3D Shape Measurement and Reflectance Analysis for Highly Specular and Interreflection-Affected Surfaces

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

Steffen Herbort

Tag der mündlichen Prüfung:	28.05.2014
Hauptreferent:	Prof. Dr. rer. nat. Christian Wöhler
Korreferent:	Prof. Dr.-Ing. Franz Kummert

Arbeitsgebiet Bildsignalverarbeitung, Technische Universität Dortmund

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Christian Wöhler for profitable discussions, support, and the space to pursue my own ideas. Then, I'd like to express my gratitude to my second examiner Prof. Franz Kummert for his interest and the effort of providing the second evaluation. Further, I'd like to thank my colleague Arne Grumpe for several professional discussions and the great time working together for almost four years.
Several students contributed through their theses and feedback to publications, research, and not at least to a pleasant experience in supervising young professionals. These were Matthias Fiegel, Britta Gerken, Malte Lenoch, Eduard Schick, Daniel Schugk, Jens Westerhoff, and Tatsiana Zakreuskaya. In addition, I was fortunate to be supported by a bunch of very reliable student workers. These were Christian Leismann, Malte Lenoch, Sebastian Sudholt, Thorsten Wilhelm, and Tatsiana Zakreuskaya.
I enjoyed working at the TU Dortmund University because of the pleasant atmosphere among my colleagues and coworkers at AG BV, AG DT, and LS KT. Thank you all for a great time and good luck and much success with your research, theses, and your future.

Kurzfassung

Die Verfügbarkeit von 3D-Daten realer Objekte zur computergestützten Auswertung und Analyse bietet vielfältige Möglichkeiten für z.B. industrielle Qualitätskontrolle, Planeten-Fernerkundung, medizinische Diagnose, Tatorterfassung und Reverse Engineering. Die zur Objektvermessung oft eingesetzten 3D-Scanner weisen jedoch verschiedene Schwächen auf, wie z.B. Messrauschen bzw. fehlende Messpunkte auf spekularen Oberflächen oder inferreflektionsbehafteten Vertiefungen. Bildbasierte 3D-Oberflächenrekonstruktionsalgorithmen, wie z.B. Photometric Stereo, sind als Alternative zu 3D-Scannern weniger von den genannten Schwächen betroffen, neigen jedoch auf großen Skalen zu systematischen Abweichungen von der korrekten Objektform. Im Kern dieser Arbeit steht daher die Präsentation eines innovativen Ansatzes, der metrische 3D-Punktewolkendaten mit bildbasiert ermittelten Oberflächengradienteninformationen fusioniert, was eine signifikante Verbesserung der genannten Schwächen ermöglicht. Thematisch verwandte Arbeiten lassen dahingegen Interreflektionen oder nicht-Lambert'sche Oberflächen oft unberücksichtigt, arbeiten nur mit synthetischen Daten und/oder verwenden den unrealistischen Fall weit entfernter Szenenbeleuchtungslichtquellen.

Insgesamt werden in dieser Arbeit drei wichtige Beiträge zum Forschungsbereich Computer Vision vorgestellt: Erstens ein neuer Ansatz zur Kalibrierung von Lichtquellenintensitäten und -positionen unter Verwendung eines einzelnen Bildes einer diffus reflektierenden Kugel. Zweitens ein Ansatz, mit dem interreflektionsbehaftete Objekte mit unbekannten nicht-Lambert'schen Reflexionseigenschaften erfolgreich rekonstruiert werden können, indem 3D Punktewolkendaten mit bildbasiert ermittelten Oberflächengradienteninformationen iterativ fusioniert werden. Drittens ein innovativer Ansatz zur Segmentierung verschiedener Oberflächenmaterialien basierend auf Multispektraldaten.

Die erzielten 3D-Objektprofile zeigen jeweils deutliche Verbesserungen verglichen mit den Profilen aus einem industriellen 3D-Scanner ohne Bilddatenfusion. Diese Ergebnisse wurden qualitativ und quantitativ auf sowohl synthetischen als auch realen Daten ermittelt. Somit ist es nun möglich, dichte 3D-Profile von metallischen, dielektrischen und lackierten Oberflächen zu erstellen, die in ihrer Genauigkeit den von 3D-Scannern gut erfassbaren diffusen Oberflächen entsprechen.

Abschließend wird ein ausführlicher Ausblick gegeben und ein Prototyp zur Stereokamera-basierten 3D-Rekonstruktion vorgestellt. Dieser ermöglicht die Vermessung von 3D-Objekten ohne Verwendung eines kostenintensiven 3D-Scanners durch Korrespondenzfindung auf Gradientenfeldern in Verbindung mit Gradienten/Tiefendaten-Fusion.

Abstract

Transferring the 3D shape of real world objects into automatic data processing and examination machines is the basis for a wide range of promising applications like industrial quality inspections, lunar expeditions, medical diagnostics, crime scene investigations, and reverse engineering. However, commonly applied techniques for 3D data acquisition come with weaknesses alongside their strengths: While photogrammetric approaches like 3D scanners yield robust 3D point cloud data, they are commonly challenged by high frequency noise, specular surfaces regions, and interreflection-affected surface concavities. Image-based shape acquisition techniques like Photometric Stereo as an alternative to 3D scanners are less affected by those impairments, but turn out to be less robust on large scales. Therefore, this thesis presents an innovative approach that fuses metric 3D point cloud data with image-based gradient information, which strongly improves measurement noise, specular surfaces, and interreflection-affected regions. Past and recent work in that field rarely regards the case of interreflection-affected non-Lambertian surfaces, operates on synthetic data only, and/or applies the unrealistic simplification of distant light sources.

On that background, three main contributions have been made to the field of computer vision: First, a novel approach for the calibration of point light source intensities and positions using a single image of a diffusely reflecting sphere. Second, a 3D surface reconstruction approach that successfully handles even unknown non-Lambertian surface reflectance behavior and interreflections in a scheme that iteratively fuses 3D point cloud data and image-based surface gradient information. Third, an innovative technique for the segmentation of surface materials based on multispectral data.

The obtained 3D profiles show overall strong improvements compared to 3D measurement results of industrial 3D scanners without additional image data incorporation. These results have been assessed qualitatively and quantitatively on both, real and synthetic data. Overall, it is now possible to obtain dense 3D profiles of metallic, dielectric, and lacquered surfaces with an accuracy that was previously only available for the much simpler case of diffuse surfaces. Additionally, it is shown how metallic and colored surface materials can be segmented feasibly using multispectral data, which provides a basis for the 3D reconstruction of inhomogeneous surfaces.

Concludingly, a thorough outlook is given that contains the presentation of a prototype for stereo-camera-based 3D reconstruction, which successfully deals with the very challenging case of measuring specular surfaces without the need for a costly 3D scanner by using correspondence establishment on gradient fields in conjunction with gradient/depth data fusion.

Publications

This thesis is partly based on the following publications:

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

2. [Grumpe et al., 2011] A. Grumpe, S. Herbort, and C. Wöhler. 3D reconstruction of non-Lambertian surfaces with non-uniform reflectance parameters by fusion of photometrically estimated surface normal data with active range scanner data. *Oldenburger 3D Tage*, 10:54-61, 2011.

3. [Herbort et al., 2011] S. Herbort, A. Grumpe, and C. Wöhler. Reconstruction of non-Lambertian surfaces by fusion of Shape from Shading and active range scanning. *IEEE International Conference on Image Processing*, 1-4, 2011.

4. [Lenoch et al., 2012] M. Lenoch, S. Herbort, and C. Wöhler. Robust and accurate light source calibration using a diffuse spherical calibration object. *Oldenburger 3D Tage*, 11:1-8, 2012.

5. [Herbort and Wöhler, 2012] S. Herbort and C. Wöhler. Self-consistent 3D surface reconstruction and reflectance model estimation of metallic surfaces. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 1-8, 2012.

6. [Herbort et al., 2013b] S. Herbort, D. Schugk and C. Wöhler. 3D reconstruction of interreflection-affected surface concavities using photometric stereo. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 1-7, 2013

7. [Herbort et al., 2013a] S. Herbort, B. Gerken, D. Schugk, and C. Wöhler. 3D range scan enhancement using image-based methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84:69-84, 2013

8. [Bachmann et al., 2013] M. Bachmann, B. Gerken, T. Mager, C. Hedayat, S. Herbort and C. Wöhler. Integration eines photometrischen Stereoverfahrens in ein Laserscanner-System zur 3D-Oberflächenrekonstruktion. *Oldenburger 3D Tage*, 12:1-8, 2013.

9. [Schick et al., 2013] E. Schick, S. Herbort, A. Grumpe and C. Wöhler. Single view single light multispectral object segmentation. *International Conference on Computer Graphics, Visualization and Computer Vision*, 1:1-7, 2013.

Supervised theses I

The following theses from the fields automotive, mining, computer vision, and medical image processing have been (co-) supervised during the course of this thesis:

1. [Bhutta, 2011] Application of polarization data for 3d reconstruction. *Master thesis*, 2011.
2. [Birkholz, 2012] Entwicklung und Validierung eines Simulationsmodells für Laserscanner im Fahrerassistenzsystem. *Master thesis*, 2012.
3. [Bleich, 2012] Untersuchung von Verfahren zur Detektion von Gesichtern im Fahrzeuginnenraum auf basis von Infrarotaufnahmen. *Bachelor thesis*, 2012.
4. [Bleich, 2013] Klassifkation dynamischer Handgesten auf Basis raum-zeitlicher time-of-flight Kameradaten. *Master thesis*, 2013.
5. [Khan, 2011] Determination and evaluation of spatially-varying reflectance parameters of surfaces with known shape. *Master thesis*, 2011.
6. [Lange, 2012] Trainingsverfahren zur Bestimmung einer lokal adaptiven Kostenfunktion für ein aktives Oberfächenmodell zur Segmentierung dreidimensionaler Gehirnregionen. *Master thesis*, 2012.
7. [Nick, 2012] Prepress-management und Bildvorverarbeitung von Direktdruckapplikationen auf konischen Oberflächen. *Bachelor thesis*, 2012.
8. [Rüesing, 2012] Detektion befahrbarer Regionen für Automobile anhand von Flächenrekonstruktionen in Stereo-Bildern. *Master thesis*, 2012
9. [Schugk, 2010] Schachbrett- und Schachbretteckendetektion in Bildern zur Integration in Kamera-Kalibrierungsalgorithmen. *Bachelor thesis*, 2010
10. [Steinhoff, 2012] Identifkation von Fahrbahnbegrenzungen in Bildinformationen basiert auf der Flussvektoranalyse und dem inverse perspective mapping. *Master thesis*, 2012.
11. [Walzok, 2011] Konzeption und Implementierung einer vollautomatisierten Objektverfolgung mit Hilfe eines Multikamerasystems. *Master thesis*, 2011.
12. [Wilhelm, 2012] Vorverarbeitung, Registrierung und Fusion von 3D-Oberfächenscans. *Bachelor thesis*, 2012.
13. [Wörner, 2012] Sensorfusion zur Volumenstrommessung auf Laufband-Fördersystemen. *Master thesis*, 2012.
14. [Zakreuskaya, 2011] Optimierung und funktionale Erweiterung eines Kamera-Kalibrierungsverfahrens. *Bachelor thesis*, 2011.

Supervised theses II

The following theses that are closely related to the research presented in this thesis have been (co-) supervised during the course of this thesis:

1. [Lenoch, 2011] Analyse der Genauigkeit gängiger Beleuchtungskalibrierverfahren und Untersuchung der Auswirkung von Kalibrierungsungenauigkeiten auf 3D Rekonstruktionsalgorithmen. *Bachelor thesis*, 2011.

2. [Schugk, 2012] D. Schugk. Modellierung von Interreflexionen zur Integration in 3D-Oberflächenrekonstruktionsalgorithmen. *Master thesis*, 2012.

3. [Schick, 2012] E. Schick. Wellenlängenabhängige Reflektanzmodellierung verschiedener Oberflächenmaterialien anhand normalisierter Multispektraldaten. *Master thesis*, 2012.

4. [Gerken, 2012] Integration und Validierung eines photometrischen Stereoverfahrens zur Oberflächenrekonstruktion. *Bachelor thesis, University of Paderborn*, 2012.

5. [Fiegel, 2013] 3D-Rekonstruktion durch Fusion von Stereo-Tiefendaten mit photometrischen Oberflächeninformationen. *Master thesis*, 2013.

6. [Westerhoff, 2013] Kamerabasierte Untersuchung von räumlich variierenden Reflektanzeigenschaften von 3D Oberflächen mit Hilfe photometrischer Methoden. *Master thesis*, 2013.

7. [Lenoch, 2013] Quantitative comparison of surface reflectance models and evaluation of their effect on photometric stereo reconstructions. *Master thesis*, 2013.

Some of their content is used throughout this thesis; those places are marked accordingly.

Contents

1. Introduction **1**

1.1. 3D depth measurement . . . 2
1.2. Image-based shape determination and depth data fusion . . . 3
1.3. Major contributions of this thesis . . . 4
1.4. Fundamental notations and coordinate system definitions . . . 5
1.5. Image formation . . . 5

2. Related work - object appearance **9**

2.1. Radiance, irradiance, and surface albedo . . . 10
2.2. Bidirectional Reflectance Distribution Function . . . 10
2.3. Dichromatic reflectance model . . . 11
2.4. Model-based BRDFs . . . 13
2.4.1. Diffuse reflectance . . . 13
2.4.2. Empirical specular reflectance . . . 14
2.4.3. Microfacet-based specular reflectance . . . 20
2.4.4. Anisotropic reflectance . . . 21
2.4.5. Summary . . . 22
2.5. BRDF parameter estimation . . . 22
2.6. BRDF measurement and data-driven BRDFs . . . 24

3. Related work - 3D reconstruction **27**

3.1. Absolute depth determination . . . 28
3.2. Surface gradient determination . . . 29
3.2.1. Shape from Shading (SfS) . . . 30
3.2.2. Photometric Stereo (PS) . . . 31
3.3. Comparison of photogrammetric and photometric approaches . . . 32
3.4. Reconstruction of the depth profile from local gradients . . . 33
3.5. Generalizations of SfS and PS . . . 35
3.5.1. Non-distant light sources . . . 35
3.5.2. Unknown illumination conditions . . . 36
3.5.3. Perspective projections . . . 38

3.5.4. Non-Lambertian surfaces . . . 38
3.5.4.1. Purely specular reflections . . . 39
3.5.4.2. Linear combination of different reflection components . . . 39
3.5.4.3. Separation of specular and diffuse components . . . 40
3.6. Treatment of interreflections . . . 42
3.7. Multiplexed photometric stereo . . . 44
3.8. 3D scan enhancement by fusion with photometric cues . . . 45
3.8.1. Iterative error functional minimization . . . 45
3.8.2. Filter-based approaches . . . 47
3.8.3. Linear problem statement . . . 47
3.8.4. Further approaches . . . 48

4. Experimental setup **51**
4.1. Setup description . . . 52
4.2. Camera calibration and configuration . . . 53
4.3. Robust high dynamic range (HDR) imaging . . . 54

5. Contribution - light source calibration **55**
5.1. Mathematical description . . . 57
5.2. Calibration object . . . 57
5.3. Light source calibration algorithm . . . 58
5.4. Experimental results . . . 59
5.4.1. Image and depth data . . . 60
5.4.2. Image data only . . . 61
5.5. Conclusion . . . 62

6. Contribution - surface reconstruction **63**
6.1. Absolute depth and gradient field fusion . . . 64
6.1.1. Variational Shape from Shading . . . 64
6.1.2. Inclusion of absolute depth data . . . 65
6.1.3. Computational efficiency . . . 66
6.1.4. Reflectance map derivatives under specular reflection . . . 67
6.1.5. Treatment of interreflections . . . 68
6.2. Simultaneous 3D surface and BRDF estimation for convex non-Lambertian surfaces . . . 69
6.2.1. Algorithm . . . 70
6.2.2. Results . . . 70
6.2.3. Synthetic evaluation . . . 73
6.2.4. Conclusion . . . 74
6.3. Photometric Stereo for interreflection-affected surfaces . . . 75
6.3.1. Algorithm . . . 76
6.3.2. Results . . . 77
6.3.2.1. Progressing change in reconstructed depth . . . 78
6.3.2.2. Qualitative evaluation . . . 78

6.3.2.3. Quantitative evaluation . . . 79
6.3.2.4. Convergence criterion . . . 80
6.3.3. Conclusion . . . 80
6.4. High accuracy surface scan enhancement . . . 81
6.4.1. Algorithm . . . 81
6.4.2. Experimental data . . . 82
6.4.3. Results . . . 83
6.4.3.1. Choice of the reflectance model . . . 83
6.4.3.2. Caliper object examination . . . 84
6.4.3.3. Application to various surface types . . . 86
6.4.4. Conclusion . . . 88

7. Contribution - spatially varying BRDFs 89
7.1. Multispectral object segmentation . . . 91
7.2. Algorithm . . . 92
7.2.1. Data preparation . . . 92
7.2.2. BRDF measurement . . . 92
7.2.3. Clustering and BRDF fitting . . . 93
7.2.4. Spectral unmixing . . . 94
7.3. Experiments and results . . . 96
7.3.1. Object with sharp region boundaries . . . 96
7.3.2. Object with gradual region boundaries . . . 99
7.4. Conclusion . . . 99

8. Outlook, conclusion and future work 101
8.1. Outlook . . . 102
8.1.1. Algorithm . . . 102
8.1.2. Data acquisition and Photometric Stereo . . . 103
8.1.3. Triangulation stereo and filling . . . 103
8.1.4. Fusion . . . 104
8.1.5. Accuracy evaluation . . . 104
8.2. Conclusion . . . 106
8.3. Future work . . . 106

A. Detailed description of typical coordinate systems applied for computer vision 123

B. Object appearance 125
B.1. Influence of distant and non-distant lightsources . . . 125
B.2. Generalized reflectance distribution functions . . . 126
B.3. Mathematic derivations . . . 126
B.3.1. Physical plausibility of BRDF models . . . 126
B.3.1.1. Lambert . . . 126
B.3.1.2. Modified Phong . . . 127

B.3.1.3. Modified Blinn . . . 128
B.4. Further details for less common reflectance models . . . 130
B.4.1. Minnaert . . . 130
B.4.2. Neumann-Neumann . . . 130
B.4.3. Beckmann-Spizzichino and Torrance-Sparrow . . . 130
B.4.4. He-Torrance . . . 131
B.5. Anisotropic reflectance models . . . 132
B.5.1. Ward . . . 132
B.5.2. Ashikhmin-Shirley . . . 133
B.5.3. Kurt . . . 133

C. Calibration **135**
C.1. Camera . . . 135
C.1.1. Dark frame compensation and hot pixels examination . . . 135
C.2. Rotation axis calibration . . . 136
C.2.1. Rotation axis calibration of an axis parallel to the rotated plane . . . 136
C.2.2. Rotation axis calibration of an excentric axis . . . 137

D. Usage permission(s) **139**
D.1. Blender . . . 139

1 Introduction

Transferring the 3D shape of real world objects into automatic data processing and examination machines bears on the one hand great benefits due to the wide range of applications, but it is on the other hand challenging due to the complexity of the 3D coordinate measurement problem. The following chapter explains how typical industrial 3D range scanners acquire real world 3D object shape data and why these are oftentimes corrupted by noise and missing measurements. It will become clear that an improvement of these aspects is desirable and how this can be achieved by combining 3D range scanner depth data with image-based photometric surface information.

This chapter is partly adapted and/or adopted from [Herbort et al., 2013a]

The measurement of spatial 3D coordinates from objects has been the basis for a wide range of scenarios like lunar expeditions [Chin et al., 2007], cultural heritage [Pavlidis et al., 2007], industrial quality inspections [Malamas et al., 2003], augmented reality [Azuma, 1997], video game control [Leyvand et al., 2011], medical diagnostics [D'Apuzzo and Mitchell, 2008], face recognition [Mian and Pears, 2012], gesture recognition [Richarz, 2012], crime scene investigations [Buck et al., 2013], and reverse engineering [Son et al., 2002]. Solutions for the problem of 3D coordinate determination exist in a large variety with individual advantages and weaknesses. These are e.g. mechanical coordinate measurement [Ashley et al., 1970], active light projection [Shirai and Suwa, 1971; Knauer et al., 2004; Peng, 2006; Bartczak and Koch, 2009], sound and radio wave reflection [Elfes, 1987; Daniels, 1996], x-ray transmission [Farmer and Collins, 1971], stereo camera triangulation [Okutomi and Kanade, 1993], motion [Tomasi and Kanade, 1992], and defocus [Subbarao and Surya, 1994]. This obviously broad applicability has motivated several branches of research that focused on technical and algorithmic developments, which aim for an improved accuracy and resolution of 3D acquisition systems [Wiora, 2001; Blais, 2004; Remondino and El-Hakim, 2006].

1.1 3D depth measurement

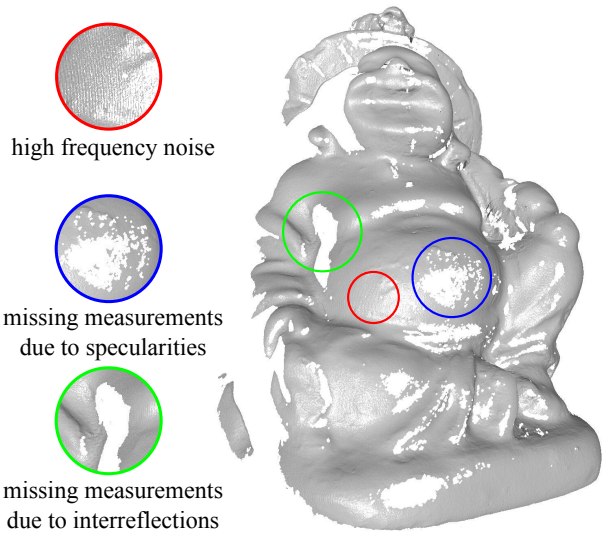
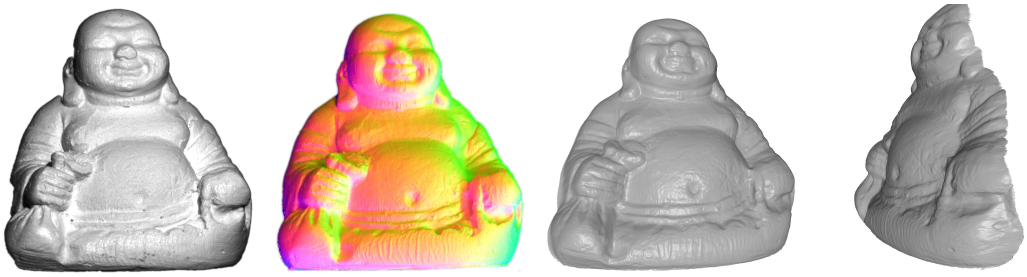


Figure 1.1.: Shaded 3D scan obtained with an industrial laser range scanner showing high frequency noise (red), missing measurements at specular highlights (blue), and missing measurements at interreflection-affected surface concavities (green).

Some of the most widespread technologies for 3D shape acquisition are (active) 3D laser triangulation scanners and (passive) stereo camera systems due to their direct practical applicability and reasonable costs. However, the 3D scan quality of active triangulation systems is limited by three main aspects (see Fig. 1.1): First, there is a considerable amount of high frequency noise present due to the neighborhood-independent measurement principle. Second, there are problems when handling light that is reflected specularly from the object surface due to either missing or saturated camera responses. As a consequence, there are no measurements available and the scan becomes holey. Third, the occurrence of interreflections impairs the scan and ultimately causes missing measurements, albeit for a different reason than for specular surface reflections.

Passive systems like stereo cameras show the same problem of high frequency noise occurrence, but face further problems for non-diffuse surfaces, since those appear significantly different when being observed from the different viewpoints. It is thus desirable to develop methods that improve the accuracy of 3D measurements obtained with an active laser range scanner or a passive stereo camera system. These are supposed to provide dense, i.e. hole free, surfaces with at best no spatial high frequency noise.

1.2 Image-based shape determination and depth data fusion



(a) exemplary grayscale image **(b)** color-coded surface normals **(c)** shaded depth profile front **(d)** shaded depth profile side

Figure 1.2.: Illustration of Photometric Stereo: A set of gray scale images (a) is analyzed in its appearance, which allows the extraction of surface normals for each pixel using Photometric Stereo (b). These can be integrated to a depth profile (c). Note, that photometric approaches are prone to systematic errors on large scales, which causes the depth profile in this example to be too flat (d).

While light-projection-based range scanners and stereo cameras are well known beyond the field of 3D computer vision, it is oftentimes disregarded that object shapes can be determined as well with a single camera based on photometric cues using e.g. Shape from Shading [Horn, 1970, "SfS"] or Photometric Stereo [Woodham, 1980, "PS"]. Instead of the geometric triangulation principle, SfS and PS analyze the intensity of light being reflected by an object (see Fig. 1.2(a)), which allows the determination of local surface gradients (see Fig. 1.2(b)) that can then be integrated to a depth profile (see Fig. 1.2(c)).
The main advantages of photometric approaches compared to triangulation-based approaches are that they are significantly less affected by local high frequency noise and that usually no measurements are missing. Unfortunately, their disadvantage concerns the fact, that the determined surface oftentimes contains low frequency errors that deform the overall shape of the estimated surface (see Fig. 1.2(d)). This is caused by systematic gradient estimation uncertainties that add up during the gradient-to-depth integration process (see Section 3.4). Nevertheless, it is important to emphasize that triangulation-based shape determination and photometric approaches are capable of mutually enhancing their individual disadvantages: While triangulation-based 3D data on the one hand is accurate on large scales, it does contain high frequency noise. Image-based techniques on the other hand suffer from large scale deviations, but provide a significantly less noisy high frequency component.
Past and recent publications in the field of surface reconstitution and depth/gradient fusion focus mainly on purely Lambertian surfaces that are free of interreflections and the authors usually apply gradient estimation and fusion process separately (see Chapter 3). In contrast, the approach presented in this thesis reconstructs depth and gradients simultaneously for the case of real non-Lambertian surfaces, which are even allowed to be convex and thus exhibit interreflections. Further, the presented evaluation regards absolute depth reconstruction accuracy and the applicability of depth/gradient fusion to 3D measurement tasks. This is considerably more complex than the typical approach of qualitative surface appearance examination used by most authors. These and further contributions of this thesis are introduced properly in the following section.

1.3 Major contributions of this thesis

Initially, this thesis gives an extensive review of former and recent publications in the fields of computer vision, computer graphics, and photogrammetry. This aims for a more complete understanding of those topics and, additionally, it illustrates the significance of the contributions that have been made to these fields. The contributions regard the following aspects:

(1) Single-image light source position and intensity calibration.
(2) Estimation and refinement of BRDF parameters within surface reconstruction frameworks.
(3) Handling interreflections in Photometric Stereo frameworks.
(4) Refinement of 3D scans of interreflection-affected non-Lambertian object surfaces by fusion with photometric cues.
(5) Multispectral material analysis and segmentation.

Overall, this thesis aims on a description of how noisy and holey 3D scans of real surfaces obtained with an industrial 3D scanner can be improved using photometrically motivated approaches. In contrast to previous work in those fields, this is applied to non-Lambertian surfaces with initially unknown reflectance behavior and possible occurrence of interreflections.
After the now following overview of some fundamental concepts, the relevant past and recent work for object appearance (Chapter 2) and 3D surface reconstruction (Chapter 3) is reviewed. Then, the experimental setup used for the acquisition of experimental data is described (Chapter 4), followed by the contributions to the fields of light source calibration (Chapter 5), surface reconstruction (Chapter 6), and multispectral material analysis (Chapter 7). Afterwards, the thesis is concluded and a thorough outlook is given (Chapter 8).

1.4 Fundamental notations and coordinate system definitions

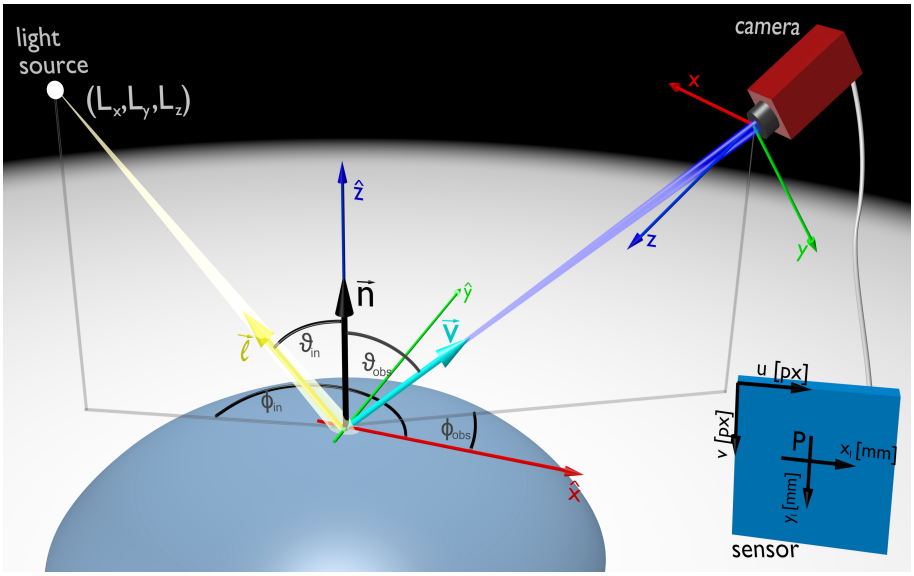


Figure 1.3.: Local scene geometry with camera coordinate system (x, y, z), local coordinate system $(\hat{x}, \hat{y}, \hat{z})$ and definitions for incident light direction $\mathbf{l}$, surface normal $\mathbf{n}$ and view direction $\mathbf{v}$.

Fig. 1.3 shows an overview of a typical single-camera computer vision scene with the most relevant coordinate systems and relevant directions therein. The main coordinate system ("CS") is the camera CS (x, y, z) located in the optical center (lens center) of the camera with the z axis pointing outwards and the x axis being aligned with the horizontal image axis. The y axis is then $y = z \times x$. Further important coordinate systems are the local CS $(\hat{x}, \hat{y}, \hat{z})$, the image CS (x_i, y_i) and the sensor CS (u, v), which are described below. Each surface point can be assigned a designated surface normal $\mathbf{n} = (n_x, n_y, n_z)$. The local viewing direction $\mathbf{v} = (v_x, v_y, v_z)$ for each surface point and the local incident light direction $\mathbf{l} = (l_x, l_y, l_z)$ point to the camera and light source, respectively. The light source is located in the camera CS at (L_x, L_y, L_z). All vectors $\mathbf{n}$, $\mathbf{v}$ and $\mathbf{l}$ are normalized to unit length. With respect to the local coordinate system, they can as well be described by their azimuth ϕ and polar angle ϑ. The subscripts in and obs denote incident and observance directions, respectively. More detailed descriptions of these coordinate systems are given in Appx. A.

1.5 Image formation

Essentially, image-based surface reconstruction approaches like SfS and PS reverse the process of how an image is generated in order to extract shape information from the images of a scene. Since this is the main topic of this thesis, it is important to understand the physical mechanisms that contribute to the formation of a 2D image as accurately as possible.

For this matter, Fig. 1.4 illustrates the components that jointly contribute to the 2D appearance of a scene. Emitted light from a light source travels through the scene environment where it interacts with participating gaseous and/or other optically active media. The light rays then hit the object surface where they are reflected and/or transmitted based on the material's light interaction properties. Light reflected towards

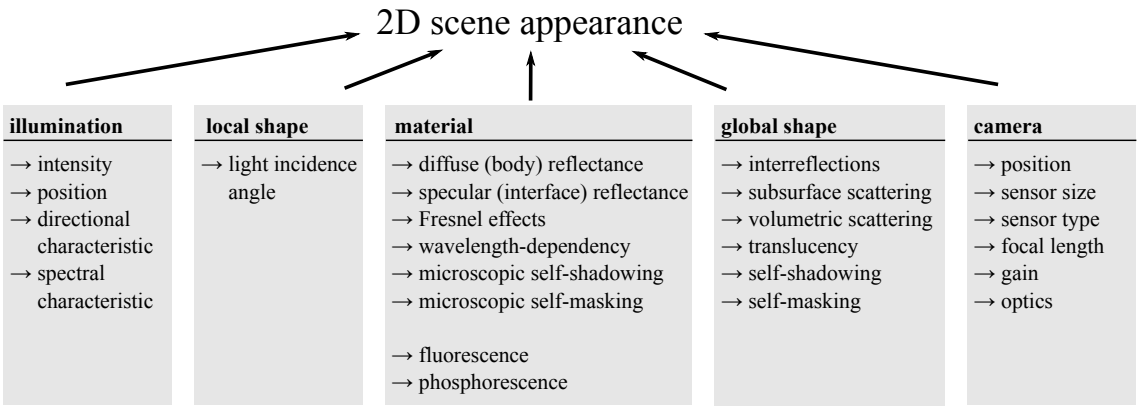
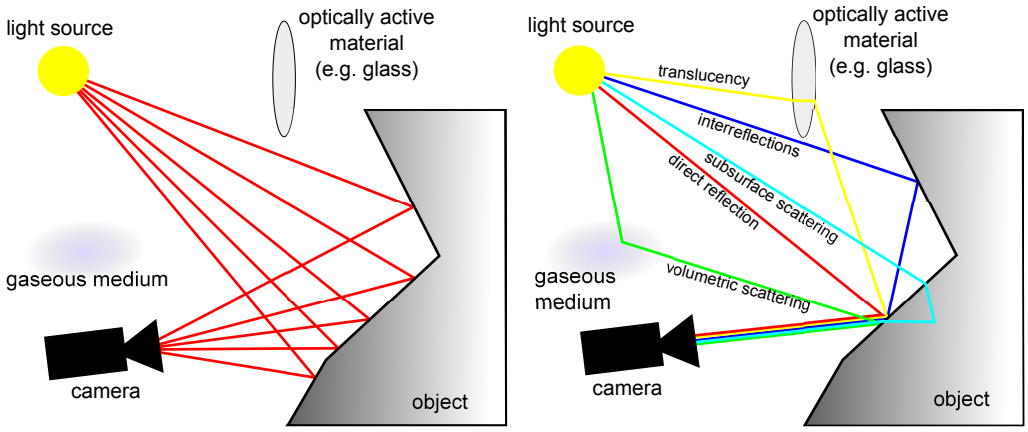


Figure 1.4.: Detailed overview of the components that contribute to the photometric appearance of a physical scene.

the camera is perceived on the camera sensor, depending on camera position, camera lens properties, focal length, sensor size, sensor position, etc. This illustrates the dimensionality of the image formation problem and gives an insight into the very complex interdependence between light, camera, shape, and material.



(a) direct illumination **(b)** global illumination

Figure 1.5.: The real appearance of a surface is a superposition of plenty of reflection effects. The "direct illumination" component alone (see (a)) is used for simplicity in several computer vision algorithms, but lacks physical accuracy. "Global illumination" (see (b)) models the influence of object geometry, subsurface scattering and optically active media between light source and object (or between object and camera). Image adopted from [Nayar et al., 2006; Schugk, 2012].

This description of the image formation process did not distinguish between (1) local material reflectance properties and (2) global light/object interactions, which jointly influence the object's appearance on its image. However, both components are important and need to be distinguished carefully: "Material reflectance" only refers to the isolated process of how light is reflected by an infinitesimally small surface patch from some incident direction to some reflection direction. In Computer Graphics, this simplification of the actual light/scene interaction is termed "direct illumination" (see Fig. 1.5(a)). Global light/object interactions,

where the light that is incident upon a surface patch possibly originates from anywhere in the scene and not only directly from the light source(s), are termed "global illumination" (see Fig. 1.5(b)). Global illumination mainly regards interreflections, subsurface scattering, translucency, and volumetric scattering [Nayar et al., 2006]. An example for direct illumination separated from interrefletions in an image can be found in Section 6.3.

For a computational simulation of these effects, there are ray tracing algorithms available, which compute an image based on scene, geometry and material. Usually, these solve the rendering equation by Kajiya [1986] in a computationally efficient way. For physically based examinations, as they are required for 3D surface reconstructions, it is crucial to use scene renderers that specifically account for energy conservation and model the reflection effects with physical plausibility (see Section 2.1). One of those tools is PBRT[1], published by Pharr and Humphreys [2010], which has been used for the experiments conducted during the course of this thesis. In Computer Graphics, there are new methods available that aim for real-time rendering global illumination, see e.g. [Papaioannou, 2011]. Global illumination as such is an increasingly important field in Computer Graphics and Computer Vision, since e.g. 3D surface reconstruction algorithms are usually based on direct illumination only while global effects are neglected, which at least decreases their accuracy (see Section 6.3). However, it remains a helping technique throughout this thesis and descriptions on how rendering algorithms work in detail thus need to be obtained from the literature. An extensive review of global illumination techniques has been issued by e.g. Ritschel et al. [2012].

[1] Physically Based Ray Tracer, http://pbrt.org

2 Related work - object appearance

It is of great importance for image-based surface reconstruction algorithms, as they are used in the following chapters, to understand how the appearance of an object is created by the light that is reflected from the surface. On that basis, it is possible to reverse the image formation process and estimate surface normals based on observed intensities, which then allows the extraction of object shape information from object images.
The following sections explain mathematical models for the description of how a surface reflects light and illustrates which mechanisms contribute to the amount of light being perceived by a camera or a human observer of a physical scene.

This chapter is partly adapted and/or adopted from [Herbort and Wöhler, 2011]

The image formation process explained in Section 1.5 distinguished between direct and global illumination components. The key principle used for surface reconstruction usually relies on the direct illumination part (see Chapter 3) and the chapter at hand thus focuses on a detailed description of that type. The consideration of interrreflections as one case of global illumination is explained thoroughly in Section 6.3.
The initial definitions of radiance, irradiance, and surface albedo (Section 2.1) are followed by a standardized approach for reflectance description (Bidirectional Reflectance Distribution Function, BDRF, Section 2.2), and the dichromatic reflection model for the description of body and interface reflectance (Section 2.3). Section 2.4 then extensively reviews model-based reflectance description with special attention upon physical plausibility. Note, that the model overview is extensive and contains some additional models to those applied during the experiments conducted during the course of this thesis. The section closes with a description of the BRDF parameter estimation problem (Section 2.5), and the concept of BRDF measurement for data-driven BRDFs (Section 2.6).

2.1 Radiance, irradiance, and surface albedo

The radiance L measures the power of the light that is emitted from a certain surface area dA, in a specific reflection direction (ϑ_r, ϕ_r), and over a certain solid angle $d\Omega$:

$$L(\vartheta_r, \phi_r) = \frac{d\Phi_r(\vartheta_r, \phi_r)}{dA \cdot d\Omega \cdot \cos(\vartheta_r)} \qquad \left[\frac{W}{m^2\ sr}\right]. \tag{2.1}$$

$\Phi(\vartheta_r, \phi_r)$ denotes the radiant flux in W emitted in the reflection direction.
The irradiance E measures the power of the light that falls on a certain surface area dA:

$$E(\vartheta_{in}, \phi_{in}) = \frac{d\Phi_{in}(\vartheta_{in}, \phi_{in})}{dA} \qquad \left[\frac{W}{m^2}\right]. \tag{2.2}$$

The surface albedo ρ of a surface point (x, y, z) is defined as the ratio between the total reflected light energy Q_r and the total incident light energy Q_i

$$\rho(\phi_{in}, \vartheta_{in}, \lambda, x, y, z) = \frac{Q_r}{Q_{in}(\phi_{in}, \vartheta_{in}, \lambda, x, y, z)} \qquad [1]. \tag{2.3}$$

With Q_r and Q_{in} measured in Joules $[J]$ and $Q_r \leq Q_{in}$ for opaque materials, the albedo ρ becomes a dimensionless unit with $\rho \in [0...1] \subset \mathbb{R}^+$. The albedo measures how much light becomes reflected collectively from a given surface point into the upper hemisphere, which usually depends on the incident light direction. The amount of absorbed light is then simply $1 - \rho$.

2.2 Bidirectional Reflectance Distribution Function

The ratio between the light that is incident from a certain direction $(\vartheta_{in}, \phi_{in})$ upon a single surface patch and the light that is reflected into a certain direction (ϑ_r, ϕ_r) can be described by a 4-dimensional function,

the Bidirectional Reflectance Distribution Function (BRDF):

$$f(\vartheta_{in}, \phi_{in}; \vartheta_r, \phi_r) = \frac{dL(\vartheta_r, \phi_r)}{dE(\vartheta_{in}, \phi_{in})} = \frac{dL(\vartheta_r, \phi_r)}{L(\vartheta_{in}, \phi_{in}) \cos \vartheta_{in} d\Omega_{in}} \qquad \left[\frac{1}{sr}\right]. \tag{2.4}$$

The term *bidirectional* implies that two directions, namely incident and reflectance direction, are involved. While the reflectance direction (ϑ_r, ϕ_r) *may* coincide with the observance direction $(\vartheta_{obs}, \phi_{obs})$, e.g. for the case of direct illumination, this is not the case in general. The BRDF thus describes reflection into any direction in the upper hemisphere, but only few of these are actually observed.
Note that the explicit dependence on the azimuth angles ϕ_{in} and ϕ_r includes the case of anisotropic reflectance. For the case of *isotropic* reflectance, the function decreases to three dimensions and the dependence from ϕ_{in} and ϕ_r reduces to the dependence on their absolute difference $|\phi_{in} - \phi_r|$, such that $f_{\text{iso}} = f(\vartheta_{in}, \vartheta_r; |\phi_{in} - \phi_r|)$.
Physically, there are two intrinsic properties of a BRDF which are universally valid. These are (1) Helmholtz reciprocity [von Helmholtz, 1924] and (2) energy conservation, e.g. [Lewis, 1994]. Helmholtz reciprocity states, that the observed light intensity remains unchanged if the incident light direction and the observance direction are exchanged, i.e.

$$f(\vartheta_{in}, \phi_{in}; \vartheta_r, \phi_r) = f(\vartheta_r, \phi_r; \vartheta_{in}, \phi_{in}). \tag{2.5}$$

Energy conservation requires that the total energy of the reflected light can not be larger than the incident energy, which means that the total hemispherical reflectance is limited by 1:

$$\int_{\Omega} f(\vartheta_{in}, \phi_{in}; \vartheta_r, \phi_r) \underbrace{\langle \mathbf{n} \cdot \mathbf{v} \rangle}_{\cos \vartheta_r} d\Omega_r \leq 1, \tag{2.6}$$

where $\mathbf{v}$ denotes the reflection direction. Note, that Eq. 2.6 assumes the observed surface to be nonemissive (does not radiate any light by itself) and opaque (no light can travel through or within the object).
While a BRDF describes the ratio between incident and reflected light, it disregards the incident light intensity I_0 and the influence of the light travel distance from a point light source to the object (r_l). Using the generalized notation of an arbitrary Reflectance Distribution Function $f(\vartheta_{in}, \phi_{in}, \vartheta_r, \phi_r, ...)$, this means that the observed intensity I_{obs} becomes

$$I_{obs} = \frac{I_0}{r_l^2} \, f(\vartheta_{in}, \phi_{in}, \vartheta_{obs}, \phi_{obs}, ...) \, \langle \mathbf{n} \cdot \mathbf{l} \rangle \, . \tag{2.7}$$

More generalized parameterizations of RDFs regarding wavelength-dependent reflectance, spatial variations, and subsurface scattering can be found in Appx. B.2.

2.3 Dichromatic reflectance model

Surface reconstruction algorithms oftentimes focus on diffusely reflecting surfaces only (see Chapter 3). However, for most real surfaces, there are at least *two* reflectance phenomena that are apparent. The first is termed *diffuse* or *body* reflection, the second *specular* or *interface* reflection. Their difference has been analyzed thoroughly by Shafer [1985], who then introduced the dichromatic reflection model for their

respective description. Fig. 2.1 illustrates the characteristics of both reflection types: Body reflections on

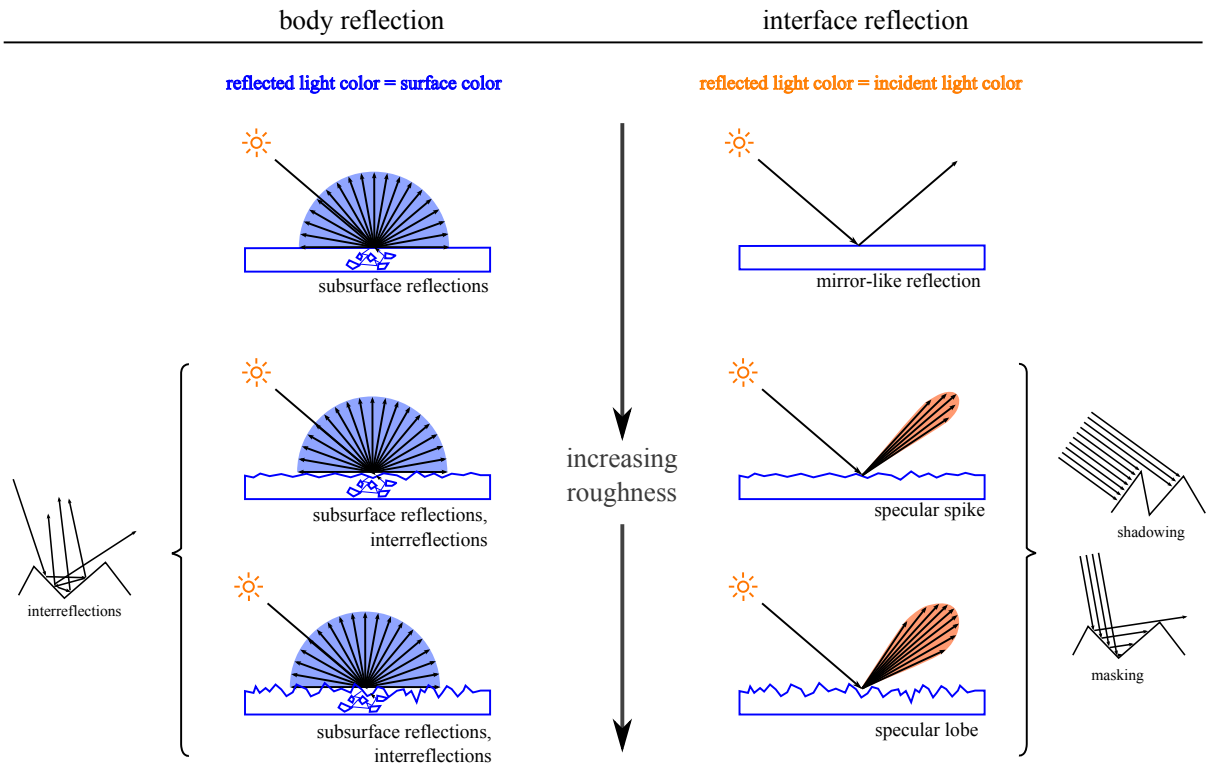


Figure 2.1.: Explanation of the difference between specular (interface) and diffuse (body) reflectance. While the incident light color remains unchanged under interface reflection, the wavelengths of the reflected light *do* depend on the surface color for the case of body reflectance. Image adapted from [Bajcsy et al., 1996].

the one hand are governed by light/material interactions where the incident light is partly absorbed and partly reflected as a result of a large number of microscopic interreflection and subsurface scattering events (see Fig. 2.1, right). The reflected light is thus "filtered" by repetitively interacting with the surface and the material, which causes the light color (wavelength spectrum) of the reflected light to be different from that of the incident light. After a large number of scattering events, the light is reflected away from the surface into the upper hemisphere. An increasing surface roughness adds interreflections on a more macroscopic scale, but has little influence on the directional characteristic of the reflected light.

Interface reflections on the other hand are caused by electromagnetic interactions between incident light and surface at the interface between the light-carrying medium (e.g. air) and the object. This causes the perceived light color (wavelength) of specular reflections to be the same as that of the incident light. As a secondary effect, it can be observed that the angular extent under which the reflected light increases with the surface roughness.

Further characteristics of specular reflections have been derived by [Nayar et al., 1991] through an analysis of the model by Beckmann and Spizzichino [1963]. That model is comprised of an electromagnetic-wave-theory based description of how plane, perpendicularly polarized light with a large wavelength compared to the surface roughness σ is reflected from a perfectly conducting surface without the regard of self-shadowing, self-masking, or interreflections. The evaluation yields the following four main insights: First, the surface reflects light under the occurrence of a broad specular lobe and a superimposed narrow specular

peak. Second, specular reflections show *off-specular* lobe peaks. Third, specular spikes occur in addition to the specular lobe only significantly if $\frac{\sigma}{\lambda} < 1.5$. Fourth, the specular lobe dominates if $\lambda \ll \sigma$. Details can be found in Appx. B.4.3.
These characteristics will become relevant throughout the course of this thesis. The description of diffuse and specular reflections is usually performed by reflection models, which are overviewed in the following sections.

2.4 Model-based BRDFs

Fig. 2.2 gives an overview of BRDF models known from the literature. Since physical plausibility becomes

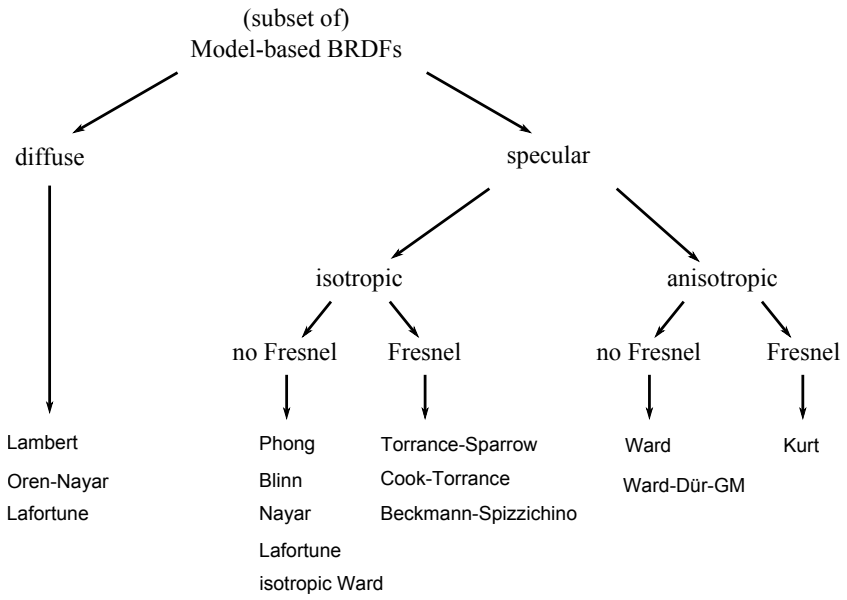


Figure 2.2.: Structured overview of different model-based reflectance models (BRDFs).

important in the following chapters, this is paid close attention to for each model.
The set of BRDFs that model specular surface appearance range from empirically motivated models like [Phong, 1975; Blinn, 1977; Nayar et al., 1990a; Lafortune et al., 1997] over models based on electromagnetic wave theory like [Beckmann and Spizzichino, 1963] to models based on geometrical optics [Torrance and Sparrow, 1967; Cook and Torrance, 1981; He et al., 1991]. The principle of the latter models lies in the assumption that the object surface is composed of a collection of microfacets. Works that summarize, compare, and/or analyze different reflection models are e.g. [Ngan et al., 2005], [Dorsey et al., 2008], and [Montes and Urena, 2012].

2.4.1 Diffuse reflectance

The most common models for *diffuse* reflectance are Lambert [1760] and Oren and Nayar [1994]. These are described and analyzed regarding their applicability in the following. The approach by Minnaert [1941] is seldom applied and a summary can be found in Appx. B.4.1.

Lambert

The property that incident light is reflected by the surface equally into all possible directions in the upper hemisphere has first been regarded by Lambert [1760]. Demanding energy conservation (Eq. 2.6) leads to a widely applied BRDF for strictly diffuse reflections

$$f_{\text{diffuse}} = \frac{\rho}{\pi}. \tag{2.8}$$

A detailed mathematical derivation can be found in Appx. B.3.1. This model is thus comprised of a constant scalar and the only parameter is the (diffuse) albedo $\rho \in [0...1] \subset \mathbb{R}$. The model is suitable for smooth non-emitting opaque surfaces, where the reflections are caused by microscopic subsurface scatterings and interreflections as described in Section 2.3. Note, that the independence of the observance direction indicates that the amount of perceived light remains unchanged even if a point is observed from a different viewpoint.
Due to its simplicity, there are lots of applications in the field of 3D surface reconstruction that build upon that model. A thorough review of several of those approaches is given in Chapter 3, but it is stressed again here that most real surfaces are insufficiently described by Lambertian reflectance alone due to the missing specular reflection component.

Oren-Nayar

Oren and Nayar [1994] generalize Lambert [1760] from smooth opaque surfaces to rough opaque surfaces, such as plaster or concrete. Their proposed BRDF is

$$\begin{aligned} f_{ON} &= \frac{\rho}{\pi}\left[1 - \frac{0.5\,\sigma^2}{\sigma^2 + 0.33} + \frac{0.45\,\sigma^2}{\sigma^2 + 0.09} \cdot \max\left[0, \cos(\phi_{in} - \phi_{obs})\right] \sin\gamma \tan\delta\right] \\ \text{with } \gamma &= \max(\vartheta_{in}, \vartheta_{obs}), \\ \delta &= \min(\vartheta_{in}, \vartheta_{obs}), \\ \sigma &= \text{surface facet slope }, \sigma \in [0...\frac{\pi}{2}] \subset \mathbb{R}^+, \end{aligned} \tag{2.9}$$

which requires the surface facet slope parameter σ in addition to the surface albedo ρ. σ is determined as the standard deviation of the angle between the microfacet normals and the macroscopic surface normal. The case $\sigma = 0$ yields $f_{ON} = \frac{\rho}{\pi}$, which is exactly the Lambertian case. The modification is motivated by self-shadowing effects, self-masking effects, and interreflections, which occur at microscopic v-shaped concavities on rough surfaces. The model is thus not only dependent on the incident direction (ϑ_{in}, ϕ_{in}), but as well on the observance direction (ϑ_{obs}, ϕ_{obs}). Compared to the Lambertian case, it can be noted that the resulting reflectivity decreases less for increasing incidence angles.
Although being a more generalized case, there have only been few publications that feature the Oren-Nayar model for 3D reconstructions, see e.g. [Ju et al., 2012; Abdelrahim et al., 2012].

2.4.2 Empirical specular reflectance

In the following, typical approaches for empirically modeling specular reflectance are discussed, which regards perfect mirrors, the pioneering work for specular reflections by Phong [1975], its variation by

Blinn [1977], the lobe/spike superposition by Nayar et al. [1990a], the generalized cosine lobe model by Lafortune et al. [1997], and the separable model by Neumann et al. [1999]. Each section examines physical plausibility whose importance has been stressed by Lewis [1994] and is required for the correct computation of interreflections as required in Section 6.3. Note, that several of the now following BRDFs are cosine-lobe models of the form

$$f(\beta) = \cos^m(\beta), \tag{2.10}$$

which inconveniently exhibit only minor changes in their value for large exponential factors. When these factors are examined e.g. in the context of Phong, Blinn, or Nayar reflectance (see below), it is thus more meaningful to examine the "full width at half maximum" (FWHM) $\beta_{\frac{1}{2}} = \arccos(\sqrt[m]{0.5})$ of the trigonometric function, which denotes the angle at which the cosine has decreased to the value 0.5. For raytracing, it is common to use some empirical surface roughness $e = \frac{1}{m}$and claiming $m \geq 1$, which is the case for reasonable highlights anyway.

Perfect mirror

The BRDF for a perfect mirror is

$$f_M = \delta_{\mathbf{r},\mathbf{v}} = \begin{cases} 1 & \text{if } \mathbf{r} = \mathbf{v} \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

with the Kronecker-Delta $\delta_{\mathbf{r},\mathbf{v}}$, viewing direction $\mathbf{v}$, and direction of mirror reflection $\mathbf{r} = \mathbf{M} \cdot \mathbf{l}$, where $\mathbf{M} = 2\mathbf{n}\mathbf{n}^T - \mathbb{I}$ denotes a Householder transform for mirroring the incident light direction $\mathbf{l}$ on a plane with

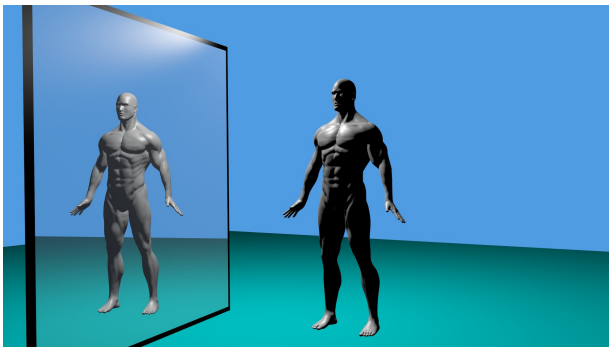
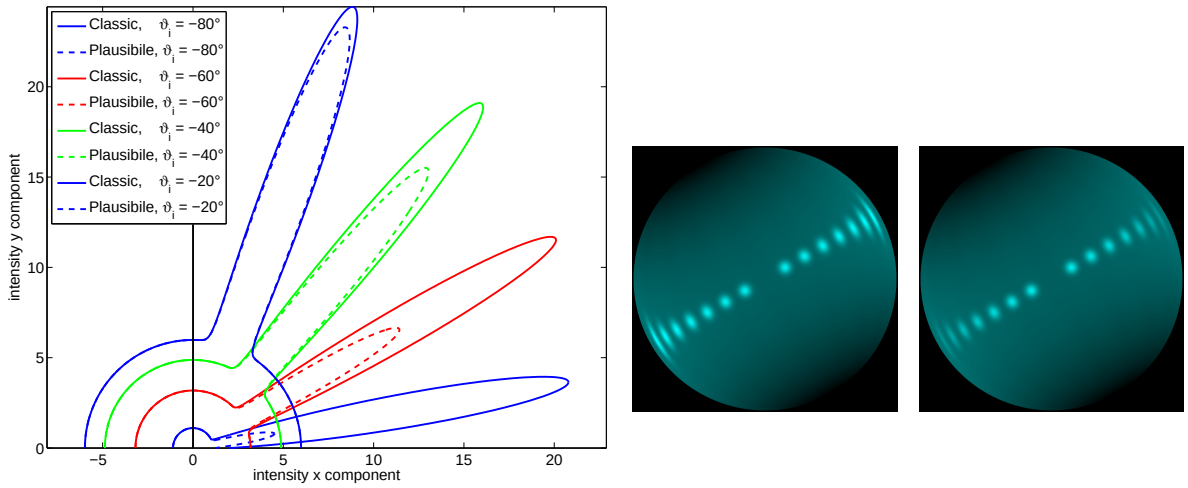


Figure 2.3.: Perfect mirror raytracing example. Image has been created with Blender[1], human model made by Nick Zuccarello[2], downloaded from Blenrig[3].

normal $\mathbf{n}$. Fig. 2.3 shows an example when raytracing a surface with perfect mirror properties. A perfect mirror is rarely found for 3D surface reconstruction based on Photometric Stereo or Shape from Shading (see Section 3.2.1 and Section 3.2.2). However, the field on deflectometric surface shape determination (see e.g. [Knauer et al., 2004]) intrinsically relies on that principle.

[1] http://www.blender.org/
[2] http://nickzucc.blogspot.de/2008/08/human-male-athletic-final-model.html
[3] For a more detailed usage permission, see Appx. D.1

Phong



(a) classic Phong model (solid line) and physically plausible Phong (broken line)

(b) Standard Phong

(c) Plausible Phong

Figure 2.4.: Comparison between the classic Phong model and its physically plausible form. Note the intensity differences between the standard and the plausible Phong towards grazing angles.

The approach by Phong [1975] focuses in its classic form mainly on qualitatively well-looking images instead of physical plausibility. In its classic form, the BRDF is

$$f_P = \frac{k_a}{\langle \mathbf{n} \cdot \mathbf{l} \rangle} + k_d + k_s \frac{\langle \mathbf{r} \cdot \mathbf{v} \rangle^m}{\langle \mathbf{n} \cdot \mathbf{l} \rangle} \tag{2.12}$$

where k_a, k_d and k_s denote the coefficients of the ambient, diffuse and specular reflection, respectively. The angular extent (width) of the observed specularity is controlled by the exponential parameter $m \in \mathbb{R}^+$. Note, that k_s is 0 if the reflected light is unobservable ($\langle \mathbf{r} \cdot \mathbf{v} \rangle < 0$) and a positive nonzero constant otherwise. Energy conservation and Helmholtz reciprocity are not fulfilled, which can be derived directly from the expression $\langle \mathbf{n} \cdot \mathbf{l} \rangle = \cos \vartheta_{in}$ in the denominator [Lewis, 1994]. The Phong model can be plausibilized by using $k_a = 0$ and applying a variation of Eq. 2.12 which is then energy conserving [Lafortune and Willems, 1994] and reciprocal:

$$f_{PP} = \frac{k_d}{\pi} + k_s \frac{2+m}{2\pi} \langle \mathbf{r} \cdot \mathbf{v} \rangle^m , \tag{2.13}$$

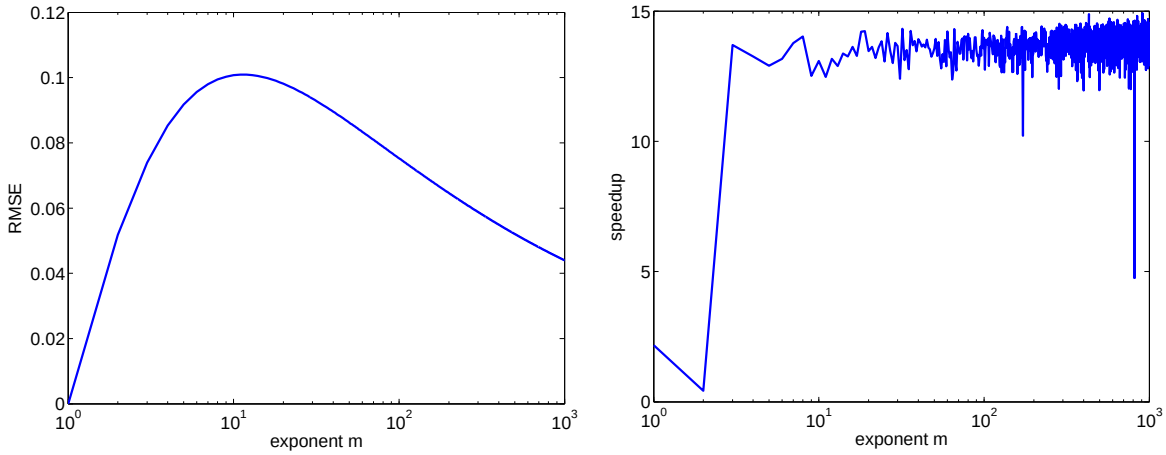
$$\text{with} \qquad k_d, k_s, m \in \mathbb{R}^+, \qquad k_d + k_s = 1. \tag{2.14}$$

An illustration of the differences between the classic and the physically plausible Phong model is given in Fig. 2.4 and a detailed derivation of Eq. 2.13 can be found in Appx. B.3.1.2.

3D surface reconstruction methods that use the Phong model are e.g. [Nayar et al., 1988, 1990a; Carceroni and Kutulakos, 2001; Smith and Hancock, 2010].

Schlick [1994b] examined the Phong cosine lobe model by means of computational efficiency. He suggested a replacement of $\langle \mathbf{r} \cdot \mathbf{v} \rangle^m$ by

$$\langle \mathbf{r} \cdot \mathbf{v} \rangle^m = \frac{\langle \mathbf{r} \cdot \mathbf{v} \rangle}{m - m \cdot \langle \mathbf{r} \cdot \mathbf{v} \rangle + \langle \mathbf{r} \cdot \mathbf{v} \rangle} \tag{2.15}$$



(a) accuracy **(b)** speedup

Figure 2.5.: (a) Schlick model accuracy compared to Phong for different exponential factors m. The largest deviation of 0.10092 occurs for $m = 11$. (b) speedup using Matlab on an Intel i7 2.8GHz PC, the average speedup amounts to 13.7.

which comes without the need for exponentiation. Fig. 2.5 illustrates the Schlick model accuracy and the speedup compared to the classic Phong model. While the speedup is significant (13.7 on average), there is a considerable inaccuracy of the model. Since computational expenses are a secondary criteria to accuracy when it comes to 3D surface reconstructions, the Schlick model has rarely been used.

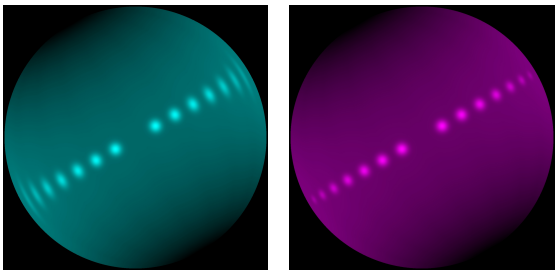
Nayar

For the goal of modeling metallic reflectance, Nayar et al. [1988] apply a superposition of a Lambertian term, a specular lobe, and a mirror-like specular spike. Using those two specular components conforms with the observations by Beckmann and Spizzichino [1963] and Nayar et al. [1991], and have been reviewed in Section 2.3. In detail, Nayar et al. [1988] apply a Torrance Sparrow lobe [Torrance and Sparrow, 1967] and use the perfect mirror BRDF (see Section 2.4.2) to represent the specular spike.
This approach has been adapted by d'Angelo and Wöhler [2008], who model both, the specular lobe and the specular spike, with a specular Phong BRDF component, which provides the possibility to control the angular extent of the specular spike. Their (physically plausibilized) BRDF thus becomes

$$f_N = \frac{k_d}{\pi} + k_l \frac{m_l + 2}{2\pi} \langle \mathbf{r} \cdot \mathbf{v} \rangle^{m_l} + k_s \frac{m_s + 2}{2\pi} \langle \mathbf{r} \cdot \mathbf{v} \rangle^{m_s} . \tag{2.16}$$

The model itself requires five parameters. These are the (diffuse) albedo k_d, the specular lobe strength k_l, the specular lobe width m_l, the specular spike strength k_s, and the specular spike width m_s. The model has been used by e.g. [d'Angelo and Wöhler, 2008; Herbort and Wöhler, 2012] for the reconstruction of 3D surfaces.

Blinn



(a) plausible Phong **(b)** plausible Blinn

Figure 2.6.: Comparison of a specular sphere shaded with the Phong and the Blinn model. Note the physically implausible oval shape of Phong's specular reflections near grazing angles.

The specular highlight model proposed by Blinn [1977] is very similar to Phong's model, but uses a different parameterization, namely the half-angle vector $\mathbf{h} = \frac{\mathbf{l}+\mathbf{v}}{||\mathbf{l}+\mathbf{v}||_2}$. This is faster to compute and actually models the radially symmetric distribution of normals, which is more physically motivated [Ngan et al., 2005]. In the plane of incidence, the half vector angle $\cos\vartheta_h = \langle \mathbf{h}\cdot\mathbf{n}\rangle$ fulfills $\vartheta_h = \frac{1}{2}\vartheta_r$, i.e. amounts to half the size of Phong's mirror reflectance angle. Outside the plane of incidence, this relation is more complex. Demanding energy conservation and reciprocity yields [Lewis, 1994]

$$f_{BP} = \frac{k_d}{\pi} + k_s \underbrace{\frac{(m+2)(m+4)}{8\pi(2^{-\frac{m}{2}}+m)}}_{N_B} \langle \mathbf{h}\cdot\mathbf{n}\rangle^m , \tag{2.17}$$

$$\text{with} \qquad k_d, k_s, m \in \mathbb{R}^+, \qquad k_d + k_s = 1 \tag{2.18}$$

which applies the energy normalization factor N_B. Some shaders and publications use different Blinn normalization aspects and factors, which are overviewed in Table 2.1. The experiments presented in this thesis rely on the factor derived in Appx. B.3.1.3.
3D surface reconstruction methods that use the Blinn model are e.g. [Malzbender et al., 2006; Higo et al., 2010; Herbort et al., 2013b,a].

factor	source	normalization aspect
$\frac{m+8}{8\pi}$	[Akenine-Möller et al., 2008]	energy (approximate)
$\frac{(m+2)(m+4)}{8\pi(2^{-\frac{m}{2}}+m)}$	Appx. B.3.1.3	energy
$\frac{m+1}{2\pi}$	[Lafortune and Willems, 1994]	microfacet distribution
$\frac{m+2}{2\pi}$	[Pharr and Humphreys, 2010, p. 457]	heightfield

Table 2.1.: Blinn normalization factors. While the first factor [Akenine-Möller et al., 2008] is used commonly, it can be shown that this is only an approximation and the derivation for the exact factor is shown in Appx. B.3.1.3. The microfacet normalization normalizes the probability density that a microfacet normal is oriented along the half vector direction $\mathbf{h}$. The so-called heightfield normalization regards the sum of the projected area of the microfacets, which is enforced to be equal to that area viewed from a large distance.

Lafortune

A generalization of the cosine lobe model [Phong, 1975; Lewis, 1994] has been presented by Lafortune et al. [1997]. The original Phong model is restricted by the specular reflection being centered around the direction of mirror reflectance $\mathbf{r} = 2\langle \mathbf{l} \cdot \mathbf{n} \rangle \mathbf{n} - \mathbf{l}$, which is revoked in the generalized model. Detailed insight into the model is given by the following derivation:
The Phong BRDF is

$$f = C_s \langle \mathbf{r} \cdot \mathbf{v} \rangle^m , \tag{2.19}$$

which can be rewritten as

$$f = C_s \left\langle (\mathbf{l} \cdot \mathbf{M}) \cdot \mathbf{v}^T \right\rangle^m \tag{2.20}$$

with the Householder transform $\mathbf{M} = 2\mathbf{n}\mathbf{n}^T - \mathbb{I}$.
In general, $\mathbf{M}$ contains three parameters of freedom, which are extracted by an eigenvalue decomposition such that $\mathbf{M} = \mathbf{Q}^T\ \mathbf{D}\ \mathbf{Q}$. Graphically, $\mathbf{Q}$ is used to transform $\mathbf{l}$ and $\mathbf{v}$ to a new local coordinate system, whose $\hat{z}$-axis is aligned with the local surface normal, and its $\hat{x}$- and $\hat{y}$-axis are aligned with the principal directions of anisotropy. The components of $\mathbf{l}' = \mathbf{l}\mathbf{Q}^T = [l'_x, l'_y, l'_z]$ and $\mathbf{v}'^T = \mathbf{Q}\mathbf{v}^T = [v'_x, v'_y, v'_z]^T$ are then weighted by the elements of the diagonal matrix $\mathbf{D}$:

$$f = C_s(D_{xx}l'_x v'_x + D_{yy}l'_y v'_y + D_{zz}l'_z v'_z)^m. \tag{2.21}$$

The configurations of the elements of $\mathbf{D} = \mathrm{diag}(D_{xx}, D_{yy}, D_{zz})$ characterize several different types of reflectance as shown in Table 2.2. The model thus comprises a wide variety of reflectance phenomena

configuration	reflection type
$D_{xx} = D_{yy}$	isotropic reflection
$D_{xx} \neq D_{yy}$	anisotropic reflection
$D_{zz} = \sqrt[m]{C_s},\ \ D_{xx} = D_{yy} = -\sqrt[m]{C_s}$	classic cosine lobe model
$m = 0$	Lambertian reflectance
$D_{xx} = D_{yy} = 0$	rotationally symmetric diffuse reflectance
$D_{zz} < -D_{xx} = -D_{yy}$	off-specular reflectance
$D_{xx} = D_{yy} > 0$	retro-reflection

Table 2.2.: Characteristic configurations of the weight factors of the generalized cosine lobe model

without needing to apply a specialized model for each case.
Regarding energy conservation, it is most feasible to normalize the elements of $\mathbf{D}$ such that the total hemispherical reflectance D_Σ of $(D_{xx}l'_x v'_x + D_{yy}l'_y v'_y + D_{zz}l'_z v'_z)^m$ equals 1, i.e.

$$D'_{ii} = \frac{D_{ii}}{\sqrt[m]{D_\Sigma}}, \quad i \in \{x, y, z\}. \tag{2.22}$$

The model has been used by e.g. [Lensch et al., 2003; Ahmed et al., 2008] for surface shape determination.

2.4.3 Microfacet-based specular reflectance

The main improvement of microfacet-based specular reflectance lies in the ability to describe off-specular peaks, which have been reported if the surface roughness σ is comparable to the incident light's wavelength λ, i.e. $\frac{\sigma}{\lambda} \gtrsim 1$ [Torrance and Sparrow, 1967][1]. Generally, the introduction of the surface roughness is physically more meaningful than the empirical exponential factor m employed by [Phong, 1975; Blinn, 1977]. Since microfacet-based models are out of the scope of this thesis, only the seminal work by Cook and Torrance [1981] will be discussed in greater detail to explain the underlying principles. Some further models [Beckmann and Spizzichino, 1963; He et al., 1991] can be found in Appx. B.4.

Cook-Torrance

In contrast to previous models Cook and Torrance [1981] use a parametrization of the surface facet distribution function D by the half vector $\mathbf{h}$ as empirically used by Blinn [1977]. Additionally, the Fresnel factor F (see below) has been optimized in contrast to [Beckmann and Spizzichino, 1963; Torrance and Sparrow, 1967] for Computer Graphics applications, while the geometrical attenuation factor G (see below) remains unchanged. The Cook-Torrance BRDF is

$$f_{CT} = \frac{k_d}{\pi} \;+\; k_{CT}\,\frac{F(n,\langle \mathbf{v}\cdot\mathbf{h}\rangle)\cdot D(m,\langle \mathbf{n}\cdot\mathbf{h}\rangle)\cdot G(\mathbf{n},\mathbf{v},\mathbf{h})}{\pi\;\langle \mathbf{n}\cdot\mathbf{v}\rangle\;\langle \mathbf{n}\cdot\mathbf{l}\rangle} \tag{2.23}$$

with

$$F = \frac{1}{2}\frac{(g-c)^2}{(g+c)^2}\left(1+\frac{(c(g+c)-1)^2}{(c(g-c)+1)^2}\right) \tag{2.24}$$

$$c = \langle \mathbf{n}\cdot\mathbf{h}\rangle \tag{2.25}$$

$$g = \sqrt{n^2+c^2-1} \tag{2.26}$$

$$n \quad \text{index of refraction} \tag{2.27}$$

$$G = \min\left\{1, \frac{2\;\langle \mathbf{n}\cdot\mathbf{h}\rangle\;\langle \mathbf{n}\cdot\mathbf{v}\rangle}{\langle \mathbf{v}\cdot\mathbf{h}\rangle}, \frac{2\;\langle \mathbf{n}\cdot\mathbf{h}\rangle\;\langle \mathbf{n}\cdot\mathbf{l}\rangle}{\langle \mathbf{v}\cdot\mathbf{h}\rangle}\right\} \tag{2.28}$$

$$D = \frac{1}{m^2\,\langle \mathbf{n}\cdot\mathbf{h}\rangle^4}\exp\left\{-\frac{\tan^2\vartheta_h}{m^2}\right\} \quad \text{(Beckman distribution)} \tag{2.29}$$

$$m \quad \text{root mean square slope} \tag{2.30}$$

F denotes the Fresnel term for unpolarized light (i.e. extinction coefficient $k = 0$). An extension that models a specular spike component in addition to the diffuse and specular lobe component only concerns the superposition of two or more (weighted) facet slope distribution functions, e.g. $\sigma_l D_l(m_l, \langle \mathbf{n}\cdot\mathbf{h}\rangle)$ and $\sigma_s D_s(m_s, \langle \mathbf{n}\cdot\mathbf{h}\rangle)$, since $G(\mathbf{n},\mathbf{v},\mathbf{h})$ only contains geometric relations and the only variable factor of $F(n,\langle \mathbf{v}\cdot\mathbf{h}\rangle)$ is the index of refraction (n), which does not change for the specular spike. One thus obtains

$$f_{CT} = \frac{k_d}{\pi} \;+\; \frac{F(n,\langle \mathbf{v}\cdot\mathbf{h}\rangle)\cdot G(\mathbf{n},\mathbf{v},\mathbf{h})}{\pi\;\langle \mathbf{n}\cdot\mathbf{v}\rangle\;\langle \mathbf{n}\cdot\mathbf{l}\rangle}\cdot\sum_{l=1}^{L} k_{CT,l} D(m_l,\langle \mathbf{n}\cdot\mathbf{h}\rangle) \tag{2.31}$$

[1] Torrance and Sparrow [1967] use the root mean square slope as a measure for surface roughness.

The model has been simplified with regard to computational efficiency by Schlick [1994a], who approximated the above mentioned Fresnel term by

$$F_S = 1 - (1 - \langle \mathbf{n} \cdot \mathbf{l} \rangle)^5, \tag{2.32}$$

which is now commonly known as the Schlick approximation. His approximation of the surface facet slope distribution function and the geometrical attenuation factor can be found in [Schlick, 1994a].

2.4.4 Anisotropic reflectance

While the above models regard isotropic reflectance where the observed light intensity depends on the difference of the azimuth angles $|\phi_{in} - \phi_{obs}|$ only, there are surfaces where the actual dependence on ϕ_{in} and ϕ_{obs} becomes important. An isotropic BRDF f_{iso} then has at least three degrees of freedom, while an anisotropic BRDF f_{an} has four:

$$f_{iso} = f_{iso}(\vartheta_{in}, \vartheta_{obs}, |\phi_{in} - \phi_{obs}|) \tag{2.33}$$

$$f_{an} = f_{an}(\vartheta_{in}, \vartheta_{obs}, \phi_{in}, \phi_{obs}). \tag{2.34}$$

Typical examples for such anisotropic materials are e.g. brushed aluminum, velvet and satin [Kurt et al., 2010], which are all materials that show a predominant direction of the surface structure.Anisotropic materials show a dependency of the observed intensity from the rotation of a point around its normal. A rendered example is shown in Fig. 2.7.

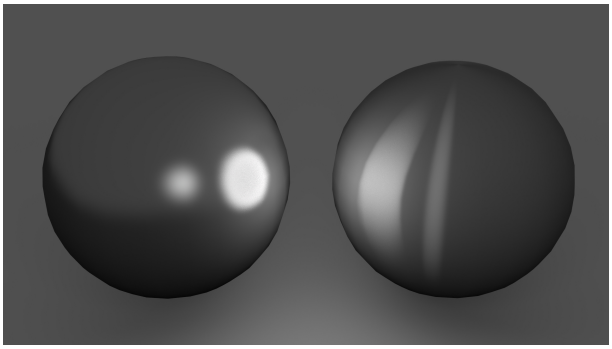


Figure 2.7.: Comparison between specular reflectance (left) and anisotropic reflectance (right).

Anisotropic reflectance requires the introduction of a reflectance value dependency on the azimuth of the half vector $\phi_h = \arccos(\langle \mathbf{h} \cdot \hat{\mathbf{x}} \rangle)$, where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ denote the directions of the local x and y axis (see Section 1.4 for coordinate system definitions). Similar to e.g. [Blinn, 1977; Torrance and Sparrow, 1967; Cook and Torrance, 1981], the surface facet distribution of the following models are parameterized by the polar half angle vector $\vartheta_h = \arccos(\langle \mathbf{n} \cdot \mathbf{h} \rangle)$.
In Appx. B.5, some of the most important anisotropic BRDFs are explained, which are the pioneering approach by Ward [1992], its physical plausibilization by Dür [2006]; Geisler-Moroder and Dür [2010], and the work by Ashikhmin and Shirley [2000] and Kurt et al. [2010]. A detailed derivation of those models is not needed throughout this thesis and Appx. B.5 thus focus on an overview rather than a detailed discussion.

2.4.5 Summary

Table 2.3 gives a summary of the models discussed above along with their main properties – energy conservation, Helmholtz reciprocity, diffuse reflection, specular reflection, anisotropy, Fresnel effects, self-shadowing and self-masking.

Model	E	R	D	S	A	F	G	publication
Lambert	yes$^{(9)}$	yes$^{(9)}$	Y	N	N	N	N	[Lambert, 1760]
Oren-Nayar	yes$^{(7)}$	yes$^{(7)}$	Y	N	N	N	Y	[Oren and Nayar, 1994]
Phong	no$^{(1)}$	no$^{(1)}$	N	Y	N	N	N	[Phong, 1975]
Modified Phong	yes$^{(5)}$	yes$^{(5)}$	N	Y	N	N	N	[Lafortune and Willems, 1994]
Blinn	no$^{(1)}$	no$^{(1)}$	N	Y	N	N	N	[Blinn, 1977]
Modified Blinn	yes$^{(1)}$	yes$^{(1)}$	N	Y	N	N	N	[Lewis, 1994]
Torrance-Sparrow	yes$^{(1)}$	yes$^{(1)}$	N	Y	N	Y	Y	[Torrance and Sparrow, 1967]
Cook-Torrance	no$^{(7)}$	yes$^{(7)}$	N	Y	N	Y	Y	[Cook and Torrance, 1981]
He-Torrance	yes$^{(6)}$	yes$^{(6)}$	Y	Y	N	Y	Y	[He et al., 1991]
Neumann-Neumann	yes$^{(1)}$	yes$^{(1)}$	N	Y	N	N	N	[Neumann and Neumann, 1989]
Minnaert	yes$^{(1)}$	yes$^{(1)}$	N	Y	N	N	N	[Minnaert, 1941]
Ward	no$^{(3)}$	yes$^{(2)}$	N	Y	Y	N	N	[Ward, 1992]
Ward-Dür	no$^{(3)}$	yes$^{(3)}$	N	Y	Y	N	N	[Dür, 2006]
Ward-GM-Dür	yes$^{(4)}$	yes$^{(4)}$	N	Y	Y	N	N	[Geisler-Moroder and Dür, 2010]
Kurt	yes$^{(8)}$	yes$^{(8)}$	N	Y	Y	Y	N	[Kurt et al., 2010]

Table 2.3.: Reflectance model properties. E - energy conservation; R - obeys Helmholtz reciprocity; D - diffuse reflection component; S - specular reflection component; A - anisotropy; F - Fresnel effects; G - self-shadowing and self-masking.
The superscripts refer to: [Lewis, 1994]$^{(1)}$, [Dorsey et al., 2008]$^{(2)}$, Dür [2006]$^{(3)}$, [Geisler-Moroder and Dür, 2010]$^{(4)}$,[Lafortune and Willems, 1994]$^{(5)}$, [Neumann et al., 1999]$^{(6)}$, [Ngan et al., 2005]$^{(7)}$, [Kurt et al., 2010]$^{(8)}$, Appx. B.3.1.1$^{(9)}$.

2.5 BRDF parameter estimation

BRDF models provide an analytical way for the description of light reflectance, but for surface reconstruction algorithms, the material-specific parameters (e.g. ρ, k_d, k_s, m, ...) need to be known. Note, that the surface reconstruction problem is faced with unknown BRDF parameters *and* an unknown surface shape. That scenario will be regarded in greater depth as one of the contributions of this thesis (see Chapter 6). In the following, different approaches are introduced that assume the underlying shape to be known prior to the parameter estimation process, which explains the principle of BRDF parameter estimation suitably.
Yu et al. [1999] propose a method for computing inverse global illumination based on a scene with known geometry and known lighting distribution. They assume an environment consisting of diffuse and specular surfaces, where the incident light is at most reflected twice before being observed by the camera. They successfully extract the respective BRDF parameters and show that this can be used for realistic scene relighting and scene modification.
Machida et al. [2003] model interreflections for dense non-uniform BRDF parameter retrieval using a radiometrically calibrated environment. Reflections are modeled using a combination of a classic single-parameter Lambertian diffuse term and a two-parameter specular Torrance-Sparrow component. Modeling interreflections allows the identification of diffuse parameters even if the intensity measurement is impaired by mutual illumination. Their experimental setup consists of a laser range finder with 60 attached light

sources with known position and intensity that can be activated independently. Their idea is to initially measure the object's 3D shape and fit quadric surfaces for noise robustness, which then allows the extraction of local geometric information (**n**,**l**,**v**). These are used to decide if diffuse, specular, both, or no BRDF parameters can be extracted based on the available **n**-**l**-**v**-combinations for any given pixel on the object surface. It is important to consider that

(1) the diffuse term requires one parameter, which can only be identified correctly if the measurement is free of specularities, which is the case if the observance direction is outside the specular lobe. The mathematical condition applied by Machida et al. for this relation is $\vartheta_r \geq \vartheta_{r,th1}$ with some threshold $\vartheta_{r,th1} = 60°$.
(2) the specular term requires $k \geq 2$ parameters and thus requires k intensity measurements *within* the specular lobe for correct identification. These mathematically need to fulfill e.g. $\vartheta_r \leq \vartheta_{r,th2}$ with some threshold $\vartheta_{r,th2}$. Machida et al. [2003] use $\vartheta_{r,th2} = 20°$.

The number and location of suitable light sources is chosen by Machida et al. [2003] iteratively until the ratio of the measurability of both components reaches 80%. Chosen suitably, they require an object-dependent number of $3…5$ light source positions for the determination of 100% of all pixelwise diffuse parameters and $10…12$ light sources for the determination of $83\%…89\%$ of all specular parameters.
Narasimhan et al. [2003] introduce a method for the segmentation of different surface colors and materials based on RGB color images with at least 3 different light incident directions. Their "photometric invariants" are computed based on the assumption that the underlying BRDF is separable into a material-dependent term and a geometry-dependent term. If then sub-determinants are computed from a matrix of intensity measurements (color $\times$ incident direction), then the geometrical term cancels out and provides a ratio that defines the local material properties. Unfortunately, it has been examined in the masters thesis by Westerhoff [2013] that the segmentations are at best qualitatively and are not suitable for detailed color or material segmentations. This result has been obtained for 3×3 RGB/light matrices as well as for 5×5 multispectral intensity / light matrices that provide larger sub-determinants, which should then be more robust than the 3×3 case.
Once BRDF measurements have been determined with an appropriate method (see Section 2.6), it remains a (in general) nonlinear optimization problem to fit a model M with a set of parameters P to the data R:

$$P^* = \arg\min_P \sum_{j=1}^{J} \left[R(j) - \tilde{R}(\mathbf{l}(j), \mathbf{v}(j), \mathbf{n}(j), M, P) \right]^2 \tag{2.35}$$

This requires to know the incident light direction **l**, viewing direction **v**, and normal direction **n** for each surface sample j. Ultimately, this yields a set of optimized parameters P^*. The algorithm used throughout this thesis for nonlinear optimizations is the trust region reflective optimization approach by Coleman and Li [1994, 1996].
Under real conditions it is not as easily possible to fit the model since the surface normals are possibly inaccurate due to low quality 3D shape measurements or the BRDF samples are corrupted by interreflection-affected measurements. Details regarding those problems are discussed in greater detail in Section 6.2, where a novel solution for that problem is presented.

2.6 BRDF measurement and data-driven BRDFs

As an alternative to analytic BRDF models whose complexity increase with every reflection effect that needs to be modeled, it is possible to *measure* the BRDF of a certain material with respect to the angles of incidence (θ_{in}, ϕ_{in}) and reflectance (θ_r, ϕ_r). This process is tedious due to the dimensionality (4D) of the problem, but it releases the need for modeling each and every reflection effect. Furthermore, it is possible to change the basis functions used for BRDF description through e.g. using spherical harmonics. In the following, different approaches for non-parametric or measured BDRFs are reviewed.

technique	probe shape	acquisition time	sample accuracy
gonioreflectometer	arbitrary	high	high
known objects	specialized	rather short	high
3D scanner and camera	arbitrary	rather short	medium

Table 2.4.: BRDF sample measurement comparison.

The measurement of BRDF samples depends on application, object shape, and available time. White et al. [1998] used a gonioreflectometer for measuring the reflectance properties depending on the angles of incidence (θ_i, ϕ_i) and the angles of reflectance (θ_r, ϕ_r), i.e., a full BRDF (see Section 2.2). With the consideration of the azimuth ϕ_i and ϕ_r, the approach even allows the acquisition of anisotropic materials. A general drawback of gonioreflectometers lies in the time that is needed to fully capture the BRDF of some material, since only one sample at a time is obtained [Weyrich et al., 2008].
With the availability of high resolution digital camera sensors, the reflection property sample rate grew immensely since they capture a whole set of angles over a single image. Ward [1992] used one of such settings for the acquisition of anisotropic BRDFs. He applied planar samples for the acquisition, which yield an easy geometry for the estimation process. Later, Marschner et al. [1999a] captured the BRDF from curved objects with known shape. The approach has been extended to arbitrary shapes [Marschner et al., 1999b], which then allows capturing the BRDF of e.g. human skin. However, the latter procedure requires the prior knowledge of the shape before the BRDF can be determined. Since such a shape determination algorithm (e.g. (active) triangulation stereo) induces additional inaccuracies, the increased generality comes at the cost of a loss in precision in comparison to objects with known shape. For isotropic materials, the object can be exchanged with a metallic sphere (again under controlled lighting and a camera as a capturing device), as presented by Matusik et al. [2003a]. The usage of sample spheres depends on the availability of the material in that shape, which is usually not the case for biological materials like skin, leaves or liquids. The advantages and disadvantages that have been discussed above are summarized in Table 2.4.
In a series of publications, Dror et al. [2001d,a,b,c] pursued the goal of surface material identification from single object images. In contrast to the above-mentioned approaches on BRDF measurement, they focused on the identification of the material that has been captured rather than its exact reflectance function. They showed that coarse reflectance properties already provide the required information to classify a material based on a set of previously determined properties.
Alldrin [2006] published results on BRDF estimation using spherical harmonics as basis functions and prior knowledge of object shape (perfectly spherical), material (completely homogeneous), lighting, negligible subsurface scattering, and isotropy. The spherical harmonic approximation to the BRDF is then obtained by

solving a system of linear equations based on the measurements from several images. Alldrin's results show that his spherical harmonics approximation produces undesired artifacts and lacks in the representation of specularities. Smith and Hancock [2010] use a facial shape prior for the estimation and then iterative refinement of BRDF parameters. This deals with the trade-off between accuracy and shape generality mentioned by Marschner et al. [1999b], since the iterative shape refinement successively yields an accurate shape basis. A somewhat similar approach that uses an iterative refinement scheme is described in Chapter 6 along with an evaluation of the accuracy of the determined BRDF.
Since model based approaches need careful attention regarding model design, if all relevant effects (diffuse reflection, specular reflection, anisotropy, back-scattering, Fresnel effect, ...) are supposed to be captured, it seems to be an alternative to use a data-driven BRDF for renderings and surface reconstructions. Weyrich et al. [2008][p. 13] argue that the concept of measured reflectance data comprises a greater variety of optical phenomena than modeled data. While this eliminates the need for modeling complex reflection effects, it comes at the cost of time efficiency and memory efficiency [Lensch et al., 2003]. However, measuring reflectance functions has gained popularity since recent technical advances now provide the technical requirements [Matusik et al., 2003b]. The BRDFs of 100 materials measured by Matusik et al. [2003a] can be found in the MERL BRDF database[2].

[2]Mitubishi Electric Research Laboratories (MERL) BRDF databse http://www.merl.com/brdf/

3 Related work - 3D reconstruction

While a human observer can intuitively infer object shapes from object images, this is considerably more complex for a machine-driven algorithm. However, human and computer-based scene evaluations differ in their expectations: Humans are usually interested in a *qualitative* knowledge of the shape, but it is desired to obtain *quantitative* shape information from a machine, i.e. precise normal directions or depth measurements in some metric unit.

The following chapter reviews important advances that have been published in the area of object shape determination, specifically regarding absolute shape measurements using photogrammetric techniques, and surface gradient determination methods using photometric approaches. It will become clear that a combination of both areas is desirable since they benefit from their mutual advantages. Additionally, it will be illustrated that most of the existing algorithms rarely deal with non-Lambertian surfaces and/or concave surfaces that exhibit interreflections.

This chapter is partly adapted and/or adopted from [Herbort and Wöhler, 2011] and [Herbort et al., 2013a]

While the 3D-world-to-2D-image principle has been known as the camera obscura since Mozi[1] [Ouellette, 2005, p.52], it is considerably more difficult to obtain 3D measurements from the real world or even more to recover the depth information from a 2D image.
Since the algorithms that are introduced later (Chapter 5 et seq.) do not rely on the actual source of the depth data, and since they focus on scan refinement rather than absolute depth determination, it is not reasonable to discuss depth determination techniques in great detail or a large variety. Instead, the following sections regard the technique that has been applied for depth data acquisition within this thesis – namely a laser stripe sensor, which belongs to the class of active triangulation techniques. A detailed categorization and discussion of range scanning techniques can be obtained from the PhD thesis by Wiora [2001].
The following sections initially describe techniques that measure 3D coordinates (Section 3.1), and then review seminal algorithms that attempt to recover shape information from an image (Section 3.2). The remaining sections regard more complex image-based shape retrieval algorithms.

3.1 Absolute depth determination

Techniques for the determination of absolute depth measurements in some metric unit exist in a large variety. Table 3.1 gives an overview of some of these techniques along with information regarding their principles and categorization. Active triangulation techniques manipulate the light that is projected on

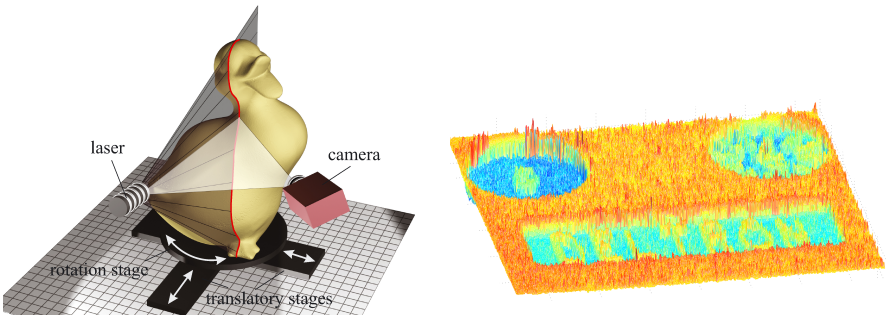
Technique	scene interaction	physical contact	category	principle	publication (e.g.)
CMM[2]	active	yes	contact	mechanics	[Ashley et al., 1970]
Active Triangulation	active	no	reflection	triangulation	[Shirai and Suwa, 1971]
Pattern projection	active	no	reflection	deformations	[Peng, 2006]
Interferometry	active	no	reflection	interferences	[Michelson and Morley, 1887]
Deflectometry	active	no	reflection	phase shift	[Knauer et al., 2004]
Time-of-flight	active	no	reflection	light speed	[Bartczak and Koch, 2009]
AM[3]	active	no	reflection	phase shift	[Besel, 1989]
Radar	active	no	reflection	radio waves	[Daniels, 1996]
Sonar	active	no	reflection	sound waves	[Elfes, 1987]
CT[4]	active	no	transmission	absorption	[Farmer and Collins, 1971]
Triangulation Stereo	passive	no	appearance	disparity	[Okutomi and Kanade, 1993]
Shape from Motion	passive	no	appearance	correlation	[Tomasi and Kanade, 1992]
Depth from defocus	passive	no	appearance	depth of field	[Subbarao and Surya, 1994]

Table 3.1.: Overview of several absolute depth determination techniques.

a camera-observed scene such that it transmits light-coded information upon the visible surfaces. This information is later used to identify each projected light ray, which then allows the geometric estimation of the 3D intersection points between light rays and object surface (see Fig. 3.1(a) for an illustration). Since the underlying principles are well known from the field of geometry and light physics, there is no seminal work that actually introduces the principle, but the review by Besl [1988] gives a good impression of early principles employed for 3D shape determination.

[1] chinese philosopher, 470 to 390 BC
[2] Coordinate Measurement Machine
[3] Computer Tomography
[4] Amplitude modulation

In practice, there are several different illumination types that are used to provide the required intersections:



(a) triangulation principle **(b)** noise induced by laser speckles

Figure 3.1.: (a) Triangulation setup with camera, laser and object[4]. (b) High frequency noise due to laser speckles. An illustration of high frequency noise, missing measurements at specular highlights, and missing measurements at interreflection-affected surface concavities on the surfaces obtained with a light stripe sensor can be found in Fig. 1.1.

It is possible to project a single flying spot [Rioux et al., 1987], single swiped laser stripes [Curless and Levoy, 1995], Gray coded stripes [Besl, 1988], or even a shadow stripe [Bouguet and Perona, 1999]. What remains challenging is the exact detection and identification of the position of the projected light in the image. Wrong detections, which are common in e.g. concave regions, degrade the scan quality severely and require special treatment as explained by Couture et al. [2011] and Gupta et al. [2011].
Additionally, there are problems whenever the measuring sensor reaches saturation, since the exact spatial intensity peak is not detectable for those locations. This occurs for large exposure times or if the observed material is specular, since specular highlights reflect lots of light into a narrow angular range. Similarly, if the light ray is reflected away from the sensor by a mirror-like material, then there is no intersection detectable.
Generally, there is a limitation of the depth accuracy of triangulation-based range scanning techniques to a multiple of the lateral resolution and the scan thus contains a significant amount of high frequency noise [Nehab et al., 2005]. An illustration of high frequency noise caused by laser speckles is shown in Fig. 3.1(b), and Fig. 1.1 already introduced the three most common measurement errors.
Range scanners are usually challenged by moving objects [Weise et al., 2007], but recent approaches show improvement when using motion compensation [Harendt, 2013] or ultra-high-frequency pattern projection [Schaffer et al., 2013].

3.2 Surface gradient determination

Absolute depth determination techniques usually rely on geometric principles or light roundtrip measurements. A different class of approaches for surface shape determination exploits photometric[5] cues, which yield local surface gradient information. Examples are Shape from Shading (SfS, see Section 3.2.1) and Photometric Stereo (PS, see Section 3.2.2). The difference between photometric and photogrammetric approaches is explained in greater detail in Section 3.3. Since surface gradients only contain information of

[4]Image source http://en.wikipedia.org/wiki/File:LaserPrinciple.png, public domain license.
[5]*photometry*: measurement of light in terms of its perceived brightness.

the surface slope, it is required to integrate those gradient fields into a final depth map (see Section 3.4) in order to determine the underlying 3D shape.

3.2.1 Shape from Shading (SfS)

SfS aims for the determination of surface gradients based on a single given intensity image. The basic mathematic problem statement is, that the observed intensity I at the surface point (x, y) is supposed to be equal to a modeled reflectance R, which is parameterized by the unknown surface normal $\mathbf{n}(x, y)$ at that point:

$$I(x, y) = R(\mathbf{n}(x, y)). \tag{3.1}$$

Note, that each normal $\mathbf{n}(x, y)$ is defined by at least two unknowns p and q for its direction, which makes the problem statement in Eq. 3.1 ill-posed and requires additional constraints to be solvable as described later.
Under the assumption of Lambertian reflectance, Eq. 3.1 becomes

$$I(x, y) = \underbrace{\frac{I_0}{r_l^2(x, y)} \cdot \frac{\rho(x, y)}{\pi} \cdot \langle \mathbf{l}(x, y) \cdot \mathbf{n}(x, y) \rangle}_{R(\mathbf{n}(x,y))} . \tag{3.2}$$

with observed intensity $I \in \mathbb{R}^+$, point light source to object distance $r_l \in \mathbb{R}^+$, surface albedo $\rho \in [0..1] \subset \mathbb{R}^+$, illumination intensity $I_0 \in \mathbb{R}^+$, and illumination direction $\mathbf{l} \in \mathbb{R}^3$.
Without loss of generality, it can be assumed that

$$I(x, y) = C(x, y) \cdot \langle \mathbf{l}(x, y) \cdot \mathbf{n}(x, y) \rangle , \tag{3.3}$$

i.e. $C(x, y) = \frac{I_0}{r_l^2(x,y)} \cdot \frac{\rho(x,y)}{\pi} \in \mathbb{R}^+$. Using the gradients $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ such that $\mathbf{n} = (p, q, -1)^T$, and $\mathbf{l} = (l_x, l_y, l_z)^T$, this can be rewritten as

$$I(x, y) = C \cdot \frac{pl_x + ql_y - l_z}{\sqrt{1 + p^2 + q^2}}. \tag{3.4}$$

The problem thus becomes a first order non-linear partial Hamilton-Jacobi differential equation [Duroua et al., 2007]:

$$I(x, y)\sqrt{1 + |\nabla z|^2} - C \cdot ((l_x, l_y)\nabla z + l_z) = 0. \tag{3.5}$$

An early approach towards solving the equation (via *characteristic strip expansion*) has been attempted by Rindfleisch [1966], but nowadays it is the work by Horn [1970], which is known as the seminal Shape from Shading (SfS). Further publications by Horn regarding SfS improvements are [Horn, 1975a,b, 1977; Horn and Sjoberg, 1978]. Other methods for solving the equation are e.g. viscosity solutions [Rouy and Tourin, 1992] or level sets [Bruckstein, 1988; Kimmel and Bruckstein, 1995].
Note that even with $\mathbf{l}(x, y)$, $\rho(x, y)$, $r_l(x, y)$, and I_0 being known, Eq. 3.4 can not be solved uniquely. Uniqueness requires more than one intensity value $I(x, y)$ per pixel, since at least two independent gra-

dient directions ($p = \frac{dz}{dx}, q = \frac{dz}{dy}$) need to be determined. Therefore, additional constraints are introduced. These are e.g. smoothness or integrability constraints [Horn, 1989]. Details can be found in the extensive reviews by Zhang et al. [1999] and Duroua et al. [2007]. The SfS formulation with an integrability regularization condition and a variational approach towards its solution [Horn and Brooks, 1986] is presented in Section 6.1.1 as a foundation for the contributions in that field (see Chapter 6).
The incorporation of additional (regularization) constraints poses the fundamental difference between SfS and PS, where the latter is usually independent of additional constraints due to a sufficiently large number of intensity measurements for each pixel.
Since SfS analyzes the amount of reflected light from an object's surface, the approach is generally vulnerable against all phenomena that influence the intensity values in the object's image. These are the object's reflection properties, the sensor's sensitivity properties, and how the object is illuminated, i.e. the number, direction, and radiation characteristics of the light sources. Further challenges are caused by interreflections, which are discussed in detail in Chapter 6.

3.2.2 Photometric Stereo (PS)

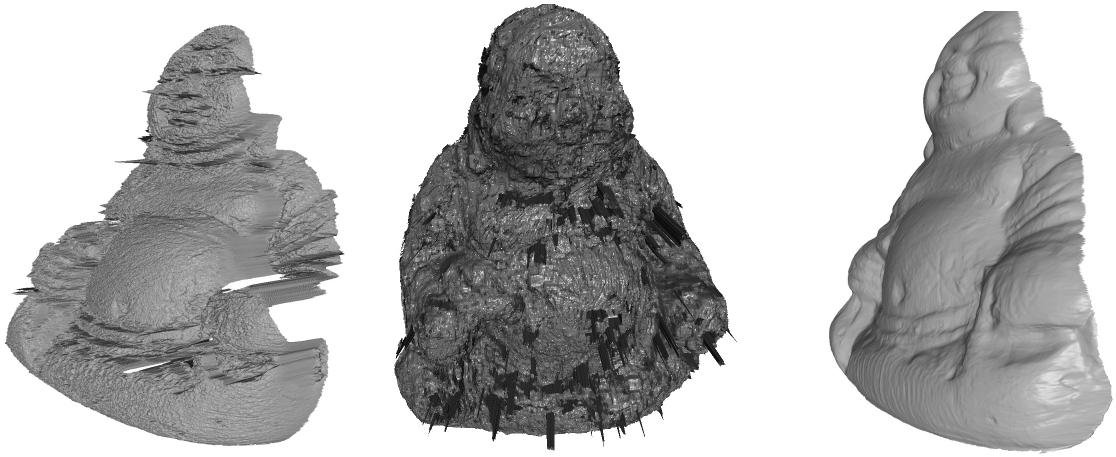
SfS's need for regularization constraints has been resolved by Woodham [1980] through the acquisition of $N \in \{2, 3, ...\} \subset \mathbb{N}$ images I_i with $i = 1...N$ of an object that is observed from direction $\mathbf{v} \in \mathbb{R}^3$, but with *varying* illumination directions $\mathbf{l}_i \in \mathbb{R}^3$ of strength $I_{0,i} \in \mathbb{R}$. This approach is commonly known as *Photometric Stereo (PS)*. In contrast to SfS, the problem results in a system of linear equations that is uniquely solvable:

$$\begin{pmatrix} I_0(x,y) \\ I_1(x,y) \\ \vdots \\ I_N(x,y) \end{pmatrix} = \frac{\rho(x,y)}{\pi} \cdot \begin{pmatrix} \leftarrow & I_{0,0} \cdot \mathbf{l}_0(x,y) & \rightarrow \\ \leftarrow & I_{0,1} \cdot \mathbf{l}_1(x,y) & \rightarrow \\ \leftarrow & \vdots & \rightarrow \\ \leftarrow & I_{0,N} \cdot \mathbf{l}_N(x,y) & \rightarrow \end{pmatrix} \cdot \mathbf{n}(x,y) = \frac{\rho}{\pi} \cdot \mathbf{L} \cdot \mathbf{n}. \tag{3.6}$$

Note, that Eq. 3.6 needs to be solved for each pixel in the image. Through determining only the direction of the surface normal such that $\mathbf{n} = (p, q, -1)^T$ and assuming the surface albedo ρ to be known a priori, this ends up with two unknowns p, q and thus requires at least two intensity measurements to be solvable. If ρ is unknown as well, then at least three images are required. Note, that for $\mathbf{L}$ to be invertible, its rows (i.e. illumination directions) need to be linearly independent. In the case of more than three acquired images, $\mathbf{L}$ is overdetermined, but the equation can then be solved in the least mean squares sense using the pseudo inverse of $\mathbf{L}$ (see e.g. [Penrose, 1955; Israel and Greville, 2003]).
While Eq. 3.6 assumes Lambertian reflectance, there are usually non-Lambertian reflection effects like specularities present in real scenes. This is discussed in greater detail in Section 3.5.4, but it is noted here that this usually requires careful consideration and modeling of these effects and then solving a non-linear problem statement, which is mathematically more complex but still follows the surface-gradients-from-image-intensities principle.
The acquisition of images is traditionally achieved sequentially in time multiplex, but frequency-multiplexed capturing by application of colored lights has been applied by e.g. Hernández et al. [2007], which is beneficial for e.g. moving objects.

PS computes – like SfS – a pixelwise surface normal field rather than absolute depth values. Techniques for the integration of that field to depth values are discussed in Section 3.4.

3.3 Comparison of photogrammetric and photometric approaches



(a) raw structured light scan **(b)** raw triangulation stereo result **(c)** photometric stereo followed by gradient field integration

Figure 3.2.: Demonstration of different surface shape determination techniques. Note the high frequency noise and outliers in structured light and triangulation stereo depth ((a) and (b)), which are almost not contained in the depth obtained from Photometric Stereo with subsequent gradient field integration (c). However, PS causes a deformation of the overall shape, which yields a profile that is too flat in comparison to the structured light scan (a). (b) has been adopted from [Fiegel, 2013].

Structured light, triangulation stereo, and photometric approaches can be compared by means of extracted data, ambient illumination influence, high frequency noise robustness, large scale shape accuracy, evaluateable surface types, and calibration requirements.

Both, structured light and triangulation stereo, estimate absolute depth data directly, while photometric approaches yield surface gradients that need to be integrated into a depth field. Photogrammetrically obtained surfaces show a considerable amount of high frequency noise and outliers (see Fig. 3.2(a) and Fig. 3.2(b)), which are significantly less present in the depth profile of photometrically estimated and integrated gradients (see Fig. 3.2(c)). Note that structured light and triangulation stereo on the one hand measure the depth values for each pixel independently, which causes high frequency noise due to small deviations in the depth estimation of adjacent pixels. Photometric approaches on the other hand estimate the surface gradients independently for each pixel, but the final depth profile contains accumulated gradient inaccuracies through gradient field integration, which yields large scale inaccuracies (i.e. biased low spatial frequency data).

An advantage of structured light approaches over appearance-based methods (triangulation stereo, photometric methods) lies in being independent from ambient illumination as long as the light patterns are perceivable despite the ambient light. The projected light patterns provide the required information/illumination and make the process robust against light that originates from elsewhere than the structured light projector.

All approaches face challenges regarding object surface materials that lead to saturated sensor responses: Non-Lambertian surfaces exhibit specularities that cause these responses for structured light, create dif-

ferent object appearances between the cameras of triangulation stereo, and induce estimation errors to photometric techniques that assume Lambertian reflectance. This can be improved by e.g. multi-exposure settings for structured light approaches, improved correspondence establishment for triangulation stereo [Wöhler and d'Angelo, 2009], and a reflectance model adjustment for photometric approaches. Note that triangulation stereo requires at least a little surface texture to determine correspondences, which is not the case for structured light and photometric approaches.
System calibration is most critical for triangulation stereo approaches, followed by structured light, and photometric approaches. Generally, all approaches require the removal of lens distortion effects. Triangulation stereo is additionally vulnerable against displacements of the cameras, since a dislocation of the epipolar lines yields systematic errors in the reconstruction process. Structured light requires the calibration of the projector position and direction relative to the camera, and photometric approaches require a calibration of the surrounding light sources (see Chapter 5).
Mainly due to the advantage regarding high spatial frequency accuracy, photometric approaches show broad application and initiated a wide field of research. However, the analysis of non-Lambertian surfaces is challenging. Recent and past attempts for the applicability to arbitrary (non-Lambertian) surfaces are discussed in detail in Section 3.5, followed by novel approaches in Chapter 6.

3.4 Reconstruction of the depth profile from local gradients

Once surface normals are obtained from photometric approaches such as SfS or PS, the final shape of the object needs to be determined from the surface gradients p and q. If the gradient field belongs to a real surface z such that $p := \frac{\partial z}{\partial x}$ and $q := \frac{\partial z}{\partial y}$, then the field is integrable, i.e. the integral along any closed path C in the field is zero:

$$\int_C p + q \, ds = 0 \;\; \forall \;\; C \tag{3.7}$$

In other words, the curl of the field is zero. Practically, this is rarely the case due to noise, inaccuracies, and possible ambiguities. It is thus needed to find the best-matching surface approximation for a given non-integrable gradient field.
Agrawal et al. [2006] exposed, that the problem is far from being unique. The degree of variation in the possible solutions results from weights applied to the gradients during the integration process. In the following, their main ideas are sketched since the derivation and description gives good insight into the topic and its applicability.
The classically applied condition is simply *integrability* and the minimization problem becomes

$$z^* = \underset{z}{\operatorname{argmin}} \iint_{x,y} \underbrace{(z_x - p)^2 + (z_y - q)^2}_{E(p,q,z_x,z_y)} \, dx \, dy, \tag{3.8}$$

and the Euler-Lagrange equation for that problem $\frac{\partial E}{\partial z} - \frac{\partial}{\partial x}\frac{\partial E}{\partial z_x} - \frac{\partial}{\partial y}\frac{\partial E}{\partial z_y} = 0$ leads to the Poisson equation

$$\nabla^2 z = \nabla (p, q)^T, \tag{3.9}$$

whose solutions are straightforward (see Section 6.1.1).
Note, that Eq. 3.8 is a specialization and that in general it is required to minimize

$$z^* = \underset{z}{\operatorname{argmin}} \iint_{x,y} E(z, p, q, z_{x^a y^b}, p_{x^c y^d}, q_{x^c y^d}, ...)\, dx\, dy, \tag{3.10}$$

where E is an n-th degree differential equation and comprises the derivatives of p, q and z with $a + b = k$ and $c + d = k - 1$ for arbitrary numbers $1 \leq k \leq n \subset \mathbb{N}$ and $z_{x^a y^b} = \frac{\partial^a}{\partial x^a} \frac{\partial^b}{\partial y^b} z$ [Agrawal et al., 2006]. Using the Euler-Lagrange equation and assuming

$$\frac{\partial E}{\partial z_x} = f_1(z_x, z_y) - f_3(p, q) \qquad \frac{\partial E}{\partial z_y} = f_2(z_x, z_y) - f_4(p, q) \tag{3.11}$$

with four weight functions $f_1 ... f_4$ then yields

$$\frac{\partial E}{\partial z_x} = \nabla\left(f_1(z_x, z_y), f_2(z_x, z_y)\right)) - \nabla\left(f_3(p, q), f_4(p, q)\right)). \tag{3.12}$$

On that basis, it is possible to trace the variety of approaches that solve the gradient-to-depth-integration problem back to different weighting functions. Some available approaches are summarized in Table 3.2 (Frankot-Chellappa and Regularization are omitted since they contain no further insight for this thesis).
The experiments conducted during the course this thesis showed that the M-estimator approach performs

approach	$f_1(z_x, z_y)$	$f_2(z_x, z_y)$	$f_3(p, q)$	$f_4(p, q)$
Poisson	z_x	z_y	p	q
α-surface	$b_x z_x$	$b_y z_y$	$b_x p$	$b_y q$
M-estimator	$w_x z_x$	$w_y z_y$	$w_x p$	$w_y q$
diffusion	$d_{11} z_x + d_{12} z_y$	$d_{21} z_x + d_{22} z_y$	$d_{11} p + d_{12} q$	$d_{21} p + d_{22} q$

Table 3.2.: Different approaches for gradient field reconstructions as identified by Agrawal et al. [2006]. Note that Poisson simply uses spatially invariant isotropic weights, that include every gradient sample without any choice or reduced/increased weight. The binary weights (b_x, b_y) for α-surfaces decide, whether a gradient sample is included in the integration based on a threshold α. The M-estimator approach uses continuous anisotropic weights (w_x, w_y) for the gradients, which are chosen by using some less than quadratic residual-based error function. The diffusion approach uses an affine transformation as weights for the gradients, which depends on the underlying gradients p, q rather than the residual.

best for most datasets based on empirical comparisons between Poisson, α-surface, M-estimator, and diffusion.
In an alternative approach, Wu et al. [2008] initially apply the "osculating arc constraint" [Gray, 1997] to the normal field, which yields robust height differences h for each pair of adjacent pixels. Then, the depth is computed by minimizing the differences between depth field differences z and height differences h using Gauss-Seidel iteration. However, the work by Wu et al. [2008] shows, that an M-estimator yields better results by means of noise suppression and qualitative shape reconstruction.
An approach that does not rely on enforcing discrete integrability as used by Agrawal et al. [2006] and most other authors has been proposed by Ng et al. [2010].

3.5 Generalizations of SfS and PS

The classic SfS and PS approaches explained in Section 3.2.1 and Section 3.2.2 are subject to several restrictions regarding e.g. illumination, surface reflection properties and camera. This impairs the applicability considerably and the following sections thus regard generalizations. These are light source distance (Section 3.5.1), uncalibrated light sources (Section 3.5.2), perspective projections (Section 3.5.3), and - most importantly - Non-Lambertian surfaces (Section 3.5.4).

3.5.1 Non-distant light sources

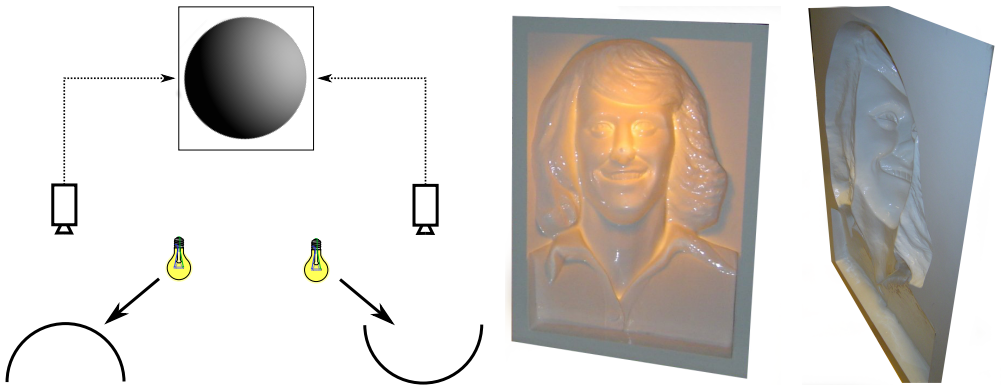
Clark [1992] introduced a method termed *Active Photometric Stereo*, which estimates absolute depth values from images that have been acquired under controlled movement of a point light source located relatively close to the object surface. In contrast to the approach by Iwahori et al. [1990], who as well applied point light sources located near the object surface, Clark's method only requires solving a linear equation (see below) and is furthermore not restricted to objects with a Lambertian surface.
Instead of just using absolute image irradiance values I, the proposed algorithm exploits irradiance changes $\nabla_{\mathbf{t}} I$ that result from translating the light source by $\mathbf{t}$. The derived equation for the computation of the absolute depth z then results in

$$z = \frac{(\nabla_{\mathbf{t}} I)^T \mathbf{t} + 2\, I}{(\nabla_{\mathbf{t}} I)^T \mathbf{p}}, \tag{3.13}$$

where $\mathbf{p} = (\frac{x}{f}, \frac{y}{f}, -1)^T$ indicates the position of the image plane point $p_I = (x, y)^T$ with respect to the world coordinate system for a camera with camera constant f, and $\nabla_{\mathbf{t}} I = (\frac{\partial I}{\partial \mathbf{t}_x}, \frac{\partial I}{\partial \mathbf{t}_y}, \frac{\partial I}{\partial \mathbf{t}_z})^T$ denotes the irradiance change rate with respect to the light source position changes. Note the independence of Eq. 3.13 from surface reflectance properties, which is a strong advantage over other approaches. Clark's evaluation of the algorithm showed a strong dependence on the image noise, which can be reduced to a certain extent by acquisition of multiple images and application of a weighted least squares approximation to the resulting depth values, or, in the biased case, a median filter. Furthermore, the accuracy of the algorithm suffers from saturated brightness values in the image.
The algorithm by Kozera and Noakes [2006] assumes non-distant light sources throughout the derivation of image noise affected PS. Similar to the algorithms that assume a perspective camera (see Section 3.5.3), they note that the depth can be computed only up to a scalar $c \in \mathbb{R}$ such that the depth $z = z + c$ underlies a so-called standard ambiguity. This vanishes once the problem is formulated using non-distant light sources.
The PS algorithms used for this thesis assume non-distant light sources by means of modeling locally varying light incident directions. This increases the reconstruction accuracy as reported by [Kozera and Noakes, 2006] and as proven by the experiments presented in Chapter 6.

3.5.2 Unknown illumination conditions

The assumption of unknown illumination directions $\mathbf{l}_i$ with $i \in \mathbb{N} < \infty$ has been analyzed by Hayakawa [1994], who examined the application of Photometric Stereo in a scene with arbitrarily moving illumination. The main insight of Hayakawa's work is, that the determination of the illumination direction and the illumination intensity from a set of images is not unique and thus requires the introduction of additional constraints.



(a) sketch **(b)** hollow face illusion **(c)** explanation (concave surface)

Figure 3.3.: Generalized Bas-Relief Ambiguity: With unknown lightning conditions, it is always possible to find an illumination condition that makes different surface profiles appear identical in the respective images as shown in (a). This is the basis for the so-called hollow face illusions (b)[2] and (c)[3].

The ambiguity problem has been examined in greater detail regarding surface reconstructions by Belhumeur et al. [1999], who concluded that without knowing the light direction, one can only solve up to the so-called Generalized Bas-Relief (GBR) ambiguity: An object's visible surface $f(x, y)$ is therefore indistinguishable from an image under a "generalized bas-relief transformation" $\bar{f}(x, y) = \lambda f(x, y) + \mu x + \nu y$, i.e. identical to a scaled version $\lambda f(x, y)$ of the surface itself with an added plane $\mu x + \nu y$, see Fig. 3.3 for a general explanation, $\lambda \in \mathbb{R}^{>0}; \mu, \nu \in \mathbb{R}$.

Solutions for the problem have been proposed by Yuille et al. [2000] through introducing and enforcing Freeman's "generic viewpoint constraint" [Freeman, 1994]. Further approaches have been published by Drbohlav and Sara [2002] through exploitation of specularities and thus enforcing the "consistent viewpoint constraint". They assume a Lambertian surface with superimposed mirror-like reflections and the specularities are analyzed as additional geometric constraints for solving the problem. Ultimately, four images obtained under varying light positions suffice to resolve the ambiguity and compute the shape of the object at hand.

A conceptually similar approach using the Torrance-Sparrow reflectance model [Torrance and Sparrow, 1967] has been introduced by Georghiades [2003], but it allows solving the GBR ambiguity only up to the binary convex/concave ambiguity. Later, Drbohlav and Chantler [2005] reduced the number of required specular pixels (i.e. number of images captured with different light positions) from four as required by

[2] Wikipedia, public domain license, http://en.wikipedia.org/wiki/File:Bjorn_Borg_Hollow_Face.jpg (cropped).
[3] Wikipedia, public domain license, http://commons.wikimedia.org/wiki/File:Bjorn_Borg_Hollow_Face-concave.jpg (cropped).

Drbohlav and Sara [2002] to two. They ultimately propose a linear scheme that solves the GBR. Tan et al. [2007] generalized the conclusion by Drbohlav and Chantler [2005] regarding the form of the reflectance function and showed that the GBR ambiguity is resolvable using an arbitrary isotropic and spatially invariant non-Lambertian reflectance function and exploiting the Helmholtz reciprocity [von Helmholtz, 1924, p. 231]. Their approach is evaluated on real and synthetic images and compared with a ground truth obtained from a calibrated PS approach. Further insight is given by Tan and Zickler [2009] along with a projective framework based on the real projective plane that allows the analysis of reflectance symmetries using isotropy and reciprocity, or half-vector symmetry. For uncalibrated PS, they show that constraints resulting from the analysis of isotropy and reciprocity in a single image allow resolving the GBR ambiguity.
Basri et al. [2007] show that uncalibrated PS can be performed with the sole constraint of distant light sources. The illumination of each image may even include an arbitrary combination of diffuse sources, point sources, and extended light sources. Under these illumination types and Lambertian surfaces, they apply spherical harmonics to model the lighting conditions. The number of required images rises from four images for a first order spherical harmonic approximization to nine images for a second order approximization with a small effect on the surface estimation accuracy. However, the approximation only applies low-order spherical harmonics, which are not suitable for the representation of specular reflectance. Determination of environmental lighting has further been studied by Shen and Tan [2009], who examined freely available internet images of a scene with varying viewpoint, noise, and unknown global illumination based on Basri et al. [2007] and Ramamoorthi [2002]. For these images, they compute the global illumination and ultimately classify the image-specific weather conditions based on the illumination characteristics.
An approach using interreflections on non-Lambertian surfaces with non-translational symmetry has been examined by Chandraker et al. [2005]. They prove mathematically that interreflections contain the required information to resolve the GBR ambiguity and then show experimental results for noise-free rendered and real images. Alldrin et al. [2007] have shown, that a minimization of the entropy of the surface albedo distribution as well resolves the ambiguity, even without geometrical considerations.
Zhou and Tan [2010] recently introduced a method for Lambertian surfaces that applies a ring light source, i.e. light sources that are distributed uniformly with a fixed distance around the camera. This provides additional prior knowledge such that the GBR is solvable up to scaling, vertical mirroring, and hyperbolic or circular rotation. For the remaining transformations, Zhou and Tan derive and discuss combinations of constraints, which solve the ambiguities successfully. Another recent algorithm by Shi et al. [2010] initially analyzes (color) images of Lambertian surfaces to identify points with equal normals and albedoes, which then allows to resolve the GBR ambiguity based on the information from at least four pixels with the same albedo but different normals. The experiments provide results for the successful determination of surface normal groups, albedo groups and, additionally, the radiometric response function of the camera.
Hernandez and Vogiatzis [2010] introduced a self-calibrating approach for a monocular facial capturing system. The actual shape retrieval is based on multi-spectral Photometric Stereo (MSPS), which needs to be calibrated prior to its application to a human face. The calibration stage applies an initial structure-from-motion algorithm to obtain the motion in the scene and then uses a multiview-stereo algorithm to estimate the coarse shape of the face. The coarse 3D shape provides sufficient information regarding the normals and intensities to estimate the lighting parameters and thus calibrate the system. Since the lighting parameters vary between different faces, the calibration procedure needs to be repeated before capturing an unknown face. After calibration, frequency-multiplexed images are captured that provide the shape re-

trieval data required for the algorithm described by Hernández et al. [2007]. Frequency-multiplexed images do not suffer from uncompensated motion like traditional time-multiplexed Photometric Stereo images. After shadow processing according to Hernandez et al. [2008], the obtained face reconstruction results show good high-frequency detail but suffer from the low frequency noise that is typical for photometric approaches.
For this thesis, the lighting conditions will be calibrated carefully in advance (see Chapter 5), which provides robust light source positions and intensities. This releases the reconstruction algorithm from solving that problem and removes a typical source of inaccuracies without introducing limiting constraints.

3.5.3 Perspective projections

Prados and Faugeras [2003] developed a formulation of the SfS problem under consideration of a perspective pinhole camera. A new partial differential equation is obtained, whose solution is found to be unique, exact, and converging using their proposed viscosity solution scheme. The experimental evaluation shows strong advantages over assumedly orthogonal projections and increased robustness against intensity noise and deviations from the illumination direction.
Tankus et al. [2005] initially discuss the significance of the reconstruction error that occurs if a projectively obtained image is reconstructed under the orthographic assumption. Furthermore, they derive a SfS image irradiance equation similar to [Prados and Faugeras, 2003] based on the natural logarithm of the depth $\ln(z)$. The solution based on the projective framework is found to be invariant to scale changes $z^* = c \cdot z$ with $c, z \in \mathbb{R}$, which is more plausible than the orthographic framework's independence of translations $z^* = z + c$. They propose a scheme to solve for the depth z using a Fast Marching algorithm [Kimmel and Sethian, 2001]. An evaluation of the solutions obtained by fast marching and viscosity solutions shows advantages of perspective algorithms over orthographic algorithms as reported by Prados and Faugeras [2003].
Similar to SfS, the orthographic camera assumption has been found to perform unfavorably for the PS reconstruction accuracy as well [Tankus, 2005]. Tankus thus propose a novel PS framework, where the perspective pinhole camera model is used for the derivation of the PS image intensity equation. Again, the perspective approach has been found to perform better over a whole set of error measures.
As a conclusion of this section, it is meaningful to design PS algorithms based on the perspective rather than the orthographic framework. This has thus been the basis for the contributions in this thesis (see Chapter 6).

3.5.4 Non-Lambertian surfaces

The importance of extending photometric approaches to non-Lambertian surfaces has been discussed by Georghiades [2003], who stresses that surfaces in general are not sufficiently described by a Lambertian diffuse term and a superimposed mirror-like specular spike with narrow extent. Real surfaces exhibit specular *lobes* rather than spikes, which is the case for e.g. cast iron or rough plastics. Contrary, the pioneering algorithms by Horn [1975a] and Woodham [1980] assume the case of strictly Lambertian surfaces (see Section 3.2). The generalization of that case requires correct handling of various reflection effects that occur in addition to diffuse reflections. In the following, approaches that consider the simultaneous appearance of superimposed reflection components are discussed. In the beginning, the focus lies upon methods that consider the case of purely specular reflectance (Section 3.5.4.1), and combinations of specular and diffuse

reflectance (Section 3.5.4.2). Then, methods for the separation of both reflection components are discussed (see Section 3.5.4.3).
It will become clear that surface reconstruction algorithms that aim for highly accurate reconstructions (see Chapter 6) need to consider non-Lambertian reflectance behavior.

3.5.4.1 Purely specular reflections

An approach for dealing with *purely* specular surfaces has been introduced by Ikeuchi [1981]. It alters the illumination setting of PS from point light sources as applied by Horn [1975a] and Woodham [1980] to extended light sources. An extended light source broadens the angular extent of the specular spike and additionally balances the relative intensities of specular and diffuse reflections. Both of these effects are advantageous for the reconstruction process since the specular spike is then visible under a larger range of angles and facilitates measuring both reflection components in a single image.
Under the assumption that the incident light is exclusively reflected specularly, Ikeuchi defines a model for perfect specular reflections at the object's surface. After image acquisition, the intensities are related to the integrated irradiance over the whole radiator area. For the determination of surface orientations, Ikeuchi applies PS with a number of reflectance maps equal to the number of illumination sources. For the relaxation process, two approaches have been applied, where the first one averages the results from two distinct surface orientation lookups and the second one applies a surface smoothness constraint in combination with an irradiance constraint [Ikeuchi, 1980]. Since the technical equipment yielded coarse $128\ px \times 128\ px$ data, the results are rather qualitative, but it is shown that the techniques allows the extraction of reasonable surface orientations from the examined objects.
Later, Morel et al. [2005] proposed a reconstruction method based on polarization data obtained from specular reflections. In their experiments, a dome light is employed to provide (almost) omnidirectional illumination, which allows exploiting Fresnel's formulae for surface gradient retrieval. The results have been compared to the data obtained by a range scanning system, for which the object had to be covered by diffuse coating to allow dense depth measurements. The proposed algorithm has been found to provide a qualitatively better result, especially in the cross-sectional profile of the object under examination.

3.5.4.2 Linear combination of different reflection components

The previously discussed Lambertian assumption (Section 3.2.1) and the pure specularity assumption (Section 3.5.4.1) are both restricted in their application, since they are specializations of a general case. Further generalization of reflectance phenomena comprises a combination of a Lambertian and a specular spike component, with optional inclusion of a specular lobe component [Ikeuchi, 1981].
Nayar et al. [1990a] initially discuss surfaces with three reflection components (diffuse, spike, lobe) but ultimately reflections are modeled as a linear combination of a Lambertian and a specular spike component. For the acquisition of the intensity images, they apply multiple *extended* light sources with uniform distribution around the object, which have already proven well-suited for capturing surfaces with purely specular reflectance (see Ikeuchi [1981] and Section 3.5.4.1).
Nayar et al. [1988] present an algorithm for the simultaneous extraction of surface orientations and reflectance parameters for plastic and metallic objects, based on a sampling constraint and a unique orientation constraint. Georghiades [2003] models the specular (lobe) reflectance behavior using the Torrance-Sparrow

reflectance model [Torrance and Sparrow, 1967] in addition to a Lambertian term. The model complexity is reduced under the assumption of small directional differences between viewing direction $\mathbf{v}$ and illumination direction $\mathbf{l}$, which allows neglecting attached shadows and Fresnel reflectivity. The reconstruction as such is an extension of their earlier work [Georghiades et al., 2001], which iteratively determines surface normals, Lambertian albedo, light sources, and parameters of the Torrance-Sparrow reflectance model in the least square sense. They demonstrate the performance of their approach with lacquered surfaces and human faces, which both benefit from accounting for specular lobes.
Goldman et al. [2005, 2010] included the isotropic Ward model [Ward, 1992] into a PS framework to estimate spatially-varying BRDFs while simultaneously recovering the object shape. They assume that the examined material is composed of a small set of fundamental materials, whose combination allows modeling the surface at hand. Their recovered rendered images show low quantitative errors and plausible material parameters, but contain some ambiguities at specular highlights due to overfitted material reflection models.
Tan et al. [2008] proposed a superresolution approach that recovers the shape of non-Lambertian surfaces with a depth resolution greater than the underlying image resolution. As in their initial work on that topic [Tan et al., 2006], they extend the description of surface reflectivity so that more than a single surface facet per pixel may contribute to the observed brightness. The process as such requires a quite large number of images ($60…70$) under varying illumination directions in order to solve for the surface. The authors present results for synthetic and real surfaces at $2\,\text{px} \times 2\,\text{px}$ and $4\,\text{px} \times 4\,\text{px}$ resolution enhancement. These show reduced angular normal errors in comparison to an a priori available ground truth and exhibit the expected increase in high frequency surface detail, but possibly contain additional noise.

3.5.4.3 Separation of specular and diffuse components

The separation of the specular and the diffuse component relies on their respective characteristic properties. The three main differences are [Tan and Ikeuchi, 2005]: (1) Specular reflections have a larger degree of polarization than diffuse reflections. (2) The intensity distributions of diffuse reflections approximately follow Lambert's Law, the intensity distributions of specular (lobe) reflections follow the Torrance-Sparrow model [Torrance and Sparrow, 1967] or the Beckmann-Spizzichino model [Beckmann and Spizzichino, 1963]. (3) In the visual light spectrum, specular reflections have been found to be largely independent of the object surface's spectral reflectance properties as noted in the dichromatic reflection model in Section 2.3. In contrast, the spectral power distribution of diffuse reflections depends on the object's spectral reflectance, which causes stronger wavelength-dependence in the visual spectrum.
A method that uses the first property (polarization) for the identification of specular reflections has been presented by e.g. Wolff [1989] and Wolff and Boult [1991], who exploit that specular reflections show a higher degree of polarization $D \in [0..1] \subset \mathbb{R}$ than diffuse reflections. D can be obtained (see Eq. 3.14) from the highest observed intensity value $I^{max}(x,y) = I(x,y,\varphi_{max})$ and the lowest observed intensity value $I^{min}(x,y) = I(x,y,\varphi_{min})$ of a pixel (x,y) in an image with respect to the polarization filter orientation φ:

$$D = \frac{I^{max} - I^{min}}{I^{max} + I^{min}}. \tag{3.14}$$

The principle has been extended using polarization *and* color by Nayar et al. [1997]. Algorithms that use color information allow a reduction of the required images for reflection component separation to a single

image. For specular spike suppression, Ikeda [2013] use polarization filters in front of light sources *and* camera.

A method for the separation of specular and diffuse reflection components on the basis of one single color image has been introduced by Tan and Ikeuchi [2003, 2005]. Their assumptions require a chromatic surface color ($R \neq G \neq B$), i.e. not a black, white, or gray color. In their specular-to-diffuse mechanism, the illumination color of the input image is initially normalized to obtain pure white specular components. Afterwards, the image pixels are examined in a maximum chromaticity over image intensity space, where the specular and diffuse components can be separated by a scalar thresholding value. Although the authors claim that their specularity-free image has an exactly identical geometrical profile as the diffuse component, the evaluated results show errors of up to 35 intensity steps, which are 13.7% for 8 bit images.

Tan et al. [2003] proposed a single color image highlight removal algorithm that estimates the underlying diffuse color by illumination-constrained inpainting. Constraints are obtained from the partial diffuse color information contained in specular highlights and their illumination color uniformity, which can be derived partially from a chromaticity analysis. Their experiments show an improvement of traditional vector fitting based or total variation based algorithms due to an improved recovery of obscured textures. Both, images with large scale texture and images with detailed texture are processed successfully.

Mallick et al. [2005] propose another separation method for dichromatic surfaces and extend their examination by performing PS on the extracted diffuse part. In contrast to the approaches described above, three images are acquired as necessary for PS and analyzed for specular highlights. These are identified and separated from the diffuse component by a data-dependent rotation of the RGB colorspace. The important property of that "SUV"-colorspace is the preservation of shading information, which ultimately allows applying PS.

The approach of Tan and Ikeuchi [2003, 2005] has been extended by Thomas and Sugimoto [2010] through elimination of Tan and Ikeuchi's illumination constraint and thus allow the number and respective direction of the light sources to be unknown. While the algorithm aims for the registration of range images, the number of acquired images can be raised unconditionally to two and thus surface normals can be recovered.

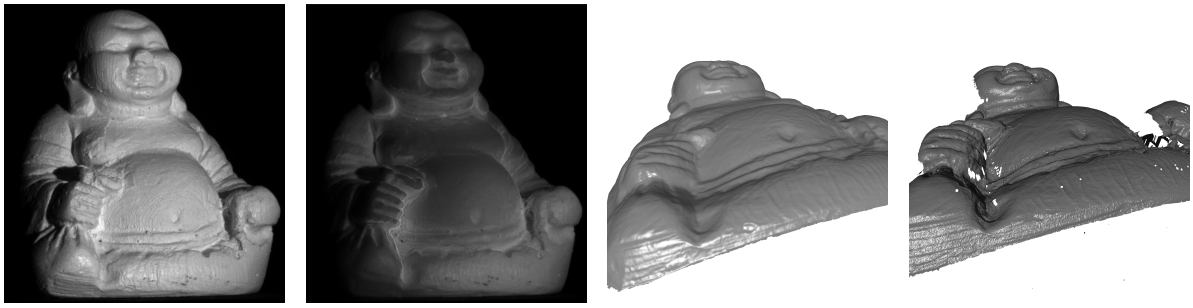
[Narasimhan et al., 2003] present "photometric invariants", which can be used for various tasks, including material segmentation and separation of diffuse/specular reflections. In detail, they aim for a separation of material from lighting and shape based on a scene description with separable BRDFs. The availability of multispectral data (color images) under different illumination conditions then allows to solve for a geometric invariant by simply computing ratios of matrix determinants. The authors show successful application of their method to isolated homogeneous objects, homogeneous objects in complex environments, objects consisting of different materials, inhomogeneous objects, and objects with a specular surface. However, the invariants require changes in illumination, viewpoint, and/or object position between consecutive acquisitions. Unfortunately, it was not possible to recreate the presented results even though the technique was carefully implemented and tested in the masters thesis by Westerhoff [2013] who identifies several critical points regarding the applicability of that algorithm.

Zickler et al. [2008] use a projection of color images with three color channels $[R_s, G_s, B_s]$ on a two-component subspace, which is then found to be independent of specular reflections. The algorithm requires illumination with $R_l \neq 0, G_l \neq 0$ and $B_l \neq 0$, and is restricted in its application to surfaces with a diffuse color $[R_s, G_s, B_s]$ that differs in more than saturation from the incident light such that $[R_s, G_s, B_s] = [\alpha R_l, \beta G_l, \gamma B_l]$ does *not* fulfill $\alpha = \beta = \gamma$. For example, if the incident light is perfectly

white $[1, 1, 1]$, then the algorithm is unable to distinguish between specularities and diffuse surface regions with any gray color $[\alpha, \alpha, \alpha], \alpha \in [0...1] \subset \mathbb{R}$. The application to color images by Zickler et al. shows removal of specular highlight and preservation of shading information that can be used for e.g. photometric surface normal determination. Westerhoff [2013] repeated the application to various surfaces. He notes problems if the specular highlights outshine the underlying diffuse component considerably, but assesses an overall very good applicability.

3.6 Treatment of interreflections

Interreflections cause major accuracy problems for all surface regions that are illuminated indirectly as shown in Fig. 3.4. Forsyth and Zisserman [1991] argue, that it is impossible to obtain accurate normal maps from SfS or PS, if the intensity measurements are impaired by mutually illuminated facets, i.e., interreflections. The impairment is lower, if the surface albedo is low such that most of the incident light is absorbed rather than reflected back into the scene. Using the Lambertian assumption, Forsyth and Zisserman [1991]



(a) image **(b)** interreflections **(c)** PS depth **(d)** scanner depth

Figure 3.4.: Illustration of the accuracy impairment by interreflections. The mutual illumination (b) contained in the image data (a) causes underestimated concavity depths in the profile obtained with PS (c) compared to the more accurate scanner profile (d)

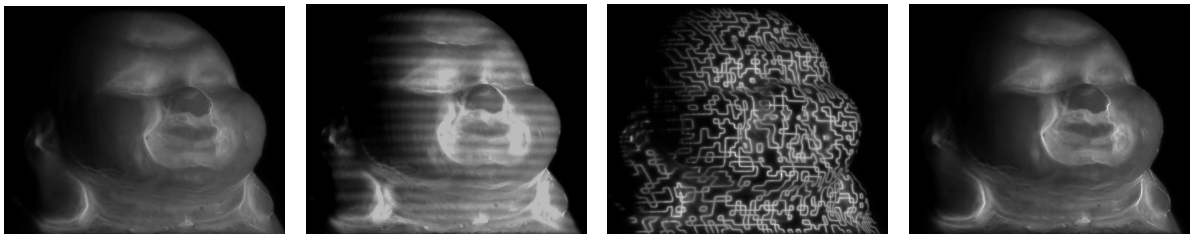
examine several basic shapes that are prone to interreflection occurrence, but they are only able to detect interreflection-affected regions rather than quantitatively analyzing them. Ultimately, they admit that they are not able to extract useful surface information for 3D shape retrieval.

The only algorithm known where interreflections are exploited for shape retrieval has been published by Nayar et al. [1990b]. They introduced an approach that iteratively refines concave v-shapes affected by interreflections. For strictly Lambertian surfaces, they are able to model the interreflections using a kernel matrix for the illumination interdependence between the facets, which allows an estimation of improved normals on the direct illumination component. However, they are restricted to simplistic shapes that are mainly long v-groves with homogeneous albedoes. While they provide a discussion of the convergence of the algorithm, they do not apply their approach to a complex 3D shape within a real scene.

Other approaches like [Seitz et al., 2005; Nayar et al., 2006] focus on the measurement-based separation of global and direct illumination, which has then been applied by Gu et al. [2011] for surface reconstruction using the direct component only: Seitz et al. [2005] propose the existence of so-called interreflection cancellation operators for arbitrarily reflecting scenes. They show that it is possible to obtain an operator matrix by analyzing images that have been acquired through point-wise illumination of the scene. With that

operator, it is possible to obtain n-bounce images, i.e. images in which the light has been reflected exactly $n = \{1, 2, 3, ...\}$ times. The presented results regard Lambertian surfaces and demonstrate the applicability of the method and the existence of interreflection cancellation operators. For practical applications, it is very time-consuming and computationally expensive to acquire the pointwise illuminated images and the required image matrix inversion requires a considerable amount of memory for an $NM \times NM$ matrix if the images are of size $N \times M$.

A more practical approach for the determination of global and direct illumination has been presented



(a) ground truth,SSIM = 1

(b) sinusoidal pattern, 1 period per $4\,px$, SSIM = 0.969

(c) random field, 1 in $4\,px$ illuminated, SSIM = 0.8

(d) 5×5 checkerboards, $1\,px$ step, SSIM = 0.977

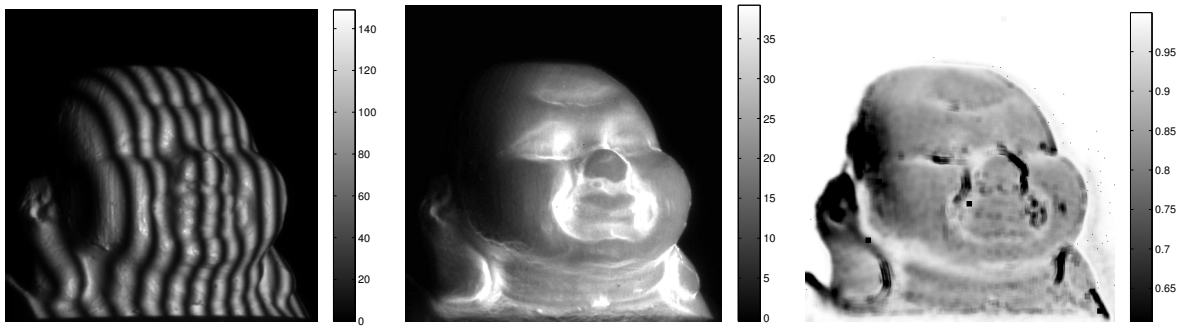
Figure 3.5.: (a) shows a ground truth acquired by successively illuminating the whole scene except for one spot, which allows highly accurate global component determination. (b) uses a bivariate sinusoidal pattern, (c) shows an example using a random projection field, and (d) uses a checkerboard pattern. The results of the checkerboard pattern show quantitatively and qualitatively the best accuracy. Images obtained from the masters thesis by Schugk [2012].

by Nayar et al. [2006]. Instead of pointwise illumination, they use checkerboard, sinusoidal, and random binary patterns, which are swiped across the scene and reduce the number of required images tremendously compared to [Seitz et al., 2005]. While the technique is not able to extract the n-bounce images, it allows the determination of an accurate direct and a global illumination component. The accuracy is limited by the spatial frequency of the projected patterns and the method depends on the quality and location of the scene illumination projector. The masters thesis of Schugk [2012], which has been supervised during the course of this thesis, shows, that the most feasible patterns are checkerboard patterns with regard to global component accuracy and number of required images as shown in Fig. 3.5.

Gu et al. [2011] use the principle of Nayar et al. [2006] for global illumination determination, but they use a lattice pattern for the illumination. Fig. 3.6 shows some evaluation results for that pattern. The problem is, that global illumination components of surface concavities that run along the same direction as the projected stripes are not determined with the accuracy as those running perpendicularly. However, the main contribution of Gu et al. [2011] lies in the idea that PS-based surface reconstructions can be performed on the direct component only, which is then free of global effects that impair the reconstruction accuracy. They show the successful transfer of that idea to various scenarios like a concave two-plane edge, a metallic cake tin, and an indoor scene model.

Note, that the techniques by Nayar et al. [2006], Gu et al. [2011], and the ground truth acquisition principle by Seitz et al. [2005] are not only restricted to global illumination caused by interreflections, but are able to determine the global component of subsurface scattering effects as well.

Ikeda [2013] account for interreflections by adaptively weighting the estimated surface gradients p and q with the curvature. Their assumption is, that interreflections occur for (large) positive curvature regions



(a) exemplary image **(b)** global component **(c)** pixelwise SSIM

Figure 3.6.: (a) exemplary acquisition image with vertical stripes. (b) obtained global illumination component. (c) pixel wise SSIM, note the deviations near vertical concavities. Images obtained from the master thesis by Schugk [2012].

and that the intensity can be modeled using the curvature. They show qualitative results but do not provide an evaluation that validates their model for the relationship between interreflection intensity and curvature, and they do not provide maps that suitably show the improvements yielded for the depth reconstruction by their interreflection compensation.
An approach for dealing with interreflections in complex non-Lambertian scenes is introduced in Chapter 6. In contrast to most of the approaches introduced above, it does not rely on specialized illumination techniques. Instead, point light sources are modeled and the interreflections are incorporated iteratively.

3.7 Multiplexed photometric stereo

The idea to replace time multiplexed image acquisition by frequency multiplexed (multispectral) acquisition has been proposed by [Drew and Kontsevich, 1994; Kontsevich et al., 1994] and [Woodham, 1994]. The application to dynamic scenes has later been presented by [Klaudiny et al., 2010] and [Hernandez and Vogiatzis, 2010].
The thee-source multispectral PS equation under the assumption of distant illumination of a Lambertian surface is [Hernandez and Vogiatzis, 2010; Anderson et al., 2011a]:

$$\begin{pmatrix} I_R \\ I_G \\ I_B \end{pmatrix} = \rho \underbrace{\begin{bmatrix} c_{R,l_1} & c_{R,l_2} & c_{R,l_3} \\ c_{G,l_1} & c_{G,l_2} & c_{G,l_3} \\ c_{B,l_1} & c_{B,l_2} & c_{B,l_3} \end{bmatrix} \cdot \begin{bmatrix} l_{1,x} & l_{1,y} & l_{1,z} \\ l_{2,x} & l_{2,y} & l_{2,z} \\ l_{3,x} & l_{3,y} & l_{3,z} \end{bmatrix}}_{\mathbf{C}\cdot\mathbf{L}} \cdot \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}, \tag{3.15}$$

with RGB intensities I_R, I_G, I_B, surface albedo ρ, unit global incident light directions $\mathbf{l}_i = [l_{i,x}, l_{i,y}, l_{i,z}]^T \in \mathbb{R}^{3\times1}$, $i = 1, 2, 3$, and unit surface normal $\mathbf{n} = [n_x, n_y, n_z]^T \in \mathbb{R}^{3\times1}$. The elements $c_{m,i}$ denote the influence of the i-th incident light direction on the m-th color channel. Note that this implies, that the arbitrarily colored light source i may have an influence on the intensity channel m. For calibrating the system, the 9 unknowns of $\mathbf{C}\cdot\mathbf{L}$ need to be found. Solutions for that problem can be obtained from [Johnson and Adelson., 2009; Hernandez and Vogiatzis, 2010; Klaudiny et al., 2010; Anderson et al., 2011a].
Note that Eq. 3.15 requires an individual matrix $\mathbf{C}\cdot\mathbf{L}$ for each region with homogeneous chromaticity

as applied by Anderson et al. [2011a]. Usually, the chromaticity is assumed to be constant over the whole scene, but that problem can be resolved by using some time multiplexing in addition to the frequency multiplexed data acquisition [Decker et al., 2009; Kim et al., 2010], or by estimation of the chromaticity and regularization of the normal field [Janko et al., 2010].

3.8 3D scan enhancement by fusion with photometric cues

The general drawback of triangulation-based algorithms is the noisy high frequency component, while the general advantage lies in the robustness of the low frequency part as described thoroughly in Section 3.3. Since these properties exist inversely for photometric approaches, i.e. accurate high frequency data but biased low-frequency data, it intuitively implies a combination of both approaches [d'Angelo and Wöhler, 2008]. The crucial problem lies in finding a way to combine both datasets that suitably fuses both data sources and exploits the mutual advantages. Deeper insight into the topic is given in the following sections, which review some important advances in that particular field of research in recent years.
The surface reconstruction related contributions presented in this thesis (see Chapter 6) rely on an iterative minimization of an extended error functional. In contrast to the algorithms presented below, the functional allows to determine depth and gradients simultaneously in an integrated scheme while accounting for large-scale deviations from the scanner depth. Details are presented in Section 6.1.

3.8.1 Iterative error functional minimization

Ikeuchi [1987] initially computes normal maps for the images obtained from *two* cameras that observe a scene. This is achieved by consecutively illuminating the scene by three light sources and then performing classic Lambertian PS for each viewpoint, respectively. Then, a coarse depth profile is established by finding region-wise correspondences between both normal fields. The normals are integrated in that coarse depth profile by iteratively minimizing the integrability error [Horn and Brooks, 1986] with an additional constraint that demands the gradients estimated for left and right camera $(p^l, q^l),\ (p^r, q^r)$ to be the same for corresponding points.

$$E = \underbrace{\int_{x,y} (z_x - p)^2 + (z_y - q)^2 \, dx \, dy}_{\text{integrability error}} + \lambda \underbrace{\int_{x,y} (T(p^r) - p^l)^2 + (T(q^r) - q^l)^2 \, dx \, dy}_{\text{normals error}} . \tag{3.16}$$

Note that this requires a transformation $T(.)$ of the right (or left) camera's gradient fields that accounts for the translational and rotational displacement of both cameras. The integrability error is explained in greater detail in Section 6.1.
[Lu et al., 2012] propose a method for fusing depth and image data with a 100-to-1 difference between image and depth resolution. They use the depth-from-normals approach by Wu et al. [2008], which is extended by a depth deviation constraint and a boundary constraint that enables patch-wise processing to

save computational expenses. The iterative procedure becomes (in principle)

$$z_i^{(n+1)} = z_i^{(n)} + \lambda \underbrace{\sum_j (h_{i,j} - (z_i^{(n)} - z_j^{(n)}))}_{\text{depth-from-normals}} + \gamma \underbrace{(f(z_i^{(n)}) - z_{RS,i})}_{\text{depth error}} + \delta \underbrace{(z_i^{(n)} - z_k^{(n)})}_{\text{boundary error}} \tag{3.17}$$

where λ, γ, δ denote weights, (n) and $(n+1)$ denote the iteration indices, i and j are depth value indices, and $h_{i,j}$ is the osculating-arc-constrained depth difference, and the depth-from-normals term is evaluated for each sample j in some neighborhood of i. $f(.)$ performs subsampling and low pass filtering to match the optimized depth z and the measured range scanner depth Z_{RS} by means of their resolution. The boundary error is only evaluated for all regions indicated by the index k where surface patches overlap. They show detailed evaluations using synthetic and real data of Lambertian surfaces, which both demonstrate high accuracy. Additionally, the results of one dataset is compared to the depth profile obtained with an industrial laser scanner and shows very small depth differences.

Wöhler and d'Angelo [2009] proposed an approach for the stereo image analysis of non-Lambertian surfaces, which exploits the advantages of three different sources of information, i.e. geometric, polarimetric and photometric data. Since pure passive triangulation techniques only yield coarse results in the presence of specular reflections, an iterative scheme is introduced that combines the data successively. The approach initially uses a blockmatching algorithm for the estimation of a sparse set of depth data points and then iteratively refines the initial result by inclusion of polarization and reflectance information. The successive refinement of the 3D surface reconstruction yields a dense and accurate representation of the examined surface. The results show small depth errors ($30...100\mu m$) for the examined region at a lateral pixel resolution of $86\mu m$. The approach is another example that the combination of (noisy) absolute depth data from correspondence-based approaches supplemented with high frequency details from photometrically motivated approaches yields very exact and robust results. This method additionally deals with non-Lambertian surfaces, which is a strong advantage over the approaches mentioned above.

Vlasic et al. [2009] use a dome of 1200 light sources and eight cameras for capturing 3D data of dynamic Lambertian shapes like moving humans at up to 60 Hz. They capture dense normals and account for low frequency deviations by performing multi-view matching followed by a thin-plate spline deformation. Initially, a calibration object is used for the creation of a normal direction look-up table similar to [Hertzmann and Seitz, 2005]. After depth-from-normals integration, the mesh is deformed based on well-performing surface matches from different views by minimizing

$$\underset{z}{\text{argmin}}\, \gamma \sum_i (z_i - \tilde{z}_i)^2 + \mu \sum_{i,j} (z_i - z_j)^2, \tag{3.18}$$

i.e., the surface z is obtained by minimizing for each sample i the distance between the current mesh samples z_i and the multi-view correspondences $\tilde{z}_i$, while the deformation is smoothly interpolated with respect to z_i's neighbors z_j.

3.8.2 Filter-based approaches

Cryer et al. [1995] combine the depth obtained from low-pass filtered triangulation stereo with depth data yielded by a SfS algorithm with subsequent high-pass filtering. The filters are designed according to the visual perception properties of the human eye [Hall and Hall, 1977], which emphasizes the high frequency component against the low frequency component. However, results are obtained for coarse 128×128 images, which are only suitable for a qualitative assessment.
Yang et al. [2007] use iterative bilateral filter for a post processing step of depth maps established using multiview stereo. Just as [Kim et al., 2009; Zhu et al., 2008], there is no explicit modeling of surface gradients and the depth maps are only improved by means of edges and smooth areas present in the image data. The approach can thus be regarded as an edge-preserving smoothing algorithm for depth maps. They show that their approach is able to improve the accuracy and ranking of several stereo algorithms in the Middlebury database.

3.8.3 Linear problem statement

Nehab et al. [2005] linearized the depth/gradient fusion problem, which allows efficiently solving for the optimized depth data z. In detail, the weighted sum of two error terms,

$$\underset{z}{\operatorname{argmin}} \quad \lambda E_p + (1-\lambda) E_n, \tag{3.19}$$

needs to be minimized. In this sum,

$$E_p = \sum_i |P(i) - P_m(i)|^2 = \sum_i \mu_i |z(i) - z_m(i)|^2 \tag{3.20}$$

denotes the position error between the optimized positions $P(i)$ and the measured positions $P_m(i)$ with $\mu_i = (\frac{x_i}{f_x})^2 + (\frac{y_i}{f_y})^2 + 1$ being a factor that accounts for the perspective projections and $i \in [1, 2, ..., N] \subset \mathbb{N}$ denotes the sample number with a total of $N \in \mathbb{N}$ samples in the dataset.

$$E_n = \sum_i [T_x(P(i)) \cdot N^c(P(i))]^2 + [T_y(P(i)) \cdot N^c(P(i))]^2 \tag{3.21}$$

determines the error in the normal directions, which uses the tangents T_x and T_y of the respective points $P(i)$ in the absolute depth data cloud, and a field of normals (N^c) that carries the high frequency information from the measured normals and low frequency information from the absolute depth data field. The error sum as such decreases if the normals and tangents are perpendicular. The tangent-based formulation allows a linear statement of the problem

$$\begin{bmatrix} \lambda\mu I \\ \vdots \\ T_x \cdot N^c \\ \vdots \\ T_y \cdot N^c \\ \vdots \end{bmatrix} \cdot \begin{bmatrix} \vdots \\ \vdots \\ z \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \lambda\mu z^m \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix}, \tag{3.22}$$

which can be solved for the depth data z linearly and thus uniquely in the least mean squares sense. The authors addotionally present several examples where the successful high frequency noise reduction and high frequency detail reconstruction becomes apparent. The approach by Nehab et al. [2005] has been applied for fusing normals obtained from multispectral PS with coarse absolute depth maps from a Kinect sensor [Anderson et al., 2011a], and for fusing similarly estimated normals with depth obtained from a stereo triangulation system [Anderson et al., 2011b].
Joshi and Kriegman [2007] present an approach that recovers the shape of a rotating object with a distantly illuminated Lambertian surface – hence under image-wise varying viewpoint and image-wise varying illumination. Their fusion step is an application of [Nehab et al., 2005] with an appended smoothness constraint $E_s = \nabla^2 z$ penalizing high second order derivatives.
Du et al. [2011] restate [Nehab et al., 2005] for application in a two-camera Lambertian PS framework. For this, they apply the "filter flow framework" [Seitz and Baker, 2009], which is based on finding space-variant linear filters and linear constraints that transform one dataset into another. In detail, they restate the stereo correspondence problem and Nehab et al.'s surface normal error (Eq. 3.21) such that both are described by linear space-variant filters and linear constraints. The filter masks are then obtained using linear programming from which the object depth can be derived. However, the efficiency of Nehab et al.'s statement is lost since linear programming is computationally more expensive. Similar to other approaches, they assess large scale shape robustness and high reconstruction detail for several objects. Zhang et al. [2012] extend the approach by Nehab et al. [2005] by adding a smoothness constraint and discontinuity-adaptive weights. This improves the reconstruction result significantly for datasets with large depth differences and/or missing data. Additionally, they show that the framework can be extended for handling datasets with ill-posed intensity equations, i.e. less than three light sources. In that case, it is required to optimize the normal field and the depth map in an alternating iterative scheme, but the results are promising. The approach is evaluated qualitatively on real and synthetic data, where it produces sharp edges and smooth planes.

3.8.4 Further approaches

Lim et al. [2005] use feature points on a rigid Lambertian surface to estimate camera parameters, construct a coarse surface, and resolve the GBR ambiguity (see Section 3.5.2). Then, the normal map is computed by collecting corresponding intensities from each image and performing classic PS on that data, followed by integrability-error-based depth-from-normal estimation. This is repeated iteratively to refine the resolution. Note, that the algorithm does not fuse coarse depth and fine-detailed gradient information. Instead, the coarse depth data is solely used to estimate the information that is required to perform PS with unknown lighting and unknown camera movements. Other approaches for the reconstruction of the shape from moving objects have been proposed by Simakov et al. [2003]; Zhang et al. [2003], but both refrain from merging depth and image data to obtain a more accurate reconstruction result.
A Markov Random Field approach has been introduced by Diebel and Thrun [2006] and has later been used for 3D television applications by Huhle et al. [2007]. The approach as such does not rely on actual reflectance modeling but improves the depth profile near coinciding image and depth edges and reduces depth noise in smooth image parts.

A graph-cut-based approach for fusing depth and image data from two viewpoints has been proposed by Hahne and Alexa [2008], who minimize the total variation of the final surface using the functional

$$E = \int_{x,y} \underbrace{(I_L - I_R)^2}_{\text{stereo error}} + \lambda \underbrace{(z - z_{RS})^2}_{\text{depth error}} + \underbrace{\gamma z_x + \delta z_y}_{\text{regularization}} \, dx \, dy. \tag{3.23}$$

Since their evaluation focuses on the appearance of depth maps rather than the appearance of shaded scenes, differences to high quality data, or evaluations on known synthetic data, it is not possible to assess the accuracy of the method suitably.

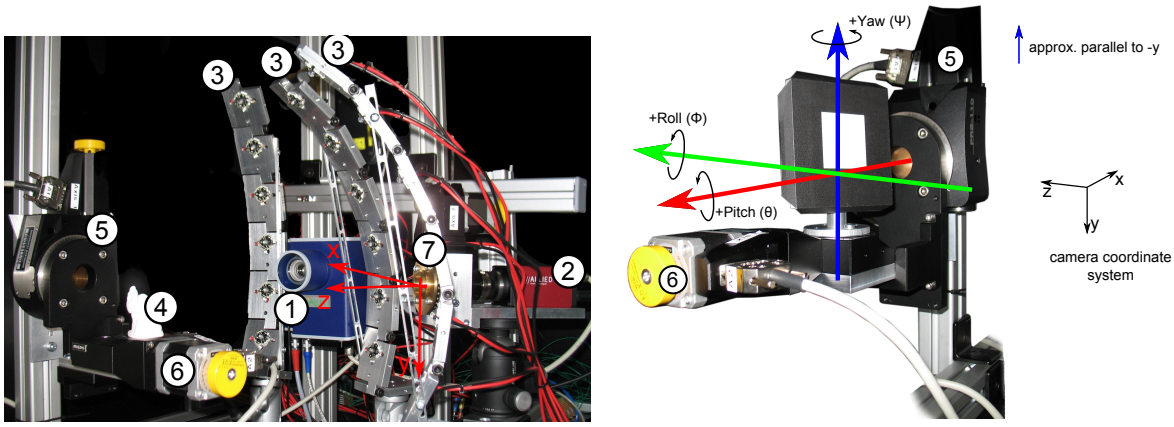
The fusion of different depth sensors (e.g. time-of-flight and multiview stereo) has been presented by Zhu et al. [2008]; Kim et al. [2009] but these approaches are inherently limited by the high frequency noise that is present in both of the data sources. Bernardini et al. [2002] acquire depth and image data, but do not fuse both data sources: The normal maps are used for rendering only.

4 Experimental setup

The experiments for this thesis have been conducted on an experimental setup that was built specifically for the research presented in this thesis. The setup contains a 3D range scanner consisting of an industrial camera and a projector, several LEDs for object illumination, and rotation stages for object positioning. The following chapter gives a description of the respective setup components and why they have been designed the way they are. It will become clear that an environment that is supposed to provide a basis for high accuracy measurements, needs exact component choice and placement.

This chapter is partly adapted and/or adopted from [Herbort et al., 2013a].

4.1 Setup description



(a) general overview **(b)** rotation axes

Figure 4.1.: (a) Overview of the experimental setup and its components: Projection unit ① and camera ② compose the 3D range scanning system. The object ④ is illuminated by 18 LEDs in three arcs ③ and can be pitched and/or yawed using the rotation stages ⑤ and ⑥. The third rotation stage ⑦ is used to adjust the orientation φ of a polarization filter positioned just in front of the camera lens. The red coordinate system indicates the camera coordinate system axes. (b) Orientation of the rotation axes in the camera coordinate system. Since the rotation stages ⑤ and ⑥ define the ultimate axis direction, these are not necessarily parallel to the camera coordinate system's axes.

An overview of the setup is shown in Fig. 4.1. The range scanner[1] consists of a projector unit (①) that directs light onto the screen using a $1024\,\text{px} \times 768\,\text{px}$ micro mirror array[2], and a $2048\,\text{px} \times 2048\,\text{px}$, 8 Bit monochrome CCD camera[3] (②) that observes the scene. The 18 light sources[4] (③) emit green light with $\lambda \approx 525\,\text{nm}$ and a luminous flux of $\approx 120\,\text{lm}$.

The rotation stages[5] (⑤, ⑥) shown in Fig. 4.1(b) allow highly accurate pitch and yaw rotations with an average angular resolution of $0.002°$, unidirectional repeatability of $0.002°$, and bi-directional repeatability of $\pm 0.01°$. Details regarding a method for the calibration of the rotation axis direction and its position in the camera coordinate system are given in Appx. C.2. A third rotation stage (⑦) controls the rotation of a polarization filter mounted in front of the camera lens.

[1] ViALUX zSnapper Vario
[2] DLP Discovery 4100
[3] AVT pike 421B with Schneider-Kreuznach 2.8/50 Xenoplan optics
[4] Seoul P4 LED
[5] miCos PRS-110, DT-65N, DT-80R

4.2 Camera calibration and configuration

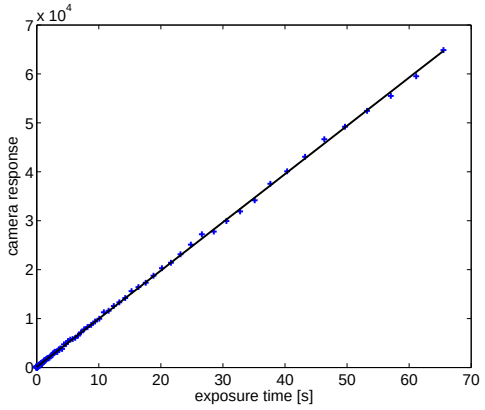
For unbiased data acquisition, it is required to examine and calibrate the camera, which includes response curve analysis, intensity noise analysis, dark frame subtraction, hot pixel analysis, aperture configuration, and lens distortion removal.

The examination of the camera response curve (see Fig. 4.2(a)) shows an almost perfect linear response curve as it is typical for CCD sensors. The intensity noise σ as such exhibits the typical $\sqrt{I}$-dependent behavior (see Fig. 4.2(b)) and relative noise $\frac{\sigma}{\mu}$ below 1% within $[3.8\%...98.7\%]$ of the dynamic range (see Fig. 4.2(c)). This is the range that has later been chosen for high dynamic range image computation (see Section 4.3 below).

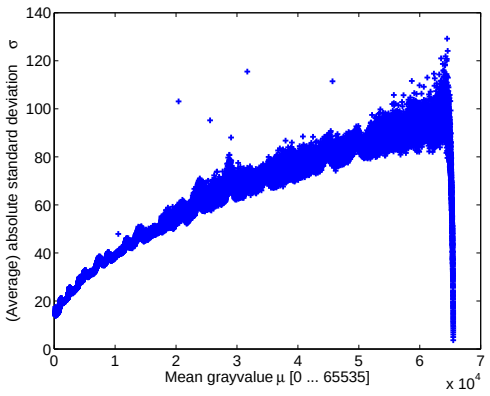
Besides the response curve, it is important to examine and compensate the sensor's dark noise. It can be deducted that the dark noise shows significant spatial and exposure time dependence as shown in Fig. 4.3. Therefore, each object image taken with some exposure time τ is compensated with dark frames of the same exposure time τ even though the actual noise level is rather low (on average ≈ 20 of 65535 levels of gray), but it becomes important for very dark image regions.

Apart from the dark noise, it is another quality criterion to examine the number of hot pixels on the sensor. Fortunately, the sensor is of high quality and there are no hot pixels apparent for exposure times of $\tau \leq 512\,\text{ms}$. Above, there are few pixels affected, but there are still only $441\,\text{px} \mathrel{\hat{=}} 0.01\%$ hot for the maximum exposure time of $\tau = 65536\,\text{ms}$; details are given in Appx. C.1.1. For our object images, hot pixels are a minor issue due to the high dynamic range procedure described below, where saturated pixels are excluded a priori.

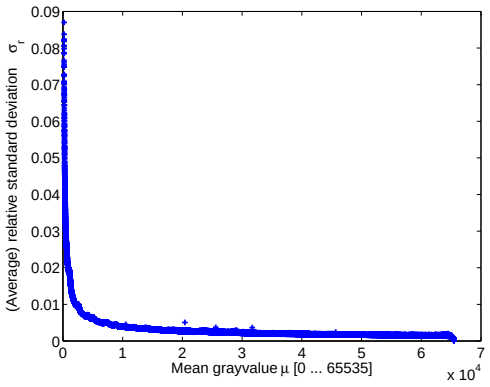
One of the important camera parameters that has a strong effect on the properties of the acquired images is the aperture f of the optical system of the camera. Generally, a widely opened aperture (small f-number) allows shorter exposure times. At the



(a) camera brightness response



(b) absolute standard deviation σ



(c) relative standard deviation $\sigma_r = \frac{\sigma}{\mu}$

Figure 4.2.: (a) 16 bit camera response, fitting $\text{RMSE}_{16\,\text{bit}} = 206.84$, fitting $\text{RMSE}_{8\,\text{bit}} = 0.802$. (b) typical $\sqrt{I}$-dependent noise behavior. (c) for $\approx 3.8\%$ of the grayvalue range or above, the noise drops below 1%. Above $\approx 98.7\%$ of the grayvalue range, the cells approach saturation and the standard deviation drops due to a lack of proper capturing of high grayvalues.

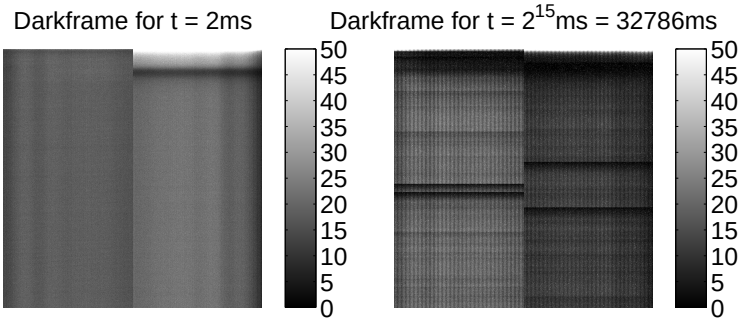


Figure 4.3.: AVT pike - examples for darkframes at different exposure times. Note the spatial variation of the sensor noise and it's dependency upon the exposure time.

same time, the depth of field decreases. Closing the aperture for a larger depth of field comes at the cost of less light passing through the optical system and thus requiring longer exposure times for the same image contrast. Note, that for very small apertures, the light rays are diffracted at the edges of the aperture. This causes globally a large depth of field, but locally these images are less crisp than those within the depth of field of a widely opened aperture. With that in mind, a medium-sized aperture of $f = 4...8$ has been chosen for image acquisitions for the experiments conducted for this thesis.
The camera distortions and parameters have been determined using Bouguet's MATLAB camera calibration toolbox [Bouguet, 2008], which is based on the principles by [Zhang, 1999; Tsai, 1987]. Due to the high quality of the camera optics, there are only very slight distortions visible near the corners of the images.

4.3 Robust high dynamic range (HDR) imaging

Since real (natural) luminous intensities range from $0.001\ \frac{cd}{m^2}$ (starlight) up to $100\ 000\ \frac{cd}{m^2}$ (sunlight), the dynamic range may amount up to $\frac{100\ 000}{0.001} = 10^8 = 160\ dB$[6] [Schulz et al., 2007]. While not having a range of $160\ dB$, the high intensities of specular spikes in our object images are still too bright to be captured with our camera. The HDR procedure thus works as follows: Since the camera response curve is linear (see Section 4.2), the perceived intensity for each pixel (u, v) is assumed to be sufficiently well described by

$$\mathbf{I} = I_{\text{HDR}}\mathbf{t}_{exp} + b \tag{4.1}$$

where $\mathbf{I} = [I_1, ..., I_L]^T$ denotes a vector of L intensity measurements corresponding to the exposure times $\mathbf{t}_{exp} = [t_1, ..., t_L]^T$. The quantity I_{HDR} denotes the slope of the exposure per pixel per unit of exposure time t_{exp} and b is an offset that compensates possible deviations from the model due to noise or measurement inaccuracies. Typically, b is very small compared to the obtained pixel intensities. Both, I_{HDR} and b are computed with a simple linear regression for each pixel (u, v):

$$I_{\text{HDR}} = \frac{\frac{1}{L}\sum_{l=1}^{L}(t_l - \bar{\mathbf{t}})(I_l - \bar{\mathbf{I}})}{\frac{1}{L}\sum_{l=1}^{L}(t_l - \bar{\mathbf{t}})^2} = \frac{\text{covar}(\mathbf{I}, \mathbf{t})}{\text{var}(\mathbf{t})} = \frac{\sigma_{\mathbf{tI}}}{\sigma_{\mathbf{t}}^2}, \qquad b = \bar{\mathbf{I}} - I_{\text{HDR}}\bar{\mathbf{t}}, \tag{4.2}$$

where $\bar{\mathbf{I}}$ and $\bar{\mathbf{t}}$ denote the arithmetic means of $\mathbf{I}$ and $\mathbf{t}$, respectively.

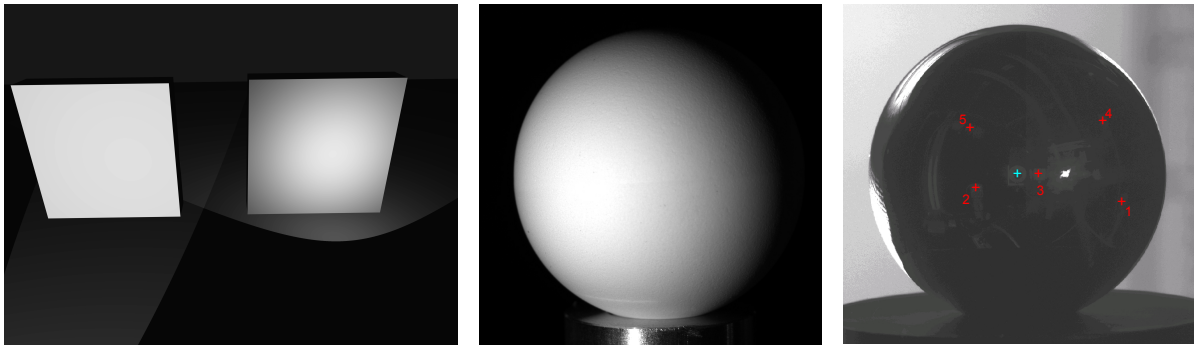
[6] $P_{dB} = 20 \cdot log_{10}(P_{linear})$

5 Contribution - light source calibration

In an image and depth data acquisition setup for surface reconstruction algorithms, it is ideal to know the incident light direction and intensity. That knowledge makes the difference between *calibrated* and *uncalibrated* PS. Since uncalibrated PS needs to determine its parameters from an arbitrary scene, it is prone to additional sources of inaccuracies (see Section 3.5.2). Therefore, calibrated PS is to be favored if the goal is high reconstruction accuracy.
This chapter introduces a new method for the calibration of light source position and intensities using a single sphere made of a diffusely reflecting material as a calibration object. The advantage of that approach for light source calibration over related methods lies in the high accuracy and the need for only a single image of a single calibration object. Accuracy and robustness are evaluated based on confidence bands and compared to common alternatives, and it is found that all of them perform less accurately.

This chapter is adapted and/or adopted from [Lenoch, Herbort, and Wöhler, 2012].

Commonly, the calibration of light sources is achieved by placing a specularly reflecting sphere in the camera's field of view, as shown e.g. by Lensch et al. [2003]. Conceptually, this approach is vulnerable to falsely determined sphere surface normals which typically result from an inaccurate segmentation of the sphere in the image. Furthermore, the approach is typically unable to determine the light source intensity since the regarded pixels are usually saturated due to bright light sources or cameras with a limited dynamic range. Even if saturation is not reached, the intensity then depends on very few pixels, which makes it very sensitive to noise as stressed by Goldman et al. [2005]. The approach is additionally restricted to the extraction of the incident *direction*, which is only reasonable for the whole image if distant light sources can be assumed. If that condition is not met (non-distant light sources), the incident light direction varies locally, which causes local intensity variations as shown in Fig. 5.1(a). While the application of several



(a) **(b)** **(c)**

Figure 5.1.: (a) influence of non-distant light sources; (b)+(c) spherical calibration targets with 30 mm in diameter. For reference, the grayvalues of the sphere (b) in the depicted 8 Bit image range from 0 to 155, sphere (c) shows saturated pixels. The indicators in the specular sphere show the location of the camera center (cyan), and the mapped positions of the LED reflections (red). Note, that the sphere center agrees with the visible reflection of the camera lens center and that the LEDs are visible as highlights / mirror reflections.

specular spheres allows the determination of the sphere position [Lensch et al., 2003], this still prohibits the calibration of the illumination strength for the above reasons and requires the error-prone segmentation of several spheres and determination of their respective positions in the real world. In this study, it is proposed to apply a single diffusely reflecting sphere[1] (see Fig. 5.1(b)) for light source calibration, which allows the determination of the light source position and its intensity from a single image of a single diffuse sphere. In contrast to [Zhou and Kambhamettu, 2002] and [Goldman et al., 2005], this does not rely on a combination of specular and diffuse reflections: It is shown that the diffuse sphere is sufficient for robust position and intensity calibration. In the following, the experimental setup is presented (Section 5.1) along with mathematical terms and reflectance properties of the applied diffuse sphere. In Section 5.3, the proposed algorithm is explained and details regarding the optimization procedure are given. The experimental results are described in Section 5.4 along with their discussion. The chapter is concluded in Section 5.5 and aspects for future work are specified.

[1] 30 mm diameter, manufactured by OptoPolymer, Munich

5.1 Mathematical description

Fig. 5.2 illustrates the camera[2]/range scanning system[3], which provides pixel-synchronous depth and intensity data $x(u,v), y(u,v), z(u,v)$, and $I(u,v)$. These are indexed by the sensor pixel coordinates u and v. The origin of the camera coordinate system (x_c, y_c, z_c) is located at the camera lens center. The camera

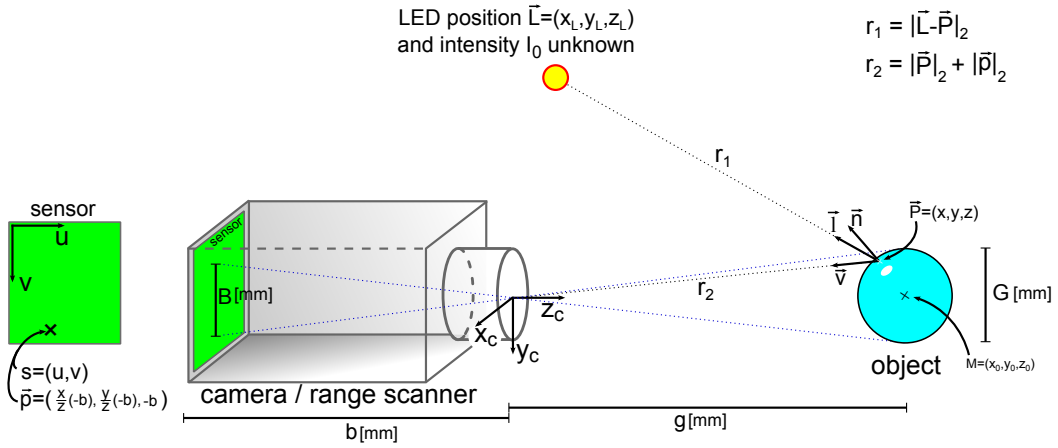


Figure 5.2.: Sketched experimental setup with mathematical definitions.

sensor is located at the camera constant distance b behind the lens. The vector $\mathbf{p} = (\frac{x}{z}(-b), \frac{y}{z}(-b), -b)$ denotes the projection of the object point $\mathbf{P} = (x, y, z)$ onto the sensor, where it is denoted $s = (u, v)$. The distance between $\mathbf{P}$ and its projection on the sensor is denoted r_2. At $\mathbf{P}$, the local illumination direction is $\mathbf{l} = \mathbf{l}(x, y, z)$, the local viewing direction is $\mathbf{v} = \mathbf{v}(x, y, z)$ and the local normal is $\mathbf{n} = \mathbf{n}(x, y, z)$. Each of these vectors is normalized such that their Euclidean norm is 1. The location of the (point) light source is $\mathbf{L} = (x_l, y_l, z_l)$, which has a distance of r_1 from a surface point $\mathbf{P}$ and exhibits an intensity of I_0. According to the pinhole camera model, the object of size G is located at a distance g from the camera and its image has size B on the sensor.

5.2 Calibration object

The applied spherical calibration object has been found to be Lambertian [Lambert, 1760] with a slight specular component, which is modeled by a specular lobe [Phong, 1975] such that the overall observed intensity becomes

$$R(u,v) = \frac{I_0}{r^2(u,v)} \left[\langle \mathbf{n} \cdot \mathbf{l} \rangle + 0.0535 \langle \mathbf{v} \cdot \mathbf{r} \rangle^{2.0648} \right]. \tag{5.1}$$

Note that the model accounts for locally varying distances between light source and object ($r(u,v)$).

[2] AVT pike 421B, 14 Bit monochrome CCD camera, $2048 \times 2048\,px$, typical measurement noise standard deviation amounts 1% of the measured intensity.

[3] ViALUX zSnapper Vario, structured/modulated light range scanner, typical measurement noise standard deviation per direction (x, y, z) is $\sigma_x = \sigma_y = 0.06$ mm, $\sigma_z = 0.12$ mm.

Exemplary rendered images are shown in Fig. 5.3 along with the pixelwise intensity error maps. The mean RMSEs over all 18 LEDs on 8 Bit images ($[0...255]$) amount 4.0 for the Lambertian case and 3.1 for the Lambert/Phong according to Eq. 5.1.

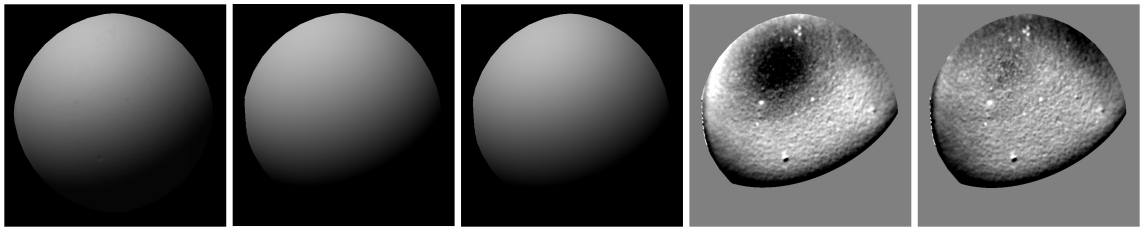


Figure 5.3.: From left to right: (1) Camera image, (2) Image rendered using Lambertian reflection only, (3) Image rendered using the Lambertian model and a Phong lobe, (4) pixelwise error for the Lambertian model, (5) pixelwise error for the Lambert + Phong model. For the error maps, the grayscale colors are: Black: <=-7.3, White: >= +7.3. For the images, the grayscale colors are: Black: 0, White: 255.

5.3 Light source calibration algorithm

The idea behind the proposed approach lies in the optimization of the appearance of the sphere in its reflectance map $R(u, v)$ compared to its image $I(u, v)$ with respect to light source position $\mathbf{L} = (x_l, y_l, z_l)$ and light source intensity I_0 such that

$$\{\mathbf{L}^*, I_0^*\} = \underset{x_l, y_l, z_l, I_0}{\operatorname{argmin}} \sum_{u,v} \left(I(u, v) - R(u, v)\right)^2 \tag{5.2}$$

In analogy to computer graphics, this corresponds to moving the light source and adjusting its intensity until the rendered image R matches the observed image I as well as possible. The general components of that algorithm are depicted in Fig. 5.4. If there are measured depth data available, we perform a linear

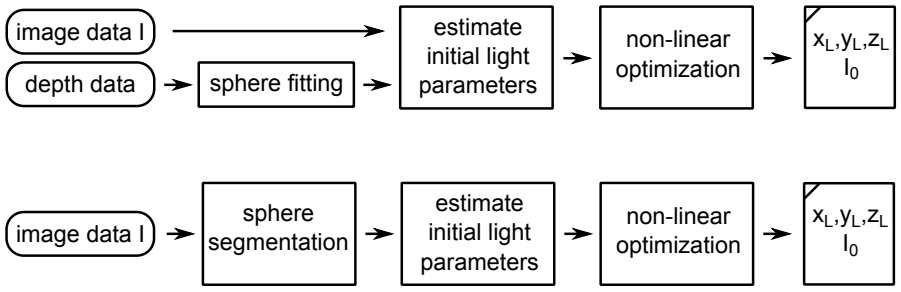


Figure 5.4.: Algorithm overview with depth and image data (top) or solely image data (bottom) available.

sphere fit to the data to make the data more robust. In fact, under the assumption of noise in the typical range of the inaccuracies of the applied range scanner[3] and over 100 iterations, the sphere center and radius are estimated with very high robustness, i.e., $\tilde{y_0} = \pm 0.32 \mu m$, $\tilde{z_0} = \pm 1.27 \mu m$, $\tilde{R} = \pm 0.87 \mu m$. This is due to the very large number of ≈ 250000 depth data measurement triplets that are available in a scanned 3D point cloud. If there is only image data available, the sphere is initially segmented from its black background using a simple thresholding approach. Afterwards, the optimal sphere position is found

by nonlinearly optimizing its position $\mathbf{M} = (x_0, y_0)$ and Radius r according to

$$\{\mathbf{M}^*, r^*\} = \underset{x_0,y_0,r}{\operatorname{argmin}} \left(|I - R|^2 + \gamma N\right) \tag{5.3}$$

where $N \in \mathbb{R}^{>0}$ denotes the number of image pixels outside the currently segmented sphere and γ denotes a Lagrange multiplier. Afterwards, the segmented sphere is used to compute the surface normals corresponding to each intensity pixel.

Initialization and optimization

An initialization for the direction and intensity of the light source is obtained by solving the linear image intensity equation $I(u,v) = \rho I_0 \langle \mathbf{n}(u,v) \cdot \mathbf{l} \rangle$ (Lambert's law) in a least-squares sense for the global illumination direction $\mathbf{l} \in \mathbb{R}^{3\times 1}$, i.e.

$$\mathbf{l}^* = \mathbf{l}\rho I_0 = [\mathbf{n}(u,v)]^{-1} \cdot I(u,v) \tag{5.4}$$

which requires a pseudoinverse $[.]^{-1}$ and uses an albedo value of $\rho = 0.99$ as measured by the sphere manufacturer. The initial source intensity value is obtained by computing the Euclidean norm $|.|_2$ of $\mathbf{l}^*$ and the direction $\mathbf{l}$ is computed by normalizing $\mathbf{l}^*$ to a length of 1. This is then used to initialize the non-linear optimization[4], which minimizes the error function according to Eq. 5.2 and thus estimates the light source position.

5.4 Experimental results

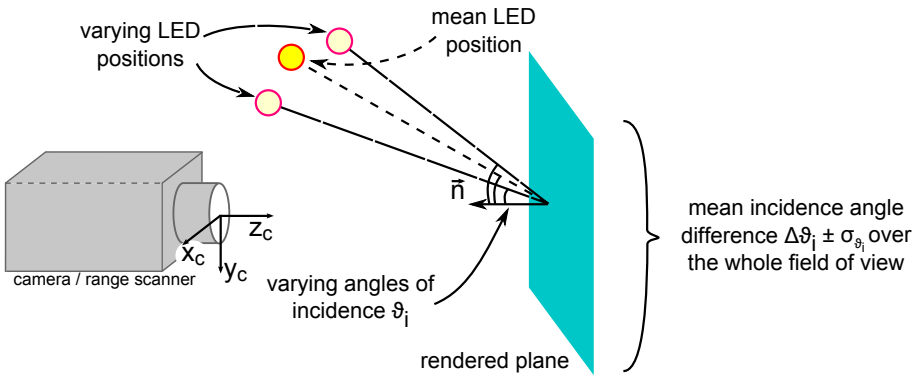


Figure 5.5.: Error assessment principle.

Two different criteria are used to assess our results. To evaluate the robustness of the approach, confidence bands are computed under repeated algorithm execution and application of synthetic measurement noise[2,3]. To evaluate the accuracy of the results, the average local incident angle differences $\Delta\vartheta_{in}$ are computed

[4]Trust-Region-Reflective Optimization using *lsqnonlin* in Matlab, see [Coleman and Li, 1994] and [Coleman and Li, 1996] for methodical details.

over a plane that fills the whole field of view of the camera:

$$\Delta\vartheta_{in} = \frac{180^\circ}{\pi \cdot U \cdot V} \sum_{k=1}^{K} \sum_{u,v}^{U,V} |\arccos(\mathbf{l}_{\text{mean}}(u,v) \cdot \mathbf{n}(u,v)) - \arccos(\mathbf{l}_k(u,v) \cdot \mathbf{n}(u,v))| \tag{5.5}$$

This assesses the average local incidence error when an object is observed. To emphasize the need for robust illumination calibration, the uncertainty of the common approach using a specular sphere is shown in Fig. 5.6(a). The sphere parameters (R, x_0, y_0) are assumed to be known up to an inaccuracy of $\sigma = 4$ pixels, which corresponds to $\approx 0.16\,\text{mm}$. Note the increasing error for specular reflections close to the border of the sphere (LEDs $1, 7, 13$), which is due to the local gradients being more sensitive against missegmentations, since the projected surface curvature increases towards the sphere border.

5.4.1 Image and depth data

The calibration with a diffuse calibration object (see Fig. 5.6(b)) shows significantly increased robustness compared to the classical approach with a specular sphere. Even in the case of depth measurement errors

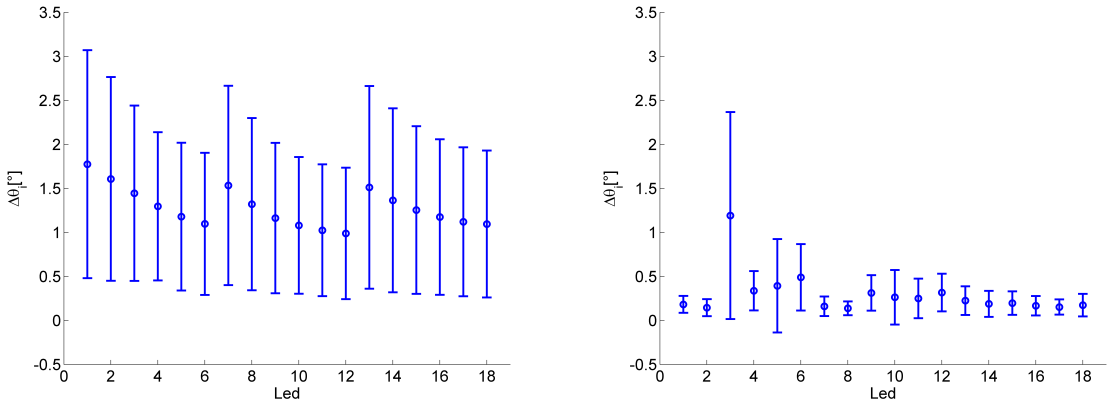


Figure 5.6.: 1σ-confidence band of $\Delta\vartheta_{in} \pm \sigma_{\vartheta_{in}}$ for calibration with a single specular sphere ((a), 1000 iterations) and a single diffuse sphere ((b), 100 iterations).

(LED 3), the robustness is still comparable to the specular case. Numeric results for an exemplary LED are shown in Table 5.1.

criterion	1σ-confidence band / value
sphere center (x_0, y_0, z_0)	$(\sigma_{x_0}, \sigma_{y_0}, \sigma_{z_0}) = (0.75, 0.32, 1.27)\,\mu\text{m}$
sphere radius (R)	$\sigma_R = 0.87\,\mu\text{m}$
light source position (x_L,y_L,z_L)	$(\sigma_{x_L}, \sigma_{y_L}, \sigma_{z_L}) = (1.1, 0.8, 0.1)\,\text{mm}$
light source intensity for 8 Bit (I_0)	$\sigma_{I_0} = 0.005$
distance to sphere (D)	$D = 264.7 \pm 0.07\,\text{mm}$
$\Delta\vartheta_{in} \pm \sigma_{\vartheta_{in}}$	$\Delta\vartheta_{in} \pm \sigma_{\vartheta_{in}} = 0.15^\circ \pm 0.097^\circ$

Table 5.1.: Robustness evaluation results for LED 2 with available depth data.

To investigate the possibility of using more than one diffuse sphere, confidence bands have been computed while altering the number of image and depth data from one sphere to a total of ten spheres. The calculation is based on 100 iterations and applies the typical noise to the image[2] and depth[3] data. The results are shown in Fig. 5.7. Clearly, using several spheres increases the robustness even more, but only minor improvements are achieved for more than 3 spheres.

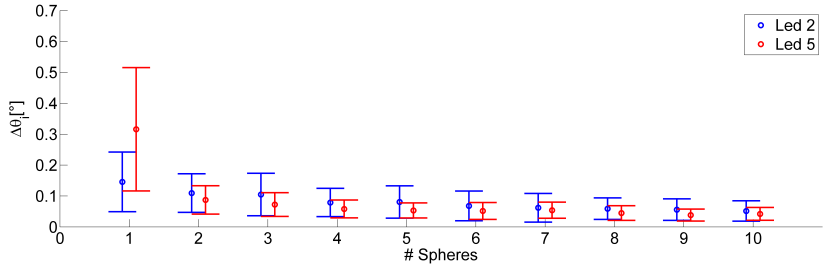


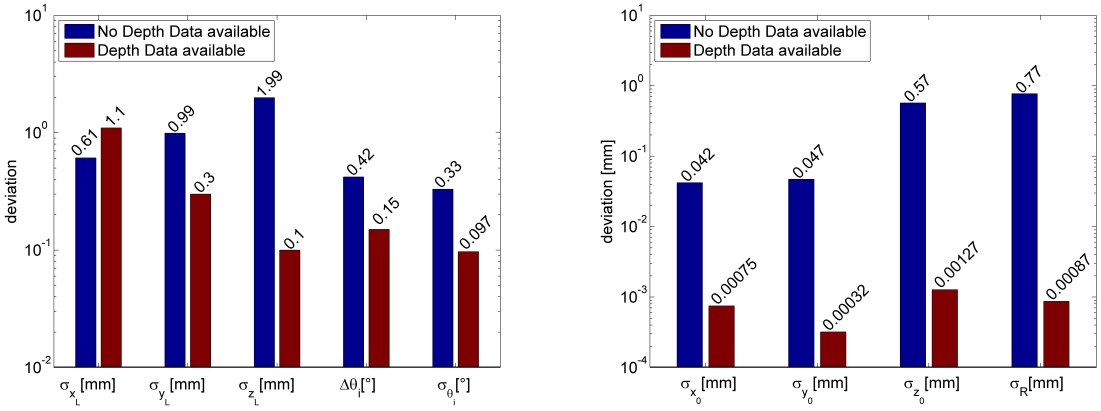
Figure 5.7.: 1σ-confidence band of $\Delta\vartheta_{in}$ in dependence on the number of spheres.

5.4.2 Image data only

1σ-confidence band / value	LED 2	LED 5
sphere center $(\sigma_{x_0}, \sigma_{y_0}, \sigma_{z_0})$	$\pm(0.042, 0.047, 0.57)$ mm	$\pm(0.07, 0.06, 1.0)$ mm
sphere radius (σ_R)	± 0.77 mm	± 0.5 mm
light source position $(\sigma_{x_L}, \sigma_{y_L}, \sigma_{z_L})$	$\pm(0.61, 0.99, 1.99)$ mm	$\pm(4.8, 1.6, 10.9)$ mm
light source intensity for 8 Bit (σ_{I_0})	± 0.23	± 0.01
distance to sphere (D)	191 ± 2.3 mm	214 ± 12.4 mm
$\Delta\vartheta_{in} \pm \sigma_{\vartheta_{in}}$	$0.42° \pm 0.33°$	$1.2° \pm 0.62°$

Table 5.2.: Robustness evaluation results for the case of unavailable depth data (200 iterations, 2% image noise).

Numeric results for the robustness of the presented approach without depth data are shown in Table 5.2. The obtained results are again robust against noise as shown by the small confidence band widths. Additionally, the average incidence angle difference over the whole field of view amounts to merely $0.42°/1.2°$. However, the distance of the light source from the calibration object shows a systematic error since the values deviate from manually measured values (225 ± 5 mm for both LEDs). This seems to appear due to unmodelled effects like the light sources' radiation characteristic. Since the obtained light source position is optimal by means of the calibration sphere appearance and assumes a point light source, the obtained positions can be regarded as virtual (point) light source positions, which are typically used for surface rendering anyway. Fig. 5.8 summarizes the difference between the case of available depth data compared to the case of unavailable depth data. As expected, the error is significantly larger, but, as discussed above, still smaller than for the classic approach.



(a) Light position confidence bands

(b) Sphere position confidence bands

Figure 5.8.: Results (LED 2) with and without available depth data. Note the logarithmic scale.

5.5 Conclusion

The results show an increased robustness compared to the classical method using specularly reflecting spheres. Additionally, positions and intensities are obtained instead of directions only, as it is the case for previously proposed algorithms. Since the approach currently models point light sources, it is not directly applicable to light sources with a strong directional characteristic. However, this can easily be overcome by incorporation of that characteristic in the reflectance model. An aspect that requires improvement is the robustness of the segmentation of the diffuse sphere. This regards e.g. inclusion of edge and gradient cues, intensity characteristics and/or active contours. The minimization algorithm as such, however, is very accurate and robust as shown in the previous sections.

6 Contribution - surface reconstruction

This chapter contains the surface-reconstruction-related contributions of this thesis. The overall goal is the development of an algorithm for robust, dense and accurate 3D surface shape determination. The presented algorithms show that it is possible to estimate and fuse gradient fields with absolute depth data, even though the BRDF of the regarded surfaces, and the amount and locations of interreflections, are unknown a priori. The obtained surfaces show the desired large scale robustness and small scale detail accuracy, which are evaluated on a whole set of surfaces with Lambertian, specular, and metallic reflectance.

This chapter is partly adapted and/or adopted from [Herbort et al., 2011], [Herbort and Wöhler, 2012], [Herbort et al., 2013b], and [Herbort et al., 2013a].

6.1 Absolute depth and gradient field fusion

The following sections give a detailed derivation of the proposed algorithm for the reconstruction of non-Lambertian surfaces with possible occurrence of interreflections. Initially, Horn's variational SfS approach [Horn, 1990] is explained, which iteratively minimizes a two-component error functional for simultaneous depth and gradient field retrieval (Section 6.1.1). As proposed in [Herbort et al., 2011], it can be extended towards the inclusion of absolute depth data by appending a third error component that penalizes deviations from the large scale shape (Section 6.1.2).
Afterwards, the aspect of computational efficiency is regarded in Section 6.1.3 by means of efficient reflectance map derivative computation. Section 6.1.4 illustrates how large values in the reflectance map derivatives impair the optimization robustness and an improvement is proposed as published in [Herbort et al., 2013a]. The algorithmic descriptions close with a discussion of how the occurrence of interreflections is implicitly contained in the variational optimization scheme, but that the alternative approach [Herbort et al., 2013b] is favorable by means of computational expenses (Section 6.1.5).

6.1.1 Variational Shape from Shading

Given an image $I(x, y) \in \mathbb{R}$ indexed by its continuous image coordinates $x \in [0...X]$ and $y \in [0...Y]$, Horn [1990] tries to find the integrable surface field $z^*(x, y) \in \mathbb{R}$ with the surface gradient fields $p(x, y)$ and $q(x, y)$ that belong to the (unnormalized) surface normal field

$$\tilde{\mathbf{n}}(x, y) = \begin{bmatrix} p(x, y) \\ q(x, y) \\ -1 \end{bmatrix} \tag{6.1}$$

and whose "reflectance map"

$$R(x, y) = R(p(x, y), q(x, y), \mathbf{l}(x, y), \mathbf{v}(x, y), P, M) \in \mathbb{R}^+ \tag{6.2}$$

best matches the observed image $I(x, y)$. The reflectance map for each pixel is computed by rendering the underlying surface based on its normals p, q, incident light direction $\mathbf{l}$, viewing direction $\mathbf{v}$ and a reflectance model M with parameters P that are assumed to be known a priori.
Mathematically stated, the problem above becomes[1]

$$\{z^*, p^*, q^*\} = \underset{p,q,z}{\arg\min} \underbrace{\int_{x,y} (I - R)^2 dx\, dy}_{E_I} + \gamma \underbrace{\int_{x,y} \left[(z_x - p)^2 + (z_y - q)^2\right] dx\, dy}_{E_{\text{int}}} \tag{6.3}$$

with the intensity error E_{I}, the integrability error E_{int} and the partial derivatives z_x and z_y of the surface z. The integrability error serves as a regularization constraint and determines the error of the integrability of the gradient fields p and q. I.e., it tries to find the surface $z(x, y)$ whose gradients z_x and z_y are most similar to p and q. According to Horn [1990], the solution for p and q is found iteratively by a combination

[1] For the sake of clarity, the indices (x, y) for I, R, p, q and z have been omitted.

of an ordinary minimization calculus and a calculus of variations in the discrete space $\{.\}$ by computing

$$\{p^{(n+1)}\} = \{z_x^n\} + \frac{1}{\gamma}(I - R)R_p \qquad \text{and} \tag{6.4}$$

$$\{q^{(n+1)}\} = \{z_y^n\} + \frac{1}{\gamma}(I - R)R_q. \tag{6.5}$$

With p and q being known, the optimized depth z is obtained from the Poisson equation[2] $\Delta z = p_x + q_y$ yielding

$$\{z^{(n+1)}\} = \{\bar{z}^n\} - \frac{1}{4}(\{p_x^{(n+1)}\} + \{q_y^{(n+1)}\}) \tag{6.6}$$

[Horn, 1986, 1990]. Note that Eq. 6.6 applies a discrete approximation of the Laplace operator using the four nearest neighbors [Sonka et al., 2008, p. 136] as proposed by Horn [1986]. On the boundary, the optimized surface satisfies

$$\mathbf{n}^T \begin{bmatrix} z_x \\ z_y \\ -1 \end{bmatrix} = \mathbf{n}^T \begin{bmatrix} p \\ q \\ -1 \end{bmatrix}. \tag{6.7}$$

6.1.2 Inclusion of absolute depth data

The approach by Horn [1990] is now extended by the possibility to enhance the result with independently measured absolute depth data. The error terms according to Eq. 6.3 are thus appended by the range scanner error

$$E_{RS} = \int_{x,y} [f(Z_x) - f(p)]^2 + [f(Z_y) - f(q)]^2 \, dx\, dy, \tag{6.8}$$

where Z denotes the measured range scanner depth data and $f(.)$ denotes a spatial low-pass filter. The error thus penalizes deviations of the measured large-scale surface gradients ($f(Z_x)$ and $f(Z_y)$) from the optimized large-scale gradients ($f(p)$ and $f(q)$). This deliberately disregards high spatial frequencies, which are incorporated by the intensity error and integrability error, and thus only prevent global shape deviations, which are typical for SfS.

The intensity error is extended to work on K images similar to Horn [1986], which merely means that

$$E_{\mathrm{I}} = \sum_{k=1}^{K} \frac{1}{K} \int_{x,y} (I_k - R_k)^2 \, dx\, dy. \tag{6.9}$$

The overall reconstruction error thus becomes

$$E_\Sigma = E_I + \gamma \, E_{int} + \delta \, E_{RS}, \tag{6.10}$$

[2] To minimize an equation of the form $\int_x \int_y f(z, z_x, z_y) \, dx \, dy$ with regard to $z(x, y)$, the Euler equation is $\frac{d}{dz} f - \partial_x \frac{d}{dz_x} f - \partial_y \frac{d}{dz_y} f = 0$, which can be rewritten as $f_z - \partial_x f_p - \partial_y f_q$ when using $z_x = \partial_x z = p$, $z_y = \partial_y z = q$, and $f_z = \frac{d}{dz} f$. Applying this to the integrability error yields the Poisson equation $\Delta z = p_x + q_y$, see [Horn, 1986, p.266]

where $\gamma, \delta \in \mathbb{R}^+$ denote weighting factors that determine the individual relative importance of all three errors. Similar to Horn [1990], the minimization of E_Σ can be achieved using an ordinary calculus, which is solved iteratively:

$$p = z_x + \frac{1}{\gamma}(I - R)R_p + \frac{\delta}{\gamma}\left[f(Z_x) - f(z_x)\right] f_p(z_x) \text{ and} \tag{6.11}$$

$$q = z_y + \frac{1}{\gamma}(I - R)R_q + \frac{\delta}{\gamma}\left[f(Z_y) - f(z_y)\right] f_q(z_y). \tag{6.12}$$

Just as Horn [1990], an iterative discrete scheme for solving those equations is applied such that

$$p^{(n+1)} = z_x^{(n)} + \frac{1}{\gamma}\underbrace{\left(I - R(p^{(n)}, q^{(n)})\right) R_p(p^{(n)}, q^{(n)})}_{\partial_p E_I} + \frac{\delta}{\gamma}\underbrace{\left(f(Z_x) - f(p^{(n)})\right) f_p(p^{(n)})}_{\partial_p E_{RS}} \tag{6.13}$$

$$q^{(n+1)} = z_y^{(n)} + \frac{1}{\gamma}\underbrace{\left(I - R(p^{(n)}, q^{(n)})\right) R_q(p^{(n)}, q^{(n)})}_{\partial_q E_I} + \frac{\delta}{\gamma}\underbrace{\left(f(Z_y) - f(q^{(n)})\right) f_q(q^{(n)})}_{\partial_q E_{RS}}. \tag{6.14}$$

Note that the curly brackets have been omitted in comparison to Eqs. 6.4 and 6.5 for clarity, but Eqs. 6.13 and 6.14 are analogously computed in the discrete domain. Better convergence properties have been experienced when evaluating $R(p^{(n)}, q^{(n)})$ and all other p, q-dependent functions at $(z_x^{(n)}, z_y^{(n)})$.
With p and q being solvable, the optimal surface z needs to be determined. Since E_I and E_{RS} are not explicitly dependent on z, it is only required to find a surface that minimizes the integrability error

$$E_{int} = \int_x \int_y (z_x(x,y) - p)^2 + (z_y(x,y) - q)^2 \, dx \, dy. \tag{6.15}$$

as defined by Horn [1986]. Equivalently to Horn [1990], Eq. 6.15 is solved using a calculus of variations which yields the same Poisson equation and the same solution for z is obtained as in Eq. 6.6. Convergence of the iterative process is reached once the error stops decreasing (and usually starts to increase).

6.1.3 Computational efficiency

For the practical application of the algorithm it is crucial to optimize the computationally expensive elements. This specifically regards the second term of Eqs. 6.13 and 6.14, which is actually $\partial_p E_{\text{RS}}$ and $\partial_q E_{\text{RS}}$, respectively. Under the assumption that f denotes a linear filter, it can be computed efficiently by using two distinct 2D linear filter operations on the whole image [Grumpe et al., 2011; Wöhler and Grumpe, 2013]. For an explanation, the discrete case of the derivative in Eq. 6.13 is examined:

$$\begin{aligned}\partial_p E_{RS}|_{u,v} = & -\sum_{i=1}^{U}\sum_{j=1}^{V} A(i,j)\left[\sum_{k=-\frac{K}{2}}^{\frac{K}{2}}\sum_{l=-\frac{L}{2}}^{\frac{L}{2}} F_{RS}(k,l)\left(\partial_x Z|_{i+k,j+l} - p(i+k,j+l)\right)\right] \\ & \cdot\left[\sum_{k=-\frac{K}{2}}^{\frac{K}{2}}\sum_{l=-\frac{L}{2}}^{\frac{L}{2}} F_{RS}(k,l)\partial_p p(i+k,j+l)|_{u,v}\right].\end{aligned} \tag{6.16}$$

where the image is assumed to be of size $U \times V$ and the filter matrix F_{RS} is of size $K \times L$ and A denotes the spatial extent of a pixel. In the following, the filter is applied to the data by a correlation ($\circ$) rather than a convolution ($*$). Note, that the summed elements are only non-zero if the indices fulfill $u = i + k$ and $v = j + l$. Those elements can thus be left out and the indices i and j are shifted to be centered at u and v:

$$\begin{aligned}\partial_p E_{RS}|_{u,v} &= \sum_{i=-\frac{K}{2}}^{\frac{K}{2}} \sum_{j=-\frac{L}{2}}^{\frac{L}{2}} F_{RS}(-i,-j) A(u+i, v+j) \\ &\quad \cdot \left[\sum_{k=-\frac{K}{2}}^{\frac{K}{2}} \sum_{l=-\frac{L}{2}}^{\frac{L}{2}} F_{RS}(k,l) \left(\partial_x Z|_{u+k,v+l} - p(u+k, v+l) \right) \right], \qquad (6.17) \\ &= [F_{RS} \cdot A] * [F_{RS} \circ (\partial_x Z - p)]. \qquad (6.18)\end{aligned}$$

$\partial_p E_{RS}$ can hence be obtained by correlating F_{RS} with the gradient difference ($\partial_x z_{RS} - p$) followed by a convolution with the filter matrix F_{RS} and the spatial extent A. The term $\partial_q E_{RS}$ is computed analogously. In summary, this approach improves the computational efficiency significantly without loosing any accuracy.

6.1.4 Reflectance map derivatives under specular reflection

Since most real surfaces exhibit specular spikes with high intensities and narrow angular extents, the derivatives of the reflectance map R with respect to the gradients p and q (see Eq. 6.4 and Eq. 6.5) exhibit large values near and at specular highlights. This is analytically correct, but during the optimization with real data, this leads to locally diverging gradients. Additionally, the large values of the surface gradients near specular spikes are of such strength that they dominate over the change in diffuse reflection and the optimization becomes instable.

As a solution, the maximum derivative of the specular component is reduced such that its influence on the optimization is the same as the influence of the diffuse component. In the following, the excessively high influence of specular intensities compared to diffuse intensities is explained by illustrating the maximum ratio between them based on the incidence angle ϑ_{in} and the "half angle" ϑ_h. This makes the obtained results more meaningful, but the actual implementation then explicitly uses pixelwise derivatives with respect to $p_{(u,v)}$ and $q_{(u,v)}$.

It is straightforward to show that the diffuse part derivative of the observed intensity using the physically plausible Lambert+Blinn model (Eq. 2.17) has a maximum of

$$\max \left[\frac{d}{d\vartheta_{in}} \left(\frac{I_0}{r_1^2} \frac{k_d}{\pi} \cos \vartheta_{in} \right) \right] = \frac{I_0}{r_1^2} \frac{k_d}{\pi}. \qquad (6.19)$$

For the specular component

$$I_s = \frac{I_0}{r_1^2} k_s \underbrace{\frac{(m+2)(m+4)}{8\pi(2^{(-m/2)} + m)}}_{1/N} \cos^m \vartheta_h \cos \vartheta_{in}, \qquad (6.20)$$

the derivative $dI_s/d\vartheta_h$ is computed under the assumption of a configuration where $d\cos\vartheta_{\text{in}}/d\vartheta_{\text{in}}$ exhibits its maximum of 1 and for a material with $m > 2$

$$\max\left(\frac{d}{d\vartheta_h}I_s\right) = \frac{I_0}{r_1^2}\,k_s\,\frac{1}{N}\,\max\left(m\cos^{(m-1)}\vartheta_h(-\sin\vartheta_h)\right) = \frac{I_0}{r_1^2}\,\frac{k_s}{N}\,\frac{m}{2}. \tag{6.21}$$

Therefore, the largest ratio S between the derivative of the diffuse component and the derivative of the specular component is

$$S = \frac{k_s\,\pi\,m}{2\,N\,k_d}. \tag{6.22}$$

For typical values ($k_d = 0.8$, $k_s = 0.2$, $m = 200$), that factor becomes $S \approx \frac{0.2\pi\cdot 200}{2\cdot(1/8.2)\cdot 0.8} \approx 644$, which is of such magnitude that it prevents the convergence and instead causes oscillations around a non-optimal configuration. Practically, $\frac{dR}{dp}$ and $\frac{dR}{dq}$ are computed analytically rather than numerically, which avoids numerical problems.

6.1.5 Treatment of interreflections

The following derivation shows that Horn's reconstruction approach [Horn and Brooks, 1986; Horn, 1990] used a special case of his actual problem formulation. Its generalization regarding interreflections is computationally very expensive and a more convenient approach for interreflection handling will be introduced. Horn did not extend Eqs. 6.4 and 6.5 to handle interreflections. For the *direct* reflection case and images of size $U \times V$, the value of R_p at (u, v) depends only on the respective gradient $p(u, v)$. R_p, R and I can thus be regarded as matrices of size $U \times V$.

For the general case, Eq. 6.4 can be restated such that $R, I, p, q \in \mathbb{R}^{U\times V}$ become row vectors of size $1 \times UV$ indexed by column-wise linear indices $k = V \cdot u + v$ and $n = V \cdot u + v$, and R_p becomes a matrix of size $UV \times UV$:

$$\tilde{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_k \end{bmatrix}, \qquad \tilde{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_k \end{bmatrix}, \qquad \tilde{R}_p = \begin{bmatrix} R_{p,1}(p_1) & R_{p,1}(p_2) & R_{p,1}(p_3) & \dots & R_{p,1}(p_n) \\ R_{p,2}(p_1) & R_{p,2}(p_2) & R_{p,2}(p_3) & \dots & R_{p,2}(p_n) \\ R_{p,3}(p_1) & R_{p,3}(p_2) & R_{p,3}(p_3) & \dots & R_{p,3}(p_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{p,k}(p_1) & R_{p,k}(p_2) & R_{p,k}(p_3) & \dots & R_{p,k}(p_n) \end{bmatrix}. \tag{6.23}$$

For the direct illumination case, each non-diagonal element $\tilde{R}_p(k, n)$ with $k \neq n$ is zero. If the surface is concave and thus displays interreflections, then $R_p(k, n)$ contains the influence of the n-th surface element with gradient p_n on the k-th surface element.

While the above solution is mathematically correct, it is computationally highly expensive and an approximation that is computed more rapidly is proposed here: Instead of the computation of all U^2V^2 elements of $\tilde{R}_p$, the interreflections are removed from each of the M camera images I_m with $m = 1...M$ and then proceed with the reconstruction using the direct illumination assumption:

$$I_m^{\text{comp}} = I_m - \underbrace{(I_m^G - I_m^D)}_{\text{interreflections}}. \tag{6.24}$$

The interreflections $(I_m^G - I_m^D)$ are determined using images rendered with direct illumination I_m^D and global illumination I_m^G, whose difference contains only the interreflections. Both, I_m^D and I_m^G, are computed by rendering the scene with the PBRT framework described in Section 1.5 using the (x, y, z) surface and BRDF data available at the current iteration of the algorithm. Applying an iterative scheme then allows a successive refinement of the interreflection-affected regions.
This concept is used for interreflection handling in a pure PS framework (see Section 6.3) and in a depth/-gradients fusion framework (see Section 6.4).

6.2 Simultaneous 3D surface and BRDF estimation for convex non-Lambertian surfaces

For image-based 3D surface reconstructions it is necessary to know the BRDF parameters prior to the reconstruction, since they are required for computing the reflectance map (see Section 6.1.1). This poses a "chicken-and-egg" problem, since the surface reconstruction requires a known BRDF, while simultaneously the BRDF estimation requires a known surface. Another challenging point is the fact that the estimation of BRDFs based on arbitrary shapes is subject to a trade-off between precision and generality [Weyrich et al., 2008]. This means that even a predetermined surface shape inevitably bears some inaccuracies due to e.g. measurement errors, and the overall BRDF fitting accuracy becomes reduced. This is especially the case for objects with "bumpy" surfaces, which exhibit rapidly changing specular intensities as shown in Fig. 6.1.

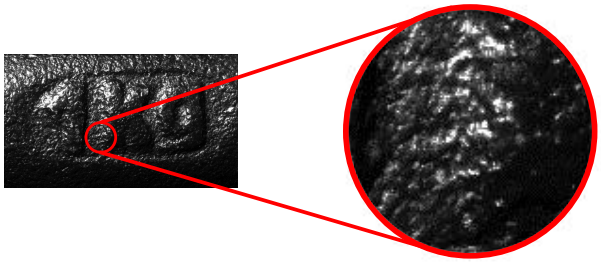


Figure 6.1.: Close-up zoom of a metallic surface. The small surface facets oriented in rapidly changing directions make correct BRDF estimation challenging, since the depth data is not known to the required precision due to measurement inaccuracies.

To overcome these problems, the availability of coarse but large-scale-robust depth data acquired with a laser scanner is exploited. This allows us to estimate the BRDF on low-pass filtered and subsampled depth and image data, which permits a reconstruction on that scale. This process can then be iterated on a refined scale as described previously. In this way, one can gradually refine the shape and BRDF accuracy, which then yields highly accurate 3D and BRDF data.
Further problems arise if the surface contains concavities, which introduce misleading intensity measurements into the data set since the fitted model only considers direct illumination.
For dealing with interreflections, it is either possible to render the surface using global illumination, or to exclude interreflection-affected measurements from the dataset. Since physically plausible rendering is usually time-consuming, the more practical approach of measurement removal is used. This is achieved based on the second derivative of low-pass filtered surface depth data. In particular, only measurements with second derivative values in the interval $[-1...1]$ are used. The set of optimal parameters (P^*) of a chosen BRDF model is then obtained by non-linear data fitting as explained in Section 2.5.

6.2.1 Algorithm

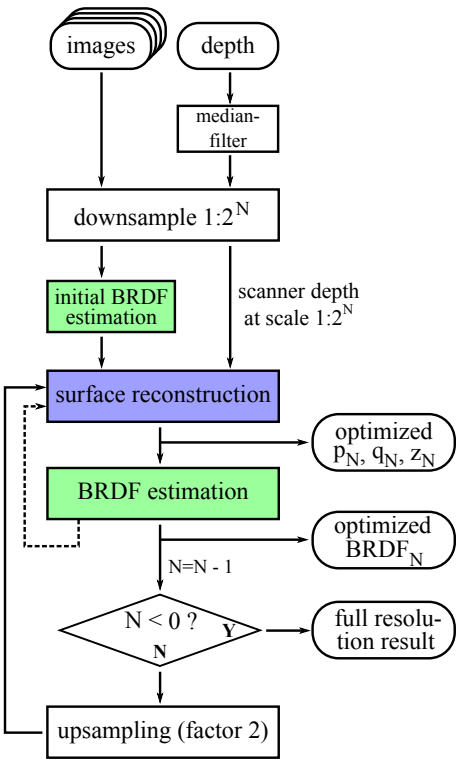


Figure 6.2.: Overview of the self-consistent algorithm for 3D surface reconstruction and reflectance parameter estimation.

Fig. 6.2 illustrates the proposed self-consistent 3D reconstruction algorithm. The main elements are the estimation of the reflectance parameters (green) and the surface reconstruction (blue). The algorithm starts with the given image and depth data, which are initially subsampled by a factor of 2^N with appropriate spatial low-pass filtering to avoid aliasing. The subsampling stage ensures the removal of spurious high spatial frequency components from the range scanner data and thus provides the basis for robust reflectance estimation on that scale. Once the initial reflectance parameters are known, the reconstruction exploits the reflectance information to incorporate the image-based depth data. The successive steps "surface reconstruction" and "BRDF estimation" can be iterated in an inner loop (cf. dashed line in Fig. 6.2) without increasing the current resolution scale. Afterwards, the result of the 3D reconstruction stage is upsampled by a factor of 2 using bicubic interpolation and then serves as the initialization for the next iteration. The algorithm terminates when the full resolution scale is reached.

For modeling the reflectance, a three-component Lambert+Phong+Phong model inspired by Nayar et al. [1990a]; Wöhler and d'Angelo [2009] has been applied with an appended directional diffuse term [Dorsey et al., 2008] according to

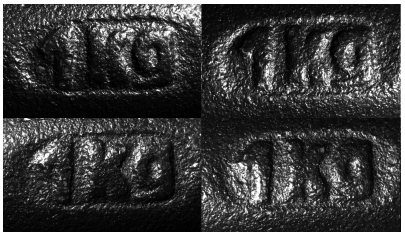
$$I = I_0\,\rho\,\Big[\underbrace{\langle \mathbf{n}\cdot\mathbf{l}\rangle}_{\text{diffuse}} + k_{ds}\ \underbrace{\langle \mathbf{n}\cdot\mathbf{l}\rangle^{m_{ds}}}_{\text{directional diffuse}} + k_l\ \underbrace{\langle \mathbf{v}\cdot\mathbf{r}\rangle^{m_l}}_{\text{specular lobe}} + k_s\ \underbrace{\langle \mathbf{v}\cdot\mathbf{r}\rangle^{m_s}}_{\text{specular spike}}\Big]. \tag{6.25}$$

Its parameters are the intensity I_0 of the incident light, the surface albedo ρ, the specular lobe strength k_l, the specular lobe width m_l, the specular spike strength k_s, and the specular spike width m_s [Wöhler and d'Angelo, 2009] with the directional diffuse width m_{ds} and strength k_{ds}. This model has proven to be flexible enough to represent the reflectance behavior, while having a feasible number of 7 parameters (the term $I_0\,\rho$ can be set to the "effective albedo" ρ_{eff}). The directional diffuse term has empirically proven to have a favorable effect on the 3D reconstruction accuracy when few light sources (i.e. images) are available.

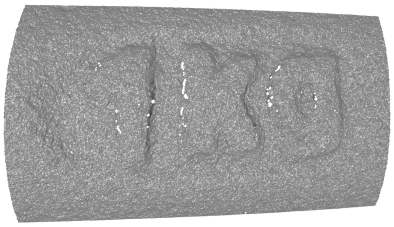
6.2.2 Results

The results for the 3D reconstruction of a mainly convex embossing on the surface of a metallic dumbbell consisting of dark cast iron are presented in Fig. 6.3. The object and the area of interest are depicted in Fig. 6.3(a), the input data for the algorithm are shown in Fig. 6.3(b) and 6.3(c). Four of the 12 images acquired under different illumination directions (cf. Fig. 6.3(b)) and the raw scanner data (cf. Fig. 6.3(c)) are shown as well. Figs. 6.3(d)–6.3(g) illustrate how the surface evolves over 4 iterations ($K = 4$) from a

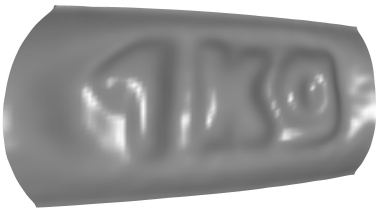
(a) Metallic dumbbell, the ROI for the reconstruction is indicated in red.



(b) 4 out of 12 HDR images, 400×700 pixels, lateral resolution 0.042 mm per pixel.

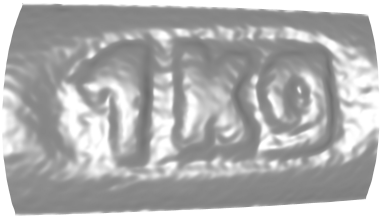


(c) Raw range scanner data.

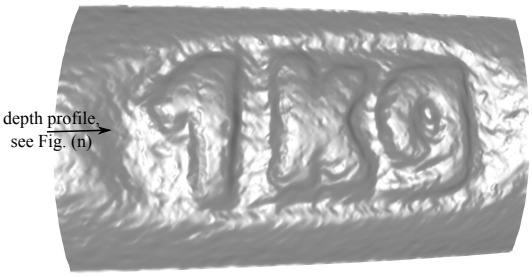


(d) Reconstruction result at scale $1/8$.

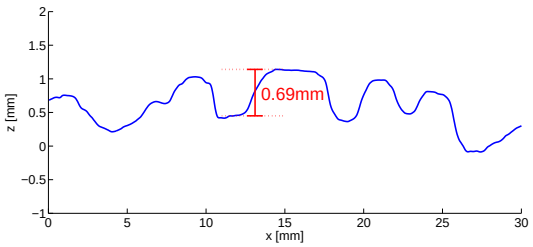
(e) reconstruction result at scale $1/4$.



(f) reconstruction result at scale $1/2$.



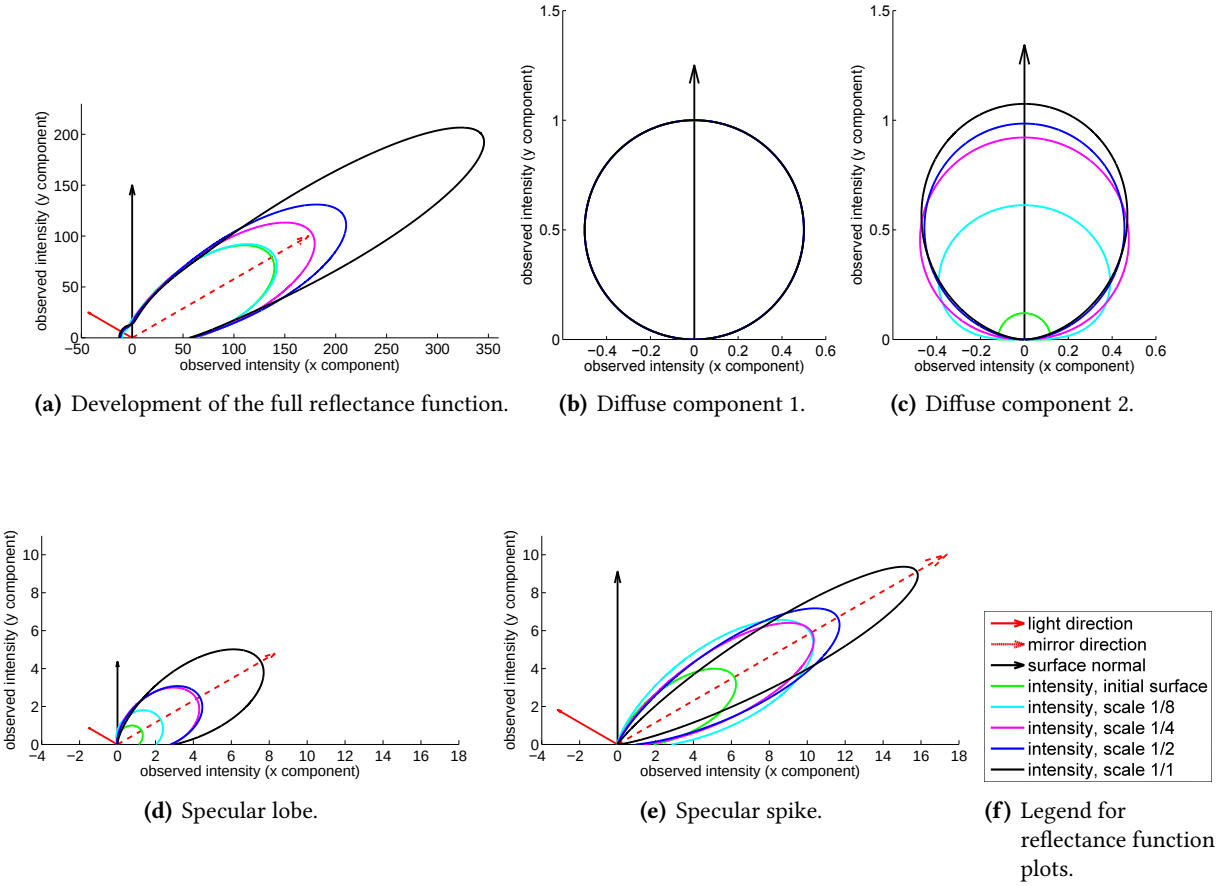
(g) reconstruction result at full scale (400×700 pixels.)



(h) Tactile depth difference measurement on a cross-sectional surface profile (cf. arrow in (g)).

Figure 6.3.: Experimental results for an embossing in a metallic dumbbell consisting of dark cast iron.

coarse surface at $1/8$ of the full scale (cf. Fig. 6.3(d)) over the intermediate scales $1/4$ (cf. Fig. 6.3(e)) and $1/2$ (cf. Fig. 6.3(f)) to the full scale where the resolution of the 3D surface reconstruction reaches the full resolution of the images (300×700 pixels at a scale of $42\ \mu$m per pixel) as shown in Fig. 6.3(g). Note that with each iteration, an increasing amount of surface detail becomes visible and is thus incorporated into the reconstructed surface. The comparison of the images and the surface shows correspondences between surface bumps and their bright or dark counterparts in the images. The same is true for the correspondences between images and reflectance maps in Fig. 6.5. An analysis of a cross-sectional profile of the surface with an indicated depth difference measurement is shown in Fig. 6.3(h), where the reconstructed surface has a depth difference of 0.69 mm, whereas a tactile reference measurement with a caliper gauge yields 0.67 ± 0.02 mm. Fig. 6.4(a)–Fig. 6.4(e) show the estimated reflectance functions for each resolution level. The full reflectance function (cf. Fig. 6.4(a)) is decomposed into its four components as shown in Fig. 6.4(b)–Fig. 6.4(e), where a normalization with respect to the effective albedo ρ_{eff} has been performed in order to demonstrate the development of the respective component without the influence of the albedo. Note that the diffuse components are plotted with the direction $\mathbf{l}$ of incident light varying over the upper hemisphere, while the specular components are plotted with the viewing direction $\mathbf{v}$ being varied while $\vartheta_{in} = 45°$. The numerical results for each component are listed in Table 6.1.



(a) Development of the full reflectance function. **(b)** Diffuse component 1. **(c)** Diffuse component 2.

(d) Specular lobe. **(e)** Specular spike. **(f)** Legend for reflectance function plots.

Figure 6.4.: In (b)–(e), the diffuse components are plotted with the direction $\mathbf{l}$ of incident light varying over the upper hemisphere, while the specular components are plotted with the viewing direction $\mathbf{v}$ being varied.

The plots show an increasing strength of the specular reflectance components while their widths decrease,

	raw weights				physically plausible weights					exponents		
scale	ρ_{eff}	σ_{ds}	σ_l	σ_s	ρ_{eff}	σ_{dd}	σ_{ds}	σ_l	σ_s	m_{ds}	m_l	m_s
initial	17.24	0.12	1.46	7.10	128.99	0.134	0.016	0.318	0.532	$1e^{-4}$	1.85	9.21
1/8	10.52	0.61	2.62	11.75	128.01	0.082	0.050	0.357	0.511	0.45	1.79	9.88
1/4	11.59	0.92	4.83	11.83	135.09	0.086	0.079	0.461	0.374	0.91	3.65	15.03
1/2	12.42	0.98	5.01	13.40	141.06	0.088	0.086	0.458	0.368	1.24	4.05	18.17
1/1	14.27	1.07	8.73	18.22	144.44	0.099	0.106	0.577	0.218	1.45	7.39	49.8

Table 6.1.: Determined reflectance function parameters for each reconstruction scale and the final result.

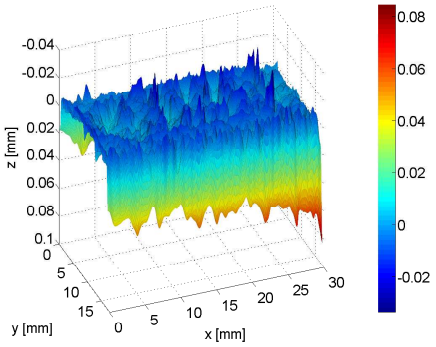
which causes the characteristic sharp and intense specular reflections on the surface apparent at full resolution. This behavior can also be observed in the resulting reflectance maps shown in Fig. 6.5. Both diffuse components are significantly lower in their intensities compared to the specular components, which is the typical behavior of metallic surfaces. In Fig. 6.5, the structural similarity (SSIM) measure known from the domain of video coding [Wang et al., 2004] is used to illustrate the similarity between the acquired images and the corresponding reflectance maps. The SSIM is a real number from the interval $[0...1] \subset \mathbb{R}$ and increases with increasing similarity.

(a) Scale $1/8$, SSIM = 0.85962 **(b)** Scale $1/4$, SSIM = 0.84261 **(c)** Scale $1/2$, SSIM = 0.88546 **(d)** Scale $1/1$, SSIM = 0.92896

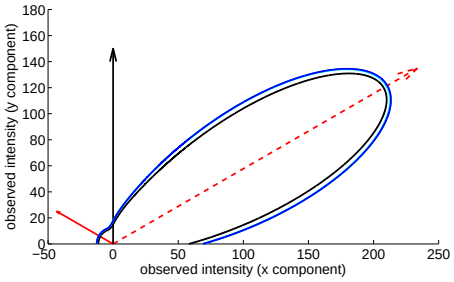
Figure 6.5.: Details regarding the final reflectance maps for different scales under illumination from one selected LED. For each image, the upper half shows the image data and the lower half the reflectance map.

6.2.3 Synthetic evaluation

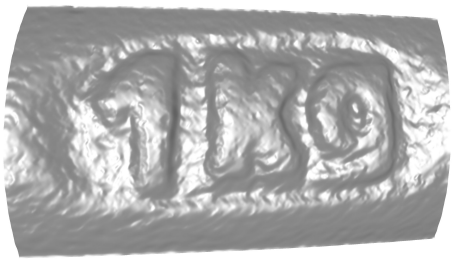
Since there is no ground truth with the required accuracy available, it is only possible to evaluate the accuracy of the approach using synthetically generated data. For this purpose, the result of the described surface reconstruction algorithm is used as the ground truth surface and the obtained reflectance maps as the corresponding synthetic images. The algorithm is then initialized with that data, and Gaussian noise with a standard deviation of $60\,\mu$m is added to imitate the inaccuracies of the range scanner. The obtained results are shown in Fig. 6.6. Note that the RMSE of the reconstructed surface with respect to the synthetic ground truth only amounts to $10.2\,\mu$m, which corresponds to approximately $1/4$ of the lateral pixel extent of $42\,\mu$m, where the highest deviations occur near the margin of the reconstructed surface section. The reflectance function estimated based on the synthetic data set is shown in Fig. 6.6(b). The estimated reflectance function resembles the ground truth closely to within a few percent. Fig. 6.6(b) shows the results for different numbers of subiterations (cf. Fig. 6.2), which have a very small effect on the inferred shape of the reflectance function. The surface is reconstructed at a high accuracy (cf. Fig. 6.6), and the rendered reflectance maps closely resemble the synthetic images used for 3D reconstruction as the SSIM corresponds to a very high value of 0.985 on the full resolution scale (cf. Fig. 6.6(d)).



(a) Deviations of the reconstructed surface from the synthetic ground truth.



(b) Full reflectance function determined using synthetic data.



(c) Reconstruction result using synthetic data for validation.

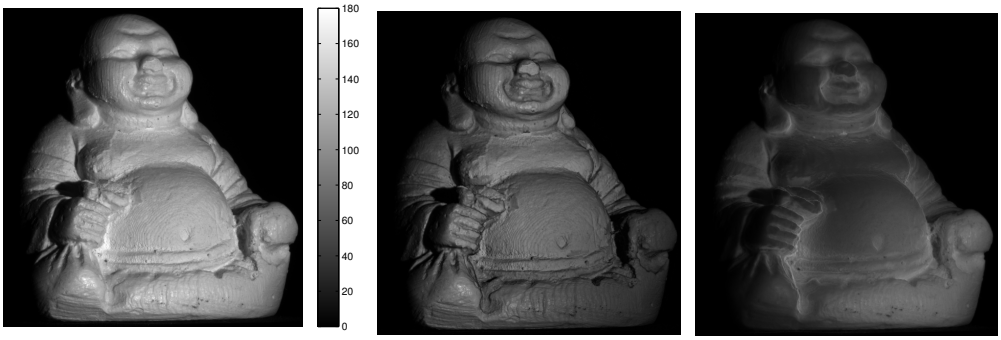
(d) Image and reflectance map.

Figure 6.6.: Validation of the proposed algorithm based on synthetic ground truth data. (a) The RMSE amounts to $10.2\,\mu$m. (b) Black: ground truth; green: 1 sub-iteration; cyan: 2 sub-iterations; blue: 4 sub-iterations. (d) Full-scale synthetic image (top) and corresponding reflectance map (bottom) of the reconstructed surface. The SSIM amounts to 0.985, thus indicating a very high similarity.

6.2.4 Conclusion

Is has been shown that the algorithm described above determines shape *and* BRDF with very high accuracy. This result is based on a qualitative evaluation and a quantitative error assessment using real and synthetic depth and image data. What remains an issue is the treatment of interreflections that occur at surface concavities. The following sections deal with a solution for that problem using photometric stereo alone (Section 6.3) and an algorithm that fuses absolute depth and image data that possibly contains interreflections (Section 6.4).

6.3 Photometric Stereo for interreflection-affected surfaces



(a) camera image **(b)** direct component **(c)** global component without direct component

Figure 6.7.: Experimentally measured illumination components using the method by Seitz et al. [2005] for determination of the direct component (b) and a variant of that method (inverse illumination pattern) for the global component (c). All images have the same grayvalue scaling.

The method introduced in the following accounts for interreflections when reconstructing surface concavities in a fully calibrated perspective photometric stereo environment. The approach as such imposes no restrictions upon the surface's reflective behavior and thus allows arbitrary and possibly even measured BRDFs.

In contrast to related publications, point light sources are used, which make illumination with specialized projectors as applied by Gu et al. [2011] unnecessary; furthermore, there is no shape restriction like the v-shapes examined by Nayar et al. [1990b] but it operates similarly in an iterative scheme. In addition, the light sources and camera are not assumed to be distant, which is not a prerequisite for the algorithm to work but it improves the reconstruction accuracy as examined by e.g. [Prados and Faugeras, 2003; Tankus et al., 2005]. Calibration of the (point) light source positions and intensities is achieved using the method by Lenoch et al. [2012] as described in Chapter 5.

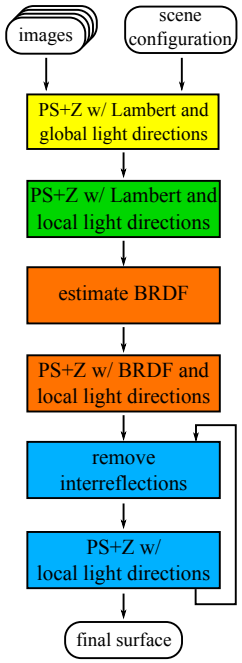


Figure 6.8.: Algorithm overview. "PS+Z" denotes photometric normal estimation and followed by depth integration. Note that each group of components (green, orange, blue) can possibly be iterated.

The key to the consideration of interreflections lies in an iterative scheme (see Fig. 6.8): (yellow) The method initially reconstructs the surface based on distant light sources and Lambertian reflectance. (green) That result becomes refined assuming non-distant light sources. The absolute surface depth data z is computed in each step using Agrawal's M-estimator approach [Agrawal et al., 2006]. The concatenation of PS and Agrawal's method is denoted "PS+Z". The resulting surface is afterwards used (orange) to obtain the parameters of a parametric BRDF, which additionally refines the surface in another reconstruction step. Lastly (blue), the consideration of interreflections is incorporated by determination of the amount of global illumination using a

physically plausible rendering system (PBRT, see Section 1.5). This allows removing a (significant) amount of interreflection-based radiance from the captured scene image data and thus to obtain a refined reconstruction result. Note that it may be beneficial to repeat some of the stages described above in order to ensure their convergence.
The following section describes algorithmic details for the respective stages. The potential of the method is then evaluated on real data for a Lambertian surface (Section 6.3.2). The validation of the correctness of these results is achieved on the one hand by qualitative assessment and on the other hand by quantitative comparison with the surface data measured by an industry-standard laser scanner.

6.3.1 Algorithm

In stage 1 (yellow), the surface normals are computed for each pixel (u, v) based on the standard PS equation for Lambertian surfaces,

$$\mathbf{I}_{obs}(u,v) = I_0 \, \rho(u,v) \, \mathbf{L} \, \mathbf{n}(u,v). \tag{6.26}$$

Note that this has been first described under the application of three light sources by Woodham [1980]. In Eq. 6.26, u and v denote the pixel indices, $\mathbf{I}_{obs}(u,v) \in \mathbb{R}^{k\times 1}$ denotes the vector of k observed intensity samples under $i = 1...k$ varying (distant) normalized illumination directions $\mathbf{l}_i$ with $||\mathbf{l}_i||_2 = 1$. The light source relative radiance is referred to as I_0 and the light source directions are aligned row-wise in $\mathbf{L} = [l_{1x}, l_{1y}, l_{1z}; l_{2x}, l_{2y}, l_{2z}; ...] \in \mathbb{R}^{3\times k}$. The surface normals are $\mathbf{n} = [n_x, n_y, n_z]^T$ with $||\mathbf{n}||_2 = 1$ and ρ denotes the surface albedo. The surface normal is then determined in the least-mean-square (LMS) sense using

$$\mathbf{n}(u,v) = \frac{1}{I_0 \, \rho(u,v)} \, \mathbf{L}^{-1} \, \mathbf{I}_{obs}(u,v) \tag{6.27}$$

with a pseudo-inverse $\mathbf{L}^{-1}$. Note, that locally varying albedos can be obtained when solving for $\mathbf{n}$ and ρ with known irradiance I_0:

$$\mathbf{n}^*(u,v) = \mathbf{L}^{-1} \, \mathbf{I}_{obs}(u,v), \quad \mathbf{n} = \frac{\mathbf{n}^*}{||\mathbf{n}^*||_2}, \quad \rho(u,v) = \frac{||\mathbf{n}^*||_2}{I_0} \tag{6.28}$$

The obtained surface normal field is then integrated to absolute depth values using the M-estimator approach described by Agrawal et al. [2006]. For the transformation into a metric depth space, the algorithm is given the approximate object distance from the camera, which suffices for the determination of pixel extents and thus metric depth scaling.
In stage 2 (green), the light sources are regarded as non-distant. This alters the image intensity equation such that

$$\mathbf{I}_{obs}(u,v) = \frac{I_0 \, \rho}{r^2(u,v)} \, \mathbf{L}(u,v) \, \mathbf{n}. \tag{6.29}$$

Note that this now assumes locally varying illumination directions $\mathbf{l}_i(u,v)$ for each pixel (u,v) and each light source $i = 1...k$, which is due to the scene being illuminated by non-distant point light sources

(*perspective* PS). Additionally, each pixel is assigned the light travel distance from the light source to the object (r). Even on this small-scaled framework (object height $\approx 50\,mm$, $\bar{r} \approx 300\,mm$ and an object-to-sensor-distance of $\approx 310\,mm$), this shows a significant effect. Up to this point, the surface has been assumed to be perfectly Lambertian and free of global illumination.

In stage 3 (orange), the BRDF of the underlying material is estimated globally based on the extracted surface from stage 2 and the obtained images. This is achieved by minimizing the RMSE of the intensity data $I = I(u, v)$ and the rendered intensities $R = R(P, u, v)$ of the surface with respect to the parameter set P of the chosen BRDF as described by Eq. 2.35 in Section 2.5. For Lambertian surfaces, the only parameter is the surface albedo ρ, which can be computed by determination of I_0 using Lenoch et al. [2012].

In stage 4 (blue), the interreflection component is (iteratively) compensated before starting the PS+Z algorithm: Two separate images of the last stage's PS+Z surface result are rendered under (1) direct $R_{d,i}$ and (2) global $R_{g,i}$ illumination for each light source $i = 1...k$. Afterwards, the global component is determined as the difference between $R_{g,i}$ and $R_{d,i}$, which is used to compensate these in the real camera images I_i:

$$I_{\text{new},i} = I_i - \underbrace{(R_{g,i} - R_{d,i})}_{\text{interreflections}} \qquad i = 1...k. \tag{6.30}$$

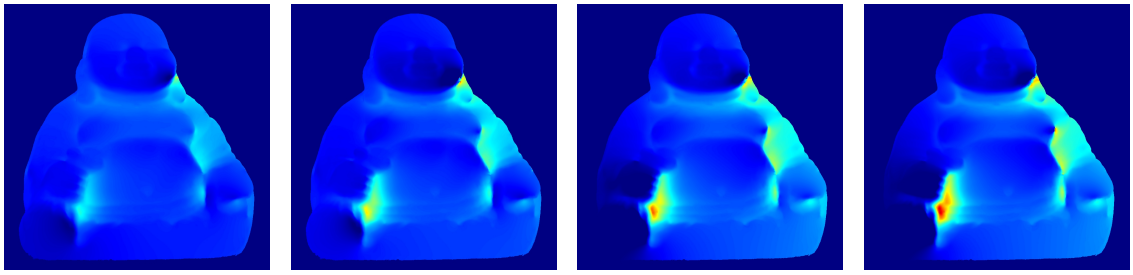
As a last step for this stage, the PS+Z algorithm is executed on the new image data basis $I_{\text{new},i}$.

6.3.2 Results

The description and discussion of the experiments is structured as follows: Initially, the progressing change in PS+Z depth with increasing algorithm iterations is analyzed (Section 6.3.2.1). Afterwards, Section 6.3.2.2 shows the beneficial effect of the approach qualitatively based on an analysis of the reflectance maps with global illumination.

Section 6.3.2.3 then describes the comparison of the PS+Z depth result with independently measured scanner depth data z_{RS} for a region of the surface that is especially affected by interreflections. To conclude the evaluation, the aspect of algorithm convergence is discussed in Section 6.3.2.4 along with suitable criteria that terminate the execution at a reasonable state.

6.3.2.1 Progressing change in reconstructed depth



(a) $z_{PS,2} - z_{PS,1}$ **(b)** $z_{PS,3} - z_{PS,1}$ **(c)** $z_{PS,4} - z_{PS,1}$ **(d)** $z_{PS,5} - z_{PS,1}$

Figure 6.9.: Difference between the naive PS depth $z_{PS,1}$ (no interreflection compensation) and the iterations using image data with compensated interreflections $z_{PS,2...5}$. Scale: $-0.5\,mm$ (dark blue) to $2.5\,mm$ (red)

Fig. 6.9 shows the differences between the naive PS+Z depth $z_{PS,1}$ (no interreflection compensation) and its iterations $z_{PS,2...5}$ using image data with compensated interreflections. It becomes clear that the compensation of the global (interreflection) component produces deeper concavities compared to the uncompensated case. The results show that less deep concavities on the one hand (e.g. neck and ear region) exhibit only a change in depth over the early iterations of the algorithm, but remain static on the late iterations. Deep concavities on the other hand progress in their depth over several iterations. This already gives a hint that the algorithm converges suitably, since regions with shallow and deeper concavities converge independently.

6.3.2.2 Qualitative evaluation

Fig. 6.10 shows reflectance maps from different stages of the algorithm. Note that at iteration 5 (Fig. 6.10(c)), the object surface appears more similar to the physical image (Fig. 6.10(a)), especially in the interreflection-affected regions. For a human observer, it is obvious that the underlying surface resembles the one of the

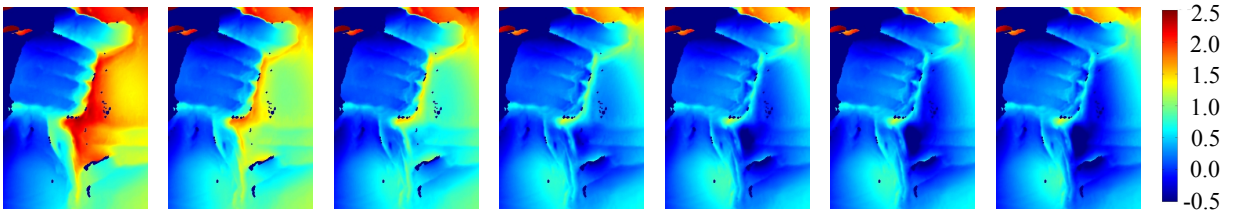
(a) image **(b)** stage 4, iteration 1 rendering result (global illumination) **(c)** stage 4, iteration 5 rendering result (global illumination)

Figure 6.10.: Rendered PS+Z results. Note the appearance difference at surface concavities.

physical image more closely due to the perceived intensities and due to the plausibility of where and how shadows are cast. In summary, the qualitative assessment result of the surface is very favorable, due to fine surface details being visible and surface concavities being shaped as expected from the real world object image.

6.3.2.3 Quantitative evaluation

Fig. 6.11 shows a comparison of the reconstructed depth $z_{\text{PS,n}}$ and quantitatively measured depth data obtained using a laser range scanner z_{RS} for the iterations $n = 1...8$. The registration of both depth profiles has been achieved by aligning the absolute depth of a convex surface area in the "hand"-region. This ensures that the PS+Z depth $z_{\text{PS,n}}$ is unaffected by interreflections in that area, and thus allows unbiased registration. With naive photometric reconstruction (no compensation of interreflections, $z_{PS,1}$, Fig. 6.11(a)), there are very strong differences visible between the reconstructed surface $z_{PS,1}$ and the ground truth scanner depth z_{RS}, which are in the range of some mm. The differences initially reach more than $2\,mm$

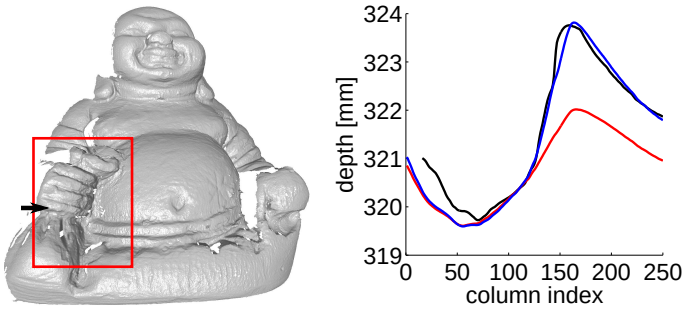


(a) $z_{\text{RS}} - z_{PS,1}$ **(b)** $z_{\text{RS}} - z_{PS,2}$ **(c)** $z_{\text{RS}} - z_{PS,3}$ **(d)** $z_{\text{RS}} - z_{PS,4}$ **(e)** $z_{\text{RS}} - z_{PS,5}$ **(f)** $z_{\text{RS}} - z_{PS,6}$ **(g)** $z_{\text{RS}} - z_{PS,7}$ **(h)** scale (mm)

Figure 6.11.: Ground truth validation. Scale: $-0.5\,mm$ (dark blue) to $2.5\,mm$ (red)

(Fig. 6.11(a)), which is unacceptable in terms of measurement accuracy for e.g. industrial applications. In the course of the iterations, the differences decrease quickly and already after the second iteration with interreflection-compensated image data ($z_{PS,3}$, Fig. 6.11(c)), the error has decreased over a wide range of the concavity.

The depth profile plot in Fig. 6.12 indicates the same improvements and shows that the final surface aligns with few but small differences with the scanner depth profile. A close inspection reveals regions where the depth error increases slightly, e.g. between the fingers. For an explanation, three main aspects seem to be important: First, there are possibly misalignments between the measured image data and the rendered interreflections using the PBRT framework. These are caused by PBRT's simulated camera, which inevitably bears some differences to the actual camera. Some part of the misalignments has already been compensated using the normalized cross correlation maximum of both images to detect a global offset, but further un-modeled distortions cause small additional misalignments, which then result in misplaced interreflection intensities. Second, the albedo value may be slightly too high, which causes interreflections that are too bright and



(a) high quality scan, region of interest **(b)** profile along row 170 (black arrow in (a))

Figure 6.12.: The plot (b) shows a depth profile along a row of the depth profile (a) with scanner depth (black), initial PS+Z depth (red) and final PS+Z depth (blue).

then cause concavities that are too deep. Third, the scanner depth data actually contains some inaccuracies, which make the evaluation infeasible for small details. Overall, the error has been reduced to a large extent.

6.3.2.4 Convergence criterion

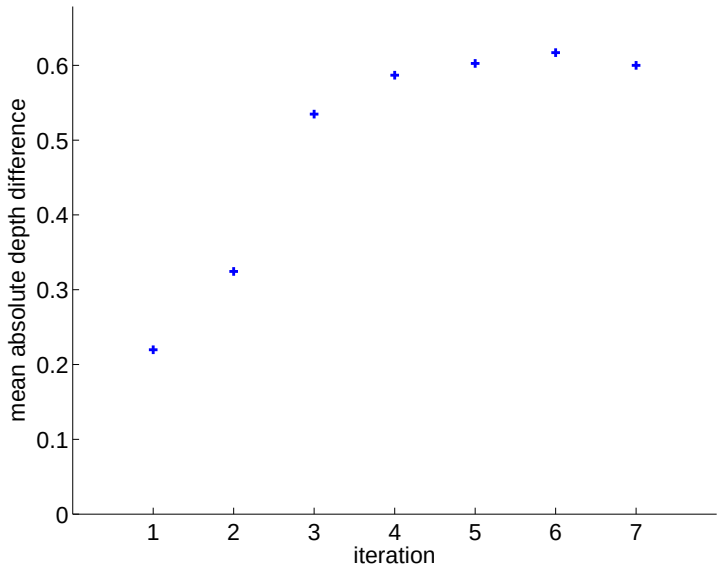


Figure 6.13.: Average change in depth $\Delta z_n = \text{mean}(|z_{PS,n} - z_{PS,1}|)$ per iteration in the “hand"-section

Although it seems to be of less importance when to stop iterating due to e.g. shallow concavities remaining unchanged while deeper concavities still deepen (see Section 6.3.2.1), it makes sense to include another criterion for robustness and for saving computational time. It has been found particularly useful to examine the mean absolute change in depth of the photometrically reconstructed depth data of iteration n (i.e. $z_{PS,n}$) compared to the initial PS+Z result without compensation of interreflections (i.e. $z_{PS,1}$) such that $\Delta z_n = \text{mean}(|z_{PS,n} - z_{PS,1}|)$. Once that change is lower than a certain threshold (here: $\Delta z_n < 0.02\,\text{mm}$), the computation is regarded as complete. In this case, convergence is then reached after the 4^{th} iteration.

6.3.3 Conclusion

The presented approach for dealing with interreflections has been found to be very beneficial for PS-based shape reconstruction accuracy. The algorithm iteratively computes the global illumination at surface concavities and then removes these from the physical image data before using that improved data for PS and absolute depth reconstruction.

The evaluation showed a qualitatively and quantitatively beneficial effect for the extracted surface. The results show furthermore, that convergence is reached already after ≈ 4 iterations (for the given surface). Additionally, the presented criterion (mean absolute depth change) is found to be suitable to determine when the computation has finished.

While this approach improves the reconstruction accuracy significantly, there are still large-scale shape deviations contained in the obtained surfaces. The following sections thus introduce an approach for handling that problem by inclusion of absolute depth data.

6.4 High accuracy surface scan enhancement

In comparison to both aforementioned algorithms, the following regards several extensions and a more unified description: Section 6.2 showed the beneficial effect of iterative BRDF estimation, but it has later been found that it experiences convergence problems for surfaces with narrow specular spikes (plastics, lacquered surfaces). A solution provides the specular spike attenuation (see Section 6.1.4), which improves the optimization scheme and increases its robustness. It is then shown how to generally account for interreflections in the same framework, which is currently used for Lambertian surfaces and PS only (see Section 6.3). With those extensions it is possible to handle a new class of surfaces, which are lacquered and plastic surfaces with possible occurrence of interreflections.
Additionally, the sections below present a more detailed description of the algorithm, show its applicability to a wide range of surface types, and evaluate the accuracy with an improved approach using caliper objects with known shape.

6.4.1 Algorithm

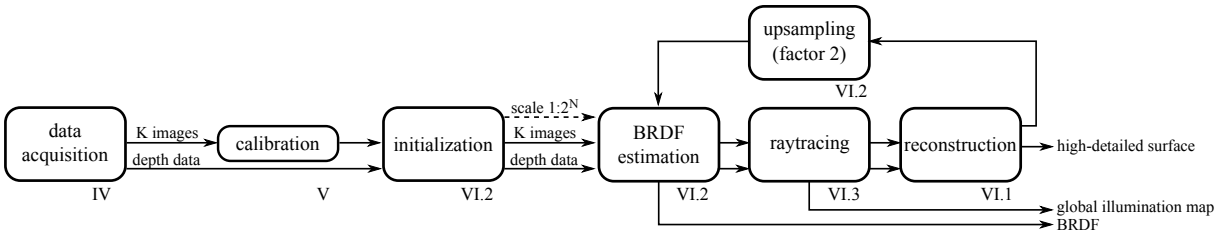


Figure 6.14.: Algorithm overview. The key idea lies in an iterative resolution refinement, which allows meaningful BRDF estimation. Each iteration then comprises BRDF estimation, compensation of interreflections and the actual reconstruction. Roman numbers indicate the chapters in which the respective components are explained in greater detail.

An overview of the proposed algorithm is provided by Fig. 6.14. Initially, the image and depth data are acquired. For the following reconstruction stages it is important to obtain pixel-synchronous depth and image data. Each image is acquired preferably with a different illumination direction, since this provides a robust data basis for the surface gradient determination. The images are then calibrated radiometrically using the technique by Lenoch et al. [2012], which provides the algorithm with information about light source strengths and metric light source positions.
The algorithm begins with an initialization stage that determines the starting values for the following stages. Mainly, this concerns the initial scale and initial guesses for depth profile and shadow maps based on the provided depth and image data. Afterwards, an iterative scheme based on [Herbort and Wöhler, 2012] determines the parameters of the underlying BRDF of the material, computes global illumination components similar to Herbort et al. [2013b] using a raytracing algorithm, and finally reconstructs a refined surface on that scale by fusing the provided depth data with the gradient information obtained from the interreflection-compensated image data. For the next iteration, the scale is doubled by upsampling the refined depth profile using bicubic interpolation to account for missing measurements. This scheme iterates until the full resolution is reached. For each stage, the algorithm provides increasingly refined global illumination maps and refined BRDF parameters. Note that the mathematic principles for each stage are explained thoroughly in Section 6.1.

6.4.2 Experimental data

(a) pumice stone **(b)** orange figure **(c)** metallic star **(d)** dumbbell

(e) Buddha WD **(f)** Buddha WS **(g)** Buddha GS

Figure 6.15.: Available objects for image and depth data acquisition. Note the different surface types, which are chosen to cover a wide variety of reflectance behaviors.

The datasets are chosen in a way that a large variety of reflectance behaviors will be examined. These are diffuse, specular, and metallic reflectance as well as the occurrence of interreflections. Fig. 6.15 shows color images of the objects used for data acquisition. Table 6.2 then gives an overview of the different reflectance behaviors contained in the respective data.

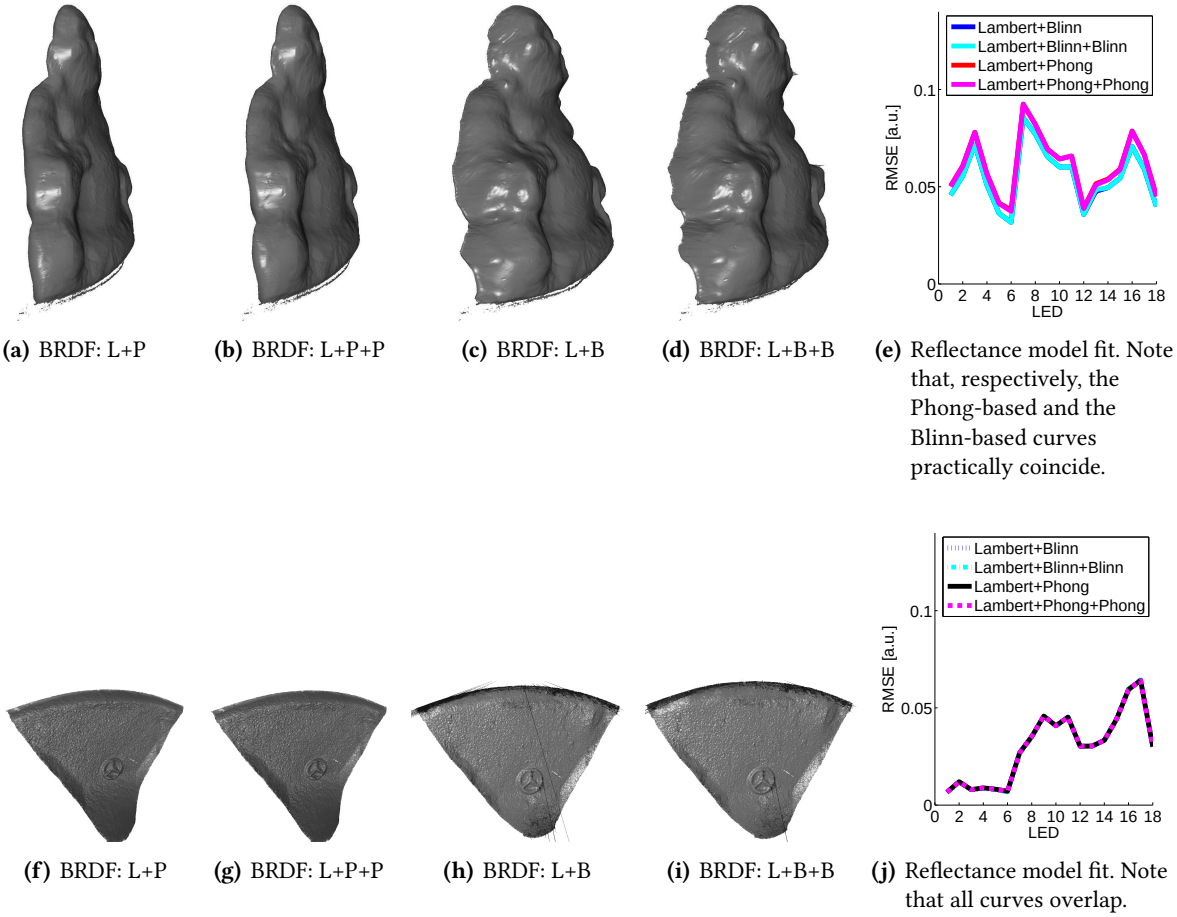
Object	surface material	reflection properties	inter-reflections	size $(u \times v)$ [px]	mean lateral pixel extents e_x, e_y [mm/px]
pumice stone	volcanic ash	diffuse	few	1800×1400	$0.041,\ 0.041$
orange figure	plastic	specular	few	600×900	$0.040,\ 0.040$
metallic star	cast iron	metallic	few	1400×1100	$0.041,\ 0.041$
dumbbell	cast iron	metallic	few	700×400	$0.040,\ 0.040$
Buddha WD	plaster	diffuse	yes	1000×1000	$0.040,\ 0.040$
Buddha WS	paint	specular	yes	1100×1150	$0.040,\ 0.039$
Buddha GS	paint	specular	yes	1100×1400	$0.040,\ 0.040$

Table 6.2.: Overview of the properties of the examined objects. Note the large variety, which is used to demonstrate the wide applicability of the presented method.

6.4.3 Results

The following sections present the conducted experiments. After an initial examination of the reflectance model choice in Section 6.4.3.1, the results of an accuracy evaluation based on caliper objects with known shape are illustrated in Section 6.4.3.2. Afterwards, it is shown in Section 6.4.3.3 that the algorithm can be applied to a wide range of surface types for several objects with different reflectance behaviors. The algorithm has already been evaluated based on synthetic input data as presented in Section 6.2 and the beneficial effect of interreflection compensation has been addressed in Section 6.3.

6.4.3.1 Choice of the reflectance model



(a) BRDF: L+P **(b)** BRDF: L+P+P **(c)** BRDF: L+B **(d)** BRDF: L+B+B **(e)** Reflectance model fit. Note that, respectively, the Phong-based and the Blinn-based curves practically coincide.

(f) BRDF: L+P **(g)** BRDF: L+P+P **(h)** BRDF: L+B **(i)** BRDF: L+B+B **(j)** Reflectance model fit. Note that all curves overlap.

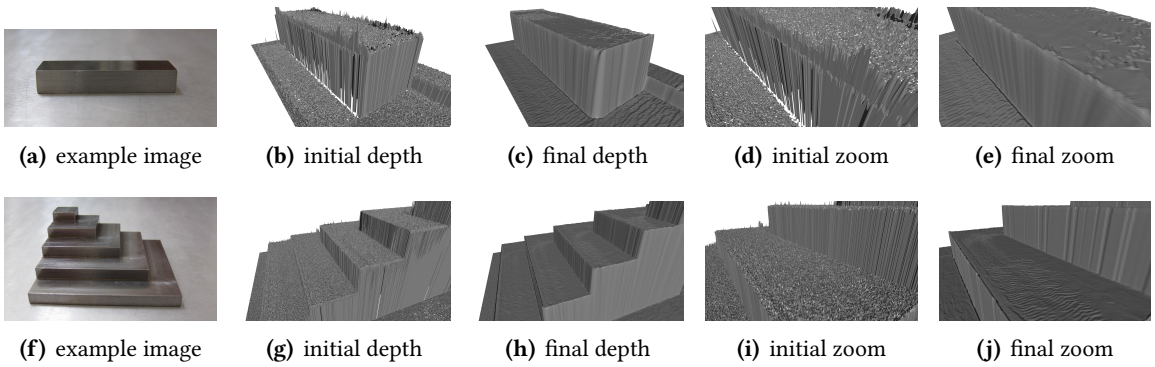
Figure 6.16.: PS+Z results for a specular surface (see (a)–(d)) and a metallic surface (see (f)–(i)) using different BRDF models. The abbreviations L+P, L+P+P, L+B, L+B+B indicate (L)ambert, (P)hong, and (B)linn with one or two specular components. The results show an advantage of Blinn over Phong, especially for specular surfaces. The application of two specular components improves the reconstruction accuracy for metallic surfaces. The RMSE values of the measured image intensity values compared to the modeled intensities in plot (e) and (j) indicate the advantage of Blinn over Phong for lacquered surfaces, but predict a similar performance for metallic surfaces, which is not the case given the PS+Z results.

Fig. 6.16 shows that lacquered surfaces are reconstructed with higher accuracy using Lambert+Blinn (L+B) rather than Lambert+Phong (L+P), since the L+B depth profile does not deviate from the actual object shape as much as the L+P equivalent. Yet, the application of more than one specular Blinn/Phong component

(L+B+B, L+P+P) has almost no effect. The examinations thus use a single Blinn lobe in addition to a Lambertian component for lacquered and plastic surfaces. The PS+Z reconstruction results for metallic surfaces are similar for the two-lobe models L+B+B and L+P+P with a slight advantage for L+B+B. Nevertheless, the L+P+P model is used for metallic surfaces to demonstrate the applicability of various reflectance models.

6.4.3.2 Caliper object examination

Due to the complex shape of the objects presented in Section 6.4.2, there are no depth data available with an accuracy that suffices for a meaningful absolute accuracy evaluation. Instead, two caliper objects are applied [Gerken, 2012], which have been manufactured with a CNC milling machine that provides knowledge of the shape at micrometer accuracy. Fig. 6.17 shows images and rendered 3D surfaces for both objects before



(a) example image **(b)** initial depth **(c)** final depth **(d)** initial zoom **(e)** final zoom

(f) example image **(g)** initial depth **(h)** final depth **(i)** initial zoom **(j)** final zoom

Figure 6.17.: Experimental results for the CNC manufactured caliper objects shown in (a) and (f). The raw scanned surfaces (b), (d), (g), and (i) show typical high frequency noise and missing measurements. The results shown in (c), (e), (h), and (j) illustrate outlier suppression and noise reduction. The caliper objects allow a quantitative accuracy analysis due to their known shape. Note that the rills visible in (h) and (j) result from the CNC milling cutter head and are actually contained in the surface.

and after processing with the presented algorithm. Fig. 6.18 presents the results of the evaluation. While the standard deviation as a measurement for the noise decreases by a factor of about 3, there is nearly no change in the large scale differences from the actual shape. This demonstrates that the algorithm refines the surface details while leaving the overall shape intact. Note that the rills visible in Fig. 6.17(h) and Fig. 6.17(j) result from the CNC milling's cutter head and are actually contained in the surface but cannot be evaluated regarding their accuracy, since depth data with that resolution are not available. A close inspection of the processed caliper pyramid identifies rounded corners of the pyramid levels, which possibly originate from the enforcement of integrability (see Eq. 6.3) as well as from the low-pass effect induced by numerically solving the Poisson equation (see Eq. 6.6) and using finite differences for the differential equation.

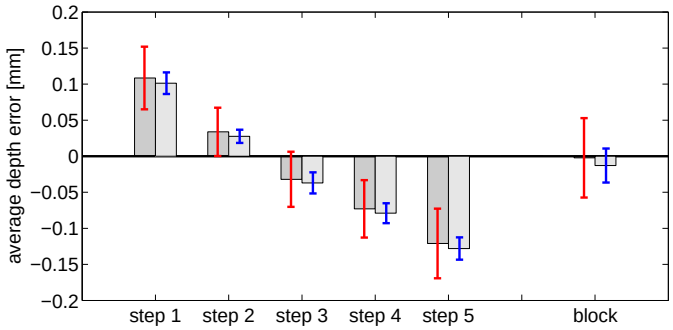
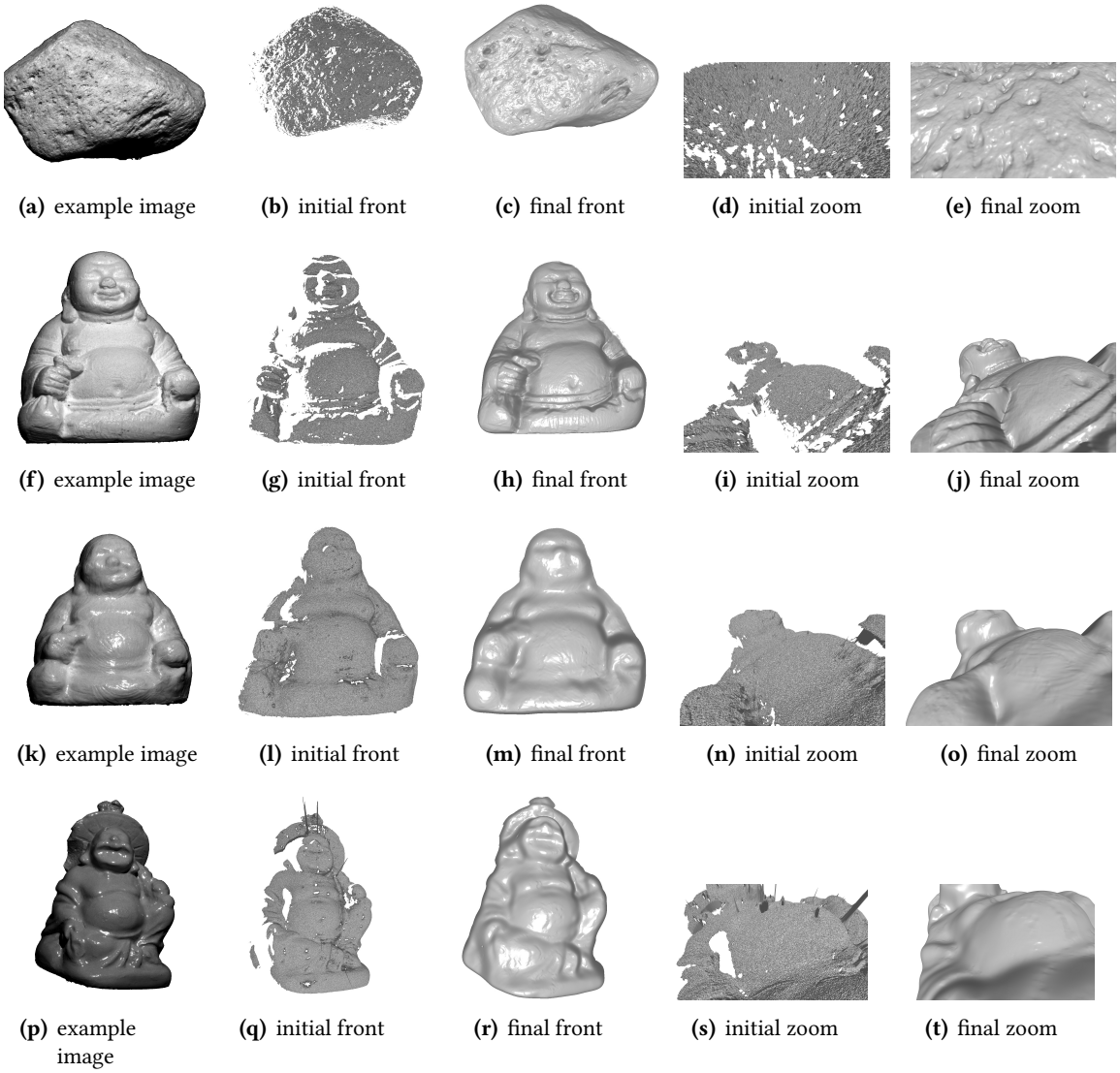


Figure 6.18.: Mean height differences between the optimized depth in Fig. 6.17(c), Fig. 6.17(h) and the known step height values (4, 8, 12, 16, 20; 5 mm) of the caliper pyramid and the caliper block. Dark gray columns: Average depth error as measured by the laser scanner. Light gray columns: Average depth error after application of the proposed algorithm. Red and blue: standard deviation of the data used for average depth error computation. While systematic deviations remain almost unchanged (almost constant mean height deviations), there is a significant improvement in the small scale details (decreasing standard deviation).

6.4.3.3 Application to various surface types

The following demonstrates the reconstruction performance qualitatively and states significant observations regarding the obtained surfaces. Figs. 6.19 and 6.20 show the reconstruction results obtained with the proposed algorithm. For each of the eight objects, an exemplary image along with raw and refined depth data viewed from the front and zoomed to show fine details are presented. Each result (column 3 and 5, respectively) shows significant improvement of its depth profile when compared to the raw laser scanner data: While the raw scan exhibits noise, outliers and missing measurements, this has all been reduced in the result and fine surface details become visible while the overall large scale shape is preserved.



(a) example image **(b)** initial front **(c)** final front **(d)** initial zoom **(e)** final zoom

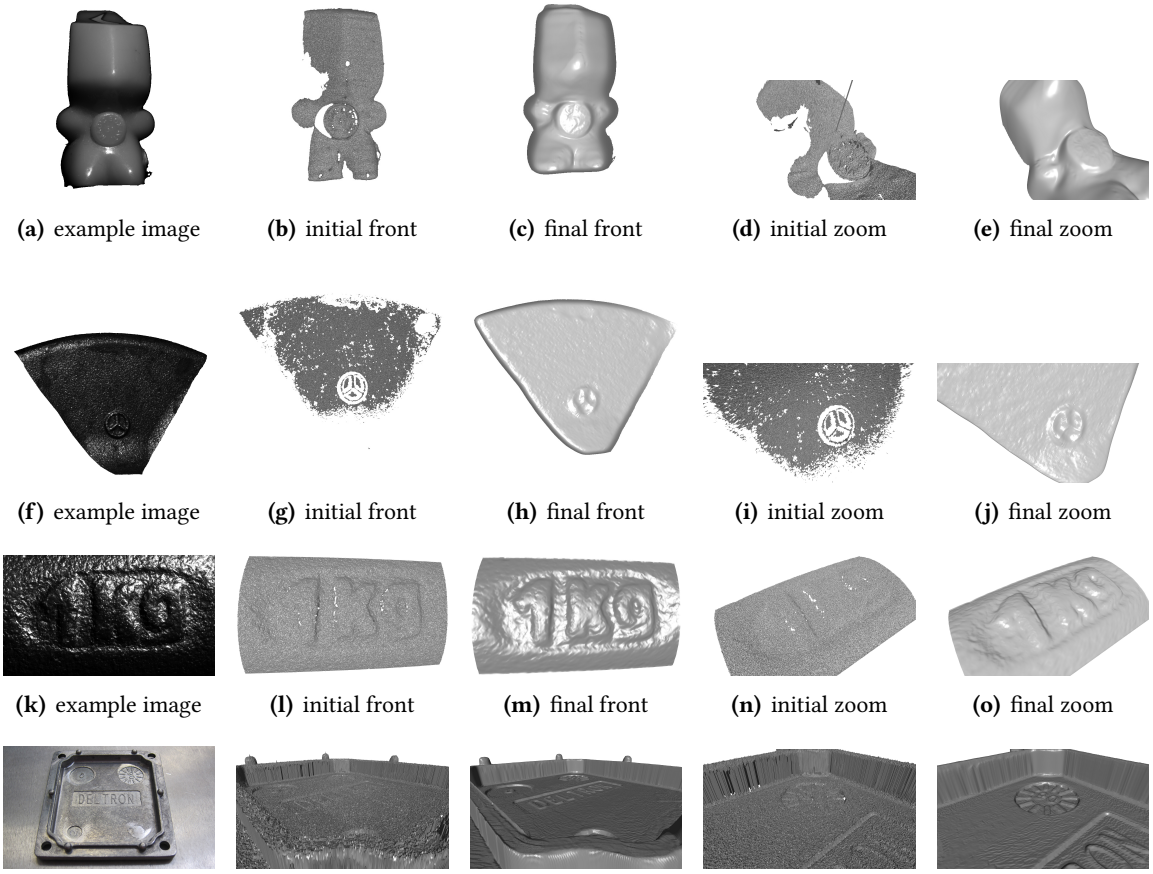
(f) example image **(g)** initial front **(h)** final front **(i)** initial zoom **(j)** final zoom

(k) example image **(l)** initial front **(m)** final front **(n)** initial zoom **(o)** final zoom

(p) example image **(q)** initial front **(r)** final front **(s)** initial zoom **(t)** final zoom

Figure 6.19.: Reconstruction results - object set 1

Especially the objects presented in Fig. 6.19 show strong interreflections and comprise diffuse and specular reflections with different surface albedoes. Nevertheless, the approach successfully obtains a finely detailed reconstructed surface, with an unaffected large scale shape.
The second set of objects (see Fig. 6.20) shows that the algorithm is even applicable to plastic and metallic surfaces, whose surfaces are reconstructed with qualitatively similar results as those in Fig. 6.19.
While each object surface exhibits the details visible in the 2D object images, note that the surface appears slightly less crisp than expected from the images. This is due to enforcing integrability (see Eq. 6.3), which smooths steep gradients, and due to the numerical Poisson solving step (see Eq. 6.6) inducing a slight low-pass filtering effect.



(a) example image **(b)** initial front **(c)** final front **(d)** initial zoom **(e)** final zoom

(f) example image **(g)** initial front **(h)** final front **(i)** initial zoom **(j)** final zoom

(k) example image **(l)** initial front **(m)** final front **(n)** initial zoom **(o)** final zoom

(p) example image **(q)** initial front **(r)** final front **(s)** initial zoom **(t)** final zoom

Figure 6.20.: Reconstruction results - object set 2

6.4.4 Conclusion

The presented algorithm allows the photometric 3D reconstruction of surfaces that exhibit a wide variety of reflectance behaviors and possible occurrences of interreflections. It is able to incorporate absolute depth data, which prevents the reconstructed surfaces from a deviation from the large scale shape and thus overcomes that characteristic problem known from Photometric Stereo (PS) or Shape from Shading (SfS). In contrast to other algorithms presented in the literature, it (1) regards non-Lambertian surfaces, (2) is independent from a priori known BRDF parameters, (3) models pixel-wise incident light and viewing directions, and (4) accounts for interreflections.

The conducted experiments assess the performance of the method with various evaluation aspects: The qualitative evaluation has shown a reduction of outliers and noise, while the large scale shape is preserved. Additionally, fine surface details have been incorporated through using image data, which were previously not visible in the laser scanner surface scans. These improvements are observed for concave, i.e. interreflection-affected, regions as well as for interreflection-free convex surface regions. However, a slight loss in sharpness can be observed as the obtained surfaces appear overall (to a small extent) smoother than expected from their images, which is probably due to low-pass filtering effects induced by numerically solving the Poisson equation, i.e. by solving the differential equation using finite differences.

The evaluation with caliper objects of known shapes showed that the standard deviation of the depth measurements (measurements noise) decreases by a factor of ≈ 3, which is very favorable in terms of measurement accuracy.

7 Contribution - spatially varying BRDFs

The reconstruction algorithms described in the chapters before rely on the assumption of homogeneous surfaces, i.e. surfaces that only consist of a single material and exhibit a uniform color. For the application to arbitrary surfaces, it is required to either determine the surface material/color in a unified reconstruction approach or to segment the observed scene based on spectral cues. The following chapter introduces an algorithm for the latter approach using k-means clustering and mean shift segmentation for the separation of various surface materials based on multispectral data.

This chapter is adapted and/or adopted from [Schick, Herbort, Grumpe, and Wöhler, 2013].

The following sections present a method for the separation of material components based on their multispectral reflectance characteristics using a static object, a single light source position, and a fixed camera position. For industrial applications, it is desirable to use as few light and camera positions as possible, since this facilitates the acquisition process and reduces the required recording time. However, this comes at the cost of the measurements covering only a small range of possible angles of observance ($\vartheta_o \in [-90° ... + 90°] \subset \mathbb{R}$) and incidence ($\vartheta_i \in [-90° ... + 90°] \subset \mathbb{R}$). While it is challenging to cluster and/or perform BRDF model fits using that data basis, it will be shown that it is possible if the underlying models are modified. In this way, it is possible to obtain correct segmentations for almost the whole object, even though there is no wide range of illumination and viewing angles available due to the single view single light configuration. Different clustering approaches (k-means clustering and mean-shift clustering) are compared for the examination of a surface with sharp material changes (color patches), and for a surface with smooth material changes (iron vs. iron oxide). For the segmentation, it is physically impossible to cluster the data based on specular reflections, since these only depend on the color of the incident light and are independent of the surface color as explained in detail by Shafer [1985]. The segmentation can thus be performed on the diffuse part only. Is is shown that it is possible to improve the segmentation results considerably by an empirical modification of the modeled reflectance behavior.
In the subsequent stage, each cluster is used to obtain the parameters of a model-based spectral BRDF[1], and the performances of the Phong and Cook-Torrance [Phong, 1975; Cook and Torrance, 1981] models are compared. Additionally, it is shown that the clustered data can be used to determine the abundance distribution of – in this case – iron and iron oxide using linear unmixing without any adjustment of the underlying BRDF.

The problem of separating material from shape and illumination has been examined by some authors previously. However, most of them use some sort of varying light source position and/or varying viewpoint. Tominaga [2002] applies multispectral data to the problem of circuit board element segmentation. Initially, two images are recorded with different illumination directions, which are used for a very coarse initial segmentation and measurements that contain specular reflections are removed. The remaining pixels are clustered based on a set of empirically defined if-then-else classification rules, which emerge from a priori knowledge of the spectral reflectance properties of the circuit board elements. In summary, Tominaga [2002] operate on the diffuse part only, require different illumination directions, and rely on a priori knowledge for the clustering process.
Lensch et al. [2003] acquire geometric and photometric data using different light source positions and viewpoints for a set of "lumitexels", which are assembled as a vector of geometric and photometric measurements. Afterwards, they fit the Lafortune BRDF model to all lumitexels to create two BRDF models along with a covariance matrix of the fitting parameters. Based on the error between lumitexels and those two BRDFs, they split the surface consecutively into two clusters. This procedure is repeated for the cluster with the greatest deviation of measured and fitted BRDF until a clear segmentation of all material components is achieved. Further techniques for the separation of specular/diffuse reflections and for the computation of photometric/geometric invariants can be found in Section 3.5.4.3.

[1] Bidirectional Reflectance Distribution Function

7.1 Multispectral object segmentation

In the following, a novel method for the segmentation of multispectral data is presented. In contrast to Lensch et al. [2003] and Narasimhan et al. [2003], this applies a fixed object, light source, and camera position. Unlike Tominaga [2002], it is not required to provide any prior knowledge of the material reflectance. Fig. 7.1 gives an overview of the algorithm. Initially, the depth data is recorded with a 3D laser pattern projector and image data for $N_\lambda = 8$ distinct wavelengths between 450 and 800 nm is acquired using bandpass interference filters[2]. The light source positions and intensities are known a priori following the calibration procedure described by Lenoch et al. [2012]. Note that in contrast to previous experiments, where narrowband LEDs have been used for scene illumination (see Chapter 6), a halogen light source with a much broader wavelength spectrum is required to allow meaningful multispectral data acquisition. In addition to the object image data, the image data of a white diffuse reflectance standard[3] is recorded, which is later used to infer the BRDF from the observed intensity data (Section 7.2).

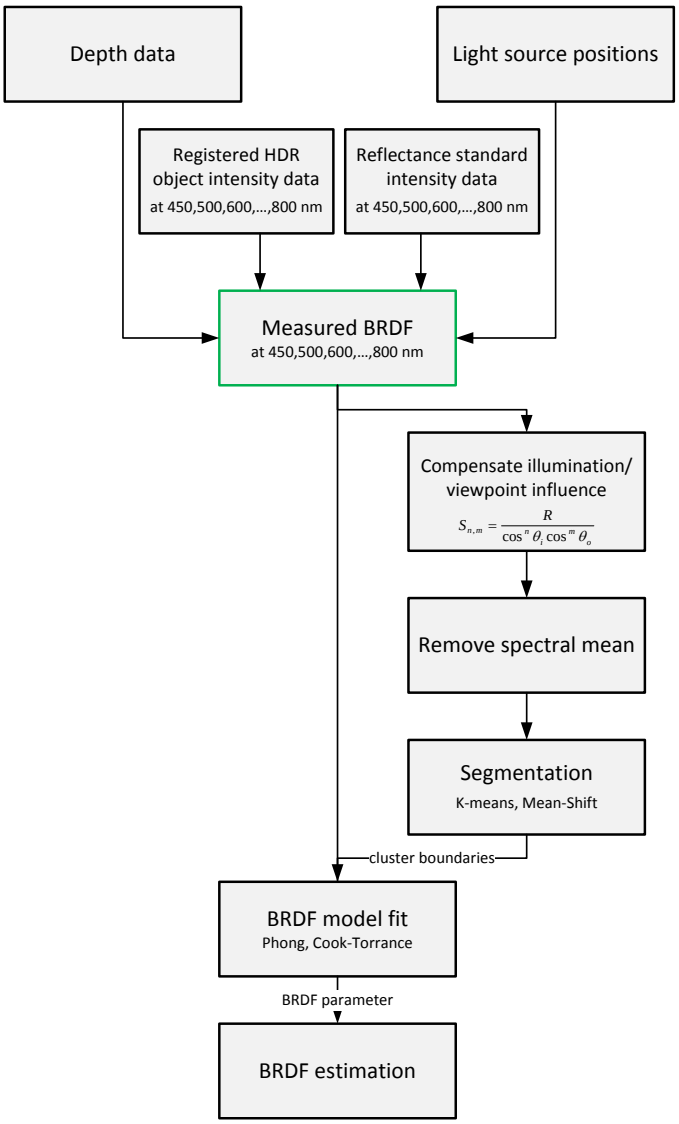


Figure 7.1.: Methodical overview.

Afterwards, the influence of the incident and viewing directions is compensated, the spectral mean is removed, and the k-means and/or mean-shift algorithms are employed for clustering. With the clustering boundaries available, it is possible to obtain BRDF parameters for each cluster and each multispectral channel within these clusters. Algorithmic details for these stages are explained in the following section.

[2]Thorlabs bandpass interference filters, center wave length CWL = $[450, 500, 550, 600, 650, 700, 750, 800]\ nm$, full width at half maximum FWHM = $10\ nm$

[3]SphereOptics Zenith Polymer Diffuse Reflectance Standard SG3052

7.2 Algorithm

7.2.1 Data preparation

In order to account for even slight camera displacements during bandpass filter replacements or by the optically active bandpass filters themselves, an image registration and subsequent affine transformation step are conducted to account for these distortions. The dynamic range of the images is increased by using high dynamic range imaging (HDR), which allows capturing bright specular regions and dark surface regions without reaching saturation or losing low-contrasted details in the camera noise.
Since the depth data are corrupted by a considerable amount of noise, a model-based fit to the acquired data is performed. The models are chosen to simply match the object shape, i.e. a cylindrical model for the "cup" dataset and a plate for the "triangle" dataset (see Fig. 7.2 and Fig. 7.7). This is achieved by a minimization of the mean squared distance between the measured data points and the model surface. A whole description of that process lies beyond the focus of this thesis and is therefore omitted, a detailed description can be found e.g. in [Lukacs et al., 1998; Eberly, 2008]. Note, that the multispectral data segmentation algorithm below is not restricted to the application of these models. The models only account for the partially severe amount of noise present in the 3D scans. The model fitting step is unnecessary, if the scanned surface contains only low noise.

7.2.2 BRDF measurement

The main idea for BRDF measurement determination lies in acquiring multispectral object and reflectance standard data, which are then related to each other. The following derivation thus aims for an expression of the observed BRDF values f_r.
The measured object intensity can be expressed as

$$I_{ob} = C \cdot \frac{I_0}{r_{l,ob}^2} \cdot f_{r,ob} \cdot \cos\vartheta_{i,ob}, \tag{7.1}$$

where the distance from the light source to the object is denoted $r_{l,ob}$. The light source intensity is I_0, the BRDF is $f_{r,ob}$. Further influencing quantities like camera gain or interference filter attenuation are all accumulated within the constant C. $\vartheta_{i,ob}$ denotes the angle of light incidence on the object surface, i.e. the angle between the local direction $\mathbf{l}$ of the incident light and the local normal vector $\mathbf{n}_{ob}$ of the surface such that $\cos\vartheta_{i,ob} = \langle \mathbf{l} \cdot \mathbf{n}_{ob} \rangle$.
Analogously, the reflectance standard is described by

$$I_{re}(x_0, y_0) = C \cdot \frac{I_0}{r_{l,re}^2} \cdot f_{r,re} \cdot \cos[\vartheta_{i,re}(x_0, y_0)] \tag{7.2}$$

In contrast to Eq. 7.1, more prior knowledge can be included here. For the reflectance standard, an approximately diffuse reflectance behavior is assumed and an a priori known spectral albedo of $\rho_{re} = 0.99$ as specified by the manufacturer. The general BRDF $f_{r,re}$ is thus replaced by a Lambertian reflectance term,

i.e. $f_{r,re} = \frac{\rho_{re}}{\pi}$, yielding

$$I_{re}(x_0, y_0) = C \cdot \frac{I_0}{r_{l,re}^2} \cdot \frac{\rho_{re}}{\pi} \cdot \cos[\vartheta_{i,re}(x_0, y_0)]. \tag{7.3}$$

Dividing Eq. 7.1 by Eq. 7.3 then leads to

$$f_{r,ob} = \frac{I_{ob}}{I_{re}} \cdot \frac{\rho_{re}}{\pi} \frac{\cos \vartheta_{i,re}(x_0, y_0)}{\cos \vartheta_{i,ob}} \cdot \frac{r_{l,ob}^2}{r_{l,re}^2}, \tag{7.4}$$

which is connected to the reflectance R_{ob} by

$$f_{r,ob} = R_{ob} \cdot \frac{1}{\cos(\vartheta_{i,ob})}. \tag{7.5}$$

It is apparent that the relation of measured object intensities with reflectance standard data provides an elegant way for BRDF measurement determination. The BRDF obtained from Eq. 7.4 can now be used to segment areas based on their spectral reflectance. Ideally, the separated areas correspond to areas with the same material characteristics. Practically, a specular reflection within the surface leads to incorrect segmentation results. Due to the lack of varying light source or viewpoint position, a precise estimation of these illumination effects can not be achieved. Instead, the measured reflectance samples R_{ob} (or, respectively, measured BRDF f_r) are modified such that

$$S_{n,m} = \frac{R_{ob}}{\cos^n \vartheta_i \cos^m \vartheta_o} = \frac{f_r}{\cos^{n+1} \vartheta_i \cos^m \vartheta_o}, \tag{7.6}$$

which has been found to alleviate that effect. The angles $\vartheta_i = \arccos(\langle \mathbf{l} \cdot \mathbf{n} \rangle)$ and $\vartheta_o = \arccos(\langle \mathbf{v} \cdot \mathbf{n} \rangle)$ denote the angles between incident light direction $\mathbf{l}$ and observance direction $\mathbf{v}$ with the local surface normal $\mathbf{n}$, respectively. Additionally, a spectral mean subtraction is applied to Eq. 7.6 prior to the segmentation since this has been found to improve the segmentation results (see Section 7.3): The spectral mean corresponds to the less significant average brightness of a point on the surface and it is naturally more meaningful to examine the point's color characteristics.
For segmentation, the k-means algorithm [Marsland, 2009] and the mean-shift algorithm [Comaniciu and Meer, 2002] are used in their classical form, i.e. without functional adaptations. For the mean-shift algorithm, a disk-shaped kernel is applied that evenly weights the involved measurements [Finkston, 2006].

7.2.3 Clustering and BRDF fitting

For clustering similar BRDF measurements, the k-means algorithm [Marsland, 2009] and the mean-shift algorithm [Comaniciu and Meer, 2002] are used in their classical form, i.e. without functional adaptations. For the mean-shift algorithm, a disk-shaped kernel is used that evenly weights the involved measurements [Finkston, 2006]. Once the clusters are obtained, a Phong [Phong, 1975] and a Cook-Torrance [Cook and Torrance, 1981] BRDF are estimated by nonlinearly fitting the $u = [1...U] \subset \mathbb{N}$ observed measurements

$f_{r,u}^{observed}$ to the corresponding model based estimations $f_{r,u}^{modeled}$ such that

$$P^* = \underset{P}{\operatorname{argmin}} \sum_{u=1}^{U} (f_{r,u}^{observed} - f_{r,u}^{modeled})^2 \tag{7.7}$$

becomes minimized with respect to the model parameters P. Both BRDF models contain a Lambertian term $\frac{k_d}{\pi}$ to describe (ideal) diffuse reflection. The applied physically plausible Phong BRDF is

$$f_r = \frac{k_d}{\pi} + k_s \frac{a+2}{2\pi} \underbrace{\langle \mathbf{r} \cdot \mathbf{v} \rangle^a}_{\cos^a \vartheta_{rv}}, \tag{7.8}$$

where a denotes an exponential coefficient for varying the angular extent of the specular component. The specular characteristics are modeled based on the angle ϑ_{rv} between the direction of mirror reflection $\mathbf{r}$ and observation $\mathbf{v}$.
The Cook-Torrance BRDF model consists the following terms:

$$f_r = k_d \frac{1}{\pi} + k_s \frac{F(n)}{\pi} \frac{D(m) \cdot G}{\cos \vartheta_i \cos \vartheta_o}. \tag{7.9}$$

The amount of reflected light depends of the three components F, D, and G, which denote the Fresnel reflection coefficient, the distribution function of the directions of the microfacets, and the geometrical attenuation factor, respectively. F improves the model exactness for grazing incidence angles, D defines the influence of surface roughness on lobe width, and G accounts for self shadowing and occlusion. The parameters of the Cook-Torrance BRDF are the weights k_d and k_s, the index of refraction n and the surface roughness m. Details can be obtained from [Cook and Torrance, 1981].

7.2.4 Spectral unmixing

Spectral unmixing denotes the decomposition of the spectrum of a compound material into the spectra of the known pure materials out of which it consists by estimating the corresponding relative frequencies [Keshava and Mustard, 2002]. To estimate a distribution of a compound of two materials ("endmembers") – in this case: iron and iron oxide (rust), see Fig. 7.7(a) – a linear unmixing approach [Keshava and Mustard, 2002] is used such that the pixel-wise spectral BRDF $\mathbf{f_r} = [f_{r,1}, ..., f_{r,N}]^T \in \mathbb{R}^{N \times 1}$ sampled at N distinct wavelengths is supposed to be composed of two components $\mathbf{x}_{\text{iron}} \in \mathbb{R}^{N \times 1}$ and $\mathbf{x}_{\text{rust}} \in \mathbb{R}^{N \times 1}$ according to

$$\mathbf{x}_{\text{iron}} + \mathbf{x}_{\text{rust}} = \mathbf{BRDF} = \mathbf{f_r}. \tag{7.10}$$

These $\mathbf{x}$, in turn, are weighted variants of the spectral BRDFs of the pure elements ($\mathbf{r}_{\text{iron}}$ and $\mathbf{r}_{\text{rust}}$):

$$\mathbf{x}_{\text{iron}} = a_{\text{iron}} \cdot \mathbf{r}_{\text{iron}} \qquad \mathbf{x}_{\text{rust}} = a_{\text{rust}} \cdot \mathbf{r}_{\text{rust}}. \tag{7.11}$$

The vectors $\mathbf{r}_{\text{iron}}$ and $\mathbf{r}_{\text{rust}}$ represent reference spectra of iron and rust, which have been determined by averaging $M \in \mathbb{N}$ manually chosen BRDF samples $r_{iron,m}$ and $r_{rust,m}$ with $m = 1...M$ from the image:

$$\mathbf{r}_{\text{iron}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{r}_{\text{iron},m}, \quad \mathbf{r}_{\text{rust}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{r}_{\text{rust},m}. \tag{7.12}$$

This approach corresponds to the concept of image-based endmember selection [Keshava and Mustard, 2002]. To estimate the actual abundance distribution $a_{\text{iron}} \in [0...1] \subset \mathbb{R}$ and $a_{\text{rust}} \in [0...1] \subset \mathbb{R}$, the following least-square transformation of Eq. 7.11 is applied:

$$a_{\text{iron}} = (\mathbf{r}_{\text{iron}}^T\mathbf{r}_{\text{iron}})^{-1}\mathbf{r}_{\text{iron}}^T\mathbf{f}_{\mathbf{r}} \tag{7.13}$$

$$a_{\text{rust}} = (\mathbf{r}_{\text{rust}}^T\mathbf{r}_{\text{rust}})^{-1}\mathbf{r}_{\text{rust}}^T\mathbf{f}_{\mathbf{r}} \tag{7.14}$$

For this to be physical meaningful, the following two constraints have to be satisfied:

$$a_{\text{iron}}(i) > 0,\ a_{\text{rust}}(i) > 0 \tag{7.15}$$

$$a_{\text{iron}}(i) + a_{\text{rust}}(i) = 1 \tag{7.16}$$

The weights $\mathbf{a} = (a_{\text{iron}}, a_{\text{rust}})$ resulting from Eq. 7.13 and Eq. 7.14 are thus normalized to unit length as required by Eq. 7.16; the non-negativity constraint (Eq. 7.15) was not required explicitly during our experiments since no negative weights occurred. This is a slight variation of classic constrained least squares [Keshava and Mustard, 2002] that comes without the need for Lagrangian multipliers. Note that the derivation shown above can be applied to the separation of two *arbitrary* materials with known or measured pure material BRDFs $\mathbf{r}_{\text{Material 1}}$ and $\mathbf{r}_{\text{Material 2}}$. In that case, the BRDF vectors $\mathbf{r}_{\text{iron}}$ and $\mathbf{r}_{\text{rust}}$ are replaced by the new BRDF vectors in the derivation above. The linear unmixing approach is also applicable to $N \in \mathbb{N} > 2$ material components.

7.3 Experiments and results

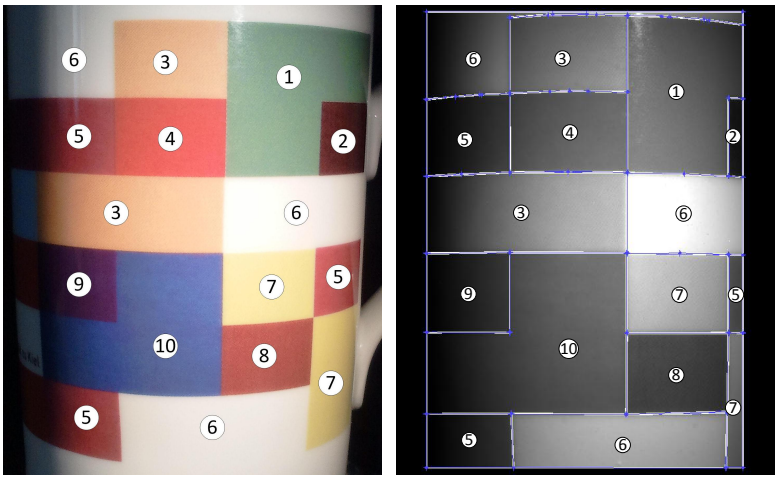
The following sections provide a detailed discussion of the “cup” dataset, which contains sharp boundaries (see Section 7.3.1), and of the “triangle” dataset, which exhibits a gradual transition between iron and iron oxide (see Section 7.3.2).

7.3.1 Object with sharp region boundaries

The first part of this work deals with the segmentation of a dataset with abruptly changing regions on the surface of a cup as shown in Fig. 7.2(a). To compare segmentation results quantitatively, a reference cluster map has been created manually, which is shown in Fig. 7.2(b). With that being available as a ground truth, the detection rate D is defined to be

$$D = \frac{\text{tp}}{\text{all}} \cdot 100\%, \tag{7.17}$$

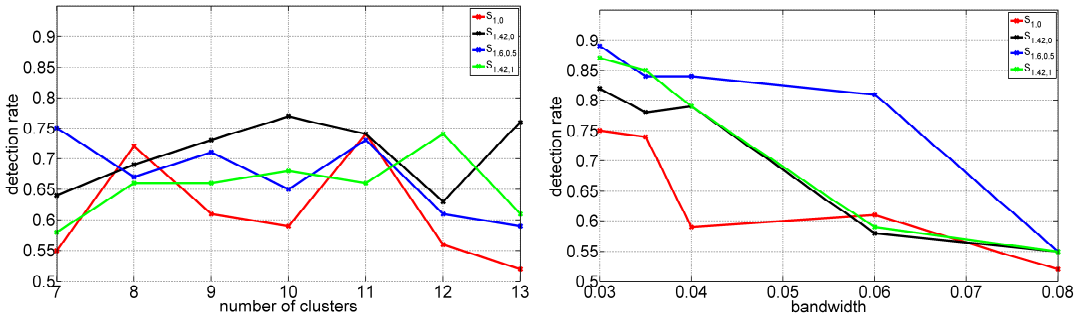
i.e. the percentage of correctly classified pixels (tp) within the whole set of classifiable pixels in the image (all). Fig. 7.3 illustrates the pixel-wise detection rate for k-means and mean-shift for a varying number of



(a) Color regions of the examined cup **(b)** Manually defined reference clusters

Figure 7.2.: Detection rate ground truth

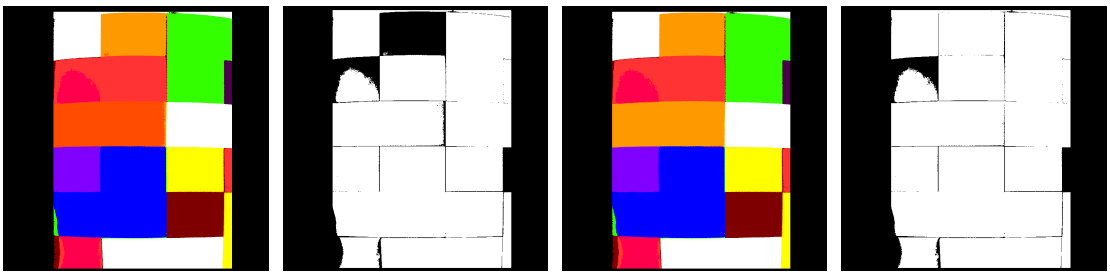
clusters, bandwidth, and BRDF modification factor $S_{n,m}$, as described in Eq. 7.6. The red curve is equivalent to a direct BRDF segmentation (spectral mean free) without any BRDF modification. Both segmentation algorithms, especially mean-shift, benefit from the reflection model modification. A detailed image of the best segmentation result yielded by mean-shift using $S_{1.6,0.5}$ is shown in Fig. 7.4. The segmentation error in the lower left part of the cup can be attributed to slightly overexposed input data originating from spurious illumination from interreflections with the experimental environment. The correct separation of the three reddish regions 4, 5 and 8 (see Fig. 7.3(a)) is especially challenging, due to the fact that the corresponding spectra are very similar and need to be distinguished based on merely 8 spectral measurements.
Additionally, an oversegmentation of the orange cluster ③ can be observed. A toleration of the oversegmentation leads to a segmentation result with a detection rate of $D \approx 95\%$ as shown in Fig. 7.4(c) and



(a) k-means detection rate over number of clusters **(b)** mean-shift detection rate over bandwidth

Figure 7.3.: Comparison of the correct segmentation rate between k-means and mean-shift for abrupt cluster transitions. Each curve (red, black, blue, green) corresponds to a different BRDF modification factor S.

Fig. 7.4(d). Note that the segmentation is correct over nearly the whole cup, i.e. even regions with steep surface gradients (cup borders), whose measurements are typically hard to obtain due to rapidly changing incidence and viewing angles, have been segmented correctly. The measured BRDF can now be approxi-



(a) Best mean-shift segmentation result without further treatment.

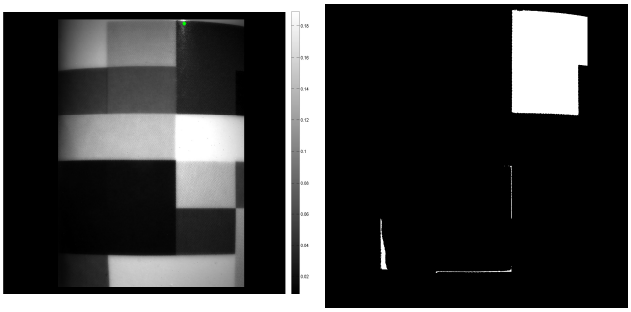
(b) Correctly detected areas (detection rate: 88.45 %).

(c) Best mean-shift segmentation result with toleration of the oversegmented orange cluster.

(d) Correctly detected areas (detection rate: 94.79 %).

Figure 7.4.: Segmentation result for $S_{1.6,0.5}$. Note the oversegmentation of the orange cluster ((a) and (b)), which can be neglected ((c) and (d)). Note that areas at the border of the cup with steep surface gradients have been segmented correctly.

mated by a Phong and/or a Cook-Torrance BRDF model for each determined segmented cluster. Using the obtained BRDF parameters allows for a determination of a BRDF value for every half polar observation angle $\vartheta_o \in [-90°, ..., 90°] \subset \mathbb{R}$, for each incident light angle $\vartheta_i \in [-90°, ..., 90°] \subset \mathbb{R}$, and for each pixel belonging to the cluster. In this case, a pixel of a BRDF at 600 nm wavelength is analyzed, which is located closely to a specular highlight. This pixel corresponds to cluster ①, which is generated by a mean-shift clustering of $S_{1.42,1}$ and visualized in Fig. 7.5(a) and Fig. 7.5(b). Fig. 7.6(a), Fig. 7.6(b) and Fig. 7.6(c) show the results of BRDF estimation for the Phong model (single lobe as well as lobe+spike variant) and the Cook-Torrance model (lobe+spike). Note that the range of incidence and viewing angles available for BRDF estimation is narrow but the estimation is robust due to the large number of pixels involved. The obtained BRDF may become inaccurate for angles a long way off from the underlying samples used during the

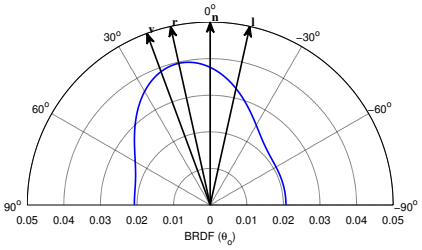


(a) Measured BRDF at a wavelength of 600 nm.

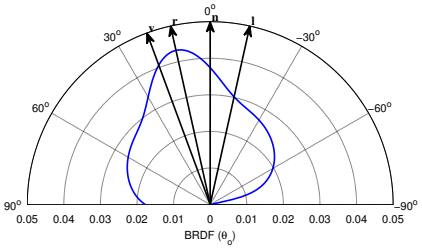
(b) Segmented cluster including observed data point (mean-shift approach with $S_{1.42,1}$).

Figure 7.5.: Data point and corresponding clusters.

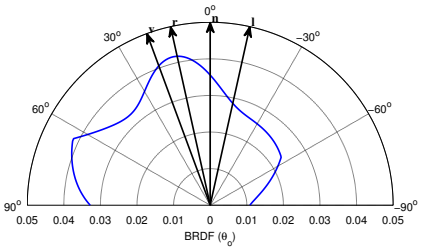
parameter estimation stage, but model usage far off the data basis should be avoided anyway.



(a) Phong fit (lobe)



(b) Phong fit (lobe+spike)



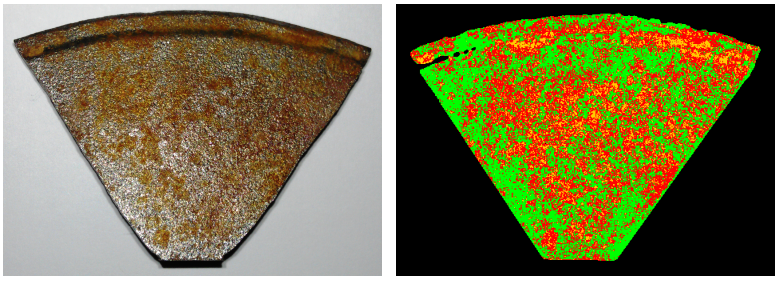
(c) Cook Torrance fit (lobe+spike)

Figure 7.6.: BRDF fitting results

An important observation is that the BRDF decreases for large absolute viewing angles $|\vartheta_o| \geq 60°$. This effect has also been observed with other objects and scene settings. Relating this effect to the varying specular part of reflection does not seem very realistic due the fact that this decrease occurs at angles far away from the ideal specular reflection angle $\vartheta_{rv} = 0°$. In fact, a non-ideal diffuse reflection for grazing viewing angles is a more plausible explanation. Oren and Nayar [1994] introduce a more detailed and complex diffuse reflectance term, but they describe an increased reflectance, which is contrary to the aforementioned observations. The reflectance in this work better corresponds with a “diffuse fall-off” described e.g. by Dorsey et al. [2008]. This effect models decreasing reflectance for larger angles of observance and is partially taken into account by Lafortune et al. [1997].

7.3.2 Object with gradual region boundaries

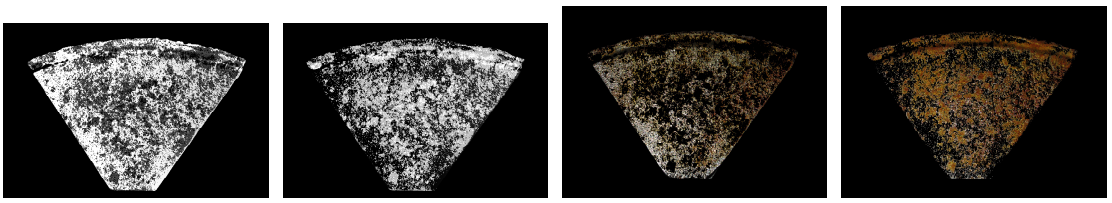
Initially, a k-means segmentation has been performed (see Fig. 7.7) of the oxidized iron object without any modifications of the BRDF, i.e. using $S_{(1,1)}$. To obtain a robust segmentation result without the influence of spurious effects like specular reflections and other illumination inhomogeneities, a planar arrangement of the object with high values of the specular angle ϑ_{rv} has been chosen in order to minimize the occurrence of specular reflections. However, due to the roughness of the surface there is an inevitable amount of small surface parts that reflect light specularly into the camera. These have been taken into account by a k-means segmentation with 3 clusters, i.e. iron, oxidized iron, and a "garbage" class that collects outliers. Based on



(a) Probed oxidized iron object. **(b)** Segmentation with k-means (3 clusters).

Figure 7.7.: Visual comparison between original and segmented clusters. A third class ("garbage class") besides iron and rust has been introduced to collect outliers.

10 manually chosen reference data points for iron and rust that that are each averaged to provide robust reference spectra $\mathbf{r}_{\text{iron}}, \mathbf{r}_{\text{rust}}$ for the linear unmixing approach described in Section 7.2.4, an abundance distribution of both materials can be estimated as shown in Fig. 7.8(a) and Fig. 7.8(b). On the basis of this distribution and an image registration approach that relates the color image Fig. 7.7(a) with the measurement data Fig. 7.8(a) and Fig. 7.8(b), a mask of the two materials can be generated as shown in Fig. 7.8(c) and Fig. 7.8(d).



(a) Generated abundance distribution of iron. **(b)** Generated abundance distribution of rust. **(c)** Areas with at least 60% iron. **(d)** Areas with at least 60% rust.

Figure 7.8.: Linear unmixing results.

7.4 Conclusion

An approach for the acquisition and segmentation of spectral BRDF data has been presented. The data are recorded by relating object intensity data and reflectance standard data, which then directly provides BRDF

measurements. Those data have been used to show that it is possible to obtain accurate segmentation results and endmember abundance estimates for objects with rapidly and with gradually changing materials, even if only a single light source position and a single viewpoint is available. The object surface with abrupt region boundaries has been classified with an accuracy of 88.5% (94.8% with some oversegmentation tolerance). The application of linear spectral unmixing to the separation of iron from rust for the gradually changing material has provided a qualitatively realistic result.
In summary, note that it is possible to determine object regions correctly even if only a single light source and a single viewpoint are available, which is very important for e.g. industrial measurement setups.
Some limitation has been observed when using raw measured depth data, which has been found to be too noisy for correct incidence and viewing angle determination. This problem has been solved by using a model-based cylindrical and plane fit for the respective objects. Using laser range scanners with higher depth and lateral resolution can overcome that limitation easily.
The Phong and Cook-Torrance BRDF models were then fitted to the previously segmented cluster data. Due to the sparse data input (limited rage of incidence and viewing directions), some uncertainties in determining a hemispherical BRDF remain, especially for obliquely viewed surface parts. Additionally, the obtained BRDF tends to exhibit lower reflectance values for large observation angles $|\vartheta_o| \geq 60°$, which can be explained by a "diffuse fall-off" [Dorsey et al., 2008]. This issue requires considerable attention in further research, since it requires a phenomenological (rather than empirical) adjustment of the reflectance model.

8 Outlook, conclusion and future work

With the results from the presented surface reconstruction approaches in mind, it is an imminent question, if the costly and technologically demanding 3D scanner is actually needed for surface reconstructions. This is motivated by the very coarse and partially sparse 3D depth data required as a basis for the reconstruction as shown in Fig. 6.19 and Fig. 6.20. Even more so, the purely PS-based reconstructions shown in Section 6.4.3.1 provide promising results that PS alone might suffice.
The following chapter introduces the idea of a new system that replaces the laser scanner by a second camera to build a stereo camera array. In this way, the system is less expensive, but still provides highly accurate 3D shape measurements. Additionally, it is shown that classic blockmatching stereo can be performed on normal fields, which allows the reconstruction of non-Lambertian surfaces that are highly challenging for stereo approaches.

This chapter is partly adapted and/or adopted from [Fiegel, 2013].

8.1 Outlook

Once the laser scanner is removed from the measurement setup, it is unlikely that PS can be used as the only reconstruction component due to the discussed large-scale shape deviations. However, if only surface smoothness needs to be measured or small-scale surface defects need to be detected, then PS fulfills the requirements, since smoothness and surface defects are small-scale features and that is where PS is accurate. The idea of removing the laser scanner can still be pursued if the projector is not just removed but replaced with a much less costly second camera. Both cameras then form a passive triangulation system, which is able to provide coarse, but large-scale-robust depth data. What remains is a problem for non-Lambertian surfaces, which exhibit specular spikes that appear at different locations in both cameras due to the changing viewpoint. This makes establishing correspondences very difficult [Wöhler and d'Angelo, 2009].
The following sections provide a solution that elegantly exploits PS data from both cameras. The approach has been examined during the master thesis by Fiegel [2013], which has been supervised by the author, and most of the results and images have been adopted from there. Matching on normal fields has been tried by Ikeuchi [1987], but he (1) deals with Lambertian surfaces, which disregards the crucial specularity invariance of the normal fields, (2) matches on a patch-level, and (3) requires several constraints.

8.1.1 Algorithm

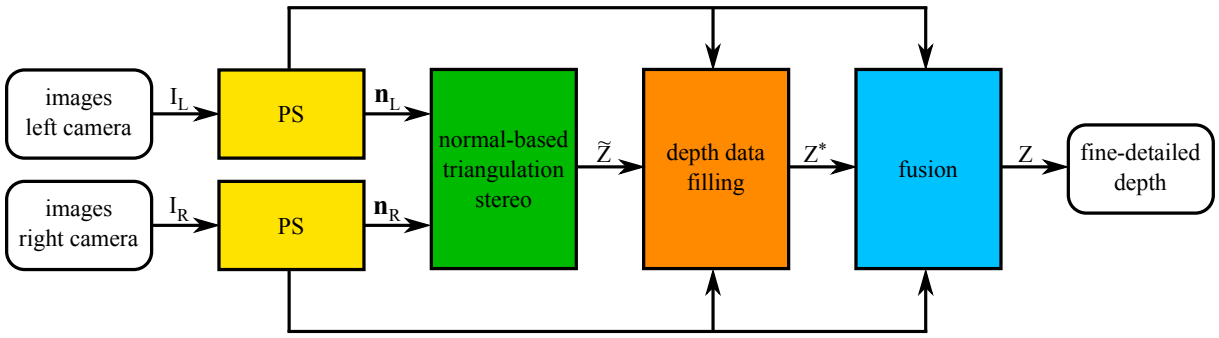
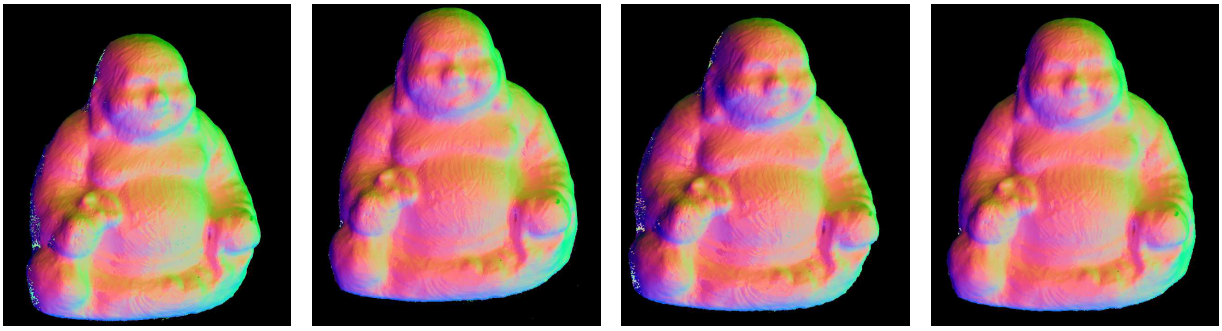


Figure 8.1.: Algorithm overview. Note the repeated usage of the surface normal data for triangulation stereo, depth data filling, and data fusion.

Fig. 8.1 gives an overview of the algorithm. Initially (yellow), a surface normal field is computed for both sets of camera images, i.e. the images from the left and from the right camera. This conforms with a transformation of the image data into the specularity-invariant surface normal space. The normal data is then used to perform a triangulation stereo approach (green), where missing measurements in the depth result $\tilde{Z}$ need to be filled (orange). Afterwards, the depth data Z^* and the normal fields are fused (blue) into a fine-detailed depth field Z. Note that the normal data (complies with surface gradient data) is used not only for the correspondences, but later on for filling and data fusion. More detailed descriptions of the respective stages are given in the following sections.



(a) $\mathbf{n}_L$ **(b)** $\mathbf{n}_R$ **(c)** rectified, rotated $\mathbf{n}_L$ **(d)** rectified $\mathbf{n}_R$

Figure 8.2.: (a) and (b) show the color-coded normals that are computed for both viewpoints with a PS approach under assumption of Lambert+Blinn+Blinn reflectance. These are both rectified using [Bouguet, 2008], which yields lens corrections and parallel epipolar lines as shown by the vertical alignment between (c) and (d). The left camera's normals are additionally rotated to match the right camera's orientation, which yields similar colors (normal orientations) for corresponding surface points between (c) and (d). Images have been adopted from [Fiegel, 2013].

8.1.2 Data acquisition and Photometric Stereo

Initially, HDR image data is acquired under successive illumination with one of the 18 point light sources as described in Section 4.3. Note, that the number of observing cameras does not increase the time required for the acquisition process since they operate simultaneously. However, the varying illumination still requires time multiplexed acquisition of the 18 images.
Then, the PS algorithm described in Section 6.1.5 is applied to compute pixelwise surface normals under assumption of Lambert+Blinn+Blinn reflectance without interreflection compensation to save computational expenses. Note, that once more no prior knowledge of the BRDF parameters is required since the algorithm computes these internally. Fig. 8.2 shows the normal field results along with the subsequent rectification and rotation of normals to match the orientation of the main camera (right camera). Both, rectification and rotation, are needed to ensure the computability of blockmatching stereo correspondences in the subsequent stage.

8.1.3 Triangulation stereo and filling

For simplicity, a classic blockmatching stereo algorithm has been employed for the establishment of correspondences between the left and the right camera's viewpoint. Note, that an intensity-image based approach tries to find matching points based on two $N \times M \times 1$ matrices where only one measurement per pixel is available. The intensity-based approach is additionally challenged by non-Lambertian surfaces since surface specularities are viewpoint-dependent and thus induce falsely positive correspondences.
Note that the normal fields $\mathbf{n}_l, \mathbf{n}_r \in \mathbb{R}^{N \times M \times 3}$ computed for the proposed approach provide three values for each pixel and are *independent* of specularities. Both properties are beneficial since they make the approach more robust due to being specularity-invariant and containing more information.
Fiegel [2013] proposed a combination of several blocksizes and similarity criteria (angle between normals, RMSE, normalized cross-correlation) and additionally secures the correspondence selection by plausibilizing it with the available surface gradient information. Note, that this means using the normal information again: The normal fields can be transformed into surface gradients (see Section 6.1.1), which contain the

(a) normal-based blockmatching stereo result **(b)** filling result

Figure 8.3.: (a) Result of the blockmatching stereo stage, (b) result after filling missing measurements with values based on an analysis of the available gradient-fields. Images adopted from [Fiegel, 2013].

depth variation between adjacent pixels. This can be used to select correspondences that have a certain similarity but additionally match the gradient requirements.
After the blockmatching stereo algorithm yielded a raw depth profile (see Fig. 8.3(a)), the missing measurements are filled by once more exploiting the available gradient information and filling the holes iteratively from their borders Fig. 8.3(b). In this way, it is possible to compute a surface that contains some noise but is robust on large scales.

8.1.4 Fusion

The triangulated depth data z_{triang} can now be fused with the depth data z_{PS} obtained from PS gradient field integration using the M-estimator approach by Agrawal et al. [2006] as explained in Section 3.4. The data fusion can be done in several ways, e.g. using the linearized approach by Nehab et al. [2005]. For simplicity, the noisy high frequency part of z_{triang} is here replaced by the significantly more accurate high frequency part of z_{PS} using a simple filter approach

$$z_{\text{fused}} = z_{\text{triang}} * G_\sigma + z_{\text{PS}} * (1 - G_\sigma) \tag{8.1}$$

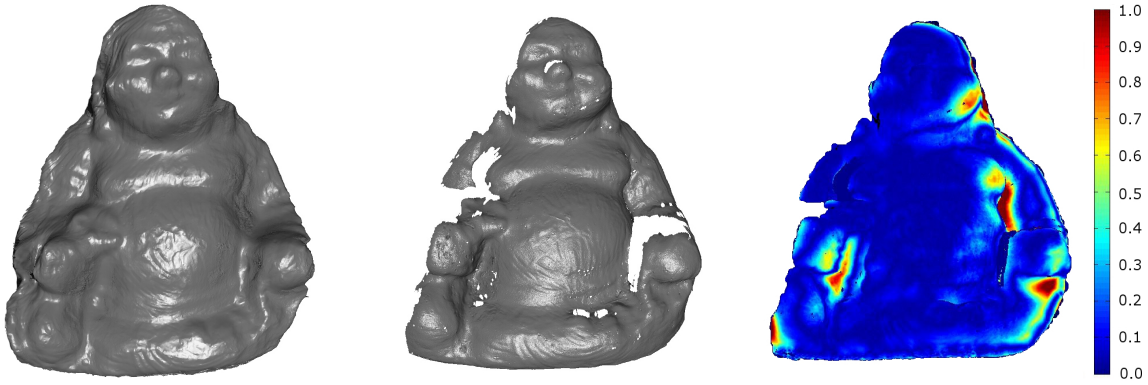
where G_σ denotes a Gaussian low-pass filter with standard deviation σ, and $(1 - G_\sigma)$ denotes its high-pass counterpart. The result is shown in Fig. 8.4(a) where one can observe the robustness of the large scale shape and the visibility of fine surface details.

8.1.5 Accuracy evaluation

The measurement accuracy is now evaluated based on the available 3D scanner data shown in Fig. 8.4(b), which serves as a ground truth such that

$$\Delta z = |z_{\text{scanner}} - z_{\text{Steteo+PS}}|. \tag{8.2}$$

It becomes clear from the pixelwise differences shown in Fig. 8.4(c) that the algorithm is very accurate



(a) fusion result **(b)** scanner ground truth **(c)** pixelwise absolute differences Δz in mm, RMSE = 0.2163 mm

Figure 8.4.: Note that (a) and (b) are shown under a slightly varying viewpoint which has been corrected using an iterative closest point algorithm before computing the pixelwise differences shown in (c). Images adopted from [Fiegel, 2013].

by means of large scale and small scale depth accuracy for most regions. However, it is obvious that the interreflection-affected surface concavities are error-prone, as shown by the yellow and red regions which mark depth errors of more than 0.6 mm. This is an expected result since the occurrence of interreflections has been disregarded as mentioned in Section 8.1.2.

For all convex regions that are mostly free of interreflections, the local depth deviations amount less than 0.1 mm, which is a very promising result. Note that the approach for the removal of interreflections described in Section 6.3 is very likely to reduce these errors tremendously.

8.2 Conclusion

All in all it can be assessed that the fusion of image-based surface shape information and absolute depth data is very beneficial in terms of reducing noise and outliers, filling missing measurements, and reconstructing surfaces with high accuracy. However, it is required that the measurement process not only captures (coarse) depth data, but also acquires image data with varying incident directions. This is more time consuming due to the sequential acquisition process, but only requires an extension of the existing setup by an array of (point) light sources.

This thesis presented three contributions that solve common problems that are encountered when reconstructing real surfaces using PS or SfS. These problems were (1) large scale shape deviations, (2) surface reflectance properties that are unknown prior to the reconstruction, and (3) the occurrence of interreflections. The large scale shape deviations have been resolved by fusing the image-based reconstruction result with absolute depth data (see Section 6.1), which has proven to be very beneficial for the reconstruction accuracy. Unknown reflectance parameters have been computed in a self consistent scheme (see Section 6.2), which yields improved parameters and increased accuracy (see Section 6.2.2). Interreflection can be handled in an iterative scheme (see Section 6.3), which again improves the reconstruction accuracy (see Section 6.3.2). Ultimately, all these techniques can be combined to form a high accuracy surface scan enhancement system (see Section 6.4) with very high reconstruction detail even for interreflection-affected surface concavities (see Section 6.4.3).

Further conclusions in related topics have been reached: The single image light source position calibration (see Section 5) provides accurate light source positions that allow the computation of locally adapted incidence angles. The BRDF models whose accuracy is the key element for detailed surface reconstructions have been reviewed extensively (see Section 2.4) and it has been shown that the choice influences the reconstruction accuracy considerably (see Section 6.4.3). Additionally, an approach for highly accurate material segmentations based on multispectral data has been presented and evaluated (see Section 7.1).

8.3 Future work

There are three main areas for future work, which have already been touched by recent or currently ongoing master theses. First, this regards locally varying reflectance parameters, since the presented system only deals with homogeneous surfaces, i.e. where the whole scene is only comprised of a single material. This is examined in the master thesis by Westerhoff [2013] and shows promising results for the estimation of a set of parameters for each pixel of a non-Lambertian surface. Second, the BRDF accuracy evaluation and the influence of BRDF accuracy on the PS reconstruction result need to be examined in greater detail. This is currently regarded by the master thesis by Lenoch [2013], which already identified BRDF properties that need to be considered for highly accurate reflectance modeling. Third, it is promising to pursue the stereo-based approach that has been examined by Fiegel [2013], which has been described above.

Bibliography

A. Abdelrahim, M. El-Melegy, and A. Farag. Realistic 3D reconstruction of the human teeth using shape from shading with shape priors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 64–69, 2012.

A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? *European Conference on Computer Vision*, 1(TR2006-021):578–591, 2006.

N. Ahmed, C. Theobalt, P. Dobrev, H.-P. Seidel, and S. Thrun. Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

T. Akenine-Möller, E. Haines, and N. Hoffman. *Real-Time Rendering 3rd Edition.* A.K. Peters Ltd, 2008.

N. G. Alldrin. Reflectance estimation under natural illumination. Technical report, University of California, San Diego, 2006.

N. G. Alldrin, S. P. Mallick, and D. J. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. *Conference on Computer Vision and Pattern Recognition*, 2007.

R. Anderson, B. Stenger, and R. Cipolla. Augmenting depth camera output using photometric stereo. *Machine Vision Applications*, 2011a.

R. Anderson, B. Stenger, and R. Cipolla. Color photometric stereo for multicolored surfaces. *International Conference on Computer Vision*, pages 2182–2189, 2011b.

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

F. R. Ashley, E. B. Murphy, and J. H. J. Savard. A computer controlled coordinate measuring machine. *The Bell System Technical Journal*, 49:2193 – 2202, 1970.

R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, 1997.

M. Bachmann, B. Gerken, T. Mager, C. Hedayat, S. Herbort, and C. Wöhler. Integration eines photometrischen Stereoverfahrens in ein Laserscanner-System zur 3D-Oberflächenrekonstruktion. *Oldenburger 3D Tage*, 12:1–8, 2013.

R. Bajcsy, S. W. Lee, and A. Leonardis. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *International Journal of Computer Vision*, 17(3):241–272, 1996.

B. Bartczak and R. Koch. Dense depth maps from low resolution time-of-flight depth and high resolution color views. *Lecture Notes in Computer Science*, 5876:228–239, 2009.

R. Basri, D. W. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Artech House Radar Library, 1963.

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

F. Bernardini, H. Rushmeier, I. M. Martin, J. Mittleman, and G. Taubin. Building a digital model of michelangelo's florentine pieta. *IEEE Computer Graphics and Applications*, 22:59–67, 2002.

P. J. Besel. *Advances in Machine Vision*, pages 1–63. Springer Verlag, 1989.

P. J. Besl. Active optical range imaging sensors. *Machine Vision and Applications*, 1:127–152, 1988.

U. Bhutta. Application of polarization data for 3D reconstruction. *Master Thesis, TU Dortmund University*, 2011.

O. Birkholz. Entwicklung und Validierung eines Simulationsmodells für Laserscanner im Fahrerassistenzsystem. *Master Thesis, Audi AG, TU Dortmund University*, 2012.

F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–240, 2004.

F. Bleich. Untersuchung von Verfahren zur Detektion von Gesichtern im Fahrzeuginnenraum auf Basis von Infrarotaufnahmen. *Bachelor Thesis, Kostal AG, TU Dortmund University*, 2012.

F. Bleich. Klassifikation dynamischer Handgesten auf Basis raum-zeitlicher Time-of-Flight Kameradaten. *Master Thesis, Kostal AG, TU Dortmund University*, 2013.

J. F. Blinn. Models of light reflection for computer synthesized pictures. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 11(2):192–198, 1977.

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

J.-Y. Bouguet and P. Perona. 3D photography using shadows in dual-space geometry. *International Journal of Computer Vision*, 35:129–149, 1999.

A. M. Bruckstein. On shape from shading. *Computer Vision, Graphics, and Image Processing*, 44(2):139–154, 1988.

U. Buck, S. Naether, B. Räss, C. Jackowski, and M. J. Thali. Accident or homicide: Virtual crime scene reconstruction using 3D methods. *Forensic Science International*, 225(1–3):75 – 84, 2013.

R. L. Carceroni and K. N. Kutulakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape reflectance. *International Journal of Computer Vision*, 49(2):175–214, 2001.

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

G. Chin, S. Brylow, M. Foote, J. Garvin, J. Kasper, J. Keller, M. Litvak, I. Mitrofanov, D. Paige, K. Raney, M. Robinson, A. Sanin, D. Smith, H. Spence, P. Spudis, S. A. Stern, and M. Zuber. Lunar reconnaissance orbiter overview: The instrument suite and mission. *Space Science Review*, 129:391–419, 2007.

H.-S. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. *CVPR 2008*, pages 1–8, 2008.

J. J. Clark. Active photometric stereo. *Conference on Computer Vision and Pattern Recognition*, pages 29–34, 1992.

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

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

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

R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 15(3):307 – 316, 1981.

V. Couture, N. Martin, and S. Roy. Unstructured light scanning to overcome interreflections. *International Conference on Computer Vision*, pages 1–8, 2011.

J. E. Cryer, P.-S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28 (7):1033–1043, 1995.

B. Curless and M. Levoy. Better optical triangulation through spacetime analysis. *International Conference on Computer Vision*, pages 897–994, 1995.

P. d'Angelo and C. Wöhler. Image-based 3D surface reconstruction by combination of photometric, geometric, and real-aperture methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(3):297–321, 2008.

D. J. Daniels. Surface-penetrating radar. *Electronics and Communication Engineering Journal*, 8:165–182, 1996.

N. D'Apuzzo and H. Mitchell. Medical applications. *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Congress Book*, 1:425–438, 2008.

B. D. Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. *Conference on Computer Vision and Pattern Recognition*, pages 2536–2543, 2009.

J. Diebel and S. Thrun. An application of markov random fields to range sensing. *Advances in Neural Information Processing Systems*, 18:291–298, 2006.

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

O. Drbohlav and M. Chantler. Can two specular pixels calibrate photometric stereo? *International Conference on Computer Vision*, 2:1850–1857, 2005.

O. Drbohlav and R. Sara. Specularities reduce ambiguity of uncalibrated photometric stereo. *European Conference on Computer Vision*, 2:46–60, 2002.

M. Drew and L. Kontsevich. Closed-form attitude determination under spectrally varying illumination. *Conference on Computer Vision and Pattern Recognition*, pages 985–990, 1994.

R. O. Dror, E. H. Adelson, and A. S. Willsky. Recognition of surface reflectance properties from a single image under unknown real-world illumination. *Proceedings of the IEEE Workshop on Identifying Objects Across Variations in Lighting: Psychophysics & Computation*, 2001a.

R. O. Dror, E. H. Adelson, and A. S. Willsky. Surface reflectance estimation and natural illumination statistics. *Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision*, 2001b.

R. O. Dror, E. H. Adelson, and A. S. Willsky. Estimating surface reflectance properties from images under unknown illumination. *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging IV,*, 4, 2001c.

R. O. Dror, T. K. Leung, E. H. Adelson, and A. S. Willsky. Statistics of real-world illumination. *Conference on Computer Vision and Pattern Recognition*, 2:164–171, 2001d.

H. Du, D. Goldman, and S. Seitz. Binocular photometric stereo. In *Proceedings of the British Machine Vision Conference*, pages 84.1–84.11, 2011.

A. Dür. An improved normalization for the ward reflectance model. *Journal of Graphics, GPU, and Game Tools*, 11:51–59, 2006.

J.-D. Duroua, M. Falconeb, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1), 2007.

D. Eberly. Fitting 3D data with a cylinder. *online http://www.geometrictools.com/Documentation/CylinderFitting.pdf*, 1:1–4, 2008.

A. Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, 3: 249–265, 1987.

F. T. Farmer and M. P. Collins. A new approach to the determination of anatomical cross-sections of the body by compton scattering of gamma-rays. *Physics in Medicine and Biology*, 16:577–586, 1971.

M. Fiegel. 3D-Rekonstruktion durch Fusion von Stereo-Tiefendaten mit photometrischen Oberflächeninformationen. *Master Thesis, TU Dortmund University*, 2013.

B. Finkston. Mean Shift Clustering. *online http://www.mathworks.com/matlabcentral/fileexchange/10161-mean-shift-clustering*, 2006.

D. A. Forsyth and A. Zisserman. Reflections on shading. *Transactions on Pattern Analysis and Machine Intelligence*, 13(7):671–679, 1991.

W. T. Freeman. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368: 542–545, 1994.

D. Geisler-Moroder and A. Dür. A new ward brdf model with bounded albedo. *Eurographics Symposium on Rendering 2010*, 29:1391–1398, 2010.

A. S. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. *International Conference on Computer Vision*, 2:816–823, 2003.

A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Transactions on Pattern Analysis and Machine Intelligence*, 23:643–660, 2001.

B. Gerken. Integration und Validierung eines photometrischen Stereoverfahrens zur Oberflächenrekonstruktion. *Bachelor Thesis, Universität Paderborn*, 2012.

F. Giesen. Phong and blinn-phong normaization factors. *online http://www.farbrausch.de/~fg/stuff/phong.pdf*, 1:1–2, 2009.

D. B. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. *International Conference on Computer Vision*, 1:341–348, 2005.

D. B. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially varying brdfs from photometric stereo. *Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010.

A. Gray. *Modern Differential Geometry of Curves and Surfaces with Mathematica.* CRC Press, 1997.

A. Grumpe, S. Herbort, and C. Wöhler. 3D reconstruction of non-lambertian surfaces with non-uniform reflectance parameters by fusion of photometrically estimated surface normal data with active range scanner data. *Oldenburger 3D Tage*, 10:54–61, 2011.

J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. *International Conference on Computer Vision*, pages 1–8, 2011.

M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, pages 1–24, 2011.

U. Hahne and M. Alexa. Combining time-of-flight depth and stereo images without accurate extrinsic calibration. *Int. J. of Intelligent Systems Technology and Applications*, 5:325–333, 2008.

C. F. Hall and E. L. Hall. A nonlinear model for the spatial characteristics of the human visual system. *Systems, Man and Cybernetics, IEEE Transactions on*, 7(3):161–170, 1977.

B. Harendt. Adaptive raumzeitliche korrelation zur 3D-vermessung dynamischer szenen. In *Oldenburger 3D Tage*, 2013.

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

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

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

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

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

S. Herbort, B. Gerken, D. Schugk, and C. Wöhler. 3D range scan enhancement using image-based methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84:69–84, 2013a.

S. Herbort, D. Schugk, and C. Wöhler. 3D reconstruction of interreflection-affected surface concavities using photometric stereo. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 1–7, 2013b.

C. Hernandez and G. Vogiatzis. Self-calibrating a real-time monocular 3D facial capture system. *Fifth International Symposium on 3D Data Processing, Visualization and Transmission*, 2010.

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

C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. *International Conference on Computer Vision*, 2007.

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

T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. *International Conference on Pattern Recognition*, 2010.

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

B. K. P. Horn. *Determining Shape from Shading (Chapter 4 in The Psychology of Computer Vision).* McGraw-Hill, New York, 1975a.

B. K. P. Horn. Image intensity understanding. Technical Report 335, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1975b.

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

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

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

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

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

B. K. P. Horn and R. W. Sjoberg. Calculating the reflectance map. Technical Report 498, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1978.

B. Huhle, S. Fleck, and A. S. In. Integrating 3D time-of-flight camera data and high resolution images for 3Dtv applications. *3DTV Conference*, pages 1–4, 2007.

O. Ikeda. Shape reconstruction in photometric stereo compensating for inter-reflection and directional diffuse reflection. *3D Research*, 4(1):1–6, 2013.

K. Ikeuchi. Numerical shape from shading and occluding contours in a single view. Technical Report 566, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1980.

K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *Transactions on Pattern Analysis and Machine Intelligence*, 3(6):661–669, 1981.

K. Ikeuchi. Determining a depth map using a dual photo-metric stereo. *International Journal of Robotics Research*, 6:1–11, 1987.

A. B. Israel and T. N. E. Greville. *Generalized Inverses Theory & Applications*. Springer, 2 edition, 2003.

Y. Iwahori, H. Sugie, and N. Ishii. Reconstructing shape from shading images under point light source illumination. *International Conference on Pattern Recognition*, 1:83–87, 1990.

Z. Janko, A. Delaunoy, and E. Prados. Colour dynamic photometric stereo for textured surfaces. *Asian Conference on Computer Vision*, pages 55–66, 2010.

M. K. Johnson and E. H. Adelson. Retrographic sensing for the measurement of surface texture and shape. *Conference on Computer Vision and Pattern Recognition*, pages 1070–1077, 2009.

N. Joshi and D. J. Kriegman. Shape from varying illumination and viewpoint. *International Conference on Computer Vision*, 2007.

Y. C. Ju, M. Breuss, A. Bruhn, and S. Galliani. Shape from shading for rough surfaces: Analysis of the oren-nayar model. In *Proceedings of the British Machine Vision Conference*, pages 104.1–104.11, 2012.

J. T. Kajiya. The rendering equation. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 13:143–150, 1986.

N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, January 2002:44–57, 2002.

A. Khan. Determination and evaluation of spatially-varying reflectance parameters of surfaces with known shape. *Master Thesis, TU Dortmund University*, 2011.

H. Kim, B. Wilburn, and M. Ben-Ezra. Photometric stereo for dynamic surface orientations. *European Conference on Computer Vision*, pages 59–72, 2010.

Y. M. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Miscusik, and S. Thrun. Multi-view image and tof sensor fusion for dense 3D reconstruction. *International Conference on Computer Vision*, pages 1542–1549, 2009.

R. Kimmel and A. M. Bruckstein. Global shape from shading. *Computer Vision and Image Understanding*, 62(3):360–369, 1995.

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

M. Klaudiny, A. Hilton, and J. Edge. High-detail 3D capture of facial performance. *Fifth International Symposium on 3D Data Processing, Visualization and Transmission*, 2010.

M. C. Knauer, J. Kaminski, and G. Häusler. Phase measuring deflectometry: a new approach to measure specular free-form surfaces. *Optical Metrology in Production Engineering*, 5457:366–376, 2004.

L. Kontsevich, A. Petrov, and I. Vergelskaya. Reconstruction of shape from shading in color images. *Journal of the Optical Society of America A*, 11:1047–1052, 1994.

R. Kozera and L. Noakes. Noise reduction in photometric stereo with non-distant light sources. *Computer Vision and Graphics, Computational Imaging and Vision*, 32:103–110, 2006.

M. Kurt, L. Szirmay-Kalos, and J. Krivánek. An anisotropic brdf model for fitting and monte carlo rendering. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 44(1):1–15, 2010.

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

E. P. F. Lafortune, S.-C. Foo, K. E. Torrance, and D. P. Greenberg. Non-linear approximation of reflectance functions. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, pages 117–126, 1997.

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

R. Lange. Trainingsverfahren zur Bestimmung einer lokal adaptiven Kostenfunktion für ein aktives Oberflächenmodell zur Segmentierung dreidimensionaler Gehirnregionen. *Master Thesis, Forschungszentrum Jülich, TU Dortmund University*, 2012.

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

M. Lenoch. Quantitative comparison of surface reflectance models and evaluation of their effect on photometric stereo reconstructions. *Master Thesis, TU Dortmund University*, 2013.

M. Lenoch, S. Herbort, and C. Wöhler. Robust and accurate light source calibration using a diffuse spherical calibration object. *Oldenburger 3D Tage*, 11:1–8, 2012.

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

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

T. Leyvand, C. Meekhof, Y.-C. Wei, J. Sun, and B. Guo. Kinect identity: Technology and experience. *Computer*, 44(4):94–96, 2011.

J. Lim, J. Ho, M.-H. Yang, and D. Kriegman. Passive photometric stereo from motion. *International Conference on Computer Vision*, 2:1635–1642, 2005.

Z. Lu, Y.-W. Tai, F. Deng, M. Ben-Ezra, and M. S. Brown. A 3D imaging framework based on high-resolution photometric-stereo and low-resolution depth. *International Journal of Computer Vision*, pages 1–15, 2012.

G. Lukacs, R. Martin, and D. Marshall. Faithful least-squares fitting of spheres, cylinders, cones and tori for reliable segmentation. *European Conference on Computer Vision*, 1:671–686, 1998.

T. Machida, N. Yokoya, and H. Takemura. Surface reflectance modeling of real objects with interreflections. *International Conference on Computer Vision*, pages 170–177, 2003.

E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J.-D. Legat. A survey on industrial vision systems, applications and tools. *Image and Vision Computing*, 21(2):171–188, 2003.

S. P. Mallick, T. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. *Conference on Computer Vision and Pattern Recognition*, 1:619–626, 2005.

T. Malzbender, B. Wilburn, D. Gelb, and B. Ambrisco. Surface enhancement using real-time photometric stereo and reflectance transformation. *Proceedings of the European Symposium on Rendering*, pages 245–250, 2006.

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

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

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

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

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

A. Mian and N. Pears. 3D face recognition. In Nick Pears, Yonghuai Liu, and Peter Bunting, editors, *3D Imaging, Analysis and Applications*, pages 311–366. Springer London, 2012.

A. Michelson and E. Morley. On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, 34(203):333–345, 1887.

M. G. J. Minnaert. The reciprocity principle in lunar photometry. *American Astronomical Society*, 93: 403–410, 1941.

R. Montes and C. Urena. An overview of brdf models. Technical report, Dept. Lenguajes y Sistemas Informaticos, University of Granada, Granada, Spain, 2012.

O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria. Polarization imaging applied to 3D reconstruction of specular metallic surfaces. *Machine Vision Applications in Industrial Inspection*, 13, 2005.

S. G. Narasimhan, V. Ramesh, and S. K. Nayar. A class of photometric invariants: Separating material from shape and illumination. *International Conference on Computer Vision*, pages 1387–, 2003.

S. K. Nayar, K. Ikeuchi, and T. Kanade. Extracting shape and reflectance of lambertian, specular and hybrid surfaces. Technical Report CMU-FU-TR-88-14, The Robotics Institute, Carnegie Mellon University, 1988.

S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(1):418–431, 1990a.

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

S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *Transactions on Pattern Analysis and Machine Intelligence*, 13:611–634, 1991.

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

S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics*, 25(3):935–944, 2006.

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

L. Neumann and A. Neumann. Photosimulation: Interreflection with arbitrary reflectance models and illumination. *Computer Graphics Forum*, 8(1):21–34, 1989.

L. Neumann, A. Neumann, and L. Szirmay-Kalos. Compact metallic reflectance models. *Computer Graphics Forum*, 18:161–172, 1999.

H.-S. Ng, T.-P. Wu, and C.-K. Tang. Surface-from-gradients without discrete integrability enforcement: A gaussian kernel approach. *IEEE Transactions on Acoustics Speech and Signal ProcessingPattern Analysis and Machine Intelligence*, 32(11):2085–2099, 2010.

A. Ngan, F. Durand, and W. Matusik. Experimental analysis of brdf models. Technical Report TR2005-151, Mitsubishi Electric Research Laboratories, 2005.

M. Nick. Prepress-Management und Bildvorverarbeitung von Direktdruckapplikationen auf konischen Oberflächen. *Bachelor Thesis, KHS Dortmund, TU Dortmund University*, 2012.

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

M. Okutomi and T. Kanade. A multiple baseline stereo. *Transactions on Pattern Analysis and Machine Intelligence*, 15:353–363, 1993.

M. Oren and S. K. Nayar. Generalization of lambert's reflectance model. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, pages 239–246, 1994.

J. Ouellette. *Black Bodies and Quantum Cats: Tales from the Annals of Physics.* London: Penguin Books Ltd, 2005.

G. Papaioannou. Real-time diffuse global illumination using radiance hints. *High Performance Graphics*, 2011.

G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, and C. Chamzas. Methods for 3D digitization of cultural heritage. *Journal of Cultural Heritage*, 8(1):93–98, 2007.

T. Peng. *Algorithms and models for 3-D shape measurement using digital fringe projections.* PhD thesis, University of Maryland, Department for Mechanical Engineering, 2006.

R. Penrose. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51: 406–413, 1955.

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

B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311 –17, 1975.

E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. *International Conference on Computer Vision*, 2:826–831, 2003.

R. Ramamoorthi. *A signal processing framework for forward and inverse rendering.* PhD thesis, Stanford University, 2002.

F. Remondino and S. El-Hakim. Image-based 3D modelling: A review. *The Photogrammetric Record*, 21(115): 269–291, 2006.

J. Richarz. *Videobasierte Gestenerkennung in einer intelligenten Umgebung.* PhD thesis, TU Dortmund University, 2012.

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

M. Rioux, G. Bechthold, D. Taylor, and M. Duggan. Design of a large depth of view three-dimensional camera for robot vision. *Optical Engineering*, 26:1245–1250, 1987.

T. Ritschel, C. Dachsbacher, T. Grosch, and J. Kautz. The state of the art in interactive global illumination. *Computer Graphics Forum*, 31(1):160–188, 2012.

E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29(3):867–884, 1992.

M. Rüesing. Detektion befahrbarer Regionen für Automobile anhand von Flächenrekonstruktionen in Stereo-Bildern. *Master Thesis, Bosch GmbH, TU Dortmund University*, 2012.

M. Schaffer, M. Große, B. Harendt, and R. Kowarschik. Phasenschiebende Streifenprojektion mit 200kHz. *Oldenburger 3D Tage*, 2013.

E. Schick. Wellenlängenabhängige Reflektanzmodellierung verschiedener Oberflächenmaterialien anhand normalisierter Multispektraldaten. *Master Thesis, TU Dortmund University*, 2012.

E. Schick, S. Herbort, A. Grumpe, and C. Wöhler. Single view single light multispectral object segmentation. *International Conference on Computer Graphics, Visualization and Computer Vision*, 1:1–7, 2013.

C. Schlick. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13(3): 233–246, 1994a.

C. Schlick. A fast alternative to Phong's specular model. *Graphics Gems*, 4:363–366, 1994b.

D. Schugk. Schachbrett- und Schachbretteckendetektion in Bildern zur Integration in Kamera-Kalibrierungsalgorithmen. *Bachelor Thesis, TU Dortmund University*, 2010.

D. Schugk. Modellierung von Interreflexionen zur Integration in 3D-Oberflächenrekonstruktionsalgorithmen. *Master Thesis, TU Dortmund University*, 2012.

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

S. M. Seitz and S. Baker. Filter flow. *International Conference on Computer Vision*, 2009.

S. M. Seitz, Y. Matasushita, and K. N. Kutulakos. A theory of inverse light transport. *International Conference on Computer Vision*, pages 1440–1447, 2005.

S. A. Shafer. Using color to separate reflection components. *Color Research and Application*, 10(4):210–218, 1985.

L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. *Conference on Computer Vision and Pattern Recognition*, 1:1850–1857, 2009.

B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. *Conference on Computer Vision and Pattern Recognition*, 2010.

Y. Shirai and M. Suwa. Recognition of polyhedrons with a range finder. *International Joint Conference on Artificial intelligence*, 2:80–87, 1971.

D. Simakov, D. Frolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. *International Conference on Computer Vision*, 2:1202, 2003.

W. A. P. Smith and E. R. Hancock. Estimating facial reflectance properties using shape-from-shading. *International Journal of Computer Vision*, 86:152–170, 2010.

S. Son, H. Park, and K. H. Lee. Automated laser scanning system for reverse engineering and inspection. *International Journal of Machine Tools and Manufacture*, 42(8):889–897, 2002.

M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*, volume 3. Thomson Learning, 2008.

M. Steinhoff. Identifikation von Fahrbahnbegrenzungen in Bildinformationen basiert auf der Flussvektoranalyse und dem Inverse Perspective Mapping. *Master Thesis, Bosch GmbH, TU Dortmund University*, 2012.

M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *International Journal of Computer Vision*, 13(13):271–294, 1994.

R. Szeliski. *Computer Vision Algorithms and Applications.* online, http://szeliski.org/Book/, 2010.

P. Tan and T. Zickler. A projective framework for radiometric image analysis. *Conference on Computer Vision and Pattern Recognition*, pages 2977–2984, 2009.

P. Tan, S. Lin, L. Quan, and H.-Y. Shum. Highlight removal by illumination-constraint inpainting. *International Conference on Computer Vision*, 1:164–169, 2003.

P. Tan, S. Lin, and L. Quan. Resolution-enhanced photometric stereo. *European Conference on Computer Vision*, 3:58–71, 2006.

P. Tan, S. P. Mallick, L. Quan, D. J. Kriegman, and T. Zickler. Isotropy, reciprocity and the generalized bas-relief ambiguity. *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

P. Tan, S. Lin, and L. Quan. Subpixel photometric stereo. *Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1460–1471, 2008.

R. T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *International Conference on Computer Vision*, 1:870–877, 2003.

R. T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *Transactions on Pattern Analysis and Machine Intelligence*, 27(2):179–193, 2005.

A. Tankus. Photometric stereo under perspective projection. In *Proceedings of the 10th IEEE International Conference on Computer Vision*, pages 611–616, 2005.

A. Tankus, N. Sochen, and Y. Yeshurun. Shape-from-shading under perspective projection. *International Journal of Computer Vision*, 63(1):21–43, 2005.

D. Thomas and A. Sugimoto. Range image registration of specular objects under complex illumination. *Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT2010)*, 2010.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

S. Tominaga. Region segmentation by multispectral imaging. *Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 238–242, 2002.

K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America A*, 57(9):1105–1114, 1967.

R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, 1987.

D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popovic, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 28(5):1–11, 2009.

H. von Helmholtz. *Handbuch der Physiologischen Optik*. Optical Society of America, 1924.

B. Walter. Notes on the Ward BRDF. Technical report, Cornell Program of Computer Graphics, 2005.

T. Walzok. Konzeption und Implementierung einer vollautomatisierten Objektverfolgung mit Hilfe eines Multikamerasystems. *Master Thesis, Caterpillar Inc., TU Dortmund University*, 2011.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.

G. J. Ward. Measuring and modeling anisotropic reflection. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 26(2):265–272, 1992.

T. Weise, B. Leibe, and L. Van Gool. Fast 3D scanning with automatic motion compensation. *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

J. Westerhoff. Kamerabasierte Untersuchung von räumlich variierenden Reflektanzeigenschaften von 3D Oberflächen mit Hilfe photometrischer Methoden. *Master Thesis, TU Dortmund University*, 2013.

T. Weyrich, J. Lawrence, H. Lensch, S. Rusinkiewicz, and T. Zickler. Principles of appearance acquisition and representation. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, pages 1–70, 2008.

D. R. White, P. Saunders, S. J. Bonsey, J. van de Ven, and H. Edgar. Reflectometer for measuring the bidirectional reflectance of rough surfaces. *Applied Optics*, 37(16):3450–3454, 1998.

T. Wilhelm. Vorverarbeitung, Registrierung und Fusion von 3D-Oberflächenscans. *Bachelor Thesis, TU Dortmund University*, 2012.

G. Wiora. *Optical 3D-Metrology: Precise Shape Measurement with an extended Fringe Projection Method.* PhD thesis, Ruprechts-Karls-University, Heidelberg, Germany, 2001.

C. Wöhler and P. d'Angelo. Stereo image analysis of non-lambertian surfaces. *International Journal of Computer Vision*, 81(2):529–540, 2009.

C. Wöhler and A. Grumpe. *Integrated DEM Construction and Calibration of Hyperspectral Imagery: A Remote Sensing Perspective.* In: M. Breuss et al. (eds.), Innovations for Shape Analysis, Mathematics and Visualization, Springer-Verlag, 2013.

L. B. Wolff. Using polarization to separate reflection components. *Conference on Computer Vision and Pattern Recognition*, 1(1):363–369, 1989.

L. B. Wolff and T. E. Boult. Constraining object features using a polarization reflectance model. *Transactions on Pattern Analysis and Machine Intelligence*, 13(7):635–657, 1991.

R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

R. J. Woodham. Gradient and curvature from the photometric-stereo method, including local confidence estimation. *Journal of the Optical Society of America A*, 11(11):3050–3068, 1994.

M. Wörner. Sensorfusion zur Volumenstrommessung auf Laufband-Fördersystemen. *Master Thesis, Caterpillar Inc., TU Dortmund University*, 2012.

T.-P. Wu, J. Sun, C.-K. Tang, and H.-Y. Shum. Interactive normal reconstruction from a single image. *ACM Transactions on Graphics*, 27(5):1–9, 2008.

Q. Yang, R. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, pages 215–224, 1999.

A. L. Yuille, J. M. Coughlan, and S. Konishi. The generic viewpoint constraint resolves the generalized bas relief ambiguity. *Conference on Information Science and Systems*, 2000.

T. Zakreuskaya. Optimierung und funktionale Erweiterung eines Kamera-Kalibrierungsverfahrens. *Bachelor Thesis, TU Dortmund University*, 2011.

L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *International Conference on Computer Vision*, 1:618–626, 2003.

Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. *Conference on Computer Vision and Pattern Recognition*, pages 2472–2479, 2012.

R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. *International Conference on Computer Vision*, 1:666–673, 1999.

Z. Zheng, L. Ma, Z. Li, and Z. Chen. An extended photometric stereo algorithm for recovering specular object shape and its reflectance properties. *Computer Science and Information Systems*, 7:1–12, 2010.

W. Zhou and C. Kambhamettu. Estimation of illuminant direction and intensity of multiple light sources. *European Conference on Computer Vision*, 1:206–220, 2002.

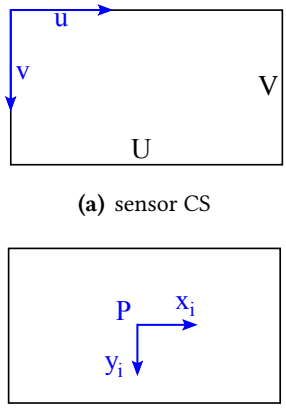
Z. Zhou and P. Tan. Ring-light photometric stereo. *European Conference on Computer Vision*, pages 1–14, 2010.

J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

T. Zickler, S. P. Mallick, D. J. Kand, and P. N. Belhumeur. Color subspaces as photometric invariants. *International Journal of Computer Vision*, 79:13–30, 2008.

A Detailed description of typical coordinate systems applied for computer vision

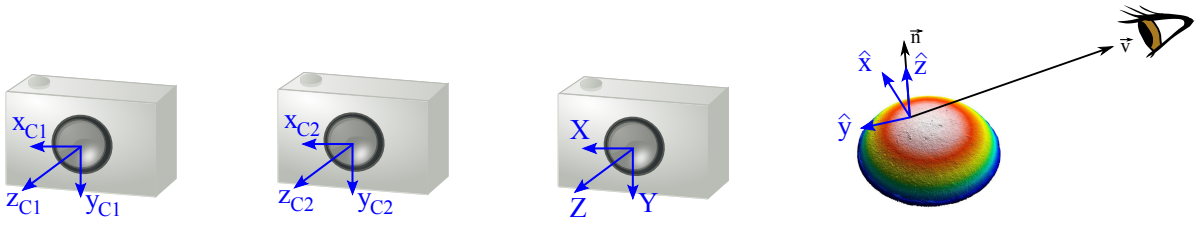
The sensor coordinate system shown in Fig. A.1(a) is the only discretely parameterized CS, and it describes a real world image that is projected onto and sampled by the pixel grid of the sensor. Its origin lies in the upper left sensor corner, and the image $I(u, v)$ is discretely sampled in pixels (px) along the axes $u \in [1...U] \subset \mathbb{N}$ and $v \in [1...V] \subset \mathbb{N}$. Both axes are aligned with the major symmetry axes of the sensor. The image CS (see Fig. A.1(b)) describes the physical image that exists in an infinitesimally small distance before the light strikes the sensor. Its origin is located in the principle point P, which is by definition located where the optical axis (perpendicularly) intersects the image plane. The image $I(x_i, y_i) \in \mathbb{R}^+$ on the sensor is parameterized by metric units $(x_i, y_i) \in \mathbb{R}^{1\times 2}$ relative to the principal point P, and the axes point along the major symmetry axes of the image/sensor plane. The camera CS (x, y, z) depicted in



(a) sensor CS

(b) image CS

Figure A.1.: Sensor and image coordinate systems



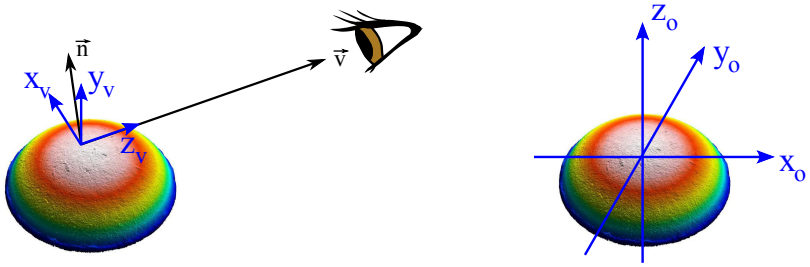
(a) camera CS **(b)** world CS **(c)** local CS

Figure A.2.: Camera, world, and local coordinate systems

Fig. A.2(a) is described in detail in Section 1.4. Note, that it is possible to have several camera CS in a scene, based on the number of available cameras. For the definition of a globally valid coordinate system, it is common to chose a designated world camera system (X, Y, Z), which is typically identical to one of the camera coordinate systems (see Fig. A.2(b)).

Another coordinate system $(\hat{x}, \hat{y}, \hat{z})$ exists locally for each point on the surface. The $\hat{z}$-axis of that CS is aligned with the locally varying surface normal $\mathbf{n} = (n_x, n_y, n_z)$ as shown in Fig. A.2(c). While it is required for $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ to lie in the plane defined by the surface point and the normal $\mathbf{n} = \hat{\mathbf{z}}$, this does not define the rotation of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ in the $\hat{\mathbf{x}}$-$\hat{\mathbf{y}}$-plane. However, the correct rotation needs to be known since it influences the values obtained by e.g. the Ward model for anisotropic reflectance (see Section 2.4.4) as it requires the $\hat{\mathbf{x}}$-axis to be aligned with the predominant surface facet direction, which defines the main anisotropy direction. The Lafortune model discussed in Section 2.4.2 uses an eigenvalue decomposition, which then yields the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$-axes to be aligned with these major axes of anisotropy. Another possible approach is to determine the rotation as an additional parameter during the estimation of the surface reflectance parameters (see Section 2.5).

In the fields of remote sensing for e.g. lunar surface examinations, it is more common to use the viewer



(a) viewer CS **(b)** object CS

Figure A.3.: Viewer and object coordinate systems

coordinates system shown in Fig. A.3(a). That system is similar to the camera CS, but rotated along the camera x-axis by $180°$. This implies that a description of the non-unit-length surface normals $\mathbf{n}$ using $p = \frac{\partial z}{\partial x} = \partial_x z = z_x$ and $q = \frac{\partial z}{\partial y} = \partial_y z = z_y$, yields $\tilde{\mathbf{n}} = (-p, -q, 1)^T$. A coordinate system used oftentimes in rendering programs is the local object CS (x_o, y_o, z_o), whose origin is located in the object's center of gravity as shown in Fig. A.3(b).

B Object appearance

This chapter is adapted and/or adopted from [Herbort and Wöhler, 2011]

The following sections give more detailed information for certain aspects regarding the appearance of surfaces. These aspects are a comparison of distant and non-distant light sources in Section B.1, a description of generalized reflectance functions in Section B.2, mathematic derivations for physically plausible BRDFs in Section B.3, details for further BRDFs (Minnaert, Neumann-Neumann, Beckmann-Spizzichino, Torrance-Sparrow, He-Torrance) in Section B.4, and details for anisotropic BRDF models in Section B.5.

B.1 Influence of distant and non-distant lightsources

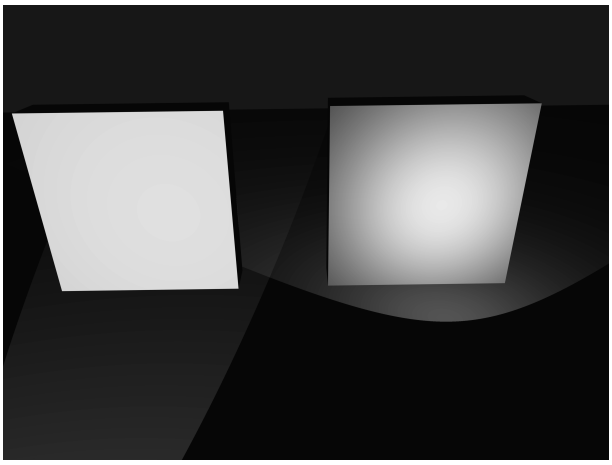


Figure B.1.: Illustration of the light source and illumination component influence on the appearance of an object. The distance of a light source changes locally the effective incident light intensity, which is clearly conceivable in the image.

A light source has a certain intensity, position, directional characteristic, and spectral characteristic. In the field of computer vision, it is a common simplification to assume light sources to be distant, such that any point on the surface is illuminated from the same light direction $\mathbf{l}$. Non-distant light sources as the LEDs used for the experiments for this thesis provide locally varying illumination directions. Fig. B.1 given an example for both cases.

B.2 Generalized reflectance distribution functions

The description of additional reflection effects requires an increased dimensionality compared to the four-dimensional BRDF (Eq. 2.4). Typical extensions are wavelength-dependency, spatial variability, and subsurface scattering [Nicodemus et al., 1977].
Wavelength-dependency is described by the spectral BRDF

$$\text{spectral BRDF} : f_r(\vartheta_{in}, \phi_{in}, \vartheta_{obs}, \phi_{obs}; \lambda). \tag{B.1}$$

The Bidirectional Spatially Varying Reflectance Distribution Function (BSVRDF) provides locally varying reflectance properties and thus depends on the surface point with coordinates (x, y) that it is assigned to.

$$\text{BSVRDF} : f_r(x, y, \vartheta_{in}, \phi_{in}, \vartheta_{obs}, \phi_{obs}) \tag{B.2}$$

As mentioned by Szeliski [2010] and surveyed by Weyrich et al. [2008], it is a largely generalized case for reflectance description to use the Bidirectional Subsurface-Scattering Reflectance-Distribution Function (BSSRDF). The BSSRDF describes the amount of light that is observed additionally depending on incident irradiance point (x_{in}, y_{in}) and outgoing radiance point (x_{obs}, y_{obs}):

$$\text{BSSRDF} : S(x_{in}, y_{in}, \vartheta_{in}, \phi_{in}, x_{obs}, y_{obs}, \vartheta_{obs}, \phi_{obs}) \tag{B.3}$$

Although this models a complex environment, the approach is far from general and further effects can be considered, which include fluorescence or phosphorescence [Weyrich et al., 2008].

B.3 Mathematic derivations

The following sections regard the problem of physical plausibility for BRDFs and show the derivations for the Lambert, Phong, and Blinn BRDF explicitly.

B.3.1 Physical plausibility of BRDF models

Enforcing physical plausibility requires checking for (1) Helmholtz reciprocity [von Helmholtz, 1924] and (2) energy conservation, e.g. [Lewis, 1994]. This is shown explicitly for the Lambertian, Phong, and Blinn BRDF in the following.

B.3.1.1 Lambert

(1) A constant BRDF (as it is the case for Lembertian reflectance) obviously obeys Helmholtz reciprocity.
(2) Starting at the condition for energy conserving BRDFs

$$\int_{\Omega} f(\vartheta_{in}, \phi_{in}; \vartheta_{obs}, \phi_{obs}) \cos\vartheta_{obs} d\Omega_{obs} \leq 1, \tag{B.4}$$

we have $f(\vartheta_{in}, \phi_{in}; \vartheta_{obs}, \phi_{obs}) = const. = f$ by definition, which yields

$$f \cdot \int_{\Omega} \cos\vartheta_{obs} d\Omega_{obs} \leq 1. \tag{B.5}$$

Integration over the upper hemisphere by substituting $d\Omega_{obs} = \sin\vartheta_{obs}\, d\vartheta_{obs}\, d\phi_{obs}$ gives

$$f \cdot \int_{\vartheta=0}^{\frac{\pi}{2}} \int_{\phi=0}^{2\pi} \cos\vartheta_{obs} \sin\vartheta_{obs}\, d\vartheta_{obs}\, d\phi_{obs} \leq 1 \tag{B.6}$$

$$\Leftrightarrow \quad f \cdot 2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \cos\vartheta_{obs} \sin\vartheta_{obs}\, d\vartheta_{obs} \leq 1 \tag{B.7}$$

$$\Leftrightarrow \quad f \cdot 2\pi \left[-\frac{1}{2}\cos^2\vartheta_{obs}\right]_0^{\frac{\pi}{2}} \leq 1 \tag{B.8}$$

$$\Leftrightarrow \quad f \cdot 2\pi \left[\frac{1}{2}\right] \leq 1 \tag{B.9}$$

$$\Leftrightarrow \quad f \cdot \leq \frac{1}{\pi}. \tag{B.10}$$

This requires the BRDF to be of the form

$$f = \frac{\varrho}{\pi} \qquad \text{with } \varrho \in [0...1] \subset \mathbb{R}, \tag{B.11}$$

and ϱ then simply denotes to the surface albedo.

B.3.1.2 Modified Phong

(1) Rewriting $\langle \mathbf{r} \cdot \mathbf{v} \rangle$ yields

$$\begin{aligned} \langle \mathbf{r} \cdot \mathbf{v} \rangle &= \cos\vartheta_r && \text{(B.12)} \\ &= 2\cos\vartheta_{in}\cos\vartheta_{obs} - \cos\alpha && \text{(B.13)} \\ &= 2\langle \mathbf{n} \cdot \mathbf{l} \rangle \cdot \langle \mathbf{n} \cdot \mathbf{v} \rangle - \langle \mathbf{l} \cdot \mathbf{v} \rangle && \text{(B.14)} \end{aligned}$$

and immediately proves reciprocity. Eq. B.13 has been adopted from [d'Angelo and Wöhler, 2008].
(2) Starting at the condition for energy conserving BRDFs

$$\int_{\Omega} f(\vartheta_{in}, \phi_{in}; \vartheta_{obs}, \phi_{obs}) \cos\vartheta_{obs} d\Omega_{obs} \leq 1, \tag{B.15}$$

we have $f(\vartheta_{in}, \phi_{in}; \vartheta_{obs}, \phi_{obs}) = \tilde{k}_d + \tilde{k}_s \langle \mathbf{r} \cdot \mathbf{v} \rangle^m$ by definition, which yields

$$\int_{\Omega} (\tilde{k}_d + \tilde{k}_s \langle \mathbf{r} \cdot \mathbf{v} \rangle^m) \cos\vartheta_{obs} d\Omega_{obs} \leq 1. \tag{B.16}$$

using $\langle \mathbf{r} \cdot \mathbf{v} \rangle^m = \cos \vartheta_r$ and (by assuming the maximum energy being reflected if $\mathbf{l} = \mathbf{n}$)[1] $\vartheta_r = \vartheta_{obs}$ under integration over the upper hemisphere ($d\Omega_{obs} = \sin \vartheta_{obs} \, d\vartheta_{obs} \, d\phi_{obs}$), we obtain

$$\int_{\vartheta=0}^{\frac{\pi}{2}} \int_{\phi=0}^{2\pi} \left(\tilde{k}_d + \tilde{k}_s \cos^m \vartheta_{obs} \cos \vartheta_{obs} \right) \sin \vartheta_{obs} \, d\vartheta_{obs} \, d\phi_{obs} \leq 1 \tag{B.17}$$

$$2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \left(\tilde{k}_d + \tilde{k}_s \cos^m \vartheta_{obs} \cos \vartheta_{obs} \right) \sin \vartheta_{obs} \, d\vartheta_{obs} \, d\phi_{obs} \leq 1 \tag{B.18}$$

$$\pi \tilde{k}_d + 2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \left(\tilde{k}_s \cos^m \vartheta_{obs} \cos \vartheta_{obs} \right) \sin \vartheta_{obs} \, d\vartheta_{obs} \, d\phi_{obs} \leq 1 \tag{B.19}$$

$$\pi \tilde{k}_d + 2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \left(\tilde{k}_s \cos^{m+1} \vartheta_{obs} \right) \sin \vartheta_{obs} \, d\vartheta_{obs} \, d\phi_{obs} \leq 1 \tag{B.20}$$

$$\pi \tilde{k}_d + 2\pi \left[-\frac{\tilde{k}_s}{m+2} \cos^{m+2} \vartheta_{obs} \right]_0^{\frac{\pi}{2}} \leq 1 \tag{B.21}$$

$$\pi \tilde{k}_d + \tilde{k}_s \frac{2\pi}{m+2} \leq 1 \tag{B.22}$$

Using $\tilde{k}_d = \frac{k_d}{\pi}$ and $\tilde{k}_s = k_s \frac{m+2}{2\pi}$ yields

$$f = \frac{k_d}{\pi} + k_s \frac{m+2}{2\pi} \cos \vartheta_r, \tag{B.23}$$

which is energy conserving if

$$k_d + k_s \leq 1 \tag{B.24}$$

The derivation can be obtained partially from e.g. [Lafortune and Willems, 1994].

B.3.1.3 Modified Blinn

(1) Rewriting $\langle \mathbf{h} \cdot \mathbf{n} \rangle$ yields

$$\langle \mathbf{h} \cdot \mathbf{n} \rangle = \left\langle \frac{\mathbf{v} + \mathbf{l}}{|\mathbf{v} + \mathbf{l}|} \cdot \mathbf{n} \right\rangle \tag{B.25}$$

$$= \left\langle \frac{\mathbf{l} + \mathbf{v}}{|\mathbf{l} + \mathbf{v}|} \cdot \mathbf{n} \right\rangle \tag{B.26}$$

and immediately proves reciprocity.

(2) Even though the following assumes $\mathbf{n}, \mathbf{l}$ and $\mathbf{v}$ to lie in the same plane, this goes without loss of generality. In that case, the Blinn angle ϑ_h is exactly half the observation angle ϑ_{obs}, i.e. $\vartheta_h = \vartheta_{obs}/2$. Additionally, the geometry is chosen such that the total reflected energy is a maximum, which is assumed to be the case[2]

[1] This is not proven here

[2] This is not proven here

for $\mathbf{l} = \mathbf{n}$. Starting at the integral for the total hemispherical reflectance

$$\int_{\Omega} f(\vartheta_{in}, \phi_{in}; \vartheta_{obs}, \phi_{obs}) \cos\vartheta_{obs} d\Omega_{obs} \qquad |f = \langle \mathbf{h} \cdot \mathbf{n} \rangle^m = \cos^m \vartheta_h \tag{B.27}$$

$$= \int_{\Omega} \cos^m \vartheta_h \cos\vartheta_{obs} d\Omega_{obs} \qquad |\cos^m \vartheta_h = \cos^m \frac{\vartheta_{obs}}{2} \tag{B.28}$$

$$= \int_{\Omega} \cos^m \frac{\vartheta_{obs}}{2} \cos\vartheta_{obs} d\Omega_{obs} \qquad |d\Omega_{obs} = \sin\vartheta_{obs}\, d\vartheta_{obs}\, d\phi_{obs} \tag{B.29}$$

$$= \int_{\vartheta=0}^{\frac{\pi}{2}} \int_{\phi=0}^{2\pi} \cos^m \frac{\vartheta_{obs}}{2} \cos\vartheta_{obs} \sin\vartheta_{obs}\, d\vartheta_{obs}\, d\phi_{obs} \qquad |\text{resolve } d\phi_{obs} \tag{B.30}$$

$$= 2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \cos^m \frac{\vartheta_{obs}}{2} \cos\vartheta_{obs}\ \sin\vartheta_{obs}\, d\vartheta_{obs} \qquad |\cos\frac{\vartheta_{obs}}{2} = \sqrt{\frac{1}{2} + \frac{\cos\vartheta_{obs}}{2}} \tag{B.31}$$

$$= 2\pi \int_{\vartheta=0}^{\frac{\pi}{2}} \sqrt{\frac{1}{2} + \frac{\cos\vartheta_{obs}}{2}}^m \cos\vartheta_{obs}\ \sin\vartheta_{obs}\, d\vartheta_{obs} \quad |z = \cos\vartheta_{obs}, \frac{dz}{d\vartheta_{obs}} = -\sin\vartheta_{obs} \tag{B.32}$$

$$= -2\pi \int_{z=0}^{1} \sqrt{\frac{1}{2} + \frac{z}{2}}^m z\, dz, \tag{B.33}$$

which can be integrated by parts using $f = z$ and $\frac{dg}{dz} = \sqrt{\frac{1}{2} + \frac{z}{2}}^m$:

$$= -2\pi \left(\left[\frac{4}{m+2} z \sqrt{\frac{1}{2} + \frac{z}{2}}^{(m+2)} \right]_{z=1}^{0} + \frac{4}{m+2} \int_{z=1}^{0} \sqrt{\frac{1}{2} + \frac{z}{2}}^{(m+2)} dz \right) \tag{B.34}$$

$$= -2\pi \left[\frac{4}{m+2} z \sqrt{\frac{1}{2} + \frac{z}{2}}^{(m+2)} \right]_{z=1}^{0} - 2\pi \left[\frac{4}{m+2} \sqrt{\frac{1}{2} + \frac{z}{2}}^{(m+4)} \right]_{z=1}^{0}. \tag{B.35}$$

Trivial mathematical logic then yields

$$= 8\pi \frac{2^{-\frac{n}{2}} + n}{(n+2)(n+4)}. \tag{B.36}$$

The derivation can be found as well in the (unpublished) online document [Giesen, 2009].

B.4 Further details for less common reflectance models

B.4.1 Minnaert

The model by Minnaert [1941] originally aimed at Lunar observations. The BRDF is given as

$$f_M = b\left(\langle \mathbf{n} \cdot \mathbf{l} \rangle \langle \mathbf{n} \cdot \mathbf{v} \rangle\right)^{(k-1)} \tag{B.37}$$

with the parameters $b \in \mathbb{R}^+$ and $k \in \mathbb{R}^+$. The results are very similar to Neumann-Neumann BRDFs as explained by [Lewis, 1994]. In comparison to Lambert [1760] and Oren and Nayar [1994], Minnaert models nonlinearly decreasing diffuse reflectance near grazing angles (diffuse fall-off) as they are described in detail by Dorsey et al. [2008]. The Minnaert model as such is contained in the generalized cosine lobe model by Lafortune et al. [1997] when using a specialized configuration (see Section 2.4.2).

B.4.2 Neumann-Neumann

The goal of Neumann and Neumann [1989] was the fast computation of reflectance values on computers with low computational power (compared to contemporary computers). Neumann and Neumann [1989] thus developed a separable BRDF model, whose components can be computed independently. Their proposed BRDF is

$$f_{NN} = c \exp\left\{-s\left(\frac{1}{\langle \mathbf{n} \cdot \mathbf{l} \rangle} + \frac{1}{\langle \mathbf{n} \cdot \mathbf{v} \rangle}\right)\right\} \tag{B.38}$$

with two parameters $c \in \mathbb{R}^+$ and $s \in \mathbb{R}^+$.

B.4.3 Beckmann-Spizzichino and Torrance-Sparrow

There are two BRDF models that have already been developed in 1963 and 1967, which aim on highly exact surface reflectance descriptions. These are the wave optics based model by Beckmann and Spizzichino [1963] and the geometric optics based model by Torrance and Sparrow [1967]. Due to their complexity, they are rarely used for 3D reconstructions and this description is thus short and focuses on their principles rather than their exact mathematical formulae. Both models have been reviewed, evaluated, and compared thoroughly by Nayar et al. [1991], who give an especially useful insight into the nature and properties of specular reflections as shown in the following.

The Beckmann and Spizzichino (BS) model is based on Maxwell's equations, which describe how a light wave interacts with the material that it falls upon. An interpretation of the BS model shows, that there is a specular lobe and a specular spike component emerging, which jointly contribute to the reflection. This is the physical justification of the Lambert/lobe/spike superposition by Nayar et al. [1988].

The derivation of the BS model makes the following assumptions [Nayar et al., 1991]:

(1) The incident wavelength λ is large compared to the surface roughness structures.
(2) The incident light is perpendicularly polarized.
(3) The incident wave is a plane wave.

(4) The surface roughness can be modeled by normally distributed height variations.
(5) The surface is a perfect conductor.
(6) No self-shadowing or self-masking occurs.
(7) No interreflections occur.

The work by Torrance and Sparrow [1967] regards geometrical optics for modeling specular reflections that originate from a rough surface. The model underlies the following assumptions [Nayar et al., 1991]:

(1) The surface roughness σ can be modeled by microscopic surface facets.
(2) The surface facet size, i.e. σ, is much larger than the incident light wavelength λ.
(3) Since $\lambda \ll \sigma$, there is no specular spike occurring.
(4) The surface facet slopes are distributed normally.
(5) Fresnel effects are modeled by the Fresnel factor.
(6) The surface facets form v-concavities, which allow modeling self-shadowing and self-masking.

An evaluation and comparison of both models yields the following main insights [Nayar et al., 1991]:

(1) Both models show off-specular lobe peaks.
(2) Specular spikes occur only significantly if $\frac{\sigma}{\lambda} < 1.5$.
(3) The specular lobe dominates if $\lambda \ll \sigma$.
(4) The Torrance-Sparrow model is a good approximation of the specular lobe described by Beckmann-Spizzichino.
(5) The specular spike component is not described well by the Torrance-Sparrow model.

Further details can be obtained from [Beckmann and Spizzichino, 1963; Torrance and Sparrow, 1967; Nayar et al., 1991]. Georghiades [2003] used the Torrance-Sparrow model for uncalibrated Photometric Stereo surface reconstructions.

B.4.4 He-Torrance

The BRDF model f_{HT} by He et al. [1991] consists of three components,

$$f_{HT} = f_d + f_{dd} + f_s. \tag{B.39}$$

These describe an ideal diffuse (f_d), directional diffuse (f_{dd}) and ideal specular (f_s) reflectance behavior. The main advantage is, that those components contain expressions for interference and diffraction effects that occur on surfaces when electromagnetic light waves are considered. The ideal diffuse behavior is assumed to be a wavelength-dependent constant

$$f_d = a(\lambda), \tag{B.40}$$

which comprises the Lambertian case where $a(\lambda) = \frac{\rho(\lambda)}{\pi}$. The directional diffuse behavior is modeled via

$$f_{dd} = \frac{F \cdot S}{\cos\vartheta_{in} \cos\vartheta_{obs}} \frac{\tau^2}{16\pi} \sum_{k=1}^{\infty} \frac{g^k \exp\{-g\}}{k! \cdot k} \exp\left\{-\frac{v^2\tau^2}{4k}\right\}, \tag{B.41}$$

with

$$g = \left[\frac{2\pi\sigma}{\lambda}(\cos\vartheta_{in} + \cos\vartheta_{obs})\right]^2, \tag{B.42}$$

wave vector change v, surface roughness σ, shadowing function S, and Fresnel reflection coefficient F. The ideal specular component is given as

$$f_s = \frac{|F|^2 \, S \exp\{-g\}}{\cos\vartheta_{in}}\Delta \tag{B.43}$$

and Δ denotes a binary function $\Delta \in \{0, 1\} \subset \mathbb{N}$ that determines if the observed light ray is located within the specular cone or not. Since the model as such is very elaborate and complex, it is feasible to refer to the original publication [He et al., 1991] for details. Note, that the model comprises polarized incident light, which is not the case for other models known from the literature. Additionally, He et al. [1991] use an improved shadowing function compared to e.g. Cook and Torrance [1981]. Due to the model complexity, He et al. [1991] require an extensive computation time, and Neumann et al. [1999] note that it is 320 times higher than the computation time of the model by Phong [1975].

B.5 Anisotropic reflectance models

B.5.1 Ward

The Ward BRDF is of the form [Ward, 1992]:

$$f_W \quad = \quad k_s \frac{D_W(\mathbf{h})}{4K_W} \tag{B.44}$$

with

$$D_W(\mathbf{h}) = \frac{1}{\pi\alpha_x\alpha_y} \exp\left\{-\tan^2\vartheta_h \left(\left(\frac{\cos\phi_h}{\alpha_x}\right)^2 + \left(\frac{\sin\phi_h}{\alpha_y}\right)^2\right)\right\} \tag{B.45}$$

and

$$K_W \quad = \quad \sqrt{\langle\mathbf{n}\cdot\mathbf{l}\rangle \, \langle\mathbf{n}\cdot\mathbf{v}\rangle}. \tag{B.46}$$

Note the explicit dependency of $D_W(\mathbf{h})$ on the azimuth angle ϕ_h of the half vector $\mathbf{h}$. It has been reported that the BRDF itself is not energy conserving [Dür, 2006], but it has recently been achieved by Geisler-Moroder and Dür [2010], who apply the normalization

$$K_G = \langle\mathbf{l}\cdot\mathbf{h}\rangle^2 \, \langle\mathbf{h}\cdot\mathbf{n}\rangle^4 . \tag{B.47}$$

While the local x- and y-axes ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$) may be defined according to the coordinate system definitions in Section 1.4, the local $\hat{\mathbf{x}}$ axis needs to be aligned with the main directions of anisotropy for the Ward model to work correctly (see e.g. Walter [2005]). Since this has an implication for the correct fitting of BRDF measurements to the model, it makes sense to introduce an additional model parameter ϕ_0 that defines

the rotation (phase) of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ with respect to $\hat{\mathbf{x}}_0 = (0, -1, 0)^T \times \hat{\mathbf{z}}$. For this, Eq. B.44 merely requires the replacement of ϕ_h by $\phi_h^{'} = \phi_h + \phi_0$. A different approach lies in the singular value decomposition applied by Lafortune et al. [1997], see Section 2.4.2, which rotates the axes to be aligned with the principle axes of anisotropy.
The Ward model has been applied by e.g. Yu et al. [1999]; Chung and Jia [2008]; Goldman et al. [2010]; Zheng et al. [2010] for 3D surface reconstructions.

B.5.2 Ashikhmin-Shirley

The contribution of the model by Ashikhmin and Shirley [2000] lies in a combination of a diffuse BRDF component with a specular and anisotropic BRDF component that furthermore models the Fresnel effect. The specialty now is, that the diffuse component is modeled such that its strength decreases as large polar incidence angles ϑ_{in} are approached. This keeps the law of energy conservation intact even if the specular strength increases according to Schlick's Fresnel term $F_S(\langle \mathbf{v} \cdot \mathbf{h} \rangle)$ [Schlick, 1994a].

$$f_{AS} = f_{AS-d} + f_{AS-s} \tag{B.48}$$

$$f_{AS-s} = \frac{\sqrt{(\alpha+1)(\beta+1)}}{8\pi} \frac{\langle \mathbf{n} \cdot \mathbf{h} \rangle^{\alpha \cos^2 \phi + \beta \sin^2 \phi}}{\langle \mathbf{h} \cdot \mathbf{l} \rangle \max(\langle \mathbf{n} \cdot \mathbf{v} \rangle, \langle \mathbf{n} \cdot \mathbf{l} \rangle)} F_s(\langle \mathbf{h} \cdot \mathbf{l} \rangle) \tag{B.49}$$

$$f_{AS-d} = \frac{28 k_d}{23\pi}(1 - k_s) \left(1 - \left(\frac{1 - \langle \mathbf{n} \cdot \mathbf{l} \rangle}{2}\right)^5\right) \left(1 - \left(\frac{1 - \langle \mathbf{n} \cdot \mathbf{v} \rangle}{2}\right)^5\right) \tag{B.50}$$

Note that Eq. B.49 uses the microfacet distribution normalization while normalizing the heightfield distribution according to Pharr and Humphreys [2010, p. 458] requires the replacement of $\sqrt{(\alpha+1)(\beta+1)}$ by $\sqrt{(\alpha+2)(\beta+2)}$. In contrast to Ward [1992], who uses a modified Beckmann distribution as a description for the microfacet distribution ($\exp\{...\}$), the Ashikhmin-Shirley model uses a parameterization similar to Blinn ($\langle \mathbf{n} \cdot \mathbf{h} \rangle$).

B.5.3 Kurt

The BRDF by Kurt et al. [2010] is of the form

$$f_K = k_s \frac{D_K(\mathbf{h})\, F_S(\langle \mathbf{v} \cdot \mathbf{h} \rangle)}{4 K_K} \tag{B.51}$$

with

$$D_K(\mathbf{h}) = \frac{1}{\pi \alpha_x \alpha_y} \exp\left\{ -\tan^2 \vartheta_h \left(\left(\frac{\cos \phi_h}{\alpha_x}\right)^2 + \left(\frac{\sin \phi_h}{\alpha_y}\right)^2 \right) \right\}, \tag{B.52}$$

the normalization factor

$$K_K = \langle \mathbf{n} \cdot \mathbf{h} \rangle^4 \langle \mathbf{v} \cdot \mathbf{h} \rangle \ (\langle \mathbf{l} \cdot \mathbf{n} \rangle \langle \mathbf{v} \cdot \mathbf{n} \rangle)^{m_K} \tag{B.53}$$

and Schlick's approximation [Schlick, 1994a] of the Fresnel term

$$F_S(\langle \mathbf{v} \cdot \mathbf{h} \rangle) = F_0 + (1 - F_0)(1 - \langle \mathbf{v} \cdot \mathbf{h} \rangle)^5, \tag{B.54}$$

where $F_0 \in [0...1] \subset \mathbb{R}$ denotes the value of the Fresnel term at $\theta_h = 0°$.
Note that the model essentially corresponds to the Ward-modification by Geisler-Moroder and Dür [2010] with an appended Fresnel term and a slightly modified normalization factor

$$K_K = K_G \frac{(\langle \mathbf{l} \cdot \mathbf{n} \rangle \langle \mathbf{v} \cdot \mathbf{n} \rangle)^{m_K}}{\langle \mathbf{v} \cdot \mathbf{h} \rangle}. \tag{B.55}$$

C Calibration

The following sections regard details for the sensor assessment of the applied monochromatic camera (AVT Pike) and for the calibration of the optionally available precision rotation stages (micos DT 65, PRS 110).

C.1 Camera

C.1.1 Dark frame compensation and hot pixels examination

The examination yielded the results shown in Table C.1 for 17 characteristic exposure times ($2^{0:1:16}\ ms = 1, 2, 4, 8, ..., 65536ms$).

Results (1) Over all exposure times, the examination of 16 Bit images showed an average brightness of 20.03 ± 1.46 before compensation and 2.30 ± 0.18 after compensation with exposure time dependent darkframes. The corresponding $SSIM_{10000,8}$ values[1] are 0.2334 ± 0.0152 before compensation and 0.9324 ± 0.0161 after compensation. The sensor noise does thus not increase with increasing exposure times, which can be concluded from the average pixel values $\bar{I}$ and the $SSIM$ values before compensation. Detailed results are listed in Table C.1. **(2)** A significant connection can be observed regarding the spatial variation as depicted in Fig. 4.3.

Conclusion For correct compensation, an exposure-dependent dark frame compensation thus proves to be most feasible. To obtain a suitable set of darkframes, 161 exposure time samples ($2^{0:0.1:16}\ ms = 1, 2, 4, 8, ..., 65536ms$) have been acquired. For compensation of an image with exposure time t_E, the darkframe with the most similar exposure time among these samples is used.

[1] The SSIM values have been computed using a virtual grayvalue maximum of 10000, which shifts the $SSIM$ to more significant values. Application of the typical $SSIM_{65535,8}$ values yields the same conclusions, but produces $SSIM$ values that are less meaningful when looking at them intuitively (e.g. 0.9993 without compensation, 0.99999991 with compensation). The reference frame for the $SSIM$ simply consists of an image with the same size and containing only zeros.

Exposure time $[ms]$	mean intensity $\bar{I}$ before	mean intensity $\bar{I}$ after	$SSIM_{1000,8}$ before	$SSIM_{1000,8}$ after	# hot pixels
1.0	18.87	2.27	0.2291	0.9381	0
2.0	19.00	2.31	0.2282	0.9330	0
4.0	19.04	2.87	0.2259	0.9060	0
8.0	18.93	2.27	0.2280	0.9381	0
16.0	19.02	2.28	0.2259	0.9376	0
32.0	19.03	2.21	0.2257	0.9410	0
64.0	19.11	2.24	0.2244	0.9397	0
128.0	19.08	2.26	0.2255	0.9388	0
256.0	19.33	2.34	0.2223	0.9345	0
512.0	19.27	2.26	0.2270	0.9381	0
1024.0	19.72	2.63	0.2249	0.9192	2
2048.0	20.19	2.29	0.2305	0.9366	3
4096.0	21.46	2.37	0.2293	0.9329	27
8192.0	21.28	2.22	0.2417	0.9404	152
16384.0	20.78	2.15	0.2427	0.9435	311
32768.0	21.67	2.03	0.2552	0.9485	361
65536.0	24.50	2.17	0.2848	0.8798	441
mean	20.03	2.30	0.2334	0.9324	-
standard deviation	1.46	0.18	0.0152	0.0161	-

Table C.1.: Darkframe compensation results for AVT pike: The SSIM has been computed over a windows of $8\ px \times 8\ px$ and a maximum pixel value of 1000 has manually been selected to provide SSIM values that are more suitable for comparison. The conclusion remains the same for a (correct) pixel maximum of 65535, but the differences are less noticeable. For the SSIM, a zero-matrix has been applied as the reference data.

C.2 Rotation axis calibration

Precision rotation stages are very advantageous in computer vision settings, since they allow a variation of the object orientation with an accuracy of (in this case) $\pm 0.01°$. In the following, the cases of a non-excentric and an excentric rotation axis are explained. Knowing the orientation and position of a rotation axis e.g. allows a computer vision algorithm to fuse 3D point clouds from different object orientations without tedious point cloud registration.

C.2.1 Rotation axis calibration of an axis parallel to the rotated plane

In the case of a rotation axis being parallel to the rotated plane, the rotation axis can be determined using the cross product such that $\mathbf{a} = \frac{\mathbf{n}_1 \times \mathbf{n}_2}{||\mathbf{n}_1 \times \mathbf{n}_2||_2}$. For this, the normal directions are determined from a least squares fit of a plane to the point cloud data to each of the orientations with outliers suppression[2].

The rotation axis offset (i.e. the position of the rotation axis in the camera coordinate system) can initially be restricted to a single degree of freedom by the following consideration (see Fig. C.1): The rotation axis is parallel to the line of intersection between the initial plane and the rotated plane. This plane is thus spanned by the rotation axis direction itself and by $\mathbf{n}_3 = -\frac{\mathbf{n}_1 + \mathbf{n}_2}{2}$, as shown by the direction of the cyan

[2]The detailed procedure is: (1) acquire 3D point cloud data, (2) determine point cloud values that correspond to points on the bright rectangle of the calibration plate, (3) determine a least squares fit through that plane, (4) iteratively (N iterations with successively decreasing threshold D) remove all measurements from the point cloud that have a distance of more than D $[mm]$ from the fitted plane and determine a new fitted plane to the remaining data, (5) determine normal direction from a final least squares fit of a plane to the remaining data points.

line in Fig. C.1. Therefore, the offset $\mathbf{o}$ of the rotation axis is given by

$$\mathbf{o} = P_0 + \alpha \cdot \mathbf{n}_3 + \beta \cdot \mathbf{a} \tag{C.1}$$

where P_0 is an arbitrary point on a line of intersection of both planes, α determines the displacement of the rotation axis with respect to P_0 and β is an arbitrary factor, which determines the translation of the axis origin in the direction of the axis itself.

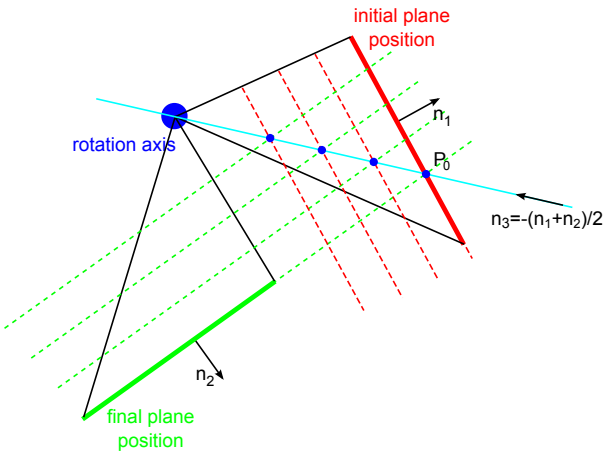


Figure C.1.: Reduction of the degrees of freedom of the rotation axis offset: The axis must lie within a plane that contains all cut lines of the rotated planes which result from equally moving these along their (negative) normal direction.

Practically, α is determined using a gradient descent implementation, which optimizes the sum of the distances of corresponding rectangle corners between corners of the rotated plane 1 and the position of the corners of plane 2. The corresponding corners have automatically been determined by the algorithm, the results (numbers “1”, “2”, “3”, “4”) are shown in Fig. C.2(a). Without correspondences, the axis location can only be determined u to the factor α. The final parameter, β, is finally adjusted such that one component of $\mathbf{p}$ is zero.

C.2.2 Rotation axis calibration of an excentric axis

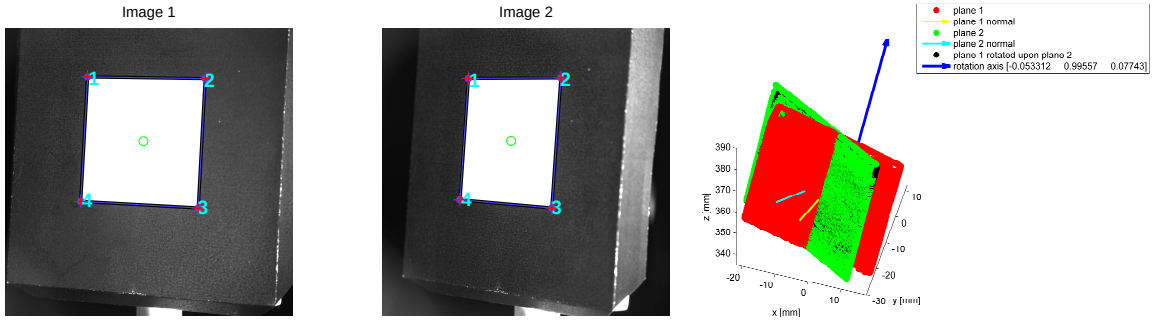
The general principle lies in the acquisition of intensity and 3D point cloud data of several orientations of a calibration target and subsequent determination of the underlying rotation axis and its offset in the camera coordinate system.

Since the rotation axis $\mathbf{a}$ is not necessarily perpendicular to the normal directions of the rotated plane[3], it needs to be computed based on tracking of interest points. Additionally, the tips of the normals of a rotated plane describe a circle that lies within a plane in the three-dimensional space. The *direction* of the rotation axis can thus be determined based on finding the normal on that plane. For this, the normal directions are determined from a least squares fit of a plane to the point cloud data to each of the orientations with

[3]The rotation axis is perpendicular to the normal directions if and only if the rotation axis is parallel to the rotated plane and thus does not intersect with the plane. In that case, the rotation axis can be determined using the cross product such that $\mathbf{a} = \frac{\mathbf{n}_1 \times \mathbf{n}_2}{||\mathbf{n}_1 \times \mathbf{n}_2||_2}$.

outliers suppression[4].

With the rotation axis direction $\mathbf{r} = (r_x, r_y, r_z)$ and the rotation angle being known, the rotation matrix



(a) Images of the calibration target with different orientations

(b) 3D point cloud data of valid points on the bright area on the calibration target

Figure C.2.: Illustration of the main rotation axis calibration aspects. The calibration target consists of a black plain plate with a $40 \times 40\ mm$ white rectangular area.

R for a rotation of α about $\mathbf{r}$ is

$$\mathbf{R}(\alpha) = \begin{pmatrix} \cos\alpha + r_x^2(1-\cos\alpha) & r_x r_y(1-\cos\alpha) - r_z\sin\alpha & r_x r_z(1-\cos\alpha) + r_y\sin\alpha \\ r_y r_x(1-\cos\alpha) + r_z\sin\alpha & \cos\alpha + r_y^2(1-\cos\alpha) & r_y r_z(1-\cos\alpha) - r_x\sin\alpha \\ r_z r_x(1-\cos\alpha) - r_y\sin\alpha & r_z r_y(1-\cos\alpha) + r_x\sin\alpha & \cos\alpha + r_z^2(1-\cos\alpha) \end{pmatrix} \tag{C.2}$$

The rotation of a point $\mathbf{p}$ is then

$$\mathbf{p}_{rotated} = R \cdot \mathbf{p} \tag{C.3}$$

And with inclusion of the offset $\mathbf{o}$

$$\mathbf{p}_{rotated} = R \cdot (\mathbf{p} - \mathbf{o}) + \mathbf{o} = R \cdot \mathbf{p} - (R - \mathbf{1}) \cdot \mathbf{o}. \tag{C.4}$$

For our purpose, R is known from the above derivations, $\mathbf{p}$ and $\mathbf{p}_{rotated}$ (column vectors) are known by tracking interest points on the rotated plane. Therefore, with $N - 1$ tracked points, $\mathbf{o}$ can be computed using least squares:

$$(\leftarrow \mathbf{o} \rightarrow) = (R - \mathbf{1})^{-1}(-(\leftarrow \mathbf{p}_{rotated} \rightarrow) + R(\leftarrow \mathbf{p} \rightarrow)). \tag{C.5}$$

[4]The detailed procedure is: (1) acquire 3D point cloud data, (2) determine point cloud values that correspond to points on the bright rectangle of the calibration plate, (3) determine a least squares fit through that plane, (4) remove all measurements from the point cloud that have a distance of more than D $[mm]$ from the fitted plane, (5) determine normal direction from a second least squares fit of a plane to the remaining data points.

D Usage permission(s)

D.1 Blender

Subject: Blender model license
From: "Steffen Herbort" <steffen.herbort@tu-dortmund.de>
Date: Mo, 4.02.2013, 10:37
To: jpbouza@gmail.com

Dear JP,

are the models, expecially the “athletic human” under an open source license and is it allowed to used them for scientfic work?

Best regards, Steffen

———–

Subject: Re: Blender model license
From: "Juan Pablo Bouza" <jpbouza@gmail.com>
Date: Mo, 4.02.2013, 16:19
To: "Steffen Herbort" <steffen.herbort@tu-dortmund.de>

Hi Steffen! Yes, the license is open source, Creative Commons. The only thing it asks for is for the user to put the original author in the credits of the work. Therefore, if your work happens to be published, you should put something like “human model used from Blenrig by Juan Pablo Bouza” and in the case of the athletic human you should also put “model by Nick Zuccarello”, cause in that case the model is from Nick. The other models were created by me.

Well, that is the only condition, it’s very similar to GPL license.

See you!