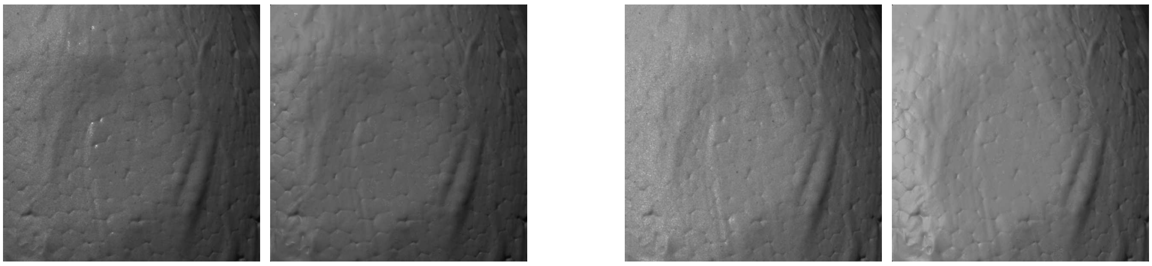


(q) LED 3.5 (r) LED 3.6

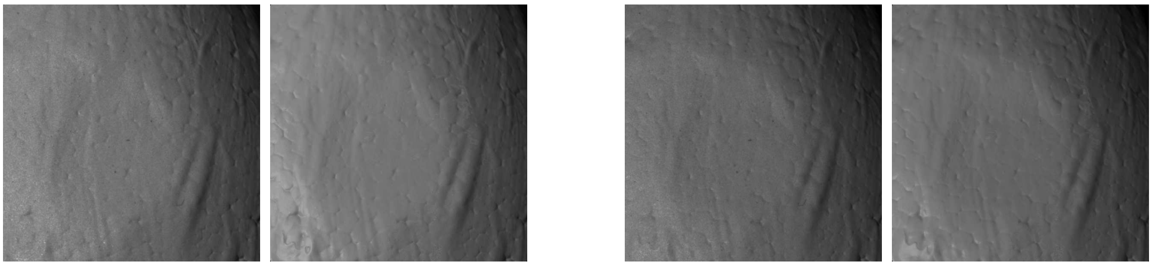
Figure A.3 (*cont.*). Input images and reflectance maps of laboratory dataset for RJMCMC evaluation. Gray value scaling from 0.0229 to 0.2951 in all images.

A.3 Surface Refinement Datasets

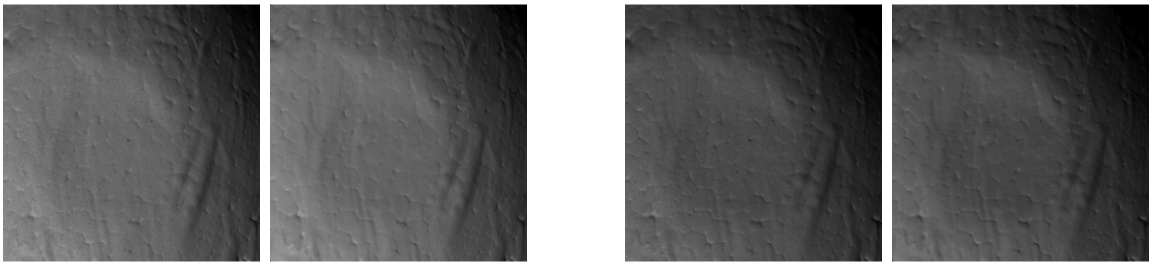
The datasets used in the evaluation of the iterative surface refinement algorithm described in Section 7.1 are depicted in Fig. A.4 and Fig. A.5. Gray values of corresponding images (left) and reflectance maps (right) are scaled identically.



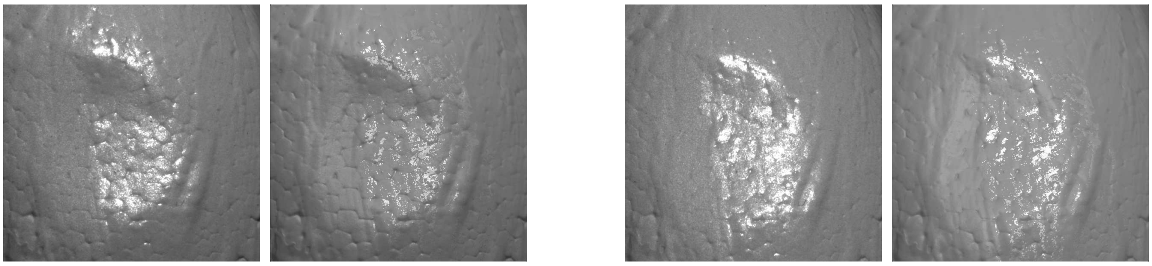
(a) LED 1.1 (b) LED 1.2



(c) LED 1.3 (d) LED 1.4

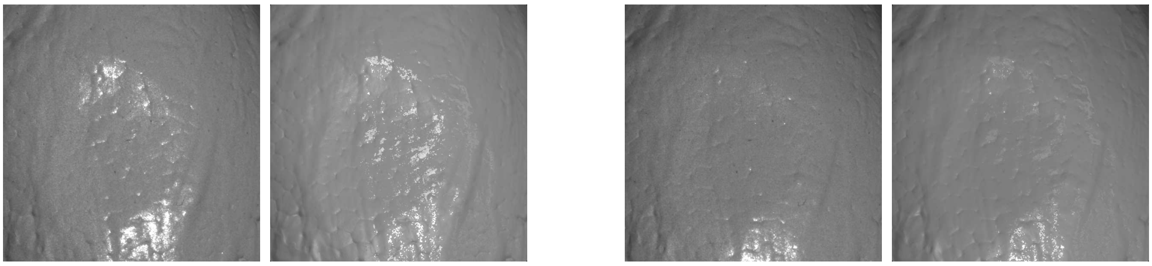


(e) LED 1.5 (f) LED 1.6

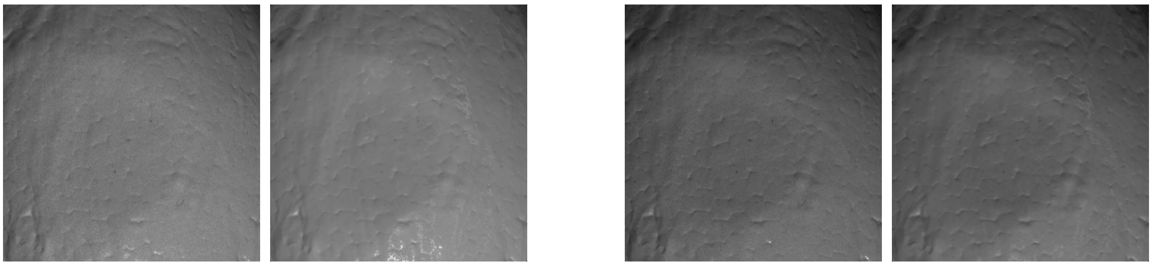


(g) LED 2.1 (h) LED 2.2

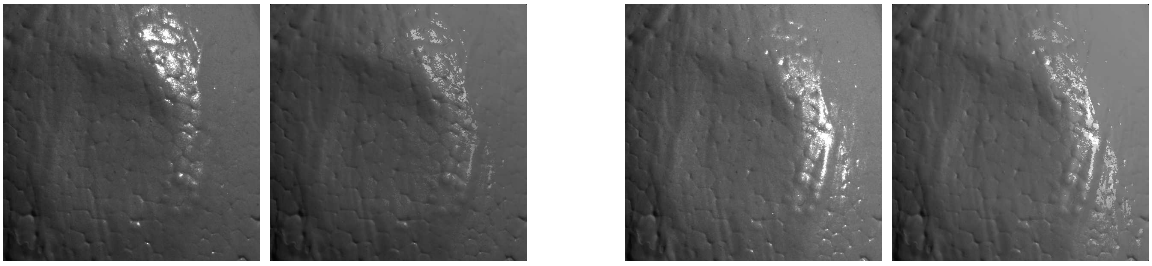
Figure A.4. Input images (left) and corresponding reflectance maps (right) on estimated surface structure of the white egg dataset for evaluation of iterative surface refinement. Same scaling in all images.



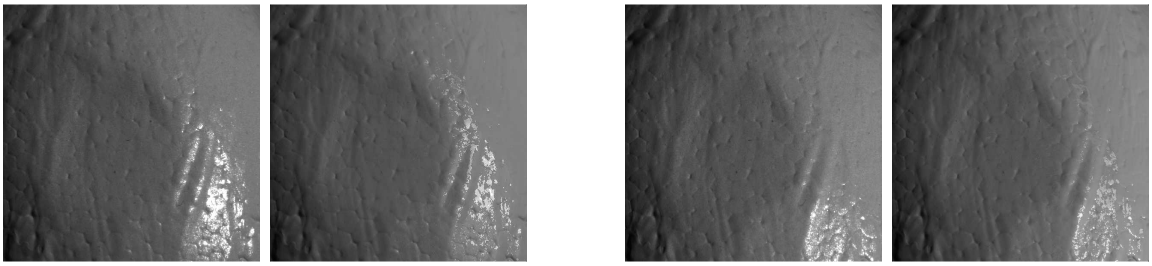
(i) LED 2.3 (j) LED 2.4



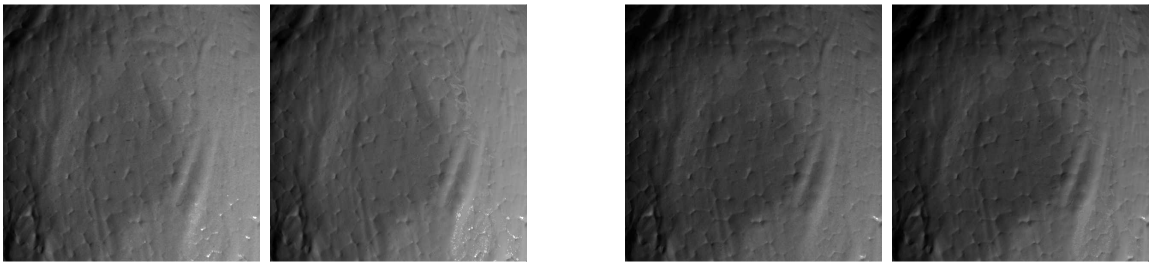
(k) LED 2.5 (l) LED 2.6



(m) LED 3.1 (n) LED 3.2

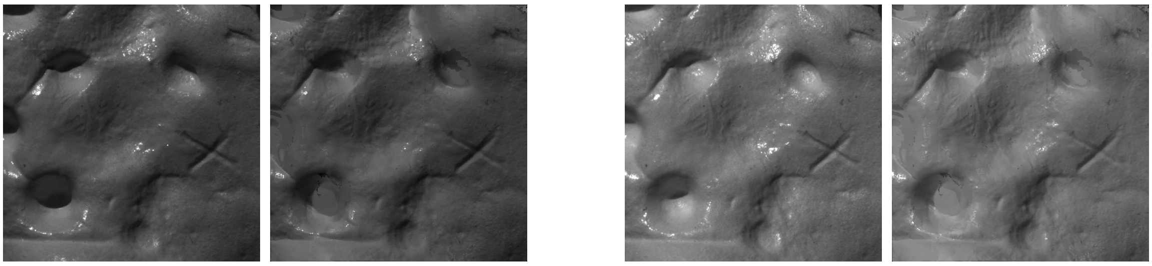


(o) LED 3.3 (p) LED 3.4



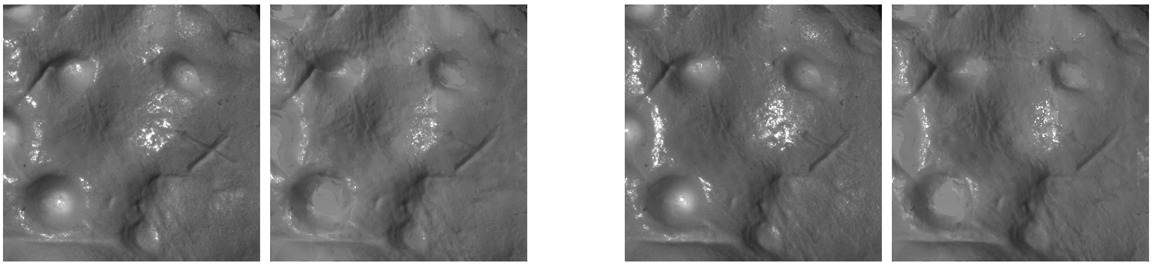
(q) LED 3.5 (r) LED 3.6

Figure A.4 (*cont.*). Input images (left) and corresponding reflectance maps (right) on estimated surface structure of the white egg dataset for evaluation of iterative surface refinement. Same scaling in all images.



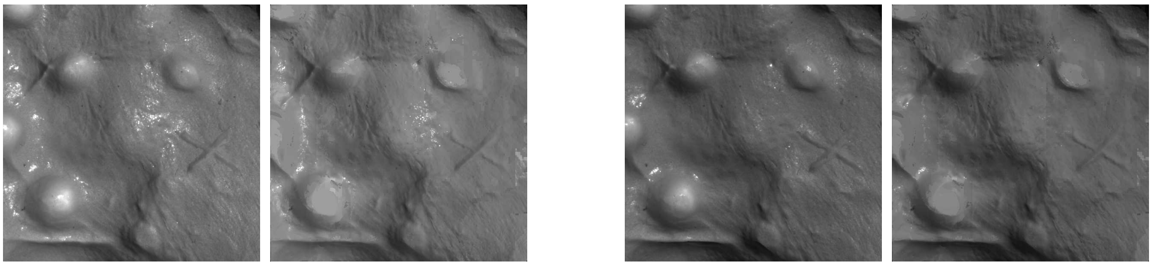
(a) LED 1.1

(b) LED 1.2



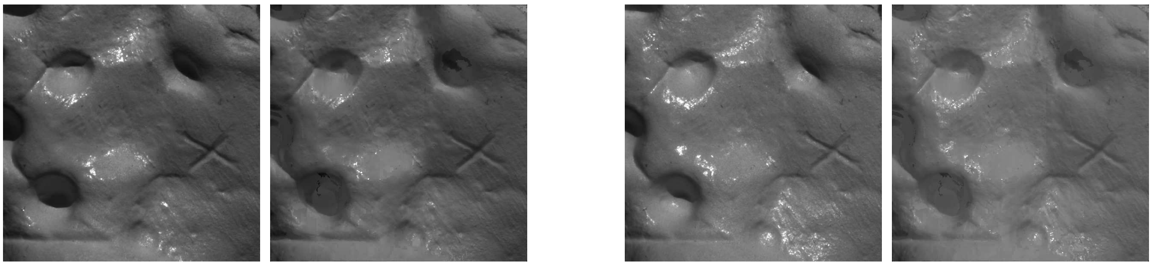
(c) LED 1.3

(d) LED 1.4



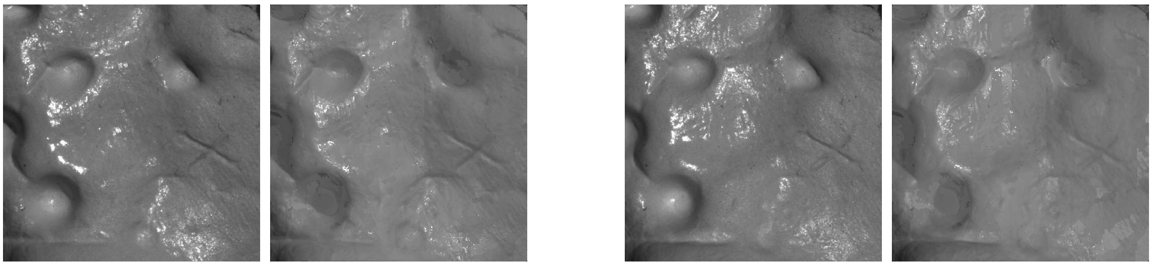
(e) LED 1.5

(f) LED 1.6



(g) LED 2.1

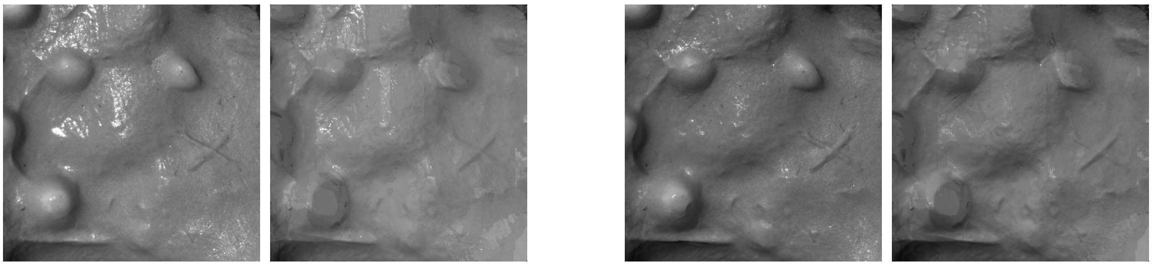
(h) LED 2.2



(i) LED 2.3

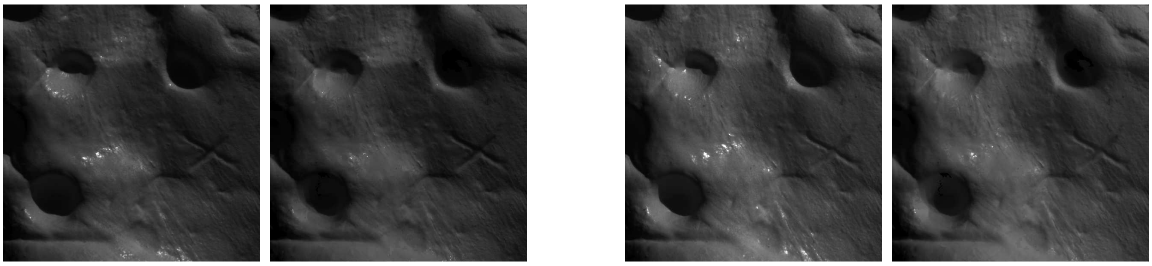
(j) LED 2.4

Figure A.5. Input images (left) and corresponding reflectance maps (right) on estimated surface structure of the white plaster dataset for evaluation of iterative surface refinement. Same scaling in all images.



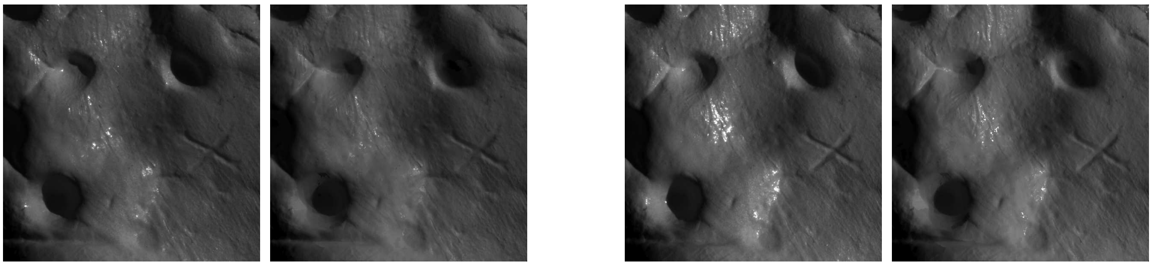
(k) LED 2.5

(l) LED 2.6



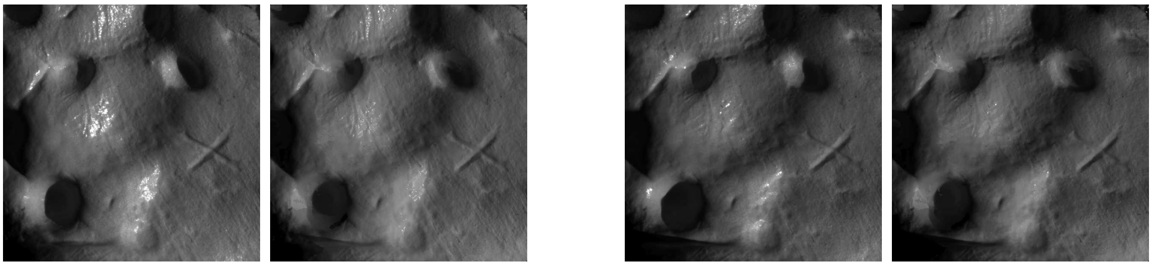
(m) LED 3.1

(n) LED 3.2



(o) LED 3.3

(p) LED 3.4



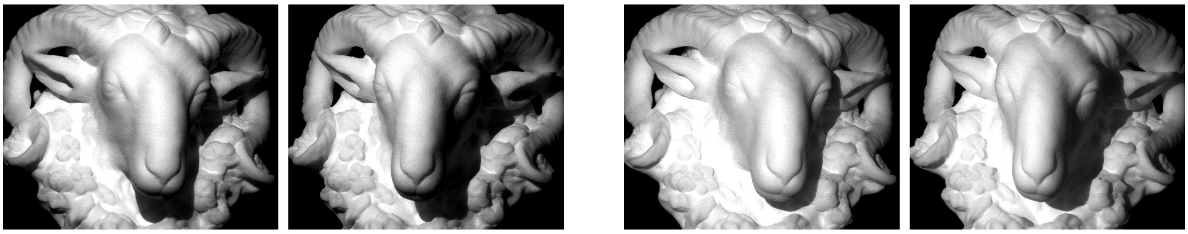
(q) LED 3.5

(r) LED 3.6

Figure A.5 (*cont.*). Input images (left) and corresponding reflectance maps (right) on estimated surface structure of the white plaster dataset for evaluation of iterative surface refinement. Same scaling in all images.

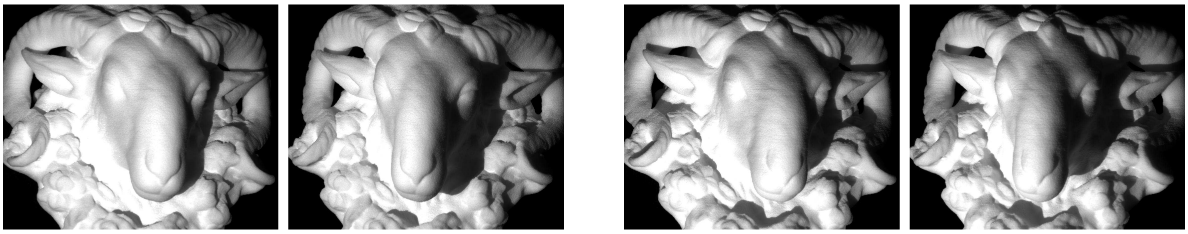
A.4 Stereo Fusion Datasets

The laboratory datasets used to evaluate the Stereo Fusion algorithm depicted in Section 7.2 are contained in this section. The stereo image pairs are scaled in the same gray value interval. The ram, Ganesh, plaster and leaf datasets are shown in Figs. A.6 to A.9.



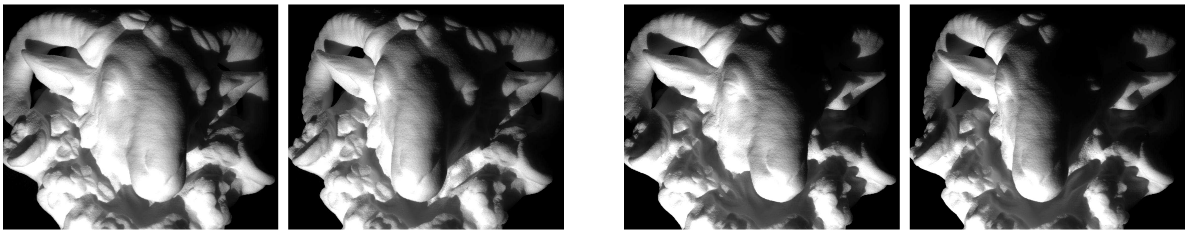
(a) LED 1.1

(b) LED 1.2



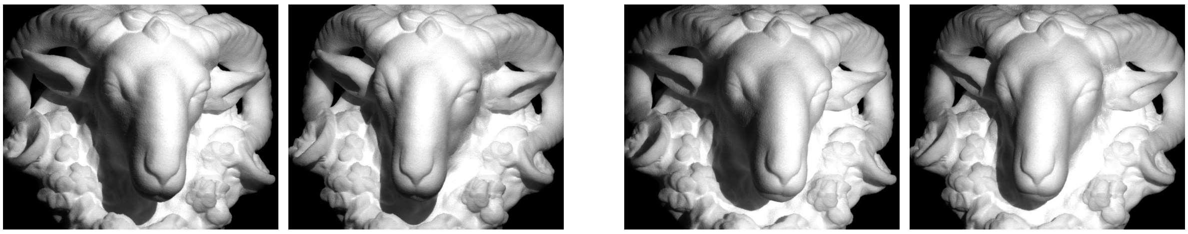
(c) LED 1.3

(d) LED 1.4



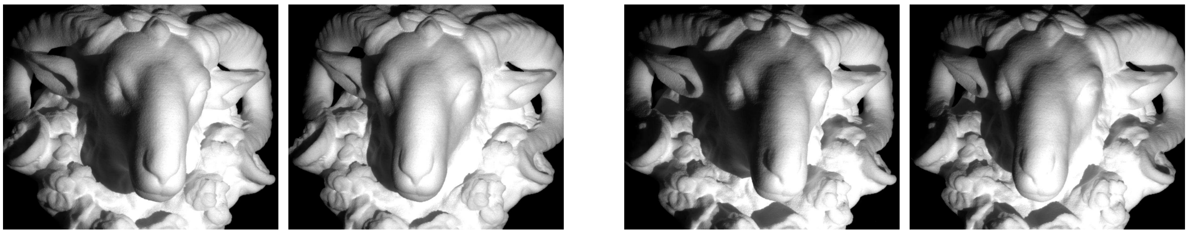
(e) LED 1.5

(f) LED 1.6



(g) LED 2.1

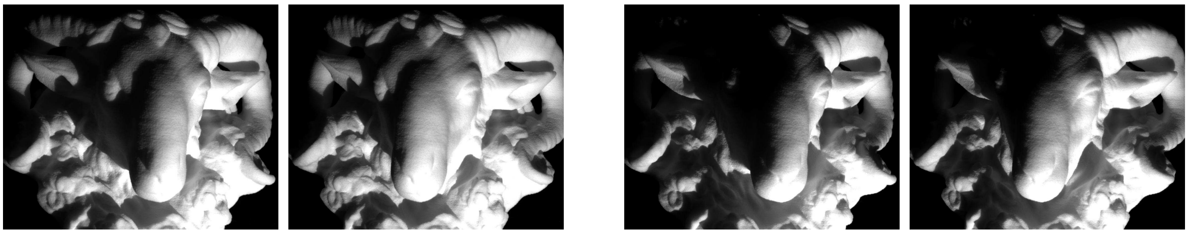
(h) LED 2.2



(i) LED 2.3

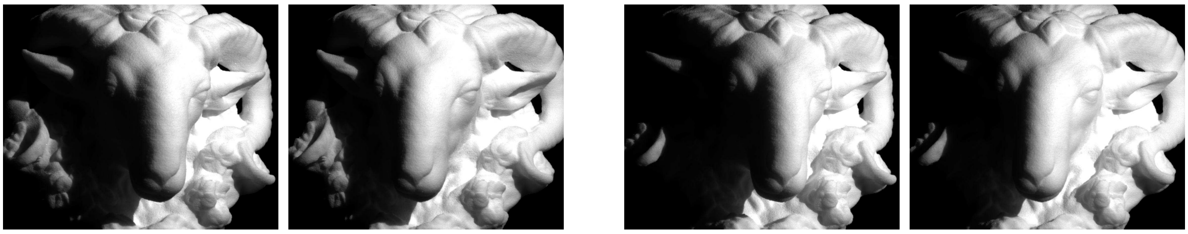
(j) LED 2.4

Figure A.6. Left and right stereo input images of the ram dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



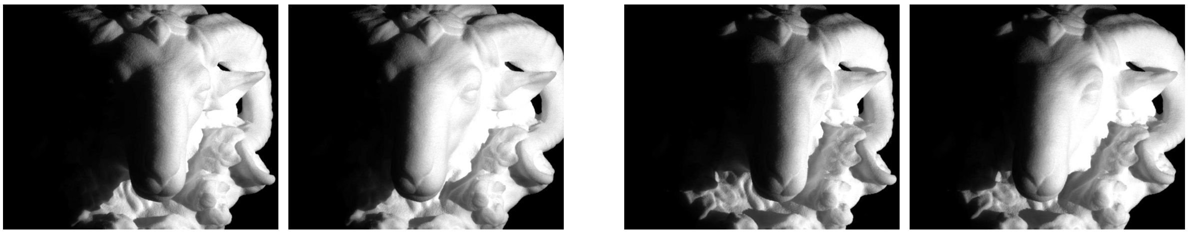
(k) LED 2.5

(l) LED 2.6



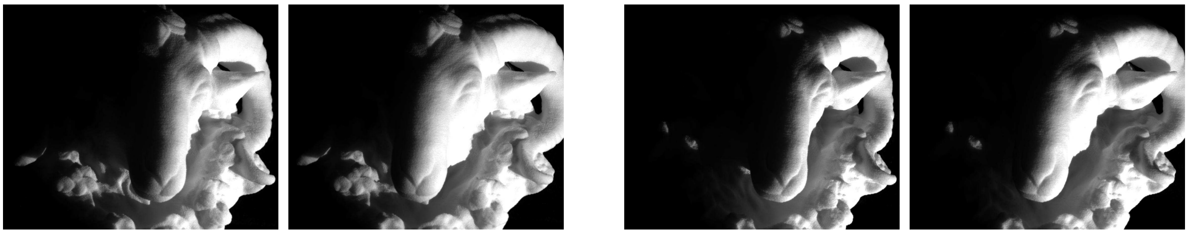
(m) LED 3.1

(n) LED 3.2



(o) LED 3.3

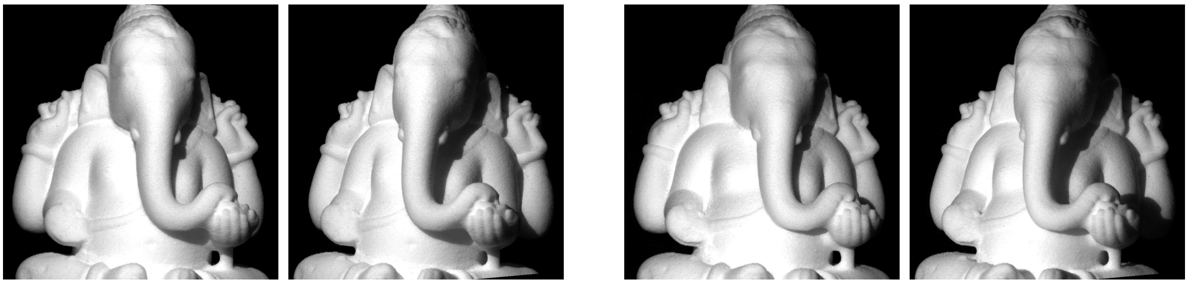
(p) LED 3.4



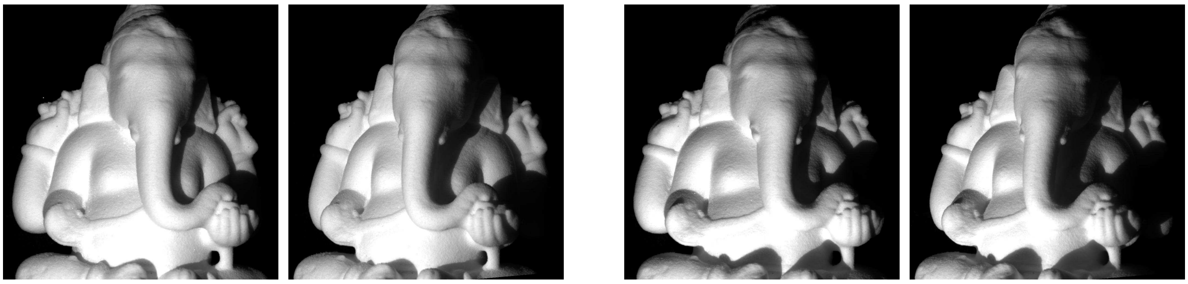
(q) LED 3.5

(r) LED 3.6

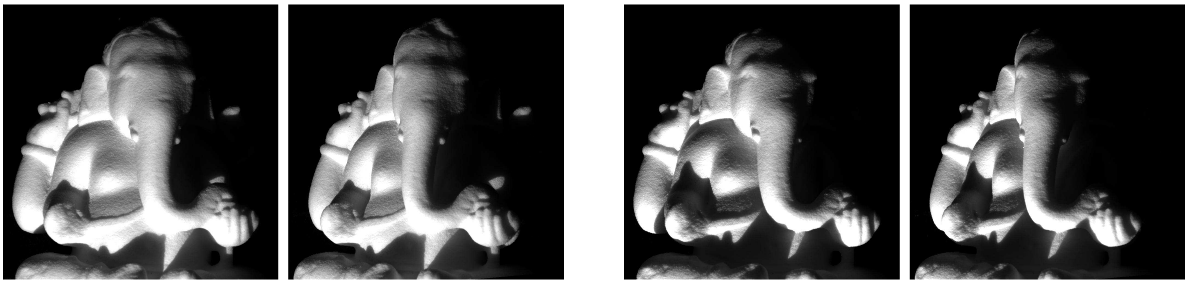
Figure A.6 (*cont.*). Left and right stereo input images of the ram dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



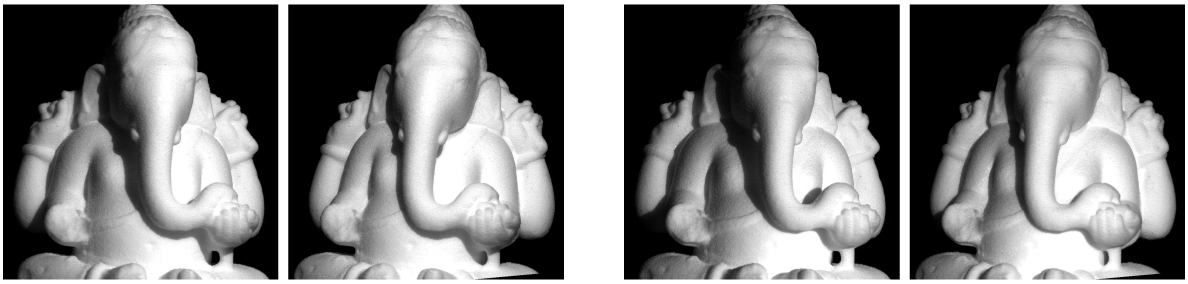
(a) LED 1.1 (b) LED 1.2



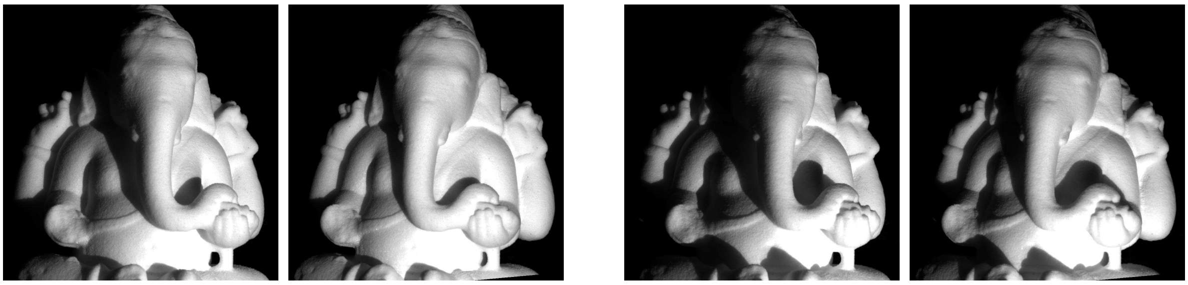
(c) LED 1.3 (d) LED 1.4



(e) LED 1.5 (f) LED 1.6

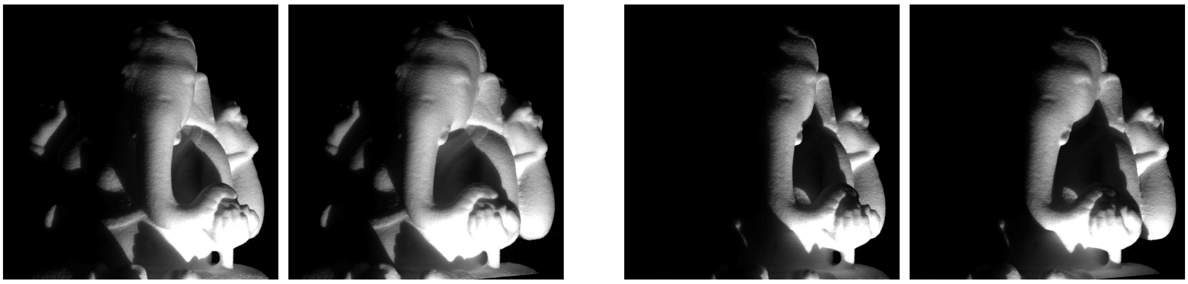


(g) LED 2.1 (h) LED 2.2



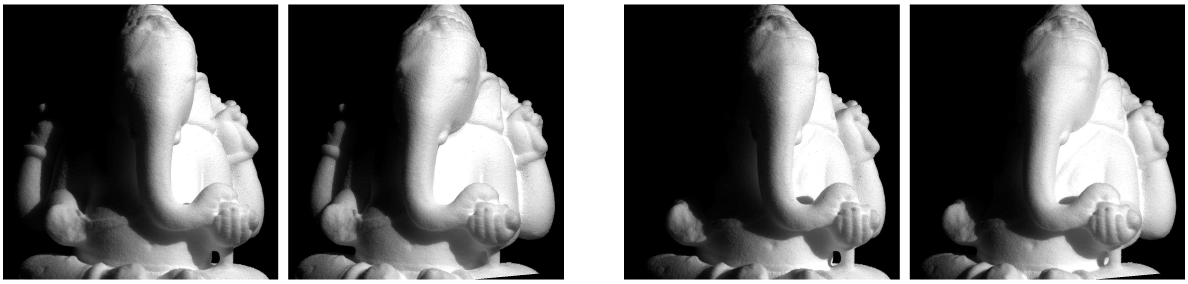
(i) LED 2.3 (j) LED 2.4

Figure A.7. Left and right stereo input images of the ganesh dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



(k) LED 2.5

(l) LED 2.6



(m) LED 3.1

(n) LED 3.2

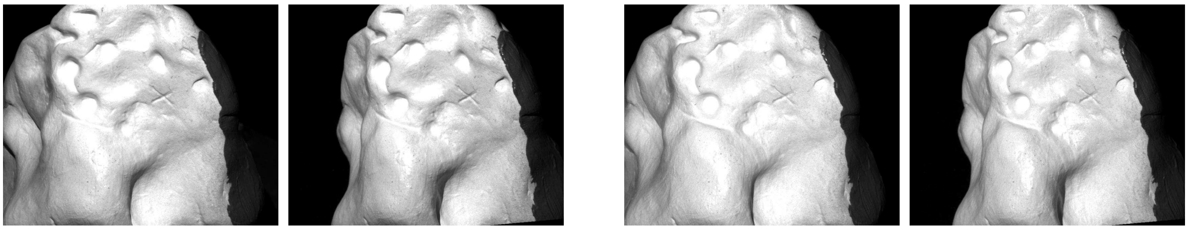
(o) LED 3.3

(p) LED 3.4

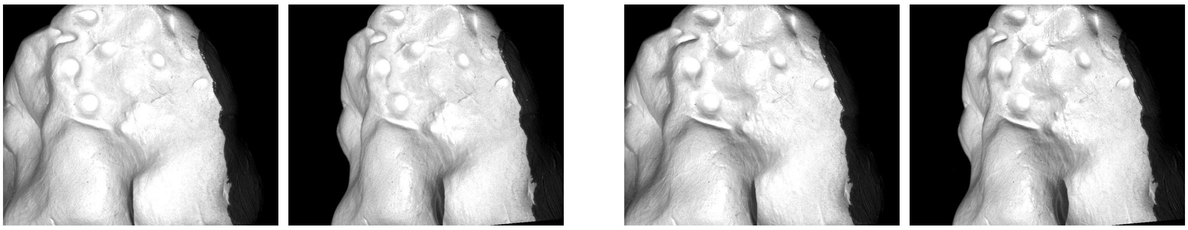
(q) LED 3.5

(r) LED 3.6

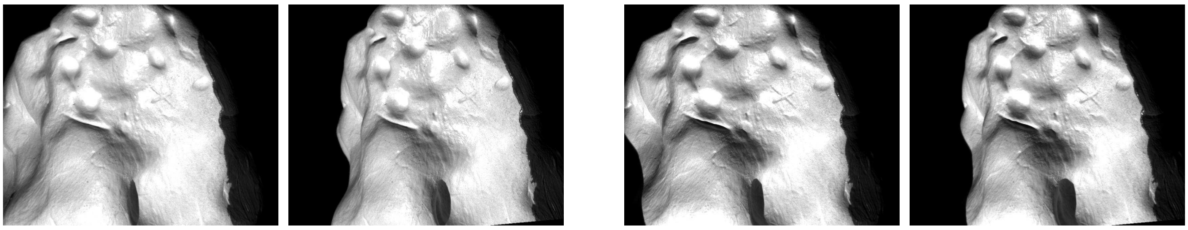
Figure A.7 (*cont.*). Left and right stereo input images of the ganesh dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



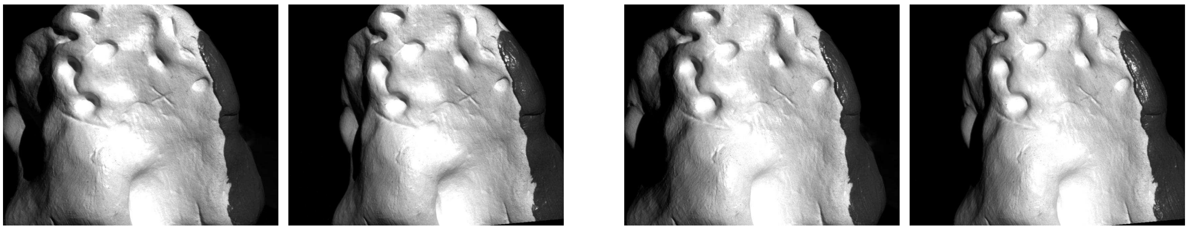
(a) LED 1.1 (b) LED 1.2



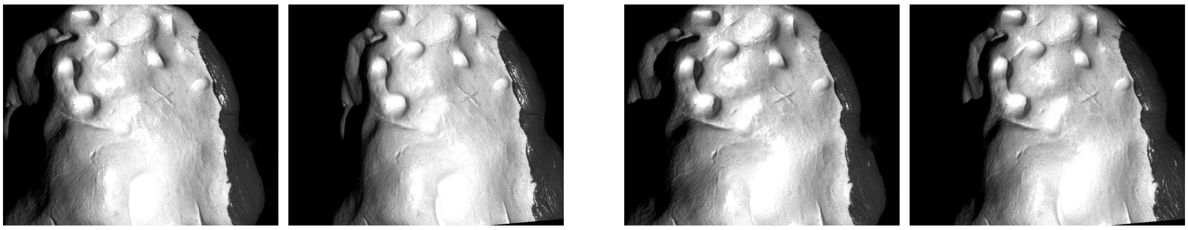
(c) LED 1.3 (d) LED 1.4



(e) LED 1.5 (f) LED 1.6

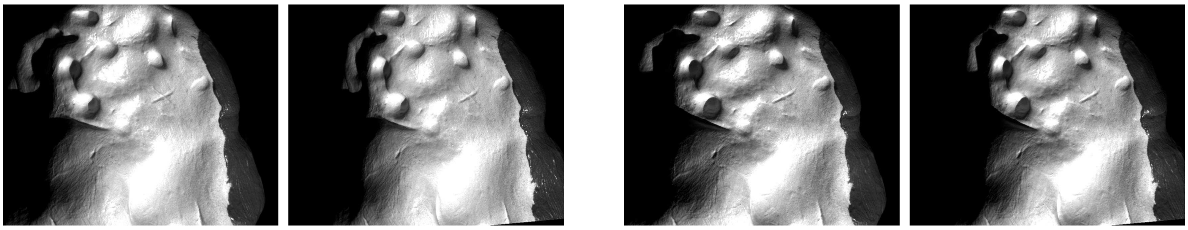


(g) LED 2.1 (h) LED 2.2

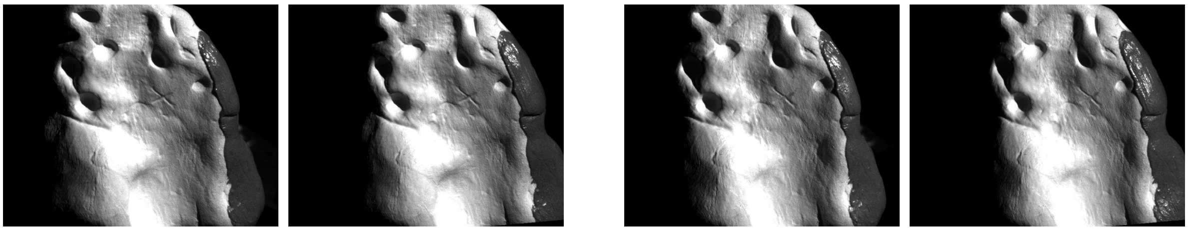


(i) LED 2.3 (j) LED 2.4

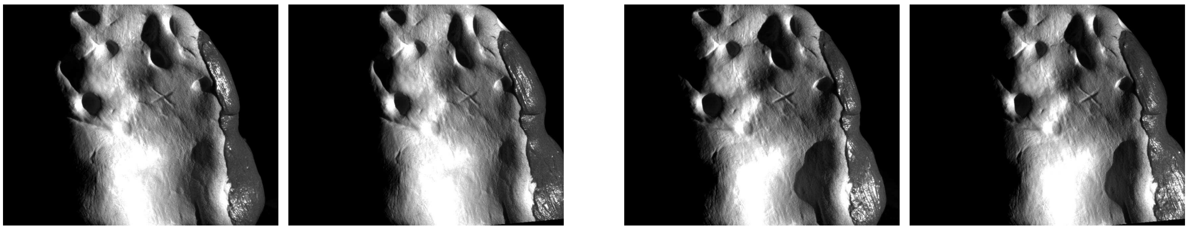
Figure A.8. Left and right stereo input images of the plaster dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



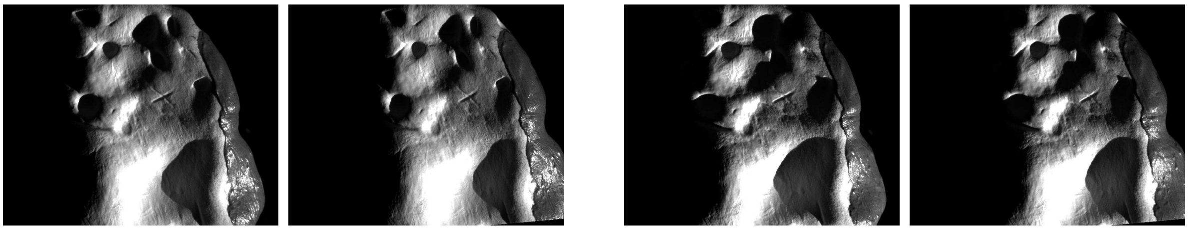
(k) LED 2.5 (l) LED 2.6



(m) LED 3.1 (n) LED 3.2

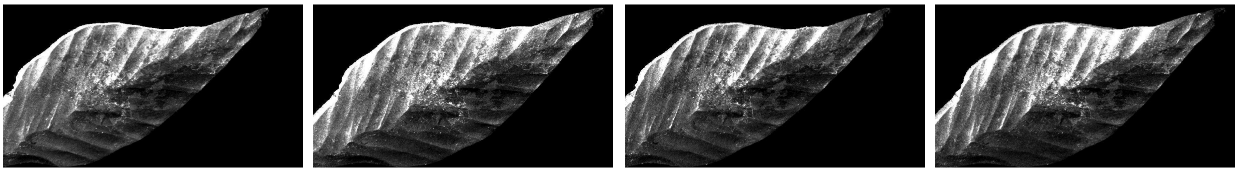


(o) LED 3.3 (p) LED 3.4

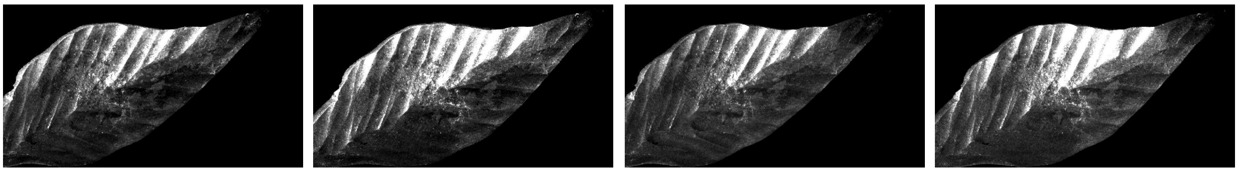


(q) LED 3.5 (r) LED 3.6

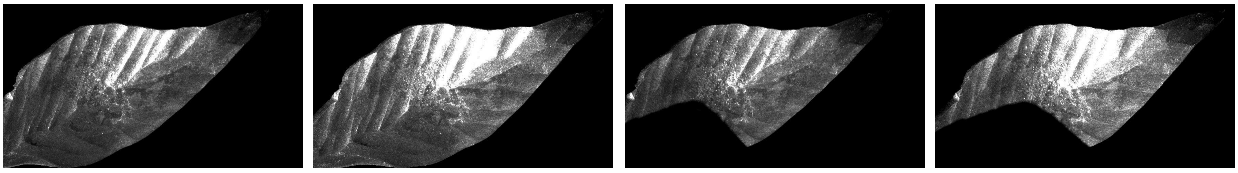
Figure A.8 (*cont.*). Left and right stereo input images of the plaster dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



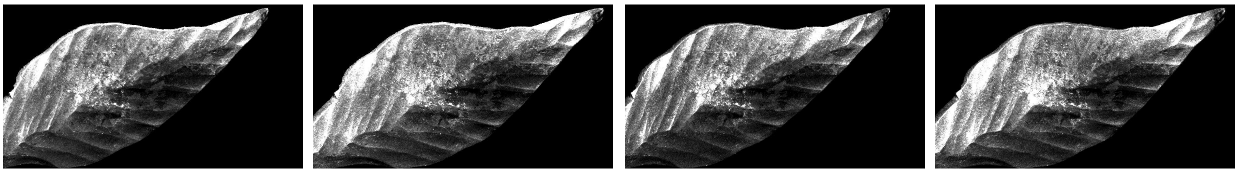
(a) LED 1.1 (b) LED 1.2



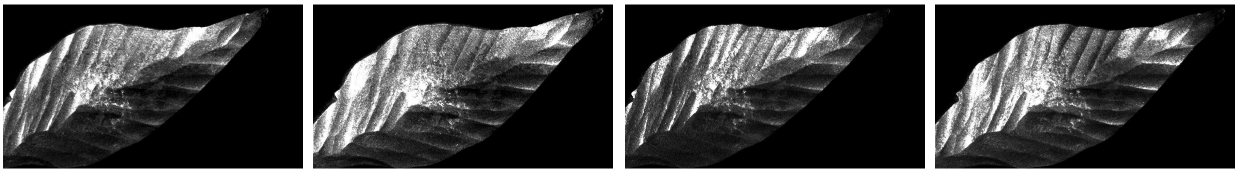
(c) LED 1.3 (d) LED 1.4



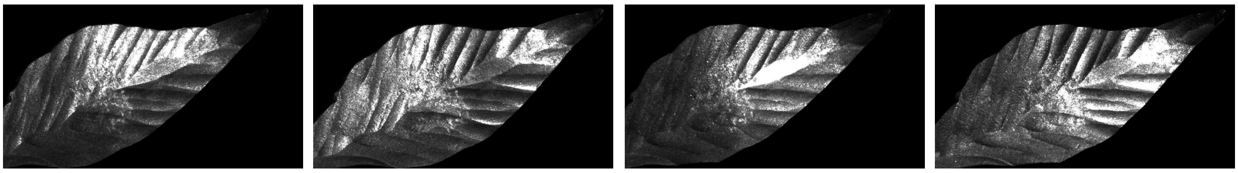
(e) LED 1.5 (f) LED 1.6



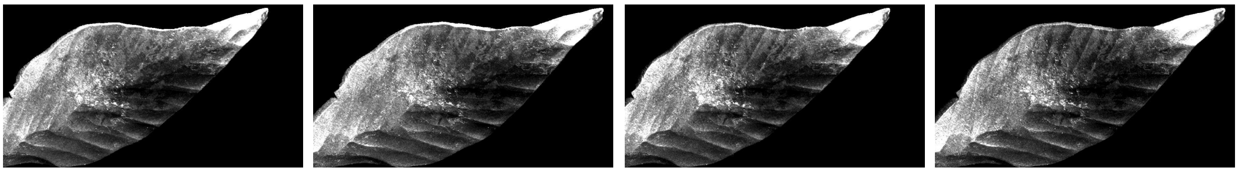
(g) LED 2.1 (h) LED 2.2



(i) LED 2.3 (j) LED 2.4

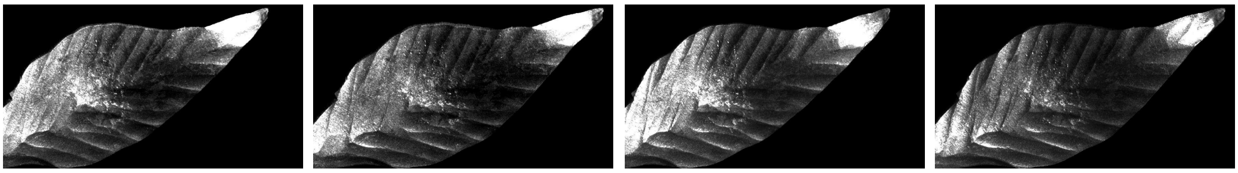


(k) LED 2.5 (l) LED 2.6



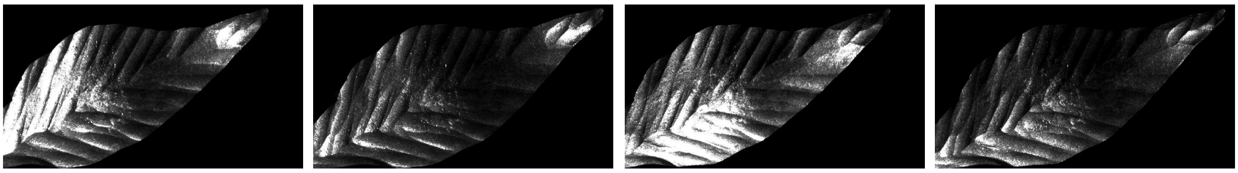
(m) LED 3.1 (n) LED 3.2

Figure A.9. Left and right stereo input images of the leaf dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.



(o) LED 3.3

(p) LED 3.4



(q) LED 3.5

(r) LED 3.6

Figure A.9 (*cont.*). Left and right stereo input images of the leaf dataset used to evaluate the Stereo Fusion algorithm. Same scaling in corresponding stereo pairs.