Doris Antensteiner, MSc. Mag.rer.soc.oec. BSc.

Fusion of Light Field with Photometric Stereo

Dissertation

to achieve the university degree of

Doctor of Philosophy

Doktoratsstudium der technischen Wissenschaften: Informatik

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dr. Thomas Pock

Institute of Computer Graphics and Vision

Graz, July 2018

Every great and deep difficulty bears in itself its own solution.
It forces us to change our thinking in order to find it.

Niels Bohr

Acknowledgments

My deepest gratitude is to my supervisor Professor Thomas Pock and my advisors Svorad Štolc and Reinhold Huber-Mörk. I have been fortunate to be provided with knowledge, inspiration and encouragement throughout the period of my research. I am also grateful to Professor Vaclav Hlaváč for acting as my second supervisor on my thesis committee.

I would like to express my appreciation to my colleagues in the High Performance Imaging group at the Austrian Institute of Technology (AIT). I was lucky to work in a very friendly, inspirational and supportive environment. Especially I want to thank the Inline Computational Imaging team including Bernhard Blaschitz, Nicole Brosch, Dorothea Heiss, Petra Thanner and Lukas Traxler.

I also want to thank the Computer Graphics and Vision group at the Technical University of Graz for many fruitful cooperations and interesting discussions. I especially want to express my gratitude to Stefan Heber for the discussions in the early phase of my research, as well as to Patrick Knöbelreiter for his cooperation.

Finally, I want to thank my family and friends for their throughout support.

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the presented doctoral thesis.

_____   _____

Date   Signature

Abstract

The acquisition of superior shape and reflectance properties of a scene is a current dilemma of computer vision long beset by a myriad of complications. In this thesis we overcome these obstacles by combining light field depth with photometric stereo normals in order to reach a refined depth reconstruction and material analysis. To realize this, we analyze light rays passing through the camera lens capturing a scene illuminated from a defined direction. Conventional imaging systems only provide limited information, since they can't capture directional radiance information and only provide the sum of light at each image position. Hence, they capture two-dimensional pictures of this world. We aim to achieve a more complete description by using both light field cameras and photometric stereo techniques.

Light field imaging offers powerful capabilities in the field of computer vision. Rays of light traveling through space are defined by a plenoptic function, which we utilize in a 5D form containing three spatial and two directional dimensions for each ray. Currently, a number of light field cameras are available including area scan devices such as plenoptic or matrix cameras and multi-line scan cameras for industrial applications. We process the provided information to obtain a 3D reconstruction of the scene and to capture the reflectance and material description of the objects.

Photometric stereo uses several images taken from a constant viewing angle and multiple illumination directions. Changes in the radiance of an object point depend on the surface orientation and reflectance properties of the material. To achieve a full description of a scene we use photometric stereo techniques to obtain results of improved quality. The fusion of light fields and photometric stereo gives improved depth reconstruction results and allows a more detailed analysis of the reflectance distribution function of the surface.

In this thesis we provide a thorough analysis and classification of methods combining surface depth and surface normal data. Based on our findings we introduce a novel generalized combinatorial formulation as well as a gradient-based method using Total Generalized Variation (TGV). The analysis is carried out for standard area-scan imaging as well as for multi-line scan data. The latter is most commonly applied by in-line industrial environments were objects are moving w.r.t. the acquisition device. We present approaches for a combined reconstruction of depth and surface orientation with missing photometric stereo evidence orthogonally to the transport direction.

Another contribution deals with feature construction for multi-line scan data. We present novel features which ensure stable results with respect to noise regardless of the regions reflective properties. We compare these features with commonly used Census Transform (CT) features.

The appearance of surfaces can be described by reflectance function. These functions depend on properties such as the position of the object, the wavelength of the illumination, the viewing angle and the illumination direction. A special case is the bidirectional reflectance distribution function (BRDF), which is described by four variables defining the viewing and illumination direction. Capturing full BRDFs is highly time consuming and not feasible. We present a machine learning approach to reconstruct full BRDFs from partial data using convolutional neural networks (CNNs).

Applications of our results lie in the field of industrial inspection tasks such as defect detection, brand protection, product security and optical inspection of materials. A highly precise 3D reconstruction allows a detailed error detection in production lines. Methods of anti-counterfeiting can be improved by a better description and analysis of the material structure.

Keywords. Light Field, Photometric Stereo, Computer Vision, Variational Methods, Reflectance Functions, Multi-Line Scanning, 3D Reconstruction

Kurzfassung

In dieser Arbeit betrachten wir die problematische Natur der hochpräzisen 3D Rekonstruktion von Objekten mittels einer Kombination von Tiefendaten von Lichtfeld mit Oberflächennormalen von photometrischem Stereo. Wir analysieren Lichtstrahlen die von Objekten reflektiert werden unter definierten Beleuchtungsbedingungen. Konventionelle bildgebende Methoden ermöglichen die Aufzeichnung von Richtungsinformation des Lichteinfalls nicht und halten die Summe des einfallenden Lichtes fest. Daher wird lediglich eine 2D Repräsentation der Szene erstellt. Wir forcieren eine umfangreichere Szenenbeschreibung mittels Lichtfeld und photometrischem Stereo.

Die Verbreitung von Lichtstrahlen im Raum wird über eine plenoptische Funktion definiert, wir nutzen diese in einer 5D Formulierung welche aus drei räumlichen und zwei richtungsgebenden Dimensionen besteht. Zum heutigen Zeitpunkt sind mehrere Lichtfeldkameras verschiedener Typen auf dem Markt erhältlich. Diese umfassen plenoptische Kameras, Matrixkameras sowie industrielle Multi-Zeilenkameras. Wir verwenden Lichtfeldinformation um eine 3D Oberflächenrekonstruktion zu erstellen sowie für die Aufzeichnung von Reflektanz- und Materialeigenschaften von Objekten.

Bei photometrischem Stereo werden mehrere Bilder unter einer konstanten Kameraansicht und variierenden Beleuchtungswinkeln aufgezeichnet. Diese Information erlaubt eine vollständigere Szenenbeschreibung von hoher Qualität. Wir berechnen hochpräzise 3D Rekonstruktionen durch die Kombination von Lichtfeld mit photometrischem Stereo. Zusätzlich erlaubt uns diese Fusion eine detailliertere Analyse der Reflektanzverteilungsfunktionen von Objekten.

In dieser Arbeit zeigen wir eine gründliche Analyse und Kategorisierung von Methoden welche die Tiefe von Objekten mit Oberflächennormalen kombinieren und führen basierend auf unserer Analyse präzisere Methoden ein. Resultierend aus unserer Analyse führen wir sowohl eine verbesserte Formulierung zur Kombination ein als auch eine Regularisierung mittels einer Variationsmethode (Total Generalized Variation). Wir analysieren diese Methoden sowohl für Flächen- als auch Multizeilenkameras. Letztere werden zumeist in industriellen in-line Umgebungen verwendet, in welchen sich Objekte auf Förderbändern mit konstanter Geschwindigkeit bewegen. Wir präsentieren Lösungen für die kombinierte 3D Oberflächenrekonstruktion von Tiefendaten mit Oberflächennormalen in industriellen in-line Umgebungen. Wir zeigen den Umgang mit fehlender photometrischer Information orthogonal zur Transportrichtung.

Ein weiterer Beitrag dieser Arbeit zeigt die Erstellung von Features für Daten von Multi-Zeilenkameras. Wir präsentieren neuartige Features, welche stabile Resultate in schwierigen Regionen die rauschen oder stark reflexive Eigenschaften aufweisen. Diese werden mit den häufig verwendeten Features der Census Transformation (CT) verglichen.

Das Erscheinungsbild von Oberflächen kann mittels Reflektanzverteilungsfunktionen beschrieben werden. Diese sind abhängig von Objekteigenschaften wie der Position auf der Objektoberfläche, der Wellenlänge der Beleuchtung und dem Betrachtungs- sowie Beleuchtungswinkel. Ein spezieller Fall ist die Bidirektionale Reflektanzverteilungsfunktion (BRDF), welche mittels vier Variablen abhängig von Lichteinfall und Sensorposition definiert wird. Vollständige BRDFs für ein bestimmtes Material aufzuzeichnen würde mehrere Jahre in Anspruch nehmen und ist daher nicht durchführbar. Wir präsentieren eine Methode zur Rekonstruktion von vollständigeren BRDFs von partiellen Daten mittels eines faltenden neuronalen Netzwerkes (CNN).

Anwendungsgebiete unserer Methoden umfassen Aufgaben in der industriellen Inspektion wie zum Beispiel in der Defekterkennung, im Markenschutz, in der Produktsicherheit und in der optischen Inspektion von Materialien. Hochpräzise 3D Rekonstruktionen erlauben eine detaillierte Fehlererkennung in der industriellen Inspektion. Weiters können Methoden zur Gewährleistung der Fälschungssicherheit

durch eine verbesserte Beschreibung und Analyse von Materialeigenschaften verbessert werden.

Schlagwörter. Lichtfeld, Photometrisches stereo, Bildverarbeitung, Variationsmethoden, Reflektanzverteilungsfunktionen, Multi-Zeilenkamera, 3D Rekonstruktion

Contents

Contents

Contents

List of Figures

1 Introduction

Contents

1.1 Motivation

In this thesis we endeavor to resolve the problematic nature of acquiring a highly precise 3D surface reconstruction of objects with a focus on industrial applications. To facilitate this end we conceived, designed and implemented highly precise optical depth measurement systems. Depth acquisition systems can be classified in active and passive approaches. While active methods interfere with the objects, e.g. by the projection of structured light, passive methods use imaging sensors to capture the radiance of a scene illuminated by light sources. We focus on passive systems, which allow the capture of 3D and texture maps simultaneously, enabling an easy setup procedure as well as high quality depth reconstruction. We present solutions for several types of light field cameras, such as in-line multi-line scan systems and area-scan devices.

Conventional stereo cameras are impaired by reconstruction errors in dark, textureless and highly reflective regions in conjunction with insufficient fine surface details. We transcend these limitations by fusing the complementary properties of light field and

Light field   Fusion of light field with photometric stereo
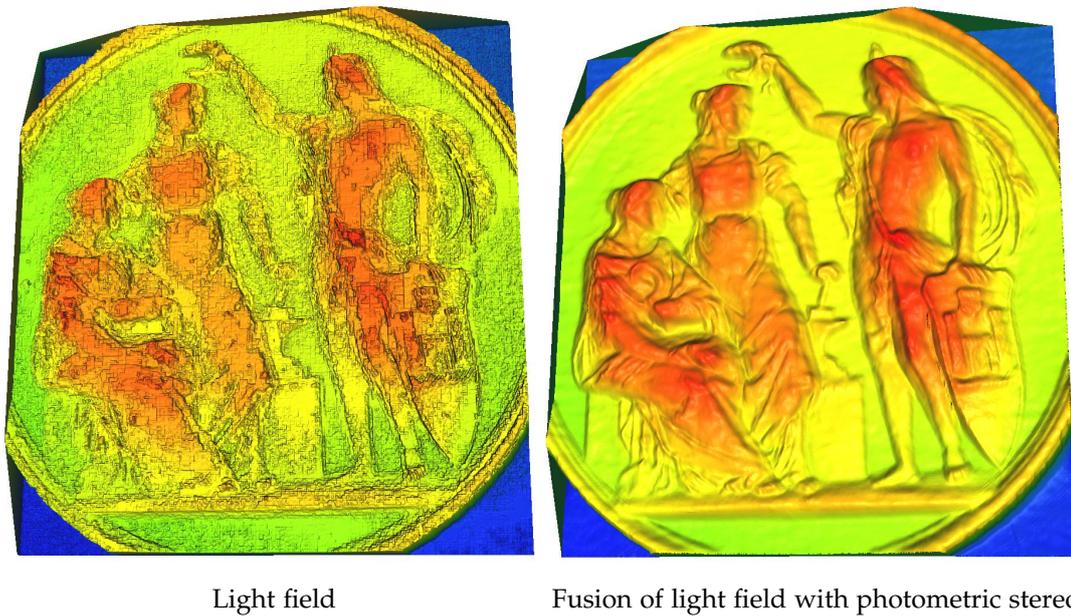
Figure 1.1: Fusion of light field and photometric stereo compared to the 3D reconstruction with light field.

photometric stereo systems. Light field cameras capture light from multiple viewing angles, allowing a more robust depth estimation of objects. This utility occurs regardless of a high variety of surface reflectance properties including partial occlusions. While this method increases global accuracy, it suffers from local imprecision e.g. in textureless regions. Photometric stereo allows the estimation of surface normals by illuminating the object from several directions. The resulting reflections depend on the orientation of the surface in respect to the light sources and on the position of the acquisition device. Hence the surface orientation can be recovered by analyzing the irradiance values captured by the sensor, independent of the surface texture. The recovered orientation values enable a locally precise depth reconstruction with a weakness in the global accuracy. The fusion of light field and photometric stereo methods is demonstrated in Fig. 1.1.

Depth reconstruction tasks can be approached with tools such as stereo, multi-view matrix or plenoptic cameras. Stereo cameras are restricted in reconstruction quality due to limited information in highly reflective regions or matching ambiguities regions

with little or no texture. Multi-view array cameras, which have a higher redundancy of information, are nevertheless burdened by the high complexity of harmonizing several sensor sensitivities, lenses and focal lengths. Plenoptic cameras reduce complexity, boasting a single sensor, one set of main lenses, and a microlens array allowing the capture of directional information. Their disadvantages lie in the limited baseline given by the physical sensor size and in a reduced spatial optical resolution. In plenoptic systems the spatial resolution is traded for angular resolution. Placing these cameras under a light dome or adjusting a ring light gives additional photometric stereo data. This combination allows a highly precise surface reconstruction for area-scan setups.

In industrial in-line systems an object is scanned while it moves w.r.t. the camera, e.g. on a conveyor belt. Such in-line computational imaging tasks work at high speeds and are computationally expensive. Increasing computational power and memory paired with better and more effective algorithms is sought after. Multi-line scanners sample a chosen number of lines from the imaging sensors (e.g. 11), each with a different viewing angle on the scene. This gives highly redundant information which is beneficial in highly reflective, dark or low texture regions. While the object moves, each sensor line reconstructs a full image, generating 3D information simultaneously with texture data. Placing a line-light source next to the camera provides a different illumination direction for each sensor line and hence gives additional photometric stereo information.

Additionally to the high precision reconstruction results of such fusion methods, multi-line scan systems are robust, convenient to install and calibrate for industrial setups. Applications cover the reduction of shadows, detection of cracks in materials, computation of all-in-focus images, inspection of soldering points as well as segmentation and forgery detection of holograms (diffractive optically variable image devices - DOVIDs) on banknotes or documents.

1.2 Scientific Contribution

The attempt to combine light field with photometric stereo yielded various obstacles. These included the calibration of the system, the combination and weighting of

the terms, and the analysis of reflectance distributions from the given data. All such variables were solved throughout our sustained enterprise. We focused on data acquired by area-scan and multi-line scan cameras. The latter allows our algorithms to be apt for a wide range of industrial applications.

Camera calibration is essential in order to achieve highly precise depth reconstruction results from light field images. We show the calibration of multi-view matrix cameras which requires a global optimization procedure to meet the epipolar constraints in all views. Multi-line scan cameras capture an object which is moving in a defined direction on a conveyor belt under the sensor. Hence calibration of such cameras has to take the movement of the object into account. We describe the line-based calibration and how to computationally deal with transport jitter in such systems. The calibration of photometric stereo setups covers the identification of the position of the light sources. This is often achieved by placing highly reflective spheres under the acquisition device.

To analyze depth from light field data, we solve a correspondence analysis problem. This can be done directly on the light field data, where a consistent match for all observation angles is found. A more robust alternative is feature based matching, where irradiance values are first processed in order to extract feature maps for each view. We introduce robust features which can handle even highly reflective materials.

We provide an in-depth comparison of several variational methods using depth and surface orientation data. We classify and evaluate weighting terms of common methods and explain their differences. Based on our findings, we introduce a novel generalized orientation weighting term which gives an improved performance. Additionally we introduce a gradient-based method which is using a Total Generalized Variation (TGV) regularization.

Objects can show a wide range of materials and surface structures. For the detection and classification of properties or defects on these objects it is essential to understand their reflectance distribution functions. Since capturing the dense functions is highly time consuming (several years) and hence is infeasible, acquisition devices only capture a small portion of them. We introduce a method to reconstruct more dense functions from partial data using machine learning methods.

1.3 Thesis Overview

Chapter 2 introduces the notation used in the thesis and gives an overview over convex optimization problems and neural networks.

Chapter 3 introduces several aspects of light field imaging, including the 3D reconstruction from light field data and the calibration of light field capturing devices. We build on the calibration of matrix cameras which we presented in [18] and the calibration of multi-line scan devices which we published in [19, 27].

Chapter 4 presents the acquisition, calibration and processing of photometric stereo data. Furthermore we address the extraction of photometric stereo data from multi-line scan acquisitions which we first presented in [9], with a refined formulation in [8].

Chapter 5 presents a highly precise depth reconstruction method using depth (e.g. from light field) and surface normal (e.g. from photometric stereo) information based on our work published in [7]. Earlier we introduced a basic frequency based combination in [5].

Chapter 6 is based on our publication in [8]. We refine our previous ideas and adapt them specifically for multi-line scanning data. Hence we introduce optimization procedures to deal with missing photometric stereo evidence orthogonally to the transport direction. Additionally we present new features which are tailored to the reflectance and noise properties of multi-line scan acquisitions.

Chapter 7 shows the measurement and representation of reflectance distribution functions and introduces the calculation of full bidirectional distribution functions (BRDFs) from partial data using CNNs, based on our work published in [4].

Chapter 8 concludes the thesis and gives an outlook to future work.

1.4 Resulting Publications

The work presented in this thesis resulted in the following publications:

- D. Antensteiner, S. Štolc, and T. Pock. "Multi-line Scan 3D Sensing With A Hybrid Light-field And Photometric Stereo Approach." In: International Conference on Pattern Recognition (ICPR). (2018).

- B. Blaschitz, D. Antensteiner, and S. Štolc. "Multi-camera Array Calibration For Light Field Depth Estimation." In: Proceedings of Austrian Association for Pattern Recognition Workshop (OAGM). (2018).

- D. Antensteiner, S. Štolc, and T. Pock. "A Review Of Depth And Normal Fusion Algorithms." In: Journal Sensors 18.2. (2018).

- N. Brosch, S. Štolc, and D. Antensteiner. "Warping-based Motion Artifact Compensation For Multi-line Scan Light Field Imaging." In: IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision, CA, USA. (2018).

- B. Blaschitz, S. Štolc, and D. Antensteiner. "Geometric Calibration And Image Rectification Of A Multi-line Scan Camera For Accurate 3D Reconstruction." In: IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision, CA, USA. (2018).

- D. Antensteiner, S. Štolc, K. Valentin, B. Blaschitz, R. Huber-Mörk, and T. Pock. "High-precision 3D Sensing With Hybrid Light Field & Photometric Stereo Approach In Multi-line Scan Framework." In: IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision. (2017).

- D. Antensteiner and S. Štolc. "Full BRDF Reconstruction Using CNNs From Partial Photometric Stereo-light Field Data." In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. (2017).

- D. Antensteiner, S. Štolc, and R. Huber-Mörk. "Depth Estimation With Light Field And Photometric Stereo Data Using Energy Minimization." In: Proceedings of 21st Ibero-American Congress on Pattern Recognition (CIAPR). Springer Lecture Notes in Computer Science. (2016).

- D. Antensteiner, B. Blaschitz, C. Eisserer, R. Huber-Mörk, J. Ruisz, S. Štolc, K. Valentin: "Line-Scan Stereo for 3D Ground Reconstruction". Fraunhofer Forum Bildverarbeitung 2016, Karlsruhe. in: "Forum Bildverarbeitung 2016", Scientific Publishing, Karlsruher Institut für Technologie (KIT), Karlsruhe. S. 209 - 220. (2016).

- D. Antensteiner, S. Štolc, and R. Huber-Mörk. "Depth Estimation Using Light Fields And Photometric Stereo With A Multi-line-scan Framework." In: Proceedings of Austrian Association for Pattern Recognition Workshop (OAGM). (2016).

2 Methodology

Contents

In this chapter we introduce the notation and methodological foundation used in this work. We start with presenting the essential notations in Sec. 2.1, which are used in the following chapters. An overview of convex optimization is given in Sec. 2.2, where convex analysis and algorithms are described. An outline of the theory on neural networks is given in Sec. 2.3. Here we present approaches to machine learning, describe basic concepts with feed-forward networks and, finally, convolutional neural networks (CNNs).

2.1 Notation

In this section we introduce the essential notations which are used in the following chapters of this thesis. By default we assume discretized images of the size of $M \times N$ pixels. In order to access the image location, we define the index set $\mathcal{I} = \{\mathbf{i} = (i,j) : 1 \leq i \leq M, 1 \leq j \leq N\}$. Consider a two-dimensional image $P \in \mathbb{R}^{M \times N}$, discrete values in each pixel are defined as follows:

$$P = (P_{i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N}. \tag{2.1}$$

A p-norm on the scalar valued image $P \in \mathbb{R}^{M \times N}$ is defined as follows:

$$||P||_p = \left(\sum_{i=1}^{M} \sum_{j=1}^{N} |P_{i,j}|^p \right)^{1/p}, \tag{2.2}$$

which we interpret as a vector. Variables with a bold font refer to vector valued images. Hence, the tensor $\boldsymbol{P}$ is defined as:

$$\boldsymbol{P} = \left(\boldsymbol{P}_{i,j} \right)_{i,j \in \mathcal{I}} \in (\mathbb{R}^K)^{M \times N} = \mathbb{R}^{M \times N \times K}, \quad \text{where} \quad \boldsymbol{P}_{i,j} = (P_{i,j,1}, ..., P_{i,j,K}). \tag{2.3}$$

For vector valued images $\boldsymbol{P} \in \mathbb{R}^{M \times N \times K}$ we consider the p,q-norm which is given as:

$$||\boldsymbol{P}||_{p,q} = \left(\sum_{i=1}^{M} \sum_{j=1}^{N} |\boldsymbol{P}_{i,j}|_p^q \right)^{1/q}, \tag{2.4}$$

where the pointwise p-vector norm $|\boldsymbol{P}_{i,j}|_p$ is defined as:

$$|\boldsymbol{P}_{i,j}|_p = \left(\sum_{k=1}^{K} |P_{i,j,k}|^p \right)^{1/p}. \tag{2.5}$$

Hence the commonly used $L_{2,1}$-norm is given by

$$||\boldsymbol{P}||_{2,1} = \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{|P_{i,j,1}|^2 + |P_{i,j,2}|^2 + ... + |P_{i,j,K}|^2}. \tag{2.6}$$

The discrete depth map of a scene is scalar valued in each pixel and defined as follows:

$$Z = (Z_{i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N}. \tag{2.7}$$

The gradient of a depth map Z can be computed using standard finite differences:

$$\nabla Z = ((\nabla Z)_{i,j})_{i,j \in \mathcal{I}}, \quad \text{where} \quad (\nabla Z)_{i,j} = ((\nabla_x Z)_{i,j}, (\nabla_y Z)_{i,j}), \tag{2.8}$$

where the gradient operator in x- and y-direction $\nabla : \mathbb{R}^{M \times N} \to \mathbb{R}^{M \times N \times 2}$ is given by:

$$(\nabla_x Z)_{i,j} = \begin{cases} Z_{i+1,j} - Z_{i,j} & \text{if } 1 \leq i < M, \\ 0, & \text{otherwise,} \end{cases}$$

$$(\nabla_y Z)_{i,j} = \begin{cases} Z_{i,j+1} - Z_{i,j} & \text{if } 1 \leq j < N, \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Figure 2.1: Illustration of a line through two points. The line through x_1 and x_2 is parameterized by $y = \theta x_1 + (1 - \theta) x_2$ with $\theta \in \mathbb{R}$.

2.2 Convex Optimization

In this section we present a rough overview about essential aspects of convex optimization. The material is based on literature. For deeper insights into the matter we refer to [24, 30, 150, 168].

Convex optimization is a well developed field since convex functions and sets have convenient properties. Strictly convex functions have a unique minimum. Since any local minimum has to be a global minimum, first-order conditionals are sufficient optimality conditions. Convex functions are continuous, discontinuities can only occur at the boundary of the domain of the function.

2.2.1 Convex Analysis

Convex Sets

The set $C \subseteq \mathbb{R}^N$ is affine if a line between any two points $x_1, x_2 \in \mathbb{R}^N$ with $x_1 \neq x_2$ in C lies in C:

$$y = \theta x_1 + (1 - \theta) x_2, \quad \text{where} \quad \theta \in \mathbb{R}. \tag{2.10}$$

An affine combination defines the line through two points x_1 and x_2, which is illustrated in Fig. 2.1.

(a) Convex set (b) Non-convex set
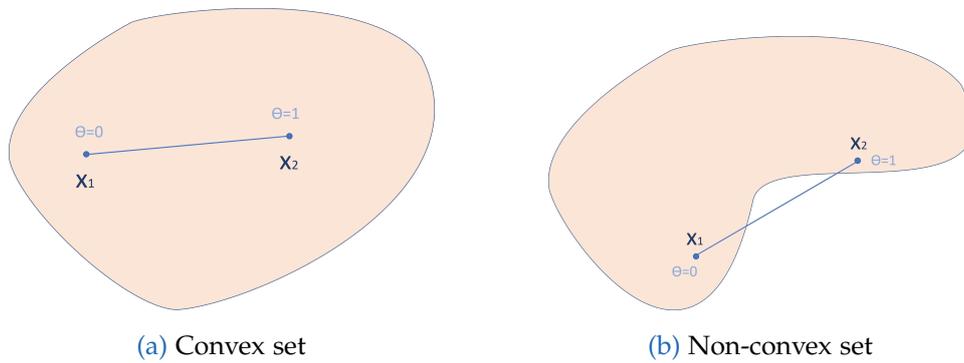
Figure 2.2: Illustration of convex and non-convex sets. In the convex set in a) the line between any two points is contained in the set. The kidney shape in b) is non-convex.

A set C is convex if a straight line between any pairs of points $x_1, x_2 \in C$ lies also in C:

$$y = \theta x_1 + (1 - \theta)x_2 \in C, \quad \text{where} \quad \theta \in [0, 1] \tag{2.11}$$

and θ forms the line between the point pairs with $y = x_1$ for $\theta = 0$ and $y = x_2$ for $\theta = 1$.

If all points in a set can be connected by a straight line which lies in the set, the set is convex. A convex and non-convex set are illustrated in Fig. 2.2. The convex combination of two points x_1 and x_2 defines the line segment between the points.

Convex Functions

A convex optimization problem minimizes a convex function F over a convex set C and has the form

$$\min_{x \in C} F(x). \tag{2.12}$$
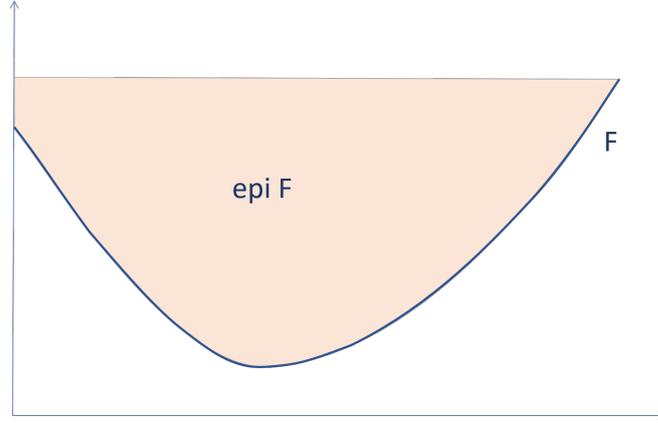
Figure 2.3: Illustration of the epigraph of a function. The epigraph (red) of a function F (blue) is a set of points lying on or above its graph.

A function is convex if and only if its epigraph is a convex set [55]. The epigraph of a function $F : \mathbb{R}^N \to \mathbb{R}$ is defined as the set of points lying above the graph as illustrated in Fig. 2.3.

$$epi\ F = \{(x, \mu) : x \in \mathbb{R}^N, \mu \in \mathbb{R}, \mu \geq F(x)\} \subseteq \mathbb{R}^{N+1} \tag{2.13}$$

A function $F : \mathbb{R}^N \to \mathbb{R}$ is proper, if $F(x) < +\infty$ for at least one x (its effective domain is nonempty) and $F(x) > -\infty$ for every x. In this case it is convex, if the domain of the function is a convex set and if all line segments $(x, F(x))$ and $(y, F(y))$ lie above the graph of F for $\theta \in [0, 1]$ and $x, y \in dom\ F$:

$$F(\theta x + (1 - \theta)y) \leq \theta F(x) + (1 - \theta)F(y). \tag{2.14}$$

This is a specific form of the Jensen's inequality which follows:

$$F(\sum_i \theta_i x_i) \quad \leq \quad \sum_i \theta_i F(x_i), \quad \text{with} \tag{2.15}$$
$$\theta_i \geq 0, \quad \text{and} \quad \sum_i \theta_i = 1.$$

A convex function is illustrated in Fig. 2.4, where the line segment between any two points of the function lies above the graph. For convex functions Eq. 2.14 holds when $x \neq y$ and $\theta \in (0, 1)$. A function F is concave if $-F$ is convex.
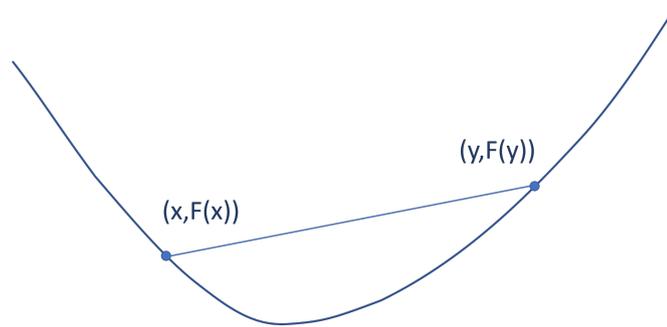
Figure 2.4: Illustration of a convex function. Any line segment between two points lies above the graph of the function.
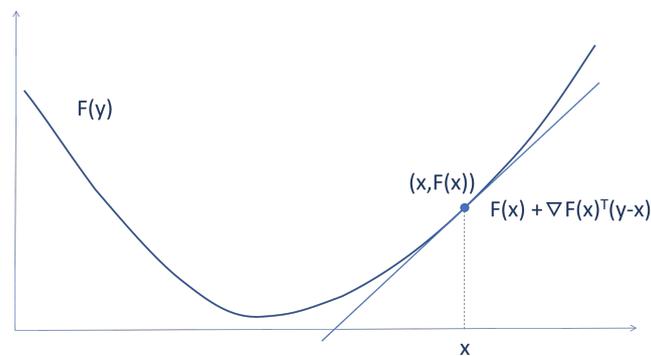
Figure 2.5: Illustration of the first order convexity condition. If the function F is convex and differentiable Eq. 2.16 holds for all $x, y \in dom\ F$.

First Order Condition

A function F is differentiable, if the gradient ∇F exists at each point in its domain. The graph of a differentiable convex function F is always above the linear approximation at a point x, hence the function is characterized by:

$$F(y) \geq F(x) + \nabla F(x)^T (y - x). \tag{2.16}$$

This describes that a function is convex, if the first order Taylor approximation of F near x is always a global underestimator of the function. Hence, convex functions allow the derivation of global information (global understimator) from local information
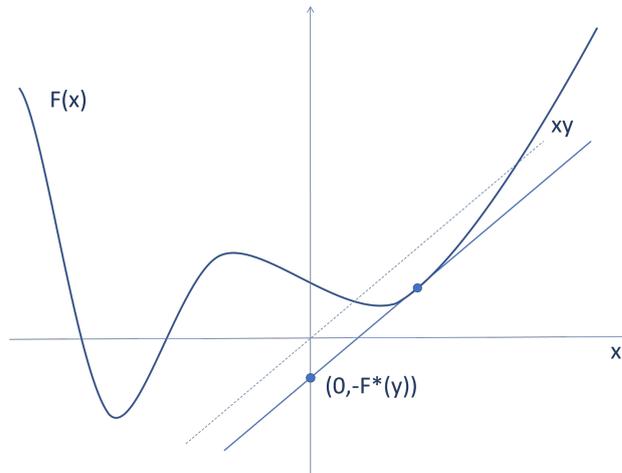
Figure 2.6: Illustration of the convex conjugate for one point. The function $F^*(y)$ can be represented as the maximum gap between a linear function xy and $F(x)$.

(first-order Taylor approximation at a point). If $\nabla F(x) = 0$, then x is a global minimizer of the function F, since $F(y) \geq F(x)$ for all $y \in dom\ F$.

Second Order Condition

A function F is twice differentiable, if the second derivative (or Hessian) $\nabla^2 F$ exists at each point in its domain. Such a function is convex, if and only if the domain of the function is convex and its Hessian is positive semidefinite (i.e. all eigenvalues are nonnegative) for all $x \in dom\ F$:

$$\nabla^2 F(x) \succeq 0. \tag{2.17}$$

This means that the graph has a non-negative curvature at the point x. If $F : \mathbb{R} \rightarrow \mathbb{R}$ this means that the 2nd derivative is nonnegative, i.e.:

$$F''(x) \geq 0. \tag{2.18}$$
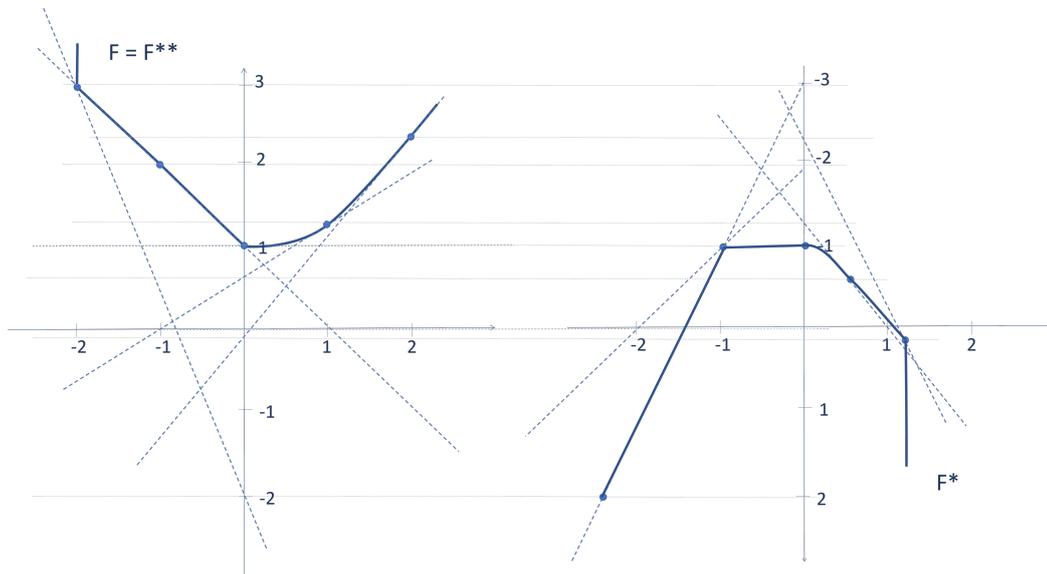
Figure 2.7: Illustration of the convex conjugate of a convex function. This shows the geometric interpretation of the convex conjugate $F*$ of the convex function F. The intersection of the tangent line intersects with the y axis defines the negative of the dual value at the value of the slope.

Convex Conjugate

The convex conjugate is used to form an optimization problem into its corresponding dual problem. If the function F is differentiable, the convex conjugate is also called Legendre transform of F. For a function $F : \mathbb{R}^N \to \mathbb{R}$ the convex conjugate is defined as:

$$F^*(y) = \sup_{x \in dom\ F} \{\langle x, y \rangle - F(x)\}. \qquad (2.19)$$

Independent of the convexity of F, the convex conjugate is always a convex function. Applying the convex conjugate twice leads to the convex bi-conjugate which satisfies the inequality $F(x) \geq F^{**}(x)$:

$$F^{**}(x) = \sup_{y \in dom\ F^*} \{\langle y, x \rangle - F^*(y)\}. \qquad (2.20)$$

Convex functions F are equal to their convex bi-conjugate $F = F^{**}$. For a specific point y, the convex conjugate can geometrically be represented as the largest gap between
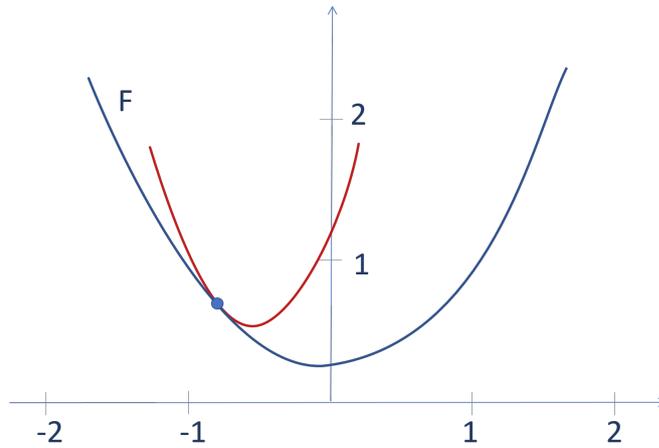
Figure 2.8: Illustration of a function with a Lipschitz continous gradient. A quadratic majorizer of the function F is shown in red.

a linear function xy and $F(x)$ (Fig. 2.6). The geometric construction of the convex conjugate of a convex function is shown in Fig. 2.7. It can also be seen that the convex bi-conjugate F^{**} and the original convex function F are equivalent.

Lipschitz Continuous Gradient

A continuously differential function $F : \mathbb{R}^N \to \mathbb{R}$ has a Lipschitz continuous gradient if there exists a constant $L > 0$ such that:

$$||\nabla F(x) - \nabla F(y)|| \leq L||x - y||, \quad \forall x, y \in domF. \tag{2.21}$$

A function F with a Lipschitz continuous gradient with the parameter L has a conjugate F^* which is strongly convex with a parameter $1/L$. Functions with a Lipschitz continuous gradient provide a quadratic majorizer (see Fig. 2.8), which is presented at the right hand side:

$$F(y) \leq F(x) + \nabla F(x)^T(y - x) + \frac{L}{2}||y - x||^2. \tag{2.22}$$

This majorizer touches the function F in one point, where both have the same slope.

(a) Orthogonal projection   (b) Proximal operator
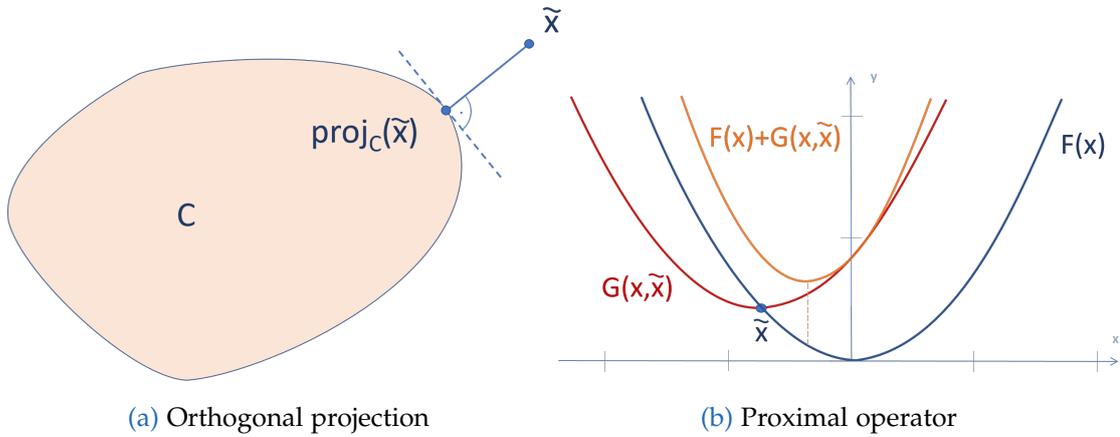
Figure 2.9: Orthogonal projection and proximal operator. a) Orthogonal projection of a point $\tilde{x}$ onto a convex set C. b) Evaluation of the proximal operator at a point $\tilde{x}$ as formulated in Eq. 2.28 is shown by the minimum position of the the combined (orange) function, the quadratic distance to the point $\tilde{x}$ is denoted as $G(x, \tilde{x})$.

Proximal Operator

Many practical optimization problems are making use of the proximal operator which was introduced in [144]. The projection of a point $\tilde{x}$ onto a convex set C is described by:

$$\operatorname*{proj}_{C}(\tilde{x}) = \arg\min_{x \in C} \frac{1}{2}||x - \tilde{x}||^2. \tag{2.23}$$

This projection is illustrated in Fig. 2.9. It is an orthogonal projection, since it relies on finding the shortest distance under the Euclidean norm between the point $\tilde{x}$ to the convex set C. We can define the projection as an unconstrained optimization problem by utilizing the indicator function:

$$
\text{proj}_C(\tilde{x}) = \arg\min_x \{||x - \tilde{x}|| + I_C(x)\} \tag{2.24}
$$

$$
= \arg\min_x \{\frac{1}{2}||x - \tilde{x}||^2 + I_C(x)\}, \quad \text{where} \tag{2.25}
$$

$$
I_C(x) = \begin{cases} 0, & \text{if } x \in C \\ \infty, & \text{otherwise.} \end{cases} \tag{2.26}
$$

Given a convex function F the proximal operator $\text{prox}_F : \mathbb{R}^N \to \mathbb{R}^N$ is given by:

$$
\text{prox}_F(\tilde{x}) = \arg\min_x \left\{ F(x) + \frac{1}{2}||x - \tilde{x}||^2 \right\}. \tag{2.27}
$$

The function on the right hand side is strongly convex and has a unique minimizer for every $\tilde{x} \in \mathbb{R}^N$. The proximal operator of the scaled function τF is given by:

$$
\text{prox}_{\tau F}(\tilde{x}) = \arg\min_x \left\{ \tau F(x) + \frac{1}{2}||x - \tilde{x}||^2 \right\}, \tag{2.28}
$$

where $\tau > 0$ controls the movement which the proximal operator gives towards the minimum of the function F.

An interpretation of the proximal operator is shown in Fig. 2.9b. The $\text{prox}_{\tau F}$ is evaluated a specific point $\tilde{x}$, as described by Eq. 2.28. The quadratic distance to the point $\tilde{x}$ is denoted as $G(x, \tilde{x})$. The evaluation of the proximal operator is shown by the position of the minimum of the sum of both terms. The extend of the movement depends on the parameter τ.

Proximal algorithms are using proximal operators to solve a convex optimization problem. This is useful in cases where the proximal operator can be evaluated quickly. A simple example is iteratively applying prox_F to the convex function F to an initial point x_0.

2.2.2 Algorithms

In this section we discuss different optimization approaches which we use to fuse depth and surface orientation as well as the type of functions they can optimize. This

section is an adapted version of the algorithm summary we previously published in [7].

Least Squares

Least squares methods can be used to solve optimization problems which are described as a system of linear equations:

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} ||Ax - b||_2^2, \tag{2.29}$$

where $x \in \mathbb{R}^N$ is an unknown vector, which is estimated from the known matrix $A \in \mathbb{R}^{M \times N}$ and the vector $b \in \mathbb{R}^M$. A unique solution x is given, when the matrix A has N linearly independent columns, hence $(A^T A)$ is invertible with $N \leq M$:

$$x = (A^T A)^{-1} A^T b \tag{2.30}$$

Otherwise, iterative methods such as the conjugate gradient method can be implemented to minimize Eq. 2.29.

Gradient Descent

Gradient descent is a first order iterative optimization method to minimize an objective function $F : \mathbb{R}^N \rightarrow \mathbb{R}$ which is convex and differentiable:

$$\min_{x \in \mathbb{R}^N} F(x) \tag{2.31}$$

An x^* so that $F(x^*) = \min_x F(x)$ can be found for a convex function $F(x)$ with a plain gradient descent method with iterative updates as follows

$$x^{k+1} \leftarrow x^k - \alpha^k \nabla F(x^k), \tag{2.32}$$

where the size of the steps in the descent direction $-\nabla F(x^k)$, for a Lipschitz continuous gradient, is controlled by $\alpha^k > 0$. This update step is repeated until an accuracy of ϵ is reached. If the function F does not fulfill the convexity condition, the algorithm can get stuck in non-optimal stationary points (local minima, local maxima, saddle points).

Accelerated Proximal Gradient Method

Proximal algorithms minimize an objective function by using proximal operators, as described in [158]. These algorithms work on general conditions as non-smooth convex functions and they can be very fast, as there exist simple proximal operators for various energy functions. Additionally they can be used in a distributed way for large-scale problems. Proximal gradient methods consider optimization problems with the objective split into two components:

$$\min_{x} F(x) + H(x), \tag{2.33}$$

where $F : \mathbb{R}^N \to \mathbb{R}$ and $H : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions and where one of the functions (F) is additionally differentiable. The proximal gradient problem uses a step size $\tau^k > 0$ and is denoted as follows:

$$x^{k+1} \leftarrow \operatorname*{prox}_{\tau^k H}(x^k - \tau^k \nabla F(x^k)). \tag{2.34}$$

If ∇F is Lipschitz continuous with a constant L and a fixed step size $\tau^k = \tau \in (0, 2/L)$ is used, the method converges in $\mathcal{O}(1/k)$. When the constant L is not known, step sizes can be found with a line search algorithm (e.g. [14]), where the values are chosen in each step. Accelerated proximal gradient methods have an additional extrapolation step in the algorithm and are denoted as follows:

$$y^{k+1} \quad \leftarrow \quad x^k + w^k(x^k - x^{k-1}), \tag{2.35}$$

$$x^{k+1} \quad \leftarrow \quad \operatorname*{prox}_{\tau^k H}(y^{k+1} - \lambda^k \nabla F(y^{k+1})), \tag{2.36}$$

where λ^k denotes the step size and w^k an extrapolation parameter which is described given by:

$$w^k = \frac{k-1}{k+2}. \tag{2.37}$$

If ∇F is Lipschitz continuous with a constant L and a fixed step size $\tau^k = \tau \in (0, 1/L]$, the method shows convergence with a rate $\mathcal{O}(1/k^2)$. The fast iterative shrinkage thresholding algorithm (FISTA) is one of the most popular accelerated proximal gradient methods it is described in detail in [13].

Primal-Dual

Primal-dual methods can be used to solve a large number of optimization problems of the form:

$$\min_{x \in X} F(Kx) + H(x), \tag{2.38}$$

where $K \in \mathbb{R}^{M \times N}$ is a continuous linear operator and $F : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$, $H : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ are convex functions with inexpensive proximal operators. The corresponding saddle point problem is described as:

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + H(x) - F^*(y), \tag{2.39}$$

where F^* denotes the convex conjugate of the convex function F. We are using the Chambolle-Pock [29] primal-dual method and compute the proximal maps for H and F^*. A proximal descent is computed in the primal variable x and, in an alternating manner, a proximal ascent in the dual variable y:

$$\begin{cases} x^{k+1} \leftarrow \text{prox}_{\tau H}(x^k - \tau K^T y^k) \\ y^{k+1} \leftarrow \text{prox}_{\sigma F^*}(y^k + \sigma K(2x^{k+1} - x^k)) \end{cases} \tag{2.40}$$

The convergence rate is $\mathcal{O}(1/k)$, and can be improved if H or F^* are uniformly convex to $\mathcal{O}(1/k^2)$. Both, primal and the dual being uniformly convex the algorithm shows a linear convergence $\mathcal{O}(w^k)$, with $w > 0$ [14, 29, 149].

2.3 Neural Networks

In this section we give a basic overview over the field of machine learning (ML). The material is based on the textbooks [16, 73, 175].

Most modern machine learning approaches focus on supervised learning with deep convolutional neural networks. In associated tasks an input vector is mapped to an output vector. First we present basic concepts and fundamental goals of machine learning. Then we describe a simple feed forward network. Finally we proceed to present more advanced techniques of convolutional neural networks.
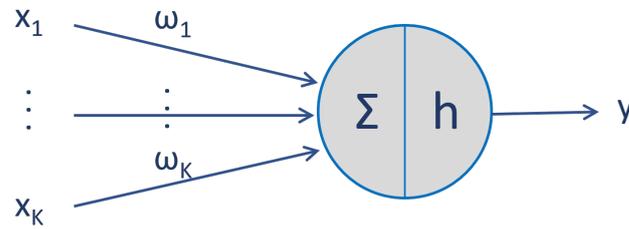
Illustration of a perceptron. The binary inputs $x = \{x_1, ..., x_K\}$ are mapped to a binary output y (see Eq. 2.41).

2.3.1 Machine Learning

In general, machine learning algorithms can be divided into supervised and unsupervised learning algorithms.

Unsupervised learning tasks learn data distributions in a set of features without prior category labels. Having several examples of an input set X the probability distribution or output set Y is learned. Hence, we learn to predict Y from X. The goal is to either cluster groups of data or to determine the data distribution in the input space, also called density estimation. Another application is the mapping of higher dimensional data to a lower dimensional space for visualization. In unsupervised learning no labeled input data is required.

Supervised learning has labels Y associated with all input vectors X which are represented in the dataset. We learn a mapping $F : X \rightarrow Y$ with a set of K training samples $\{(x_i, y_i) | i = 1, ..., K\} \subseteq X \times Y$. Applications are divided in classification and regression problems. A classification example is the categorization of images, where the images are labeled with a finite number of categories $Y = \{0, ..., n\}$ (e.g. car, apple, ...) and the network learns the representation of that category. In the field or regression we have a continuous classification (e.g. depth estimation with $Y \subseteq \mathbb{R}$). The goal of supervised learning tasks is to infer the output for elements X which were not represented in the learned labeled dataset.
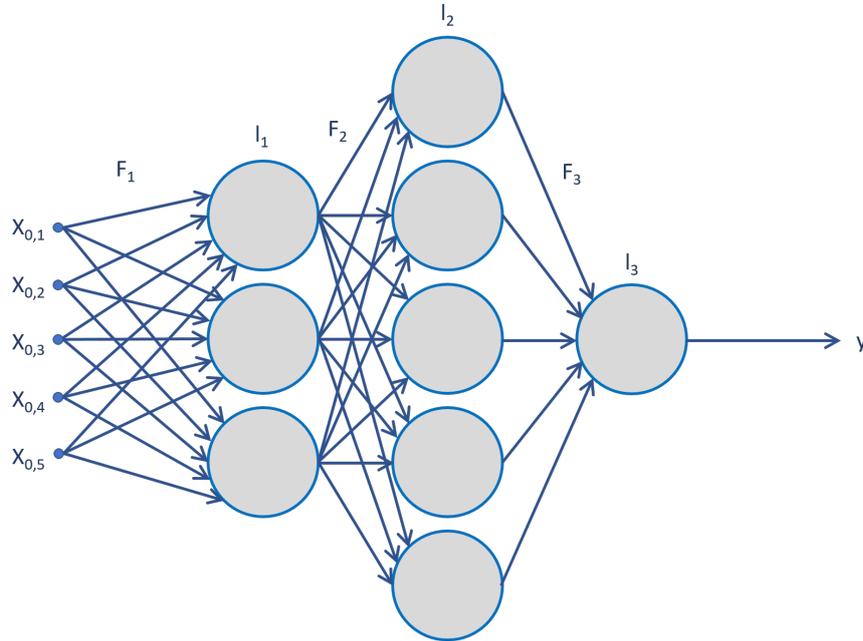
Figure 2.11: Illustration of a multilayer perceptron. The example inputs $x_0 = \{x_{0,1}, x_{0,2}, x_{0,3}, x_{0,4}, x_{0,5}\} \in X$ are mapped to an output $y \in Y$ using three layers l.

2.3.2 Feed-Forward Neural Network

Feed-forward networks emerged from the machine learning concept called perceptron, which was developed by Frank Rosenblatt [169], inspired by [142]. The perceptron estimates an output $Y = \{0, 1\}$ from a linear combination of binary inputs $X = \{x_1, ..., x_n\}$ followed by a nonlinear function h:

$$y = h\left(\sum_j w_j x_j\right). \tag{2.41}$$

Weights $w = \{w_1, ..., w_n\}$ determine the importance of an input to the output to the final output value $y \in Y$. An illustration is shown in Fig. 2.10.

Multilayer perceptrons (MLPs) combine several layers l of perceptrons to approximate a function to map a sample from X to an output from Y. A fully connected network is

shown in Fig. 2.11, where each neuron (perceptron) is connected to every neuron from the previous layer. The output of a single layer l is represented as:

$$F_l(x_{l-1}) = h \circ a_l(x_{l-1}), \quad \text{with} \tag{2.42}$$

$$a_l(x) = w_l x + b_l \quad \text{and} \tag{2.43}$$

$$h_l(x) = (h_{l,1}(x_1), ..., h_{l,K}(x_k)), \tag{2.44}$$

where $a_l(x) : \mathbb{R}^k_{l-1} \to \mathbb{R}^k_l$ represents an affine function, $h_l(x) : \mathbb{R}^k_l \to \mathbb{R}^k_l$ a non-linear function and $\circ$ denotes the composition of functions. The output of multiple layers is defined as:

$$F(x_0) = (h_L \circ a_L \circ ... \circ h_1 \circ a_1)(x_0), \tag{2.45}$$

where x_0 denotes the input samples. A large value of L describes a deep network. MLPs are a specific class in the group of feed-forward networks. Where networks of this group, contrary to the more specific MLPs, usually use nonlinear activation functions. Using perceptrons, a small change in the weights or the bias of the network can result in a drastic change in the output from 0 to 1. For learning tasks we want a small change in the weights w to result in a small change in the output y. This allows the network to learn by iteratively updating the weights so that in each step the output moves closer to the correct result. To achieve this, usually neurons with a nonlinear activation function h for each neuron z_i are utilized (e.g. sigmoid function shown in Fig. 2.12a, hyperbolic tangent function shown in Fig. 2.12b). We compute the activation value for each neuron as follows:

$$z_i = h \left(\sum_{j=1}^{K} w_{ij} x_j + b_i \right), \tag{2.46}$$

with a bias equal to the negative threshold $b_i = -t$ for each neuron. In case of using a sigmoid h_σ or tanh h_t activation functions (see Fig. 2.12a), the output of a neuron is given by:

$$h_\sigma(z_i) = \frac{1}{1 + \exp(-z_i)} \quad \text{or} \tag{2.47}$$

$$h_t(z_i) = \tanh(z_i). \tag{2.48}$$

For multi-class problems usually the softmax activation function h_s is used, which

(a) Sigmoid function

(b) Tanh function

(c) ReLU

(d) Leaky ReLU

(e) ELU
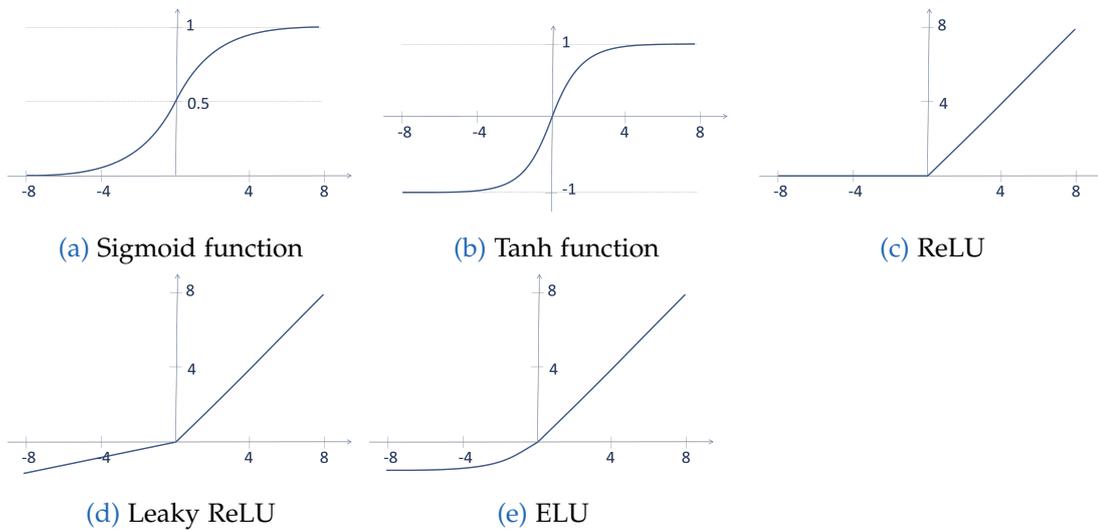
Figure 2.12: Illustration of activation functions. a) Sigmoid function maps real numbers to a range of $[0, 1]$. b) Hyperbolic tangent function maps real numbers to a range of $[-1, 1]$. c) Rectified linear units have slope of 1 for $x > 0$ and 0 otherwise. d) Leaky ReLUs have a range of $[-\infty, \infty]$. e) Exponential linear unit.

normalizes an output vector $Z \in \mathbb{R}^N$ to the range $(0, 1)$, with all K values adding up to 1.

$$h_s(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^{K} \exp(z_j)}. \tag{2.49}$$

Softmax activation functions give probability distribution in categories with a probability for each class.

Rectified linear units (ReLUs) are defined as $h_{ReLU}(x) = \max(0, x)$, as shown in Fig. 2.12c. They are the most frequently used activation functions in neural networks. In [114] it was shown that ReLUs allow a faster convergence of the network compared to using sigmoid or tanh functions, as any negative input is mapped to zero. ReLUs have the drawback that neurons can deactivate when the learning rate was too high, which would cause them to continuously output the same value regardless of the input. A later activation during the learning phase is unlikely, since the gradient remains 0 and the weights won't be changed. This problem was addressed by introducing the variations such as leaky ReLUs with $h_{LReLU}(x) = \max(x, ax)$ for $a < 1$ or exponential

linear units (ELUs), which are shown in Fig. 2.12d and Fig. 2.12e respectively. Leaky Re-LUs have a linear negative function and are superior to ReLUs [136]. Other approaches proposed the learning of the slope in the negative part (parametric ReLUs) or randomly sampling the slope (randomized leaky ReLUs). It was shown in [39] that ELUs can lead to faster learning and to a significantly better generalization performance than ReLUs or leaky ReLUs.

The result of the activation in each layer depends on the previous layers, but not on the following ones. In a forward-propagation step we start evaluating the activations of the first hidden layer and continue until the output layer is reached.

In order to train a neural network a loss function is defined, e.g.:

$$E(w,b) = \frac{1}{2} \sum_{j=1}^{K} ||f(x_j, w_j, b) - y_j||^2, \tag{2.50}$$

where x_n denotes the input vectors and y_k denotes the corresponding output vectors from the network. The weights of the network are denoted with w and the biases with b. These weights are first initialized randomly and then updated based on the learning rate and gradient in each iteration during the training process in order to minimize the loss function. This is usually done with an iterative gradient descent method (see Sec. 2.2.2). The gradients are usually estimated with the backpropagation method.

2.3.3 Convolutional Neural Network

For image processing tasks adjacent pixels are usually correlated with each other. This can be exploited using convolutional neural networks (CNNs) [118]. They are widely used for computer vision applications such as character recognition, object detection and recognition, analysis of medical images (e.g. tumor detection) or depth reconstruction. The difference to feed forward networks is the use of local receptive fields, weight sharing, subsampling and pooling layers in earlier networks. CNNs are arranging the neurons in 3 dimensions (width, height, depth), these layers are also called feature maps. The first layers of CNNs are usually convolutional layers. These

(a) Receptive field   (b) Pooling
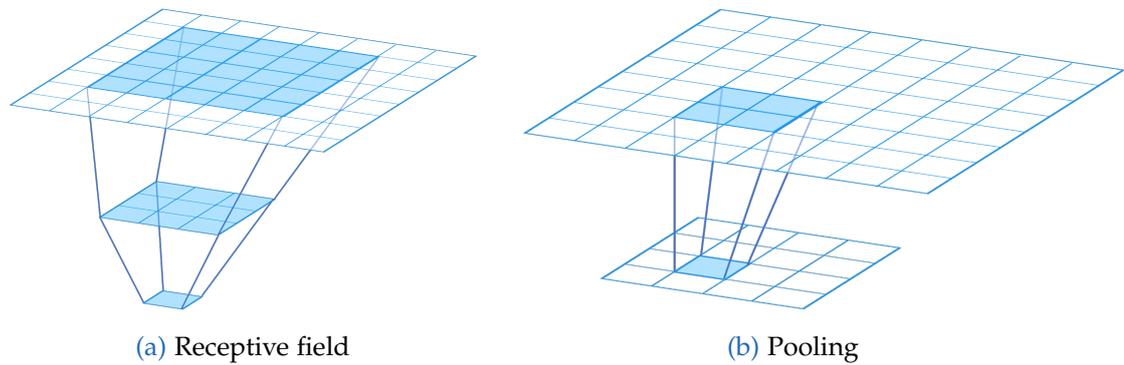
Figure 2.13: Illustration of CNN feature maps. a) The convolution kernel has a size $k = 3 \times 3$. Applying the convolution in a 5×5 input map produces a 3×3 feature map and in the next step a single feature. One feature is therefor related to a defined region in the input map, which is called receptive field. b) Pooling creates a more condensed feature map. Max-pooling takes feature maps from a defined region (e.g. 2×2) and output the maximum activation in the region.

perform convolutions on the image with a filter kernel of a defined size (e.g. the input image has a size of 256×256 and the filter kernel 5×5), where the depth of the filter is equal to the depth of the input. The receptive field of a specific feature is illustrated in Fig. 2.13a. It is defined by the center location and the size of the field. Sharing weights reduces the amount of parameters in the network. The assumption is that if a feature is useful in a specific location, it should also generalize to other locations. Pooling is used to determine if a specific feature is found in the image. It increases the robustness against minor object transformations. It describes the grouping of local features from adjacent pixels, as illustrated in Fig. 2.13b. A common method is max-pooling which takes the maximum response from the chosen feature region. L2-pooling takes the square root of the sum of the squares of the activations in the feature region. Fully connected layers are usually used for the last layers. In the case of a classification problem, they form the output of the previous layer to a probability for each output class. For example, a network which predicts an image to be a car has high values for the activation maps which represent high level features such as wheels or doors. The fully connected layers take the high level features which strongly represent a specific
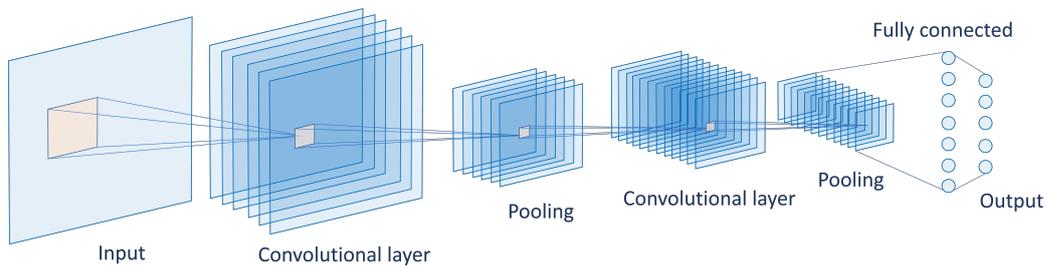
Figure 2.14: Illustration of a convolutional neural network.

object class. The weights are set in a way that the product between the weights and the features in the preceding layer gives the probabilities for each object class.

A CNN can be trained by backpropagation which is also used for feed-forward networks. An example of such a network is shown in Fig. 2.14, comprising two convolutional layers followed by pooling layers. The last layer is fully connected, where all neurons between two consecutive layers are connected to each other.

CNNs exploit the 2D structure of input images. They are often organized in multiple layers which are alternating between convolutional- and pooling layers, which leads to a deep convolutional architecture. Convolutional filters in the first layer represent low-level features and represent e.g. edge filters. In higher layers more complicated structures can be learned by the network. Examples of features learned in 3 different layers (low-level, mid-level, high-level) are represented in Fig. 2.15.

(a) Low-level (b) Mid-level (c) High-level
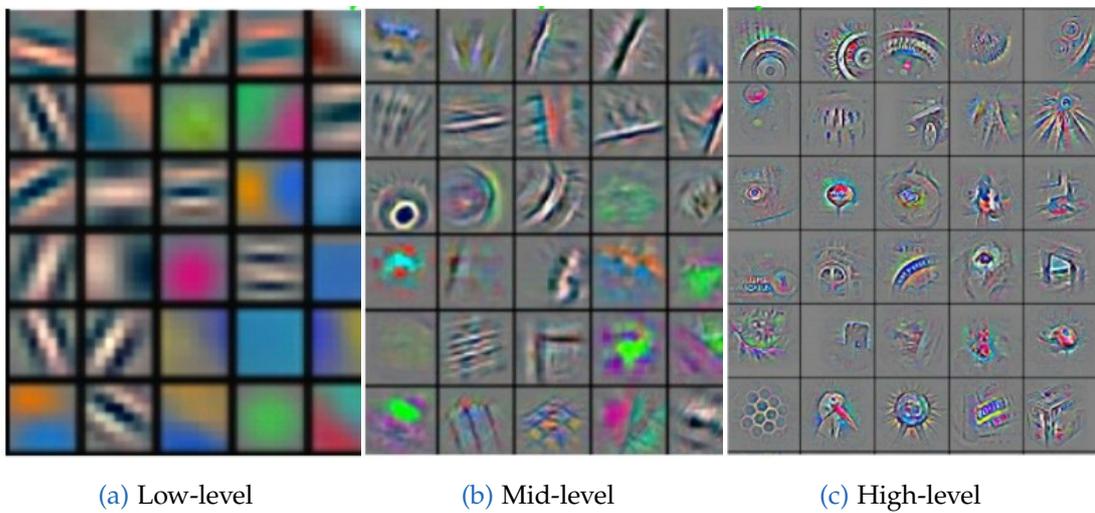
Figure 2.15: Illustration of features learned by a CNN [124]. The features are extracted from three different layers on the network. Low-level features are formed in the first layer and high-level features in the last layer of the CNN.

3 Light Field Imaging

Contents

In this chapter we will introduce several aspects of light field imaging. First we present the current state-of-the-art in Sec. 3.1, then we discuss the acquisition of light field data in Sec. 3.2. We describe the calibration of capturing devices with a focus on industrial applications in Sec. 3.3. Here we build on our calibration of matrix cameras presented in [18] as well as on our calibration of multi-line scan systems presented in [19, 27]. Finally, we discuss the depth reconstruction from light field in Sec. 3.4. This is related to our work based on the 3D reconstruction in multi-line scan systems presented in [5, 6].

3.1 State-of-the-art

The modern interpretation of the flow of light through space as fields, similar to magnetic fields, traces back to Micheal Faraday's lecture to the Royal Society in 1847 [53]. This idea was not further formalized by him. The concept of light fields was first defined by Arun Gershun in his well known paper in 1936 [70], where he defined the flow of light through each point in space. He identified that light shows a smooth
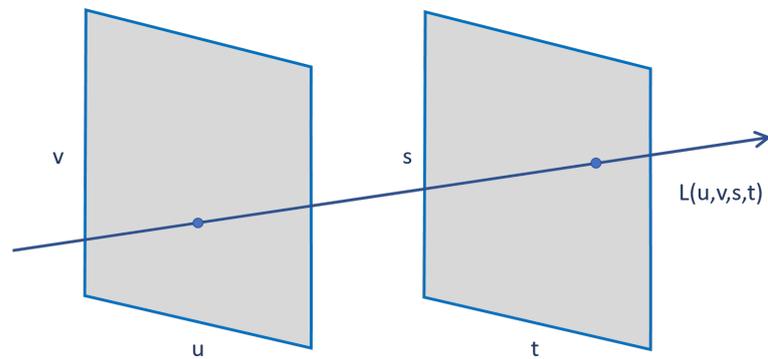
Figure 3.1: Light field parametrization example. Light rays are defined by their intersection of two regions, where (u,v) represents the first plane and (s,t) the second plane.

variation in spacial distances, which was characterized using calculus and analytic geometry before the time of digital processing and hence without the possibility to measure light fields.

Today we have the technology to capture and analyze light fields. They are used for computational imaging tasks as refocusing images after the acquisition, reconstructing the 3D shape of a scene out of a single camera shot, analyzing the surface material structure from a photograph or changing the viewpoint after an image was acquired.

Light fields can be parameterized by a 5D plenoptic function $L(x, y, z, \phi, \theta)$, which defines the position (x, y, z) and orientation angles (ϕ, θ) of the radiance along all rays in space with a fixed arrangement of lights [121]. Since the radiance along each ray remains constant, the function contains one redundant dimension. This results in a 4D light field function which can be denoted by $L(u, v, s, t)$, where the light rays are parameterized by their intersection of two planes as illustrated in Fig. 3.1.

Light field cameras capture a scene from different viewpoints. This can be achieved by camera arrays, where several cameras are placed next to each other and each camera has a different viewing perspective on the scene. An example of a light field camera is the acclaimed Stanford camera array introduced in [219]. Plenoptic cameras consist of a single camera sensor and usually have a microlens array placed

in front of the sensor to acquire angular dependent reflectance information. This system was proposed by Adelson [1] and implemented by Ng [151]. Another way to acquire light fields is moving a single camera in order to acquire several images, while observing a static scene [122] [74]. Four single cameras were used to capture a scene in [140] while the processing tasks were distributed over five computers. A dome of cameras was introduced in [107], where each camera is fixed on a dome structure and captures the scene from a different viewpoint. Coded aperture cameras use a patterned occluder within the aperture of the camera lens [120]. The radiation is blocked in a specific pattern and casts a shadow on the camera sensor. The original image can be reconstructed computationally in addition to the depth of a scene. Multi-line scan cameras, as described in [189], allow high speed and highly accurate light field acquisitions for in-line applications. From a camera sensor several lines are acquired while an object is moving below the camera on a conveyor belt. Each sensor line gives a different viewing angle on the acquired scene.

Applications for light field cameras include the depth reconstruction of scenes, computational refocusing and the dynamic depth of field rendering.

In this chapter we discuss light field acquisition devices and their calibration, as well as applications and the computational 3D reconstruction.

3.2 Acquisition Devices

Light fields can be acquired in several ways, each of which show different properties in precision, baseline, speed and calibration. In this section we present various camera types and setups.

3.2.1 Matrix Cameras

Light fields can be captured by a matrix of conventional cameras, as illustrated in Fig. 3.2a. Each camera captures a slice of the 4D light field function with a specific position and orientation. An early linear array of cameras was used in [201], which

(a) Matrix camera

(b) Plenoptic camera

(c) Ring camera

(d) Coded aperture
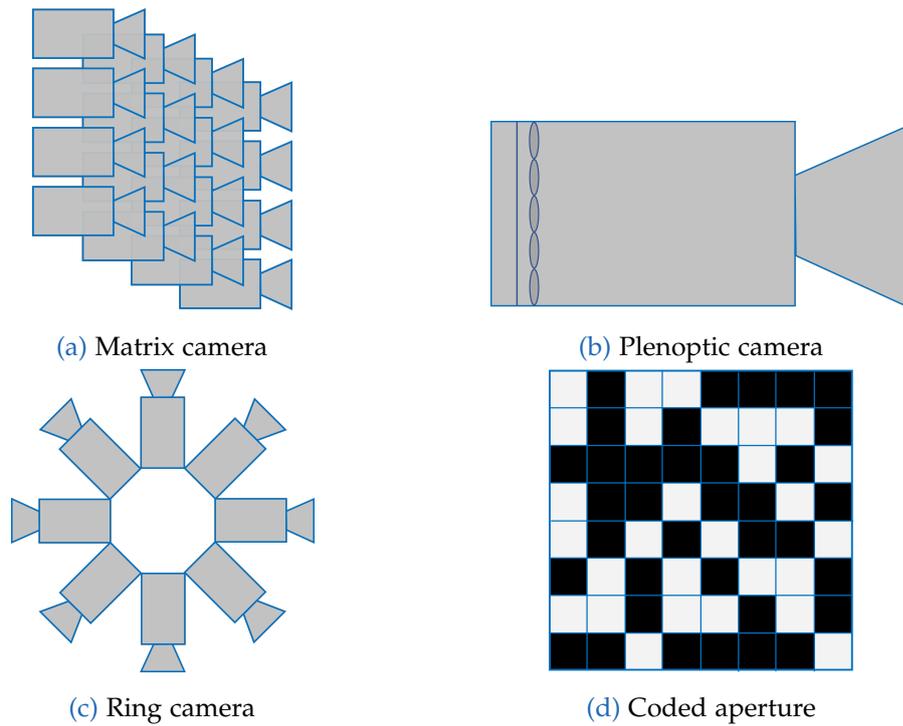
Figure 3.2: Illustration of light field camera concepts. a) 4x4 matrix camera. b) Plenoptic camera, where a microlens array in front of the sensor splits light rays to provide directional information. c) 360 degree ring camera. d) Patterened occluder as used for a coded aperture.

had the aim to create the illusion of a linear camera movement by iterating through the views of a static setup. A camera array holding 64 cameras was introduced in [230], the work focused on creating a scalable light field system designed to use a minimum data bandwidth. A self configurable array with 48 cameras as introduced in [237]. Using a color consistency score allowed a real time calibration of the cameras during motion. The prominent Stanford camera array was then introduced in [219] for computational imaging tasks such as increased resolution, dynamic range or simulating a larger aperture. The setup consists of 100 cameras and allows the capture images or video data.

A compact matrix camera for consumer devices was presented in [208]. This passive camera improves previous approaches such as [196, 197] in image quality. It supports
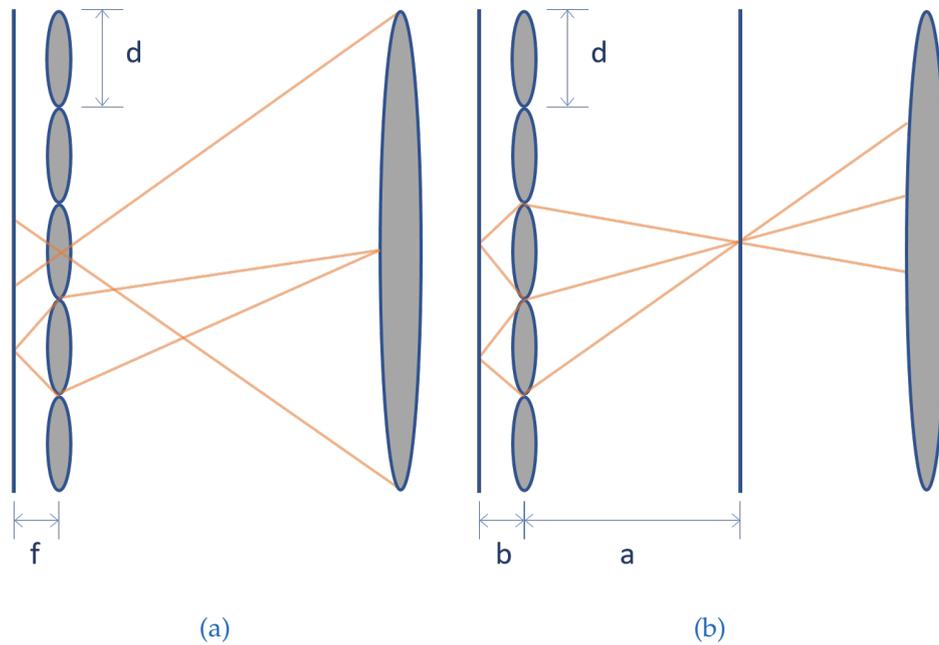
Figure 3.3: Examples of plenoptic cameras. a) Plenoptic 1.0 camera. The main lens is focused on the microlens array and the microlens array is focused at optical infinity. b) Plenoptic 2.0 camera. The microlenses are focused at the image plane of the main lens.

still images and video and low light acquisitions, due to its thin and portable structure it is well suited for use in mobile phones.

Xapt introduced a 4x4 matrix camera [227]. The camera showed to be apt for industrial inspection tasks [19]. We will discuss the calibration of matrix devices using the example of the XApt camera with a focus on industrial acquisitions in Sec. 3.3.2.

3.2.2 Plenoptic Cameras

Standard 2D cameras capture irradiance from different directions at a defined viewpoint. Plenoptic cameras have a microlens array placed in front of the camera sensor, with which they capture 4D light field data. This is illustrated in Fig. 3.2b. In 1908 the concept of microlens arrays was introduced by Lippmann [130] and later refined by

Ives [99]. Adelson [1] introduced a camera in 1992, where the main lens was focused on the microlenses, which is now known as plenoptic 1.0. The concept is illustrated in Fig. 3.3a. This idea had been considered and improved by [57, 133, 152], which resulted in plenoptic 2.0 cameras. The concept is illustrated in Fig. 3.3b, where the microlenses are placed in a distance b from the sensor and the microlenses are focused on an image plane in distance a. This approach allows the increase of spatial resolution, which is given as $\frac{b}{a}$ of the sensor resolution. Further increasing the resolution computationally and inferring super-resolution images from plenoptic 2.0 cameras was introduced in [67]. The Lytro plenoptic camera was commercially available in 2011 followed by a second generation in 2014. Raytrix produces light field cameras for industrial applications, with a focus on the production of plenoptic 2.0 cameras.

3.2.3 Coded Aperture Cameras

Another method to acquire light field data with a conventional camera and a single sensor is using coded apertures. A patterned occluder (see Fig. 3.2d) should reside in the aperture of a standard camera. The known obstruction pattern lets less light to the sensor and creates a different point spread function. The closer the object moves to the focal plane, the smaller the pattern projected on the sensor becomes. The resulting image on the sensor has to be decoded to produce a final image.

A coded aperture with a Fresnel zone plate was introduced in 1961 by [143]. In 1978 the idea was extended to using uniformly redundant arrays by Fanimore and Cannon [56]. These arrays are nonrandom patterns of openings with specific features. Striving for the reduction of noise in the image. Wavefront coding was used in [28, 46] in order to get an increased depth of focus and increase the depth of field. These early coded aperture approaches recover either the depth or the standard image. The depth reconstruction combined with the retrieval of the standard image using a coded aperture camera was shown in [120].

3.2.4 360 Degree Cameras

Recently the growing interest in virtual reality (VR) applications led to an increasing development of 360 degree light field cameras. These setups should allow rendering stereo views and depth cues over a 360 degree capture of a scene. To acquire a full circular viewing range, cameras are usually placed on ring structures as illustrated in Fig. 3.2c. Such models have been introduced by e.g. Google [3] and Facebook [54], which also allows rendering scenes from a single vantage point. Algorithms were extended to facilitate a limited head-motion range in [129], which is an important factor in VR applications. A second generation camera was recently introduced, where the cameras are placed on a sphere for an extended viewing range. Other 360 degree camera setups were introduced which are using a mirror rig to capture panoramic videos for immersive applications [71, 195, 217].

Since we focus on light field cameras which can be applied in industrial inspection tasks, acquisition devices used for gaming or immersive applications will not be further considered.

3.2.5 Multi-Line Scan Cameras

Single line camera sensors have already been used commercially since the introduction of fax machines on the market. Today, line scan cameras are frequently applied in industrial inspection tasks and embedded in commercial camera bodies. Contrary to standard line scan cameras, multi-line scan devices can capture light field data and consequently allow computational imaging applications. This is equivalent to using multiple line scan cameras observing a scene from a different position, which was previously shown in [190].

We use multi-line scan light field for 3D inspection of fine surface structures. The setup was previously described in [5, 190], and an illustration is shown in Fig. 3.4. From an area-scan camera a chosen number of sensor lines is extracted. Each sensor line allows a different viewing angle on the scene, where an object is moving in a defined transport direction with a constant speed on a conveyor belt. Each sensor line observes
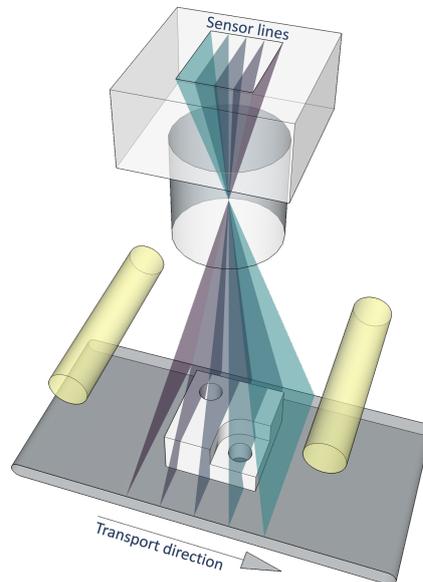
Figure 3.4: Illustration of a multi-line scan camera. The object is moving on a conveyor belt in the transport direction under the camera sensor, illuminated by two linear light sources. Each sensor line acquires an image from a different viewing angle over time.

the conveyor belt in a different viewing angle and captures the object at a certain position. For each viewing angle, by sampling single lines over time, a full image will be generated from the object. This allows acquiring light field data in real-time and in-line, which is crucial for industrial inspections.

We introduced calibration approaches for multi-line scan cameras in [19, 27], which we describe in more detail in Sec. 3.3.4.

3.3 Calibration

3.3.1 Concept

The calibration of light field systems is important for applications such as precise depth estimation, 3D reconstruction, digital refocusing or reflectance analysis. Calibrating the

combination of a camera with a lens starts with the acquisition of a calibration target, such as a checkerboard pattern or a grid of dots with known dimensions. This allows the calculation of the extrinsic and intrinsic camera parameters and consecutively the rectification of the image. Extrinsic camera parameters describe the camera position and rotation, while intrinsic camera parameters describe the geometric property of the camera, defined by the focal length, the principal point offset and the axis skew.

Multi-view camera systems additionally require a global optimization procedure, such as a bundle adjustment step, in order to bring all cameras in the array in one common coordinate system.

Plenoptic cameras have an additional parameter, which is the distance of the lenslet array and the camera sensor. These cameras also require a decoding step prior to the calibration. The precise placement of the lenslet array is unknown and the position of the center of all lenslets has the be determined before each view can be separated. In the rectification process for plenoptic cameras, it has to be considered that all views are captured on the same sensor and with the same main lens.

Multi-line scan cameras capture objects which are moving under the camera sensor. In industrial applications, these objects move on conveyor belts in a defined direction. The calibration of such systems has to take the transport jitter into account, as well as the unique structure of an image, which is formed from a specific view and thereby from a specific sensor line. This means, that each sensor line has the same distortion from the main lens along the transport direction.

In the following section we will discuss the calibration of those methods which are relevant for industrial applications.

3.3.2 Matrix Cameras

Multiple views of a scene can be captured by matrix camera systems, where all cameras are fixed on a plane in a defined distance from each other. Concepts of multi-camera calibration systems have previously been presented in [156, 192, 205]. An efficient space-sweep approach has been proposed in [41], which showed a linear algorithmic
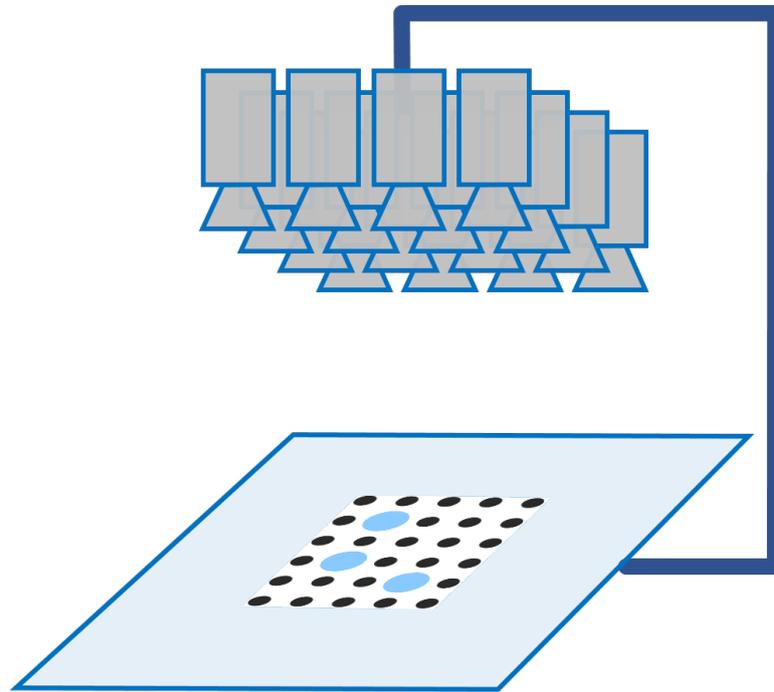
Figure 3.5: Illustration of our experimental setup with a 4x4 matrix camera. The calibration pattern with a three dot central pattern is described in Sec. 3.3.4.

complexity with the number of taken images. In [75] Graber et al. proposed a system for the interactive reconstruction of a scene from monocular video making use of convex optimization.

In this section discuss a multi-camera array calibration approach which we introduced in [18]. This calibration is targeted for the use of matrix cameras in industrial applications with an arbitrary multi-camera geometry. An example of our experimental calibration setup with a 4x4 matrix camera (such as the Xapt Eye-sect XA) is shown in Fig. 3.5.

First, each camera is calibrated with a calibration target to identify the intrinsic and extrinsic camera parameters [38, 240]. Intrinsic camera parameters include the focal length, principal point, camera skew, radial distortion and tangential distortion coefficients. Extrinsic camera parameters comprise the position and the rotation of the camera in a global coordinate system.

point in space

reprojected

detected

reprojected detected

detected

reprojected

camera 1
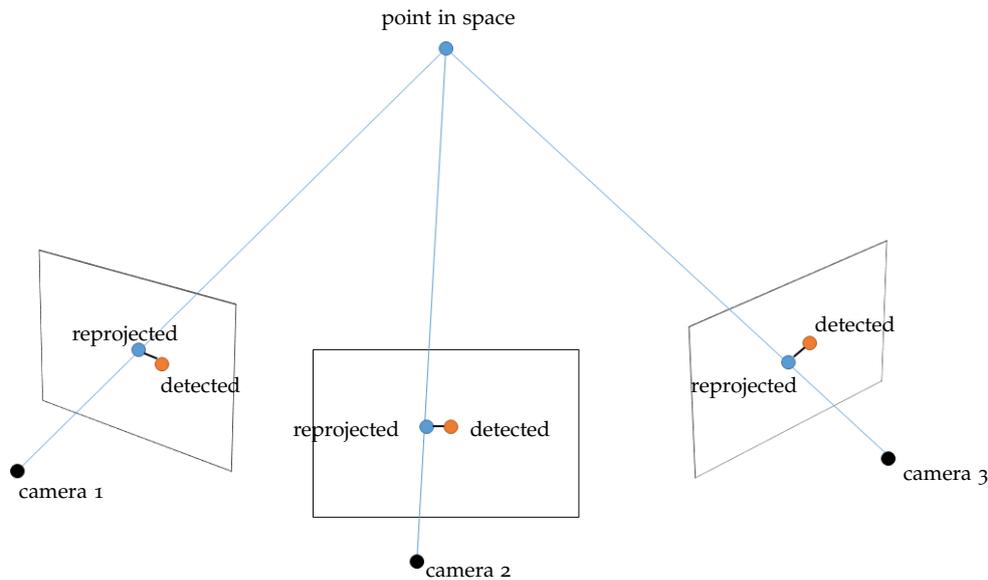
camera 2

camera 3

Figure 3.6: Illustration of the reprojection error of the matrix camera calibration. The sum of the difference of the error between the detected points and the reprojected points for all cameras is called bundle adjustment.

Second, the reprojection error is minimized by an optimization procedure as illustrated in Fig. 3.6. An initial pairwise multi-view correspondence analysis is followed by a bundle adjustment routine, where the reprojection error is optimized over all cameras and image points [63].

Third, the images are rectified making use of the obtained calibration model in order to allow further computational imaging applications. Usually all cameras point to different directions and their locations are seldom coplanar, hence we cannot use standard stereo image rectification procedures as described in [64, 111]. The camera views have to be reprojected to one common regression plane, as discussed in [41]. This allows simple translation and scaling operations instead of otherwise costly warping procedures for the cross-comparison between multiple images. The image manipulation procudure is reduced to a simple translation between multiple cameras in case all camera centers are coplanar and the chosen regression plane is parallel to the camera plane. This gives a significant computational and algorithmic advantage compared to the standard stereo vision approach.

The rectification of a matrix camera through calibration is shown with an example

(a)

(b)
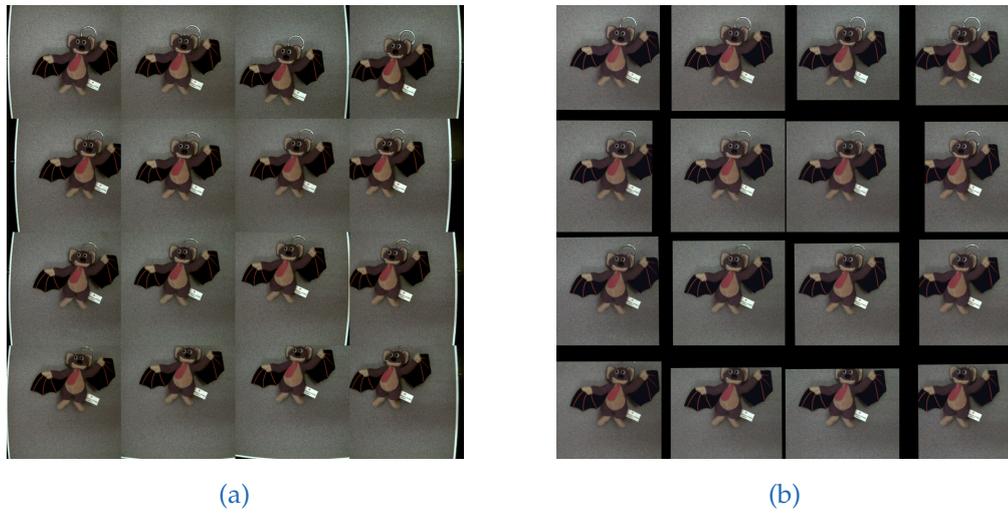
Figure 3.7: Matrix camera calibration example [18]. a) Input image acquired with the Xapt Eye-sect XA camera. b) Undistorted rectified image.

from the Xapt Eye-sect XA camera in Fig. 3.7.

3.3.3 Plenoptic Cameras

Plenoptic cameras have a microlens array in front of the camera sensor, which allows encoding the direction of incoming light rays and thereby multiple views on a scene, which we described in Sec. 3.2.2. A plenoptic camera model has been derived with ray transfer matrix analysis in [68] to describe the light path through the system. The work of [45] refined the model, including the effects of lens distortion and the projection through the microlens array.

The rectification process for light field can be implemented by sampling the light field data on the sensor in such a way, that the lens distortion is reversed. The calibration of plenoptic 2.0 cameras was described in [102]. Here the intrinsic and extrinsic camera parameters were estimated by optimizing an energy formulation, which is based on a thin lens equation. The approach allows the correction of radial lateral and radial depth distortions.

(a)   (b)
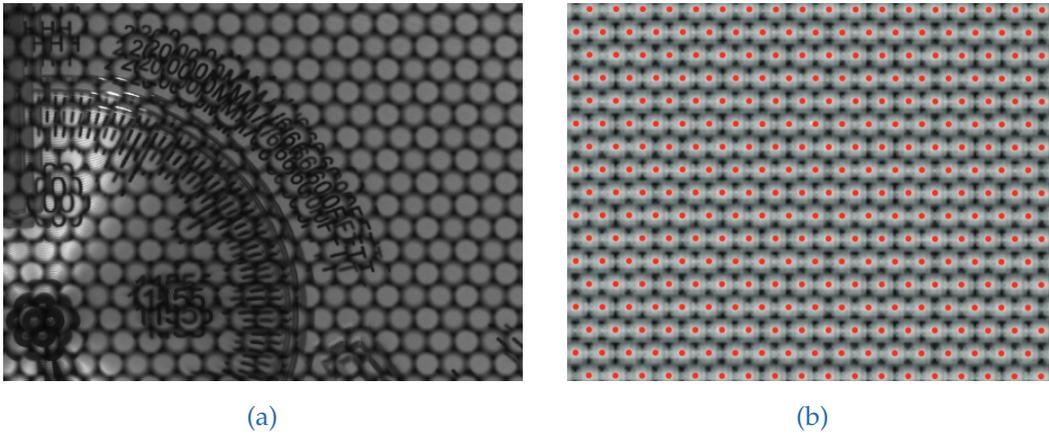
Figure 3.8: Plenoptic camera example images. a) Image acquisition with a plenoptic
2.0 camera [102]. b) White image acquisition with microlens centers [45].

The input for each calibration for plenoptic cameras is the raw 2D microlens image
from the sensor (see Fig. 3.8a), which is decoded into a 4D light field image. Therefor
the precise position of each microlens is determined by analyzing a white image
acquisition and optimizing the position of the center of each microlens (see Fig 3.8b).
Converting the hexagonal grid to a regular image grid was explained in [42]. Each pixel
position in a microlens corresponds to a spatial ray. For all well-visible pixel points in
the microlens regions an image with a defined view on the scene can be sampled. Initial
intrinsic and extrinsic parameters are estimated with a standard camera calibration
process for each sampled image. The lens distortion model is computed including
decentering, radial distortion and the undistortion of the ray directions.

3.3.4 Multi-Line Scan Cameras

Multi-line scan cameras are very well suited for high speed industrial inline inspection
applications. These cameras capture multiple views on one sensor by acquiring images
from several sensor lines over time, as described in Sec. 3.2.5.

A calibration for multi-line scan cameras was introduced by [19]. Since a highly
accurate camera calibration is especially important for high magnification acquisitions,
additionally a high precision calibration target was introduced. Before, these cameras
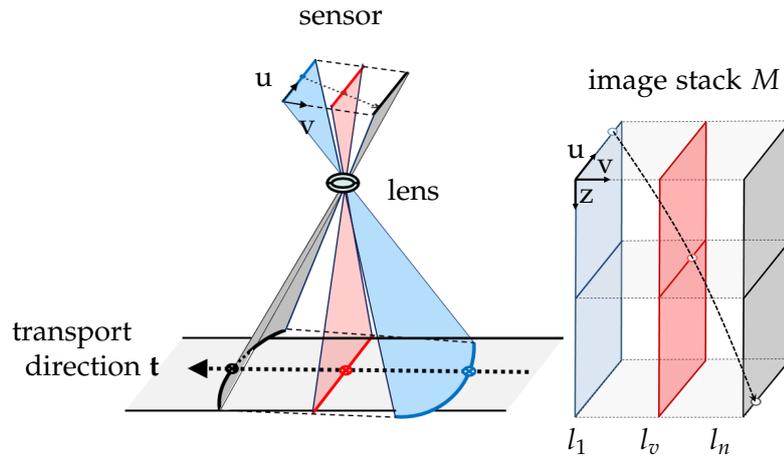
Figure 3.9: Multi-line scan camera calibration [19]. The setup (left) shows a camera sensor rotated w.r.t. the transport direction of the transport stage. The views which are captured in that setup are saved in an image stack (right). As depicted, in an uncalibrated system object points can travel along curves in the image stack.

were usually equipped with high quality industrial lenses, without a calibration tailored to the multi-line camera properties. In a standard multi-line scan setup, an object is moving on a conveyor belt below the camera. The synchronization of this transport stage and the camera is crucial for highly precise measurements. A computational correction of transport jitters was introduced in [27].

Camera Calibration

The multi-line scan camera calibration, as introduced in [19], determines the intrinsic and extrinsic parameters of the camera for image rectification in an industrial setup. Previously several approaches were introduced for the calibration of single-line scan cameras, using structured light [126, 134], an additional laser finder [239], or the assumption that the transport is orthogonal to the camera [35, 77, 96].

Internally the multi-line scan camera is using a fast area sensor. Hence it can be configured to read out all sensor lines, which allows using standard calibration approaches [23, 25] to calculate the intrinsic camera parameters.
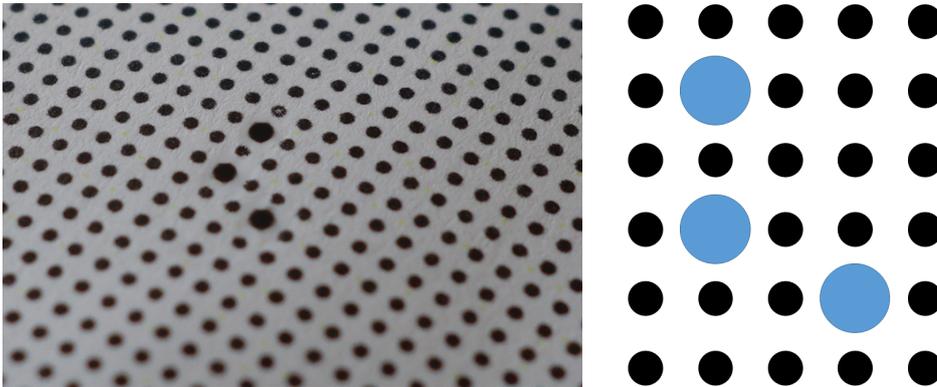
Figure 3.10: AIT calibration target [19]. To recognize the pattern, only the central three dots need to be visible.

For the remaining calibration n sensor lines $(l_1, \cdots , l_n)$ are sampled, while an object is moving below the camera in a defined transport direction t (see Fig. 3.9). Each sensor line corresponds to a specific viewing angle on the scene. Note that the sensor distortion is constant in z-direction, since the lens distortion is constant over time for the same scan line. The camera cannot be assumed to be perfectly aligned with the direction of the transport stage, which leads to a perspective warping between arbitrary lines l_v and l_{v+1}. After determining the intrinsic and extrinsic camera parameters, including the focal length, the principal point and the distortion vector - the original image is warped to a virtual camera plane, undistorted, resampled w.r.t. the u-coordinate and rectified.

Since there is no guarentee that the whole caibration pattern will be visible, especially at high magnifications, a new high precision grid distortion target was introduced in [19] (see Fig. 3.10). This marker is robust and easy to detect, while providing information about the orientation of the pattern and image mirroring. For further details about the specific multi-line calibration approach consider [19].

Transport Jitter

Transport jitter occurs for moving objects when the transport stage is not triggering the camera perfectly, which leads to uneven distances between the scene samples.
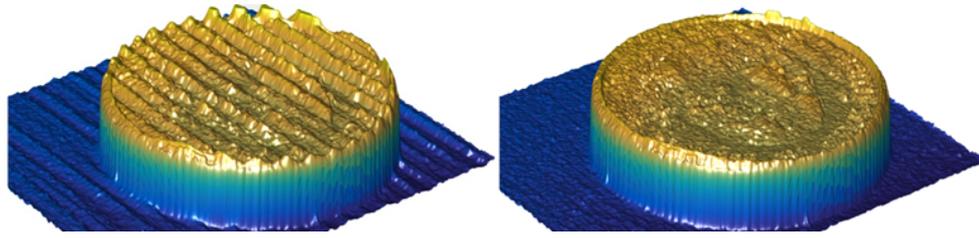
Figure 3.11: Transport jitter artifacts effects in 3D reconstruction [27]. The depth map resulting from the uncorrected input is shown left and the depth map after a jitter correction on the right.

Especially for high precision measurements at high magnifications this has a serious impact. When using the data for 3D reconstruction the transport jitter causes ripple artifacts in the depth map, as shown in in Fig. 3.11. There are several approaches to reduce transport jitter, such as letting the transport stage trigger the camera or capturing an optical measurement marker additionally to the transported objects.

In [216] the motion error was compensated with a closed-form expression on a pixel level in the 3D reconstruction. A smartphone based approach was introduced in [115], where the front camera captures a control pattern while the back camera acquires images from the scene. Internal reflections were used in [131].

An optimization approach with a pixel warping function was introduced in [27], which corrects the jitter using light field data of the scene and is tailored to multi-line scan data correction. It works under the assumption, that the backwards and forwards disparity should be equal between different views. Additionally the result is penalized for deviations from the original solution and smoothness is enforced. These terms are formulated in an energy function and optimized with a standard least squares approach.

The transport jitter correction part in the calibration of multi-line scan systems is especially important when the transport stage cannot be perfectly synchronized with the camera, high magnifications are used or there is a need for highly precise measurements.

3.4 3D Reconstruction

3D reconstructing methods using light fields have been around for decades. It became popular due to the works of Bolles *et al.* [21, 22], which also introduced the analysis of light fields using epipolar image plane (EPI) structures. They use line fitting approaches to find linear structures in these EPI stacks, which was frequently adopted. The analysis of EPI stacks was used for densely sampled, high resolution light field data in [109, 110], where smooth regions are handeled in a hierarchical approach.

Techniques from classical stereo matching were regularly used for the analysis of light field data. Robust patch-based methods were developed in [17, 65, 206, 237]. Clustering approaches were used in [12, 61], where points with a matching depth range are clustered together. Venkataraman et al. [208] used pattern matching between different views, i.e. for a discrete number of hypothesized depths the sum of absolute differences (SAD) of radiances between different views is calculated.

Depth from focus techniques were presented in [76, 162], which were customized for light field data in [128, 199, 200]. Those techniques show inaccuracies at boundary regions and the precision is limited by the size of the aperture.

Shearing approaches describe finding the correct lines by shifting the EPI stack (see Fig. 3.12) for all possible depth hypothesis. The correct depth is found, where in a vertical line (over all views) the image values correspond best to each other. The shearing can cause depth estimation errors for objects which are in a far distance from the focus plane [198], where large shifts are necessary.

Optimization approaches to calculate depth from light field data were presented in [87, 125, 188]. A variational approach to estimate the depth in linear EPI structures via analysis of dominant local orientations through structure tensors was introduced in [213].

In this section we describe the structure of the EPI representation of light field stacks as well as a light field depth reconstruction approach.
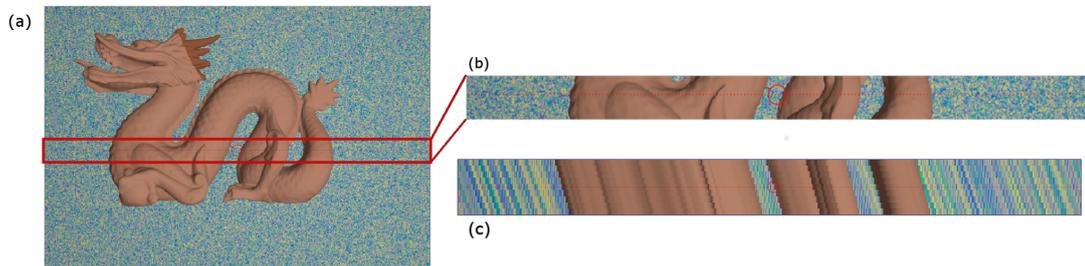
Figure 3.12: EPI stack illustration [6]. (a) A Stanford dragon object [186] scanned with a simulated light field camera. (b) Zoom of the red marked region. (c) EPI stack along the red dotted line.

3.4.1 EPI Representation

Consider the 4D light field representation as described in Sec. 3.1, which consists of two directional and two spatial dimensions describing the propagation of light through space. These light fields can be analyzed in the EPI domain. All light field images which are usually acquired with a specific illumination direction form a light field stack. A cut through this stack shows linear structures, where the slope angles correspond to a defined distance to the camera of the corresponding object point in the scene. Fig 3.12 shows an EPI structure from one image line, constructed using 9 camera viewpoints.

The depth of each point in the scene can be found by solving a correspondence problem between all views to find the correct disparities (slope angles), and thereby depth values.

3.4.2 Slope Analysis

The calibration of light field cameras was described in Sec. 3.3. Analyzing the slopes in EPI stacks (see Fig 3.12), allows for a well calibrated camera a pixel-wise depth reconstruction. This is contrary to standard stereo methods, which need to analyze neighboring regions to reach a reasonable matching confidence. A refined result can

(a) No occlusion   (b) Occlusion
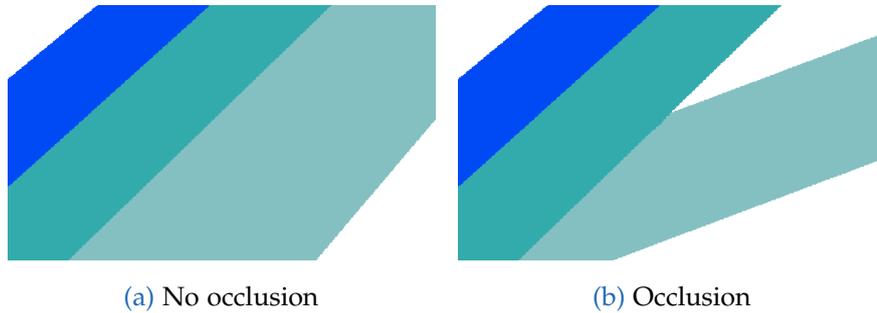
Figure 3.13: EPI correspondence with occluded objects.

Figure 3.14: EPI stack with uniform color. Uniform color (as in the red marked region)
leads to an ambiguous slope analysis result. The red lines mark the central
view and the center of the image position.

be achieved by pyramid analysis methods as shown in [109], as well as by tensor or
variational approaches [213].

The 4D light field representation $L(u, v, s, t)$, as described in Sec. 3.1, holds the intensity
of a light ray parameterized by two planes. This can be analyzed in the EPI domain
(shown in Sec. 3.4.1) by fixing two coordinates, namely either (v, t) or (u, s) which
results in a horizontal or vertical EPI stack respectively.

Occlusions effect the matching result, which has been previously discussed [34, 141,
200, 210]. An example is shown in Fig. 3.13. When no occlusion is present, the
assumption that depth values of the slopes can be analyzed over all $k \in n$ views does
not hold anymore. This influences the reliability of the matching result.

When objects have little or no surface texture (as shown in Fig. 3.14), depth from
light field additionally leads to ambiguous results. Using additional measurements
such as photometric stereo (which we describe in Sec. 4) can give highly confident
results in those regions to compensate the restrictions of multi view imaging. A
confidence measure can be used to determine if a calculated depth value is reliable.

Blur estimation was used with a tensor approach in [229] to estimate the confidence of a light field measurement. Local and global confidence metrics were introduced in [34]. Where the local confidence gives more confidence to the depth values estimated in textured regions and the global confidence was computed by the gap between the global depth estimation and the second best local solution, the smaller the gap the lower is the confidence. The depth for all pixels with a low confidence was marked as unknown. A peak ratio measure was introduced by [94], here the best and second best correlation values are compared, where a low distance between the two points indicates a lower confidence.

A correspondence measure, as shown in [199], can be applied in the EPI stack by shearing the stack $L(u, v, k)$ in an angle α and computing the variance for a given pixel position (u, v).

$$L_\alpha(u, v, k) = L(u + k(1 - \frac{1}{\alpha}), v, k) \tag{3.1}$$

$$\sigma_\alpha(u, v)^2 = \frac{1}{n} \sum_{k=1}^{n} (L_\alpha(u, v, k) - \overline{L}_\alpha(u, v))^2 \tag{3.2}$$

Where k denotes the index of an image from a defined viewing position in a light field stack. $\overline{L}$ is the refocused images for a sheared stack over the angle α and therefor the average over all views in each pixel location (u, v). For robustness the correspondence analysis result can be averaged over a region with a window size w.

$$C_\alpha(u, v) = \frac{1}{|w|} \sum_{u,v \in w} \sigma_\alpha(u, v) \tag{3.3}$$

We need to find a shear angle α which minimized the correspondence measure, which is given by:

$$\alpha_C(u, v) = \arg \min_\alpha C_\alpha(u, v). \tag{3.4}$$
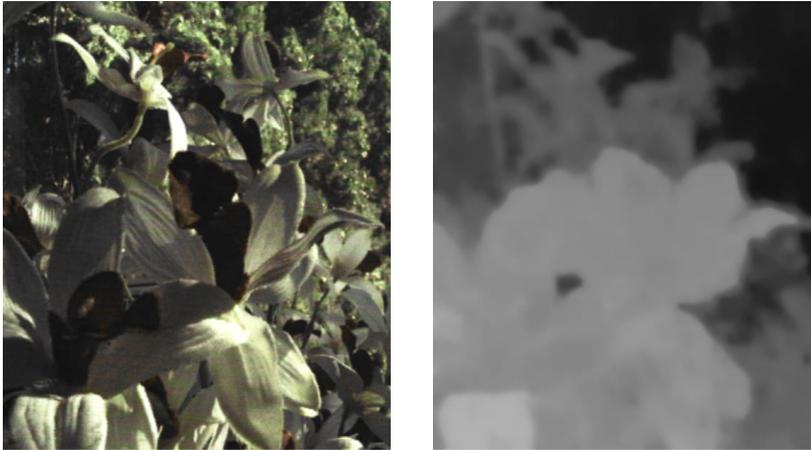
Figure 3.15: Light field depth estimation example [198].

This angle is in linear relation to the depth Z. The confidence for the best result is given by:

$$C_1(u,v) = \min_\alpha C_\alpha(u,v).$$ (3.5)

A simple peak ratio measurement can be implemented by comparing the first and the second best matching result.

$$C_{conf} = \frac{C_2(u,v) - C_1(u,v)}{C_1(u,v)}$$ (3.6)

Where $C_2(u,v)$ denotes the second best correlation value.

An example of a depth result is shown in Fig. 3.15, where several occlusions occur.

Depth estimation with light field data has limitations, lying in the baseline of the capturing system, the angular resolution, and the fact that homogeneous regions cannot be reliably estimated. Especially in scenes with fine details or for surfaces with little or no pattern, the construction result can be greatly improved by combining light field methods with photometric stereo. In the next Chapter we describe photometric stereo acquisition frameworks, calibration approaches and depth reconstruction methods in more detail.

3.5 Conclusion

In this chapter we introduced light field methods with a focus on the application in industrial 3D reconstruction. In Sec. 3.2 we showed light field acquisition devices for area scan applications (matrix/array cameras, plenoptic cameras, 360 degree cameras, coded aperture cameras) as well as for in-line applications (multi-line scan camera).

In Sec. 3.3 we discussed in detail the calibration of camera types relevant for industrial applications, namely matrix/array cameras, plenoptic cameras and the multi-line scan camera. The presented calibration of matrix cameras [18] and multi-line scan systems [19, 27] is based on our previous publications.

We showed the analysis and depth reconstruction of light field data in Sec. 3.4. We presented the 3D reconstruction from light field data for multi-line scan systems previously in [5, 6]. In the next Chapter we will discuss photometric stereo methods and their application in 3D reconstruction.

4 Photometric Stereo

Contents

In this chapter we present the acquisition, calibration and processing of photometric stereo data. First we describe the current state-of-the-art techniques in Sec. 4.1. Then methods to acquire photometric stereo data are discussed in Sec. 4.2, including our method to extract photometric stereo data from light field stacks which we first presented in [9] with a refined formulation in [8]. The calibration of photometric stereo data is explained in Sec. 4.3. Finally, we describe the reconstruction of depth information from photometric stereo data in Sec. 4.4.

The combination of surface normals with depth from light field will be discussed in Chapter 5 in general and specifically for multi-line scan data in Chapter 6.

4.1 State-of-the-art

The term photometric stereo was introduced by Woodham [222] to describe the surface orientations in a scene (see Fig. 4.1). While the viewing direction remains constant, the incident illumination is varied between image acquisitions. This allows one to determine the surface orientations in a scene through the radiance values in
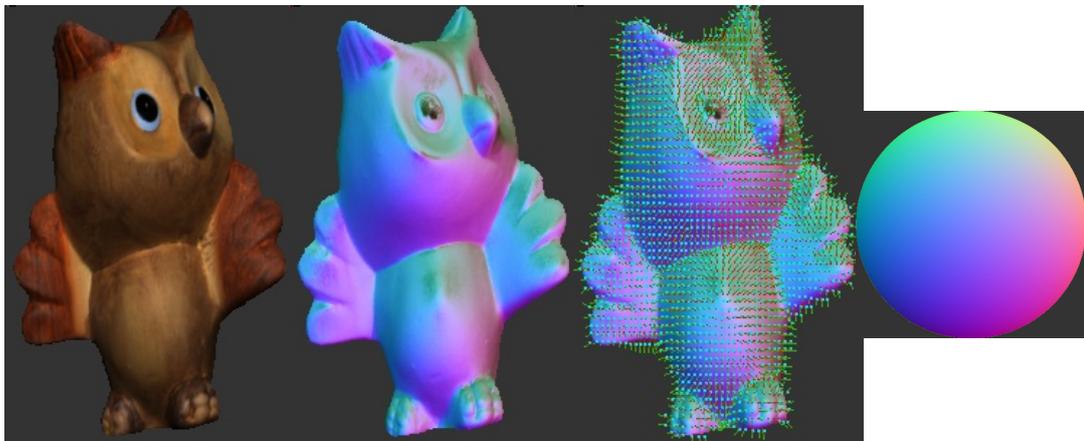
(a) Texture   (b) Normal map   (c) Normal vectors   (d) Reference Sphere

Figure 4.1: Illustration of surface normals [180].

each location of the images. Due to the constant viewing angle over all acquisitions the correspondence between the images is straight forward. In Woodhams original manuscript it was suggested that the surface orientation could be recovered from Lambertian surfaces, when at least three independent illumination directions are used.

The limitation to Lambertian surface reflection was addressed by later methods which introduced methods to handle specular reflections or shadows [11, 40, 182]. For these advanced reconstructions four images were used to search for outliers or determine the roughness of the surface structure. Other approaches modeled the reflection lobes by a linear combination of Lambertian and specular reflections [66, 146, 194].

The acquisition of partial bidirectional reflectance distribution functions (BRDFs) was approached with photometric stereo methods in [31, 72, 90]. We describe BRDFs in more detail in Chapter 7.

Other methods are handling different reflection functions without defining a reflectance model. In [91] the number of possible surface normals is reduced by weak constraints assuming monotonicity, visibility (valid for all normals in a half-space) and isotropy.

The approach presented in [36] uses only dense specularities, contrary to most methods which treat them as outliers.

The methods have in common, that they allow the estimation of surface normal vectors. By integrating the surface normals a depth map can be reconstructed. The resulting depth map is locally sensitive, but globally inaccurate if no further constraints are considered.

Globally accurate methods were previously combined with approaches that show a high local accuracy, e.g. by combining depth maps with surface normals taking into account the frequency domain [5, 148], polarization normals were combined to enhance depth maps in [106], or by combining the depth from RGB-D cameras with shape-from-shading techniques [235].

4.2 Acquisition

Photometric stereo setups allow one to illuminate a scene from different incident illumination directions. This can be achieved by moving the illumination source or by fixed light sources, which are turned on either separately or with light configuration patterns. Additionally, polarization filters have frequently been used in photometric stereo to identify diffuse and specular reflectance structures [106, 135, 145, 221, 233].

Mobile photometric stereo devices were previously frequently discussed [89, 92, 184, 233]. A popular application of portable devices is the field of cultural heritage or the mobile inspection of security structures on documents or banknotes. Ring lights were used in [241] with 20 light sources evenly placed in a circle. An auto calibration method for ring lights was introduced in [178] to recover the positions of the light sources.

Light dome setups have light sources (LEDs) placed in a dome structure, while usually a fixed camera is observing the scene from the top center position. Since the light sources are fixed, the acquisition can be fast and controlled which is necessary in industrial applications. This rigid structure also simplifies the calibration. A portable dome was used in [164, 215, 220], where for each acquisition one out of 256 light
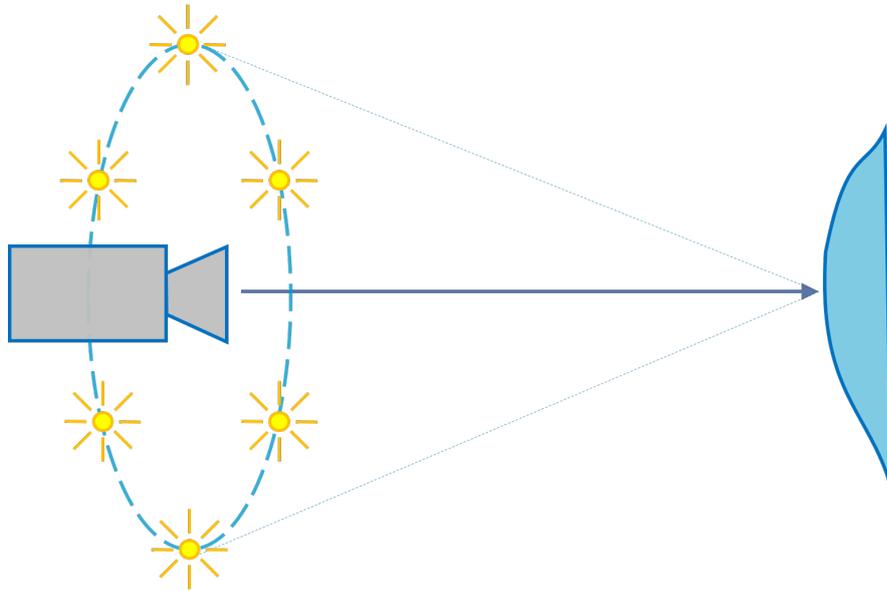
Figure 4.2: Illustration of a ring light photometric stereo setup. The light sources are placed on a ring. In the illustration the light sources are oriented to the center of a cone.

sources was illuminated separately. A light dome setup with 32 illumination sources was used in [6, 185] for industrial applications in depth reconstruction and analyzing diffractive optically variable image devices (DOVIDs).

The analysis of photometric stereo data from a multi-line setup was introduced in [9].

4.2.1 Ring Light

When a bulky photometric stereo setup cannot be used, a mobile ring light can be suitable for taking acquisitions.

For ring lights the position of the light sources is partially known [241], since the light sources are placed on defined positions on a ring. A mobile ring light was shown in [184] for the application of hologram verification. The setup comprises 24 individually operable LEDs and is mounted on a mobile phone with a 12.3 mega pixel camera. During the acquisition the ring module is placed on the object, the small range

(a) Setup　　　　　　(b) Ring light　　　　　　(c) Cross-section
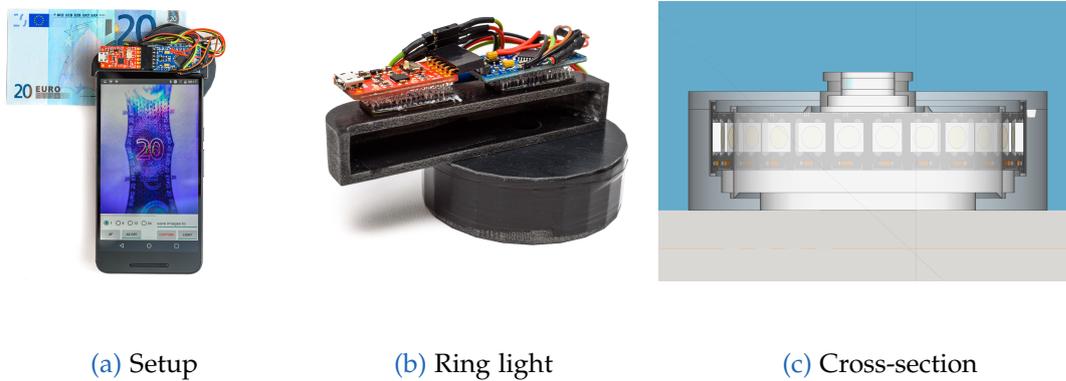
Figure 4.3: Mobile photometric stereo acquisition setup [184]. a) Setup mounted on a mobile phone. b) LED ring light module. c) Cross-section of the device.

(distance to the object) allows large illumination angles. Due to the small range, a macro and wide-angle lens is mounted on the phone. The setup was used to detect face holograms on banknotes by training a convolutional neural network (CNN) to learn relevant feature descriptors.

For non mobile but nonetheless portable applications, a light dome is well suited for photometric stereo applications.

4.2.2 Light Dome

Light domes with a high precision and a portable structure are well suited for industrial environments, where the object can be placed for a certain time at the same position, contrary to an object transported with a constant speed on a conveyor belt. In this section we introduce a specific setup which we used for industrial experiments.

The high performance light dome setting NUSTEP LightDome32D by NUSTEP Ltd. was paired with a light field camera in [6, 185]. The setup is shown in Fig. 4.4 and consists of two components. Namely a light dome which is a half-sphere with 32 LED light sources (approximately 10 W, 1000 lm) and centered camera. The half sphere has an inner diameter of approximately 30 cm with a minimum working distance of 5 cm. The lights are arranged in circular patterns at three height levels around the object,

(a) Light dome　　　　　(b) Light placement　　　　　(c) Top view
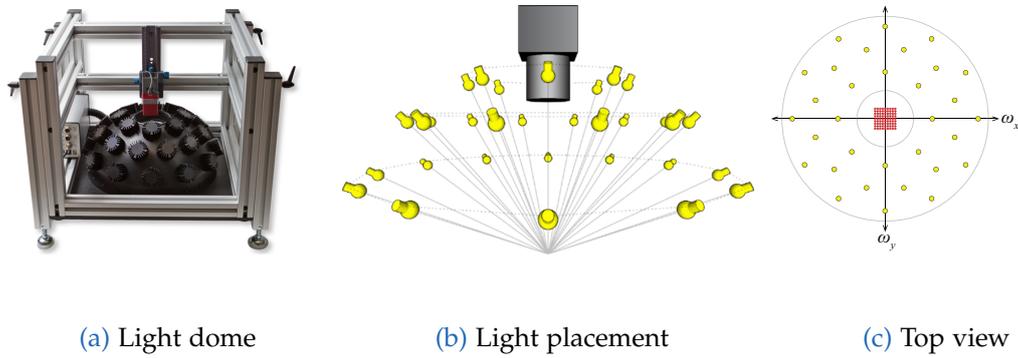
Figure 4.4: Photometric light field acquisition setup [6].

with 8, 12 and 12 LEDs at each height and an angular distance of 25°, 43° and 61°. A light field camera is placed in the domes' top center. Thereby the scene is captured from slightly different viewpoints for each illumination direction. Fig. 4.4c shows the top view of the dome in respect to the domes' directions ω_x and ω_y, red dots indicate the viewing directions from the light field camera. Note that the light field camera can be replaced by a single-view or any other multi-view camera. An internal diffuser allows a diffuse directional illumination from the light dome. In Chapter 7 we describe the measurement of (bidirectional) reflectance distribution functions, where Fig. 7.10 illustrates the portion of the reflectance which can be measured with the light dome coupled with a light field camera, which corresponds to the setup illustrated in Fig. 4.4.

For applications in which the object is moving with a constant speed on a conveyor belt the multi-line scan setup works as a measurement device, the setup was described in Sec. 3.2.5. In the following section we describe how such equipment can be used to measure photometric stereo data.

4.2.3 Multi-Line Scan Setup

We introduced the analysis of photometric stereo data from a multi-line-scan setup in [9]. In a traditional photometric stereo setup the still object is observed multiple times from one viewing perspective under different illumination conditions. This

(a) Multi-Line Scan Setup

(b) EPI Image Stack
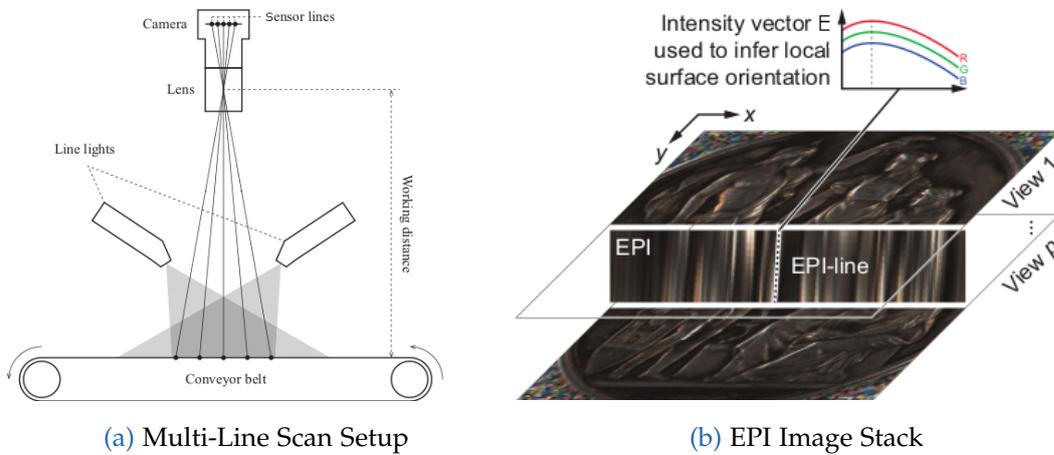
Figure 4.5: Multi-line scan setup for photometric stereo analysis. (a) Schematic depiction of the multi-line scan setup with a top-view camera and two illumination sources. (b) The resulting EPI stack which is acquired with constant light field illumination. The intensity values along the EPI line yield information about the surface orientation.

allows one to derive local surface orientations (i.e. surface normals) from observed intensities making use of known illumination angles [223]. Often this is done under the Lambertian reflectance assumption, resulting in a simple cosine law for determination of surface normals. Assuming Lambertian reflectance is entirely valid only for matte materials. It can be used to derive a simple cosine law for the determination of surface normals. Because the cosine law works quite well even for materials that slightly violate this assumption, we employ this model for surface normal estimates for data from our multi-line scan setup.

In contrast to traditional photometric stereo with multiple switched or strobed light sources, our approach uses two constant line light sources. In our setup, the light sources are located symmetrically next to the optical axis in the transport direction in order to illuminate the observed area from two usually flat angles. As illustrated in Fig. 4.6, such an arrangement gives rise to different illumination configurations in every observed line $(P_1, ..., P_3)$. Since the employed line lights have very homogeneous emissions along the sensor lines, all pixels in the same sensor line are illuminated almost equally and therefore we assume they share the same illumination parameters
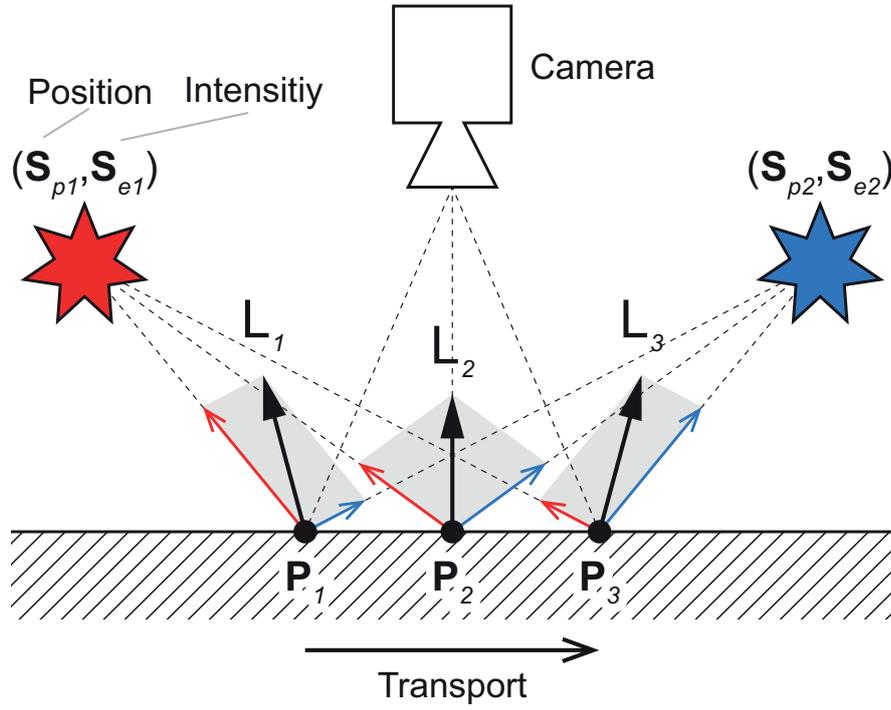
Figure 4.6: Reflectance behaviour for multi-line scan setup. Assumed illumination model comprised of two constant line light sources at the positions S_{p_2} and S_{p_2}, with the scalar intensities S_{e_1} and S_{e_2}, respectively. Due to the inverse-square law, the integral of the two illuminations at the point P_i results in an effective illumination vector L_i, that is different in each observed point (i.e. sensor line).

(i.e. the light direction and intensity). The downside of this illumination geometry is the lack of illuminations spanning orthogonally to the transport direction, resulting in a collinear set of illumination vectors. Consequently, the surface normals can be inferred only in the transport direction (i.e. x-dimension).

We assume a discretized surface with the size of $\mathcal{M} \times \mathcal{N}$ pixels. As previously described in [9], under the assumption of Lambertian reflectance, the surface normals $N_{i,j} \in \mathbb{R}^3$, for all pixel locations $(i, j) \in \mathcal{I}$ (as defined in Sec. 2.1), and the albedo $\rho_{i,j} \in \mathbb{R}$ can be retrieved from multi-line scan data with a constant illumination (without using strobe or switched light sources). For this, the matrix of illumination vectors $L \in \mathbb{R}^{n \times 3}$ and the observed intensities $E_{i,j} \in \mathbb{R}^n$ with a number of n intensities $\forall i, j \in \mathcal{I}$

are utilized. The following tensors hold vectors in each pixel location and are denoted with bold characters:

$$M_{i,j} = \rho_{i,j} N_{i,j}, \tag{4.1}$$

$$M_{i,j} = (M_{i,j,x}, M_{i,j,y}, M_{i,j,z}), \tag{4.2}$$

$$N_{i,j} = (N_{i,j,x}, N_{i,j,y}, N_{i,j,z}), \tag{4.3}$$

$$E_{i,j} = (E_{i,j,1}, ..., E_{i,j,n}), \quad \forall i,j \in \mathcal{I}. \tag{4.4}$$

The reconstructed values are denoted as $M_{i,j} \in \mathbb{R}^3$. Under the assumption of a constant light matrix the illumination vectors are defined as follows:

$$L = (L_{k,x}, L_{k,y}, L_{k,z})_{k=1}^n, \tag{4.5}$$

$$(L_k)_{k=1}^n = \sum_{l=1}^q \frac{S_{p_l} - P_k}{|S_{p_l} - P_k|} \cdot \frac{S_{e_l}}{|S_{p_l} - P_k|^2} = \tag{4.6}$$

$$= \sum_{l=1}^q \frac{S_{e_l}(S_{p_l} - P_k)}{|S_{p_l} - P_k|^3}.$$

This was previously shown in [9]. Considering our illustrative example in Fig. 4.6 we get the following for the effective illumination vector L_1 at the position P_1 with two light sources $q = 2$:

$$L_1 = \frac{S_{e_1}(S_{p_1} - P_1)}{|S_{p_1} - P_1|^3} + \frac{S_{e_1}(S_{p_1} - P_2)}{|S_{p_1} - P_2|^3}. \tag{4.7}$$

The $q \in \mathbb{N}$ light sources are placed at the positions $S_p \in \mathbb{R}^{q \times 3}$ with scalar light intensities $S_e \in \mathbb{R}^q$ at observed sensor line points (corresponds to viewing angles) $P_k \in \mathbb{R}^3$. We assume a simple illumination model based on the Lambertian assumption. Under this constraint, the effective illumination vector L_k in an observed sensor line P_k is given as the sum of all elementary illumination vectors that contribute to that point (shown as thin red and blue arrows in Fig. 4.6). Due to the inverse-square law, the elementary light vectors are different in each observed line, which results in different effective illumination vectors as well. Due to the object's movement during the capture, each object point is eventually observed under every available illumination condition.

In order to recover normals from observed intensities, the least squares problem

follows:

$$\min_{M_{i,j}} \frac{1}{2} ||L \cdot M_{i,j} - E_{i,j}||^2, \quad \text{therefore} \tag{4.8}$$

$$M_{i,j} = \rho_{i,j} \cdot N_{i,j} = (L)^+ \cdot E_{i,j}, \quad \forall i,j \in \mathcal{I} \tag{4.9}$$

$$M_{i,j} = [(L)^T L]^{-1} (L)^T \cdot E_{i,j}, \tag{4.10}$$

where $(\cdot)^T$ denotes the transpose and $(\cdot)^+$ denotes the pseudo-inverse. Since per definition normals are unit vectors:

$$\sqrt{N_{i,j,x}^2 + N_{i,j,y}^2 + N_{i,j,z}^2} = 1, \quad \forall i,j \in \mathcal{I}, \tag{4.11}$$

the length of the vector $M_{i,j}$ is defined by the albedo $\rho_{i,j}$. Therefore, we have the following correspondence:

$$\rho_{i,j} = \sqrt{M_{i,j,x}^2 + M_{i,j,y}^2 + M_{i,j,z}^2}, \quad \forall i,j \in \mathcal{I}. \tag{4.12}$$

It can be shown that in the multi-line scan framework, only a lower bound of the albedo can be estimated due to the missing photometric component that is orthogonal to the transport direction N_y. This can be shown as follows:

$$\rho_{i,j} = \sqrt{(\rho_{i,j} \cdot N_{i,j,x})^2 + (\rho_{i,j} \cdot N_{i,j,y})^2 + (\rho_{i,j} \cdot N_{i,j,z})^2}, \tag{4.13}$$

$$\rho_{i,j}^2 = \frac{M_{i,j,x}^2 + M_{i,j,z}^2}{1 - N_{i,j,y}^2}, \tag{4.14}$$

$$\rho_{i,j} = \sqrt{\frac{M_{i,j,x}^2 + M_{i,j,z}^2}{1 - N_{i,j,y}^2}}, \quad \forall i,j \in \mathcal{I}, \tag{4.15}$$

where $M_{i,j,x} = \rho_{i,j} N_{i,j,x}$ and $M_{i,j,z} = \rho_{i,j} N_{i,j,z}$ are given through our measurements (Eq. 4.9). Then the lower bound of ρ results when $N_{i,j,y} = 0$, $\forall i,j \in \mathcal{I}$:

$$\rho_{i,j} \geq \sqrt{M_{i,j,x}^2 + M_{i,j,z}^2}. \tag{4.16}$$

We can perform photometric stereo within the multi-line scan framework without the necessity of switching or strobing the illumination during the acquisition process.
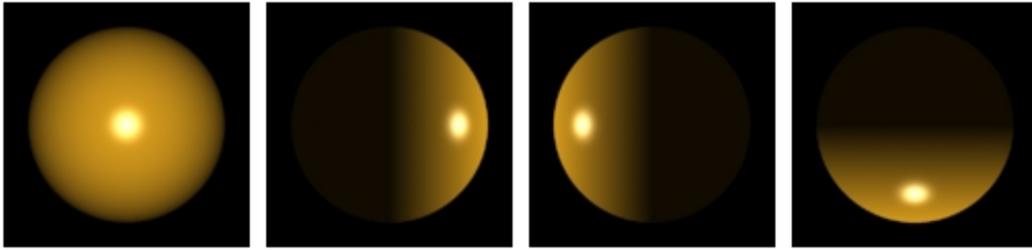
Figure 4.7: Light direction calibration for photometric stereo [180].

As a result of parallax comprised in the light field, intensities associated with the same object location occur along an EPI-line, the slope of which depends on the absolute distance of that location from the camera (see Fig. 4.5b). Therefore, with our approach the surface normal estimates are inherently linked with respective depth estimates. Hence, it is necessary to utilize a preliminary depth model to calculate surface normals, which can afterwards be used to improve the depth model, etc.

4.3 Calibration

Before calculating the normal map and further using it for 3D reconstruction, the direction of the light sources has to be estimated. Highly reflective spheres are usually utilized to identify the illumination positions [47, 165, 166, 180, 226]. A rendered example of a sphere, illuminated from four different directions, is shown in Fig. 4.7.

Under the assumption that the edge of the sphere will be presented as an ellipse in the captured image, the coordinates for any point on the surface as well as the surface normal on that specific position can be identified [166]. By detecting the light reflection on the sphere, the illumination direction can therefore be determined.

Placing a second sphere allows the reconstruction of the position of the light source through triangulation, as described in [165].

(a) Normal Map   (b) Reconstructed Depth   (c) Textured Depth
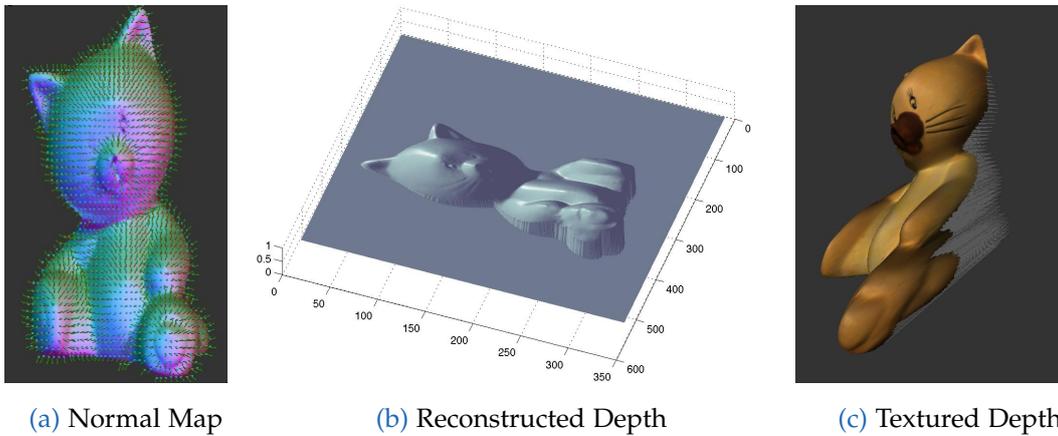
Figure 4.8: 3D reconstruction from normal map [180]. a) Normal map with normal vectors of an object. b) Reconstructed depth map. c) Rendered depth map with texture.

4.4 3D reconstruction

Through the shading variation on objects, illuminated from different positions with a static camera viewpoint, the surface normal map and the resulting scene depth can be estimated. Usually a uniform albedo and reflectance of the scene are assumed [193] when calculating the depth map from surface reflectance. The positions of the light sources are either known or estimated by utilizing a highly reflective sphere object as discussed in Sec. 4.3. The light sources are placed in a structured (on a sphere, on a ring, in a line) or in an unstructured (freely movable) way around the object. Changes in the illumination direction are leading to pixel intensity changes, depending on the orientation and reflectance characteristics of the object. The captured irradiance values $E_{(i,j)} \in \mathbb{R}$ are usually described using a reflectance map $R_{(i,j)} \in \mathbb{R}$ as introduced in [222] for all $i, j \in \mathcal{I}$:
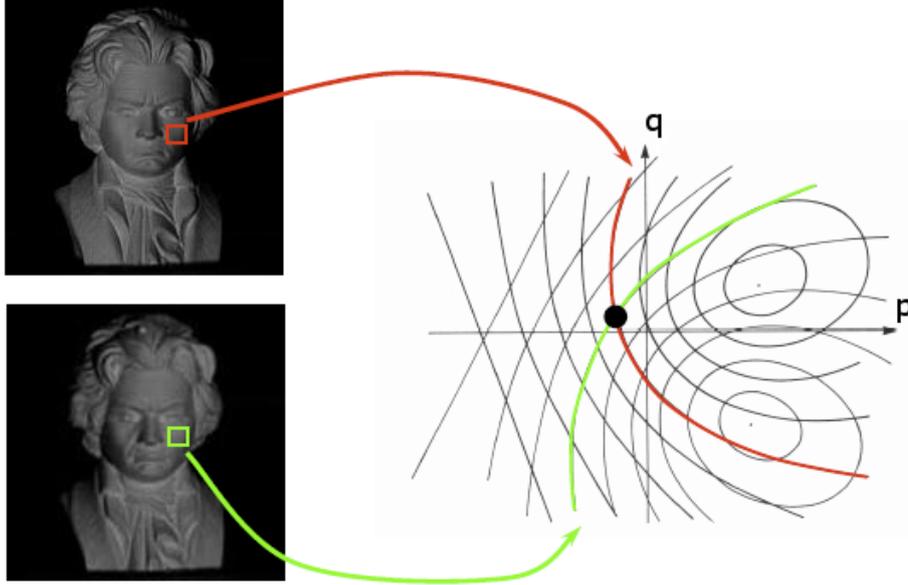
Figure 4.9: Illustration of reflectance maps from two illumination directions [69]. The shaded reflectance is shown left and the resulting reflectance map on the right for a region in a scene. Using two point source illuminations the intersection of both resulting level lines gives at most two surface orientations which would produce that pair of brightness values. A third light source gives a unique result.

$$
\begin{aligned}
E_{(i,j)} &= R(\boldsymbol{G})_{(i,j)}, \quad \text{where} & (4.17) \\
\boldsymbol{G}_{(i,j)} &= (G_{i,j,x}, G_{i,j,y}), & (4.18) \\
G_{(i,j,x)} &= \nabla_x Z_{(i,j)}, \quad \text{and} & (4.19) \\
G_{(i,j,y)} &= \nabla_y Z_{(i,j)} & (4.20)
\end{aligned}
$$

are the depth map derivates and represent two points in the gradient space of the depth map $Z_{(i,j)} \in \mathbb{R}$. The gradient operator ∇ was previously defined in Eq. 2.9. An example of a reflectance map is shown in Fig. 4.9. Since we assume a Lambertian reflectance model and a constant albedo, the irradiance results from a dot product between the surface normal unit vector $\boldsymbol{N}_{(i,j)} \in \mathbb{R}^3$ and the direction of the light source

$L_{(i,j)} \in \mathbb{R}^3$, scaled by the albedo $\rho_{(i,j)} \in \mathbb{R}$:

$$E_{(i,j)} = \rho \cdot L_{(i,j)} \cdot N_{(i,j)}, \quad \text{where} \tag{4.21}$$

$$N_{(i,j)} = \frac{(G_{i,j,x}, G_{i,j,y}, 1)}{\sqrt{1 + G_{i,j,x}^2 + G_{i,j,y}^2}}, \quad \text{and} \tag{4.22}$$

$$L_{(i,j)} = (L_{i,j,x}, L_{i,j,y}, L_{i,j,z}). \tag{4.23}$$

The reflectance map is therefore calculated as follows:

$$R(G_x, G_y) = max\left(0, \rho \frac{G_x L_x + G_y L_y + L_z}{\sqrt{1 + G_x^2 + G_y^2}}\right) \tag{4.24}$$

In photometric stereo we have n light sources, which results in the reflectance maps $(R_1, ..., R_n)$, and the intensity vectors $(L_1, ..., L_n)$. This allows one to get a set of linear equations $E_k = \rho L_k \cdot N$ for all $k \in n$ which can be solved for $\rho \cdot N$. When at least three vectors E_k are linearly independent the equations are well conditioned.

The recovered surface normal vectors can be integrated into a depth map, with e.g. the algorithm of Frankot and Chelappa [62], recent work has been presented in [86, 148]. For specular surfaces or the occurrence of inter-reflections, more than three light sources are necessary to reliably estimate the surface orientation and depth, as shown in [147].

It is well known [5, 48] that the resulting depth values show a reliable relative depth accuracy, but have an absolute depth offset, as shown in Fig. 4.10. The surface normal vectors in our energy term are used to refine the surface structure of our depth result.

4.5 Conclusion

In Chapter 3 we introduced light field methods and their application in depth reconstruction tasks. Now, in this chapter we described photometric stereo methods with

(a) Image

(b) Depth Map
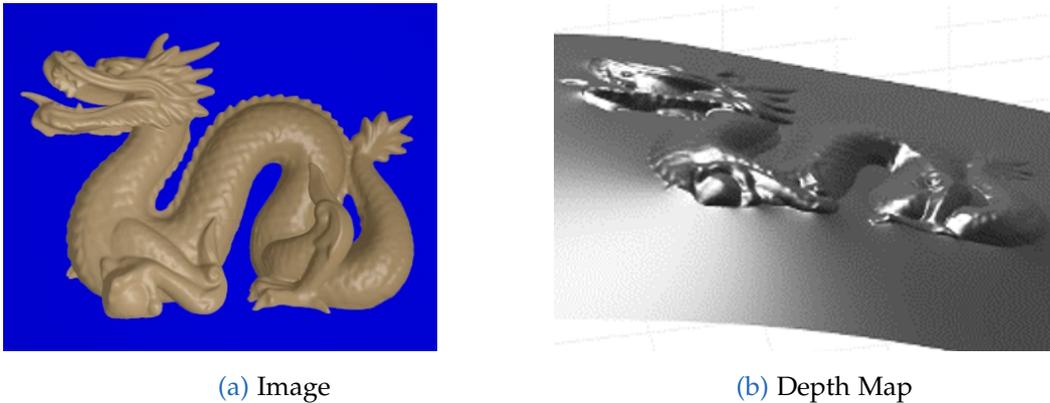
Figure 4.10: Depth reconstruction from surface normals. a) One sample acquisition out of *n* photometric stereo images. The model from the Stanford object database [186] rendered with our virtual photometric stereo setup (described in [6]) using POVRay [163]. ase b) Depth reconstruction shows global offset (without additional restrictions).

a focus on the industrial applicability. In Sec. 4.2 we showed devices and setups to capture photometric stereo data, from the mobile ring light to the industrial inline method using a multi-line scan camera. Here we presented our previous work on the retrieval of photometric stereo data from an EPI stack constructed from a multi-line scan system [9] with a refined formulation from [8]. In Sec. 4.3 we described the calibration of the illumination direction using highly reflective sphere objects. In Sec. 4.4 we discussed recovering surface normals from photometric stereo data and integrating them to achieve a 3D reconstruction.

Light field methods achieve a stable absolute depth measurement, while photometric stereo data allows the reconstruction of fine surface details but fails in global absolute depth accuracy. Combining depth (e.g. from light field data) with surface normals (e.g. from photometric stereo) allows a locally and globally accurate 3D reconstruction result. In the next chapter we take a look on the details of the combination of depth and surface normal data in order to reach a refined depth reconstruction.

5 Hybrid Depth and Surface Normal Approaches

Contents

Refining 3D data by combining depth maps with surface normals allows a globally and locally accurate surface reconstruction. In this chapter we present our work published in [7]. We motivate the research in Sec. 5.1. Sec. 5.2 addresses the retrieval of depth and surfaces normal cues. The notation is defined in Sec. 5.3. A survey and evaluation of optimization based depth and surface normal fusion algorithms for area scan data is shown in Sec. 5.4. Here we also present two novel methods, namely a generalized method of Nehab in Sec. 5.4.6 and a Total Generalized Variation (TGV) based method in Sec. 5.4.7.

5.1 Motivation

Measuring the depth of a scene accurately is essential for many tasks including applications in industrial environments, object recognition and security assurance. Usually the depth is measured by stereo cameras, structure from motion (SfM), time of flight

(ToF) sensors, or light field cameras. These methods show accurate absolute depth maps but lack detail in high frequency depth structures. The reason for this lies in the dependency on the presence of structural information in the image as well as in the analysis routine, which is usually done by hypothesis testing and therefore limited in range and step sizes. Stereo matching methods include correlation based techniques [94], semi global matching [81, 93], block based matching [100, 231] and stereo matching for micro array cameras [37]. Several methods were introduced to retrieve depth information from light field data using epipolar image slope analysis [22], structure tensors [214], fine-to-coarse approaches [110, 204] and by line consistency metrics [198]. Structure from motion and time of flight techniques were presented in [132, 159, 181], respectively. In contrast to depth based approaches, methods that recover surface normals, such as photometric stereo [223], show high frequency details but lack an absolute depth reference. Shape from shading was used to retrieve surface normals by [62]. A robust normal reconstruction using photometric stereo information with a Markov Random Field (MRF) was introduced in [226]. Previous methods have been presented, which retrieve surface normals from a calibrated stereo setup. This can be achieved e.g., by estimating the homography between two matched patches [113, 137] or by using the affine transform data between two projections [10] additionally to the reconstruction of a sparse depth map retrieved from a stereo correspondence analysis. A learning-based method using a tandem of convolutional neural networks to estimate depth and surface normals from image data simultaneously was introduced in [51]. Combining depth and surface normal data allows precise depth reconstructions for low- as well as high-frequency components in the depth map.

Depth maps and surface normals were previously combined in various ways. Shape from shading was used under general illumination in [224], photometric stereo normals were incorporated in [52, 105, 238]. Another approach was presented in [148], where the tangent plane of the given normals was projected into the measured normal field. This normal constraint was previously used in several algorithms (e.g., [33, 84, 85, 160]). The method described in [177] uses a standard depth constraint and forces the Laplacian of the optimal solution to be in the proximity of the derivative of the given normals. In [234] a depth and photometric stereo fusion algorithm was introduced, which uses additional Laplacian smoothing term and adaptive pixel-wise weighting

parameters to preserve surface discontinuities. The Laplacian smoothing term was also added in [202]. A extended penalty is chosen in [235], where the normal is enforced to be close to the normal from the initial depth map, while 2nd-order spherical harmonics are used to constrain the normals according to the observed shading in the input image, a smoothness function enforces the similarity of 1st-order neighbors and an additional term constrains the normals to unit length. A method to refine depth by photometric stereo information using RGB-D cameras was introduced in [225], where an energy function is optimizing for the depth, smoothness, shading and temporal aliasing of a scene. Surface normals from polarization cues were used to enhance the depth map in [106], in an iterative process the depth is refined with corrected surface normal information and a depth fidelity constraint, which enforces consistency between the surface from normals and accurate regions in the depth map. An original approach for inferring about the surface normals from light field data as well as a hybrid setup combining depth maps with surface normals using a block coordinate descent algorithm was demonstrated in [9].

Even though several methods to combine surface depth with surface orientation data previously emerged, a thorough analysis and classification of the properties of those approaches was missing. In this section, our first main contribution comprises an in-depth comparison of several variational methods using depth maps and surface orientation data, as well as a classification and evaluation of weighting terms for surface orientations using gradients or surface normals. We analyze orientation weighting terms of common methods and explain their differences in respect to the geodesic distance weighting. We show that methods which behave closer to this natural surface normal weighting term show a better performance, especially in regions with steep depth edges. Based on the findings we introduce our second main contribution, a new generalized formulation of a previously introduced method [148], which outperforms other methods regarding the error in the depth domain. Our third main contribution is a novel gradient-based approach, which is using TGV and outperforms other methods in the domain of the geodesic error of the resulting normals.

5.2 Depth and Surface Normal Cues

At present, 3D models are used for a wide range of analysis tasks. Depth models are being constructed by acquisition devices using stereo systems, light field cameras, time of flight (ToF), or other range scanning techniques. Common methods show a high precision in the absolute depth measurement, but a low quality in fine relative details. These errors are major obstructions for tasks such as finding defects in objects. Measuring the normal fields of objects by using methods such as photometric stereo [223] or shape from shading [97] will allow the reconstruction of surfaces with highly precise local details. On the downside, those methods show errors in the low frequency domain and therefore result in a low absolute depth accuracy.

Combining depth maps with surface normal information allows an exact 3D reconstruction both in absolute depth and fine surface details. This can be achieved by optimizing energy functions by variational methods, where the solution is penalized for deviating from the depth model and from the surface normals. In state-of-the-art techniques, the surface normal component is either used directly or by converting it to gradient information, where the x- and y-component can be treated independently. Such an independent treatment can be beneficial for applications where data components are missing, as for example line-scanners [190].

5.3 Notations and Preliminaries

In this section, we introduce the essential notations used across this Chapter, based on the notation introduced in Sec. 2.1. By default we assume discretized surface structures of the size of $M \times N$ pixels. The index set $\mathcal{I}$ was defined in Sec. 2.1.

The discrete depth map of our scene is scalar valued in each pixel and defined as follows:

$$Z = (Z_{i,j})_{i,j\in\mathcal{I}} \in \mathbb{R}^{M\times N}. \tag{5.1}$$

Variables with a bold font refer to surface structures where each pixel is vector valued. Hence, the surface gradient field $\boldsymbol{G}$ in x- and y-direction is defined as:

$$\boldsymbol{G} = \left(\boldsymbol{G}_{i,j}\right)_{i,j\in\mathcal{I}} \in \mathbb{R}^{M\times N\times 2}, \quad \text{where} \quad \boldsymbol{G}_{i,j} = (G_{i,j,x}, G_{i,j,y}). \tag{5.2}$$

The gradient of the depth map Z as well as the gradient operator ∇ were previously defined in Eq. 2.9.

The surface normal field is defined as follows:

$$\begin{aligned}
\boldsymbol{N} &= \left(\boldsymbol{N}_{i,j}\right)_{i,j\in\mathcal{I}} \in \mathbb{R}^{M\times N\times 3}, \quad \text{where} \\
\boldsymbol{N}_{i,j} &= (N_{i,j,x}, N_{i,j,x}, N_{i,j,z}) \in \mathbb{R}^3.
\end{aligned} \tag{5.3}$$

By definition, we have $|\boldsymbol{N}_{i,j}|_2 = 1$. The relation between the surface gradient estimation and the surface normals is defined for all $i,j \in \mathcal{I}$ as follows:

$$\begin{aligned}
N_{i,j,x} &= \frac{-(\nabla_x Z)_{i,j}}{|(-(\nabla Z)_{i,j}, 1)|_2}, \\
N_{i,j,y} &= \frac{-(\nabla_y Z)_{i,j}}{|(-(\nabla Z)_{i,j}, 1)|_2}, \quad \text{and} \\
N_{i,j,z} &= \frac{1}{|(-(\nabla Z)_{i,j}, 1)|_2}.
\end{aligned} \tag{5.4}$$

Furthermore, we are using two specific surface tangent vectors, which are aligned with the x- and y-vector respectively and defined as follows:

$$\boldsymbol{T} = (\boldsymbol{T}_{i,j})_{i,j\in\mathcal{I}} \in \mathbb{R}^{M\times N\times 2\times 3}, \quad \text{where} \tag{5.5}$$

$$\boldsymbol{T}_{i,j} = (\boldsymbol{T}_{i,j,x}, \boldsymbol{T}_{i,j,y}) \quad \text{and} \tag{5.6}$$

$$\boldsymbol{T}_{i,j,x} = (1, 0, (\nabla_x Z)_{i,j}) \quad \text{and} \quad \boldsymbol{T}_{i,j,y} = (0, 1, (\nabla_y Z)_{i,j}). \tag{5.7}$$

Hereinafter, the surface normals estimated by the described method will be referred to as measured surface normals, since they are assessed directly from the recorded data. On the other hand, normals derived from the reconstructed depth models will be denoted as model surface normals.

5.4 Area Scan Depth and Surface Fusion Algorithms

In this section, we analyze state-of-the-art methods in a systematic way and introduce two novel approaches. The described hybrid depth and surface normal methods are categorized in terms of their penalty functions. State-of-the-art approaches used similar depth penalty terms and differed in the surface orientation weighting and regularization. We organize the described methods in two categories: (i) gradient-based and (ii) normal-based approaches as well as with respect to their treatment of flat and steep surface regions. While the presented methods show similar behaviors in flat areas, they differ in the penalization of steep regions. We show the quadratic penalty functions of the methods presented in Figs. 5.1 and 5.2 for lateral and polar deviations, which are illustrated in Fig. 5.3. We will show that the behavior of the energy function for different inclination angles correlates with the quantitative and qualitative depth reconstruction performance of each individual method. We argue that the geodesic distance function shows the most natural behavior with the favorable property of penalizing the distance of the normal orientation independent of the steepness of the edges. Due to this ideal behavior, we use the function to evaluate other distance measures in Section 5.4.8. Without loss of generality, we assume a dense depth and normal map for the algorithms described in this paper. In case of sparse input data, we suggest an extension based on Poisson surface reconstruction [108] to deal with sparse data.

An overview of the presented methods is given in Tab. 5.1. The first method we present in Section 5.4.2 is the construction of depth from only the gradient surface orientation information (i.e., no prior information about the absolute depth is being used). Using only the surface orientation for the depth reconstruction results in large-scale low-frequency errors (and therefore depth offsets). Later we overcome this problem by introducing an additional depth constraint in all following methods.

Second, we introduce the gradient-based approach with a depth constraint formulated as a least squares problem in Section 5.4.3. The respective contour plot of the orientation distance measure shows a strong penalization of steep edges in contrast to flat surface regions. It is easy to see (Fig. 5.1), that the error from the same angular deviation due

to noisy normals may generate from small up to infinity error in the gradient domain, depending on the inclination angle. For demonstrative purposes, we additionally show two extensions of the gradient-based method with regularization terms. One forces gradient-based smoothing. The other one enforces smoothness with a Laplacian term and can be used for the reconstruction with sparse depth and surface normal data.

Third, the method of Heber [87] for combining depth with surface orientations is shown in Section 5.4.4, which scales the given normal by the length of the optimized gradient.

Forth, a review of the method of Nehab [148] is given in Section 5.4.5, which reprojects the tangents of the optimized surface onto the given normal.

Fifth, as one of our main contributions in this paper, we introduce a new penalty function in form of a generalized method of Nehab in Section 5.4.6. Using a novel parametrization moves the penalty function closer to the geodesic normal energy and hence penalizes deviations in steep edges and flat regions more equally.

Last, we introduce our second main contribution, a novel Total Generalized Variation (TGV) model in Section 5.4.7, which penalizes the distance of the gradients of the surface orientation and gives significantly improved reconstruction results. The reason for this lies in the decoupling of the gradient through the TGV term.

5.4.1 Geodesic Distance

In the 3D space, the geodesic distance is the most natural surface normal penalty as distances between surface normals are weighted equally, independent of the surface slope angle with respect to the observer. Therefore it is used in this paper as a comparison measure for the evaluation in Section 5.4.8.

The geodesic distance $d_{i,j}$ is defined as the inverse cosine of the point-wise dot product between the given normal $\hat{N} \in \mathbb{R}^{M \times N \times 3}$ and the estimated solution normal

Method	Depth Penalty	Orientation Penalty	Balance at Flat Regions	Balance at Steep Regions
Surface orientation only	✗	Gradient-based	✓	✗
Gradient-based	✓	Gradient-based	✓	✗
Gradient-based	✓	Gradient-based + regularization with zero Laplace (Eq. (5.21))	✓	✗
Gradient-based	✓	Gradient-based + regularization with derivative of the gradient (Eq. (5.19))	✓	✗
Method of Heber	✓	Scaled normal	✓	✗
Method of Nehab	✓	Projection of surface tangents to the given normal field	✓	⌣
Generalized Nehab (ours)	✓	Projection of surface tangents to the given normal field with additional N_z weighting	✓	✓
TGV (ours)	✓	Gradient-based + TGV	✓	✓

Table 5.1: Overview of the presented methods. The behavior of the surface orientation constraint in regions with different orientations is visualized for each method in Figs. 5.1 and 5.2.

field $N \in \mathbb{R}^{M \times N \times 3}$ as follows:

$$d_{i,j} = acos(\langle \hat{N}_{i,j}, N_{i,j} \rangle), \quad \forall i,j \in \mathcal{I}, \tag{5.8}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard dot product, which is defined as $\langle a, b \rangle = \sum_{i=1}^{n} a_i\, b_i = ||a||\,||b||\,cos(\phi)$ and ϕ describes the angle between a and b. We can formulate the distance in Eq. (5.8) by utilizing the surface gradient estimation $\nabla Z \in \mathbb{R}^{M \times N \times 2}$ for the surface normal N with the relations shown in Eq. (5.4) as follows:

$$d_{i,j} = acos\left(\frac{\langle \hat{N}_{i,j}, (-(\nabla Z)_{i,j}, 1) \rangle}{|(-(\nabla Z)_{i,j}, 1)|_2} \right). \tag{5.9}$$

The surface orientation weighting of the geodesic distance is illustrated in the contour plot in the first row of Fig. 5.1. The first column shows the polar deviation of the coordinates and the second column shows the lateral deviation, parameterized by the inclination angle α and the deviation angle β, as illustrated in Fig. 5.3. Colors of the
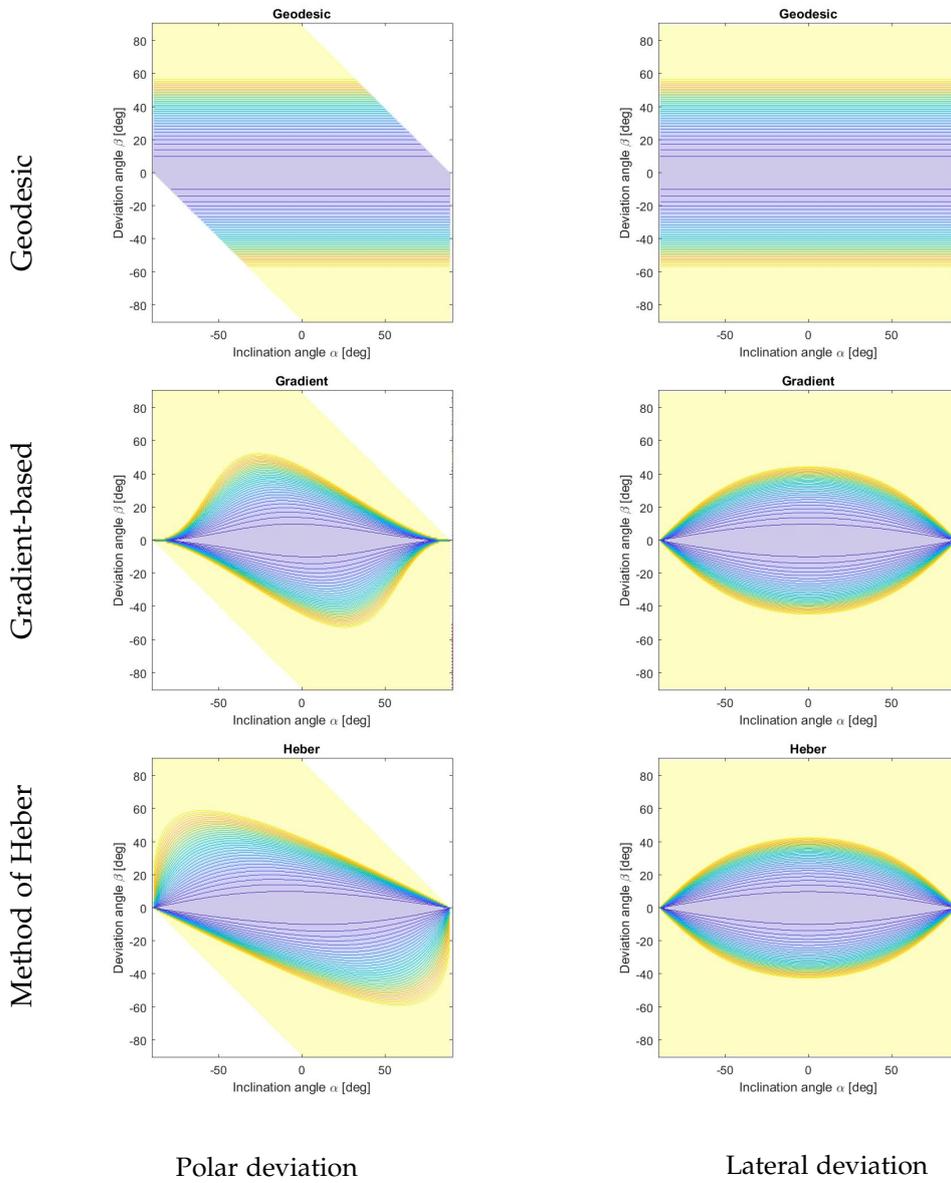
Figure 5.1: Quadratic penalty functions of the surface orientation constraint. Visualized for deviations in polar and lateral directions for the geodesic energy, the gradient-based method and the method of Heber. Colors of the contours indicate the respective penalty values. Note that the range is clipped between 0 (blue) and 1 (yellow). See Fig. 5.3 for the explanation of the parametrization used.

Polar deviation   Lateral deviation
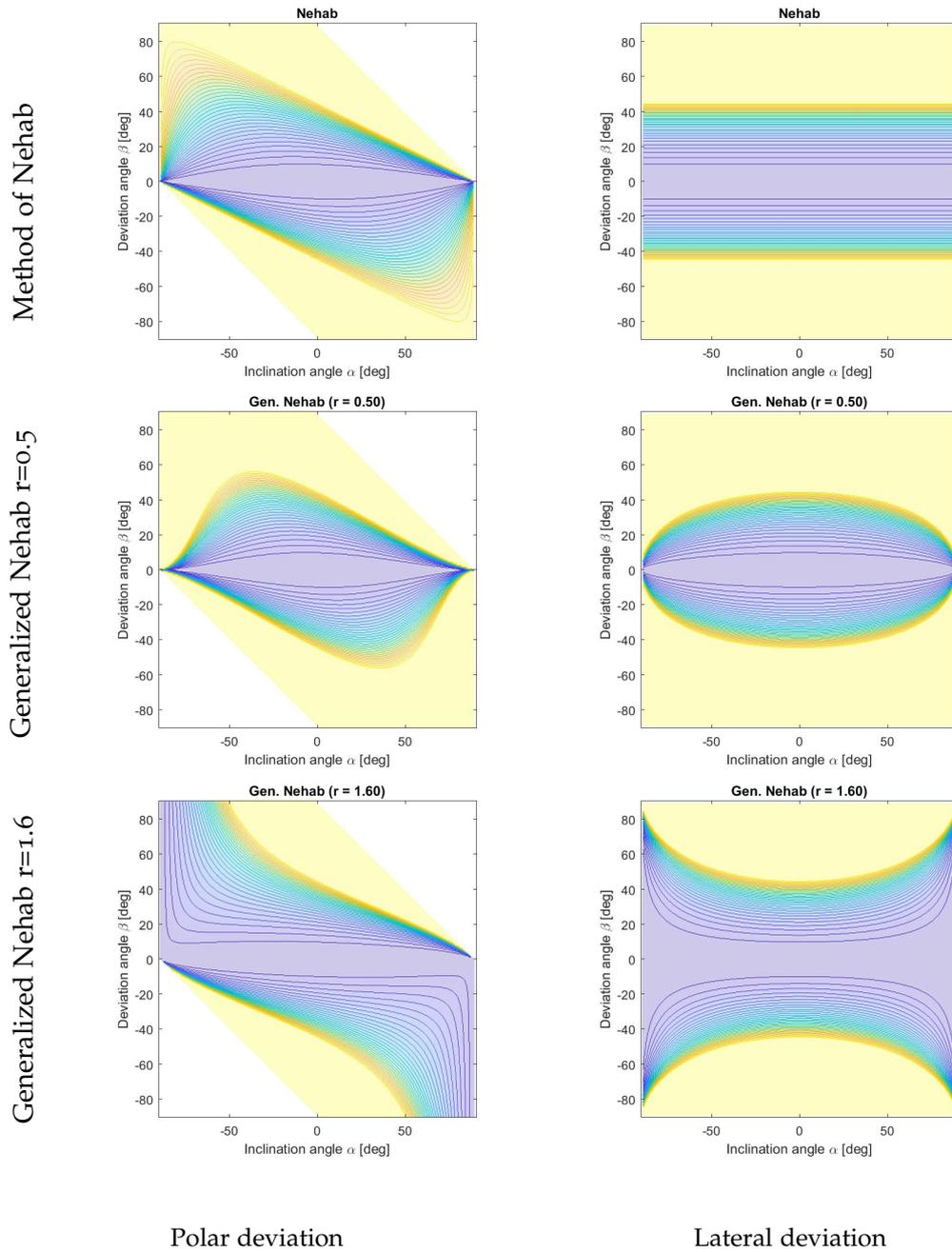
Figure 5.2: Quadratic penalty functions of the surface orientation constraint. Visualized for deviations in polar and lateral directions. Shown are the method of Nehab, the generalized method of Nehab with $r = 0.5$ as well as $r = 1.6$. Colors of the contours indicate the respective penalty values. Note that the range is clipped between 0 (blue) and 1 (yellow). See Fig. 5.3 for the explanation of the parametrization used.
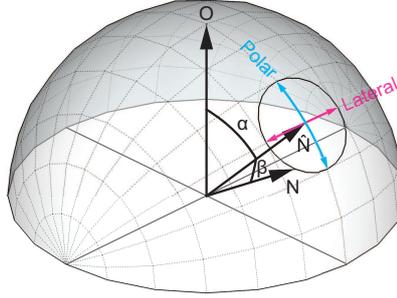
Figure 5.3: Horizontal polar coordinate system. Expressing the distance between a two surface normals $\hat{N}$ and N. The distances in Figs. 5.1 and 5.2 are measured in lateral and polar directions using angles α and β. The former describes the angle between the given normal $\hat{N}$ and the the upright vector O, the latter defines the angular deviation between the normals $\hat{N}$ and N.

contours indicate the respective penalty values obtained for different combinations of the inclination and deviation angles. Note that the range is clipped between 0 (blue) and 1 (yellow).

For the following methods a balanced weighting over all inclination angles both for the polar and lateral deviation is favored, as provided by the geodesic distance.

5.4.2 Gradient-Based Method with Surface Orientation Constraint Only

As typical for photometric stereo methods, depth can be partly recovered from surface orientations only. In order to provide a complete context for the method considered in this paper, we show here a method that is using solely gradient-based data. Gradient-based methods have previously been frequently utilized for depth reconstruction (e.g., [2, 49, 62, 86, 98, 170, 172]). Given a gradient field $\hat{G} \in \mathbb{R}^{M \times N \times 2}$, we calculate the surface gradients for the estimated depth map Z in x- $(\nabla_x Z)$ and y-direction $(\nabla_y Z)$ respectively. Combining relations from Eq. (5.4), the relations of surface normals and gradients are as follows:

$$|(-(\nabla Z)_{i,j}, 1)|_2 = -\frac{(\nabla_x Z)_{i,j}}{N_{i,j,x}} = -\frac{(\nabla_y Z)_{i,j}}{N_{i,j,y}} = \frac{1}{N_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \qquad (5.10)$$

hence the surface gradients are given as:

$$(\nabla_x Z)_{i,j} = -\frac{N_{i,j,x}}{N_{i,j,z}} \quad \text{and} \quad (\nabla_y Z)_{i,j} = -\frac{N_{i,j,y}}{N_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \tag{5.11}$$

and the given gradient fields correspond to the surface normals by:

$$\hat{G}_{i,j,x} = -\frac{\hat{N}_{i,j,x}}{\hat{N}_{i,j,z}} \quad \text{and} \quad \hat{G}_{i,j,y} = -\frac{\hat{N}_{i,j,y}}{\hat{N}_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \tag{5.12}$$

in x- and y- direction respectively. Our goal is to compute a depth map Z such that

$$(\nabla Z)_{i,j} \approx \hat{G}_{i,j}, \quad \forall i,j \in \mathcal{I}. \tag{5.13}$$

The comparison of the resulting penalty between the measured and the given gradients is illustrated in Fig. 5.4a. Since Eq. (5.13) is an overdetermined system of linear equations, it can be solved as the following least squares problem:

$$\min_{Z} \frac{1}{2} ||\nabla Z - \hat{G}||^2, \tag{5.14}$$

whose global minimizer Z_{min} satisfies the first order sufficient optimality condition:

$$\nabla^*(\nabla Z_{min} - \hat{G}) = 0, \tag{5.15}$$

where $\nabla^* : \mathbb{R}^{M \times N \times 2} \to \mathbb{R}^{M \times N}$ denotes the adjoint of the ∇ operator, with $\nabla^* = (\nabla_x^*, \nabla_y^*)$. We compute the minimizer using a standard conjugate gradient method.

It is well known that reconstructing the depth using only surface normal data usually results in errors in the low frequency domain. In the past, this has been improved by different approaches, such as introducing additional boundary conditions [86]. We resolve this problem by hybrid depth and surface normal formulations which proved very efficient in finding an accurate surface reconstruction.
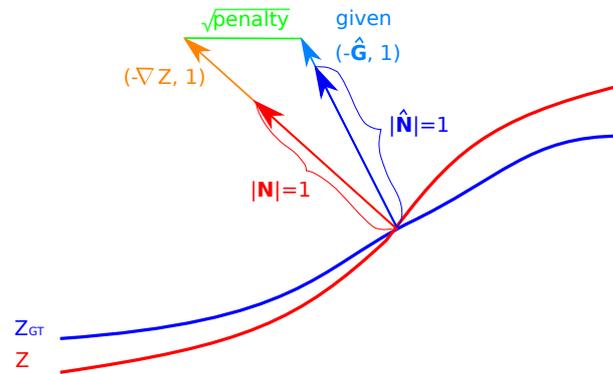
(a) Gradient-based method
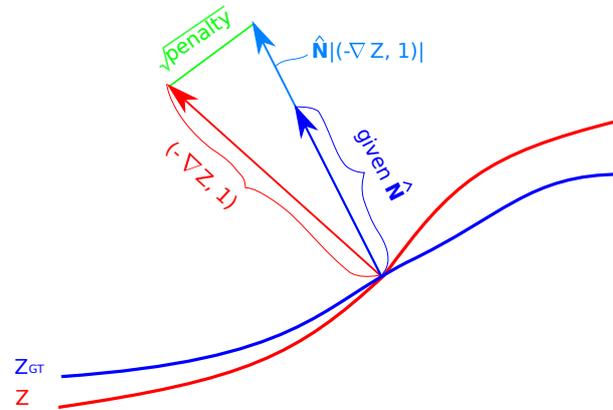
(b) Method of Heber
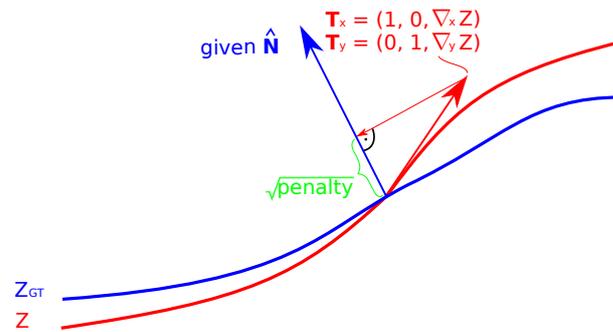
(c) Method of Nehab

Figure 5.4: Comparison of the penalty. (**a**) the gradient-based method; (**b**) the method of Heber and (**c**) the method of Nehab. Gradient-based method: comparison of the measured gradient field components $(-\nabla_x Z, -\nabla_y Z)$ with the given gradient field $(-\hat{G}_x, -\hat{G}_y)$. Method of Heber: scaling of the normal vector $\hat{N}$ by the length of $(-\nabla_x Z, -\nabla_y Z, 1)$. Method of Nehab: distance by projection of the tangents T_x and T_y of the optimized surface onto the given normal field $\hat{N}$.

5.4.3 Gradient-Based Method

In this section, we discuss a hybrid gradient-based method which reconstructs depth using a gradient-based algorithm (similar to the one from Section 5.4.2) extended by the use of the given initial depth $\hat{Z}$. Also here the gradient of the estimated depth map ∇Z is forced to be in the proximity of a measured gradient $\hat{G}$ in x- and in y-direction. Therefore, we formulate an overdetermined system of equations as follows:

$$Z \approx \hat{Z} \quad \text{and} \quad (\nabla Z)_{i,j} \approx \hat{G}_{i,j}, \quad \forall i,j \in \mathcal{I}. \tag{5.16}$$

The corresponding least squares problem is given as:

$$\min_Z \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}||\nabla Z - \hat{G}||^2, \tag{5.17}$$

where $\lambda > 0$ is used to balance between the depth and the orientation constraints. The global optimizer Z_{min} is found by a standard conjugate gradient method, with the optimality condition given as follows:

$$Z_{min} - \hat{Z} + \lambda \nabla^*(\nabla Z_{min} - \hat{G}) = 0. \tag{5.18}$$

The contour plots of the surface orientation penalty function corresponding to the gradient-based method are shown in Fig. 5.1. Note that with this method, the penalty is notably stronger for deviations in steep regions than in flat regions.

For demonstrative purposes, we introduce the extension of the gradient-based method in Eq. (5.17) with a regularization term. We use a Laplace 2nd-order method which is driven by the derivative of the given gradient field. We formulate the following least squares problem:

$$\min_Z \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}||\nabla Z - \hat{G}||^2 + \frac{\lambda_R}{2}||\Delta Z - (-\nabla^*\hat{G})||^2, \tag{5.19}$$

where $\lambda_R > 0$ balances the regularization, $\nabla^*\hat{G}$ can be decomposed into $(\nabla_x^*\hat{G}_x, \nabla_y^*\hat{G}_y)$ and Δ denotes the Laplace operator which is defined as follows:

$$\Delta_x = -\nabla_x^*\nabla_x \quad \text{and} \quad \Delta_y = -\nabla_y^*\nabla_y. \tag{5.20}$$

Such a gradient-based regularization can be applied to all presented methods.

The presented gradient-based approaches require dense depth and surface orientation data. A possibility to cope with sparse data is an additional smoothness assumption coupled with pixel-wise weighting parameters. Hence, we add another term to Eq. (5.17) with a Laplacian smoothness assumption as follows:

$$\min_Z \frac{1}{2}||\lambda_1 \odot (Z - \hat{Z})||^2 + \frac{1}{2}||\lambda_2 \odot (\nabla Z - \hat{G})||^2 + \frac{1}{2}||\lambda_3 \odot (\Delta Z)||^2. \tag{5.21}$$

The weighting parameters λ^1 and λ^2 can be given a priori by the confidence of a data point and λ^3 by the inverse confidence and are defined as follows:

$$\lambda_1 = \left(\lambda_{1,i,j}\right)_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N}, \tag{5.22}$$

$$\lambda_2 = \left(\lambda_{2,i,j}\right)_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2}, \quad \text{where} \quad \lambda_{2,i,j} = (\lambda_{2,i,j,x}, \lambda_{2,i,j,y}), \tag{5.23}$$

$$\text{and}$$

$$\lambda_3 = \left(\lambda_{3,i,j}\right)_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2}, \quad \text{where} \quad \lambda_{3,i,j} = (\lambda_{3,i,j,x}, \lambda_{3,i,j,y}). \tag{5.24}$$

In case of stereo or light-field methods for depth reconstruction the parameters can be assessed base on the matching confidence. Unknown points would have a confidence of zero. The same extension for sparse data is applicable for all following methods. Therefore, without loss of generality we discuss the weighting of the surface orientation term with a focus on dense depth and surface orientation data without an additional smoothness assumption.

5.4.4 The Method of Heber

A hybrid variational refinement model was described by Heber [87], where an initial rough depth $\hat{Z}$ is given and refined with surface normal information:

$$\min_Z \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}\left|\left|(-\nabla Z, 1) - \hat{N} \odot |(-\nabla Z, 1)|_2\right|\right|^2_2, \quad \text{where} \tag{5.25}$$

$$|(-\nabla Z, 1)|_2 = \left(|(-(\nabla Z)_{1,1}, 1)|_2, |(-(\nabla Z)_{1,2}, 1)|_2, ..., |(-(\nabla Z)_{M,N}, 1)|_2\right) \tag{5.26}$$

defines the vector of pointwise 2-norms and the symbol $\odot$ denotes the operator for the point-wise product, also known as Hadamard product. This method is conceptually similar to the gradient-based method described in Section 5.4.3, also here the same depth constraint ensures a result in the proximity of an initial depth solution $\hat{Z}$. However, the surface orientation constraint used in Heber's method exploits the given normal field $\hat{N}$ directly instead of the gradient field $\hat{G}$. Here the normalization of ∇Z by division by the length of the vector is overcome by multiplying the term on one side by $|(-\nabla Z, 1)|_2$, which leads to a convex problem [87].

An illustration of the comparison of the measured ∇Z and the given normal $\hat{N}$ is shown in Fig. 5.4b. The given normal is scaled to the length of the measured gradients ∇Z. The contour map in Fig. 5.1 demonstrates the normal weighting of the Heber's energy term. Here, similar to the gradient-based method, a deviation in steep edges is penalized more than a deviation in flat regions. Also, a different weighting applies whether the estimated gradients are steeper or flatter than the given values.

As the energy function from Eq. (5.25) is convex and differentiable, this algorithm can be solved by an (accelerated) gradient descent method or a (fast) proximal gradient approach. For our evaluation, we used a plain gradient descent approach.Nesterov [150] proposed an accelerated gradient descent method with a simple weighted gradient step, followed by additional sliding, based on the last estimation. A fast proximal gradient method has an additional extrapolation step compared to the proximal gradient method. An example is the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [13] (see Appendix 2.2.2).

5.4.5 The Method of Nehab

The method of Nehab [148] combines depth and surface normals by solving a system of linear equations, consisting of depth and surface orientation constraints. This method is similar to the gradient-based method described in Section 5.4.3, only the surface normal information is leveraged in a different way, making different trade-offs between flat and steep gradients.

In this method, the surface normal constraint optimizes the sum of squared projections of a set of surface tangent vectors T, as defined in Eq. (5.5) through Eq. (5.7), of the reconstructed surface Z onto the given normal field $\hat{N}$. This surface normal penalty has been adopted previously in several approaches (e.g., [33, 51, 84, 85, 160]). The projection is illustrated in Fig. 5.4c. Note that the lowest penalty 0 is reached when the tangent vector is precisely orthogonal to the given normal vector $\hat{N}$. We first consider the formulation as described by Nehab [148] by formulating an overdetermined linear system of sparse equations:

$$
\begin{aligned}
Z &\approx \hat{Z}, \\
\langle \hat{N}_{i,j}, T_{i,j,x} \rangle = \hat{N}_{i,j,x} + \hat{N}_{i,j,z}(\nabla_x Z)_{i,j} &\approx 0, \quad \text{and} \\
\langle \hat{N}_{i,j}, T_{i,j,y} \rangle = \hat{N}_{i,j,y} + \hat{N}_{i,j,z}(\nabla_y Z)_{i,j} &\approx 0, \quad \forall i,j \in \mathcal{I},
\end{aligned}
\tag{5.27}
$$

which leads to the following least squares problem:

$$
\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}\left|\left|\hat{N}_z \odot \nabla_x Z + \hat{N}_x\right|\right|^2 + \frac{\lambda}{2}\left|\left|\hat{N}_z \odot \nabla_y Z + \hat{N}_y\right|\right|^2,
\tag{5.28}
$$

where the parameter $\lambda \in [0,1]$ weights the influence of the initial depth and the given normals. The global optimizer Z_{min} is calculated with a standard conjugate gradient method, with the optimality condition given as follows:

$$
Z_{min} - \hat{Z} + \lambda \Big(\nabla_x^*(\hat{N}_z \odot (\hat{N}_z \odot \nabla_x Z_{min} + \tag{5.29}
$$
$$
\hat{N}_x)) + \nabla_y^*(\hat{N}_z \odot (\hat{N}_z \odot \nabla_y Z_{min} + \hat{N}_y))\Big) = 0.
$$

The weighting of the surface normal information is explained in more detail in Fig. 5.4c and further illustrated in a contour plot in Fig. 5.2. The polar deviation to the given surface orientation in steep and flat regions is penalized similarly to the method of Heber (see Fig. 5.1, bottom row, left). However, the lateral deviations are weighted equally which is the same behavior as the ideal geodesic distance function (see Fig. 5.1, top row, right).

Note that the approach of Nehab, as described in Eq. (5.27), corresponds to the gradient-based method with an additional local N_z-weighting applied to both sides

of the surface normal constraint of Eq. (5.16). This can be shown by utilizing the equivalences defined in Eqs. (5.11) and (5.12) as follows:

$$\hat{N}_{i,j,x} = \hat{N}_{i,j,z} \, (\nabla_x Z)_{i,j} \quad \text{and} \quad \hat{N}_{i,j,y} = \hat{N}_{i,j,z} \, (\nabla_y Z)_{i,j}, \quad \forall i,j \in \mathcal{I}, \qquad (5.30)$$

and formulating a corresponding overdetermined linear system of equations:

$$
\begin{aligned}
Z &\approx \hat{Z}, \\
\hat{N}_{i,j,z} \, (\nabla_x Z)_{i,j} &\approx \hat{N}_{i,j,z} \, \hat{G}_{i,j,x}, \quad \text{and} \\
\hat{N}_{i,j,z} \, (\nabla_y Z)_{i,j} &\approx \hat{N}_{i,j,z} \, \hat{G}_{i,j,y}, \quad \forall i,j \in \mathcal{I}.
\end{aligned}
\qquad (5.31)
$$

Compared with the gradient-based method, the N_z-weighting inherent to Nehab's method improves the behavior of the penalty function by weakening the influence of regions with steep edges. Rationale behind the N_z-weighting follows from the fact that flat regions exhibit a higher value of N_z (close to 1) while steep regions receive low N_z values (close to 0). Such a local weighting helps preventing the over-penalization in steep regions as observed in the gradient-based method.

5.4.6 Generalized Nehab

In order to allow the normal weighting to reach a closer proximity to the geodesic distance, we propose a generalization of the method of Nehab. We extend on the concept of the gradient N_z-weighting as explained in Eq. (5.31) by introducing an additional exponent r that controls influence of the N_z-weighting such that:

$$
\begin{aligned}
Z &\approx \hat{Z}, \\
(\hat{N}_{i,j,z})^r \, (\nabla_x Z)_{i,j} &\approx (\hat{N}_{i,j,z})^r \, \hat{G}_{i,j,x}, \quad \text{and} \\
(\hat{N}_{i,j,z})^r \, (\nabla_y Z)_{i,j} &\approx (\hat{N}_{i,j,z})^r \, \hat{G}_{i,j,y}, \quad \forall i,j \in \mathcal{I}.
\end{aligned}
\qquad (5.32)
$$

We optimize the corresponding least squares problem with a standard conjugate

gradient method:

$$\min_{Z} \quad \tfrac{1}{2} \ ||Z - \hat{Z}||^2 + \frac{\lambda}{2} \left|\left| (\hat{N}_z)^r \odot \nabla_x Z - (\hat{N}_z)^r \odot \hat{G}_x \right|\right|^2 + \tag{5.33}$$

$$\tfrac{\lambda}{2} \ \left|\left| (\hat{N}_z)^r \odot \nabla_y Z - (\hat{N}_z)^r \odot \hat{G}_y \right|\right|^2,$$

where we find a global optimizer Z_{min} that satisfies the following optimality condition:

$$Z_{min} - \hat{Z} \ + \ \lambda \left(\nabla_x^* \left((\hat{N}_z)^r \odot \left((\hat{N}_z)^r \odot \nabla_x Z_{min} - (\hat{N}_z)^r \odot \hat{G}_x \right) \right) \right. \tag{5.34}$$

$$\left. + \ \nabla_y^* \left((\hat{N}_z)^r \odot \left((\hat{N}_z)^r \odot \nabla_y Z_{min} - (\hat{N}_z)^r \odot \hat{G}_y \right) \right) \right) = 0.$$

The influence of varying r on the behavior of the surface orientation normal penalty function is shown in Fig. 5.2. It can be seen that with $r = 0.5$ the generalized method of Nehab penalizes steeper slopes stronger than the original method of Nehab (equivalent to $r = 1$) and weaker than the gradient-based method (equivalent to $r = 0$). On the other hand, with $r = 1.6$ the proposed method exhibits a more tolerant behavior towards steeper inclination angles than the original method of Nehab, resulting in a global behavior that is reasonably close to the geodesic penalty in both the polar and lateral deviation directions. A more natural weighting, which is in the proximity of the geodesic distance, can be achieved by the adaption of the parameter r. A well balanced distance measure such as with $r = 1.6$ has the potential of performing better in steep regions (high inclination angle) while preserving performance in the flat regions. The parameter might still be further optimized.

Regarding the optimal choice of r, we have recognized that $r = 1.6$ behaves in the vicinity to the geodesic distance. If the structure of the data shows strong noise on edges steeper than approximately 50 degrees, a lower choice of r can be considered.

5.4.7 Total Generalized Variation

In this section, we introduce another novel method to reconstruct a refined surface with a hybrid depth and surface normal approach that is based on a Total Generalized Variation (TGV) approach as introduced in [26]. The TGV is an extension of the Total Variation (TV), which is a very popular and efficient regularization technique used currently in many image processing applications, however, it is known for producing staircasing artifacts in slope regions of the solution. In contrast, the TGV overcomes the staircasing problem by allowing solutions of higher order. Our formulation is a gradient-based approach, which restricts the surface to be close with a quadratic penalty to an initial depth solution $\hat{Z}$, while enforcing the auxiliary gradient field G to be in the proximity of the given gradient field $\hat{G}$, propagated through the TGV penalty. Thereby, we simultaneously reconstruct the surface Z and the auxiliary gradient field G, such that our discrete model is given as follows:

$$\min_{Z,G} \alpha_1 ||\nabla Z - G||_{2,1} + \alpha_0 ||\boldsymbol{\nabla} G||_{2,1} + \frac{\alpha}{2}||Z - \hat{Z}||^2 + \frac{\beta}{2}||G - \hat{G}||^2, \qquad (5.35)$$

where the gradient operator $\boldsymbol{\nabla} : \mathbb{R}^{M \times N \times 2} \to \mathbb{R}^{M \times N \times 4}$ computes finite differences and $\boldsymbol{\nabla} G$ can be decomposed into $(\nabla G_x, \nabla G_y)$, where ∇ was defined in Eq. (2.9). Therefore, the first and second order components of our TGV regularization have the following form:

$$||\nabla Z - G||_{2,1} = \sum_{i,j} \sqrt{(\nabla_x Z)_{i,j}^2 + (G_{i,j,x})^2 + (\nabla_y Z)_{i,j}^2 + (G_{i,j,y})^2}, \quad \text{and} \qquad (5.36)$$

$$||\boldsymbol{\nabla} G||_{2,1} = \sum_{i,j} \sqrt{(\nabla_x G_x)_{i,j}^2 + (\nabla_y G_x)_{i,j}^2 + (\nabla_x G_y)_{i,j}^2 + (\nabla_y G_y)_{i,j}^2}. \qquad (5.37)$$

The function in Eq. (5.35) consists of a depth and an orientation constraint as well as the TGV regularization terms. The depth constraint (i.e., the third term of Eq. (5.35)) enforces the depth solution Z to stay in the vicinity of our noisy initial depth map $\hat{Z}$. The orientation constraint (i.e., the fourth term of Eq. (5.35)) enforces G to stay in the proximity of the given surface gradient estimation $\hat{G}$. The TGV regularization part comprises a first and a second order term (first two terms of Eq. (5.35)). The latter represents the TV component and penalizes the l_1-norm of the second order

gradient field, which forces the gradient field G to be piecewise constant. The former enforces the approximated gradient of the depth map ∇Z to stay in the proximity of the auxiliary gradient field G also through an l_1 penalty.

Given our primal problem in Eq. (5.35) we formulate a primal-dual (PD) problem, which belongs to the class of saddle-point problems, as follows:

$$\min_x \max_y (Kx)^T y - F^*(y) + H(x),\tag{5.38}$$

where H describes the depth and orientation constraints, F defines the TGV component, with its convex conjugate F^* and the linear operator K is defined as:

$$K = \begin{pmatrix} \nabla & -I \\ 0 & \nabla \end{pmatrix}.\tag{5.39}$$

Our primal variable is denoted as $x \in \mathbb{R}^{M \times N \times 3}$ and $y \in \mathbb{R}^{M \times N \times 6}$ represents the dual variable:

$$x = \begin{bmatrix} Z \\ G_x \\ G_y \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad y_1 = \begin{bmatrix} y_{1,1} \\ y_{1,2} \end{bmatrix}, \quad y_2 = \begin{bmatrix} y_{2,1} \\ y_{2,2} \\ y_{2,3} \\ y_{2,4} \end{bmatrix}.\tag{5.40}$$

where y_1 and y_2 hold the dual variables of the first and the second term in Eq. (5.35) respectively, with y_1 holding two terms for each ∇_x and ∇_y, while y_2 holds four terms, each for ∇_x and ∇_y in G_x and G_y. For $F(x) = \alpha||x||_p$, where $||\cdot||_p$ is an l_p-norm, we calculate the convex conjugate as follows:

$$F^*(y) = \delta_{||\cdot||_{2,\infty} \leq \alpha}(y) = \begin{cases} 0, & \text{if } |y_{i,j}|_2 \leq \alpha, \forall i,j \\ \infty, & \text{otherwise,} \end{cases}\tag{5.41}$$

which is the indicator function of the polar ball. Therefore, F^* and H are defined as follows:

$$F^*(y) = \delta_{||\cdot||_{2,\infty} \leq \alpha_1}(y_1) + \delta_{||\cdot||_{2,\infty} \leq \alpha_0}(y_2), \quad \text{and}\tag{5.42}$$

$$H(x) = \frac{\alpha}{2}||Z - \hat{Z}||^2 + \frac{\beta}{2}||G - \hat{G}||^2.\tag{5.43}$$
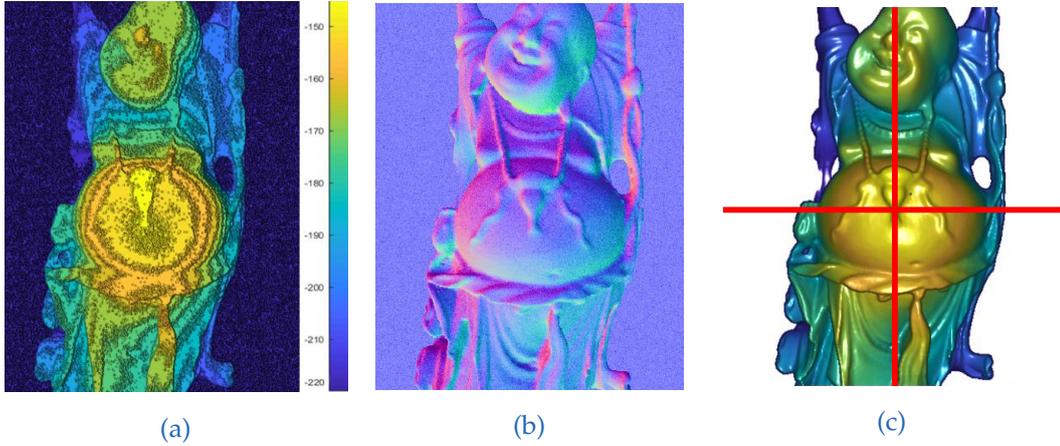
Figure 5.5: Example evaluation data. Consisting of a noisy quantized depth $\hat{Z}$ and noisy surface normals $\hat{N}$. The ground truth depth $\hat{Z}_{GT}$ is shown with a horizontal and vertical cross-section line, which indicates the positions of evaluations in the following Fig. 5.6. (**a**) Initial depth $\hat{Z}$; (**b**) Given surface normals $\hat{N}$; (**c**) Ground truth depth $\hat{Z}_{GT}$.

An optimal solution to our hybrid formulation is found with the PD algorithm, as described in Appendix 2.2.2. For a more detailed description of the PD algorithm, see [30].

5.4.8 Evaluation

To evaluate all considered algorithms we need two data structures, namely an initial depth map $\hat{Z}$ and surface orientation information such as surface normals $\hat{N}$ or gradients $\hat{G}$. Initial depth maps can be provided by stereo or light field correspondence analysis or other depth scanning methods. An estimate of the surface orientation is calculated using e.g., photometric stereo or polarization imaging. Our discussed hybrid algorithms differ mainly in how the surface orientation information is taken into account. In particular, the penalty functions associated with the surface orientation constraints vary significantly for different methods, as illustrated in Fig. 5.1 through Fig. 5.4c. We proposed two novel methods in Sections 5.4.6 and 5.4.7, namely the generalized method of Nehab and the TGV approach. The former balances the surface orientation information with an improved weighting function withing the least squares

(a) Depth result

(b) Vertical line

(c) Horizontal line

(d) Error map
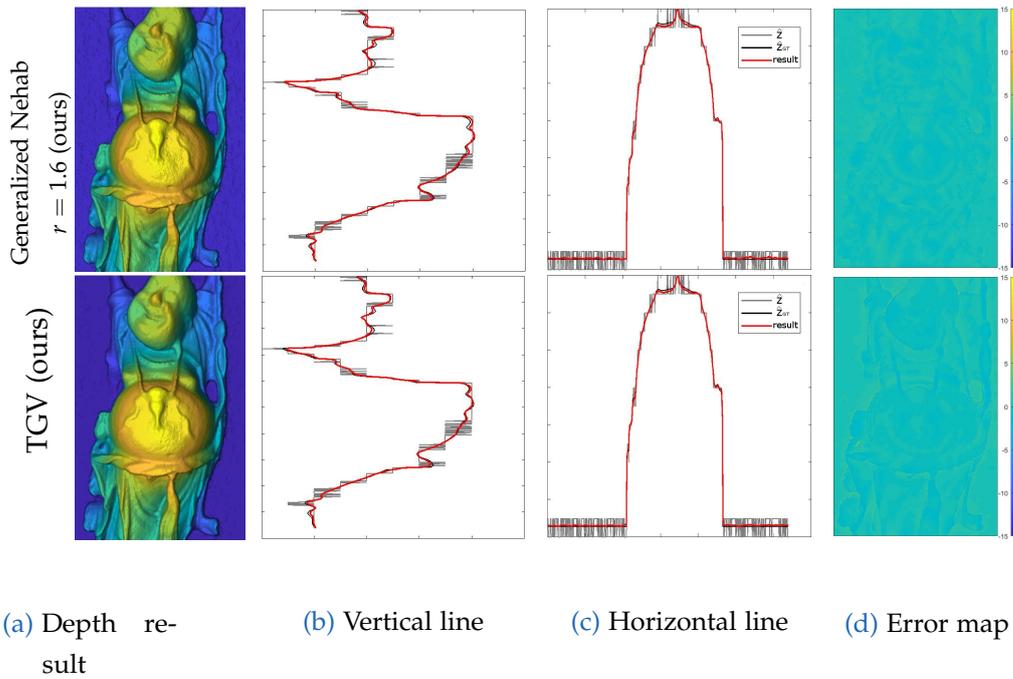
Figure 5.6: Qualitative evaluation of our novel approaches. Previously introduced in Sections 5.4.6 and 5.4.7. The following methods are demonstrated: (1st row) the generalized method of Nehab with $r = 1.6$, (2nd row) the TGV approach. The left column shows the color coded depth reconstructions as delivered by different methods. The middle two columns show the vertical and horizontal depth profiles as marked in Fig. 5.5c. In each of these plots, the red, gray and black lines indicate the reconstructed depth Z, the initial depth $\hat{Z}$ and the ground truth depth $\hat{Z}_{GT}$, respectively. The error maps showing signed distances to the ground truth depth are provided in the right column.
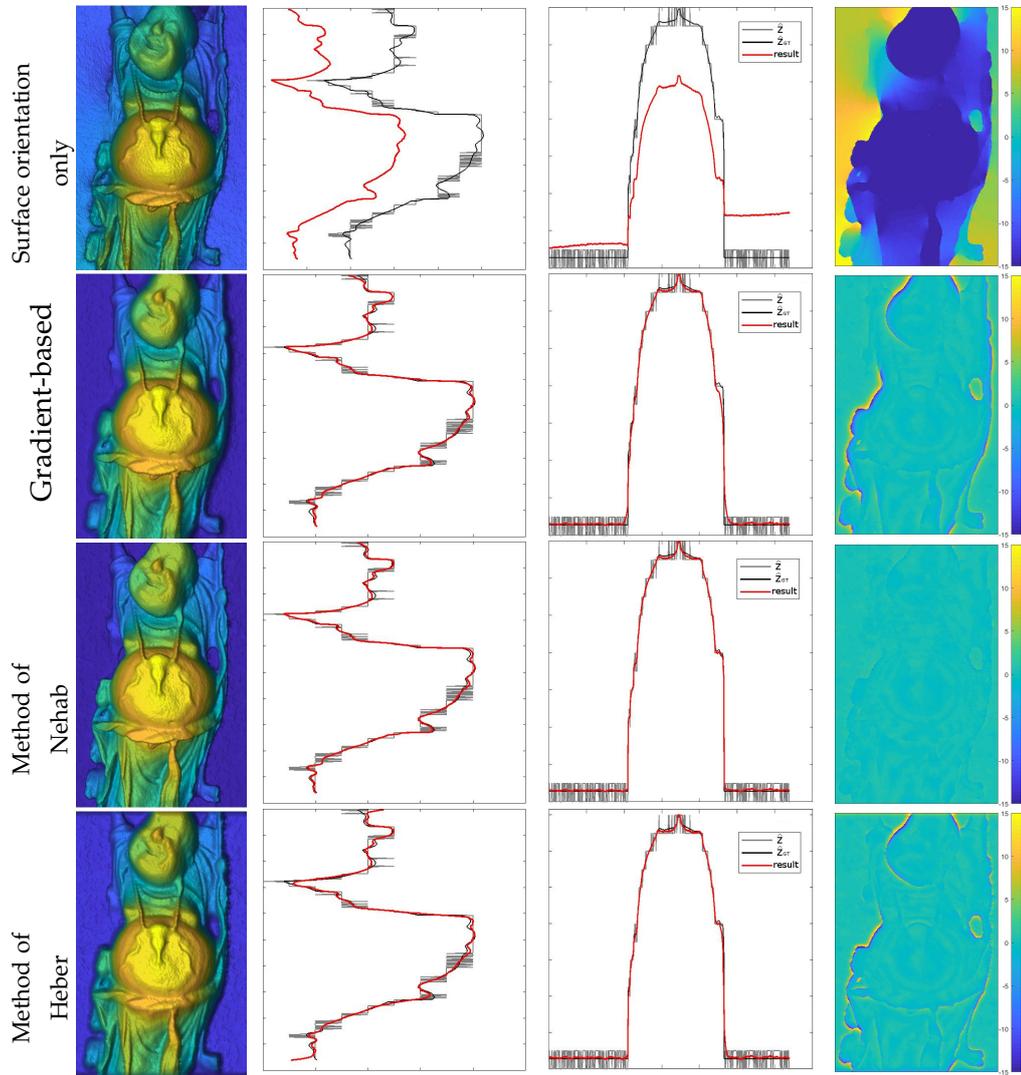
(a) Depth result

(b) Vertical line

(c) Horizontal line

(d) Error map

Figure 5.7: Qualitative evaluation of the state-of-the-art methods. Demonstrated methods: (1st row) the gradient-based method using surface orientations only, (2nd row) the combined gradient-based method, (3rd row) the method of Nehab, and (4th row) the method of Heber. The illustration is organized as Fig. 5.6.

framework. The latter uses a TGV regularization term combined with a gradient-based approach.

Qualitative and Quantitative Evaluation

In order to perform a comprehensive quantitative evaluation of the above mentioned methods, we considered using synthetic evaluation data that comprises full ground truth (GT) information. The initial depth $\hat{Z}$ for the synthetic evaluation data is given by a ground truth depth map $\hat{Z}_{GT}$, which is thresholded after adding noise:

$$\hat{Z} = \frac{1}{k}\left[k \cdot (\hat{Z}_{GT} + noise)\right]. \tag{5.44}$$

In this study, we considered a normally distributed additive noise as this type of noise well simulates the behavior of matching errors of stereo or light field methods combined with image sensor noise. The noise used in our evaluations has a maximum amplitude of 7% of the depth range. The constant k defines the number of discretization steps and $[\cdot]$ rounds to the nearest integer number $n \in \mathbb{Z}$. These discretization artifacts attempt to simulate the output of the discrete regularized correspondence analysis, which is often applied in real-world scenarios (e.g., as described in [9]).

For the orientation constraints, we assume surface normals $\hat{N}$ derived from the ground truth depth model $\hat{N}_{GT}$ by adding a normally distributed noise with a maximum amplitude of 23% of the normal range in the spherical coordinate system:

$$\hat{N} = \frac{\hat{N}_{GT} + noise}{||\hat{N}_{GT} + noise||} \tag{5.45}$$

In Fig. 5.5, examples of the evaluation data $\hat{Z}$ and $\hat{N}$ as well as the corresponding ground truth depth $\hat{Z}_{GT}$ are shown. All 3D datasets used for evaluations were taken from the Stanford 3D scanning repository [186] and rendered with POV-Ray [163].

Qualitative comparisons of the methods described in Section 5.4 are presented in Fig.s 5.6 and 5.7. The left column shows the color coded depth reconstructions as delivered by different methods. The middle two columns show the vertical and

horizontal depth profiles as marked in Fig. 5.5c. In each of these plots, the red, gray and black lines indicate the reconstructed depth Z, the initial depth $\hat{Z}$ and the ground truth depth $\hat{Z}_{GT}$, respectively. The error maps showing signed distances to the ground truth depth are provided in the right column. The corresponding geodesic distances from the ground truth surface normals $\hat{N}_{GT}$ for each method are displayed in Fig. 5.8. Results of the quantitative evaluations are shown in Tab. 5.2 for gradient-based methods and in Tab. 5.3 for normal based methods. For the evaluation we used three datasets from [186]: Buddha (shown in qualitative analysis), Dragon, and Armadillo. The tables hold fractional precisions of two digits in the depth evaluation and a higher fractional precision of four digits in the normal evaluation due to different ranges. The average is calculated using 4 digits after the comma, rounding errors can cause differences in the last digit. The MSE distance in the depth domain is calculated by the quadratic distance between the ground truth depth and the depth result:

$$MSE_Z = d_Z = \frac{1}{MN}||\hat{Z}_{GT} - Z||^2 \tag{5.46}$$

The geodesic distance is calculated as described in Section 5.4.1 by using the following equation:

$$GEO_N = d_N = \frac{1}{MN} \sum_{i,j}^{MN} acos \left(\frac{\langle \hat{N}_{i,j}, (-(\nabla_x Z_{GT})_{i,j}, -(\nabla_y Z_{GT})_{i,j}, 1) \rangle}{|(-(\nabla_x Z_{GT})_{i,j}, -(\nabla_y Z_{GT})_{i,j}, 1)|_2} \right) \tag{5.47}$$

As can be seen in Figs. 5.7 (1st row) and 5.8b, using surface orientation information alone in the gradient-based formulation provides visually pleasing detail reconstructions ($d_N = 0.2365$), though it is performing worst in terms of the absolute distance to the ground truth depth with an average MSE of $d_Z = 66.68$. In Figs. 5.7 (2nd row) and 5.8c, one can see that adding a depth constraint to the same gradient-based formulation improves the result drastically ($d_Z = 2.04$ and $d_N = 0.2951$). Nevertheless, this method still shows somewhat low performance around steep edges. Note that each of the demonstrated methods could be used to reconstruct surfaces from surface normals only solely by dropping the depth term. The method of Nehab shown in Figs. 5.7 (3rd row) and 5.8d exhibits the capability to improve the result over the previous methods exploiting a better surface orientation weighting strategy ($d_Z = 0.14$, $d_N = 0.2532$). These methods are optimized using a least squares solver. The evaluation
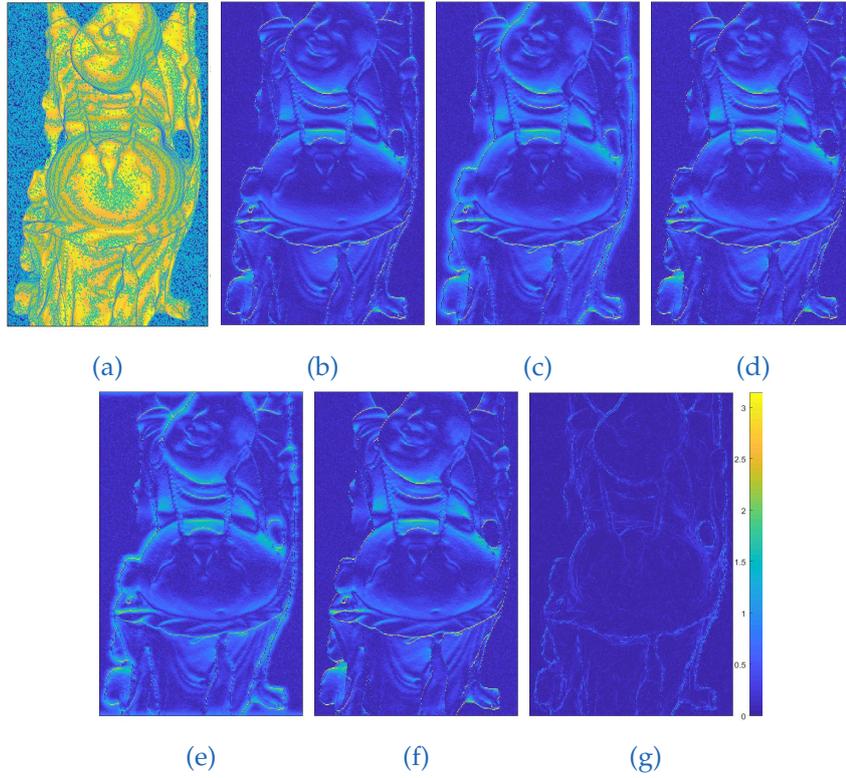
Figure 5.8: Geodesic distances of the ground truth normals. Compared to the normals from the initial depth map $\hat{Z}$ as well as the distances to the normal results of the presented methods (in the same order as displayed in Figs. 5.6 and 5.7). (**a**) Initial depth $\hat{Z}$; (**b**) Surface orientation only; (**c**) Gradient-based; (**d**) Method of Nehab; (**e**) Method of Heber; (**f**) Generalized Nehab $r = 1.6$ (ours); (**g**) TGV (ours).

of Heber's method is shown in Figs. 5.7(4th row) and 5.8e. The method is optimized using gradient descent and reaches average results of $d_Z = 1.79$ and $d_N = 0.3049$, which is significantly worse than the method of Nehab. The results of our generalized method of Nehab with a parametrization $r = 1.6$ are shown in in Figs. 5.6(1st row) and 5.8f. This method improves robustness over the standard method of Nehab against noise in surface normals and outperforms all other evaluated methods in the absolute depth error domain with $d_Z = 0.11$. As the normal weighting here is in closer vicinity to the geodesic penalty than the method of Nehab, this approach reaches an improved normal error of $d_N = 0.2442$. This method is optimized using a least squares solver.

Gradient-Based	Dataset Method	$\hat{Z}$	Surface Orientation Only	Gradient Based	Gradient Based + Reg. with Laplacian Smoothness (Eq. (5.21))	Gradient Based + Reg. with Gradient (Eq. (5.19))	TGV (Ours)
Depth [MSE_Z]	Dragon	4.23	34.05	2.04	2.15	1.85	**0.19**
	Buddha	4.85	117.29	2.12	2.25	2.01	**0.22**
	Armadillo	4.60	48.71	1.95	2.06	1.83	**0.18**
	Average	4.53	66.68	2.04	2.15	1.90	**0.20**
Normals [GEO_N]	Dragon	0.8226	0.2776	0.3344	0.3200	0.3017	**0.0664**
	Buddha	0.8767	0.1922	0.2535	0.2339	0.2125	**0.0668**
	Armadillo	0.8611	0.2397	0.2973	0.2797	0.2599	**0.0666**
	Average	0.8535	0.2365	0.2951	0.2779	0.2580	**0.0666**

Table 5.2: Gradient-based: quantitative evaluation of the distance of the optimized depth values to the ground truth depth. The evaluations were performed on objects from the Stanford database [186], which were rendered using POV-Ray [163]. Evaluated are the MSE to the ground truth depth and the geodesic distance to the ground truth normals.

Our novel TGV method, shown in Figs. 5.6(2nd row) and 5.8g, provides by far the best normal accuracy of $d_N = 0.0666$ and performs among the best in the depth domain ($d_Z = 0.20$). It is optimized with a primal-dual algorithm.

For completion, we additionally show the results of two regularized gradient methods in Tab. 5.3. First, we regularize with smoothness as shown in Eq. (5.21), which can be used for sparse data. In our dense case, where we used scalars for the weighting parameters, the normal accuracy shows a minor improvement due to smoothing in return of a weaker depth accuracy. Second, we regularize with the gradient, as shown in Eq. (5.19). With this we can reach an improvement both in the depth and normal accuracy. These regularization terms could be used for all presented methods, which would exceed the scope of this paper. Note that our novel TGV method shows a significantly better performance in both accuracy measurements.

As can be seen in Tabs. 5.2 and 5.3, the two best performing methods are our generalized method of Nehab with $r = 1.6$ (normal based) and our novel TGV method (gradient-based). We show the convergence of both in Fig. 5.9 with respect to the depth error defined in Eq. (5.46) and the normal error as defined in Eq. (5.47). While

Normal Based	Dataset Method	$\hat{Z}$	Method of Heber	Method of Nehab	Generalized Nehab $r = 1.6$ (Ours)
Depth $[MSE_Z]$	Dragon	4.23	2.01	0.13	**0.10**
	Buddha	4.85	1.60	0.15	**0.12**
	Armadillo	4.60	1.75	0.13	**0.10**
	Average	4.53	1.79	0.14	**0.11**
Normals $[GEO_N]$	Dragon	0.8226	0.3474	0.2941	**0.2849**
	Buddha	0.8767	0.2579	0.2102	**0.2013**
	Armadillo	0.8611	0.3094	0.2553	**0.2464**
	Average	0.8535	0.3049	0.2532	**0.2442**

Table 5.3: Normal based: quantitative evaluation of the distance of the optimized depth values to the ground truth depth. The evaluations were performed on objects from the Stanford database [186], which were rendered using POV-Ray [163]. Evaluated are the MSE to the ground truth depth and the geodesic distance to the ground truth normals.

the generalized Nehab converges in both terms after approximately 25 iterations, the TGV settles at around the same iteration step with a depth error which is performing slightly worse than the other method, but continues to highly optimize the normal error. Note that a gradient-based formulation was chosen to demonstrate graceful properties of our novel TGV approach. It significantly improved the results compared with the standard gradient-based method. Alternatively, for even better performance, other penalty functions such as the generalized method of Nehab, can be chosen and will be a matter of future research.

We demonstrated the described algorithms with different surface orientation weighting on a real world example, as shown in Fig. 5.10. This object was acquired with a multi-line scanner, the hybrid light field - photometric stereo acquisition framework described in [9]. In this very specific case we only have one normal direction. An example how to deal with this data in the gradient-based approach is weighting the gradient vector in Eq. (5.21) in the missing direction with zero. Specifics of the multi-line-scanning environment are out of the scope of this paper and are a matter of future research.

(a) Generalized Nehab $r = 1.6$ (ours)　　　　　(b) TGV (ours)
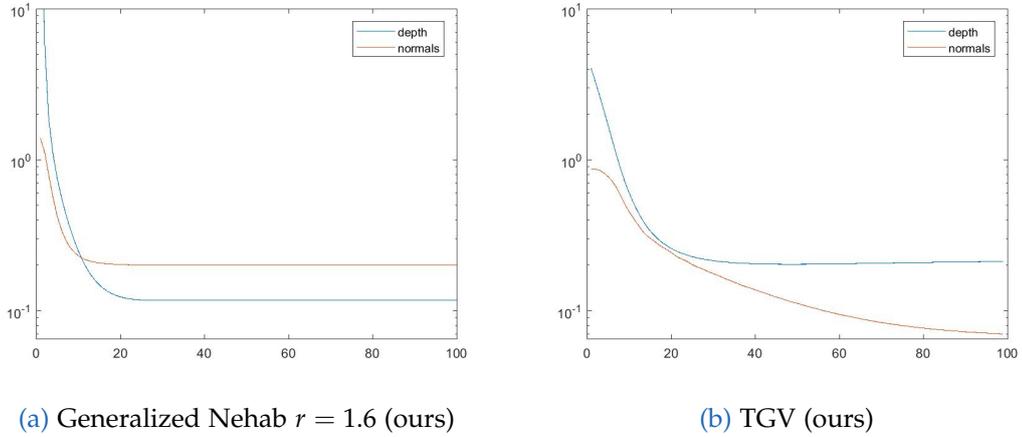
Figure 5.9: Convergence analysis. Comparison of the two winning methods: the generalized Nehab with $r = 1.6$ (**a**) and the TGV (**b**). Distances to the ground truth depth and surface normals are shown after each iteration, plotted with a logarithmic y-axis for better visibility.

5.5 Conclusions

We presented a review and classification of methods combining depth and surface orientation data (normals or gradients), in order to reach an improved surface depth estimation. State-of-the-art methods differ mostly in the formulation of the surface orientation constraint (see Section 5.4) and capabilities of the method-specific solvers.

We illustrated the differences between various formulations of the surface orientation constraint and explained performance discrepancies. Furthermore, we used our findings to introduce a generalization of the method of Nehab (see Section 5.4.6) that significantly outperforms other methods in terms of absolute depth accuracy. Additionally, we introduced a novel method based on TGV (see Section 5.4.7), which outperforms all other methods in the surface normal domain and shows a competitive performance in the depth accuracy. While our generalized Nehab method converges faster (see Fig. 5.9) and gives the most accurate result in the depth domain, our TGV based approach refines the surface orientation further and converges at the most accurate orientation result with a high accuracy in the depth domain.
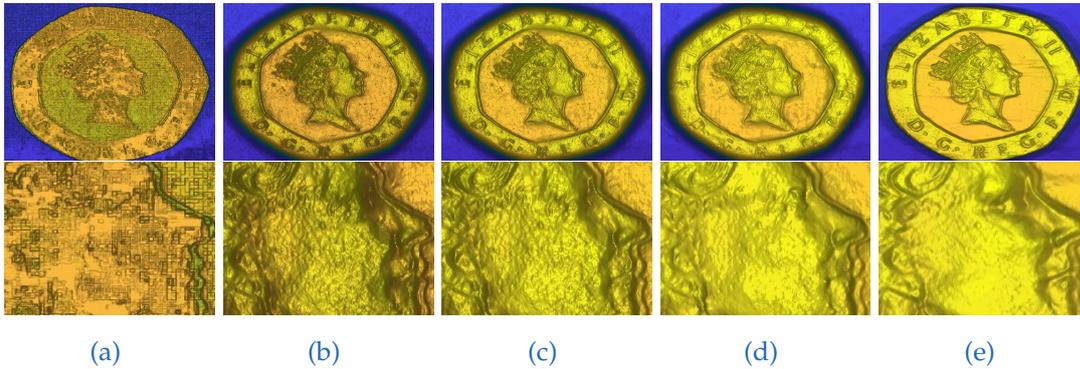
Figure 5.10: Real world evaluation of a coin object. The object was acquired by a multi-line scanner. The acquisition setup was previously described in detail in [9]. (**a**) $\hat{Z}$; (**b**) Gradient-based; (**c**) Method of Nehab; (**d**) Generalized Nehab (ours); (**e**) TGV (ours).

Further research will include the specialization on line-scanning algorithms, TGV weighting adaption and computational acceleration. With specialized hybrid algorithms that fit data from line scanning sensors we will determine a solution with incomplete surface orientation data. The surface orientation constraint of our TGV formulation is currently gradient-based. Plugging in another formulation with a better balanced normal weighting could improve the results even further and will be a matter of future research. Furthermore we will focus on computational acceleration of the proposed algorithms, where we will exploit their inherent structure to achieve efficient parallelization.

6 Multi-Line Scan Depth and Surface Normal Fusion Algorithms

Contents

In this chapter we present depth refinement approaches for multi-line scan systems. This is an extended version of our previously published papers. In [9] we introduced the analysis of photometric stereo data from light field stacks and the combined depth reconstruction. We extended the idea in [8], evaluated several state-of-the-art optimization algorithms on the task and introduced two novel variational approaches.

Before we dive into the details of depth optimization in Sec. 6.6 (illustrated in Fig. 6.1 step 5), we describe our hybrid multi-line scan setup in Sec. 6.2, the depth analysis (step 2) in Sec. 6.2.1 and the retrieval of photometric stereo information in Sec. 6.2.2. We discuss the feature generation for multi-line scan data in Sec. 6.3 and introduce our novel hand crafted features which are highly suitable to process industrial in-line

light field acquisitions. The computation of a regularized initial depth estimation is described in Sec. 6.4 and of initial photometric stereo normals using the initial depth map is explained in Sec. 6.5. Our novel optimization methods for depth refinement in multi-line scan data are evaluated in comparison to the state-of-the-art in Sec.6.7 on three synthetic datasets as well as a real world example.

6.1 Motivation

Recent work has shown the improved depth reconstruction by combining depth and surface normal information. In this chapter, we build on the findings and introduce novel variational methods for a refined depth reconstruction using a multi-line scanner. In this specific setup, the object is acquired while moving on a conveyor belt in a defined direction under the camera, which simultaneously captures light field and photometric stereo data as the object is transported. We perform our experiments on virtual and real-world data and achieve significantly improved results over state-of-the-art methods both in depth and surface normal accuracy.

In this chapter, we combine the advantages of a light field depth estimation that is accurate on large scales with strong local surface orientation estimates from photometric stereo. We adapt state-of-the art algorithms to work on data comprising only surface normal data in x-direction, as provided by multi-line scanners, for which we furthermore introduce a novel total generalized variation (TGV) approach as well as an improved normal weighting for our energy terms.

Analyzing light field data allows the depth reconstruction of a scene. In Sec. 3.4 we illustrated such a correspondence analysis using irradience values captured on the camera sensor. In feature based matching approaches, these irradiance values are first processed in order to extract feature maps for each view. Using features allows more robust matching, with a higher invariance to image transformations. We present new features which are tailored to the reflectance and noise properties of multi-line scan acquisitions.

An initial depth estimate is calculated by a multi-view correspondence analysis in the light field data, the estimate is optimized by a conditional random field (CRF). Then the normals at each position are estimated using a photometric stereo analysis. We reformulate state-of-the art approaches to combine depth with surface normals for multi-line-scan data and finally we refine the depth estimate by using a total generalized variation (TGV) method. The results are compared to a combinational approach from Nehab [148], which we adapt to the properties of our multi-line-scan data.

Several methods for depth analysis from light field data were previously presented. Most methods exploit the structure of Epipolar Plane Images (EPIs), which were described by Bolles *et al.* [22], where depth analysis is performed by fitting straight lines through EPIs. Wanner *et al.* [214] used a structure tensor to give a better local depth estimate. Kim *et al.* [110] introduced a fine-to-coarse method, which first estimates boundaries ray-wise and then propagates the information to more homogeneous interior regions. A dense depth estimation method is described by Tosic *et al.* [204]. Multi-scale light field analysis is performed by exploiting light field scale and depth (Lisad) spaces, which allows the depth reconstruction in uniform as well as textured regions. Tao *et al.* [198] proposed a line consistency metric for depth estimation with light field data to overcome Lambertian restrictions.

Photometric stereo analysis was introduced by Woodham [223]. Subsequently, several methods were developed for dealing with non-Lambertian surface structures [123, 138, 179]. The analysis of normals from photometric stereo in a multi-line scan setup was described in [9].

The combination of depth with surface normals is usually done by constructing an energy term, containing depth and orientation constraints. The depth constraint enforces the result to be close to an initial depth estimate. The weighting of the surface orientation constraints differentiates in alternative approaches. Gradient based techniques, as shown in [62, 87], enforce the surface gradient to be close to an estimated gradient. Nehab *et al.* [148] optimized the sum of squared projections of the surface tangents onto a given normal.
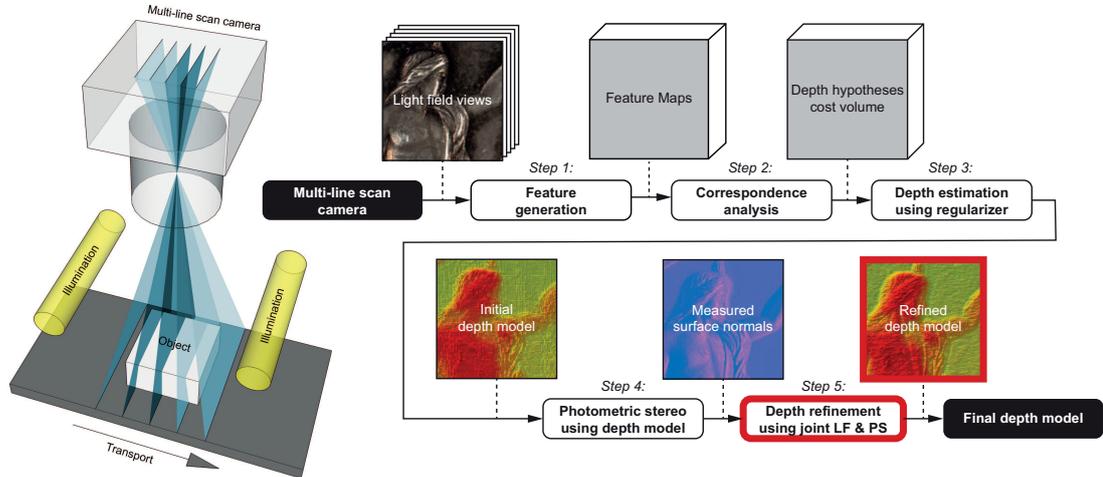
Figure 6.1: Illustration of our multi-line scan setup with the depth refinement pipeline. The setup consists of a top-view camera and two illumination sources. In this chapter we focus on depth refinement techniques.

We introduce an adaption of state-of-the art hybrid methods which were formally presented in [7] to fit the restricted structure of the multi-line scan data as well as a tailored TGV approach, which enforces closeness of the x-gradient while simultaneously reconstructing a y-component. Additionally we show an improved weighting of the surface orientation constraint.

6.2 Hybrid Setup

We acquire light field and photometric stereo data with an in-line multi-line scan system, which was described in [190]. A schematic representation of the system is shown in Fig. 6.1 (left). A top-center mounted multi-line scan camera (AVT BONITO CL-400C) extracts frames with multiple scan lines with frame rates of up to 50 kHz, the camera is equipped with a Schneider-Kreuznach APO-COMPONON 4/45 lens. During the acquisition process an object is moving on a conveyor belt (Thorlabs LTS300/M)

orthogonally to the camera's optical axis and sensor line orientation. Two static line light sources (Chromasens CORONA II) are used to illuminate the scene.

6.2.1 Light Field Cues

Contrary to the common 4D light field description, we capture a linear light field with our multi-line scan system, consisting of two spatial and one directional dimension. During the image acquisition, an object is moving in a defined direction on a conveyor belt. From the image sensor several equidistant lines are read out at each time step. Between two acquisitions the object is moving exactly the distance equivalent to one pixel, which ensures equivalent resolution in both the x- and y-dimension of the image. For the light field depth analysis, these acquisitions are represented as an EPI stack (shown in Fig. 4.5b).

The slope angles of the EPI structure correspond to defined depths for each image position. We analyze the depth with a multi-view correspondence analysis in the EPI domain normalized for brightness and contrast to achieve robustness against variations due to photometric stereo effects.

Starting from the compound light field & photometric stereo data obtained using the multi-line scan camera, we first perform a multi-view correspondence analysis in the EPI domain using image features, which are described in detail in Sec. 6.3. This is performed in order to reach a robustness against brightness and contrast variations in different views, which happens quite often due to the presence of photometric effects in our data.

During the correspondence analysis a number of $\mathcal{Z}$ disparity hypotheses are tested in each pixel location from which results in a cost volume $C \in \mathbb{R}^{\mathcal{M} \times \mathcal{N} \times \mathcal{Z}}$, on a discretized surface with the size of $\mathcal{M} \times \mathcal{N}$ pixels. Where each value reflects the similarity of visual structures at the corresponding locations in the light field views. For convenience, values in the cost volume are normalized to fit in the interval $[0, 1]$.

6.2.2 Photometric Stereo Cues

Light field depth estimation performs well in terms of the absolute depth accuracy. Failures occur on fine scales, textureless regions and on surfaces with specular reflections. Using a photometric stereo method to derive surface orientation from local shading information was described by [223]. Surface shape can be retrieved by using only the surface orientation (e.g. [62]), but such methods are prone to errors and result in a low frequency offset. We combine the depth estimation with surface orientation data to achieve an overall improved reconstruction accuracy.

In Chapter 4.2.3 we presented the photometric stereo analysis from light field data. The calculation of the surface normals and the albedo was presented in Eqs. (4.8) to (4.11). We also showed that in the multi-line scan framework we can estimate only a lower bound of the albedo due to the missing photometric component which is orthogonal to the transport direction (in our case the y-component).

The surface normals can be expressed as a disparity gradient field $G \in \mathbb{R}^{\mathcal{M} \times \mathcal{N} \times 2}$ as follows:

$$G_{i,j,x} = -\frac{N_{i,j,x}}{N_{i,j,z}}, \quad G_{i,j,y} = -\frac{N_{i,j,y}}{N_{i,j,z}}, \quad \forall i, j \in \mathcal{I}. \tag{6.1}$$

An upper bound for N_z can be estimated from Eq. (4.11) under an absence of the y-component (i.e. N_y is unknown), for $N_y = 0$. Our proposed models utilize an initial depth $\hat{Z} \in \mathbb{R}^{\mathcal{M} \times \mathcal{N}}$ and surface orientation in the form of surface normals $\hat{N} \in \mathbb{R}^{\mathcal{M} \times \mathcal{N} \times 3}$ or surface gradients $\hat{G} \in \mathbb{R}^{\mathcal{M} \times \mathcal{N} \times 2}$. To account for the missing y-component, in some methods we assume flatness in that direction, i.e. N_y and G_y both equal to zero.

6.3 Features for Multi-Line Scan Data

Analyzing light field cues allows the depth reconstruction of a scene. In Sec. 3.4 we illustrated such a correspondence analysis using irradience values captured on the camera sensor. In feature based matching approaches, these irradiance values are first
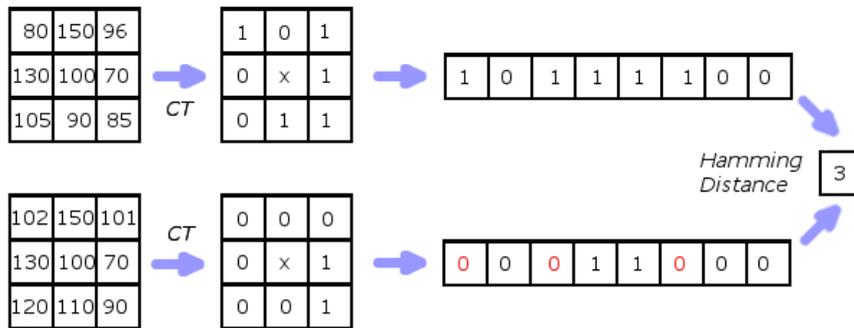
Figure 6.2: Census transform and comparison for stereo images.

processed in order to extract feature maps for each view. Using features allows more robust matching, with a higher invariance to image transformations.

Classical features are using detectors to identify structures (e.g. lines, corners, circles, curves) [157], combinations of structure features [127, 218] or local transformations, such as the census transform (CT) [236], the sign bit of the image after the convolution with the Laplacian [155] or the direction of the intensity gradient [174]. An insensitive manual matching approach was introduced in [15].

In this section we describe two manually designed feature types, which previously showed a superior performance regarding computational speed and matching accuracy. First, we present the well-known CT features, which revealed a problematic structure when processing industrial light field data. Then we introduce our novel extended high-pass features, which are specifically adapted and hence highly suitable for data captured by our multi-line scanner.

6.3.1 Census Transform

The census transform (CT) is a non-parametric local transform which was introduced in [236]. It uses relative intensities of input images which makes it robust to intensity changes. The transform maps a 8-connected pixel neighborhood to a bit string signature. Using the Hamming distance allows a fast identification of similar regions, as illustrated in Fig. 6.2. This was frequently used for stereo [103, 112, 161] or light field

matching applications [9, 183, 203]. The CT features $F_l \in \mathbb{R}^{M \times N \times T}$, for T feature and $l = (1, ..., L)$ color channels, at a pixel position p is defined as:

$$F_l(p) = \otimes_{(i,j) \in D} \xi(p, p + (i,j)), \quad \forall i, j \in \mathcal{I}, \tag{6.2}$$

where $\otimes$ denotes the concatenation, D defines the window around p and ξ is given as the relationship between the intensities I of the pixels p and $p + (i, j)$:

$$\xi(p, p + (i,j)) = \begin{cases} 1, & \text{if } I(p) \leq I(p + (i,j)) \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

The Hamming distance is used to compare the distance of the resulting feature images. Between two images, the best correspondence is found where the distance is minimal.

The census transform is not ideally suited for industrial light field applications, since it cannot be interpolated in a fast and easy way. An interpolation is necessary when testing disparity hypotheses along a slope angle, since tested values frequently lie between pixels. An interpolation method was introduced in [207] which shows speed detriment in the light field application as it requires four times as many comparisons. Our extended HP features introduced in Sec. 6.3.2 show superior properties and are therefore chosen as the state-of-the-art method for the evaluation in Sec. 6.3.3.

6.3.2 Extended High-Pass Medium-Contrast Approach

We designed features which are suitable for light field data with a low computational complexity which favors medium contrast regions. Two specifically problematic behaviors can be enclosed in line-scan light field data, namely low texture- or highly reflective surfaces. In the former matching ambiguities occur in low amplitude regions, hence the result has a low confidence. The latter shows in high amplitude regions, where dark and very bright areas converge. Resulting from these two observations, we designed our features to operate in a medium contrast range with an amplitude between chosen values r_{min} and r_{max}.

Using a high-pass (HP) filter to generate local high frequency features was previously done for stereo matching [83, 104, 232] as well as light field matching applications [32, 187]. Using high frequency matching focuses the matching on the edge regions of an image. The function computing Gaussian filter coefficients for a kernel size k is defined as:

$$H = \alpha \, e^{-\frac{(w-(k-1)/2)^2}{2\sigma^2}}, \quad \text{with} \quad w = (0,...,k-1), \tag{6.4}$$

where α denotes the scale factor to ensure $\sum_w H_w = 1$. The high-pass image $F \in \mathbb{R}^{M \times N}$ is calculated as a difference of the gray valued input image $I \in \mathbb{R}^{M \times N}$ and the Gaussian filtered result using a convolution with the filter coefficient:

$$F_{i,j} = I_{i,j} - (H * I)_{i,j}, \quad \forall i,j \in \mathcal{I}, \tag{6.5}$$

where $*$ denotes the convolution operator:

$$(H * I)_{i,j} = \sum_{u=-k}^{u=k} \sum_{v=-k}^{v=k} H_{u,v} I_{i-u,j-v}, \quad \forall i,j \in \mathcal{I}. \tag{6.6}$$

To ensure a stable and image wide matching procedure we use a weighted high frequency matching method, for which we define an amplitude matrix $A \in \mathbb{R}^{M \times N}$:

$$A_{i,j} = (H * |F|)_{i,j}, \quad \forall i,j \in \mathcal{I}, \tag{6.7}$$

and we normalize our high pass feature map:

$$\bar{F}_{i,j} = \frac{F_{i,j}}{A_{i,j}}, \quad \forall i,j \in \mathcal{I}. \tag{6.8}$$

With further processing we ensure a stable result in respect to noise regardless of the regions extensive reflective properties. We create a confidence map $C \in \mathbb{R}^{M \times N}$, which analyses the local deviation in order to assign a pixel-wise confidence (Eq. 6.9). Areas with a medium contrast between a defined range $r = (r_{min}, r_{max})$ are trusted most. Applying the confidence map on our normalized high pass features results in our high-pass medium-contrast features $M \in \mathbb{R}^{M \times N}$, we inject independent and identically distributed (IID) noise in low confidence regions in order to avoid wrong matching results.

$$C_{i,j} = e^{-\left(\frac{A_{i,j}}{r_{max}+\epsilon}\right)^2} - e^{-\left(\frac{A_{i,j}}{r_{min}+\epsilon}\right)^2} \tag{6.9}$$

$$M_{i,j} = C_{i,j}\bar{F}_{i,j} + (1-C_{i,j})\mathcal{N}_{i,j}, \quad \forall i,j \in \mathcal{I} \tag{6.10}$$

(a) Census   (b) Extended HP
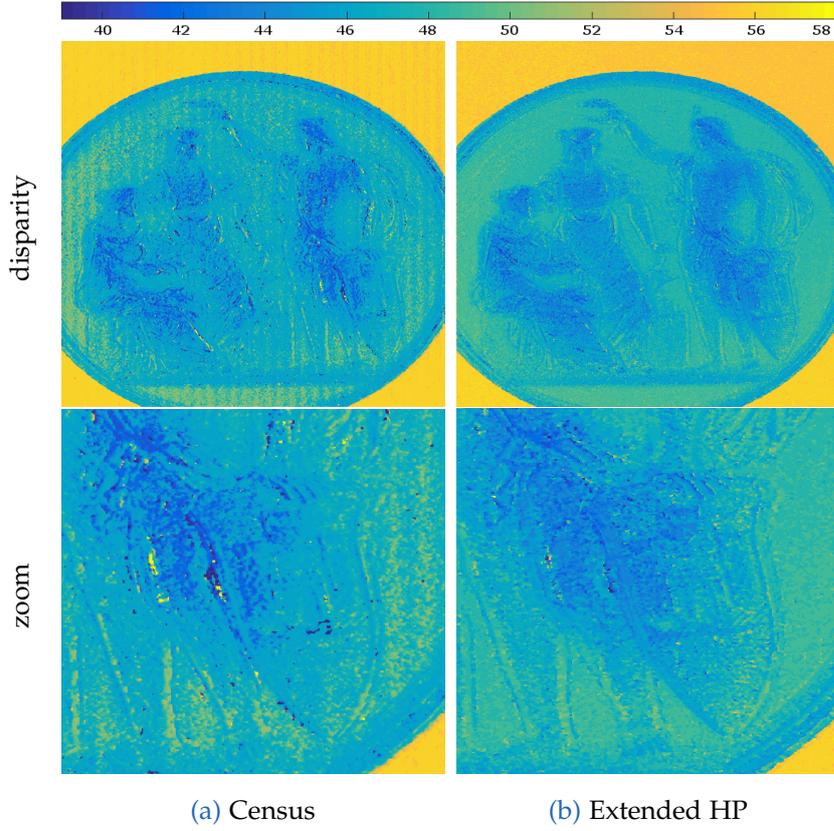
Figure 6.3: Comparison of disparity estimation results with different features.

Where $\mathcal{N} \in \mathbb{R}^{M \times N}$ is a noise matrix and ϵ denotes a small number.

We propose to compute these features at three different levels, while at each level a larger kernel size k is applied. We specifically used $k = (3,5,7)$. The redundancy in light field data is additionally used to reduce the ambiguity between matching features.

6.3.3 Comparison

We qualitatively compare the features on our multi-line scan light field stack. Both methods have $K = 3$ features for each light field view. We analyzed the correlation

volume V to get the best matching disparity for each pixel position p.

$$X_p \in \arg\min_k V_p(k) \tag{6.11}$$

The result is shown in Fig. 6.3. Note that the created features have a different precision. While our extended HP features have 8 bit, the binary CT features are converted to a single integer value with 32 bit (5×5 neighborhood).

Features from the CT are not ideal for the disparity hypotheses testing in light field image stacks. The reason for this lies in the fact that these binary features are not well suited for interpolation. Resulting disparity estimations are noisy and enhance transport jitter artifacts from the linear transport stage (see acquisition setup in Fig. 6.1). The correction of transport jitters was previous discussed in [27], we showed details in Sec. 3.3.4. Furthermore, our extended high-pass features benefit from a special treatment for specularities, which favors medium-contrast regions.

6.4 Initial Depth Estimation Using CRF

In order to assess surface normals in the next step, an initial approximate depth model must first be obtained. Given the pre-calculated hypothesis costs, we employ the discrete-continuous optimization algorithm based on conditional random fields (hereinafter referred to as the CRF algorithm, as described in [176]) to determine a quick yet accurate approximation of the global solution (i.e. globally consistent depth map), under the generalized first-order total variation (TV) prior.

Let $V = M \times N$ be a set of nodes, where each node $i \in V$ corresponds to a pixel location. Moreover, let $E \subset V \times V$ be an edge set, where each $ij \in E$ corresponds to an edge connecting the two pixel locations i and j. For efficiency reasons, the edge set is restricted to a 4-neighborhood, but the extension to a higher degree of connectivity is straight forward. To each pixel $i \in V$ we associate a discrete disparity label $Z_i \in \mathcal{Z}$.

Our goal is now to find an optimal discrete labeling $\hat{Z} \in \mathcal{Z}^V$ that minimizes the following conditional random field (CRF) energy:

$$\min_{\hat{Z}} \sum_{i \in V} F_i(z_i) + \sum_{ij \in E} F_{ij}(Z_i, Z_j), \tag{6.12}$$

where $F_i(Z_i)$ are the unary terms that are given by the CT matching costs for each depth hypothesis Z_i and F_{ij} are the binary terms that apply a smoothness constraint to the optimal labeling. We use the generalized TV function from [176]. In other words, by minimizing the CRF energy, we try to find an optimal labeling $\hat{Z}$ that provides a trade off between minimizing the matching costs and minimizing the smoothness constraint.

The CRF minimization is a combinatorial NP-hard problem. However, as shown in [176] we can compute very good approximate solutions by means of the dual minorize maximize (DMM) algorithm. The idea of the DMM algorithm is to decompose the CRF energy into distinct chain problems for which the CRF energy can be minimized efficiently using dynamic programming. Then we consider the Lagrangian function obtained from introducing a vector of Lagrange multipliers that force the solutions of the distinct chain problems to agree in the optimum. The dual problem associated with the Lagrangian function is continuous, piecewise linear and concave and provides a lower bound to the original problem. The idea of the DMM algorithm is now to iteratively construct a sequence of minorants to the dual problem which can be efficiently maximized. Once the dual problem is solved, the primal solution (and hence the depth map) is computed from the dual solution.

6.5 Initial Photometric Stereo Normals Using Previously Assessed Depth Model

As soon as a previous discrete or continuous disparity labeling $\hat{Z} \in \mathbb{R}^{\mathcal{M} \times \mathcal{N}}$ is available (as described in Sec. 6.4), it can be used to extract the observed intensities $E_{i,j}$ in each pixel location along the corresponding EPI-lines by the provided disparity model. Subsequently these vectors are used to generate the surface normal field $\hat{N}_{i,j} \in \mathbb{R}^3$ by applying Eq. (4.9) in all pixel locations. Details are presented in Sec. 4.2.3.

Due to the lack of the photometric stereo evidence orthogonally to the transport direction, values in G_y (i.e. gradients in y-dimension) are always estimated to be zero. Hence, the first-order TV smoothness prior is implicitly applied in y-dimension.

6.6 Optimized Fusion of Light Field and Photometric Stereo for Missing Data

In Sec. 5.2 we presented the setup and algorithmic approaches to reconstruct the depth from light field and the surface normals from photometric stereo with a missing y-component. Next we describe details of hybrid algorithms suitable for our multi-line scan setup, including a gradient-based method and our generalized method of Nehab. In these algorithms we introduce a flatness assumption in the y-direction. Additionally, we introduce a novel TGV approach for the multi-line scan case.

6.6.1 Gradient-Based Method

Using a gradient-based method allows us to use the measured data directly without necessity of any additional assumptions. This is an obvious benefit for processing data from our multi-line scanner. In this method, the deviation of the depth reconstruction Z to an initial light field depth estimation $\hat{Z}$ is penalized and the gradient $\nabla_x Z$ is enforced to be close to a photometric stereo gradient $\hat{G}_x$. Moreover, flatness is enforced in y-direction (see Fig. 6.4b). We formulate the gradient-based optimization approach as a least squares problem as follows:

$$\min_Z \frac{1}{2}||\hat{Z} - Z||^2 + \frac{\lambda_x}{2}||\hat{G}_x - \nabla_x Z||^2 + \frac{\lambda_y}{2}||\nabla_y Z||^2, \tag{6.13}$$
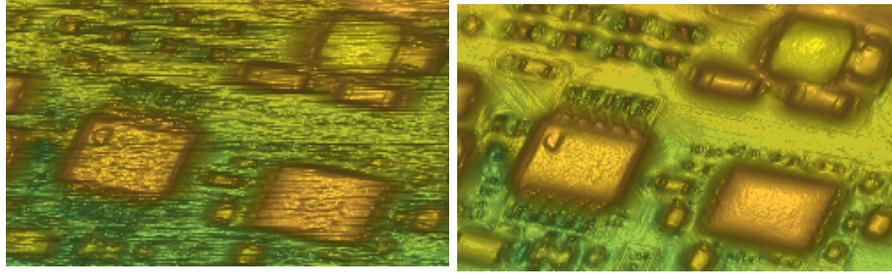
(a) G_x without y-flatness prior   (b) G_x with y-flatness prior

Figure 6.4: Depth reconstruction example using the gradient-based method. (a) Without and (b) with the y-flatness prior as described in Sec. 6.6.1

where λ_x and λ_y are weighting terms of the surface orientation in x- and y-direction and the gradient operator $\nabla : \mathbb{R}^{\mathcal{M} \times \mathcal{N}} \to \mathbb{R}^{\mathcal{M} \times \mathcal{N} \times 2}$ is given by:

$$(\nabla_x Z)_{i,j} = \begin{cases} Z_{i+1,j} - Z_{i,j} & \text{if } 1 \leq i < \mathcal{M}, \\ 0, & \text{otherwise}, \end{cases}$$

$$(\nabla_y Z)_{i,j} = \begin{cases} Z_{i,j+1} - Z_{i,j} & \text{if } 1 \leq j < \mathcal{N}, \\ 0, & \text{otherwise}. \end{cases} \tag{6.14}$$

We find the optimum of the least squares formulation by a standard *conjugate gradient* approach.

6.6.2 Generalized Nehab

As an extension of the gradient-based method, here we describe a generalized version of the method of Nehab [148], as shown in [7] for the area-scan case. Our least squares formulation of the generalized method of Nehab adapted for multi-line scanning data follows:

$$\min_Z \frac{1}{2} ||\hat{Z} - Z||^2 + \frac{\lambda_x}{2} ||(\hat{N}_z)^r \odot \hat{G}_x - (\hat{N}_z)^r \odot \nabla_x Z||^2 + \frac{\lambda_y}{2} ||(\hat{N}_z)^r \odot \nabla_y Z||^2, \tag{6.15}$$

where $\odot$ denotes the element-wise multiplication operator. A weighting of $r = 1$ corresponds to the original method of Nehab with flatness enforced in the y-direction and $r = 0$ corresponds to the gradient-based method (Eq. (6.13)). By weighting the

	GT	$\hat{Z}$	N only	Gradient-based	Method of Nehab	Generalized Nehab r=1.6 (ours)	TGV using gradients (ours)	TGV using GN r=1.6 (ours)
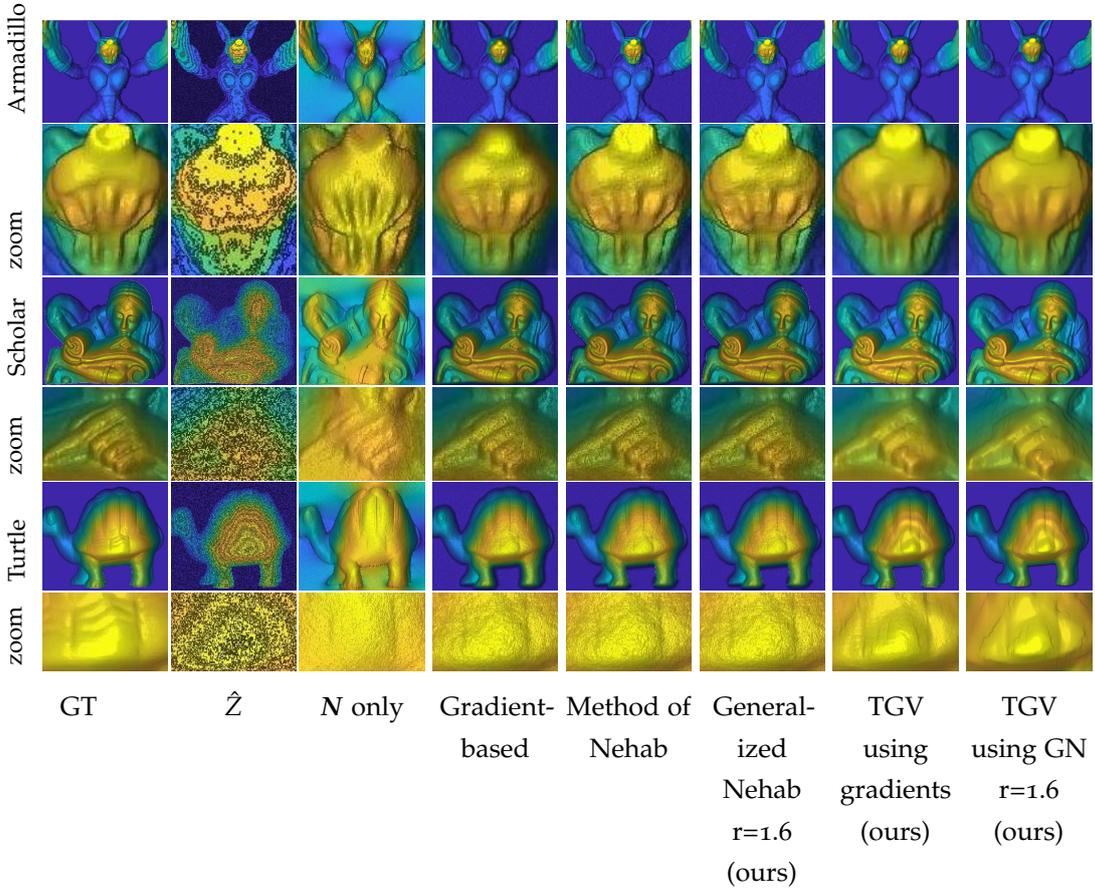
Figure 6.5: Examples of results obtained for different synthetic datasets.

surface orientation components by $(\hat{N}_z)^r$, with a parameter $r \in \mathbb{R}^+$, we can reach an improved reconstruction accuracy in the region of steep edges. The reason for this lies in an over-penalization of steep regions in the original formulation where errors in normals usually have more severe impact. This is reflected by the z-component of the normal, which holds lower values in steep regions and higher values in flat areas. As discussed in Sec. 6.2.2, we estimate an upper bound of $\hat{N}_z$ due to the incomplete photometric stereo information in our multi-line scanning data. Note that the correspondence from gradients to normals was shown in Eq. (6.1). In our experiments, we assume a weighting of $r = 1.6$, which proved to lead to better reconstructions in regions with steep edges. The minimizers to the least squares

formulations in Eq. (6.15) are found with a standard *conjugate gradient* method.

6.6.3 Total Generalized Variation

In this section we introduce a novel method to combine light field depth with photometric stereo surface orientation information for multi-line scanning data, which is an adaptation of the method [7] originally introduced for the area-scan case. The approach is based on the Total Generalized Variation (TGV) method [29], which is an extension to the popular Total Variation (TV) approach and that evades unwelcome staircasing artifacts. Our approach makes a direct use of gradients $\hat{G}$, where the auxiliary gradient G and the depth map Z are reconstructed simultaneously is formulated as follows:

$$\min_{Z,G} \alpha_1 ||\nabla Z - G||_{2,1} + \alpha_0 ||\boldsymbol{\nabla} G||_{2,1} + \frac{1}{2} ||\hat{Z} - Z||^2 + \frac{\beta}{2} ||\hat{G}_x - G_x||^2, \tag{6.16}$$

where finite differences are computed with the gradient operator $\boldsymbol{\nabla} : \mathbb{R}^{M \times N \times 2} \to \mathbb{R}^{M \times N \times 4}$ and $\boldsymbol{\nabla} G$ can be decomposed into $(\nabla G_x, \nabla G_y)$, with the ∇ operator as defined in Eq. (6.14). Our penalty function consists of a TGV term (first and second term in Eq. (6.16)), a light field constraint (third term in Eq. (6.16)) and a gradient constraint in x-direction (fourth term in Eq. (6.16)). The TGV regularization models a piecewise linear depth map Z making use a 2nd-order prior while indirectly (with the auxiliary G and parametrized by $\alpha_{0,1} > 0$) enforcing the gradient ∇Z to be in the proximity of the given gradients $\hat{G}_x$. The light field constraint enforces the final solution to be in the proximity of an initial noisy depth estimate $\hat{Z}$. The photometric stereo constraint is weighted with the parameter $\beta \geq 0$ and enforces the gradient in x-direction G_x of the solution to be close to an initial estimate $\hat{G}_x$. Additionally, we present an extension of the TGV method using the generalized Nehab's (GN) formulation introduced in Sec. 6.6.2. It is described by the following penalty:

$$\min_{Z,G} \quad \alpha_1 ||\nabla Z - G||_{2,1} + \alpha_0 ||\boldsymbol{\nabla} G||_{2,1} + \tag{6.17}$$
$$\frac{1}{2} ||Z - \hat{Z}||^2 + \frac{\beta}{2} ||(\hat{N}_z)^r \odot \hat{G}_x - (\hat{N}_z)^r \odot G_x||^2,$$

where $\hat{N}_z$ is the upper bound normal estimation in z-direction with an element-wise exponent operator with a parametrization $r = 1.6$. We optimize our TGV formulations of Eq. (6.16) and Eq. (6.18) using a primal-dual (PD) algorithm, as introduced in [29].

(a) Depth: MSE to $\hat{Z}_{GT}$ (range $[0,\infty]$)

Method / Dataset	$\hat{Z}$	N only	Gradient-based	Method of Nehab	Generalized Nehab r=1.6 (ours)	TGV using grad. (ours)	TGV using GN r=1.6 (ours)
Dragon	4.2259	134.5353	2.4709	0.1312	**0.1288**	**0.2443**	**0.2442**
Buddha	4.8448	146.3337	2.2938	0.1302	**0.1277**	**0.2603**	**0.2596**
Armadillo	4.5959	103.2253	2.2159	0.1224	**0.1208**	**0.2449**	**0.2443**
Cat	2.3707	154.5484	1.8337	1.0173	**0.9542**	**0.1916**	**0.1928**
Hippo	3.6345	364.4030	3.8063	1.9914	**1.8733**	**0.2134**	**0.2145**
Pig	2.7122	238.0685	1.7046	1.1181	**1.0884**	**0.3200**	**0.3201**
Scholar	3.0767	469.6256	3.6177	1.7873	**1.7404**	**0.2667**	**0.2673**
Turtle	2.3229	251.2245	3.1177	2.0097	**1.9619**	**0.1713**	**0.1714**
Boxes	3.1436	166.0757	22.1343	0.7611	**0.2461**	**1.5070**	**1.4355**
Cotton	3.2348	474.5878	16.5724	0.1988	**0.1729**	**0.2483**	**0.2472**
Dino	3.0853	139.2516	4.6089	0.1023	**0.0842**	**0.3205**	**0.3199**
Sideboard	3.0756	174.4863	16.1101	0.4228	**0.2189**	**0.5383**	**0.5613**
Antinous	3.1373	530.6002	22.5113	0.2707	**0.2106**	**0.3067**	**0.3057**
Town	3.0952	283.2612	8.3575	0.1927	**0.1482**	**0.3664**	**0.3665**
Medieval	3.0803	151.8558	4.6061	0.2170	**0.1896**	**0.2745**	**0.2735**
Greek	3.0759	386.7557	24.8074	0.2154	**0.1514**	**0.2805**	**0.2794**
Average	3.2945	260.5524	8.7980	0.6680	**0.5886**	**0.3597**	**0.3565**

Absolute Depth [MSE]

(b) Normals: geodesic distance to $\hat{N}_{GT}$ (range $[0,\pi]$)

	$\hat{Z}$	N only	Gradient-based	Method of Nehab	Generalized Nehab r=1.6 (ours)	TGV using grad. (ours)	TGV using GN r=1.6 (ours)
Dragon	0.1873	0.1901	0.3109	0.2657	**0.2632**	**0.0630**	**0.0634**
Buddha	0.1260	0.1296	0.2187	0.1826	**0.1807**	**0.0513**	**0.0513**
Armadillo	0.1591	0.1794	0.2669	0.2270	**0.2250**	**0.0607**	**0.0608**
Cat	0.4495	0.1772	0.2239	0.2137	**0.2109**	**0.0485**	**0.0499**
Hippo	0.7112	0.2082	0.2576	0.2481	**0.2459**	**0.0493**	**0.0506**
Pig	0.4948	0.1822	0.2191	0.2143	**0.2122**	**0.0394**	**0.0401**
Scholar	0.6066	0.2090	0.2691	0.2608	**0.2590**	**0.0621**	**0.0631**
Turtle	0.4438	0.2068	0.2502	0.2421	**0.2397**	**0.0451**	**0.0459**
Boxes	0.6449	0.1685	0.2102	0.0780	**0.0672**	**0.1056**	**0.1052**
Cotton	0.5792	0.0831	0.1534	0.0626	**0.0558**	**0.0457**	**0.0454**
Dino	0.6499	0.1462	0.1336	0.0710	**0.0660**	**0.0589**	**0.0590**
Sideboard	0.6742	0.1992	0.2147	0.0825	**0.0756**	**0.0850**	**0.0928**
Antinous	0.5969	0.1482	0.1677	0.0681	**0.0601**	**0.0420**	**0.0418**
Town	0.5875	0.1216	0.1664	0.0708	**0.0664**	**0.0879**	**0.0852**
Medieval	0.5476	0.1118	0.1214	0.0656	**0.0591**	**0.0478**	**0.0474**
Greek	0.6061	0.1514	0.2087	0.0725	**0.0639**	**0.0609**	**0.0607**
Average	0.5040	0.1633	0.2120	0.1516	**0.1469**	**0.0596**	**0.0602**

Surface normals [Geodesic distance]

Table 6.1: Quantitative comparison of the analyzed methods. Where our methods are marked bold. The accuracy of the reconstructed depth is evaluated w.r.t. the ground truth in absolute terms by means of the mean squared error (MSE) and locally on the level of surface normals using the geodesic distance. The evaluations were performed on objects from the Stanford database [186] (1-3), which were rendered using POV-Ray [163] as well as on depth maps retrieved from the photometric stereo dataset presented in [228] (4-8) and the objects of the light field dataset presented in [95] (9-16). The exceptional dataset Boxes (Fig. 6.6) is marked red.

6.7 Evaluation

To evaluate the proposed algorithms, we need an initial depth map $\hat{Z}$ and an estimate of the gradient in x-direction $\hat{G}_x$. Due to missing y-components in the multi-line scan data, the flatness assumption is considered in y-direction in the gradient ($\hat{G}_y = 0$) or normal ($\hat{N}_y = 0$) domains in order to estimate an upper bound for $\hat{N}_z$. We conduct experiments for the depth reconstruction with multi-line-scan data. The introduced hybrid light field and photometric stereo variational methods are first evaluated on synthetic datasets for a quantitative evaluation and later demonstrated on real world multi-line scan data for a qualitative evaluation.

6.7.1 Synthetic Data

We compared the results of several methods which we tailored for the use of multi-line scan data. Namely, (i) a gradient-based method, (ii) the original method of Nehab, (iii) our generalized method of Nehab and our TGV approaches with (iv) the simple gradient-based and (v) the generalized Nehab surface normal penalty formulation (r=1.6). For illustrative purposes show one additional method: the gradient-based depth reconstruction using the photometric stereo only, which gives a comparable accuracy in the normal domain but has a weak performance in the absolute depth. The obtained quantitative results are shown in Tab. 6.1, with examples of the corresponding depth maps demonstrated in Fig. 6.5.

The initial depth map $\hat{Z}$ is constructed by adding normally distributed noise of a maximum strength of 7% of the depth range to the ground truth depth map and thresholding the result. This corresponds well to results which are usually achieved by data acquired by our multi-line scan system after a normalization with a CRF (Conditional Random Field).

$$\hat{Z}_{i,j} = \frac{1}{k}\left[k\left(\hat{Z}_{GT_{i,j}} + noise_{i,j}\right)\right], \quad \forall i, j \in \mathcal{I} \tag{6.18}$$

Where $[\cdot]$ rounds to the nearest integer number and the constant k controls the number of discretization steps. The initial normals $\hat{N}$ are constructed by adding normally

distributed noise of a maximum extent of 23% of the normal range to the ground truth.

$$\hat{\mathbf{N}}_{i,j} = \frac{\hat{\mathbf{N}}_{GT_{i,j}} + noise_{i,j}}{|\hat{\mathbf{N}}_{GT_{i,j}} + noise_{i,j}|}, \quad \forall i,j \in \mathcal{I}. \tag{6.19}$$

The gradient-based reconstruction uses the gradient data generated by the multi-line scanner directly, but it shows a weak performance on steep edges. On flat regions, as often observed in real-world data, this method can still yield a quite competitive performance. The method of Nehab behaves better than the gradient-based method especially on steep edges. This is further improved by our generalized method of Nehab. Both our TGV approaches for multi-line scan data globally outperformed all other methods in the geodesic normal distance domain as well as in the depth MSE (mean squared error). Introducing the generalized Nehab penalty to TGV does not significantly change the result, which is already well-balanced through the TGV term even using simple gradients.

$$\hat{Z}_{GT} \qquad\qquad \hat{Z} \qquad\qquad \text{Result (TGV)}$$
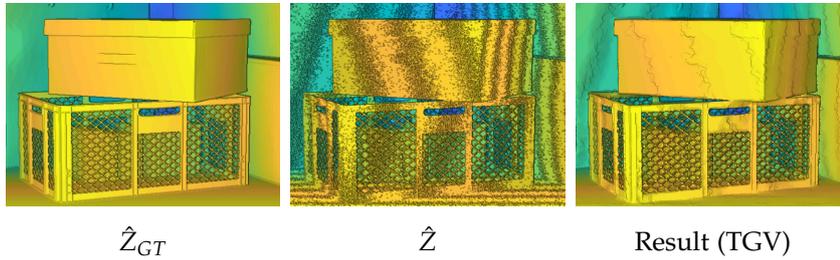
Figure 6.6: Reconstruction of boxes with our TGV method.

There was only one exceptional dataset, "Boxes" from [95], in which our TGV methods seemed to struggle (see Fig. 6.6). This scene contains locally frequent strong abrupt depth changes combined with large flat tilted planes and a noisy and quantized depth initialization. In this special case the best result was obtained with our generalized method of Nehab. In future research we intend to exploit regional weighting parameters including local smoothing based on color schemes in the image, similar to bilateral filtering.
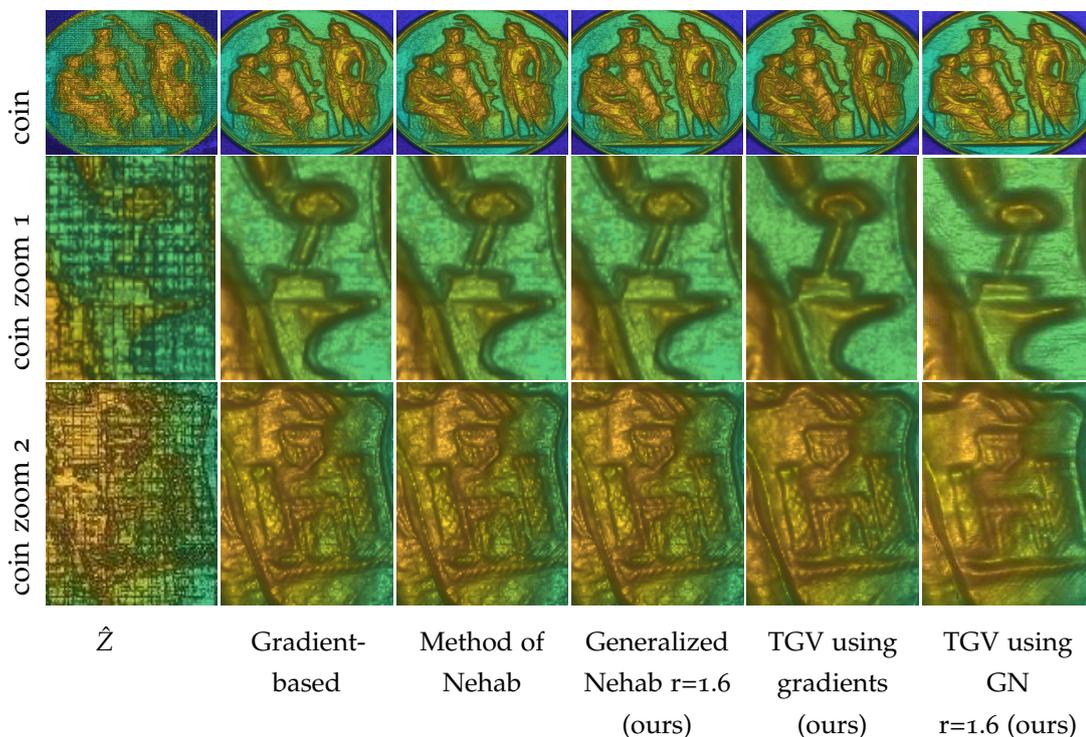
Figure 6.7: Depth reconstruction results from real world data acquired with our multi line scan setup. The setup was described in 3.2.5.

6.7.2 Real World Evaluation for Multi-Line Scan Data

With the multi-line scan setup, the initial depth maps can be provided by light field correspondence analysis. Regularizing the initial depth map allows a more reliable initial photometric stereo estimation, as described in Sec. 6.2.2. In [9] a discrete-continuous optimization algorithm was suggested, based on conditional random fields (CRFs) for a fast and accurate initial depth solution. Such an example is shown in Fig. 6.7, which shows the initial depth $\hat{Z}$ and results obtained by the methods analyzed in this chapter. The data was acquired with the setup described in Sec. 6.2, the discretization artifacts in $\hat{Z}$ stem from the noisy light field depth estimation and the CRF regularization. Normals are estimated from the light field stack as discussed in Sec. 6.2.2. One can see that on the provided real-world example the TGV approach outperforms the original method of Nehab as well as the generalized method of

Nehab.

6.8 Conclusion

In this chapter we explored variational methods for the combination of light field depth and photometric stereo surface orientation tailored for data generated by the multi-line scanner as previously presented in [8].

We showed how to tackle the problem of missing surface normal component when using the gradient-based approach, the original method of Nehab [148] or the generalized method of Nehab [7]. Both variants of the Nehab's approach exploit an upper bound estimation on $\hat{N}_z$ but still show an improved reconstruction accuracy in regions of steep edges over the gradient-based method. Furthermore, we introduced a novel hybrid TGV method, demonstrated with two different weighting functions for the photometric stereo constraint: (i) the gradient-based and (ii) the generalized Nehab. The methods were evaluated on 16 synthetic examples from three datasets, which contain a variety of different object geometries.

We showed that both TGV approaches outperform all other algorithms in the accuracy of the reconstructed depth and surface orientation. Practical applicability of the presented methods was illustrated on a real-world example obtained with our multi-line scan framework.

7 Reflectance Distribution Capture

Contents

In this chapter we will address the issue of measuring, representing and completing sparse reflectance distribution functions. First we will introduce several reflectance distribution functions and motivate their use in computational imaging tasks in Sec. 7.1. Then we introduce the visual structure we use to represent bidirectional reflectance distribution functions (BRDFs) in Sec. 7.2. In Sec. 7.3 we show several ways to capture these functions, where different acquisition methods differ in speed, accuracy and the level of density (/sparsity) of the function they are able to capture. In Sec. 7.4 we present our work published in [4], where we focus on learning dense reflectance functions from sparse data using CNNs.

7.1 Motivation

Many computer vision applications require the classification, segmentation or reconstruction of surface materials. The appearance of materials can depend on properties such as the position of the object, the wavelength of the illumination, the illumination angle (irradiance incident) and the viewing angle (reflected radiance). For a dense

Figure 7.1: Illustration of the BSSRDF.

acquisition of the reflectance function, many combinations of these parameters have to be measured over the region of the surface.

A general reflectance function has 16 variables and is currently unfeasible to measure [80], it contains information about the irradiance incident and reflected spherical radiance angles, on time and the light spectrum. It models for each material the unique appearance under any lighting direction, intensity and spectrum. In practical models assumptions are used which limit the complexity.

The bidirectional scattering-surface reflectance distribution function (BSSRDF), as illustrated in Fig. 7.1, describes the relation between the irradiance incident and the reflected radiance between arbitrary two rays that hit a surface, including the properties of subsurface scattering by the variables $(x_i, \theta_i, \phi_i; x_r, \theta_v, \phi_v)$, which denote the angles of the incident flux at the point x_i and reflected radiant flux at the point x_v.

As a specific case of the BSSRDF, restricted to opaque materials, the bidirectional reflectance distribution function (BRDF) was formally specified by [153]. The BRDF, as illustrated in Fig. 7.2, is described by four independent variables $(\theta_i, \phi_i; \theta_v, \phi_v)$, where the light is assumed to enter and leave at the same point. Therefore it cannot model subsurface scattering as it appears in translucent materials (e.g. marble, snow, skin, milk), as described in [101]. The BRDF maps the irradiance incident from a specific direction to its influence towards the reflected radiance of another direction. The BRDF shows the properties of view and illumination direction reciprocity and energy conversation, as described in [58].
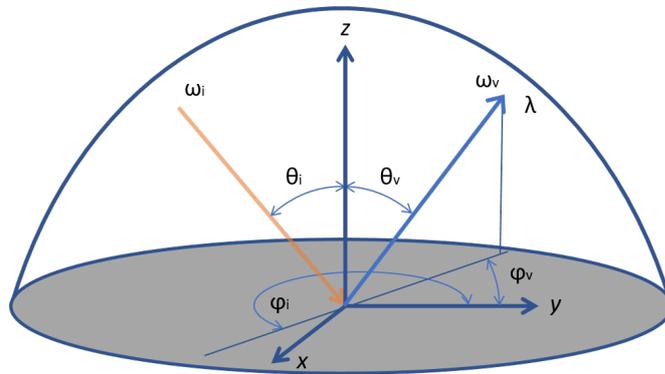
Figure 7.2: Illustration of the BRDF.
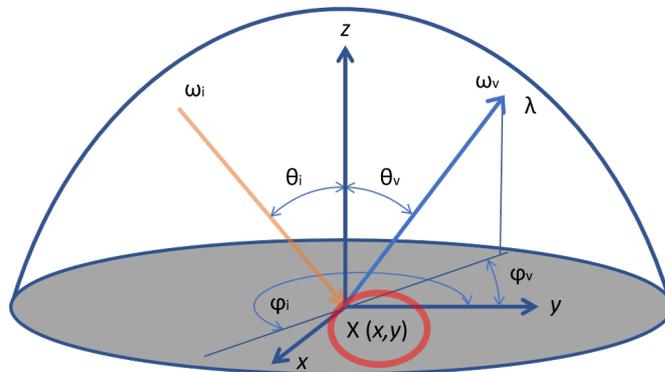
Figure 7.3: Illustration of the BTF. The difference to the BRDF in Fig. 7.2 is the position dependency marked in red.

The (multi-spectral) bidirectional texture distribution function (BTF), as illustrated in Fig. 7.3, was later introduced by [44], which describes a texture image parametrized by the viewing and illumination angle and therefore a spatially varying BRDF. The BTF is described by the variables $(p; \theta_i, \phi_i; \theta_v, \phi_v)$, where $p = [p_1, p_2, p_3]$ defines the position on the material sample (vertical/horizontal) as well as the spectral index. BTFs are very data/storage intensive but can model structures including self-occlusions, subsurface scattering, rough surfaces or self-shadowing [58].

Understanding reflectance distribution functions is essential for the detection and classification of properties or defects of objects which can have different materials and surface structures. In this chapter, we first present the structure we use to visualize
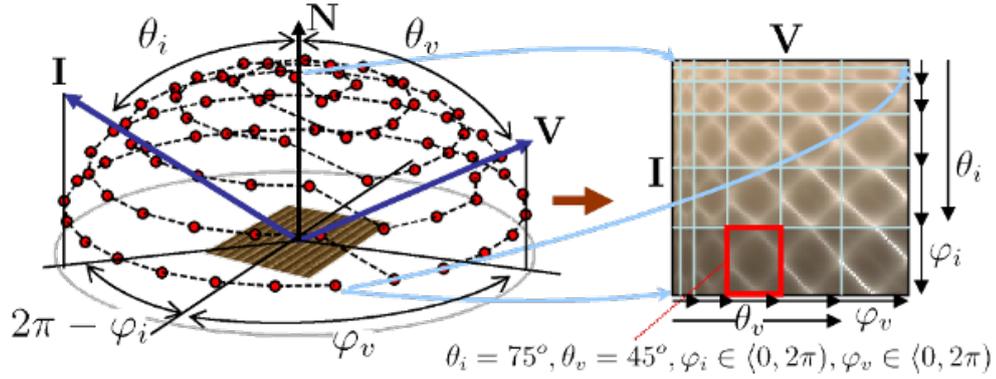
Figure 7.4: BRDF acquisition setup and data arrangement. This illustration is courtesy of UTIA [80].

BRDFs. Then we show properties of measurement systems for reflectance distribution functions. These methods can capture reflectance properties to different degrees of details, where capturing a high degree of reflectance details is highly time consuming and can take several years. We would like to capture a full BRDF which is not feasible due to the time consumption. Common methods measure parts of BRDFs. We introduce a novel method to computationally reconstruct full BRDFs from partial data using convolutional neural networks (CNNs).

7.2 BRDF Capture

In this section the structure of the BRDF, the variables and the representation are presented as used in the following chapter.

The data arrangement is illustrated in Fig. 7.4. The BRDF is described as a 4D function of $(\theta_i, \varphi_i; \theta_v, \varphi_v)$, where θ denotes the elevation- and φ the azimuth angle. These angles are defined for both the illumination direction I and the camera direction V in respect to the object normal N. The object normal points towards the top center of the acquisition device.
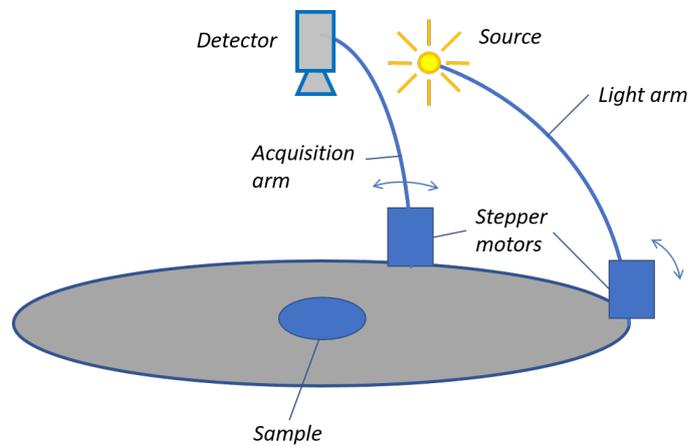
Figure 7.5: Illustration of a mechanical gonioreflectometer construction.

7.3 Measurement

Common methods to capture the reflection behavior of different materials differ in the flexibility of detail they measure, in the type of reflectance distribution function they can model, the dynamic range of the used sensors and in the time an acquisition consumes.

7.3.1 Gonioreflectometers

Gonioreflectometers use an independently moving light source and camera (single pixel detector). With this setup all combinations of light source and detector are possible down to the precision of the moving arms, with a restriction of combinations where light source and detector would occlude each other. Such a system is illustrated in Fig. 7.5. The measurement process using a gonioreflectometer is time consuming. Approaches to speed up the process included placing four cameras on the acquisition arm [191] or computational optimizations, where an algorithm plans the necessary acquisitions in order to reduce the overall uncertainty of the modeled parameters iteratively [117].
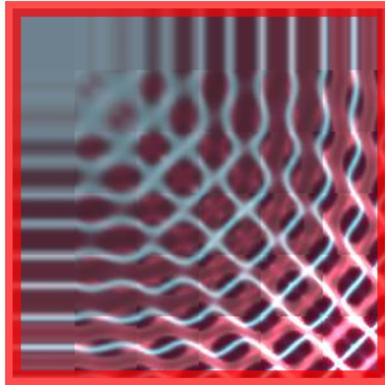
Figure 7.6: Illustration of the BRDF captured by a gonioreflectometer. The portion of the BRDF which can be measured is marked in red.

Each acquisition approach allows a different level of completion of the BRDF that can be captured. Gonioreflectometers can capture full BRDFs, as illustrated in Fig. 7.6, if enough time is invested in the procedure.

Gonioreflectometer setups can be extended to acquire BTF data by allowing the material sample to move. An approach where the object is moved using a robotic arm was presented by [44]. An extension was introduced by [171], where the camera is additionally placed on a half-circle rail.

7.3.2 Mirror-Based Setups

Mechanical parts can be reduced using mirror-based setups. Here, several views are simultaneously acquired in a single image by using hemispherical, parabolic or ellipsoidal mirrors [80]. An example with an ellipsoidal mirror is illustrated in Fig.7.7. A material sample is placed in front of the mirror and the object is illuminated using a projector emitting a multiplexed pattern. The camera captures different viewing angles on the same optical path as the projector through a half mirror. Such a setup allows a fast acquisition with a lower accuracy, higher noise and a lower range of light source and detector angles. The portion of the BRDF which can potentially be captured is equal to the Gonioreflectometer (Sec. 7.3.1) with a noise and higher errors. A BRDF can be acquired only within minutes. Due to the fast acquisition process it is feasible to
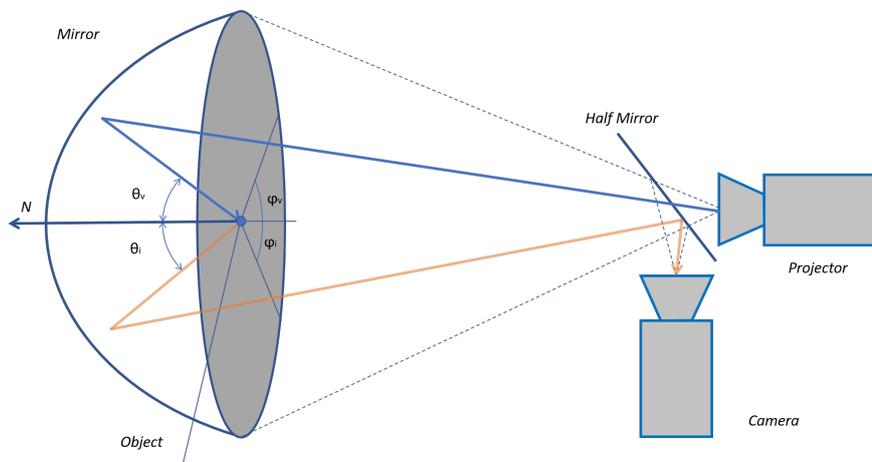
Figure 7.7: Ellipsoidal mirror-based BRDF acquisition setup.

acquire BTFs, additionally to BRDFs. A mirror based system using a kaleidoscope was presented by [82], where each triangular sub-image of the kaleidoscope corresponds to a different camera viewpoint.

7.3.3 Single Acquisition

The simplest approach is a single acquisition, where an object is placed under an area camera with a single illumination source. Aligning the camera with the surface normal allows homogeneity across the image. Examples for collections of single texture acquisitions are the vistex database [116] with over 100 homogeneous textures in frontal and oblique perspectives or the Brodatz texture library [167] with 112 acquisitions.

A single acquisition allows to capture a portion of the BRDF of one viewing angle and one illumination angle an example is illustrated in Fig. 7.8. Using a diffuse illumination would allow to acquire the sum of the first column in the illustration. Capturing several points of the same material with an area scan camera where the surface shows different and known surface normals adds several of these acquisitions together and allows a more detailed but still sparse BRDF representation.
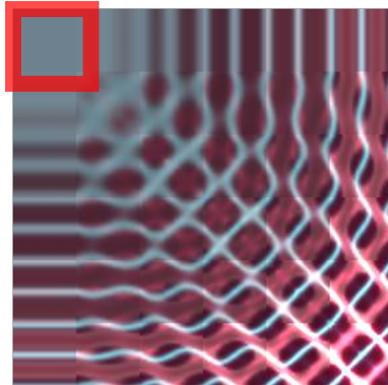
Figure 7.8: Illustration of the BRDF captured by a single acquisition. An example of a portion of the BRDF, where the reflection is measured and illuminated orthogonally to the surface is marked in red.
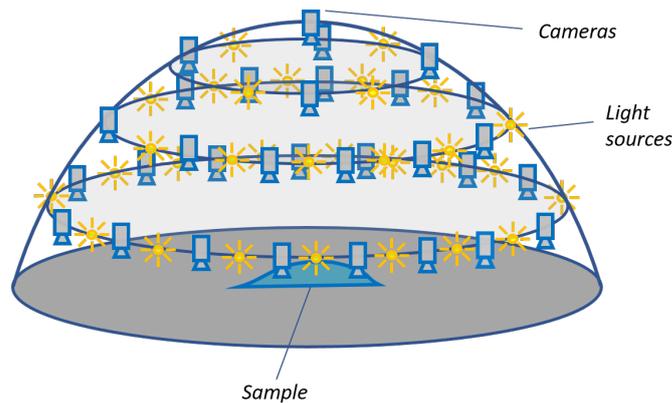
Figure 7.9: Illustration of a dome to capture BRDFs. On three levels cameras and light sources are mounted to capture a specific object point.

7.3.4 Light Dome

This setup consists of a light dome with fixed led illuminations and cameras or a single fixed camera placed at the top of the dome structure. An illustration is shown in Fig. 7.9. These devices are often build in portable sizes and used in industrial applications for on-side measurements of BRDFs. The benefits are very fast acquisitions without the problem of occlusions of the light sources and the capturing device, with the price of sparse fixed light sources which only allow to measure a partial BRDF.
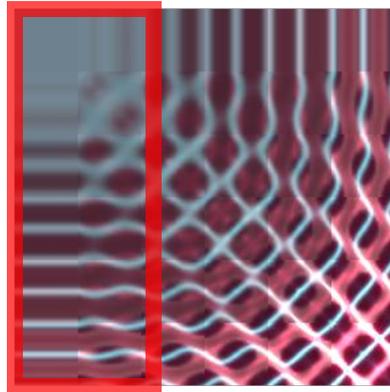
Figure 7.10: Illustration of the BRDF captured by a light dome and light field camera. The portion of the BRDF which can be measured is marked in red.

A light dome to capture (BRDFs) was presented in [173], where 151 compact cameras are mounted on the dome structure. Pairing a light dome with a light field camera (described in Chapter 3) allows to capture a wide range of illumination angles with a limited range of viewing angles as illustrated in Fig. 7.10.

In Sec. 7.4 we introduce a novel approach to deal with the limitation of sparse data by learning the behavior of BRDFs with a convolutional neural network (CNN) in order to reach a more complete description.

7.3.5 Multi-Line Scan Acquisition

The multi-line scan setup was previously described in Sec. 3.2.5. A partial BRDF is acquired by our multi-line scan system which consists of a multi-line scan camera, which acquires objects from different viewing angles, and two static illumination sources (shown in Fig. 3.4). An object is moving on a linear stage below the acquisition setup and captured after moving a pixel distance further on the sensor. At each position, the object is illuminated from a different direction. Due to limited viewing and illumination angles the BRDF acquisition is very spare. The benefit is that this method is very fast, as the object's partial BRDF can be acquired while it is moving with a speed of $> 130 km/h$ [20].
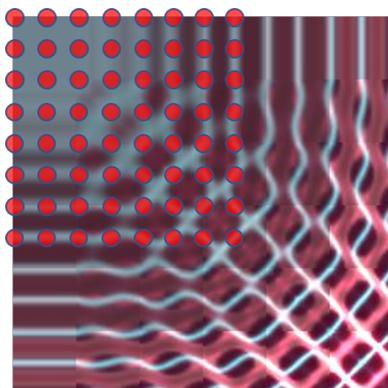
Figure 7.11: Illustration of the BRDF captured by a multi-line scanning system. The portion of the BRDF which can be measured is marked with red dots.

A multi-line scan system with one or two illumination sources allows to capture a part of the BRDF, where θ_i and θ_v vary in a certain range and only an angular rotation of $180°$ is enabled, as illustrated in Fig. 7.11.

The multi-line scan setup allows the acquisition of light field data, which was described in Chapter 3.2.5 and the acquisition of photometric stereo information, as presented in Chapter 4.2.3.

7.4 Learning BRDFs

It was shown, that the acquisition of BRDFs is either precise, complete and highly time consuming or a specific portion of the BRDF can be captured. Which part can be captured and the scanning speed depends on the system which is used. In this section we present our work published in [4] about learning dense (full) BRDFs from sparse acquisitions.

We motivate the research in Sec. 7.4.1. Then we discuss the connection of light fields with BRDFs in Sec. 7.4.2. The used BRDF dataset is described in Sec. 7.4.3. Our chosen regression CNN structure is introduced and discussed in Sec. 7.4.4. What follows clearly stated is the behaviour of the reconstruction from 1/6 to 5/6 of the dense known BRDF using CNNs in Sec. 7.4.5. There, we show quantitative and qualitative

evaluations of our dense BRDF reconstruction experiments. A summary of the paper and our key results as well as an outlook to our further work is given in Sec. 7.5.

7.4.1 Motivation

The acquisition of partial BRDF measurements using light field cameras and several illumination directions raises critical questions regarding the accuracy of inferences based on that data. Therefore, we attempt to verify the quality of the reconstruction of a full BRDF using partial input data. A dataset that provides a densely sampled BRDF was used, both in viewing and illumination directions. We show the reconstruction of dense BRDFs when the viewing angles are limited to top central regions, while the illumination angles are not reduced and are positioned in the shape of a half sphere around the material object, these properties are characteristic of data provided by plenoptic cameras paired with a photometric light dome. The partial reconstruction of the dense BRDF out of data is achieved by utilizing convolutional neural networks. We obtain a competitive full reconstruction when up to 2/3 of the BRDF is unknown.

Material appearances of structured materials such as metal, wood, or plastic can be uniquely characterized by the bidirectional reflectance distribution function (BRDF), as described by Nicodemus *et al.* [153]. The acquisition of BRDFs can be tedious. Several databases were constructed for research and academic use, such as the CUReT BRDF database [43] which provides sparse samples with 200 measurements. The MERL BRDF Database [139] was obtained with a light source moving in a circle at the same level as the camera, in respect to the surface normal. Since both illumination and camera position are restricted to a defined circle, it neither allows the simulation of viewing angles similar to plenotpic cameras nor the positioning of the illuminations in the shape of a half sphere. The BRDFs from BTFs Dataset from UTIA [78] provides densely sampled high precision BRDF measurements with independent camera and illumination positions, which allow both capturing anisotropic material behaviour as well as reducing the dimensionality for both, illumination and camera separately.

Using a very sparse BRDF for the task of image classification was discussed by Wang *et al.* [209], where a fixed traditional camera was placed at the top center of a light dome

(a) GT   (b) not transposed   (c) transposed
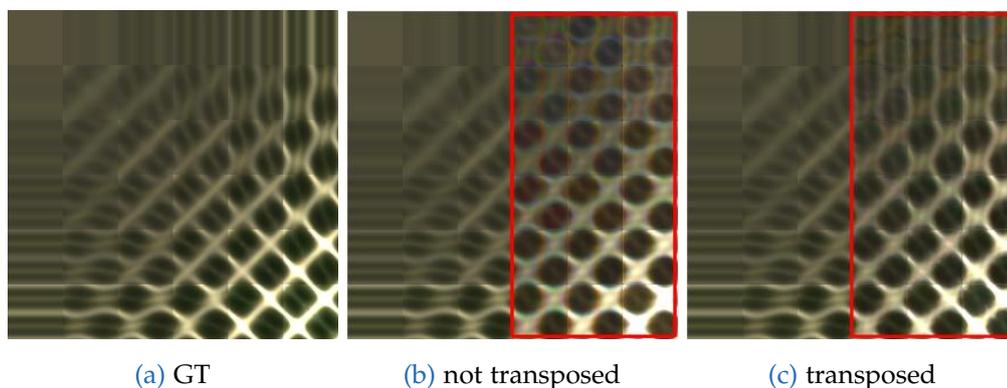
Figure 7.12: BRDF input data stack. (a) Ground truth BRDF, (b) reconstruction with input data dimensions $48 \times 48 \times u$ without the transposed channels with an SSIM=0.73 of the estimation (for the shown BRDF sample), (c) reconstruction with additional transposed input data added to the original input channels, which leads to input data dimensions $48 \times 48 \times 2u$ with an SSIM=0.88.

in order to acquire images. Each pixel observation is fit to a 2nd order hemispherical harmonic model. By Wang *et al.* [211] convolutional neural networks (CNNs) [119] were used to classify materials from plenoptic acquisitions without illumination variations. It was shown that material recognition can be improved through light field data, compared to single image acquisitions. A compressive sensing approach was implemented by Zupanic and Soler [242], in order to reconstruct BRDFs from a single image with known normals and illumination directions for isotropic and spatially constant materials. Reconstructing the BRDF from a limited set of samples was addressed previously by Nielsen *et al.* [154], where the MERL database was used. The dimensionality was reduced based on a principal component analysis in order to retrieve the most influential regions.

This paper demonstrates the ability to reconstruct dense BRDFs for restricted observation angles on the scene, as occurring in plenoptic cameras, while the illumination sources can be triggered from a wide range of positions around the object. The goal is to reconstruct dense BRDFs out of sparse data, with restricted viewing angles concentrated around the top down view.

(a) fabric	(b) glitter	(c) carpet	(d) corduroy	(e) stone

(f) wood	(g) car paint	(h) scotchlite	(i) leather	(j) paper

(k) plaster	(l) plastic	(m) plastic tape	(n) wallpaper
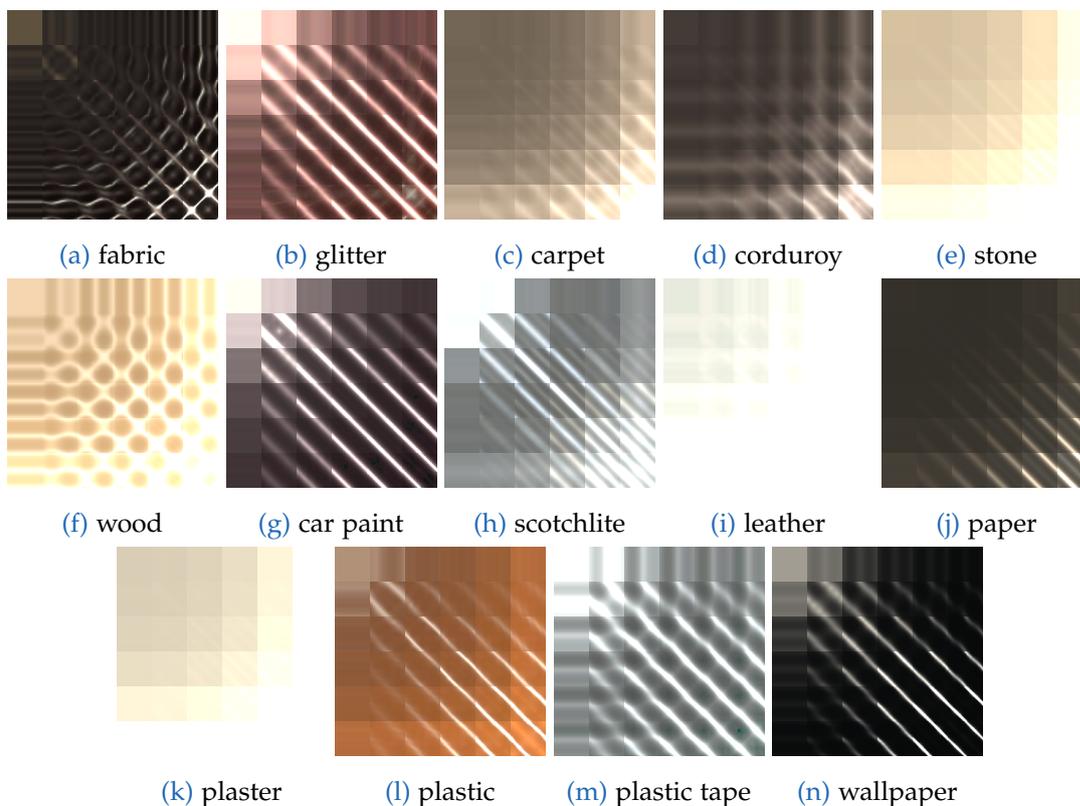
Figure 7.13: Examples of measured BRDFsof materials. Illustrated as contained in the UTIA dataset [80].

7.4.2 Light Field and BRDF

In order to measure BRDFs, usually gonioreflectometers are utilized. Varying viewing and illumination angles allow the measurement of various incoming and outgoing light rays. For the measurement of anisotropic BRDFs all four dimensions have to be sequentially sampled [59].

Plenoptic cameras capture light fields by acquiring irradiance values from different viewing directions on scenes, using a lenslet array in front of the image sensor, as described in [151]. These light fields can be represented by two directional and two spatial dimensions. Thereby they provide 4D information of the flow of light through space in a static scene. Using multiple illumination sources allows both the estimation

of local surface orientations (photometric stereo) as well as capturing additional reflectance properties.

Previously, a setup combining plenoptic acquisitions with photometric stereo, where the illumination sources were placed on a half sphere around the scene, was used in [6] in order to achieve an improved depth reconstructions of scene. Utilizing such a setup, comprising plenoptic cameras with several illumination sources, shows an ideal tool for measuring sparse BRDFs. Since the viewing directions are restricted by the cameras sensor size and position, one must address the question of the estimation of a dense BRDF from the data acquired from light field cameras.

For evaluating our reconstruction, a dense BRDF dataset is utilized which was captured with a gonioreflectometer. We simulate the reduction of viewing angles by reducing the data accordingly. This dataset is described in the following section.

7.4.3 BRDF Dataset

The BRDFs from UTIA's BTF Dataset [78] were used, which were measured with a high precision gonioreflectometer [79], with an angular precision of 0.03 degrees and a spatial resolution of 1000 DPI. Data is then sampled at fixed azimuth and elevation angles for both the illumination directions and the camera position [60]. The dataset consists of 150 samples, comprising the materials of wood, fabric, scotchlite, car paint, carpet, corduroy, glitter, leather, paper, plaster, plastic, stone, wallpaper and plastic tape, as demonstrated in Fig. 7.13.

The data acquisition Framework and data arrangement is shown in Fig. 7.4 and described in Sec. 7.2. The data contained in the dataset was interpolated in order to propagate information to missing parts of the BRDF subspace by an adapted swept surface technique (see [60]).

Limiting the cameras elevation angles towards the top center approximates capturing the scene with a plenoptic camera. Though it is obvious, that plenoptic cameras provide a more dense representation in a narrow angle range, we will be able to infer strong indications regarding the reconstruction abilities of BRDF data which is limited
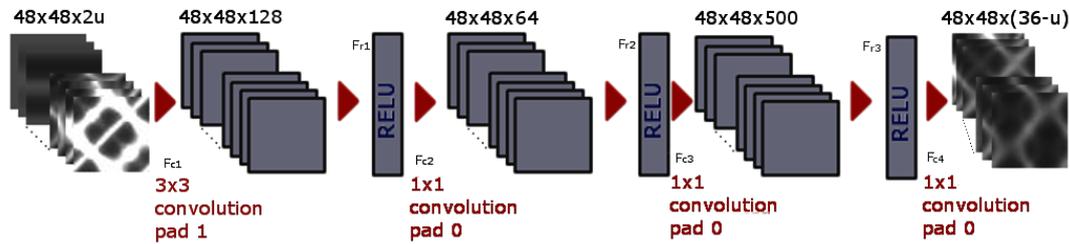
Figure 7.14: CNN network structure for BRDF reconstruction.

in its observation angles. Reducing the cameras elevation angles affect the material representation, as certain reflectance behaviour is unknown, as well as some of the representation of specular peaks and of anisotropic behaviour. Anisotropic patterns and specular peaks are visible in axial and diagonal slopes respectively, in a square $\theta_i \times \theta_v$ of the BRDF.

Each input feature stack for our CNN contains all azimuth variations (φ_i, φ_v), where $\varphi_i = (0...2\pi)$ and $\varphi_v = (0...2\pi)$, which each cover 48 pixels at defined elevation angles (θ_i, θ_v).

A description follows how the described dataset is used to approximate a dense BRDF from sparse samples.

7.4.4 BRDF Estimation

This section introduces our regression neural network and discusses the depth of the network structure. Using regression, one can estimate those parts of the BRDF which are unknown due to a restricted elevation viewing angle on the scene.

Network Structure

We consider the problem of estimating a full BRDF consisting of all measured views and illumination angles, which will further be referred to as a dense BRDF. This estimation is inferred from BRDF slices that are limited in their viewing directions.

More specifically, we reduce the dimension θ_v from $15°$ to $75°$, in step sizes of $15°$. In the representation in Fig. 7.4 this is equal to removing 1 to 5 columns with each a width of 48 pixels, starting from the right side of the image. These unknown parts are estimated from the residual known data. For this regression problem, we make use of a CNN. The network can be represented as a function $F(x)$ that transforms the input data into a stack of output feature maps:

$$F(x) = (h_{c4} \circ a_{c4} \circ ... \circ h_{c1} \circ a_{c1})(x), \tag{7.1}$$

where h denotes the nonlinear activation function and $a_l = w_l x + b_l$ an affine function for a layer l with a weight w and a bias b, as described in Sec. 2.3.2 and Eq. 2.43. The function relation follows:

$$F : \mathbb{R}^{48 \times 48 \times (2u)} \rightarrow \mathbb{R}^{48 \times 48 \times (36-u)}. \tag{7.2}$$

The input data x of the dimension $48 \times 48 \times (2u)$ leads to a number of $2u$ input maps, where the number of input channels $2u$ contains $u = \{6, 12, 18, 24, 30\}$ BRDF slices of defined elevation angles (θ_i, θ_v) as well as their transposed counterparts. The benefit of that input data structure is depicted in Fig. 7.12, which shows an improved reconstruction when using enriched input data. Transposing the input channels supports the convolutional neural network to follow the Helmholtz reciprocity, which improves the reconstruction qualitatively and quantitatively (see Fig. 7.15), especially of the top right regions. This is described in more detail in Sec. 7.4.4.

The input data is then mapped to $n = 36 - u$ output channels, using convolutional layers F_{c*}, followed by rectified linear units (ReLUs) F_{r*}. The convolutional layers F_{c2}, F_{c3} and F_{c4} are 1×1 convolutional layers. The networks weight parameters $w = (w_{c1}, w_{c2}, w_{c3}, w_{c4})$ are initialized with normally distributed random numbers and learned from the input data in order to solve the problem of reconstructing the unknown BRDF regions. The 1×1 convolutional layers followed by ReLUs benefit the reconstruction of the BRDF. We train the model to minimize a quadratic loss function
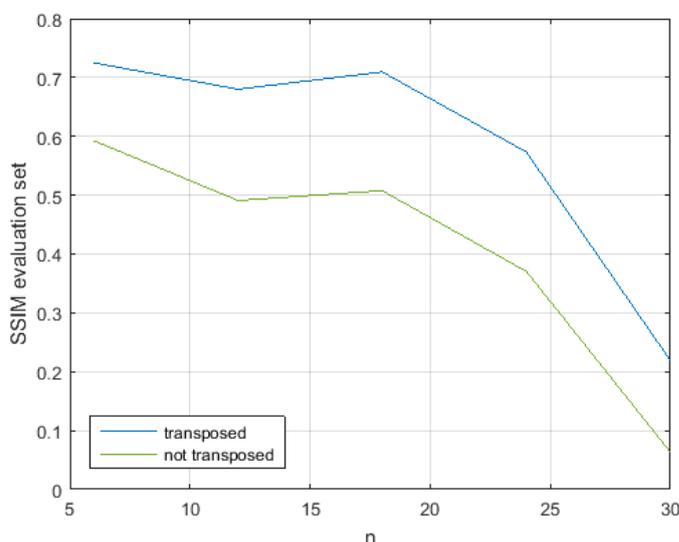
Figure 7.15: Comparison of the SSIM performance. Our proposed network, with trans-
posed input data added to the original, is compared to the original input
data (not transposed). All evaluations were performed at epoch $e = 1300$.

$\sum_i^N (L_i - F(x_i))^2$, where l is the ground truth BRDF value and $F(x_i)$ the estimation for
the input x_i.

A network architecture is used as shown in Fig. 7.14, the input consists of the known
BRDF parts and their transposed versions.

Deep Neural Networks

For the regression task, we tested deeper convolutional network architectures, where
we used the structure of the first 3×3 convolutional layer F_{c1} up to a number of 5 times.
Multiple insertions of F_{c1}, each followed by the ReLU F_{r1}, are used to gain a deeper
network structure. The network is depicted in Fig. 7.14. The depicted structure contains
one 3×3 convolutional layer F_{c1}, while more insertions of that layer demonstrate the
performance of deeper networks. While the first convolutional layer is sensitive to local
features as e.g. edge structures, deeper layers represent more complex combinations
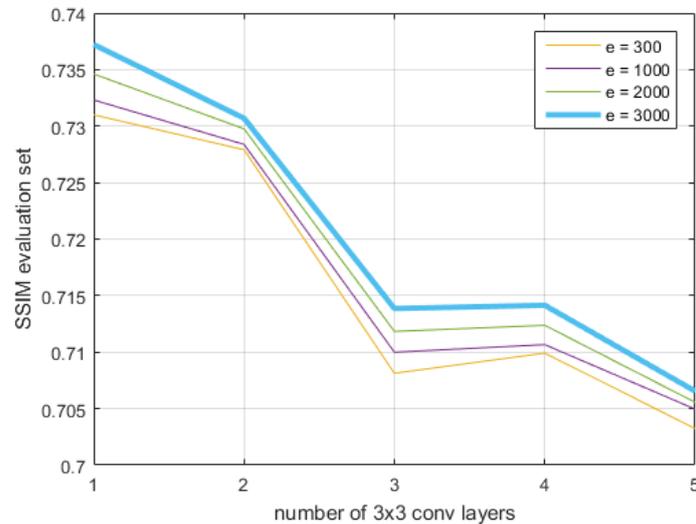
Figure 7.16: Evaluation of deep CNN structures for BRDF reconstruction. Evaluted up to epoch 3000 on the evaluation set, with $u = 30$, where the structure of the first 3×3 convolutional layer is used up to 5 times by additional insertions. The training set behaves in a similar way. The network was trained twice up to $e = 3000$, the second result showed the same tendency. Additionally we tested the mean absolute error (MAE) performance, which showed a comparable behaviour. Therefore the second result and the MAE are not displayed.

of lower level local features. For each network of increased depth, we observed the accuracy on the evaluation set. In Fig. 7.16, the performances of the networks are compared at several epochs e up to $e = 3000$. A performance drop is visible with each additional 3×3 convolutional layer, with a structural similarity (SSIM, described in detail in Sec. 7.4.5). Therefore, a deeper network of that type would not support us with our current reconstruction task. We chose a network with one 3×3 convolutional layer, as depicted in Fig. 7.14, as it shows superior performance compared to deeper network structures for the used dataset and architecture type.

Other kernel sizes were tested during our experiments. Where sizes up to 7×7 didn't improve the result in any significant way, a fully connected network showed to be

infeasible due to memory limitations. The latter was addressed by enriching the input by the transposed data.

Enriched Input Data

The input data is enriched to help the network with the reconstruction task. In order to provide a higher variaty in the input data, each BRDF slice is transposed and added to the original BRDF data. This is reasonable since the Helmholtz reciprocity holds for BRDFsof most realistic materials [50]. The principle describes that the BRDF value would be unmodified, when swapping the camera and illumination positions: $BRDF(\theta_i, \varphi_i; \theta_v, \varphi_v) = BRDF(\theta_v, \varphi_v; \theta_i, \varphi_i)$.

The results for both enriched and non-enriched input data are illustrated in Fig. 7.15, with output channels $n = \{6, 12, 18, 24, 30\}$. Using a transposed input additionally to the original input channels allows a more realistic reconstruction, especially in regions where the camera and illumination positions are swapped and either of both is represented in the input data stack.

Training

The CNN architecture, which we described in Sec. 7.4.4, was trained by learning its weights using back propagation with the 150 BRDF examples from the dataset described in Sec.7.4.3, which was randomly divided into a training and an evaluation set. The probability of a sample to belong to the training set is 80%. The input data was enriched as described in Sec. 7.4.4. Data augmentation, which is frequently used in CNNs, was not implemented because of the strict reflectance behaviour of BRDFs.

7.4.5 Experimental Results

In this section the qualitative and quantitative evaluation performances of the results of our dense BRDF reconstruction are demonstrated.

Quantitative Evaluation

(a) SSIM evaluation set

(b) SSIM training set

(c) MAE evaluation set
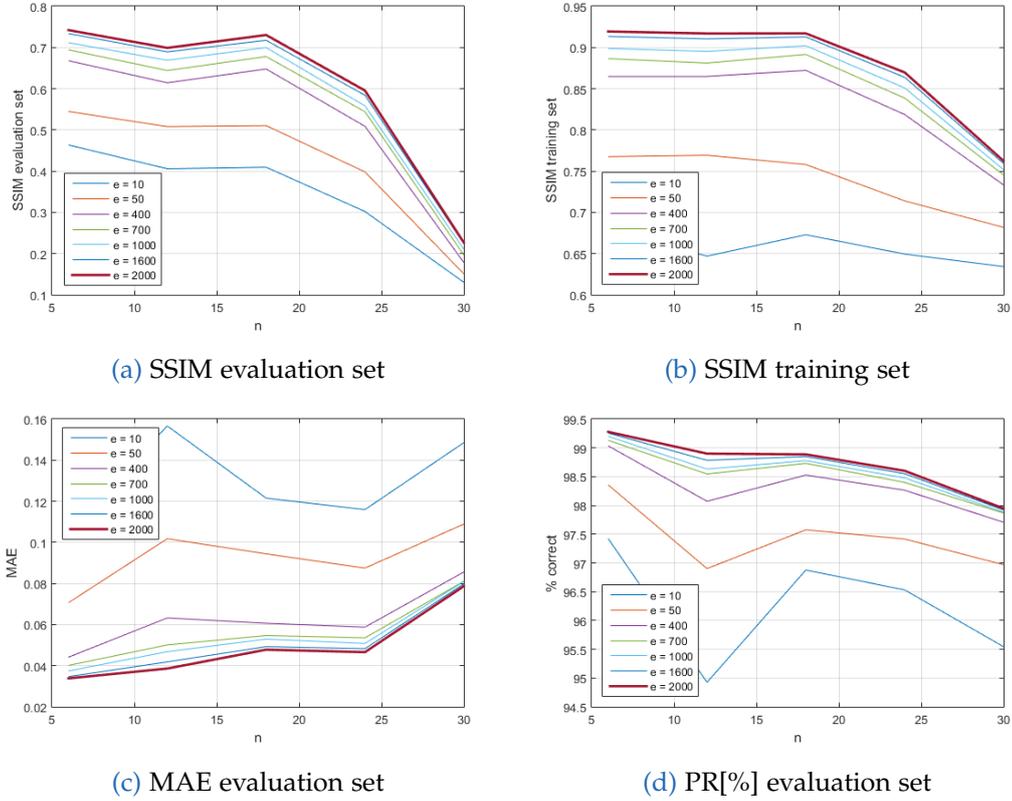
(d) PR[%] evaluation set

Figure 7.17: BRDF learning performance evaluation. Starting from epoch e = 10 up to 2000. The SSIM rate on the evaluation set is shown in (a), (b) depicts the SSIM training error. The MAE is shown in (c), and (d) demonstrates the correct classification in % compared to the full range in each BRDF slice.

Performance evaluations are pursued for defined numbers of known elevation angles in the following order of $u = \{30, 24, 18, 12, 6\}$, which corresponds to maximum viewing elevation angles $\theta_v = \{60°, 45°, 30°, 15°, 0°\}$. Fig. 7.17 shows the the reconstruction using three different evaluation methods, namely the structural similarity (SSIM), the mean absolute error (MAE), and a percentage of correct reconstruction compared to the total range of the ground truth square. In our evaluation representation, the x-axis denotes the number of unknown estimated BRDF data $n = 36 - u$ and the y-axis shows the performance measurements.

type	set	n=6	n=12	n=18	n=24	n=30
SSIM	eval	0.742	0.699	0.730	0.596	0.227
MAE	eval	0.034	0.039	0.047	0.047	0.079
PR [%]	eval	99.28	98.90	98.89	98.60	97.94
SSIM	train	0.919	0.917	0.917	0.870	0.762
MAE	train	0.024	0.028	0.030	0.037	0.067
PR [%]	train	99.26	99.11	99.05	98.88	98.09

Table 7.1: BRDF error evaluation. Performed at epoch 2000 of the evaluation set, for $n = \{6, 12, 18, 24, 30\}$.

The SSIM is shown in Fig. 7.17a and Fig. 7.17b for the evaluation and training set respectively. We compute the SSIM as suggested by Wang *et al.* [212], while taking the mean value over our N estimated slices for all pixel values, as follows:

$$SSIM = \frac{1}{N \cdot P} \sum_{p=1}^{P} \sum_{i=1}^{N} (I_i(p))^\alpha \cdot (c_i(p))^\beta \cdot (s_i(p))^\gamma, \tag{7.3}$$

with the pixel position $p := (x, y)$ and a maximum pixel index P. While I is denoting the luminance, c the contrast and s a structural term, where weights α, β, γ are set to 1. The SSIM covers values from -1 to 1, while a value of 1 describes two identical images.

The MAE is computed by the mean absolute differences between our estimation F_p and the ground truth value L for each slice i and pixel position p:

$$MAE = \frac{1}{N \cdot P} \sum_{p=1}^{P} \sum_{i=1}^{N} |L_{i,p} - F_{i,p}|. \tag{7.4}$$

MAE evaluation results are shown in Fig. 7.17c. Here, a higher value represents a less accurate matching result.

The percentage rate PR of correct classification is defined by the distance from our estimation to the ground truth value in relation to the full value range of a specific slice:

$$PR = 100 \cdot \left(1 - \frac{1}{N \cdot P} \sum_{p=1}^{P} \sum_{i=1}^{N} \frac{|L_{i,p} - F_{i,p}|}{L_{i_{max}} - L_{i_{min}}} \right),$$ (7.5)

where the maximum value over each slice is defined by $L_{i_{max}} = \max_{p \in P} L_{i,p}$ and the minimum value by $L_{i_{min}} = \min_{p \in P} L_{i,p}$ for each slice $i \in N$. The PR evaluation results are shown in Fig. 7.17d. A value of 100% corresponds to a perfect match, while a value of 0% would mean that each pixel in our estimation is on average wrong by the full value range of that slice.

Deep convolutional layers were evaluated, where we used the structure of the first 3×3 convolutional layer F_{c1} up to 5 times. We observed the network for each increased depth up to epoch $e = 3000$ (see Fig. 7.16). Using our specific network type and input data, deeper structures showed a drop in performance with a range from the best result ($e = 3000$) of $SSIM = 0.737$ (one 3×3 convolutional layer) to $SSIM = 0.707$ (five 3×3 convolutional layers). The steepest drop was observed between the use of two and three of these layers. Therefore, a deeper network of that type would not support us with our current reconstruction task. Therefore, we chose a network with one 3×3 convolutional layer.

We tested the reconstruction with our trained networks for $u = 6, 12, 18, 24, 30$ at several iterations, up to the epoch $e = 2000$. The used dataset consists of 150 samples and was split in a training and evaluation set by random choice with a probability of 80% for a sample to belong to the training set. We show the performance on our evaluation set at several iterations as well as the SSIM on the training set (see Fig. 7.17). Notably, a higher number of iterations shows better performance, which indicates, that our network didn't overfit on the training data. The performance at iteration $e = 2000$ is numerically displayed in Tab. 7.1, using the SSIM, MAE and PR performance measurements. Competitive reconstruction results are shown when 1/6 ($n = 6$) to 3/6 ($n = 18$) of the BRDF is estimated, with an SSIM performance on the evaluation set of 0.742, 0.699 and 0.730 respectively. These results are represented in
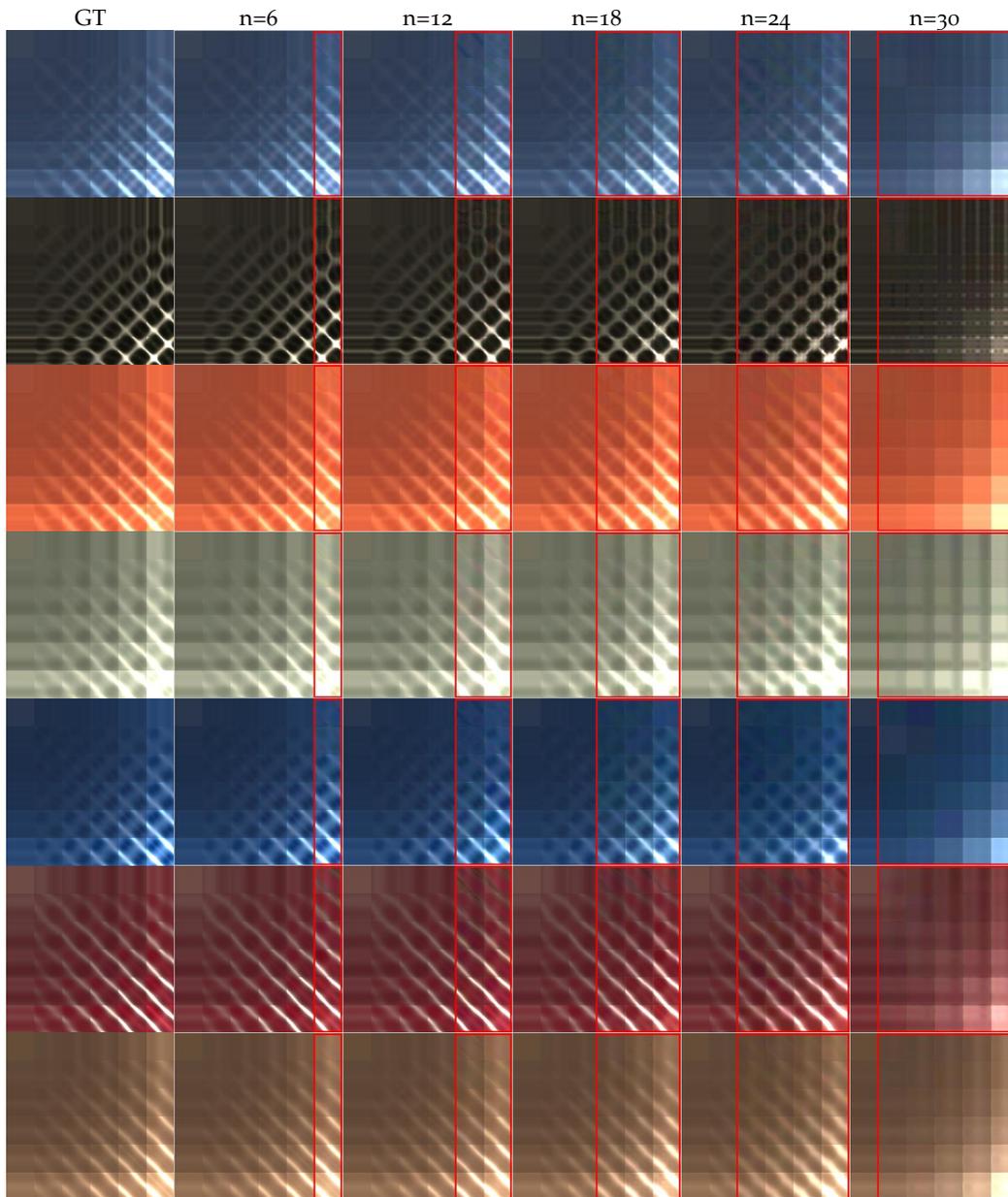
Figure 7.18: BRDF reconstructions where 5/6 to 1/6 of the dense BRDF are known. Columns from left to right: ground truth, $n = \{6, 12, 18, 24, 30\}$ unknown elements. The estimated parts of the BRDF data are framed with a red colour.

column 2 and 3 in Fig. 7.18, where even complex structures at a high angle θ_i and θ_v, shown at the bottom right of the BRDF structures, are well reconstructed.

Qualitative Evaluation

In order to evaluate the qualitative performance of our reconstruction, we show results from the introduced CNN structure, trained with input channels of the size $48 \times 48 \times (2u)$ (see Fig. 7.18). In terms of qualitative visual similarity, the second column ($u = 30$) with $n = 6$ unknown slices up to the the fourth column ($u = 18$) with $n = 18$ unknown slices, are very close to the ground truth. The last column ($u = 6$) with $n = 30$ unknown slices only a general tendency was learned with difficulties to reconstruct detailed structures. A major improvement was achieved by adding transposed image data to the input channels from the used dataset (see Fig. 7.12).

7.5 Conclusion

Capturing light field and photometric stereo data allows the partial reconstruction of reflectance functions. First, we introduced several reflectance functions and described their differences. In Sec. 7.3 we discussed surface reflectance measurement systems, which differ in the flexibility of detail they can capture, the reflectance distributions they can model, the dynamic range of the sensors and the time an acquisition consumes.

In Sec. 7.4 we introduced our work presented in [4]. We introduced the reconstruction of dense BRDFs from limited observation angles, as characteristic for plenoptic image acquisitions. Using light field cameras, the scene is captured from different viewing angles, which are concentrated around a main viewing direction. Using a half sphere structure, various illumination sources can be placed around the scene at several illumination angles. The ability to infer from thereby acquired sparse BRDF data to a more dense representation is of high interest.

We showed the performance evaluation of several convolutional neural networks on BRDF data as well as the reconstruction ability and limitations. We showed the estimation of dense BRDF structures at five different sparsity levels on the UTIA dataset, where up to 5/6 of the BRDF structure was unknown.

Most notably, we gained an improved reconstruction of dense BRDF values from sparse BRDF data of low angles θ_i and high angles θ_v, which are represented towards the top right edge in Fig. 7.18, by using transposed images additionally in the input data. This means that the propagation of the Helmholtz reciprocity from the input layers to the output was supported.

This type of CNN architecture has strong local connections and therefore cannot capture complex behaviors such as the transpose, which is used to follow the Helmholtz reciprocity. We believe that other types of networks can improve the results in regions where the locality relation is weakened. Additional refinements can be achieved by further enriching the input data by useful structural information.

Future work will cover more sophisticated network structures for the reconstruction of the dense BRDF as well as partly reusing pre-learned networks of different topologies. U-Shaped networks were previously utilized for the depth estimation from light field data by Heber *et al.* [88], we will work on an adapted version of such a network structure with a weakened local orientation dependency. Our experiments will be additionally extended to more real-world objects for two different setups, namely a plenoptic camera with a light dome and an industrial acquisition setup with a multi-line-scan camera. For ground truth evaluations we will render BRDFswitch our camera setups and reconstruct more complete BRDF data additionally.

8 Summary and Outlook

Contents

8.1 Summary

In this thesis we focused on the problematic acquisition and reconstruction of highly precise 3D surfaces. To achieve this we combined the globally accurate but locally imprecise light field depth map with surface normal data from photometric stereo.

We provided an in-depth comparison of several variational methods combining depth with surface orientation data. We showed that the weighting of the surface orientation greatly influences the performance on steep edges. Based on these findings, we introduced a more ideal weighting term and provide a gradient based combination which is using a TGV regularization. We showed the evaluation and categorization of these algorithms both on area-scan and multi-line scan data. The latter is applied in in-line industrial setups, where the object is moving at a constant speed on a conveyor belt. We proposed a method to extract photometric stereo surface normal information from multi-line scan light field data. As lies in the nature of the setup, the acquired photometric stereo information is reduced by one dimension. We presented a solution to cope with missing photometric stereo evidence orthogonally to the transport direction. Additionally we presented novel engineered features for industrial light field imaging, which are used in order to achieve a more robust result.

Capturing BRDFs is highly time consuming and unfeasible. We presented the reconstruction of full BRDFs from partial data using a CNN. This approach was evaluated on several degrees of reduction.

The calibration of light field camera systems is important for highly precise depth estimation tasks. We showed the calibration of matrix camera systems with three steps, namely the initial camera-wise calibration, a pairwise calibration and a bundle adjustment step. Multi-line scan calibration has to consider distortions along the scan lines for moving objects. We discussed the geometric calibration of such a system and addressed the transport jitter problem.

8.2 Outlook

For the industrial implementation of our hybrid light field and photometric stereo depth estimation solutions, a detailed evaluation of the algorithm complexity, allocated resources and computational speed are essential. The inherent structure of the algorithms should be exploited to achieve an efficient parallelization. A further assessment of regularization terms and the implementation of machine learning approaches is of high interest. Confidence maps combined with variational methods with regularization terms can further improve the results, especially in regions with highly frequent jumps in depth (e.g grid structures).

Reconstructing full BRDFs from partial data can improve material and defect analysis of objects. BRDFs show strong connections of regions which are non-locally related in the visual function representation. The reason for this lies in complex behaviors such as the Helmholtz reciprocity. Constructing a network which can represent non-local connection in an improved way represents a potential future direction of research.

The computational fast inference of high quality features for light field and specifically multi-line scan data is an important research topic. Learning features specifically for multi-line scanning data using highly precise data should allow the reduction of blurring effects and show a higher robustness in glossy surface regions.

Appendix

Abbreviations

AIT Austrian Institute of Technology

BRDF Bidirectional Reflectance Distribution Function

BSSRDF Bidirectional Scattering-Surface Reflectance Distribution Function

BTF Bidirectional Texture distribution Function

CNN Convolutional Neural Network

CRF Conditional Random Field

CT Census Transform

DMM Dual Minorize Maximize

DOVID Diffractive Optically Variable Image Devices

ELU Exponential Linear Unit

EPI Epipolar Plane Image

FISTA Fast Iterative Shrinkage Thresholding Algorithm

GN Generalized Nehab

GT Ground Truth

HP High-Pass

IID Independent and Identically Distributed

MERL Mitsubishi Electric Research Laboratories

ML Machine Learning

MLP Multi Layer Perceptron

MRF Markov Random Field

MSE Mean Squared Error

PD Primal-Dual

PR Percentage Rate

ReLU Rectified Linear Unit

SfM Structure from Motion

SSIM Structural Similarity Index

TGV Total Generalized Variation

ToF Time of Flight

TV Total Variation

VR Virtual Reality

Bibliography

[1]   E. H. Adelson and J. Y. A. Wang. "Single Lens Stereo With A Plenoptic Camera."
       In: vol. 14. 2. Feb. 1992, pp. 99–106. DOI: 10.1109/34.121783.

[2]   A. Agrawal, R. Chellappa, and R. Raskar. "An Algebraic Approach To Surface
       Reconstruction From Gradient Fields." In: *Tenth IEEE International Conference
       on Computer Vision (ICCV'05) Volume 1*. Vol. 1. Oct. 2005, 174–181 Vol. 1. DOI:
       10.1109/ICCV.2005.31.

[3]   R. Anderson et al. "Jump: Virtual Reality Video." In: 2016.

[4]   D. Antensteiner and S. Štolc. "Full BRDF Reconstruction Using CNNs From
       Partial Photometric Stereo-light Field Data." In: *The IEEE Conference on Computer
       Vision and Pattern Recognition (CVPR) Workshops*. 2017.

[5]   D. Antensteiner, S. Štolc, and R. Huber-Mörk. "Depth Estimation Using Light
       Fields And Photometric Stereo With A Multi-line-scan Framework." In: *Proceed-
       ings of Austrian Association for Pattern Recognition Workshop (OAGM)*. 2016.

[6]   D. Antensteiner, S. Štolc, and R. Huber-Mörk. "Depth Estimation With Light
       Field And Photometric Stereo Data Using Energy Minimization." In: *Proceedings
       of 21st Ibero-American Congress on Pattern Recognition (CIAPR)*. Springer Lecture
       Notes in Computer Science, 2016.

[7]   D. Antensteiner, S. Štolc, and T. Pock. "A Review Of Depth And Normal
       Fusion Algorithms." In: vol. 18. 2. 2018. DOI: 10.3390/s18020431. URL: http:
       //www.mdpi.com/1424-8220/18/2/431.

[8]   D. Antensteiner, S. Štolc, and T. Pock. "Multi-line Scan 3D Sensing With
       A Hybrid Light-field And Photometric Stereo Approach." In: *International
       Conference on Pattern Recognition (ICPR)*. 2018.

[9] D. Antensteiner et al. "High-precision 3D Sensing With Hybrid Light Field & Photometric Stereo Approach In Multi-line Scan Framework." In: 9. 2017, pp. 52–60. DOI: doi:10.2352/ISSN.2470-1173.2017.9.IRIACV-268. URL: https://www.ingentaconnect.com/content/ist/ei/2017/00002017/00000009/art00009.

[10] D. Barath, J. Molnar, and L. Hajder. "Novel Methods For Estimating Surface Normals From Affine Transformations." In: *Computer Vision, Imaging and Computer Graphics Theory and Applications: 10th International Joint Conference, VISIGRAPP 2015, Berlin, Germany, March 11–14, 2015, Revised Selected Papers*. Ed. by J. Braz et al. Cham: Springer International Publishing, 2016, pp. 316–337. ISBN: 978-3-319-29971-6. DOI: 10.1007/978-3-319-29971-6_17. URL: https://doi.org/10.1007/978-3-319-29971-6_17.

[11] S. Barsky and M. Petrou. "The 4-source Photometric Stereo Technique For Three-dimensional Surfaces In The Presence Of Highlights And Shadows." In: vol. 25. 10. Oct. 2003, pp. 1239–1252. DOI: 10.1109/TPAMI.2003.1233898.

[12] T. Basha et al. "Structure And Motion From Scene Registration." In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. June 2012, pp. 1426–1433. DOI: 10.1109/CVPR.2012.6247830.

[13] A. Beck and M. Teboulle. "A Fast Iterative Shrinkage-thresholding Algorithm For Linear Inverse Problems." In: vol. 2. 1. 2009, pp. 183–202. DOI: 10.1137/080716542. eprint: https://doi.org/10.1137/080716542. URL: https://doi.org/10.1137/080716542.

[14] A. Beck and M. Teboulle. "Gradient-based Algorithms With Applications To Signal Recovery Problems." In: 2010, pp. 42–88.

[15] S. Birchfield and C. Tomasi. "A Pixel Dissimilarity Measure That Is Insensitive To Image Sampling." In: vol. 20. 4. Apr. 1998, pp. 401–406. DOI: 10.1109/34.677269.

[16] C. M. Bishop. *Pattern Recognition And Machine Learning (information Science And Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN: 0387310738.

[17]   T. Bishop, S. Zanetti, and P. Favaro. "Light Field Superresolution." In: *Proc. ICCP '09. IEEE International Conference on Computational Photography*. 2009.

[18]   B. Blaschitz, D. Antensteiner, and S. Štolc. "Multi-camera Array Calibration For Light Field Depth Estimation." In: *Proceedings of Austrian Association for Pattern Recognition Workshop (OAGM)*. 2018.

[19]   B. Blaschitz, S. Štolc, and D. Antensteiner. "Geometric Calibration And Image Rectification Of A Multi-line Scan Camera For Accurate 3D Reconstruction." In: *IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision 2018, Burlingame, CA, USA*. 2018.

[20]   E. Bodenstorfer. *Xposure - More Than Just the World's Fastest Multi-line Scan Camera*. https://www.ait.ac.at/fileadmin/mc/digital_safety_security/downloads/1140_BODENSTORFER_svd2016_xposure_Camera_v4.pdf, [Accessed: 2018-03-02].

[21]   R. C. Bolles and H. H. Baker. "Readings In Computer Vision: Issues, Problems, Principles, And Paradigms." In: ed. by M. A. Fischler and O. Firschein. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987. Chap. Epipolar-plane Image Analysis: A Technique for Analyzing Motion Sequences, pp. 26–36. ISBN: 0-934613-33-8. URL: http://dl.acm.org/citation.cfm?id=33517.33520.

[22]   R. C. Bolles, H. H. Baker, and D. H. Marimont. "Epipolar-plane Image Analysis: An Approach To Determining Structure From Motion." In: vol. 1. 1. Springer, 1987, pp. 7–55.

[23]   J.-Y. Bouguet. "Camera Calibration Toolbox For Matlab." In: 2004.

[24]   S. Boyd and L. Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004. ISBN: 0521833787.

[25]   G. Bradski. "Opencv Toolbox." In: 2000.

[26]   K. Bredies, K. Kunisch, and T. Pock. "Total Generalized Variation." In: vol. 3. 3. 2010, pp. 492–526. DOI: 10.1137/090769521. eprint: https://doi.org/10.1137/090769521. URL: https://doi.org/10.1137/090769521.

[27] N. Brosch, S. Štolc, and D. Antensteiner. "Warping-based Motion Artifact Compensation For Multi-line Scan Light Field Imaging." In: *IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision 2018, Burlingame, CA, USA*. 2018.

[28] W. T. Cathey and E. R. Dowski. "New Paradigm For Imaging Systems." In: vol. 41. 29. OSA, 2002, pp. 6080–6092. DOI: 10.1364/AO.41.006080. URL: http://ao.osa.org/abstract.cfm?URI=ao-41-29-6080.

[29] A. Chambolle and T. Pock. "A First-order Primal-dual Algorithm For Convex Problems With Applications To Imaging." In: *Journal of Mathematical Imaging and Vision*. Vol. 40. Springer Netherlands, 2011, pp. 120–145.

[30] A. Chambolle and T. Pock. "An Introduction To Continuous Optimization For Imaging." In: vol. 25. Cambridge University Press, 2016, pp. 161–319. DOI: 10.1017/S096249291600009X.

[31] M. Chandraker, J. Bai, and R. Ramamoorthi. "A Theory Of Differential Photometric Stereo For Unknown Isotropic BRDFs." In: *CVPR 2011*. June 2011, pp. 2505–2512. DOI: 10.1109/CVPR.2011.5995603.

[32] C.-L. Chang et al. "Light Field Compression Using Disparity-compensated Lifting And Shape Adaptation." In: vol. 15. 4. Apr. 2006, pp. 793–806. DOI: 10.1109/TIP.2005.863954.

[33] A. Chatterjee and V. M. Govindu. "Photometric Refinement Of Depth Maps For Multi-albedo Objects." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 933–941. DOI: 10.1109/CVPR.2015.7298695.

[34] C. Chen et al. "Light Field Stereo Matching Using Bilateral Statistics Of Surface Cameras." In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 1518–1525. DOI: 10.1109/CVPR.2014.197.

[35] T. Chen, R. Shibasaki, and Z. Lin. "A Rigorous Laboratory Calibration Method For Interior Orientation Of An Airborne Linear Push-broom Camera." In: vol. 73. 4. 2007, pp. 369–374. DOI: doi:10.14358/PERS.73.4.369. URL: https://www.ingentaconnect.com/content/asprs/pers/2007/00000073/00000004/art00002.

[36] T. Chen, M. Goesele, and H. P. Seidel. "Mesostructure From Specularity." In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. 2006, pp. 1825–1832. DOI: 10.1109/CVPR.2006.182.

[37] X. Chen, D. Li, and J. Zou. "Depth Estimation Of Stereo Matching Based On Microarray Camera." In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. June 2017, pp. 108–112. DOI: 10.1109/ICIVC.2017.7984528.

[38] F. Ciurea et al. "Adaptive Geometric Calibration Correction For Camera Array." In: vol. 2016. 13. Society for Imaging Science and Technology, 2016, pp. 1–6.

[39] D. Clevert, T. Unterthiner, and S. Hochreiter. "Fast And Accurate Deep Network Learning By Exponential Linear Units (elus)." In: vol. abs/1511.07289. 2015. arXiv: 1511.07289. URL: http://arxiv.org/abs/1511.07289.

[40] E. N. Coleman and R. Jain. "Obtaining 3-dimensional Shape Of Textured And Specular Surfaces Using Four-source Photometry." In: vol. 18. 4. 1982, pp. 309–328. DOI: https://doi.org/10.1016/0146-664X(82)90001-6. URL: http://www.sciencedirect.com/science/article/pii/0146664X82900016.

[41] R. T. Collins. "A Space-sweep Approach To True Multi-image Matching." In: *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*. IEEE. 1996, pp. 358–363.

[42] L. Condat, B. Forster-Heinlein, and D. V. D. Ville. "H2o: Reversible Hexagonal-orthogonal Grid Conversion By 1-d Filtering." In: *2007 IEEE International Conference on Image Processing*. Vol. 2. Sept. 2007, pp. II - 73-II - 76. DOI: 10.1109/ICIP.2007.4379095.

[43] CURET. "BRDF Database." In: http://www.cs.columbia.edu/CAVE/curet/, last viewed 2. March 2017.

[44] K. J. Dana et al. "Reflectance And Texture Of Real-world Surfaces." In: vol. 18. 1. New York, NY, USA: ACM, Jan. 1999, pp. 1–34. DOI: 10.1145/300776.300778. URL: http://doi.acm.org/10.1145/300776.300778.

[45] D. G. Dansereau, O. Pizarro, and S. B. Williams. "Decoding, Calibration And Rectification For Lenselet-based Plenoptic Cameras." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. June 2013, pp. 1027–1034. DOI: 10.1109/CVPR.2013.137.

[46] E. R. Dowski and W. T. Cathey. "Single-lens Single-image Incoherent Passive-ranging Systems." In: vol. 33. 29. OSA, Oct. 1994, pp. 6762–6773. DOI: 10.1364/AO.33.006762. URL: http://ao.osa.org/abstract.cfm?URI=ao-33-29-6762.

[47] O. Drbohlav and M. Chaniler. "Can Two Specular Pixels Calibrate Photometric Stereo?" In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Vol. 2. Oct. 2005, 1850–1857 Vol. 2. DOI: 10.1109/ICCV.2005.53.

[48] H. Du, D. Goldman, and S. Seitz. "Binocular Photometric Stereo." In: *Proceedings of British Machine Vision Conference (BMVC)*. 2011, pp. 1–11. ISBN: 1-901725-43-X. DOI: 10.5244/C.25.84.

[49] Z. Du, A. Robles-Kelly, and F. Lu. "Robust Surface Reconstruction From Gradient Field Using The L1 Norm." In: *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*. Dec. 2007, pp. 203–209. DOI: 10.1109/DICTA.2007.4426797.

[50] B. Duvenhage, K. Bouatouch, and D. G. Kourie. "Numerical Verification Of Bidirectional Reflectance Distribution Functions For Physical Plausibility." In: *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. SAICSIT '13. East London, South Africa: ACM, 2013, pp. 200–208. ISBN: 978-1-4503-2112-9. DOI: 10.1145/2513456.2513499. URL: http://doi.acm.org/10.1145/2513456.2513499.

[51] D. Eigen and R. Fergus. "Predicting Depth, Surface Normals And Semantic Labels With A Common Multi-scale Convolutional Architecture." In: vol. abs/1411.4734. 2014. arXiv: 1411.4734. URL: http://arxiv.org/abs/1411.4734.

[52] C. H. Esteban, G. Vogiatzis, and R. Cipolla. "Multiview Photometric Stereo." In: vol. 30. 3. Mar. 2008, pp. 548–554. DOI: 10.1109/TPAMI.2007.70820.

[53] M. F. E. D. F.R.S. "Liv. Thoughts On Ray-vibrations." In: vol. 28. 188. Taylor & Francis, 1846, pp. 345–350. DOI: 10.1080/14786444608645431. eprint: https://doi.org/10.1080/14786444608645431. URL: https://doi.org/10.1080/14786444608645431.

[54] *Facebook Surround-360*. https://facebook360.fb.com/facebook-surround-360/, [Acccessed: 2018-03-09].

[55]   W. Fenchel, D. Blackett, and P. U. D. of Mathematics. Logistics Research Project. *Convex Cones, Sets, And Functions*. Princeton University - Lecture Notes. Princeton University, Department of Mathematics, 1953. URL: https://books.google.at/books?id=qOQ-AAAAIAAJ.

[56]   E. E. Fenimore and T. M. Cannon. "Coded Aperture Imaging With Uniformly Redundant Arrays." In: vol. 17. 3. OSA, Feb. 1978, pp. 337–347. DOI: 10.1364/AO.17.000337. URL: http://ao.osa.org/abstract.cfm?URI=ao-17-3-337.

[57]   K. Fife, E. Gamal, and W. H.-S.P. "A Multi-Aperture Image Sensor with 0.7um Pixels in 0.11um CMOS Technology." In: Dec. 2008, pp. 2990–3005.

[58]   J. Filip and M. Haindl. "Bidirectional Texture Function Modeling: A State Of The Art Survey." In: vol. 31. 11. Nov. 2009, pp. 1921–1940. DOI: 10.1109/TPAMI.2008.246.

[59]   J. Filip, R. Vávra, and M. Havlícek. "Effective Acquisition Of Dense Anisotropic BRDF." In: *International Conference on Pattern Recognition (ICPR)*. Aug. 2014, pp. 2047–2052. DOI: 10.1109/ICPR.2014.357.

[60]   J. Filip et al. "BRDF Slices: Accurate Adaptive Anisotropic Appearance Acquisition." In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. June 2013, pp. 1468–1473. DOI: 10.1109/CVPR.2013.193.

[61]   A. Fitzgibbon, Y. Wexler, and A. Zisserman. "Image-based Rendering Using Image-based Priors." In: vol. 63. 2. July 1, 2005, pp. 141–151. DOI: 10.1007/s11263-005-6643-9. URL: https://doi.org/10.1007/s11263-005-6643-9.

[62]   R. T. Frankot and R. Chellappa. "A Method For Enforcing Integrability In Shape From Shading Algorithms." In: vol. 10. 1988, pp. 439–451.

[63]   Y. Furukawa and J. Ponce. "Accurate Camera Calibration From Multi-view Stereo And Bundle Adjustment." In: vol. 84. 3. Springer, 2009, pp. 257–268.

[64]   A. Fusiello and L. Irsara. "Quasi-euclidean Uncalibrated Epipolar Rectification." In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE. 2008, pp. 1–4.

[65] T. Georgeiv et al. "Spatio-angular Resolution Tradeoffs In Integral Photography." In: *Proceedings of the 17th Eurographics Conference on Rendering Techniques*. EGSR '06. Nicosia, Cyprus: Eurographics Association, 2006, pp. 263–272. ISBN: 3-905673-35-5. DOI: `10.2312/EGWR/EGSR06/263-272`. URL: `http://dx.doi.org/10.2312/EGWR/EGSR06/263-272`.

[66] A. S. Georghiades. "Incorporating The Torrance And Sparrow Model Of Reflectance In Uncalibrated Photometric Stereo." In: *Proceedings Ninth IEEE International Conference on Computer Vision*. Oct. 2003, 816–823 vol.2. DOI: `10.1109/ICCV.2003.1238432`.

[67] T. G. Georgiev and A. Lumsdaine. "Superresolution With Plenoptic 2.0 Cameras." In: *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest*. Optical Society of America, 2009, STuA6. DOI: `10.1364/SRS.2009.STuA6`. URL: `http://www.osapublishing.org/abstract.cfm?URI=SRS-2009-STuA6`.

[68] T. Georgiev, A. Lumsdaine, and S. Goma. "Plenoptic Principal Planes." In: *Imaging and Applied Optics*. Optical Society of America, 2011, JTuD3. DOI: `10.1364/COSI.2011.JTuD3`. URL: `http://www.osapublishing.org/abstract.cfm?URI=COSI-2011-JTuD3`.

[69] G. Gerig. *Lecture Slides: Photometric Stereo, Shape From Shading Sfs Chapter 2 New F&P*. 1st. Utah, USA: Scientific Computing and Imaging Institute, 2013.

[70] A. Gershun, P. Moon, and G. Timoshenko. *The Light Field*. Massachusetts Institute of Technology, 1939. URL: `https://books.google.at/books?id=DyoynQEACAAJ`.

[71] J. Gluckman, S. K. Nayar, and K. J. Thoresz. "Real-time Omnidirectional And Panoramic Stereo." In: *In Proceedings of the 1998 DARPA Image Understanding Workshop*. Morgan Kaufmann, 1998, pp. 299–303.

[72] D. B. Goldman et al. "Shape And Spatially-varying BRDFs From Photometric Stereo." In: vol. 32. 6. June 2010, pp. 1060–1071. DOI: `10.1109/TPAMI.2009.102`.

[73] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[74] S. J. Gortler et al. "The Lumigraph." In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 43–54. ISBN: 0-89791-746-4. DOI: 10.1145/237170.237200. URL: http://doi.acm.org/10.1145/237170.237200.

[75] G. Graber, T. Pock, and H. Bischof. "Online 3D Reconstruction Using Convex Optimization." In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE. 2011, pp. 708–711.

[76] P. Grossmann. "Depth From Focus." In: vol. 5. 1. New York, NY, USA: Elsevier Science Inc., Jan. 1987, pp. 63–69. DOI: 10.1016/0167-8655(87)90026-2. URL: http://dx.doi.org/10.1016/0167-8655(87)90026-2.

[77] R. Gupta and R. I. Hartley. "Linear Pushbroom Cameras." In: vol. 19. 9. Sept. 1997, pp. 963–975. DOI: 10.1109/34.615446.

[78] M. Haindl, J. Filip, and R. Vávra. "Digital Material Appearance: The Curse Of Tera-bytes." In: 90. 2012, pp. 49–50.

[79] M. Haindl, J.Filip, and R. Vávra. "Advanced Textural Representation Of Materials Appearance," in: 90. 2011, 1:1–1:84.

[80] M. Haindl and J. Filip. *Visual Texture. Accurate Material Appearance Measurement, Representation and Modeling*. Advances in Computer Vision and Pattern Recognition. London: Springer-Verlag London, Jan. 2013, pp. 1–284. ISBN: 978-1-4471-4901-9 (Print) 978-1-4471-4902-6 (Online). DOI: 10.1007/978-1-4471-4902-6.

[81] I. Haller et al. "Real-time Semi-global Dense Stereo Solution With Improved Sub-pixel Accuracy." In: *2010 IEEE Intelligent Vehicles Symposium*. June 2010, pp. 369–376. DOI: 10.1109/IVS.2010.5548104.

[82] J. Y. Han and K. Perlin. "Measuring Bidirectional Texture Reflectance With A Kaleidoscope." In: vol. 22. 3. New York, NY, USA: ACM, July 2003, pp. 741–748. DOI: 10.1145/882262.882341. URL: http://doi.acm.org/10.1145/882262.882341.

[83] K.-P. Han, T.-M. Bae, and Y.-H. Ha. "Hybrid Stereo Matching With A New Relaxation Scheme Of Preserving Disparity Discontinuity." In: vol. 33. 5. 2000, pp. 767–785. DOI: https://doi.org/10.1016/S0031-3203(99)00095-3. URL: http://www.sciencedirect.com/science/article/pii/S0031320399000953.

[84] Y. Han, J. Y. Lee, and I. S. Kweon. "High Quality Shape From A Single Rgb-d Image Under Uncalibrated Natural Illumination." In: *2013 IEEE International Conference on Computer Vision*. Dec. 2013, pp. 1617–1624. DOI: 10.1109/ICCV.2013.204.

[85] S. M. Haque, A. Chatterjee, and V. M. Govindu. "High Quality Photometric Reconstruction Using A Depth Camera." In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 2283–2290. DOI: 10.1109/CVPR.2014.292.

[86] M. Harker and P. O'Leary. "Least Squares Surface Reconstruction from Measured Gradient Fields." In: 2008. ISBN: 9781424422432. DOI: 10.1109/CVPR.2008.4587414.

[87] S. Heber. "Variational Shape From Unconventional Imaging Devices." In: *Graz University of Technology, Dissertation Thesis*. May 2015.

[88] S. Heber, W. Yu, and T. Pock. "U-shaped Networks For Shape From Light Field." In: *2016 British Machine Vision Conference BMVC*. 2016.

[89] F. Hernandez-Rodriguez and M. Castelan. "A Mobile Data Acquisition Platform For Photometric Stereo." In: *2009 Electronics, Robotics and Automotive Mechanics Conference (CERMA)*. Sept. 2009, pp. 163–168. DOI: 10.1109/CERMA.2009.76.

[90] A. Hertzmann and S. M. Seitz. "Example-based Photometric Stereo: Shape Reconstruction With General, Varying BRDFs." In: vol. 27. 8. Aug. 2005, pp. 1254–1264. DOI: 10.1109/TPAMI.2005.158.

[91] T. Higo, Y. Matsushita, and K. Ikeuchi. "Consensus Photometric Stereo." In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 2010, pp. 1157–1164. DOI: 10.1109/CVPR.2010.5540084.

[92] T. Higo et al. "A Hand-held Photometric Stereo Camera For 3-D Modeling." In: *2009 IEEE 12th International Conference on Computer Vision*. Sept. 2009, pp. 1234–1241. DOI: 10.1109/ICCV.2009.5459331.

[93] H. Hirschmuller. "Stereo Processing By Semiglobal Matching And Mutual Information." In: vol. 30. 2. Feb. 2008, pp. 328–341. DOI: 10.1109/TPAMI.2007.1166.

[94] H. Hirschmüller, P. R. Innocent, and J. Garibaldi. "Real-time Correlation-based Stereo Vision With Reduced Border Errors." In: vol. 47. 1. Apr. 1, 2002, pp. 229–246. DOI: 10.1023/A:1014554110407. URL: https://doi.org/10.1023/A:1014554110407.

[95] K. Honauer et al. "A Dataset And Evaluation Methodology For Depth Estimation On 4D Light Fields." In: *ACCV: Asian Conference on Computer Vision*. Ed. by S.-H. Lai. Springer. Springer, 2016.

[96] R. Horaud, R. Mohr, and B. Lorecki. "On Single-scanline Camera Calibration." In: vol. 9. 1. Feb. 1993, pp. 71–75. DOI: 10.1109/70.210796.

[97] B. K. P. Horn and M. J. Brooks, eds. *Shape From Shading*. Cambridge, MA, USA: MIT Press, 1989. ISBN: 0-262-08183-0.

[98] B. K. P. Horn and M. J. Brooks. "The Variational Approach To Shape From Shading." In: vol. 33. 2. San Diego, CA, USA: Academic Press Professional, Inc., Feb. 1986, pp. 174–208. DOI: 10.1016/0734-189X(86)90114-3. URL: http://dx.doi.org/10.1016/0734-189X(86)90114-3.

[99] H. Ives. "A camera for making parallax panoramagrams." In: vol. 17, 4. Dec. 1928, p. 4.

[100] C. Je and H.-M. Park. "Optimized Hierarchical Block Matching For Fast And Accurate Image Registration." In: vol. 28. 7. 2013, pp. 779–791. DOI: https://doi.org/10.1016/j.image.2013.04.002. URL: http://www.sciencedirect.com/science/article/pii/S0923596513000581.

[101] H. W. Jensen et al. "A Practical Model For Subsurface Light Transport." In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 511–518. ISBN: 1-58113-374-X. DOI: 10.1145/383259.383319. URL: http://doi.acm.org/10.1145/383259.383319.

[102] O. Johannsen et al. "On The Calibration Of Focused Plenoptic Cameras." In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications: Dagstuhl 2012 Seminar on Time-of-Flight Imaging and GCPR 2013 Workshop on Imaging New Modalities.* Ed. by M. Grzegorzek et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 302–317. ISBN: 978-3-642-44964-2. DOI: 10.1007/978-3-642-44964-2_15. URL: https://doi.org/10.1007/978-3-642-44964-2_15.

[103] L. Jongchul et al. "Improved Census Transform For Noise Robust Stereo Matching." In: vol. 55. 2016, pp. 55 - 55 - 10. DOI: 10.1117/1.OE.55.6.063107.

[104] J. R. Jordan and A. C. Bovik. "Using Chromatic Information In Edge-based Stereo Correspondence." In: vol. 54. 1. 1991, pp. 98–118. DOI: https://doi.org/10.1016/1049-9660(91)90077-3. URL: http://www.sciencedirect.com/science/article/pii/1049966091900773.

[105] N. Joshi and D. J. Kriegman. "Shape From Varying Illumination And Viewpoint." In: *2007 IEEE 11th International Conference on Computer Vision.* Oct. 2007, pp. 1–7. DOI: 10.1109/ICCV.2007.4409021.

[106] A. Kadambi et al. "Polarized 3D: High-quality Depth Sensing With Polarization Cues." In: *Proceedings of IEEE International Conference on Computer Vision (ICCV).* 2015, pp. 3370–3378. DOI: 10.1109/ICCV.2015.385.

[107] T. Kanade, P. Rander, and P. J. Narayanan. "Virtualized Reality: Constructing Virtual Worlds From Real Scenes." In: vol. 4. 1. Jan. 1997, pp. 34–47. DOI: 10.1109/93.580394.

[108] M. Kazhdan, M. Bolitho, and H. Hoppe. "Poisson Surface Reconstruction." In: *Proceedings of the Fourth Eurographics Symposium on Geometry Processing.* SGP '06. Cagliari, Sardinia, Italy: Eurographics Association, 2006, pp. 61–70. ISBN: 3-905673-36-3. URL: http://dl.acm.org/citation.cfm?id=1281957.1281965.

[109] C. Kim. "3D Reconstruction And Rendering From High Resolution Light Fields." In: ETH Zurich, 2015, pp. -. DOI: 10.3929/ethz-a-010554038. URL: http://hdl.handle.net/20.500.11850/106586.

[110] C. Kim et al. "Scene Reconstruction From High Spatio-angular Resolution Light Fields." In: vol. 32. 4. New York, NY, USA: ACM, July 2013, 73:1–73:12. DOI:

10.1145/2461912.2461926. URL: http://doi.acm.org/10.1145/2461912.2461926.

[111] R. Klette. *Concise Computer Vision*. Springer, 2014.

[112] J. Ko and Y. S. Ho. "Stereo Matching Using Census Transform Of Adaptive Window Sizes With Gradient Images." In: *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. Dec. 2016, pp. 1–4. DOI: 10.1109/APSIPA.2016.7820827.

[113] K. Köser. "Geometric Estimation With Local Affine Frames And Free-form Surfaces." In: *Christian-Albrechts-Universität zu Kiel, Dissertation Thesis*. 2008.

[114] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet Classification With Deep Convolutional Neural Networks." In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[115] B. Krolla, M. Diebold, and D. Stricker. "Light Field From Smartphone-based Dual Video." In: *Computer Vision - ECCV 2014 Workshops*. Ed. by L. Agapito, M. M. Bronstein, and C. Rother. Cham: Springer International Publishing, 2015, pp. 600–610. ISBN: 978-3-319-16181-5.

[116] M. Lab. "VisTex Database." In: http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html, last viewed 2. March 2018.

[117] J. Lang, H.-P. Seidel, and H. P. A. Lensch. "View Planning For BRDF Acquisition." In: *ACM SIGGRAPH 2003 Sketches &Amp; Applications*. SIGGRAPH '03. San Diego, California: ACM, 2003, pp. 1–1. DOI: 10.1145/965400.965407. URL: http://doi.acm.org/10.1145/965400.965407.

[118] Y. LeCun et al. "Gradient-based Learning Applied To Document Recognition." In: vol. 86. 11. Nov. 1998, pp. 2278–2324.

[119] Y. LeCun et al. "Object Recognition With Gradient-based Learning." In: *Shape, Contour and Grouping in Computer Vision*. London, UK, UK: Springer-Verlag, 1999, pp. 319–. ISBN: 3-540-66722-9. URL: http://dl.acm.org/citation.cfm?id=646469.691875.

[120]   A. Levin et al. "Image And Depth From A Conventional Camera With A Coded Aperture." In: vol. 26. 3. New York, NY, USA: ACM, July 2007. DOI: 10.1145/1276377.1276464. URL: http://doi.acm.org/10.1145/1276377.1276464.

[121]   M. Levoy. "Light Fields And Computational Imaging." In: vol. 39. 8. Aug. 2006, pp. 46–55. DOI: 10.1109/MC.2006.270.

[122]   M. Levoy and P. Hanrahan. "Light Field Rendering." In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. ISBN: 0-89791-746-4. DOI: 10.1145/237170.237199. URL: http://doi.acm.org/10.1145/237170.237199.

[123]   B. Li and T. Furukawa. "Photometric Stereo Under Dichromatic Reflectance Framework Dealing With Non-lambertian Surfaces." In: IEEE, 2015, pp. 139–144.

[124]   F.-F. Li, J. Johnson, and S. Yeung. "Stanford CNN Course." In: Soring 2017. Stanford, USA: University of Stanford. URL: cs231n.stanford.edu/.

[125]   C.-K. Liang et al. "Programmable Aperture Photography: Multiplexed Light Field Acquisition." In: vol. 27. 3. New York, NY, USA: ACM, Aug. 2008, 55:1–55:10. DOI: 10.1145/1360612.1360654. URL: http://doi.acm.org/10.1145/1360612.1360654.

[126]   E. Lilienblum, A. Al-Hamadi, and B. Michaelis. "A Coded 3D Calibration Method For Line-scan Cameras." In: *Pattern Recognition*. Ed. by J. Weickert, M. Hein, and B. Schiele. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 81–90. ISBN: 978-3-642-40602-7.

[127]   H. S. Lim and T. O. Binford. "Stereo Correspondence: A Hierarchical Approach." In: *Proceedings of the Image Understanding Workshop*. Vol. 1. 1987, pp. 234–241.

[128]   H. Lin et al. "Depth Recovery From Light Field Using Focal Stack Symmetry." In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015, pp. 3451–3459. DOI: 10.1109/ICCV.2015.394.

[129]   D. Lindell and J. Thatte. "Virtual Reality Motion Parallax With The Facebook Surround-360." In: 2017.

[130] G. Lippmann. "Épreuves Réversibles- Photographies Intégrales." In: Mar. 1908, pp. 446–451.

[131] G. Liu et al. "Postprocessing Algorithms To Minimize Fixed-pattern Artifact And Reduce Trigger Jitter In Swept Source Optical Coherence Tomography." In: *Opt. Express*. Vol. 23. Apr. 2015, pp. 9824–9834.

[132] H. C. Longuet-Higgins. "Readings In Computer Vision: Issues, Problems, Principles, And Paradigms." In: ed. by M. A. Fischler and O. Firschein. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, pp. 61–62. ISBN: 0-934613-33-8. URL: http://dl.acm.org/citation.cfm?id=33517.33523.

[133] A. Lumsdaine and T. Georgiev. "Full resolution lightfield rendering." In: 2008.

[134] C. A. Luna et al. "Calibration Of Line-scan Cameras." In: vol. 59. 8. Aug. 2010, pp. 2185–2190. DOI: 10.1109/TIM.2009.2031344.

[135] W.-C. Ma et al. "Rapid Acquisition Of Specular And Diffuse Normal Maps From Polarized Spherical Gradient Illumination." In: *Proceedings of the 18th Eurographics Conference on Rendering Techniques*. EGSR'07. Grenoble, France: Eurographics Association, 2007, pp. 183–194. ISBN: 978-3-905673-52-4. DOI: 10.2312/EGWR/EGSR07/183-194. URL: http://dx.doi.org/10.2312/EGWR/EGSR07/183-194.

[136] A. L. Maas, A. Y. Hannun, and A. Y. Ng. "Rectifier Nonlinearities Improve Neural Network Acoustic Models." In: *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*. 2013.

[137] E. Malis and M. Vargas. *Deeper Understanding Of The Homography Decomposition For Vision-based Control*. Research Report RR-6303. INRIA, 2007, p. 90. URL: https://hal.inria.fr/inria-00174036.

[138] S. P. Mallick et al. "Beyond Lambert: Reconstructing Specular Surfaces Using Color." In: *IEEE CVPR*. Vol. 2. June 2005, 619–626 vol. 2.

[139] W. Matusik et al. "A Data-driven Reflectance Model." In: vol. 22. 3. July 2003, pp. 759–769.

[140] W. Matusik et al. "Image-based Visual Hulls." In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 369–374. ISBN: 1-58113-208-5. DOI: 10.1145/344779.344951. URL: http://dx.doi.org/10.1145/344779.344951.

[141] S. McCloskey. "Masking Light Fields To Remove Partial Occlusion." In: *International Conference on Pattern Recognition (ICPR)*. Aug. 2014, pp. 2053–2058. DOI: 10.1109/ICPR.2014.358.

[142] W. S. McCulloch and W. Pitts. "A Logical Calculus Of The Ideas Immanent In Nervous Activity." In: vol. 5. 4. Dec. 1, 1943, pp. 115–133. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259.

[143] L. Mertz and N. O. Young. "Fresnel transformations of images." In: 1961, p. 305.

[144] J. J. Moreau. "Proximité et dualité dans un espace hilbertien." In: vol. 93. 1965, pp. 273–299.

[145] S. K. Nayar, X. S. Fang, and T. Boult. "Removal Of Specularities Using Color And Polarization." In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. June 1993, pp. 583–590. DOI: 10.1109/CVPR.1993.341071.

[146] S. K. Nayar, K. Ikeuchi, and T. Kanade. "Determining Shape And Reflectance Of Hybrid Surfaces By Photometric Sampling." In: vol. 6. 4. Aug. 1990, pp. 418–431. DOI: 10.1109/70.59367.

[147] S. K. Nayar, K. Ikeuchi, and T. Kanade. "Shape From Interreflections." In: vol. 6. 3. Aug. 1, 1991, pp. 173–195. DOI: 10.1007/BF00115695. URL: https://doi.org/10.1007/BF00115695.

[148] D. Nehab et al. "Efficiently Combining Positions And Normals For Precise 3D Geometry." In: vol. 24. 3. 2005, p. 536. ISBN: 0730-0301. DOI: 10.1145/1073204.1073226.

[149] Y. Nesterov. "Smooth Minimization Of Non-smooth Functions." In: vol. 103. 1. May 1, 2005, pp. 127–152. DOI: 10.1007/s10107-004-0552-5. URL: https://doi.org/10.1007/s10107-004-0552-5.

[150] Y. Nesterov. *Introductory Lectures On Convex Optimization: A Basic Course*. 1st ed. Springer Publishing Company, Incorporated, 2014. ISBN: 1461346916, 9781461346913.

[151] R. Ng et al. *Light Field Photography With A Hand-held Plenoptic Camera*. Tech. rep. CSTR 2005-02. Stanford University, Apr. 2005.

[152] R. Ng. "Digital Light Field Photography." AAI3219345. PhD thesis. Stanford, CA, USA, 2006. ISBN: 978-0-542-70779-7.

[153] F. Nicodemus et al. "Geometrical Considerations And Nomenclature For Reflectance." In: NBS Monograph 160, U.S. Department of Commerce, National Bureau of Standards. 1977.

[154] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi. "On Optimal, Minimal BRDF Sampling For Reflectance Acquisition." In: vol. 34. 6. New York, NY, USA: ACM, Oct. 2015, 186:1–186:11. DOI: 10.1145/2816795.2818085. URL: http://doi.acm.org/10.1145/2816795.2818085.

[155] H. K. Nishihara. *Prism: A Practical Real-time Imaging Stereo Matcher*. Tech. rep. Cambridge, MA, USA, 1984.

[156] M. Okutomi and T. Kanade. "A Multiple-baseline Stereo." In: vol. 15. 4. IEEE, 1993, pp. 353–363.

[157] R. Owens. "Computer Vision Lectures." In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2004. URL: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/.

[158] N. Parikh and S. Boyd. *Proximal Algorithms*. Now Publishers Inc., 2013. ISBN: 1601987161, 9781601987167.

[159] J. Park et al. "High Quality Depth Map Upsampling For 3D-tof Cameras." In: *2011 International Conference on Computer Vision*. Nov. 2011, pp. 1623–1630. DOI: 10.1109/ICCV.2011.6126423.

[160] J. Park et al. "Robust Multiview Photometric Stereo Using Planar Mesh Parameterization." In: vol. 39. 8. Aug. 2017, pp. 1591–1604. DOI: 10.1109/TPAMI.2016.2608944.

[161] S. C. Pei and Y. Y. Wang. "Color Invariant Census Transform For Stereo Matching Algorithm." In: *2013 IEEE International Symposium on Consumer Electronics (ISCE)*. June 2013, pp. 209–210. DOI: 10.1109/ISCE.2013.6570188.

[162] A. P. Pentland. "A New Sense For Depth Of Field." In: vol. 9. 4. Washington, DC, USA: IEEE Computer Society, Apr. 1987, pp. 523–531. DOI: 10.1109/TPAMI.1987.4767940. URL: http://dx.doi.org/10.1109/TPAMI.1987.4767940.

[163] *Persistence of Vision Pty. Ltd. (2004) [Computer software]*. http://www.povray.org/download/, [Acccessed: 2017-11-10].

[164] A. Pevar et al. "Real-time Photometric Stereo." In: *Photogrammetric Week edition*. 2015.

[165] M. W. Powell, S. Sarkar, and D. Goldgof. "A Simple Strategy For Calibrating The Geometry Of Light Sources." In: vol. 23. 9. Washington, DC, USA: IEEE Computer Society, Sept. 2001, pp. 1022–1027. DOI: 10.1109/34.955114. URL: http://dx.doi.org/10.1109/34.955114.

[166] Y. Quéau et al. "Led-based Photometric Stereo: Modeling, Calibration And Numerical Solution." In: vol. abs/1707.01018. 2017. arXiv: 1707.01018. URL: http://arxiv.org/abs/1707.01018.

[167] T. Randen. "Brodatz Textures." In: http://www.ux.uis.no/ tranden/brodatz.html, last viewed 2. March 2018.

[168] R. T. Rockafellar. *Convex Analysis*. 1970.

[169] F. Rosenblatt. "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain." In: 1958, pp. 65–386.

[170] M. Rostami, O. Michailovich, and Z. Wang. "Gradient-based Surface Reconstruction Using Compressed Sensing." In: *2012 19th IEEE International Conference on Image Processing*. Sept. 2012, pp. 913–916. DOI: 10.1109/ICIP.2012.6467009.

[171] M. Sattler, R. Sarlette, and R. Klein. "Efficient And Realistic Visualization Of Cloth." In: *Proceedings of the 14th Eurographics Workshop on Rendering*. EGRW '03. Leuven, Belgium: Eurographics Association, 2003, pp. 167–177. ISBN: 3-905673-03-7. URL: http://dl.acm.org/citation.cfm?id=882404.882429.

[172] T. Scherr. "Gradient-based Surface Reconstruction And The Application To Wind Waves." In: *Ruprecht-Karls-University Heidelberg, Master Thesis*. 2017.

[173]   C. Schwartz et al. "Design And Implementation Of Practical Bidirectional Texture Function Measurement Devices Focusing On The Developments At The University Of Bonn." In: vol. 14. 5. 2014, pp. 7753–7819. DOI: 10.3390/s140507753. URL: http://www.mdpi.com/1424-8220/14/5/7753.

[174]   P. Seitz. *Using Local Orientational Information As Image Primitive For Robust Object Recognition*. 1989. DOI: 10.1117/12.970173. URL: https://doi.org/10.1117/12.970173.

[175]   S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory To Algorithms*. New York, NY, USA: Cambridge University Press, 2014. ISBN: 1107057132, 9781107057135.

[176]   A. Shekhovtsov et al. "Solving Dense Image Matching In Real-time Using Discrete-continuous Optimization." In: *Proceedings of 21st Computer Vision Winter Workshop (CVWW)*. 2016, p. 13. ISBN: 978-3-85125-388-7.

[177]   B. Shi et al. "Photometric Stereo Using Internet Images." In: *3DV*. 2014.

[178]   R. Shiradkar, P. Tan, and S. H. Ong. "Auto-calibrating Photometric Stereo Using Ring Light Constraints." In: vol. 25. 3. Apr. 1, 2014, pp. 801–809. DOI: 10.1007/s00138-013-0591-0. URL: https://doi.org/10.1007/s00138-013-0591-0.

[179]   N. Shroff et al. "Finding A Needle In A Specular Haystack." In: *IEEE ICRA*. IEEE, May 2011, pp. 5963–5970.

[180]   C. Singh Verma and M.-J. Wu. In: last viewed 28. March 2018. Madison, WI: University of Wisconsin-Madison. URL: http://pages.cs.wisc.edu/~csverma/CS766_09/Stereo/stereo.html.

[181]   N. Snavely, S. M. Seitz, and R. Szeliski. "Photo Tourism: Exploring Photo Collections In 3D." In: vol. 25. 3. New York, NY, USA: ACM, July 2006, pp. 835–846. DOI: 10.1145/1141911.1141964. URL: http://doi.acm.org/10.1145/1141911.1141964.

[182]   F. Solomon and K. Ikeuchi. "Extracting The Shape And Roughness Of Specular Lobe Objects Using Four Light Photometric Stereo." In: *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 1992, pp. 466–471. DOI: 10.1109/CVPR.1992.223149.

[183] D. Soukup et al. "Depth Estimation Within A Multi-line-scan Light-field Framework." In: *Advances in Visual Computing*. Ed. by G. Bebis et al. Cham: Springer International Publishing, 2014, pp. 471–481. ISBN: 978-3-319-14364-4.

[184] D. Soukup and R. Huber-Mörk. "Mobile Hologram Verification With Deep Learning." In: 2017, pp. 169–172.

[185] D. Soukup, S. Štolc, and R. Huber-Mörk. "On Optimal Illumination For Dovid Description Using Photometric Stereo." In: *Advanced Concepts for Intelligent Vision Systems*. Ed. by S. Battiato et al. Cham: Springer International Publishing, 2015, pp. 553–565. ISBN: 978-3-319-25903-1.

[186] *The Stanford 3D Scanning Repository*. http://graphics.stanford.edu/data/3Dscanrep/, [Acccessed: 2017-11-10].

[187] J. Stewart et al. "A New Reconstruction Filter For Undersampled Light Fields." In: *Proceedings of the 14th Eurographics Workshop on Rendering*. EGRW '03. Leuven, Belgium: Eurographics Association, 2003, pp. 150–156. ISBN: 3-905673-03-7. URL: http://dl.acm.org/citation.cfm?id=882404.882426.

[188] T. Stich, A. Tevs, and M. Magnor. "Global Depth From Epipolar Volumes–a General Framework For Reconstructing Non-lambertian Surfaces." In: *3D Data Processing, Visualization, and Transmission, Third International Symposium on*. June 2006, pp. 916–923. DOI: 10.1109/3DPVT.2006.69.

[189] S. Štolc and R. Huber-Mörk. "Computational Imaging Using A Multi-linescan Light-field Camera." In: *Laser Focus World*. 2014, 506:19–23.

[190] S. Štolc et al. "Depth And All-in-focus Imaging By A Multi-line-scan Light-field Camera." In: *Journal of Electronic Imaging*. Vol. 23. 5. 2014, pp. 19–23. DOI: 10.1117/1.JEI.23.5.053020. URL: http://dx.doi.org/10.1117/1.JEI.23.5.053020.

[191] B. Sun et al. "Time-varying BRDFs." In: vol. 13. 3. Piscataway, NJ, USA: IEEE Educational Activities Department, May 2007, pp. 595–609. DOI: 10.1109/TVCG.2007.1013. URL: http://dx.doi.org/10.1109/TVCG.2007.1013.

[192] T. Svoboda, D. Martinec, and T. Pajdla. "A Convenient Multicamera Self-calibration For Virtual Environments." In: vol. 14. 4. Aug. 2005, pp. 407–422. DOI: 10.1162/105474605774785325.

[193] R. Szeliski. *Computer Vision: Algorithms And Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc., 2010. ISBN: 1848829345, 9781848829343.

[194] H. D. Tagare and R. J. P. deFigueiredo. "A Theory Of Photometric Stereo For A Class Of Diffuse Non-lambertian Surfaces." In: vol. 13. 2. Feb. 1991, pp. 133–152. DOI: 10.1109/34.67643.

[195] K. Tanaka and S. Tachi. "Tornado: Omnistereo Video Imaging With Rotating Optics." In: vol. 11. 6. Piscataway, NJ, USA: IEEE Educational Activities Department, Nov. 2005, pp. 614–625. DOI: 10.1109/TVCG.2005.107. URL: http://dx.doi.org/10.1109/TVCG.2005.107.

[196] J. Tanida et al. "Color Imaging With An Integrated Compound Imaging System." In: vol. 11. 18. OSA, Sept. 2003, pp. 2109–2117. DOI: 10.1364/OE.11.002109. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-11-18-2109.

[197] J. Tanida et al. "Thin Observation Module By Bound Optics (tombo): Concept And Experimentalverification." In: vol. 40. 11. OSA, Apr. 2001, pp. 1806–1813. DOI: 10.1364/AO.40.001806. URL: http://ao.osa.org/abstract.cfm?URI=ao-40-11-1806.

[198] M. W. Tao et al. "Depth Estimation For Glossy Surfaces With Light-field Cameras." In: *ECCV Workshop*. Springer International Publishing, 2015, pp. 533–547.

[199] M. W. Tao et al. "Depth From Combining Defocus And Correspondence Using Light-field Cameras." In: Dec. 2013. URL: http://graphics.berkeley.edu/papers/Tao-DFC-2013-12/.

[200] M. W. Tao and P. P. Srinivasan. "Depth From Shading, Defocus, And Correspondence Using Light-field Angular Coherence." In: *Proc. Comp. Vis. and Pat. Rec. (CVPR)*. IEEE, June 2015. ISBN: 9781467369640. DOI: 10.1109/CVPR.2015.7298804.

[201] D. Taylor. "Virtual Camera Movement: The Way Of The Future?" In: 77. American Cinematographer, Sept. 1996, pp. 93–100.

[202]   C. Ti et al. "Simultaneous Time-of-flight Sensing And Photometric Stereo With A Single Tof Sensor." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 4334–4342. DOI: 10.1109/CVPR.2015. 7299062.

[203]   T. Tomioka et al. "Depth Map Estimation Using Census Transform For Light Field Cameras." In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2016, pp. 1641–1645. DOI: 10.1109/ICASSP. 2016.7471955.

[204]   I. Tosic and K. Berkner. "Light Field Scale-Depth Space Transform for Dense Depth Estimation." In: IEEE, 2014, pp. 435–442. ISBN: 9781479943081. DOI: 10.1109/CVPRW.2014.71.

[205]   V. Vaish et al. "Using Plane + Parallax For Calibrating Dense Camera Arrays." In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. June 2004, I-2-I-9 Vol.1. DOI: 10.1109/CVPR.2004.1315006.

[206]   V. Vaish et al. "Reconstructing Occluded Surfaces Using Synthetic Apertures: Stereo, Focus And Robust Measures." In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*. CVPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2331–2338. ISBN: 0-7695-2597-0. DOI: 10.1109/CVPR.2006.244. URL: http://dx.doi.org/10. 1109/CVPR.2006.244.

[207]   K. Valentín, S. Štolc, and R. Huber-Mörk. "Improved Cost Computation With Local Binary Features In A Multi-view Block Matching Framework." In: *Proceedings of 10th International Conference on Measurement*. 2015, pp. 21–24.

[208]   K. Venkataraman et al. "PiCam: An Ultra-thin High Performance Monolithic Camera Array." In: vol. 32. 5. New York, NY, USA: ACM, Nov. 2013, 166:1–166:13. DOI: 10.1145/2508363.2508390. URL: http://doi.acm.org/10.1145/ 2508363.2508390.

[209]   O. Wang et al. "Material Classification Using BRDF Slices." In: *In: Proc. CVPR*. 2009.

[210] T. C. Wang, A. A. Efros, and R. Ramamoorthi. "Occlusion-aware Depth Estimation Using Light-field Cameras." In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015, pp. 3487–3495. DOI: 10.1109/ICCV.2015.398.

[211] T. Wang et al. "A 4D Light-field Dataset And CNN Architectures For Material Recognition." In: vol. abs/1608.06985. 2016. URL: http://arxiv.org/abs/1608.06985.

[212] Z. Wang et al. "Image Quality Assessment: From Error Visibility To Structural Similarity." In: vol. 13. 4. Apr. 2004, pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[213] S. Wanner and B. Goldlücke. "Globally Consistent Depth Labeling Of 4D Light Fields." In: *Proc. of Comp. Vis. Pat. Rec. (CVPR)*. 2012, pp. 41–48.

[214] S. Wanner, C. Straehle, and B. Goldluecke. "Globally Consistent Multi-label Assignment On The Ray Space Of 4D Light Fields." In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 1011–1018. ISBN: 978-0-7695-4989-7. DOI: 10.1109/CVPR.2013.135.

[215] L. Watteeuw et al. "Imaging Characteristics Of Graphic Materials With The Minidome (rich)." In: *ICOM-CC graphic documents working group interim meeting*. 2013.

[216] T. Weise, B. Leibe, and L. V. Gool. "Fast 3D Scanning With Automatic Motion Compensation." In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. June 2007, pp. 1–8. DOI: 10.1109/CVPR.2007.383291.

[217] C. Weissig et al. "The Ultimate Immersive Experience: Panoramic 3D Video Acquisition." In: *Advances in Multimedia Modeling*. Ed. by K. Schoeffmann et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 671–681. ISBN: 978-3-642-27355-1.

[218] J. Weng et al. "Estimating Motion/structure From Line Correspondences: A Robust Linear Algorithm And Uniqueness Theorems." In: *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*. June 1988, pp. 387–392. DOI: 10.1109/CVPR.1988.196264.

[219]   B. Wilburn et al. "High Performance Imaging Using Large Camera Arrays."
        In: vol. 24. 3. New York, NY, USA: ACM, July 2005, pp. 765–776. DOI: 10.1145/
        1073204.1073259. URL: http://doi.acm.org/10.1145/1073204.1073259.

[220]   G. Willems et al. "Easy And Cost-effective Cuneiform Digitizing." In: *In Short
        and Project Papers Proceedings of 6th International Symposium on Virtual Reality.
        Archaeology and Cultural Heritage.* 2005.

[221]   L. B. Wolff. "Using Polarization To Separate Reflection Components." In: *Com-
        puter Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer
        Society Conference on.* June 1989, pp. 363–369. DOI: 10.1109/CVPR.1989.37873.

[222]   R. J. Woodham. "Photometric Method For Determining Surface Orientation
        From Multiple Images." In: vol. 19. 1. 1980, pp. 139–144.

[223]   R. J. Woodham. "Shape From Shading." In: ed. by B. K. P. Horn and M. J.
        Brooks. Cambridge, MA, USA: MIT Press, 1989. Chap. Photometric Method for
        Determining Surface Orientation from Multiple Images, pp. 513–531.

[224]   C. Wu et al. "High-quality Shape From Multi-view Stereo And Shading Under
        General Illumination." In: *Computer Vision and Pattern Recognition (CVPR), 2011
        IEEE Conference on.* IEEE, June 2011, pp. 969–976.

[225]   C. Wu et al. "Real-time Shading-based Refinement For Consumer Depth Cam-
        eras." In: vol. 33. Citeseer, 2014, p. 3.

[226]   T.-P. Wu and C.-K. Tang. "Dense Photometric Stereo Using A Mirror Sphere
        And Graph Cut." In: *2005 IEEE Computer Society Conference on Computer Vision
        and Pattern Recognition (CVPR).* Vol. 1. IEEE, June 2005, 140–147 vol. 1. DOI:
        10.1109/CVPR.2005.123.

[227]   *Xapt Eye-sect Xa.* http://eye-sect.com/en/products/cameras/eye-sect-xa,
        [Acccessed: 2018-03-09].

[228]   Y. Xiong et al. "From Shading To Local Shape." In: vol. 37. 1. Jan. 2015, pp. 67–
        79.

[229]   Y. Xu, X. Jin, and Q. Dai. "Depth Fused From Intensity Range And Blur
        Estimation For Light-field Cameras." In: *2016 IEEE International Conference on
        Acoustics, Speech and Signal Processing (ICASSP).* Mar. 2016, pp. 2857–2861. DOI:
        10.1109/ICASSP.2016.7472199.

[230] J. C. Yang et al. "A Real-time Distributed Light Field Camera." In: *Proceedings of the 13th Eurographics Workshop on Rendering*. EGRW '02. Pisa, Italy: Eurographics Association, 2002, pp. 77–86. ISBN: 1-58113-534-3. URL: http://dl.acm.org/citation.cfm?id=581896.581907.

[231] X. Yang, X. Chen, and J. Xi. "Block Based Dense Stereo Matching Using Adaptive Cost Aggregation And Limited Disparity Estimation." In: *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. May 2017, pp. 1–6. DOI: 10.1109/I2MTC.2017.7969972.

[232] R. Yassine, H. Mohamed, and H. Issa. "Global Techniques For Edge Based Stereo Matching, Scene Reconstruction Pose Estimation And Tracking." In: DOI: 10.5772/4944.

[233] C. K. Yeh et al. "Shape-from-shifting: Uncalibrated Photometric Stereo With A Mobile Device." In: *2017 IEEE 13th International Conference on e-Science (e-Science)*. Oct. 2017, pp. 551–558. DOI: 10.1109/eScience.2017.89.

[234] H. Yu. "Edge-preserving Photometric Stereo Via Depth Fusion." In: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 2472–2479. ISBN: 978-1-4673-1226-4. URL: http://dl.acm.org/citation.cfm?id=2354409.2354863.

[235] L.-F. Yu et al. "Shading-based Shape Refinement Of RGB-D Images." In: *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*. CVPR '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1415–1422. ISBN: 978-0-7695-4989-7. DOI: 10.1109/CVPR.2013.186. URL: http://dx.doi.org/10.1109/CVPR.2013.186.

[236] R. Zabih and J. Woodfill. "Non-parametric Local Transforms For Computing Visual Correspondence." In: *Computer Vision — ECCV '94*. Ed. by J.-O. Eklundh. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 151–158. ISBN: 978-3-540-48400-4.

[237] C. Zhang and T. Chen. "A Self-reconfigurable Camera Array." In: *ACM SIGGRAPH 2004 Sketches*. SIGGRAPH '04. Los Angeles, California: ACM, 2004,

pp. 151–. ISBN: 1-58113-896-2. DOI: 10.1145/1186223.1186412. URL: http://doi.acm.org/10.1145/1186223.1186412.

[238]   L. Zhang et al. "Shape And Motion Under Varying Illumination: Unifying Structure From Motion, Photometric Stereo, And Multi-view Stereo." In: *The 9th IEEE International Conference on Computer Vision*. Nice, France, Oct. 2003, pp. 618–625.

[239]   Q. Zhang and R. Pless. "Extrinsic Calibration Of A Camera And Laser Range Finder (improves Camera Calibration)." In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*. Vol. 3. Sept. 2004, 2301–2306 vol.3. DOI: 10.1109/IROS.2004.1389752.

[240]   Z. Zhang. "Flexible Camera Calibration By Viewing A Plane From Unknown Orientations." In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Vol. 1. Ieee. 1999, pp. 666–673.

[241]   Z. Zhou and P. Tan. "Ring-light Photometric Stereo." In: *Computer Vision – ECCV 2010*. Ed. by K. Daniilidis, P. Maragos, and N. Paragios. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 265–279. ISBN: 978-3-642-15552-9.

[242]   B. Zupancic and C. Soler. "Sparse BRDF Approximation Using Compressive Sensing." In: *SIGGRAPH Asia 2013 Posters*. SA '13. Hong Kong: ACM, 2013, 42:1–42:1. ISBN: 978-1-4503-2634-6. DOI: 10.1145/2542302.2542352. URL: http://doi.acm.org/10.1145/2542302.2542352.