

Dissertation  
submitted to the  
Combined Faculties for the  
Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

put forward by  
M.Sc. Marcel Gutsche  
born in Frankfurt am Main, Germany  
Date of oral exam: July, 16th 2018



Light Fields

Reconstructing Geometry and  
Reflectance Properties

---

Referees:  
PD Dr. rer. nat. Christoph Garbe  
Prof. Dr. rer. nat. Karl-Heinz Brenner

**Zusammenfassung:** Maschinelles Sehen spielt eine wichtige Rolle im Fortschritt der Automatisierung und der Digitalisierung in unserer Gesellschaft. Die Konstruktion exakter 3D-Modelle ist dabei eine zentrale Herausforderung. Das Lichtfeld beschreibt dabei für jeden Raumpunkt und jeden Raumwinkel die Lichtstrahlen. Diese Datenfülle die Lichtfelder bieten erlaubt präzise Tiefenschätzungen, erfordert aber auch die Entwicklung neuer Algorithmen.

Insbesondere spekulare Reflexionen bereiten existierenden Algorithmen Probleme. Der Grund dafür liegt in der vereinfachenden Annahme der meisten Algorithmen, dass die Helligkeit eines Objektpunktes konstant über verschiedene Ansichten bleibt. Die meisten Oberflächen erzeugen teilweise spekulare Reflektionen, was bei der Erstellung von Tiefenkarten berücksichtigt werden muss.

In dieser Arbeit entwickeln wir verbesserte Algorithmen, die basierend auf Glanzlichtern die Oberflächenneigung schätzen. Dazu werden Epipolarbilder untersucht, die mit Lichtfeldaufbauten gewonnen werden. Lichtfelder bieten die Möglichkeit, Reflektionseigenschaften anhand von Intensitätsvariationen im Epipolarraum zu charakterisieren. Dieser Raum wird analysiert und mit der erwarteten Reflektanz verglichen, die mit Hilfe der Rendergleichung und verschiedenen bidirektionalen Reflektanzverteilungsfunktionen modelliert wird. Damit können nicht nur hoch präzise Oberflächennormalen und Tiefenkarten bestimmt werden, sondern auch Materialeigenschaften, welche in den Reflexionscharakteristiken enthalten sind.

Die Ergebnisse dieser Arbeit zeigen, dass mit den neuen Algorithmen genauere Tiefenkarten erstellt und zusätzlich Materialeigenschaften gemessen werden können, wenn mehrere Kameras in einem Lichtfeldaufbau benutzt werden.

---

**Abstract:**

Computer vision plays an important role in the progress of automation and digitalization of our society. One of the key challenges is the creation of accurate 3D representations of our environment. The rich information in light fields can enable highly accurate depth estimates, but requires the development of new algorithms.

Especially specular reflections pose a challenge for many reconstruction algorithms. This is due to the violation of the brightness consistency assumption, which only holds for Lambertian surfaces. Most surfaces are to some extent specular and an appropriate handling is central to avoid erroneous depth maps.

In this thesis we explore the potential of using specular highlights to determine the orientation of surfaces. To this end, we examine epipolar images in light field set ups. In light field data, reflectance properties can be characterized by intensity variations in the epipolar plane space. This space is analysed and compared to the expected reflectance, which is modelled using the render equation with different bidirectional reflection distribution functions.

This approach allows us to infer highly accurate surface normals and depth estimates. Furthermore, it reveals material properties encoded in the reflectance by inspecting the intensity profile. Our results demonstrate the potential to increase the accuracy of the depth maps. Multiple cameras in a light field set up let us retrieve additional material properties encoded in the reflectance.





# Acknowledgements

The research I did from 2014 to 2017 at the HCI in Heidelberg would not have been possible without all the great people I had the honour to work with.

Special thanks to Dr. Christoph Garbe, who made this dissertation possible and whose guidance and practical views were always valuable and encouraging. My gratitude also to Prof. Brenner, who agreed to be my second referee.

It was a pleasure working with the whole Light Field group / HD Vision Systems consisting of Christoph Garbe, Maximilian Diebold, Hendrik Schilling, Hamza Aziz, Alessandro Vianello, Sven Wanner and Wolfgang Mischler and our Ground Truth co-group Karsten Krispin, Katrin Honauer, Oliver Zendel, Alexander Brock and Burkhard Güssefeld. They did not only provide intelligent counsel in every subject related matter, but also were fantastic table football companions.

I am thankful for the financial support to Prof. Katja Mombaur, Prof. Bernd Jähne and the HGS MathComp which allowed me to fully concentrate on the research.

I also thank my friends and my family for their unconditional support during all these years. Especially I want to thank Johann Brehmer, Merle Reinhart, Florian Dinger, Svenja Reith, Karsten Krispin and Hendrik Schilling for taking their time to proofread parts of this thesis.



# Contents

<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Related Work . . . . .	4
1.3 Contribution . . . . .	7
1.4 Outline . . . . .	8
<b>2 Theory of Light Fields and Reflectance</b>	<b>9</b>
2.1 Light Transport . . . . .	9
2.1.1 The Bidirectional Reflectance Distribution Function . . . . .	13
2.1.2 Light Model . . . . .	18
2.2 From Stereo to Light Fields . . . . .	19
2.2.1 Stereo . . . . .	20
2.2.2 Light Fields . . . . .	21
2.2.3 Structure Tensor . . . . .	24
2.2.4 Hough Transform . . . . .	25
<b>3 Capturing the Light Field</b>	<b>27</b>
3.1 Capturing Devices . . . . .	27
3.1.1 Translation Stage . . . . .	28
3.1.2 Camera Array . . . . .	28
3.1.3 Microlens Arrays . . . . .	29
3.2 Optimal Trajectories through the BRDF Space . . . . .	30
3.2.1 Geometric Consideration of the Camera Setup . . . . .	31
3.2.2 Half-way Vector Parametrization . . . . .	32
3.2.3 Linear Camera Motion and Fixed Light Position . . . . .	34
3.2.4 Linear Camera Motion and a Co-Moving Light Source . . . . .	35

3.2.5	Fixed Camera and Light Source, Rotating Object . . . . .	36
3.2.6	Optimal Acquisition . . . . .	36
<b>4</b>	<b>Surface Normal Estimation via Geometrical Optics</b>	<b>39</b>
4.1	The Inverse Problem . . . . .	39
4.2	Solving the Inverse Problem by Geometry . . . . .	41
4.3	Results . . . . .	43
<b>5</b>	<b>Simultaneous BRDF and Surface Normal Extraction</b>	<b>47</b>
5.1	Solving the Inverse Problem using a BRDF Model . . . . .	47
5.2	Implementation for Real and Synthetic Light Field Data . . . . .	51
5.2.1	Preprocessing . . . . .	52
5.2.2	Occlusion Handling . . . . .	54
5.2.3	Specular Highlight Detection . . . . .	55
5.2.4	Joint Estimation of Surface Normals and BRDF Parameters	55
5.2.5	Initialization and Optimization . . . . .	58
5.3	Results . . . . .	59
5.3.1	Metrics for Surface Normal Analysis . . . . .	60
5.3.2	Comparison of Fresnel Effects for Different Materials . . . . .	61
5.3.3	Synthetic Evaluation . . . . .	62
5.3.4	Evaluation on Real World Objects . . . . .	65
<b>6</b>	<b>Consistent Estimation of BRDF, Surface Normals and Depth</b>	<b>71</b>
6.1	Sparse Depth Estimation . . . . .	74
6.1.1	Zero Crossings . . . . .	74
6.1.2	Line Fitting . . . . .	74
6.2	Optimization with an Occlusion Aware Bilateral Regularization Scheme . . . . .	77
6.2.1	Initialization . . . . .	77
6.2.2	Objective Function for Diffuse Regions . . . . .	78
6.2.3	Occlusions . . . . .	80
6.2.4	Iteration Scheme . . . . .	80

6.2.5	Activation of the BRDF-Surface Normal Optimizer . . . .	81
6.3	Results . . . . .	82
6.3.1	Synthetic . . . . .	82
6.3.2	Real . . . . .	84
<b>7</b>	<b>Conclusion</b>	<b>87</b>
7.1	Summary . . . . .	87
7.2	Conclusions and Outlook . . . . .	88
<b>A</b>	<b>Appendix</b>	<b>91</b>
	<b>List of Own Publications</b>	<b>93</b>
	<b>List of Figures</b>	<b>95</b>
	<b>Bibliography</b>	<b>101</b>



# 1

## Introduction

### 1.1 Motivation

One of the central challenges in computer vision is to recover the constituents of a 3D scene, including the geometry, the illumination and material properties. Disentangling all these elements from 2D images is commonly referred to as *inverse rendering* and a holy grail in computer vision.

These physical qualities of our surrounding surfaces, namely distance, orientation and material properties can be summarized by the term *early vision*, as stated by Poggio, Torre, and Koch [45]. In contrast, high-level vision tasks are more concerned with image semantics, meaning object detection, segmentation and classification. This thesis deals with the former, the reconstruction of distance, orientation and material properties using light fields. Knowing all the constituents of a scene in terms of geometry, illumination and reflection properties would tremendously help with a variety of tasks.

The applications of robust and precise reconstruction algorithms are almost endless and cover general navigation tasks through 3D space as well as content creation of real objects for more and more widespread 3D printers. There is currently a booming market for autonomous vehicles, such as drones, robots or self-driving cars, all in need for 3D maps of their surroundings to navigate safely.

In blockbuster films, computer generated imagery (CGI), based on 3D models of faces is used to generate digital make up with astonishing results, see Figure 1.1. Many high-level computer vision tasks that are challenging based on only 2D imagery become almost trivial when additional depth information is incorporated, such as the segmentation of foreground and background objects. These findings have led to an enormous interest in research and camera development, such that even consumer devices are equipped with RGBD cameras, where additionally to the red, blue and green channel also depth is recorded. Recovering the underlying 3D world from 2D images is not only useful in a variety of applications, but it also provides interesting and challenging questions in the area of inverse rendering, the process of recovering the geometry, material properties and illumination leading to the observed images. A hard problem, which is still unsolved.

In particular the effect of reflective materials on the appearance of objects is complex, handling these specular reflections is challenging. One key ingredient for the recovery of the 3D information is the relation of the same 3D point in, at least, two different views. The simplest approach is to assume constant brightness between the images. But many materials like metals or plastics do not reflect light equally in all directions. This aspect of non-uniform light distribution is discussed in great detail and possible solutions are addressed in this thesis.

The problem affects many applications. Examples include automated defect detection, where metallic surfaces are challenging to handle, and dental reconstruction techniques for the creation of custom-fit dentures. Non-uniform light distributions are the central topic of this thesis. We discuss it in detail and explore possible solutions.

An elegant way of representing and thinking about these non-uniform distributions is the light field representation. It describes the amount of light for each point in space and for each direction. In addition to the 3D geometry, this formalism also describes the reflectance properties of the surfaces. There are many ways to capture part of the light field. For still scenes, gantries with a single



(a) Digital make up based on the 3D location of a face featuring Davy Jones for Disney's Pirates of the Caribbean. Image taken from [14].



(b) Parrot drone for creating 3D models from above. Image taken from [43].



(c) 3D printer for metal objects. Image taken from [54].

Figure 1.1: Many technologies today depend on accurate 3D models.

mounted camera are used. Camera arrays allow for dynamic scenes, but require more sophisticated calibration procedures. Also, light field cameras for the consumer market are available. They provide the angular separation using microlens arrays in front of the image sensor.

The huge amount of data in light fields is both a blessing and a challenge: while it is the reason behind their usefulness for many applications, it also makes them notoriously difficult to process. But according to Moore’s law, these computational challenges will rather sooner than later pose a less urgent problem. Since many real world objects have specular characteristics, handling these phenomena correctly in the 3D reconstruction process helps to build robust navigation algorithms which do not falter in uncontrolled environments.

Motivated by the possible applications and the holy grail of perfect inverse rendering in mind, this thesis addresses the question of how to take advantage of specular highlights in the 3D reconstruction process, especially in combination with light field capture.

## 1.2 Related Work

A plethora of methods for 3D reconstruction have been established. Many methods rely on triangulation, similar to the human vision system. Active systems are another important group of techniques. They are based on the duration between emitting and receiving a signal. Examples include time of flight cameras (ToF [15]), radar, lidar, sonar and ultrasound. In contrast, passive systems do not emit a signal and only capture incoming light with one, two or more cameras.

Methods have been proposed to estimate scene geometry from a single image [32, 33]. These techniques require either strong prior knowledge or use deep neural networks with a need for large training data sets. Due to their reliance

on a single image they suffer from low accuracy and can only provide depth estimates up to a scale factor, also referred to as scale ambiguity.

Other interesting techniques use more images from a single view. In methods based on photometric stereo, images are acquired from the same position under changing light conditions. This approach, first proposed by Woodham [64], uses the information that the intensity varies with the angle between the surface normal and the direction of the incoming light. Using enough measurements with different light positions, the surface normal can be constrained to a unique solution. Other work also includes specular highlights in their photometric stereo approaches [22, 35].

Depth from defocus uses multiple views from the same position but with different foci to estimate depth. The amount of blurring caused by the distance to the focal plane corresponds to the depth. This is especially used to get depth information in microscopy.

The methods presented in this thesis are based on 3D reconstruction from structured light fields. In general, light fields can be described by the plenoptic function [2], which describes the light as intensity as a function of both position and direction. It represents a high-dimensional function which contains all information about the light emitted or reflected at an object surface. To reduce its dimensionality, Gortler *et al.* [18] and Levoy *et al.* [31] introduce the use of a 4D subset from the full light field which enables sampling of the light field using multiple cameras. Over the last years, state of the art light field scene reconstruction methods provide an increasing precision of depth estimates, even for difficult materials, at the cost of high runtimes [39, 65, 56, 13].

The accurate description of material properties is a topic often discussed in the computer graphics community with its interest in realistic rendering. Many models in particular for forward reflections have been developed. But typically they prioritize the aesthetics of the rendering as well as the computational effi-

ciency over the physical accuracy. In Chapter 2 we will discuss the theory behind material properties encoded in the bidirectional distribution function (BRDF) in more detail.

In the computer vision community specularities are often treated as undesired component. Hence, first publications were concerned with the detection of specular highlights and the exclusion of such areas to avoid failure of algorithms based on Lambertian assumptions [4, 16, 29].

Early examinations of the information in specular highlights available for a moving observer with a static light source were conducted by Zisserman *et al.* [66]. They concluded that the information contained in the specular highlight by two or more images is sufficient to solve the convex/concave ambiguity, *i.e.* distinguishing if a surface is bending inward or outward. However, they could not further constrain the curve generated by the moving observer on the surface. Ramamoorthi and Hanrahan [46] describe the reflected light fields as the convolution of the light distribution with the BRDF, and the reconstruction of the scene geometry as deconvolution. Jin *et al.* [25] proposed another approach which utilizes a rank based cost function in a multi-view stereo setting. Other multi-view methods exploit multiple orientations in epipolar images, or similar features [10, 62, 11, 59].

Similar to the methods presented in this thesis is the work of Adato *et al.* [1] but it is restricted to mirror like surfaces. They use the apparent displacement of the surface highlight in an optical flow frame work. Nair *et al.* [38] incorporate reflections and material properties in their stereo framework, but they do not actually handle lighting. Instead, they attribute all shading effects to the diffuse colour.

Oxholm *et al.* use a multi-view setup to infer geometry and reflectance properties by means of a probabilistic model [42]. In contrast to the method presented here they need a full illumination model, whereas our method only needs the

position of the strongest light source, which can be inferred by many different methods.

Jachnik *et al.* use methods based on self-localization and mapping algorithms (SLAM) to recover the specular and diffuse components of surfaces to apply augmented reality [23]. However, they assume planar surfaces and do not provide surface normal accuracies.

Wang *et al.* proposed a BRDF invariant theory to recover shape from light fields [60]. Their method generates depth and surface normals from input images, but relies on quadratic shape priors and smoothness constraints in neighbouring regions, whereas the method developed in this thesis provides pixel-wise depth and surface normal estimates.

## 1.3 Contribution

In this thesis, we develop methods for 3D reconstruction in the situation where the position of the light source and cameras are known. While we do not present a one-step solution for the holy grail of inverse rendering, which would include inferring the position of cameras and light sources, we make the following novel contributions:

- A novel algorithm for estimating surface normals based on the position of specular highlights in a light field camera set up.
- A procedure to estimate surface normals and BRDF parameters based on intensity variations in epipolar plane images.
- The integration of the surface normal and BRDF parameter estimator in general framework for deriving depth, material properties and surface nor-

mals.

## 1.4 Outline

In Chapter 2 the theoretical background for this thesis is provided. Starting from the mechanics of light transport, we explain the interaction processes, in particular the reflection on surfaces, and discuss the theory of 3D reconstruction based on light field data.

Chapter 3 is devoted to the light field acquisition process. We present the hardware used for gathering real data in this thesis, and show results for an optimal capturing setup.

In Chapter 4, we develop methods to reconstruct the surface normal from specular highlights in a cross shaped light field array. Here we only take the maximum position of the highlight for estimation. This leads to an interesting geometrical optimization problem.

In Chapter 5 we extend these techniques of by also taking into account the total intensity distribution. We demonstrate that this allows us to extract surface material properties in addition to the surface normal orientation.

Chapter 6 is the culmination of our work: We finally develop techniques to simultaneously estimate the depth, the material properties and the surface normal orientation.

Chapter 7 gives a brief summary of this thesis and addresses possible future work.

# 2

## Theory of Light Fields and Reflectance

In this chapter an introduction about the theory of light transport and 3D reconstructions is presented. It does not go into all the details, but gives the basic insights and formulas which this thesis is build upon. A more throughout introduction and more details can be found in the books *Digitale Bildverarbeitung, 6th revised and extended edition* from Jähne [24] or in *Multiple View Geometry in Computer Vision* from Hartley and Zisserman[21]. Most of the explanations with regards to the bidirectional distribution functions (BRDFs) can be found in more detail in “An Overview of BRDF Models” from Montes Soldado and Ureña Almagro[37].

### 2.1 Light Transport

For the understanding of the image formation process the concepts of the interaction between light and matter are highlighted. Depending on the material light will be partly transmitted, absorbed or reflected, see Figure 2.1. The reflection and transmission processes are governed by the Fresnel equations. They depend on the polarization of the light, the angle between the incoming light ray  $\omega_i$  and the surface normal  $\mathbf{n}$ , as well as on the material properties summarized by the

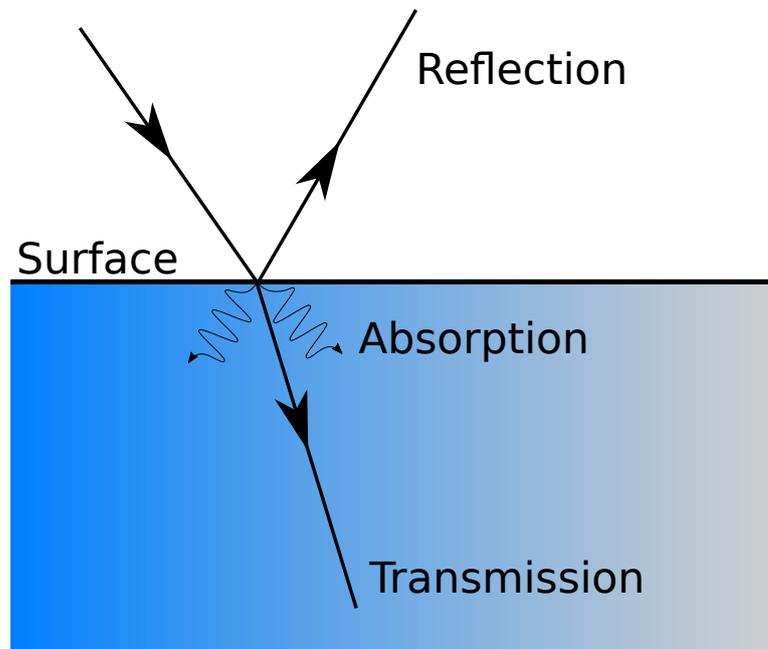


Figure 2.1: Light at an interface can be transmitted, absorbed or reflected. The total energy of the light is conserved. Second order effects, such as subsurface scattering, are not shown here.

refractive index  $n$  of the media at the interface. Additionally photons can be absorbed either by exciting vibrational modes of molecules or by exciting energy levels in the atom. We will ignore quantum specifics of light such as diffraction, but concentrate on the effects essential in geometrical optics. Before going into the details some radiometric terms are presented.

The energy of a photon is given by

$$Q_\gamma = \frac{hc}{\lambda}, \quad (2.1)$$

where  $h$  is Planck's constant and  $c$  the speed of light. The flux or power is hence

$$\Phi = \frac{dQ}{dt}. \quad (2.2)$$

The irradiance is the flux per surface area, given as

$$E = \frac{d\Phi}{dA}. \quad (2.3)$$

For example the radiation of the Sun causes an irradiance above the Earth of  $1.361 \text{ kW m}^{-2}$ , known as the solar constant.

The radiance, given as

$$L_{e,\Omega} = \frac{d^2\Phi_e}{d\Omega dA \cos\theta}, \quad (2.4)$$

describes the amount of energy per time per unit area per solid angle and therefore has the units  $\text{W sr}^{-1} \text{m}^{-2}$ . The factor  $A \cos \theta$  represents the projected area under which the radiance is emitted, *e.g.*

$$A_{\text{projected}} = A \cdot (\mathbf{n} \cdot \boldsymbol{\omega}) = A \cos \theta, \quad (2.5)$$

where  $\mathbf{n}$  is the normalized surface normal of  $A$  and  $\boldsymbol{\omega}$  the normalized direction of the in or outgoing light ray. The radiance is constant along geometrical lines of sight and does not fall off in contrast to the irradiance. This means that when a surface emits a radiance  $L_e$  the same radiance will be received by an optical system. However, this is not valid for point light sources, which are not fully resolvable in the imaging system. Here we would have to take the inverse square law into account.

We distinguish between diffuse and specular reflection. For a purely specular reflection the angle of the incident light is equal to the angle of reflection, *e.g.* a mirror. Diffuse reflection occurs on rough surfaces or due to scattering centers beneath the surface. If the light is distributed in all directions with equal likelihood we have a purely diffuse material. Almost all material have some mixture of a specular and a diffuse component. Plaster or marble are examples for very diffuse materials, while thin aluminium layers are strongly specular in the visible light. Diffuse surfaces are also called Lambertian reflectors, since they obey the Lambertian cosine law

$$L_o = \frac{k_d}{\pi} L_i \cos \theta_i, \quad (2.6)$$

where  $L_i$  is the incident radiance and  $\theta_i$  the angle between the surface normal and the incoming light,  $k_d \in [0, 1]$  is the diffuse albedo of the material and  $L_o$  the outgoing radiance. The factor of  $\frac{1}{\pi}$  is due to the integration of the cosine weighted hemisphere.

### 2.1.1 The Bidirectional Reflectance Distribution Function

Non Lambertian surfaces have a more complicated behaviour. Depending on the material we have larger reflection contributions towards certain directions, mostly in the direction of total reflection. This phenomenon is called specularly and in its most extreme form represents a perfect mirror, and all incoming rays obey the law of reflection. The bidirectional reflectance distribution function (BRDF) describes the ratio of reflected radiance  $L_r$  to the incoming irradiance  $E_i$ , depending on the in and outgoing directions  $(\omega_i, \omega_o)$  of the light. It is defined as

$$f_r(\omega_i, \omega_o) = \frac{dL_o(\omega_o)}{L_i(\omega_i) \cos(\theta_i) d\omega_i} \quad (2.7)$$

We note, that if  $f_r(\omega_i, \omega_o)$  is constant, integration yields the diffuse reflection, such that

$$f_{r, \text{diffuse}}(\omega_i, \omega_o) = \frac{k_d}{\pi}. \quad (2.8)$$

To be physically plausible, a BRDF has to have the following properties:

- The BRDF must conserve energy,  $\forall \omega_i, \int_{\Omega} f_r(\omega_i, \omega_o) \cos \theta_r d\omega_o \leq 1$ ,
- it must be positive, *e.g.*  $f_r(\omega_i, \omega_o) \geq 0$
- and it must satisfy the Helmholtz reciprocity:  $f_r(\omega_i, \omega_o) = f_r(\omega_o, \omega_i)$ .

This four dimensional function can additionally simplified by assuming isotropic reflection, meaning, that the BRDF is invariant with respect to rotations around the surface normal. Figure 2.2 shows an overview of different BRDF models,

classifying them in theoretical and empirical models. Theoretical models try to mirror the physical interactions between photons and matter as exact as possible while empirical models try to create realistically looking models which can be computed efficiently. A third group not shown in the image are measured BRDFs. Instead of figuring out a parametric model a goniometer, a device which automatically drives light source and sensor to a wide range on the hemisphere, is used to capture the BRDF for many combinations of incoming and outgoing directions. One examples for such a database for different materials is the MERL database[36]. Of course, such measurements are demanding and have a limited angular resolution.

In computer graphics BRDFs are often not physically plausible to allow efficient computation. The most simple one is the Phong model[44]. It is given by

$$f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) = k_d + k_s (\boldsymbol{\omega}_r \cdot \boldsymbol{\omega}_o)^\alpha, \quad (2.9)$$

where  $\boldsymbol{\omega}_r$  is the reflection vector, calculated by  $\boldsymbol{\omega}_r = 2(\boldsymbol{\omega}_i \cdot \mathbf{n})\mathbf{n} - \boldsymbol{\omega}_i$ . The constants  $k_d$ ,  $k_s$  and  $\alpha$  are non physical and refer to the diffuse and specular reflectivity and  $\alpha$  controls the sharpness of the specular peak. Higher values represent a more narrow intensity maximum.

Instead of using the incoming and outgoing light vectors, see (Figure 2.3a) one can establish another parametrization for isotropic BRDFs using the halfway vector  $\mathbf{h}$ . The halfway vector parametrization was proposed by [48] mainly to simplify calculations for isotropic BRDFs. In the parametrization the halfway vector is given by

$$\mathbf{h} = \frac{\boldsymbol{\omega}_i + \boldsymbol{\omega}_o}{\|\boldsymbol{\omega}_i + \boldsymbol{\omega}_o\|}.$$

The direction of the halfway vector is described by the angles  $\Theta_h$  and  $\Phi_h$ . The

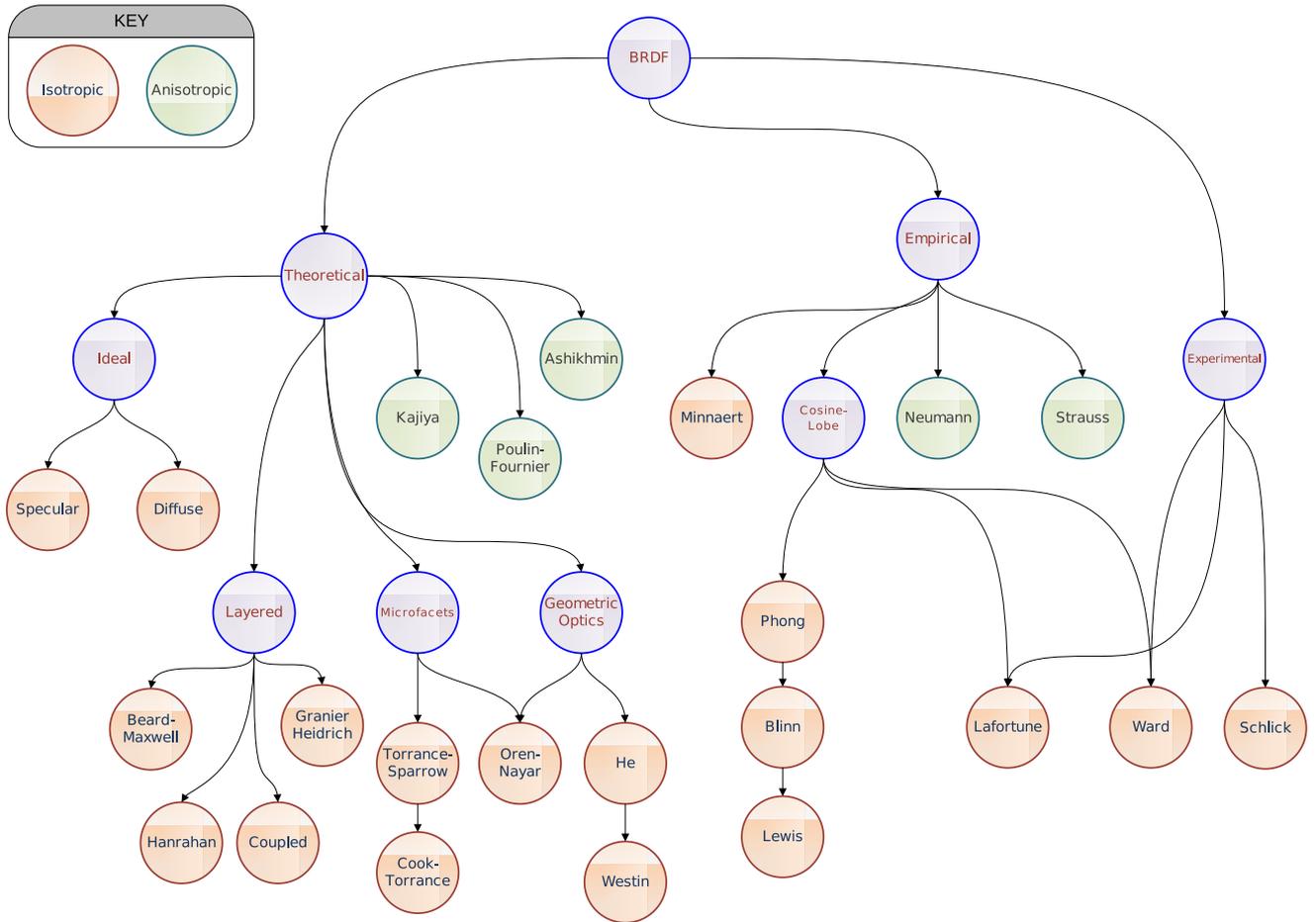


Figure 2.2: There is a plethora of different BRDF models, which can be categorized according to different attributes [37].

second pair of angles  $\Theta_d$  and  $\Phi_d$  describe the orientation of the incoming ray relative to the half vector (Figure 2.3b).

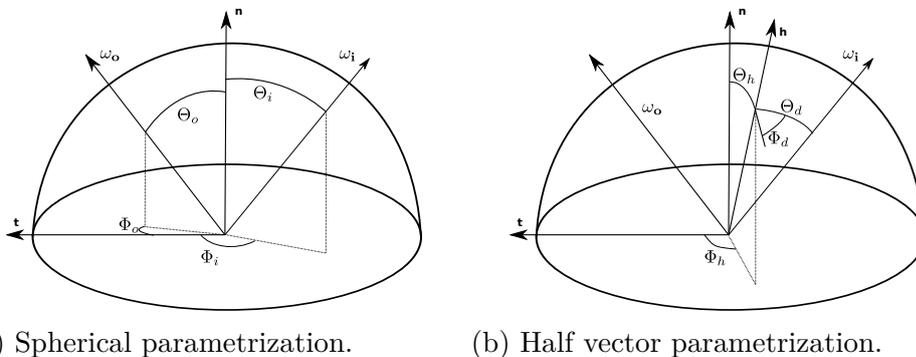


Figure 2.3: The halfway vector parameterization allows for a more efficient computation of the Blinn-Phong model and is especially useful for isotropic BRDFs.

Building up on this parametrization is the Blinn-Phong model where the dot product  $\omega_r \cdot \omega_o$  is replaced by  $\mathbf{h} \cdot \mathbf{n}$ . This approximation is exact as long as  $\omega_o, \omega_i, \omega_r$  and  $\mathbf{n}$  all lie in the same plane. Otherwise we get small deviations from the Phong model. The Blinn-Phong model can hence be written as

$$f_r(\omega_i, \omega_o) = k_d + k_s (\mathbf{n} \cdot \mathbf{h})^m, \quad (2.10)$$

where  $m$  has the same function as  $\alpha$  for the pure Phong model.

From here on many different BRDF models have been proposed, such as the Lommel-Seeliger model[55], which is interesting, because it explains why the Moon does not appear to be a Lambertian reflector. Other more physically correct models build on the assumption of a microfacet model. The surface is modelled by randomly distributed micro reflectors, which act as tiny mirrors, see Figure 2.4. A number of different models use this underlying microfacet model, for example the Torrance-Sparrow[58], Cook-Torrance[58], Ward[63], Oren-Nayar[41]

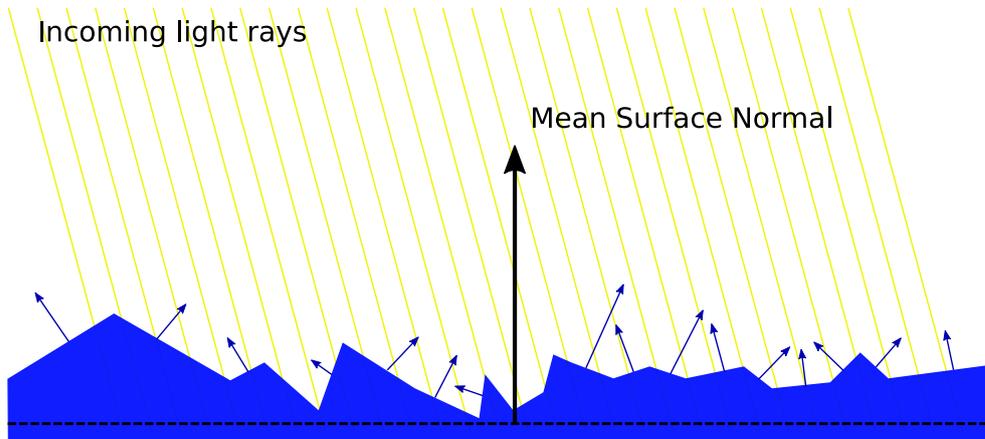


Figure 2.4: The microfacet model describes reflection properties by modeling a surface with randomly oriented microspheres. The underlying probability distribution, as well as the roughness of the surface dominate the appearance of a surface.

and the Ashikhmin-Shirley model[3]. Exemplary the Cook-Torrance model is presented in a little more detail. The BRDF term for this model is given by

$$f_{\text{r}}(\boldsymbol{\omega}_{\text{i}}, \boldsymbol{\omega}_{\text{o}}) = k_{\text{d}} + k_{\text{s}} \frac{D(\mathbf{h})F(\boldsymbol{\omega}_{\text{o}})G(\boldsymbol{\omega}_{\text{o}}, \boldsymbol{\omega}_{\text{i}})}{4(\boldsymbol{\omega}_{\text{o}} \cdot \mathbf{n})(\mathbf{n} \cdot \boldsymbol{\omega}_{\text{i}})}, \quad (2.11)$$

where  $D$  is the microfacet distribution,  $F$  the Fresnel factor and  $G$  the geometric attenuation factor, which expresses the ratio of light which is occluded by other microfacets. The distribution term  $D$  controls how the microfacets are distributed, often a Gaussian distribution function is used. Another common function is the Beckmann distribution[6], given by

$$D(\mathbf{h}) = \frac{\exp(-\tan^2(\alpha)/m^2)}{\pi m^2 \cos^4(\alpha)}, \quad \alpha = \arccos(\mathbf{n} \cdot \mathbf{h}), \quad (2.12)$$

where  $m$  describes the roughness of the material. A larger roughness value leads

to a more diffuse appearance. The Fresnel factor describes how much of the incident light is transmitted or reflected depending on the refraction indices  $n_1, n_2$  of the media at the interface. Additionally, the Fresnel equation takes into account polarization. We do not want to go into the details of the Fresnel equations, but just present the commonly used Schlick[53] approximation

$$F(\theta) = F_0 + (1 - F_0)(1 - \cos \theta)^5 \quad (2.13)$$

$$F_0 = \left( \frac{n_1 - n_2}{n_1 + n_2} \right)^2, \quad (2.14)$$

where  $\theta$  is the angle between the surface normal and the incident light ray.

The geometric attenuation term expresses that neighboring parts can block of some light, that would otherwise reach a microfacet. This can be the case either when the light enters or exits the surface. Hence, it is given as

$$G = \min \left( 1, \frac{2(\mathbf{h} \cdot \mathbf{n})(\boldsymbol{\omega}_o \cdot \mathbf{n})}{\boldsymbol{\omega}_o \cdot \mathbf{h}}, \frac{2(\mathbf{h} \cdot \mathbf{n})(\boldsymbol{\omega}_i \cdot \mathbf{n})}{\boldsymbol{\omega}_o \cdot \mathbf{h}} \right). \quad (2.15)$$

In practice these different models are often used combined to get realistic looking materials in computer graphics.

### 2.1.2 Light Model

Having all ingredients to describe the light interaction at a surface, the question remains how the flow of light in a complex scene can be described. Again, we

will point to ideas commonly used in computer graphics. The governing equation for the light transport is given by the render equation. It describes the amount of radiance emitted from a point in space  $\mathbf{x}$  in direction  $\boldsymbol{\omega}_o$  of an observer. The amount of reflected light depends on the material properties, as discussed earlier. The incident angle of the light with respect to the surface normals is represented by the shading term  $\mathbf{n} \cdot \boldsymbol{\omega}_i$  and is heavily used in techniques such as shape from shading. The full equation is given by

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = L_e(\boldsymbol{\omega}_o) + \int_{\Omega} f_r(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L(\boldsymbol{\omega}_i) (\mathbf{n} \cdot \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i, \quad (2.16)$$

where  $L_e$  is an emission term, *e.g.* for a light source and  $\Omega$  is the half sphere around the surface normal. In practice it is impossible to find an analytical solution to this integral. This is due to the fact, that we need to take into account the incoming light contributions from all directions in the upper half sphere. The calculated radiance can then again contribute to the radiance of other points and depending on the geometry we have a complicated interaction scheme with a recursive definition. In computer graphics this problem can be solved by ray tracing, where light rays are cast from an observer into the scene. Therefore, a Monte Carlo sampling method is used. The inverse process, the recovery of the light model parameters from several images is an even harder task, and mainly the goal of this thesis. But before introducing the techniques for recovering the BRDF and surface normals from multiple views, we will review the foundations of 3D reconstruction.

## 2.2 From Stereo to Light Fields

We will now present the theoretical foundation of 3D reconstruction in light fields. Before going into the details of light field imaging we recapture simple depth estimation via triangulation, which is the basis for stereo methods using only two views.

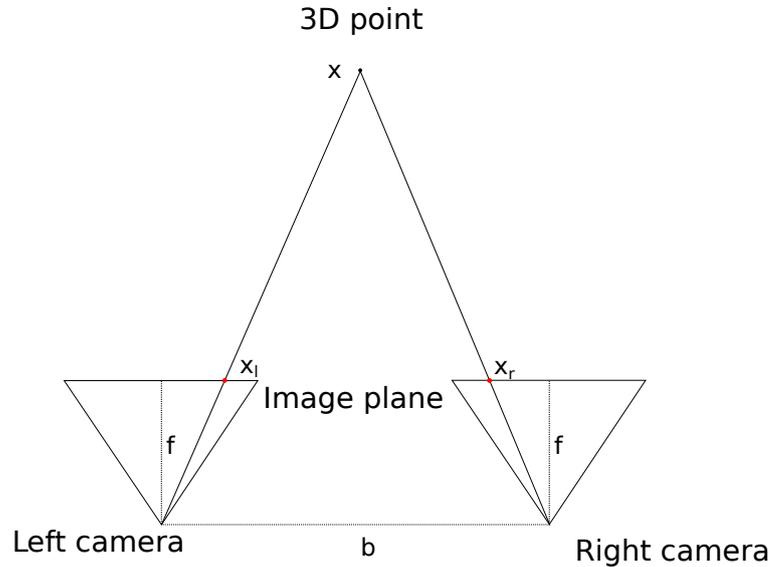


Figure 2.5: Scene geometry for estimating depth by triangulation. The different positions  $x_l$  and  $x_r$  for the projections of the 3D point  $\mathbf{x}$  result in the disparity  $d = x_l - x_r$ . The depth  $z$  can then be inferred by  $z = \frac{bf}{d}$ .

### 2.2.1 Stereo

When we image the same 3D scene from two different view points we have some constraints between these two projections. The underlying mathematical theory is the epipolar geometry. We assume rectified, calibrated images and a pinhole camera model. This means, that all epipolar lines are parallel to both image planes. Given two parallel, identical cameras we can infer the depth of a point  $\mathbf{x}$  using the different positions in the image, under which this object appears, see Figure 2.5. This difference in location on the image sensor is called parallax or disparity  $d$  and relates to the depth  $z$  via

$$z = \frac{b \cdot f}{d}, \tag{2.17}$$

where  $f$  is the focal length of the camera in px and the baseline  $b$  between the two camera projection centres. The greatest challenge is to actually find the same 3D point in both images. To do so, many methods rely on feature points, based on edges or corners. There is a whole family of these features, such as SIFT[34], SURF[5], ORB[47] and many others. In principal, all pixels have to be touched to find these feature points in the other image, which – with increasing image resolution – can be very time consuming, *e.g.*  $\mathcal{O}(nm)$ . Due to the rectification the search space can be constrained to the same epipolar lines, *e.g.* the same image row.

There are three major problems in finding good feature points, or to retrieve them.

**Occlusions** appear when a 3D point is visible in one view, but not the other. For these points it is not possible to infer the depth via triangulation.

**Featureless regions**, such as blank walls, are difficult to handle as well. The lack of a significant feature prevents the correspondence search. Usually, interpolation methods are established to fill these regions.

**Changing brightness** for specular materials from one view to another can in principal be modelled and hence accounted for, but requires more information about the scene. This is the main focus of this thesis.

### 2.2.2 Light Fields

Instead of two views, light fields utilize many views to enhance the disparity estimation, see Figure 2.6. Stacking the same image row from each view on top of another we build a so called epipolar plane image (EPI), see Figure 2.7. The slopes of the resulting lines directly relate to the disparity between the different views.

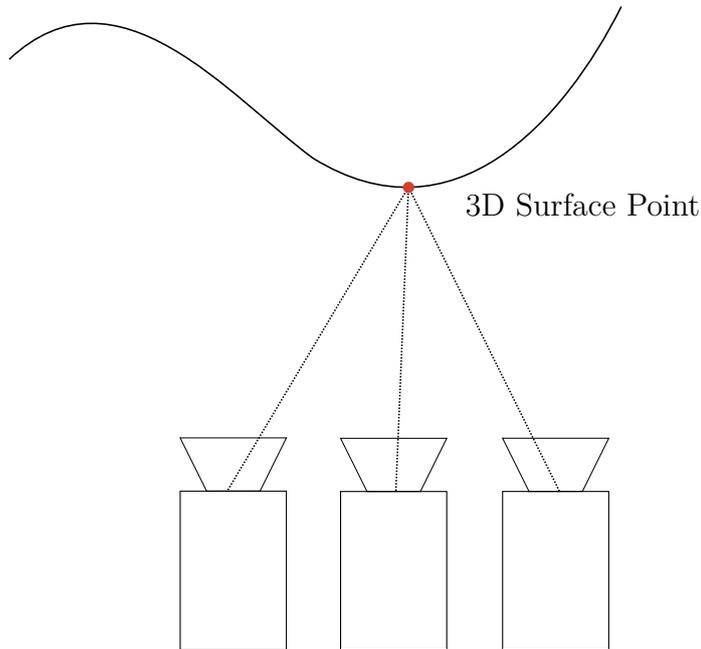


Figure 2.6: To capture a structured light field, we add more cameras. For the ease of computation we keep the baseline between camera pairs constant.

A common parametrization for a light field is the so called Lumigraph [17, 30]. It is defined using two parallel planes  $\Omega$  and  $\Pi$ . The first plane  $\Omega$  addresses coordinates  $(x, y) \in \Omega$  in the image domain. The second plane  $\Pi$  contains the focal points  $(s, t) \in \Pi$  of all cameras. The intensity for each pixel in each view is hence encoded in the light field

$$L : \Omega \times \Pi \rightarrow \mathbb{R} \quad (s, t, x, y) \mapsto L(s, t, x, y). \quad (2.18)$$

To slice out an epipolar plane image (EPI), we fix the image dimension corresponding to the camera direction, *i.e.* for the horizontal direction we fix  $y = y^*$  and  $t = t^*$ . Thus an EPI is defined as

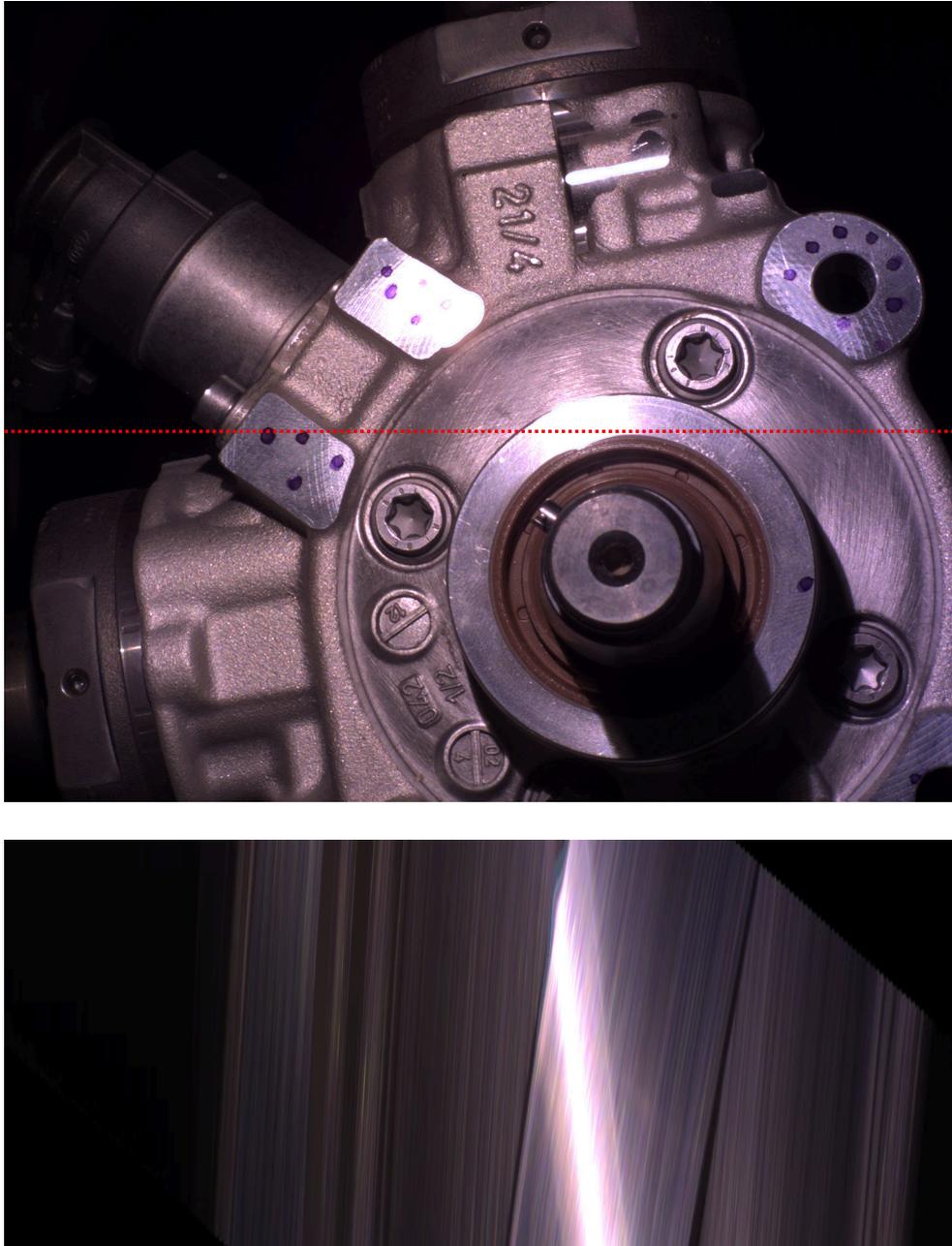


Figure 2.7: Object with a specular highlight. An epipolar plane image was built from different views and the same image row indicated by the red line. The EPI was reshifted for better visibility of the line structures.

$$S_{t^*,y^*}(s, x) := L(s, t^*, x, y^*). \quad (2.19)$$

The orientation in these EPIs directly encodes the disparity and hence the depth.

### 2.2.3 Structure Tensor

A very interesting opportunity arises now. Instead of searching for similar features among the different views, we can directly compute the local orientation in the EPI. One common used operator for orientation estimation is the structure tensor  $S$  which relates a tensor based on the partial image gradients  $\mathbf{I}_x, \mathbf{I}_y$  to each pixel  $p$ , *e.g.*

$$S(p) = \begin{bmatrix} (\mathbf{I}_x(p))^2 & \mathbf{I}_x(p)\mathbf{I}_y(p) \\ \mathbf{I}_x(p)\mathbf{I}_y(p) & (\mathbf{I}_y(p))^2 \end{bmatrix} =: \begin{bmatrix} \mathbf{J}_{xx} & \mathbf{J}_{xy} \\ \mathbf{J}_{xy} & \mathbf{J}_{yy} \end{bmatrix}, \quad (2.20)$$

where the partial gradients are usually a result of a convolution with one of the common edge detection filters, *e.g.* Sobel or Scharr, where the latter provides a better rotational symmetry. The disparity is then given by

$$d = \tan \left( \frac{1}{2} \arctan \left( \frac{2\mathbf{J}_{xy}}{\mathbf{J}_{xx} - \mathbf{J}_{yy}} \right) \right), \quad (2.21)$$

as stated by S. Wanner and B. Goldluecke [49].

Additionally, a confidence measure called the coherence  $c \in [0, 1]$  [8] can be calculated by

$$c = \sqrt{\frac{(\mathbf{J}_{xx} - \mathbf{J}_{yy})^2 + 4(\mathbf{J}_{xy})^2}{(\mathbf{J}_{xx} + \mathbf{J}_{yy})^2}}, \quad (2.22)$$

where  $c = 1$  relates to a sharp orientation in a specific orientation and  $c = 0$  relates to a uniform region where no clear orientation is visible.

### 2.2.4 Hough Transform

Besides, estimating the orientation only locally, also global methods, which take the whole EPI information into account are available. The Hough Transform is a widely known method which maps image features, such as lines or circles, to a parameter space, see Figure 2.8. Lines, circles or other parameterizable shapes become then easily identifiable in the parameter space. However, this comes at a great computational cost, since all possible parameters must be traversed. Additionally, the classical Hough transform suffers from the implicit discretization in the parameter space. To reduce computational overhead and address the discretization issue, random sampling methods are of great use and referred to as probabilistic Hough transform [57, 27].

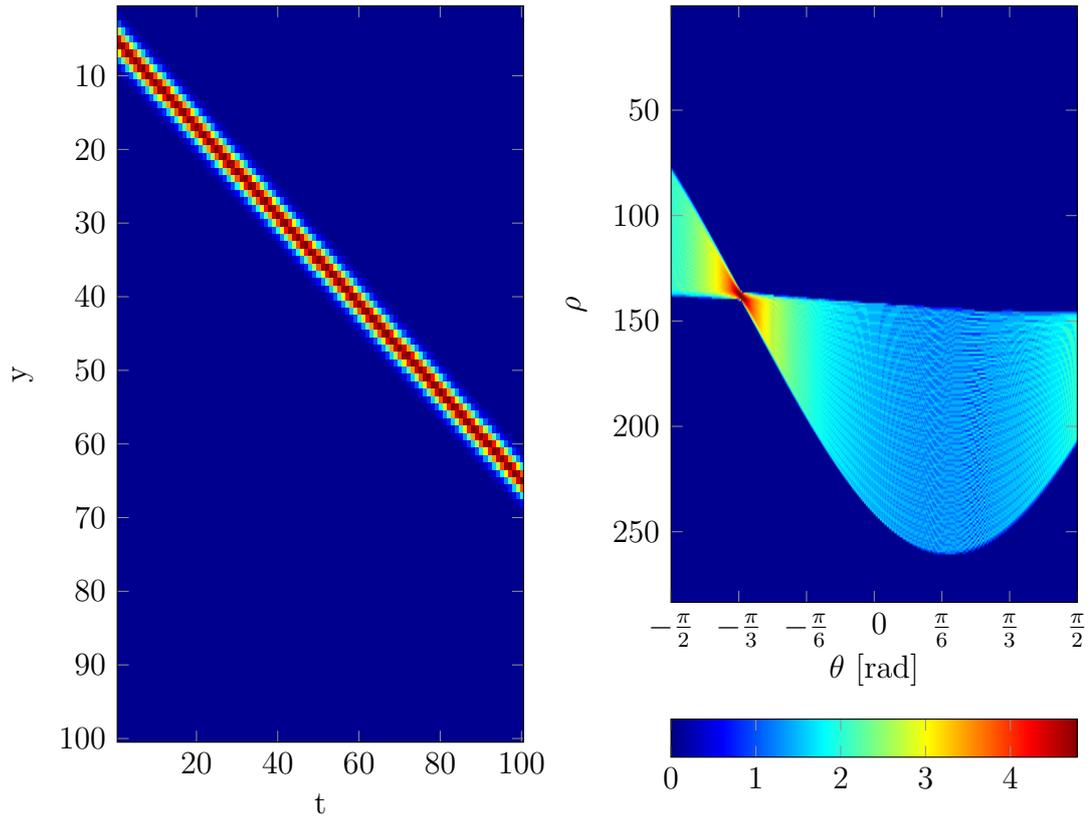


Figure 2.8: Representation of a straight line in Hough space in the  $yt$ -plane. On the left hand side the line in image space is visualized. The right hand side shows the corresponding hough transformation with intensity values on a logarithmic scale. Image taken from [19].

# 3

## Capturing the Light Field

The following chapter deals with the aspects of capturing light field data, especially in combination with specular highlights. Part of it emerged in collaboration with Bosch to develop an understanding of the challenges of light field capture for specular highlights. The first part deals with different hardware setups to capture a light field. The second analyses specific trajectories to probe a sufficient part of the object to reconstruct the BRDF.

Especially for light field capturing a precise camera calibration is necessary. For this thesis we used fractal calibration targets to get robust results, especially at the image edges [51].

### 3.1 Capturing Devices

In the following, different measurement techniques to capture a light field are presented.

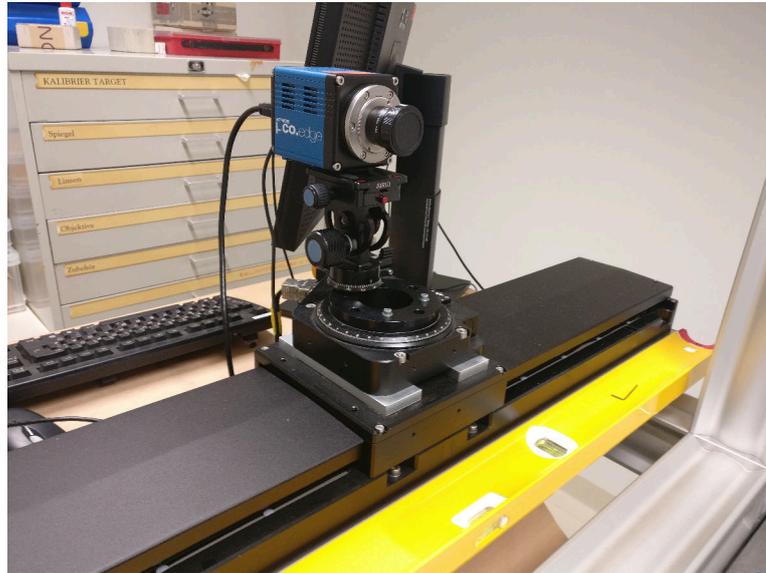


Figure 3.1: Light fields can be captured using a translation stage with a mounted camera. This is a cost-effective setup, but only static scenes can be captured.

#### 3.1.1 Translation Stage

A camera can be mounted on a movable platform, also known as translation stage. The accuracy is in the order of micrometers for translation ranges of up to a meter. An example for a linear translation stage can be seen in Figure 3.1. The need for only a single camera makes it cost efficient, and easier to calibrate since the intrinsic parameters do not change from view to view. Dynamic scenes can not be captured. Special care must be taken to prevent changing illumination, which would negatively affect the analysis of the material properties.

#### 3.1.2 Camera Array

In contrast to a translation stage a camera array consists of multiple cameras which have a fixed relation to each other. This way, it is possible to capture dynamic scenes. Here precise calibration between the cameras is critical. An



Figure 3.2: With a linear camera array in a verged setup also dynamic scenes can be captured.

technical hurdle is the synchronized capture and the storage of the often enormous amounts of data in parallel. A linear camera array is displayed in Figure 3.2, where the cameras are verged to one focal point.

#### 3.1.3 Microlens Arrays

A third option, not used in this thesis, are microlens arrays in front of a single image sensor, see Figure 3.3. Each photon is redirected to a different set of pixels on the sensor depending on its direction. The direction of the light ray can be inferred by the position on the pixel grid. The advantage is the low cost, but it suffers from a small baseline and a greatly reduced effective resolution. Nonetheless, this technique made its way into the consumer market, where the user can profit from certain light field applications, such as refocusing. Precise 3D reconstruction is not achievable, due to the small baseline. Recently ideas have emerged to use transparent photo-detectors based on graphene [40] to capture

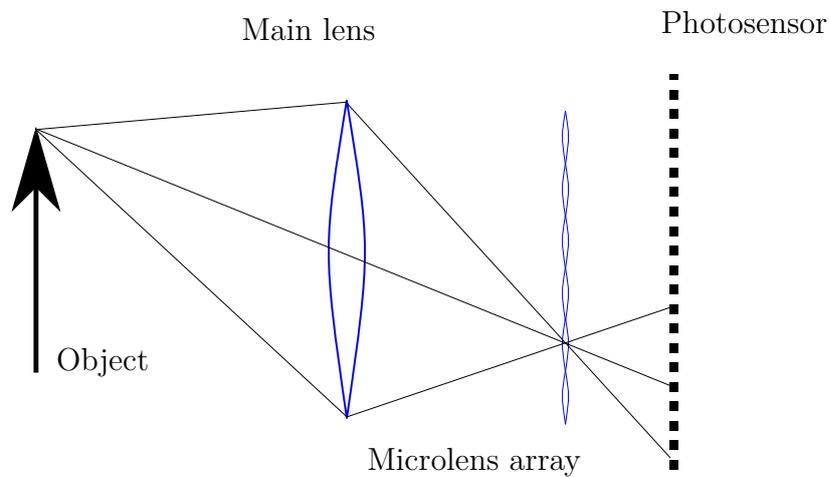


Figure 3.3: Consumer cameras such as the Lytro use a microlens array to capture a light field. Proportions are not to scale.

a light field. This would circumvent the inherent trade-off between image and angular resolutions for single sensor platforms.

## 3.2 Optimal Trajectories through the BRDF Space

An optimal acquisition setup should make it possible to acquire as much information about the surface properties as possible. The diffuse part is not changing much from view to view, so the most interesting changes are happening close to the specular lobe. In the following we will discuss how we can ascertain that a large portion in the light field contains information relevant for BRDF extraction.

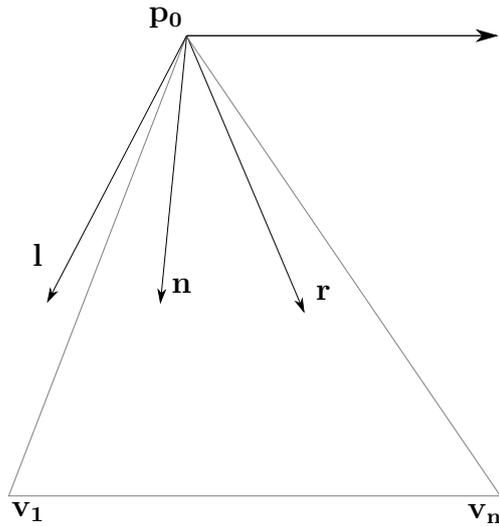


Figure 3.4: The surface normal range which can be recovered by a light field setup is limited by the geometry between cameras  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , object and light source.

### 3.2.1 Geometric Consideration of the Camera Setup

Before we go into details about optimality we want to answer the simple question which surface normal range can be, in principle, recovered from a light field setup. Let us assume a linear light field setup with  $n$  cameras and a baseline  $b$  therefore spanning a distance of  $Ds = nb$ . A point of a specular surface of interest may be at the location  $\mathbf{p}_0$  and a point light source at the position  $\mathbf{L}$ , see Figure 3.4.

Assuming perfect reflection, a light ray  $\mathbf{l}$  will be reflected into the direction  $\mathbf{r}$  according to

$$\mathbf{r} = \mathbf{l} - 2(\mathbf{l} \cdot \mathbf{n})\mathbf{n}, \quad (3.1)$$

where  $\mathbf{n}$  is the surface normal. To capture the maximum of the specular part,  $\mathbf{r}$  must reside within  $\mathbf{v}_1$  and  $\mathbf{v}_n$ . Otherwise, we still can see some part of the lobe, but the exact reconstruction becomes more difficult as we can see in chapter 4.

### 3.2.2 Half-way Vector Parametrization

There are several remarks which highlight the most important implications: We can expect specular reflection if the surface normal and the half vector are aligned ( $\Theta_h = 0$ ). For isotropic BRDFs we have no dependence on  $\Phi_h$ , reducing the BRDF to three dimensions. Viewing 2D slices along the  $\Phi_d$  direction yields, that most information is contained at  $\Phi_d = \frac{\pi}{2}$  and the other slices look like similar versions with some parts excluded. Another interesting characteristic is that the specular reflection varies mostly along the  $\Theta_h$  axis and less so along the other axes. In Figure 3.5 a subspace of BRDF from three different materials is shown.

By capturing light fields with a single moving camera we essentially cut out trajectories from the BRDF subspace. The trajectory depends on the camera path and the position of the light source. A desirable path would move in the 2D BRDF slice along the  $\Theta_h$  dimension with a small and constant  $\Theta_d$ , to guarantee specular highlights.

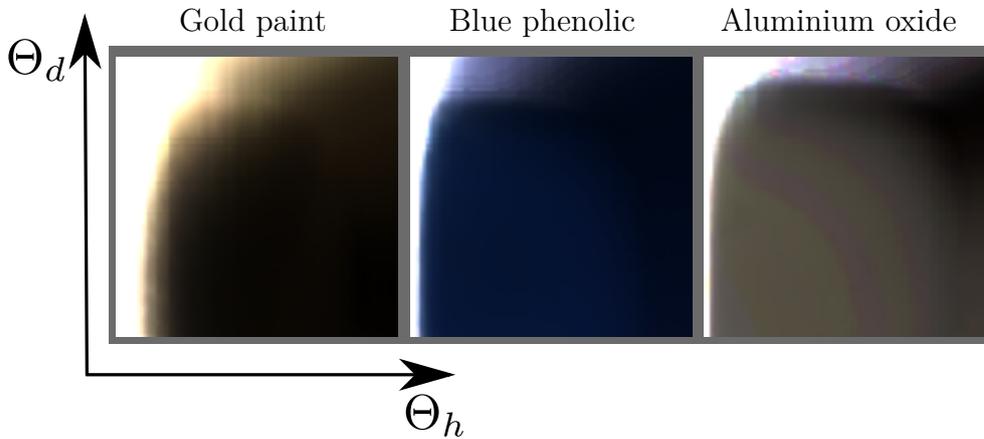


Figure 3.5: Inspecting 2D slices of BRDFs in half-way vector space we can see how the characteristics of different materials differ. For example the fall off rate for aluminum oxide is much larger than for gold paint. The data is scaled logarithmically for better visibility and taken from the MERL database[36].

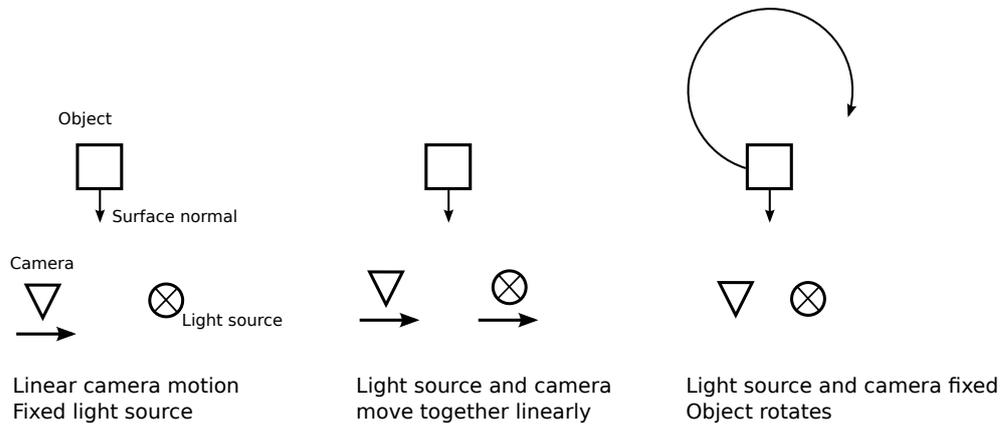


Figure 3.6: We investigate three different acquisition setups for light field data.

### 3.2.3 Linear Camera Motion and Fixed Light Position

To illustrate the importance of knowing the surface normal to estimate the BRDF of a given object we will investigate the trajectories through the 2D BRDF space defined by the two angles in half-way vector parametrization. In Figure 3.7 we see different trajectories based on various normal orientations for an acquisition setup where the camera is moving linearly. If we recall that the specular area of the BRDF runs along the  $\Theta_d$ -axis, we notice that the red trajectory – where the surface normal is slightly out of the plane spanned by the object point, the camera and the light source – does not touch the y-axis, effectively resulting in a reduced peak width in the corresponding 1D-BRDF slice taken from the EPI.

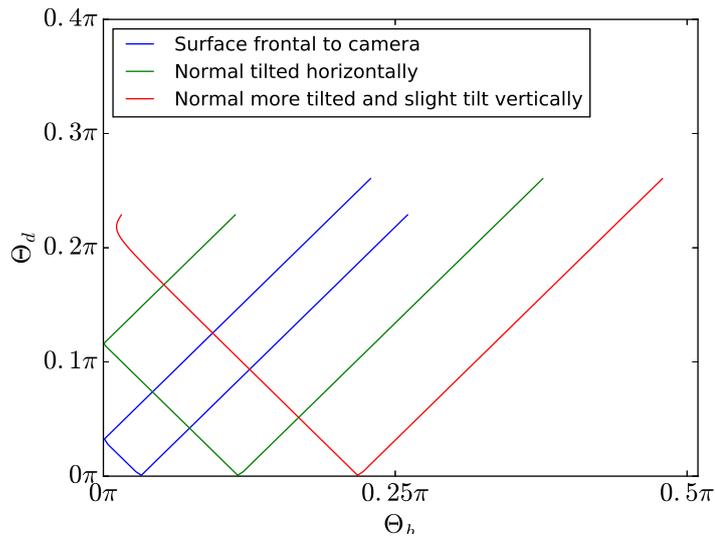


Figure 3.7: Trajectories through the 2D BRDF space for a straight moving camera and a fixed light source. The camera moves in the x direction perpendicular to its viewing direction anti-parallel to the surface normal (in the blue case). The green trajectory describes a path where the surface normal is tilted by  $22^\circ$  around the y-axis (pointing upwards). The red trajectory is tilted by  $45^\circ$  around the y-axis and small amount around the x-axis facing now upwards.

### 3.2.4 Linear Camera Motion and a Co-Moving Light Source

For the case of a moving light source, depicted in Figure 3.8, we see that the trajectory through the 2D BRDF slice crosses a much narrower part along the  $\Theta_d$  axis than in the case of a fixed light source. That the trajectory is still crossing along the  $\Theta_d$  axis stems from the fact, that the distance between the camera and light source pair and the object point is changing and therefore the difference angle increases with decreasing distance to the object point. Basically, the larger the distance between light source and camera, the greater is the coverage in the  $\Theta_d$  direction. This effect could be suppressed entirely using a confocal setup, where the light source and the camera are basically at the same position. This would open up the possibility to use only 1D BRDF slices, since we would set  $\Theta_d$  to zero.

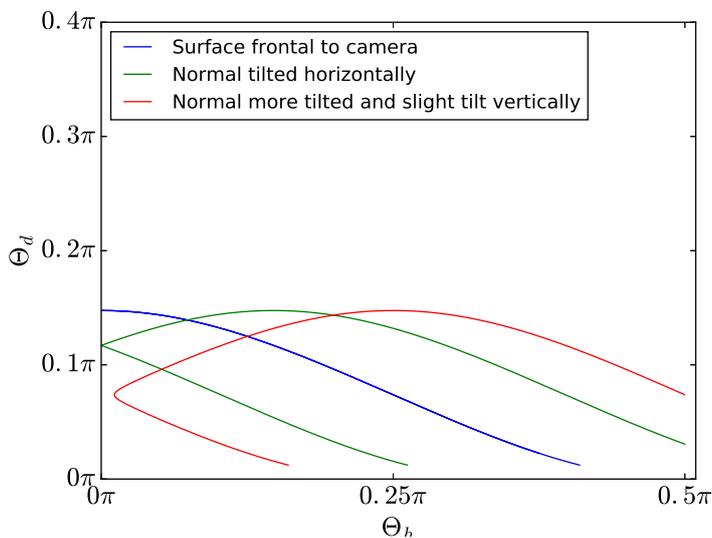


Figure 3.8: Trajectories through the 2D BRDF space for a straight moving camera and a co-moving light source. The different colors encode the same normal displacements as in Figure 3.7.

### 3.2.5 Fixed Camera and Light Source, Rotating Object

In Figure 3.9 we see the BRDF trajectories for an object describing a circular motion in front of the camera. The extension along the  $\Theta_d$  axis is much smaller than previous cases. Again the BRDF trajectory for the slightly vertical tilted normal is not reaching  $\Theta_h = 0$ .

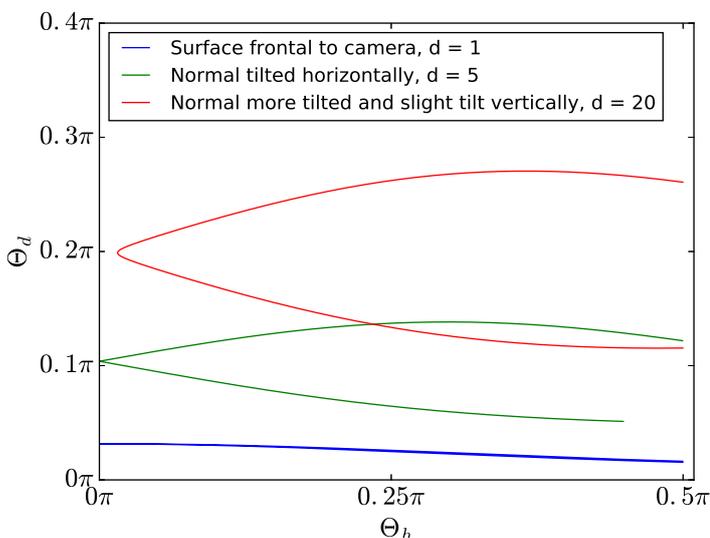


Figure 3.9: Trajectories for circular motion. The distance between camera and light source is given by  $d$ .

### 3.2.6 Optimal Acquisition

To capture all of the BRDF information multiple combinations of light source and camera position must be probed. A possible implementation would involve a circular light field with a fixed light source. To probe along the  $\Theta_d$ -axis, the light position must be varied, e.g. in a circular fashion around the object. Figure 3.10 shows a resulting BRDF coverage by such a setup. It must be noted that surface

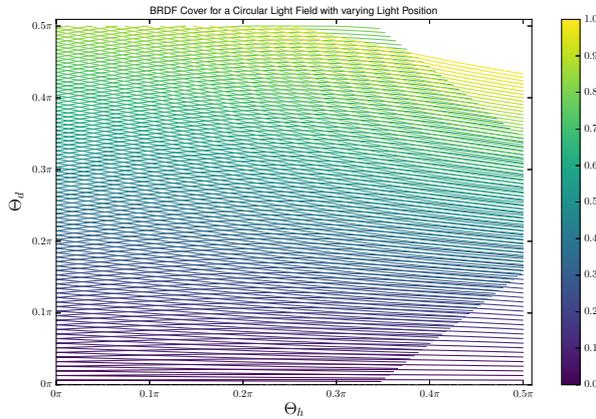


Figure 3.10: BRDF coverage of a circular light field with different light positions. Differently colored lines encode the light position. The light position is also varied in a circular fashion around the object. Zero indicates the start position close to the camera and one the opposite site of the circle. The camera is stationary and the surface normal is in the same plane spanned by the light source and camera.

normals which lie not in the same plane as the camera and the light source will not fully reach the  $\Theta_d$ -axis and therefore have reduced intensities. To address this challenge the object needs to turn around an additional axis. Basically, what we have then is a goniometer. To limit the search area to the high intensity regions of the specular peak and the Fresnel effect, two light positions are sufficient. The first one very close to the camera and the second in opposite direction, to capture the Fresnel contribution.



# 4

## Surface Normal Estimation via Geometrical Optics

In this chapter we will discuss a method to extract surface normals from specular highlights for glossy materials. We assume a given depth and illumination. We make no assumptions about the shape of the BRDF but instead focus on the highlight position.

### 4.1 The Inverse Problem

Reconstructing the geometry, light source and material properties is way more challenging than the forward process, where everything is known and images can easily be rendered. This becomes especially evident when considering how many different combinations of materials, textures, geometries and lighting conditions could lead to the same image. This intrinsic ambiguity is difficult to resolve and human brains have a broad collection of heuristics to deal with it, even if they close one eye, and stay still, and do not defocus.

Using two or more images allows us, with some assumptions, to use triangulation to infer the depth of a scene as explained in section 2.2. The assumption of brightness constancy, which only holds for Lambertian surfaces, is in reality



Figure 4.1: EPI with specular highlight.

often violated. Triangulation entirely fails if a point in 3D space is only visible in one image. This leads to a tricky trade-off, especially in stereo vision, where a large base line leads to higher accuracy but also to more occluded areas.

Here light fields provide a nice best-of-all solution, since they can cover a large base line to give better accuracy but also provide information in between to reduce occlusions. They also offer a possibility to reconstruct non Lambertian surfaces. Specular highlights appear almost everywhere in real life and are due to the microscopic structure of physical objects. As explained in section 2.1, diffuse materials have a random distribution of microfacets around the surface normal and hence appear the same from different vantage points. If this distribution favors a direction, we see a highlight when the viewing direction aligns with the resulting reflectance angle. Here, standard 3D reconstruction algorithms based on the brightness constancy start to break down. A bright pixel in one view appears less intense in another. Without any further assumptions about the reflectance properties the only way to handle this missing correspondence, is to use neighbouring regions where correspondences have been found.

In the EPI, see Figure 4.1, we can observe the gradual intensity change from one view to another.

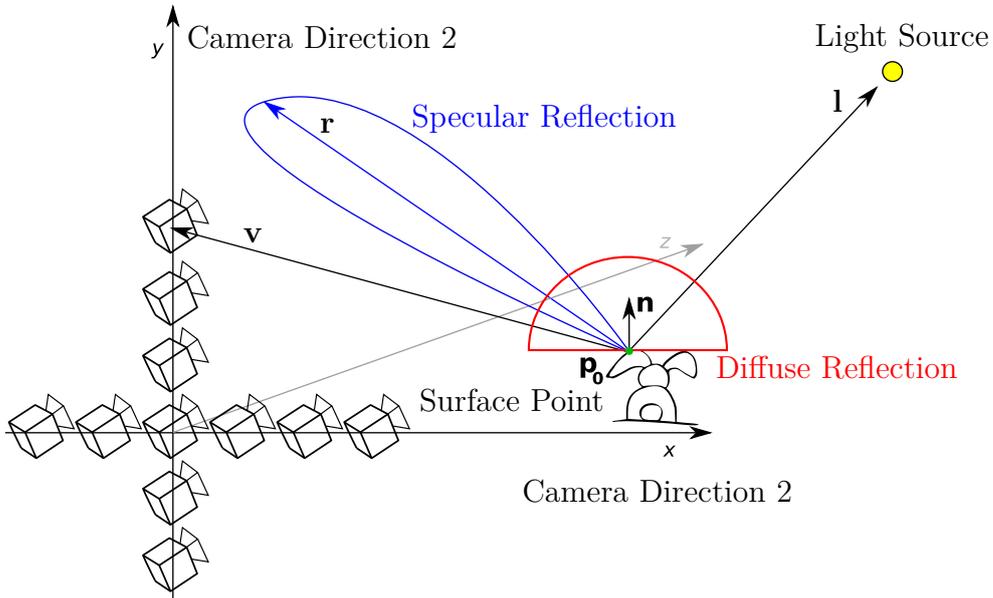


Figure 4.2: Sketch of the cross array setup.  $q_1$  and  $q_2$  are the nearest points to  $r$ . Hence, in the closest view we will see an intensity maximum.

## 4.2 Solving the Inverse Problem by Geometry

The surface normal  $\mathbf{n}$  can be calculated if the direction of the light source  $\mathbf{l}$  and the direction of reflection  $\mathbf{r}$  is known

$$\mathbf{n} = \frac{\mathbf{r} + \mathbf{l}}{\|\mathbf{r} + \mathbf{l}\|}. \quad (4.1)$$

Unfortunately, we do not know neither the direction of reflection  $\mathbf{r}$  nor the surface normal  $\mathbf{n}$ . All we have are our different camera positions  $\omega_o$  and intensity maxima along each direction of our cross setup which we can relate to a 3D point  $p_0$ . We make two assumptions about the reflecting light cone.

1. The intensity is highest in the direction of the reflection.
2. The decrease in intensity away from the reflection ray is rotationally symmetric.

Using these assumptions we can infer from two intensity maximum positions the direction of reflection  $\mathbf{r}$ . We suppose that  $\omega_i$  as well as our 3D surface point  $\mathbf{p}_0$  is known. The camera centres lie on two independent lines with directions  $\mathbf{g}_1 = (1, 0, 0)^T$  and  $\mathbf{g}_2 = (0, 1, 0)^T$  and a support point at  $(0, 0, 0)^T$ .

When observing a specular peak along  $\mathbf{g}_1$  and  $\mathbf{g}_2$ , we eventually spot a maximum in intensity in the view closest to  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , see Figure 4.2. If we presume isotropic reflection, meaning that the intensity lobe has a rotational symmetry around  $\mathbf{r}$ , we can conclude that the points  $\mathbf{q}_1$  and  $\mathbf{q}_2$  have minimal distance to the yet unknown  $\mathbf{r}$ . This leads to an interesting geometric problem, where we want to solve for  $\mathbf{r}$  and hence for the surface normal  $\mathbf{n}$ . This can be solved analytically, but leads to a very long unusable expression (see section A.1 for details). Thus, we formulate this as an optimization problem. For a given reflection vector we expect the specular highlights at certain points in our camera geometry. Hence we minimize the difference of these proposed highlight positions  $\mathbf{q}'_i$  to the measured positions  $\mathbf{q}_i$ . The objective function  $f$  is then given by

$$\operatorname{argmin}_{\mathbf{q}'_i} f = \sum_i (\mathbf{q}_i - \mathbf{q}'_i)^2. \quad (4.2)$$

We do not limit ourselves to only two independent directions, since we can have, at least in theory, arbitrary many. The  $\mathbf{q}'_i$  are calculated by

$$\mathbf{q}'_i = \mathbf{p}_i + \frac{(\mathbf{p}_0 - \mathbf{p}_i) \cdot \mathbf{s}}{\mathbf{g}_i \cdot \mathbf{s}} \mathbf{g}_i \quad (4.3)$$

with

$$\mathbf{s} = \mathbf{r} \times (\mathbf{g}_i \times \mathbf{r}). \quad (4.4)$$

and  $\mathbf{p}_i$  is the support vector for the camera directions which is in general set to zero for all  $i$ . For two independent directions of observers the solution is unique up to the magnitude of the direction vector of  $\mathbf{r}$ .

## 4.3 Results

To quantify the effects of noise in the image signal and the number of cameras the intensity distribution for a single 3D point was simulated. The intensity distribution is generated with a Blinn-Phong model with the parameters  $k_d = 10.0$ ,  $k_s = 10000.0$  and  $m = 3.0$ . The geometry is given by a single 3D point at  $(0, 0, -10)^T$ , a point light source at  $(0, 0, 0)^T$  and a camera cross setup with a distance of 50 between the most left and most right camera, all in arbitrary units. We randomly draw 100 surface normals with an angle up to  $45^\circ$  with respect to the virtual camera ray. In one experiment we take a fixed number of 17 cameras per view direction and increase the relative noise from  $10^{-4}$  up to  $10^2$ , see Figure 4.3. The angular error remains below  $4^\circ$  up to a relative noise of 0.01 and increases then up to  $17^\circ$  for a relative noise of 4.0. This limit is due to the initial range of possible surface normals, where by rare chance it is very unlikely to produce an angle error larger than  $20^\circ$ .

In Figure 4.4 a similar experiment was carried out. Instead of increasing the noise, the noise is kept constant to 0.05 of the intensity signal and the number of cameras used is increased. The decrease of the angular error is to be expected, since errors here result mostly from the insufficient angular sampling of the cameras. By adding more and more cameras the error goes down up to one degree.

In principal we can calculate the surface normal in a simple an efficient fashion, but we face three major practical issues:

1. At least two peaks must be visible. If one of the peaks is outside the visible range we need to extrapolate, which will likely lead to poor results.
2. The result suffers angular discretization depending on the number of views. This can be avoided by using an interpolation method. This directly leads to questions about how to model intensity distributions in light fields and is tackled in section 5.1.
3. Due to the unique solution we can not evaluate the "goodness" of the approximation.

To circumvent all these issues we will derive a method which includes all intensity values from all views by applying an appropriate reflection model. None the less it provides a fast method for deriving the surface normal components if specular highlights are visible.

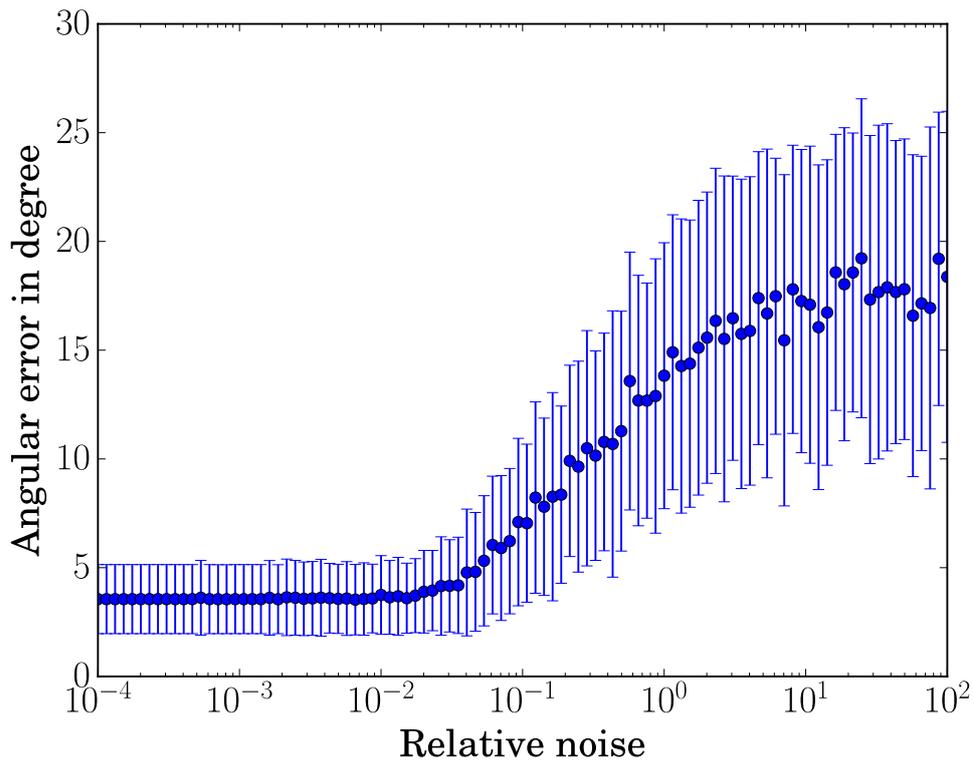


Figure 4.3: Accuracy depending on the relative signal noise. The error-bars represent the standard deviation given by  $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$ . The angular error does not approach zero due to quantization of the highlight position by the fixed number of cameras.

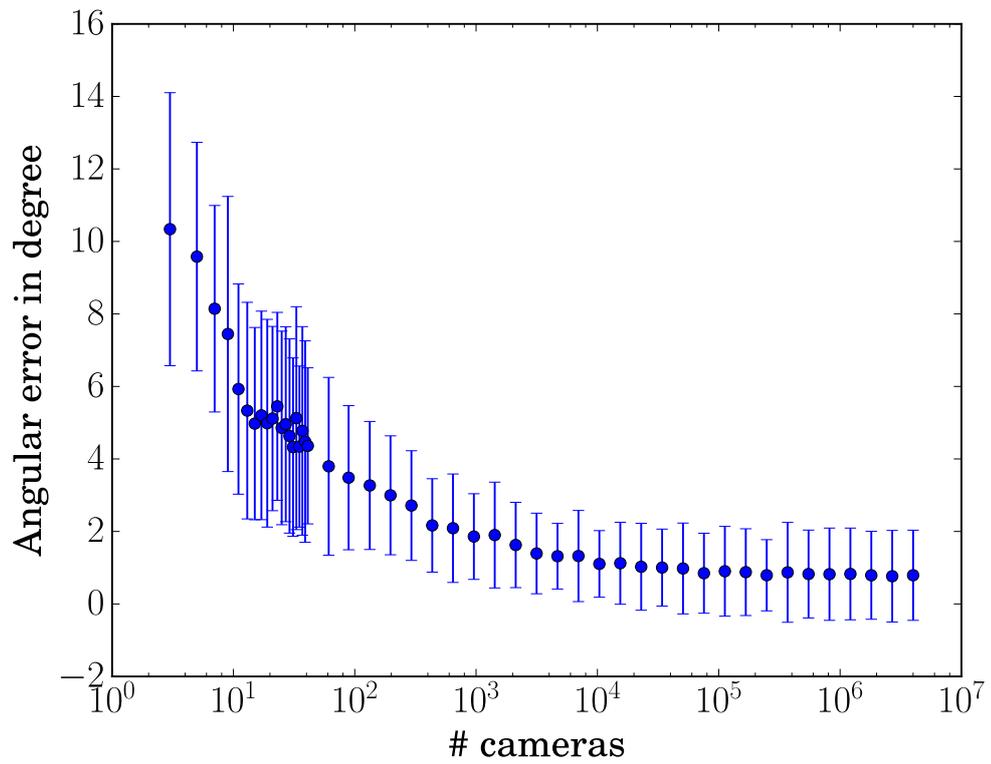


Figure 4.4: Accuracy depending on the number of cameras used.

# 5

## Simultaneous BRDF and Surface Normal Extraction

The following chapter has been published in part in [20]. Additional results are highlighted if they appear in the thesis for the first time.

### 5.1 Solving the Inverse Problem using a BRDF Model

As seen in chapter 4, we can infer surface normals using the maxima of two linearly independent camera directions. While this is already useful if both directions include the reflection lobe, we are at a loss if this is not the case. To infer normals, even when the reflection peak is outside the viewing geometry, we use not only the position of the maximum, but the full intensity distribution. The intensity variation along one viewing direction constrains the surface normal in one dimension. Therefore, we use an orthogonal second direction, which is given by a cross-shaped camera array, to constrain the surface normal in two dimensions. In our approach we use the intensity distribution as observed from both acquisition directions. We can picture the different intensity distributions as cuts through the 3D specular lobe. This means that we can still deliver estimates,

even if only a part of the specular lobe is visible.

Recalling chapter 2, we can model the incoming intensity from a point  $\mathbf{x}$  by

$$L(\mathbf{x}, \omega_{\mathbf{o}}) = L_e(\omega_{\mathbf{o}}) + \int_{\Omega} f_r(\mathbf{n}, \omega_{\mathbf{o}}, \omega_{\mathbf{i}}) L(\omega_{\mathbf{i}}) (\mathbf{n} \cdot \omega_{\mathbf{i}}) d\omega_{\mathbf{i}}. \quad (2.16 \text{ rev.})$$

There are two major challenges in solving this equation. Firstly, we need to integrate over all incoming light directions, and hence need to know the incoming radiation for all points. Secondly, light bounces of each surface, leading to a complicated interaction between active light sources, reflecting surfaces and the geometric relation between them. While the first challenge could be tackled by capturing the light using a dome, multiple light bounces are very difficult to resolve inversely. To simplify the problem we assume single light bounces and dominant light sources. We separate the incoming light dependency into a dominant term  $L_d$  and a perturbation term  $L_p$

$$L = L_d + L_p, \quad (5.1)$$

where we assume, that the perturbation is negligible.

Assuming a discrete number  $N$  of dominant incoming light directions and assuming single light bounces we yield

$$L_d = \sum_{j=1}^N f_r(\mathbf{n}, \omega_{\mathbf{o}}, \omega_{\mathbf{i},j}) L(\omega_{\mathbf{i},j}) (\mathbf{n} \cdot \omega_{\mathbf{i},j}). \quad (5.2)$$

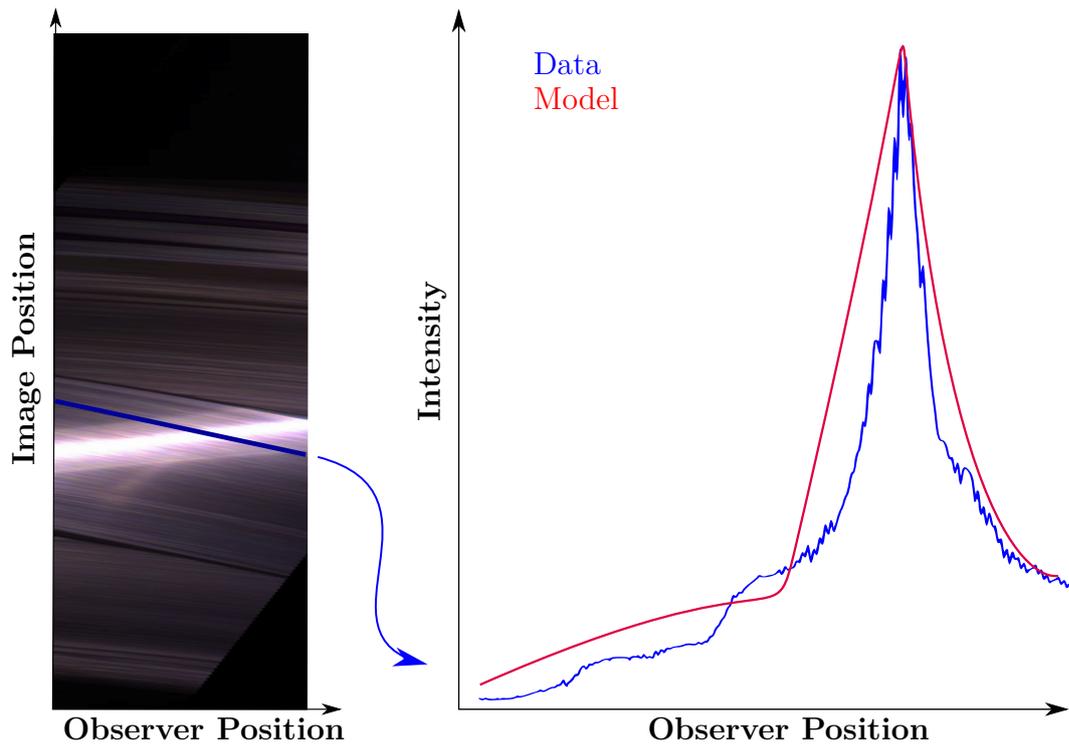


Figure 5.1: An intensity distribution can be obtained using the disparity in the EPI. Given a correct disparity, the intensity distribution depends on the relative orientation of the surface, the material properties and the illumination.

This leads to a nonlinear regression problem in the form of

$$\underset{\Theta, \mathbf{n}}{\operatorname{argmin}} f = \sum_{x=1}^M \|L_d(\Theta, \mathbf{n}, \omega_{\mathbf{o},x}) - I(\omega_{\mathbf{o},x})\|, \quad (5.3)$$

where  $\Theta$  are the BRDF parameters and  $I(\omega_{\mathbf{o},x})$  is the measured intensity for the 3D point in the  $x$ -th view, considering  $M$  different views.

Using traditional solving schemes, such as the Levenberg-Marquardt algorithm, we need to provide parameter initializations which allow for a successful convergence. For the BRDF parameter these depend on the width and relative height of the peak, see Figure 5.1. A good initialization for the surface normal is given by

$$\mathbf{n}_{\text{init}} = \frac{\mathbf{w}_i + \mathbf{w}_{\mathbf{o},x}}{\|\mathbf{w}_i + \mathbf{w}_{\mathbf{o},x}\|}, \quad (5.4)$$

where the  $x$ -th optical system is usually in the center of the cross setup. As can be seen from the form of objective function, minimization can be trapped in a local minimum, see Figure 5.3. To make the convergence more robust, we restart the optimization with random normals, in case the residual is above a certain threshold. The surface normals are very accurate as can be seen in subsection 5.3.3, but we have to keep in mind that we assume the 3D point a priori, which is in general not given. Especially for regions where specular reflection occurs the reconstruction of the geometry is very difficult. So this method can only be applied if the geometry is known beforehand, *e.g.* by using a CAD model of the object of interest. Therefore, the next chapter deals with estimating also the disparity simultaneously.

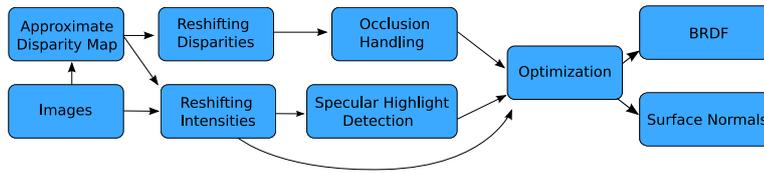


Figure 5.2: Process of simultaneously estimating surface normals and BRDF parameters.

## 5.2 Implementation for Real and Synthetic Light Field Data

The processing pipeline is depicted in Figure 5.2. As input the method requires the calibrated camera and light source positions, as well as approximate depth estimates. To calculate an initial disparity map we use the structure tensor method proposed by Wanner *et al.* [61].

EPIs encode, besides the depth information, also the intensity distribution, along each orientation line, needed to determine material properties.

Regarding the surface points visible in the center view, we use the disparity, to map the input intensities to an image stack  $L'(k, x, y)$  where  $L'_{x^*, y^*}(k)$  represents the intensity distribution for a single surface point as observed from different views, see subsection 5.2.1. A second stack is generated accordingly using the disparity as input data. This way, occlusion maps can be computed easily, as detailed in subsection 5.2.2. They prevent the algorithm from mixing foreground and background information. By regarding the unoccluded intensity values, we can identify surface points which exhibit specular characteristics by measuring the intensity change along the views, see subsection 5.2.3.

Finally, we optimize the BRDF parameters and the surface normal independently for each pixel, for the regions where we identified specular intensity variations, see subsection 5.2.5. Our method works on the following assumptions:

- We assume an approximate disparity map, in our implementation we utilize the structure tensor.
- The cameras, as well as the light source is calibrated in terms of location and camera intrinsics.
- The light transport is dominated by the single-bounce reflection of a point light source.
- At least two independent viewing directions are needed, *e.g.* by using a cross setup.

### 5.2.1 Preprocessing

To compute the intensity changes of object points in an efficient manner, we need to address all pixels related to a specific surface point. This principle is similar to the resifting in EPI processing as introduced by Diebold and Goldluecke [12]. The specific shifting is given for the horizontal light field by s

$$I'(k, x, y) = I(k, x + d \cdot (c - k), y), \quad (5.5)$$

which maps all related pixels to a vertical line, which is easy to access for further processing, see Figure 5.4. The variable  $c$  corresponds to the index of the center view and  $d$  to the disparity of the center view at  $(x, y)$ . The vertical viewing direction is handled analogously. It is important to note here, that the method is especially well suited to smooth, texture-less regions. Here, erroneous disparities have a low impact on the accuracy of the normal estimation, since even if intensity information from neighbouring surface points are mistakenly used, the material properties change slowly compared to the error in the disparity. These areas are particularly difficult for conventional methods, which rely on structural information, such as strong image gradients. This way we can utilize even ap-

## 5.2 Implementation for Real and Synthetic Light Field Data

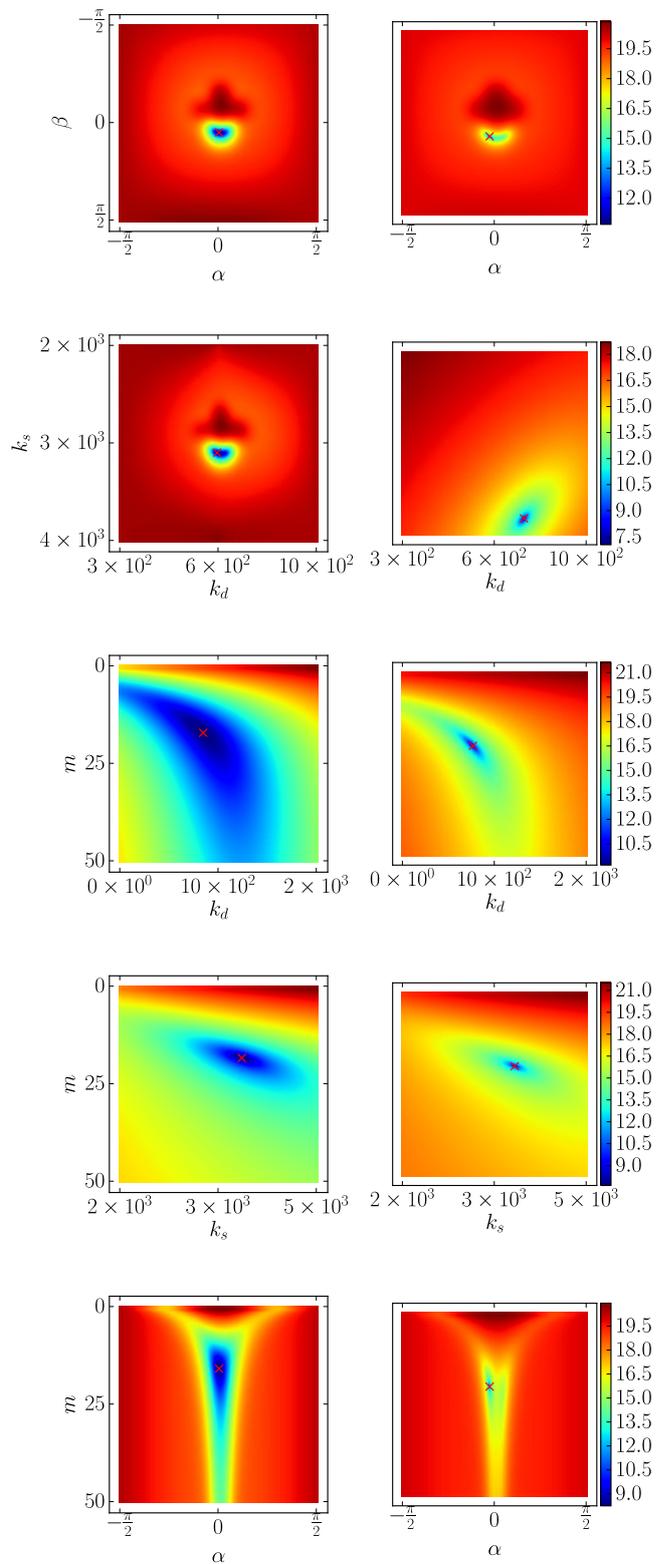


Figure 5.3: The residual of the objective function as heat maps for different pairs of variables. On the left side all variables are initialised according to a best guess. On the right hand side we see the residual after optimisation. The minimum is highlighted with a red cross. If we compare the left and right hand side we see that there are possibilities for getting stuck in a local minimum.

proximate disparities. In the case of strong material changes, we expect precise disparity measurements of the structure tensor, and expect that these methods complement each other particularly well.

### 5.2.2 Occlusion Handling

To get useful intensity variations from the surface points along  $I'(k)$ , occlusions need to be treated carefully. To this end we handle these explicitly. Due to the dense information provided by the light field this can be done by examining the approximate disparity maps which are denoted by  $d(k, x, y)$ . The disparity maps are shifted in the same way as image intensities to calculate occlusion boundaries. Given

$$d'(k, x, y) = d(k, x + d \cdot (c - k), y), \quad (5.6)$$

we can identify occlusions easily by looking for strong changes along  $d'(k)$ . To this end we apply a Gaussian derivative filter  $g$  on  $d'(k)$  and threshold it depending on the noisiness of our input disparity. Our occlusion map  $\Gamma$  is then given by

$$a(x, y) = \frac{\partial}{\partial k}(g * d'(k, x, y)) \quad (5.7)$$

$$\Gamma(k, x, y) = \begin{cases} 1, & \text{if } a \geq t_o \\ 0, & \text{otherwise} \end{cases}, \quad (5.8)$$

where  $t_o$  is a threshold depending on the noise ratio in the disparity map. This way we can incorporate even multiple occlusions in our framework.

### 5.2.3 Specular Highlight Detection

Since we can only extract reliable information from regions with specular highlights, see subsection 5.2.4, we need to decide for which regions of the image we should carry out our analysis. To this end we test if the intensity variation along  $I'(k)$  is above a certain threshold  $t_s$  which is calculated analogous to Equation 5.7. Then, we only apply our algorithm at locations where our confidence mask  $\chi$  is one, see Figure 5.14 c), where the masked regions are white.

$$a(u, v) = \sum_k \frac{\partial}{\partial k} (g * I'(u, v, k)) \quad (5.9)$$

$$\chi(u, v) = \begin{cases} 1, & \text{if } a(u, v) \geq t_o \\ 0, & \text{otherwise} \end{cases} \quad (5.10)$$

### 5.2.4 Joint Estimation of Surface Normals and BRDF Parameters

Given the visible maximum of the intensity of one surface point, as observed by a linear camera array, we can constrain the surface normal by the plane spanned by the object point and the different cameras. Given a second maximum by a linearly independent camera array we tighten that constraint to get the complete 2D surface normal. An analytical solution for this problem exists, but has the disadvantage, that both highlight maxima must be visible. To circumvent this restriction and to leverage the information contained in the light field, we optimize the surface normals of one surface point jointly with the BRDF parameters

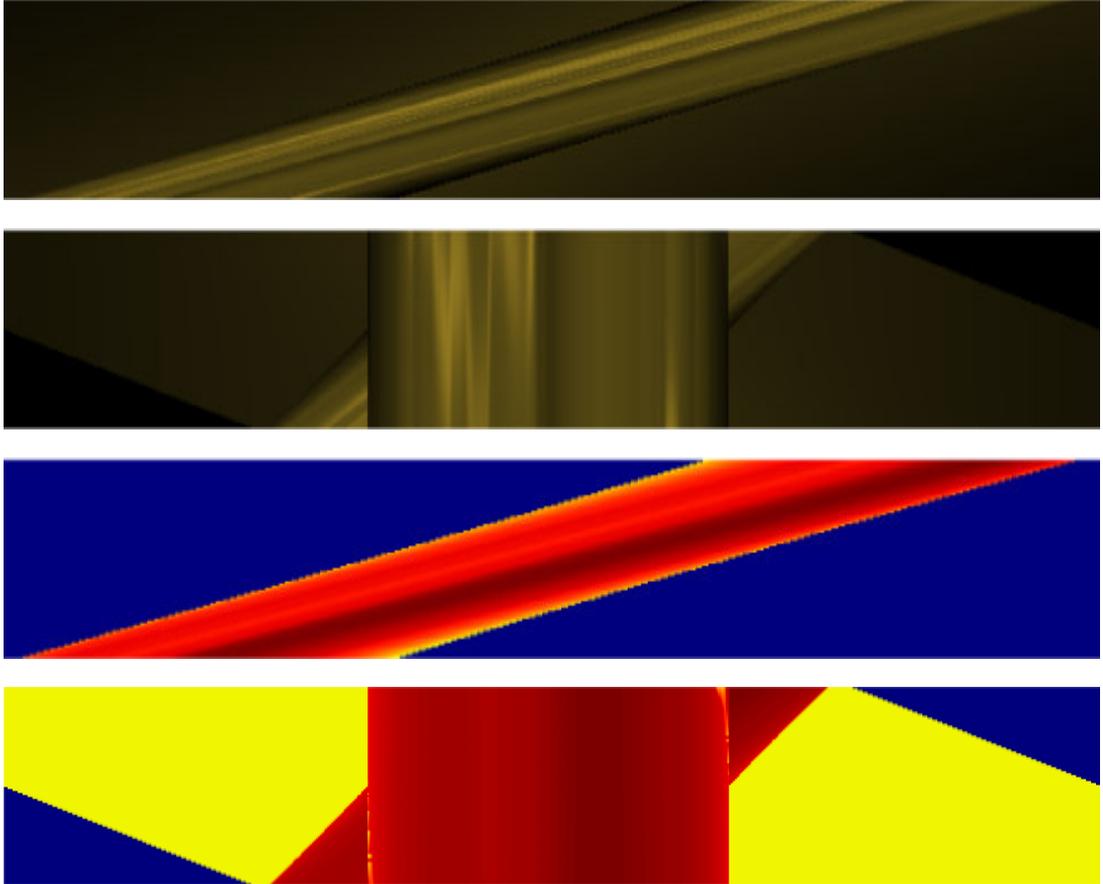
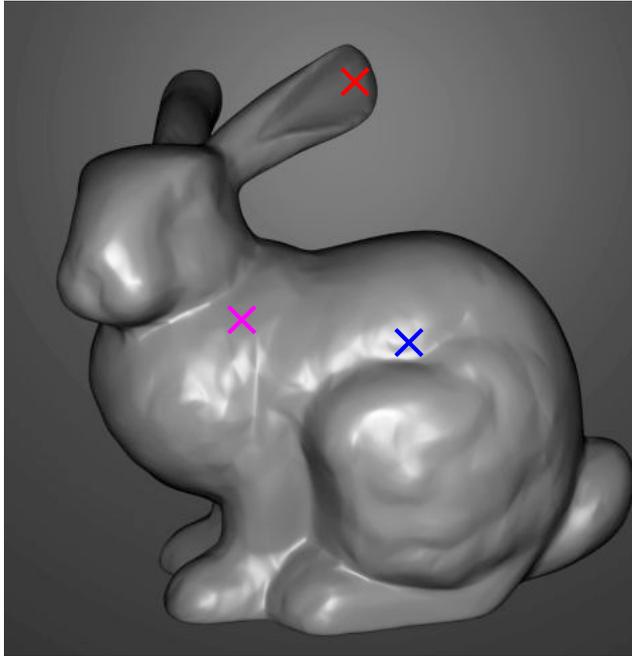
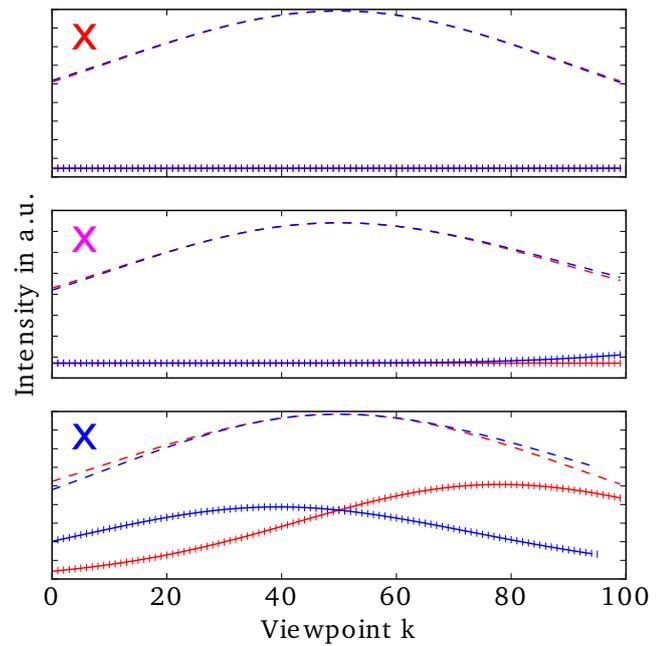


Figure 5.4: a) The orientation in the horizontal EPI directly relates to the disparity of the input image. b) Reshifting the EPI by the extracted disparity allows to investigate the intensity distribution by moving along the views which are here along the y-direction. c) Similarly the disparity for all views can be viewed in a "disparity EPI". d) Reshifting it leads to a straight forward computation of occlusions by looking at gradients in the y-direction.



(a) Center view of the Stanford Bunny. The markers indicate the positions of the corresponding intensity distributions depicted in Figure 5.5b. The red marker identifies a region with Lambertian reflection, the purple one with specular components only in the vertical direction and in the last one both intensity maxima are clearly visible.



(b) Intensity distribution for different surface normals. Red corresponds to the horizontal camera array and blue to the vertical. The dashed lines indicate the model initialization, and the solid line the recovered model. From top to bottom: No intensity variation visible – the Lambertian case (marked red in Figure 5.5a). Variation along the vertical direction visible (purple). Both peaks visible (blue). The corresponding angular errors for the normal reconstruction at these points are  $54.6^\circ$ ,  $4.2^\circ$  and  $0.5^\circ$ .

Figure 5.5: Overview of different intensity distributions.

taken into account all observations. This way we can take advantage of even slight intensity variations from the tails of the specular highlights and do not require the observation of the specular peak in both viewing directions. Of course some part of the specular region must still be visible. In essence there are three different situations, see Figure 5.5b:

1. Along both viewing directions no intensity changes are visible. Thus, we cannot estimate the surface normal because we only see the diffuse part of our surface point. Since we have no specularities, while our method can not compute normals in this case. However, conventional methods can be applied in these cases without problems.
2. One viewing direction exhibits some intensity variation. We can say that the surface normal must be on one curve constraint by our model and can exclude all normals which would lead to an intensity variation in the second viewing direction. So basically, the wider the peak the lesser the surface normal error.
3. Both viewing directions exhibit some intensity variation. In theory, the surface normal is uniquely identifiable.

### 5.2.5 Initialization and Optimization

It is very important to choose a reasonable initial guess to ensure convergence to the global optimum. We initialize the surface normals by the half-way vector between the light source and the central camera. If the surface normal coincides with the half-way vector we have the largest possible BRDF peak, compare Figure 5.5b. The BRDF parameter  $k_d$  and  $k_s$  are initialized to match very roughly

the given intensity distribution by calculating

$$k_s = \max(I'(k)), \quad (5.11)$$

$$k_d = \min(I'(k)). \quad (5.12)$$

To avoid ambiguous solutions we remove the common constraint of  $k_s + k_d \leq 1$  and set the amount of incoming light  $L_0$  equal to one. So the outgoing light will be completely determined by  $k_s$  and  $k_d$ . For each pixel we have as many measurements  $\mathbf{y}_i$  as viewpoints  $k$ . Thus, we minimize

$$\arg \min_{\mathbf{n}, k_d, k_s, m} \sum_i \|\mathbf{y}_i - L_i(\mathbf{v})\|, \quad (5.13)$$

where  $i$  is the index variable for each view, where we have no occlusion with respect to the center view. To ease the solving procedure the surface normal is parametrized in spherical coordinates  $\theta$  and  $\phi$ . We first keep the BRDF parameter constant and vary only the surface normals. In a second step we vary the BRDF parameters and the surface normal jointly. This way we force the solver to explain a constant intensity distribution by moving the surface normal in contrast to simply reducing  $k_s$ . This uses information where we have no intensity variation and can exclude a range of surface normals which otherwise would create a visible highlight.

## 5.3 Results

The resulting intensity distributions for each surface point obtained by the light-field acquisition has sufficient information to achieve high-accuracy surface normals and determine BRDF parameters.

### 5.3.1 Metrics for Surface Normal Analysis

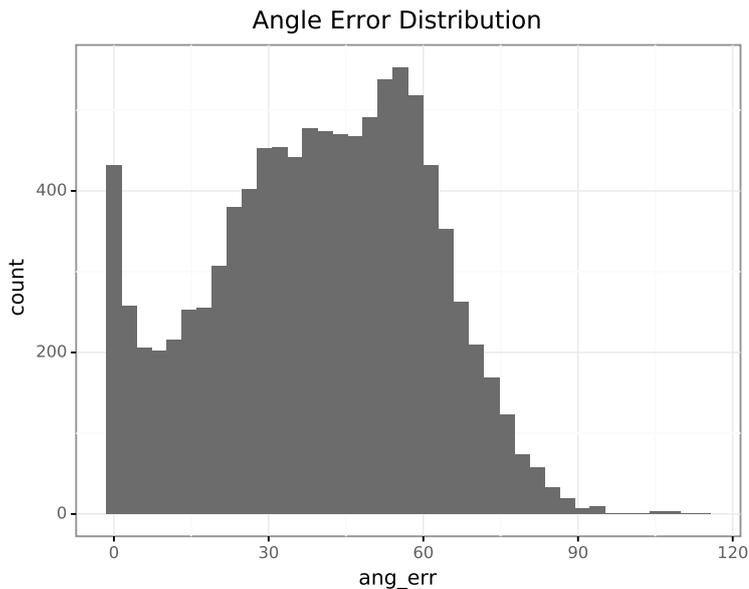


Figure 5.6: Distribution of the angular error in degree before filtering out diffuse regions. The algorithm fails for diffuse regions, where no intensity variation is present. Hence we will suppress the activation on those regions.

To quantify the reliability of the surface normal estimation described in section 5.1 we produced ten thousand random normals with random BRDF parameters. We simulate the light reflection process captured by a cross camera array and an additional relative noise of 0.02 to the resulting intensities. The surface normals can point up to an angle of  $60^\circ$  away from the center camera. This way we ensure that we also take into account that a specular surface appears diffuse if the inclination angle is relatively large in comparison to the bulb of the highlight. In Figure 5.6 we see the distribution of the angular error in the prediction. In the distribution are 9267 observations with an angular error of  $5^\circ$  or more. There are two main error sources:

- 
1. The signal consists only of the non-varying part of the intensity distribution.
  2. The regression is stuck in a local minimum.

To differentiate these two cases we look at the residual and at the relative intensity difference  $r$  which we define as

$$r = \frac{I_{\max} - I_{\min}}{I_{\max}}, \quad (5.14)$$

with  $r \in [0, 1]$ . A relative intensity difference close to zero indicates the first case, namely that we are probing a seemingly diffuse part. In Figure 5.7 we see the color coded angular error depending on the residual and the relative intensity difference. We see that a number of observations with low residual have a high error but a low relative intensity difference. Combining both metrics allows for efficient classification of correctly determined surface normals. In this artificial case we would neglect all observations with a residual larger than 0.1 and a relative intensity difference lower than 0.2. Applying these thresholds leads to the angular error distribution depicted in Figure 5.8 consisting of 180 observations. In general these thresholds need to be adjusted depending on the noise prevalent in the data.

### 5.3.2 Comparison of Fresnel Effects for Different Materials

Figure 5.9 shows the BRDF along the  $\Theta_h$  axis for different materials (mostly which exhibits specular features). Metallic materials exhibit similar BRDFs for both investigated cases ( $\Theta_d = 0$  and  $\Theta_d = 90^\circ$ ), but differ greatly in the amount of reflected light. In the case where the Fresnel effect plays an additional role the distinction between different materials is more evident since the absolute values

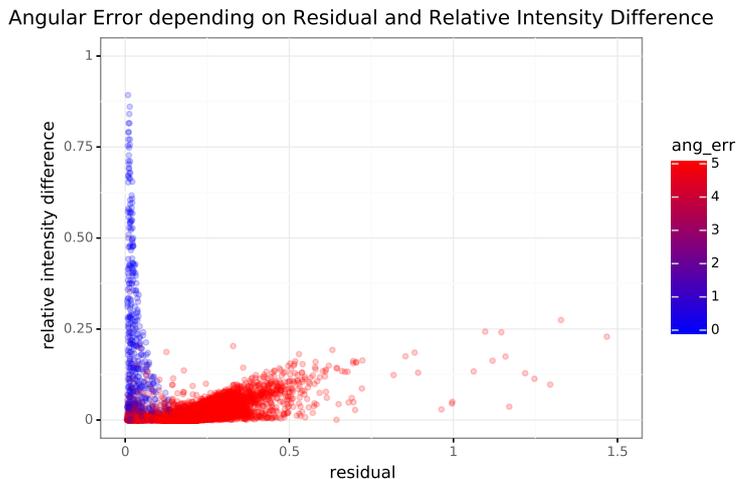


Figure 5.7: Angular error in degree depending on the residual and the relative intensity difference. The relative intensity difference is defined as  $r = \frac{I_{\max} - I_{\min}}{I_{\max}}$ . We see that we can effectively filter out regions where the residual is low, while the intensity difference is only reliable if  $r > 0.25$ .

of the intensities are several orders of magnitude higher.

### 5.3.3 Synthetic Evaluation

For the evaluation of our algorithm we used rendered images obtained with Blender [9]. Our cross array consists of 101 cameras in horizontal as well as in vertical direction. In Figure 5.11 we show how the angular error decreases with the number of views. To evaluate the error in the BRDF we use the following metric to quantify a relative error:

$$e = \sum_i \sqrt{\left(\frac{p_{i,\text{real}} - p_{i,\text{est}}}{p_{i,\text{real}}}\right)^2}, \quad (5.15)$$

where  $p_i$  stands for the  $i$ -th BRDF parameter.

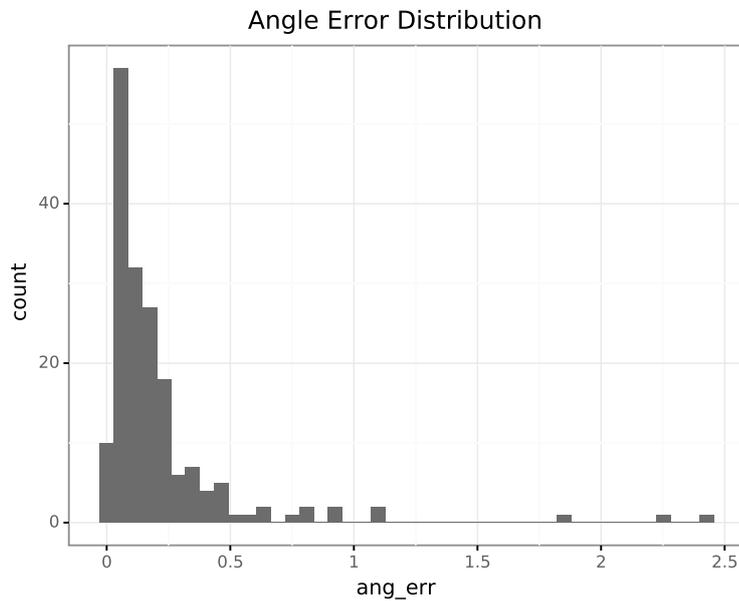


Figure 5.8: Distribution of the angular error after excluding diffuse areas and non converged optimizations.

The evaluation of the synthetic data shows that many surface normals have an angular error of less than  $1^\circ$  and almost all areas have an angular error of less than  $5^\circ$  (Figure 5.14). Surface normals where the intensity variation is not above a certain threshold, were not calculated.

In Figure 5.12 we used a constant disparity for each view to simulate unknown depth. We keep the disparity map constant at roughly the average distance of the bunny. Thus, we have a greater impact on the accuracy in the background, where the relative disparity error is larger. Despite the introduced inaccuracies the angular error is rather small (still around  $1^\circ$ ) for regions with strong specular highlights. Due to the constant disparity, we could not create a sensible specular mask, so we omitted the mask completely for the inaccurate case.

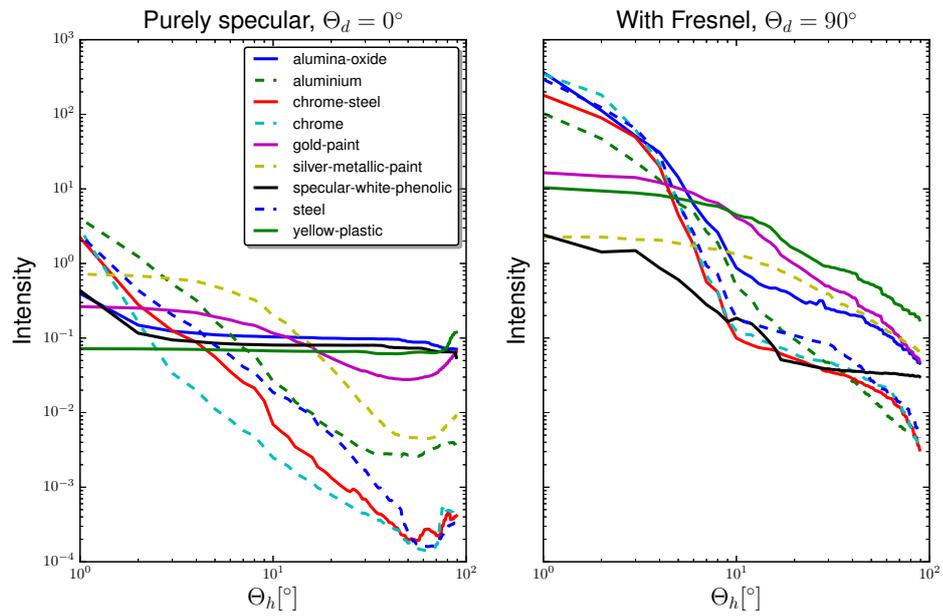


Figure 5.9: Comparison of BRDFs of different materials from the MERL database. On the left hand side only the purely specular component is present. On the right hand side Fresnel effects become apparent. For better visibility both axes are scaled logarithmically.

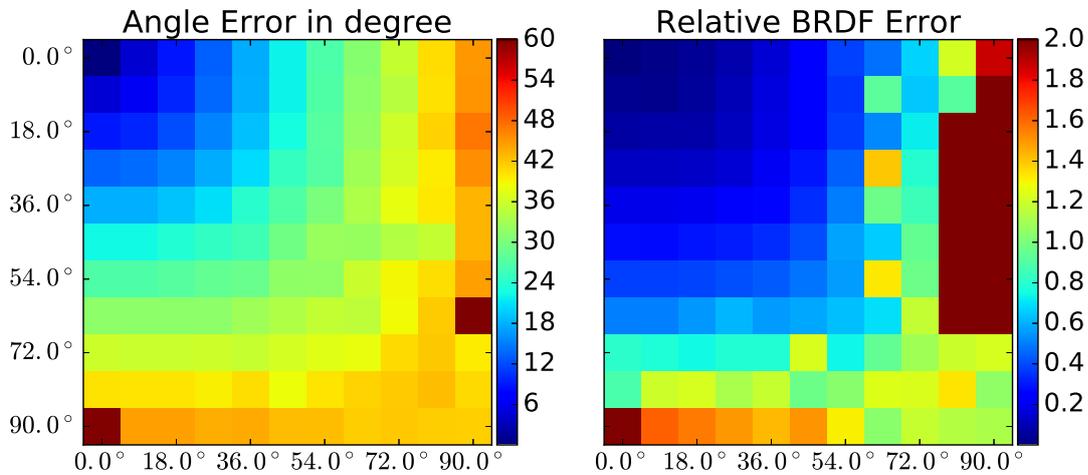


Figure 5.10: Analysis of the angular error and the relative error in the BRDF parameters, dependent on the angle error of the light direction. We rotated the light direction around the x and the y-axis up to 90°. We took 10 different samples and took the median for each different light orientation.

### 5.3.4 Evaluation on Real World Objects

To examine the feasibility of our concept for real world data we used a strongly specular coin (see Figure 5.15). To capture the data we use a gantry with two orthogonal axes to capture images in a cross setup. We used 101 images for the horizontal and for vertical axis respectively. The surface normals reveal slight scratches. The intensity distribution together with the fit shows, that the Blinn-Phong model is not well suited to fully explain the observed intensities.

So far our method only makes use of local information. Introducing smoothness constraints should improve the estimates. Towards this end, it would be reasonable to integrate this BRDF based normal estimation method with a conventional depth estimation technique, within a unified framework which jointly optimizes depth, BRDF parameters and surface normals. This is done and will be explained in chapter 6.

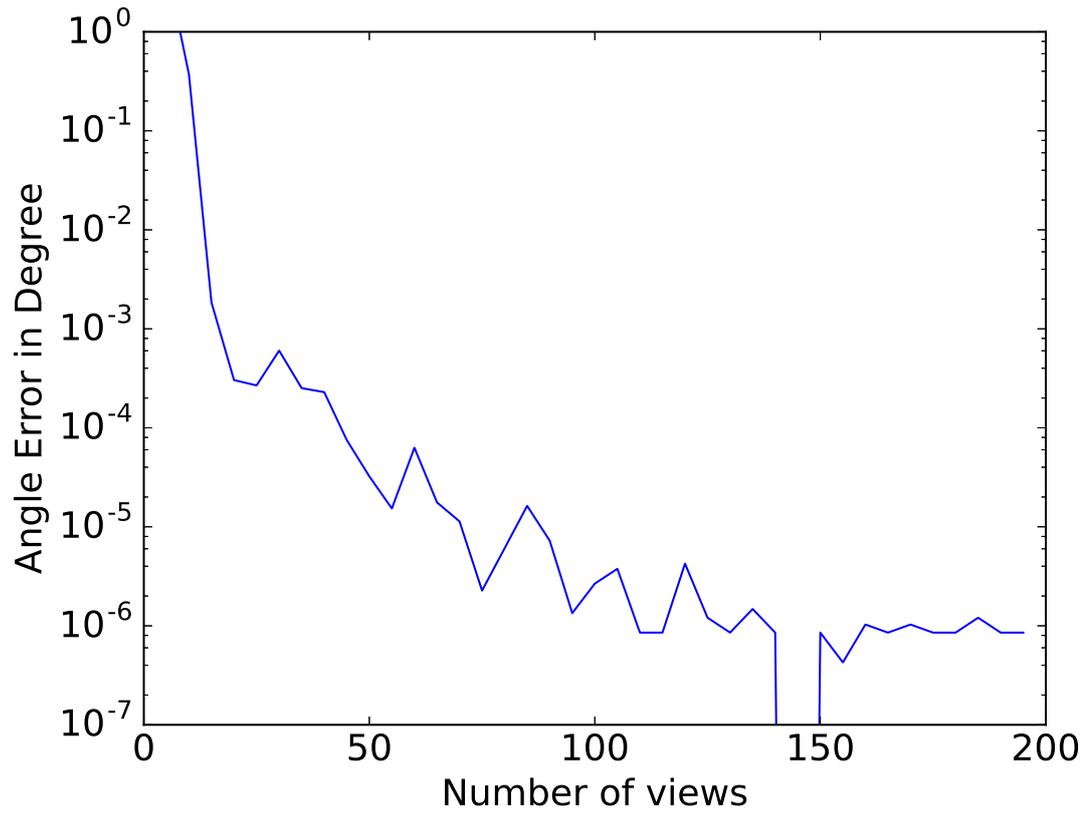


Figure 5.11: Analysis of the angular error, dependent on the number of views. The precisions increases remarkably up to 30 views. With more than 150 views discretization errors play a significant role and we can't increase the precision anymore.

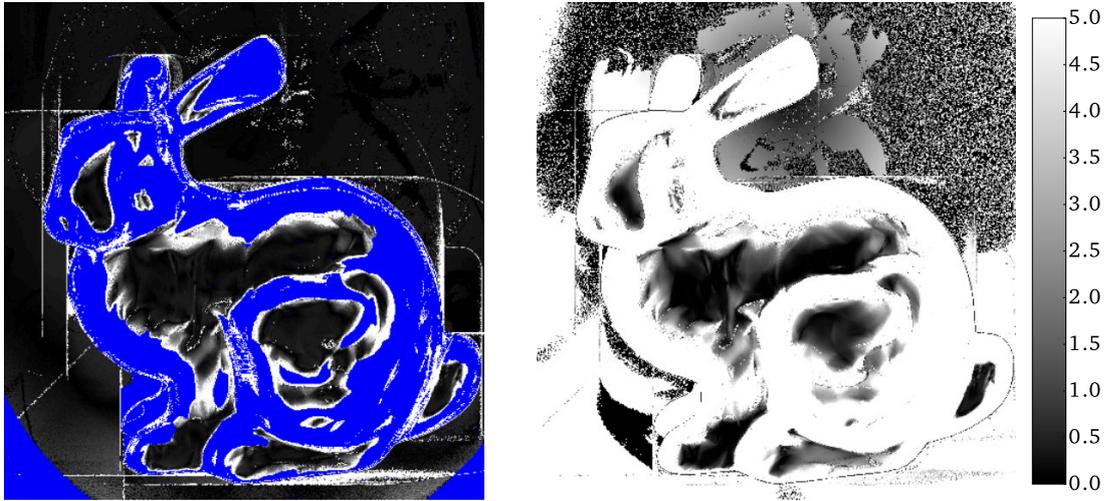


Figure 5.12: Comparison of angular errors for the reconstruction with correct disparity maps (left) and with a constant disparity roughly in the plane of the bunny.

This means that the framework can be used in a wide range of applications to improve the geometry reconstruction in the presence of specular reflections.

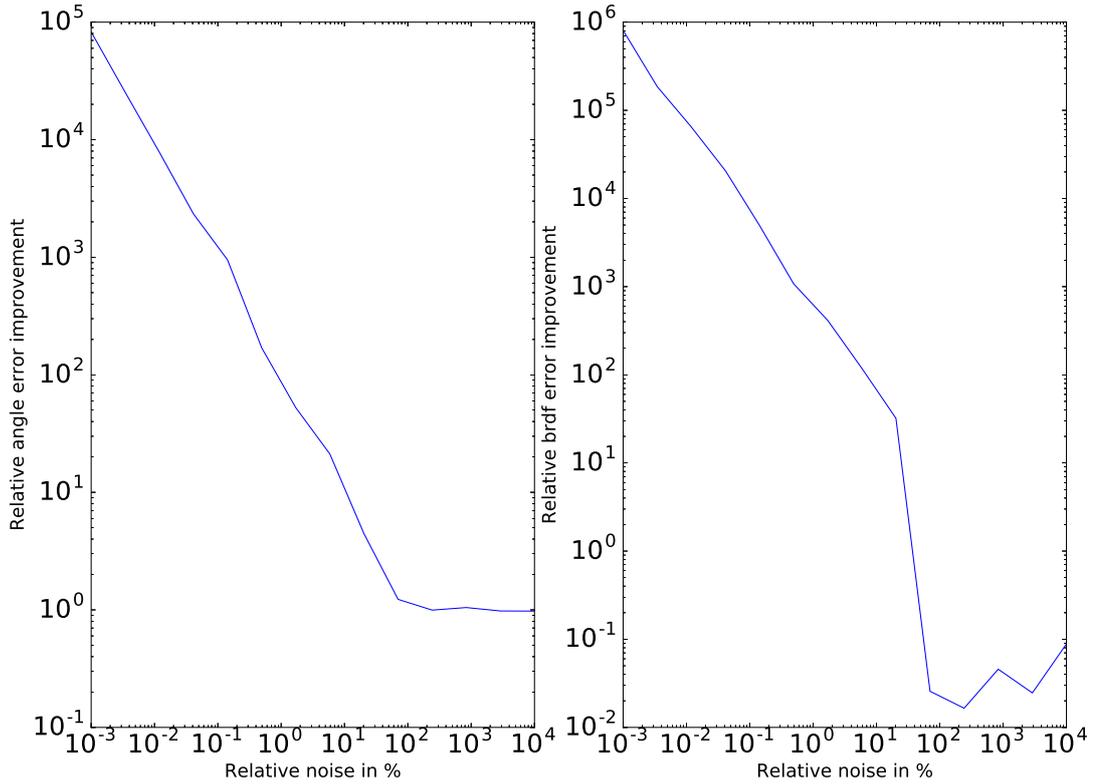


Figure 5.13: Comparison of angular and BRDF error improvements for the reconstruction with increasing amount of noise on the intensity values. Since the optimization may get stuck to a local minimum more easily in the presence of noise, we use the improvement yielded by the optimization with respect to the initialization as metric. The improvement is calculated by  $p = \frac{v_{\text{init}}}{v_{\text{opt}}}$ , where  $v$  stands for the angular error or the BRDF error accordingly.

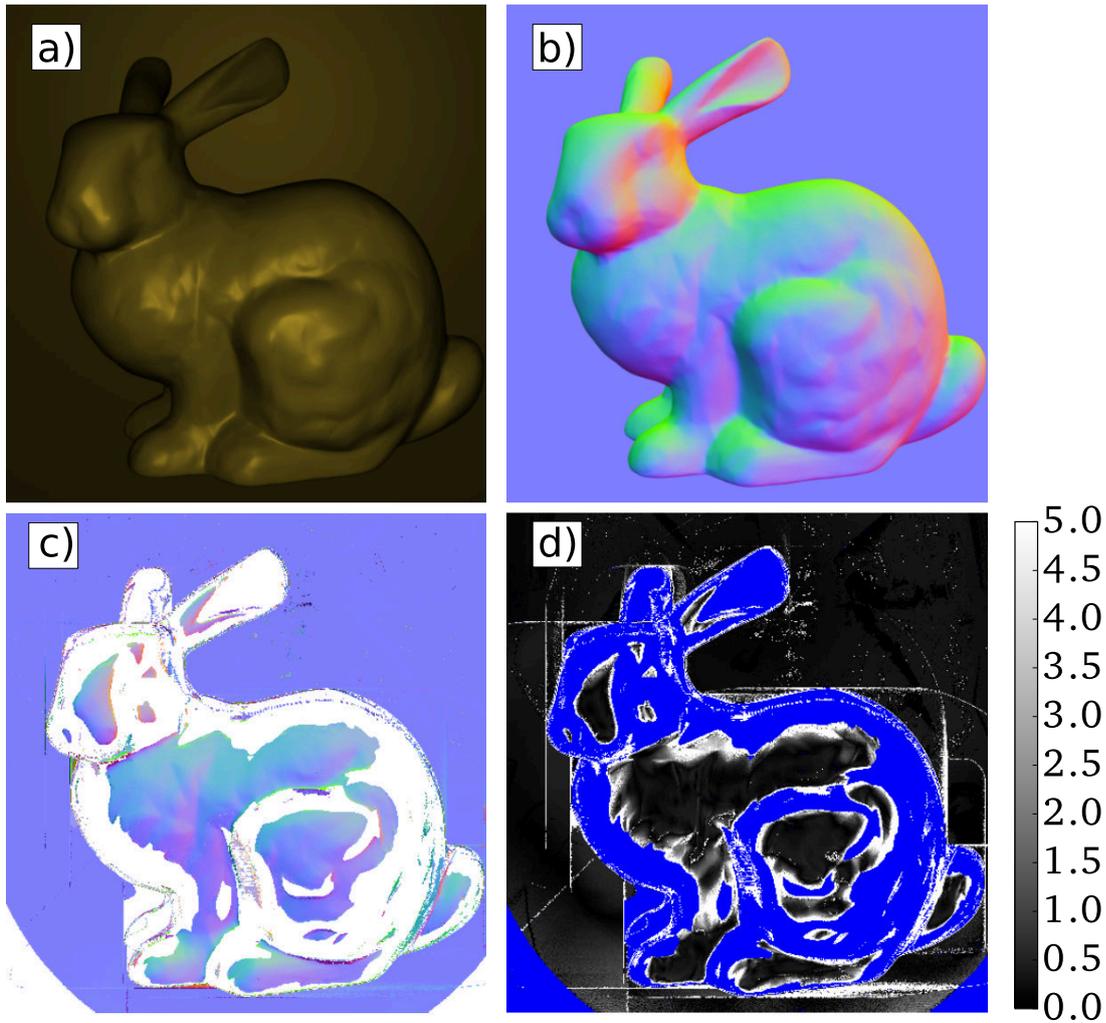


Figure 5.14: *a)* Original center view image. *b)* Color coded ground truth surface normals. *c)* Reconstructed surface normals. White areas were omitted, since there was not enough signal to calculate reliable estimates. *d)* Angular error in degree.

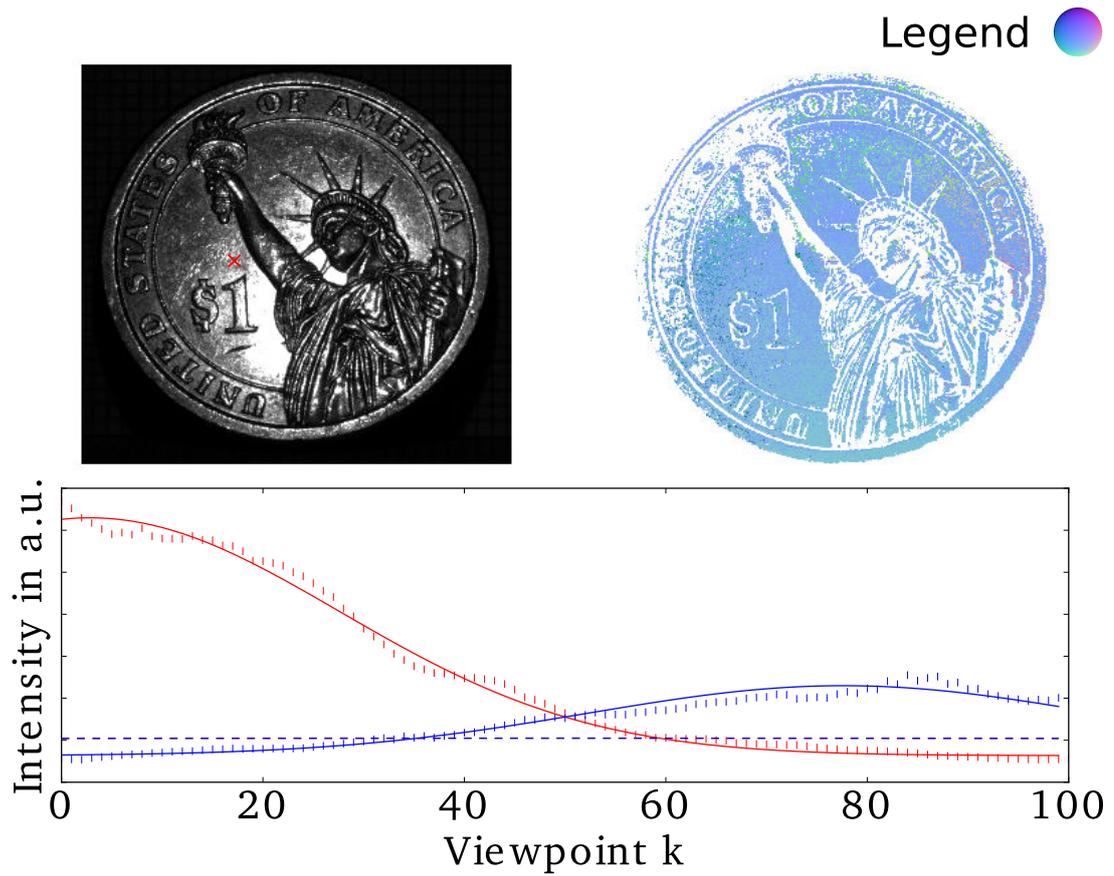


Figure 5.15: *a)* Original center view image. *b)* Color coded surface normals. White areas were omitted by our specular detection routine. *c)* Intensity curves for the horizontal and vertical view. The straight lines indicate the estimated model and the dashed line the initialization.

# 6

## Consistent Estimation of BRDF, Surface Normals and Depth

We have seen that for a given depth and given light sources, surface normal estimates are very precise. Unfortunately, most of the time we have limited knowledge of the scene depth for specular regions. To this end, the following method estimates disparity, BRDF and surface normals at the same time. The underlying algorithm for estimating occlusion aware depth estimates has been published in [52, 26], where also the synthetic results in this chapter of BRDF model incorporation are presented. The results on real world data have not been published yet.

As can be seen in Figure 6.1, depending on the disparity the resulting intensity distributions vary. This is a major issue, since this could imply that intensities are sampled from neighbourhood instead of distinguished, isolated 3D point. It is also challenging for the optimization, due to the non-differentiability of this change. Approaches for Lambertian surfaces use a Hough based method to minimize the intensity variance along EPI lines[28]. Using a BRDF model as metric for minimization in a voting based algorithm is unfeasible for two reasons:

1. Different disparities can lead to two equally likely hypothesis.
2. The additional non-linear regression step makes the algorithm infeasible

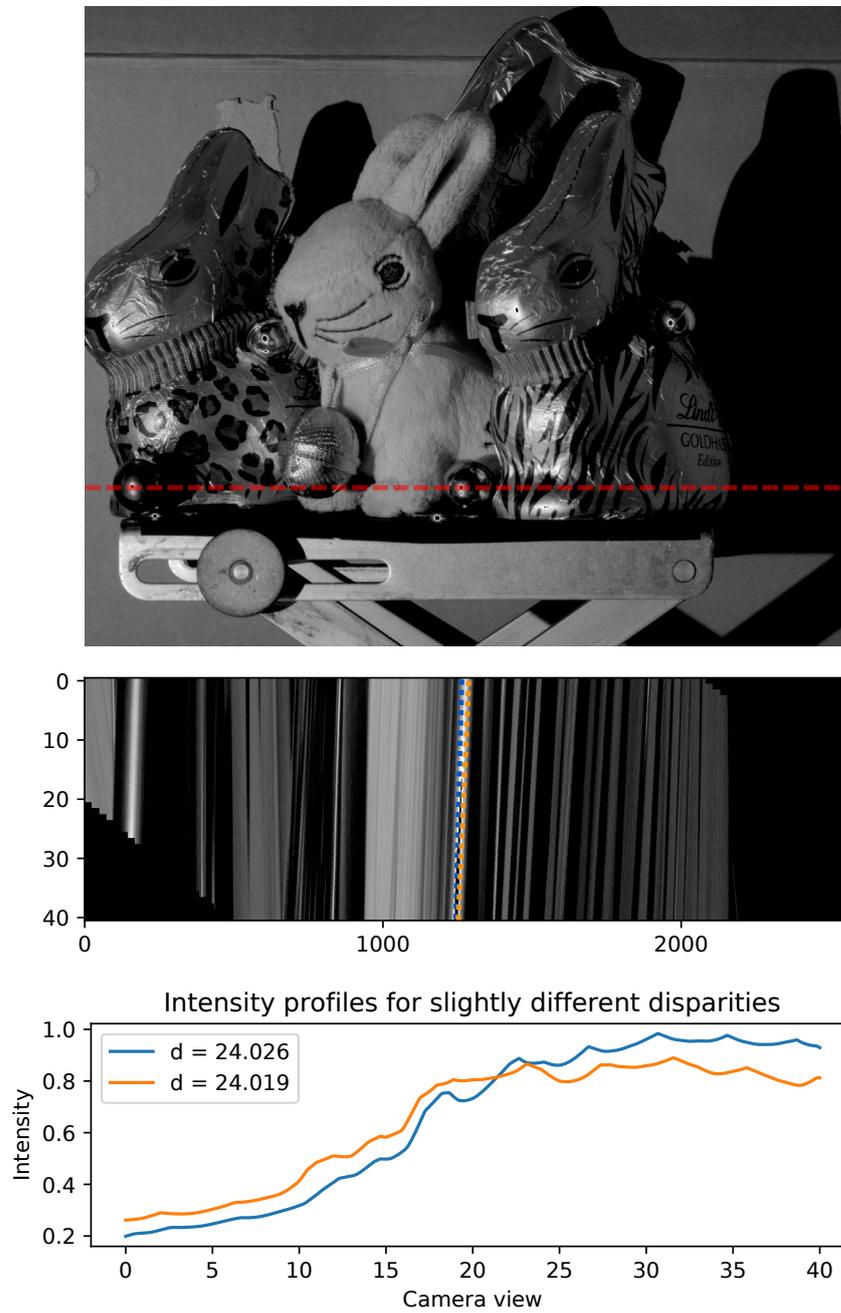


Figure 6.1: Centre view, EPI and sampled intensities for two slightly different disparity values. The EPI is taken from the red dotted row, was reshifted by a focus value of 24 and scaled in height, to improve the visibility of the line features. The line profiles are taken from the blue/orange marked lines.

---

slow.

**Data:** Light field data from multiple orientations, obeying at least two independent linear light field constraints. ( $lf_i$ )

**Result:** Disparity and surface normal map ( $disp\_map$ ,  $normal\_map$ ).

$disp\_map \leftarrow init\_with\_zerocrossings(lf_i)$ ;

$disp\_map \leftarrow fill\_by\_neighbors(lf_i, disp\_map)$ ;

$residual \leftarrow objective\_function(lf_i, disp\_map)$ ;

**while**  $residual > threshold$  or  $max$  iterations reached **do**

    //Minimize objective function using Patchmatch

**foreach** pixel in center view **do**

**foreach** disparity in semi\_random\_orientations **do**

$new\_residual \leftarrow objective\_function(lf_i, disp\_map)$ ;

**if**  $new\_residual < residual$  **then**

$disp\_map \leftarrow update\_disp\_map()$ ;

**end**

**end**

**if**  $residual > threshold$  and  $max$  lambertian iterations reached **then**

            switch objective function for current pixel to BRDF model and

            repeat current loop;

**end**

**end**

**end**

**Algorithm 1:** Computation of coherent normal and disparity map.

Instead a randomized algorithm similar to PatchMatch[7] is used. We first construct a sparse depth map based on image gradient features. Assume brightness constancy this sparse map is filled with neighbouring information. As soon as we can not find a meaningful hypothesis for a Lambertian surface, we switch to a specular reflection model. This way we limit expensive computations only to regions where specular highlights are likely. The different steps are shown in algorithm 1 and are elaborated in the next sections.

## 6.1 Sparse Depth Estimation

Our method builds on the tracking of sparse features along lines on the EPI to derive orientations. Figure 6.2 shows a visualization of the approach.

### 6.1.1 Zero Crossings

As features we use the location of zero crossings of the second derivative in the intensity domain. The vertical direction  $s$  in an EPI corresponds to the camera location, while the horizontal direction  $u$  corresponds to the image position. Thus, we localize image features in horizontal direction only, by calculating the 1D second order derivative, as depicted Figure 6.2, and detect zero-crossings with subpixel accuracy using linear interpolation.

In classical image processing, the position of a feature in an image can be determined by its location in memory and vice versa. Since it is not possible this way to save extracted features with floating point precision, they have to be saved in designated lists. While this means that conventional image processing implementations can't be utilized, it also means that there is a much larger degree of freedom when it comes to designing the evaluation of features.

### 6.1.2 Line Fitting

Since the distribution of features is quite dense and the amount of zero crossings per EPI is in the order of  $\sim 10^5$ , for an EPI with dimensions of  $100 \times 1000$  pixels, it is necessary to classify related features by a number of constraint beforehand. Thus, a classification similar to a tree search is implemented, allowing for dynamic adjustments to the constraints during classification and providing an efficient

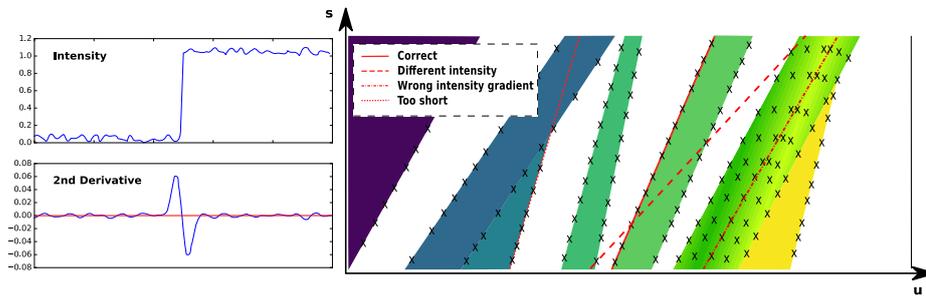


Figure 6.2: Left: Intensity with corresponding subpixel accurate second order derivative features (crosses). Right: Simplified EPI with zero crossings and line hypothesis. The solid line corresponds to a correct match, the dashed and dotted lines refer to rejected candidates due to violated constraints.

way of keeping track of already processed zero crossings. To establish a valid list of corresponding zero crossings which lie on a line we use a variant of the RANSAC algorithm. Because the number of zero-crossings can be quite high, especially for noisy images and smooth surfaces, we filter candidates using the image gradient. Our method iterates all feature points in the first view in the EPI. For each of these starting points, all zero-crossings in the last view which lie within the disparity range and whose gradient difference to the starting point is below a threshold, a candidate line is constructed. Then, all points in the EPI with a line-distance (in  $u$  direction) and a gradient difference below respective thresholds, are added to the line. From all candidates, the one with the highest number of supporting zero crossings is selected. In case of multiple lines with the same number of support points, the line with the lowest  $u$ -variance is chosen.

Here we take all combinations of two feature points within the disparity range from the first and last view to construct a model hypothesis. Each point with a distance lower than a certain threshold will support this hypothesis and the corresponding least square model will be calculated. Since we assume perfectness in the vertical EPI direction, where changes corresponds to different camera views, we are only interested in the point line distance in the horizontal EPI direction. Hence, for an arbitrary parametrized line model  $x = f(y, \theta)$ , where  $y$

is the y-coordinate of the zero crossing in the EPI and  $\theta$  the parameter vector, we define our residual for the RANSAC algorithm for each point  $(x_i, y_i)$  as

$$r_i = f(y_i, \theta) - x_i. \quad (6.1)$$

It turns out that even though it is possible to constrain most lines to an unambiguous set of zero crossings, there are areas in the EPIs where multiple zero crossings are valid options to include. For these cases a consistent additional metric is needed in order to decide which choice of zero crossings presents a best fit to the pool.

One way to identify zero crossings belonging to the same line is to look for zero crossings with similar intensity values. However, this is not useful in areas of uniform intensity, so additionally a uniform gradient is enforced. The gradient includes information about how brightness changes in a local neighborhood, so even for features with similar color the gradient can vary. There are not only way more possible numerical values for the gradient than for the intensity, allowing for a more precise constraint, there are also other possible usages for gradient information.

In areas of unvarying intensity, noise or specularities, a large number of trivial zero crossings might be registered impacting the fitting performance. Depending on noise distribution, gradient information can be used to help identify those zero crossings. Finally, we keep the lines with the most zero crossings and the lowest total variance defined as  $\sum r_i$  in a solution space.

The resulting slope of the line can be converted to a disparity  $d$  between different views. Given the line parametrization in normal form

$$r = x \cdot \cos(\alpha) + y \cdot \sin(\alpha), \quad (6.2)$$

where  $\alpha$  is the inclination angle and  $r$  the length of the normal segment. The

disparity is then given by

$$d = -\tan(\alpha)^{-1} \quad (6.3)$$

## 6.2 Optimization with an Occlusion Aware Bilateral Regularization Scheme

The line fitting method based on zero crossings, as presented above, provides high quality estimates as long as image features (edges, texture) are present. However, for a full scene, additional processing is required to determine depth estimates in areas with no feature points. Both the data and the smoothness term of the minimization metric are highly non-differentiable due to the sharp occlusion boundaries and hard cutoff thresholds in the smoothness term. Therefore, the optimization scheme is based on a randomized approach.

### 6.2.1 Initialization

Initially the sparse disparity map is filled based on the nearest evaluated disparity neighbors  $l$  and  $r$  with disparity  $d_l$  and  $d_r$ . The resulting disparity  $d_c$  is weighted based on the color difference with respect to the current pixel  $c$

$$d_c = w_l d_l + w_r d_r, \quad (6.4)$$

$$w_i = \frac{\|\mathbf{I}_c - \mathbf{I}_i\|}{\sum_i w_i}, \quad (6.5)$$

where  $\mathbf{I}_i$  denotes the color at the  $i$ -th pixel. Additionally a quality metric  $q$  is used to decide, if the disparity should be taken from the horizontal or the vertical part of the cross light field. It is defined as

$$q = \|\mathbf{I}_c - \mathbf{I}_l\| \|\mathbf{I}_c - \mathbf{I}_r\| \|d_c - d_l\| \|d_c - d_r\|. \quad (6.6)$$

This metric takes into account the color difference as well as the disparity differences of the respective pixels. The final disparity for the so far uninitialized pixel is taken from the light field direction where this score is lower.

### 6.2.2 Objective Function for Diffuse Regions

Since the values are probably not correct, an iteration scheme will optimize these starting disparities. As regularization term we use a modified bilateral filter  $B$  to preserve sharp disparity discontinuities. Similar to a canonical bilateral filter, we smooth values not only depending on the distance to the current pixel but also based on the distance in the disparity and the color space. To further keep crisp edges, we apply a hard cutoff when either the disparity difference is above a certain threshold  $t_d$  or the color difference is above a certain threshold  $t_c$ .

$$B(u, d) = \frac{1}{\omega_u} \sum_{q \in S} (f_n(\|u - q\|) f_I(\|\mathbf{I}(u) - \mathbf{I}(q)\|) f_d(\|d(u) - d(q)\|))^{-1} \xi(d(q)), \quad (6.7)$$

where  $\frac{1}{\omega_u}$  is the normalization term and the different  $f_i$  are slight modifications, such as a hard cutoff threshold,

$$f_i(x) = \begin{cases} x, & \text{if } x \leq t_i \\ 0, & \text{otherwise} \end{cases}. \quad (6.8)$$

For the spatial distance this is given by the filter mask size. Weighting neighbouring disparities for the current pixel would only make sense for a planar surface coplanar to the image plane. Therefore, the function  $\xi$  corrects the neighboring disparities by a surface normal estimation, *e.g.*

$$\xi(d(q)) = d(q) + n_d(q) \cdot (u - q), \quad (6.9)$$

where  $n_d$  is the surface normal in disparity space at the neighbour pixel  $q$ . The surface normal in disparity space is a weighted disparity gradient, again, with a hard cutoff threshold, to preserve sharp edges.

In a randomized fashion we want to minimize the following objective function with respect to the disparity

$$\operatorname{argmin}_d \sigma_{\text{diffuse}}^2(u, d) + \lambda(d - B(u, d))^2, \quad (6.10)$$

where the squared error of the data term in EPI space is given by

$$\sigma_{\text{diffuse}}^2(u, d) = \left( \sum_k I(u, c) - I(u + d \cdot (c - k), k) \right)^2. \quad (6.11)$$

Normally,  $c$  corresponds to the centre view, but could be in principle any other

view. It minimizes the variance  $\sigma^2$  of the colour intensity  $I$  along a corresponding line defined by an estimated disparity  $d$  in the EPI. The weighting  $\lambda$  between the smoothness term and the data term is slowly ramped up from iteration to iteration. This low start up of the regularization allows to correct obvious errors from the initialization and to fill between the sparse initial estimates which would otherwise be slowed down by the smoothing term.

### 6.2.3 Occlusions

Occlusion is handled using information from our depth model. This way we can simply use a threshold  $\Theta_d$  when evaluating the neighbours to the left and right in the centre view, respectively top and bottom for the vertical direction of the cross light field. A disparity sample  $d$  in the disparity map is considered to be occluded by another sample  $d_i$  if

$$d_i - d > \Theta_d. \tag{6.12}$$

The calculation deciding whether a sample is occluded or not is carried out during the evaluation of the cost terms. Thus, updates to the model during an iteration of the optimization directly affects the costs of all future evaluations, speeding up the propagation of locally good solutions.

### 6.2.4 Iteration Scheme

What is left for a randomized iteration scheme for determining disparity values, is the sampling of hypothesis. We limit the solution range to a minimal and maximal disparity  $d_{\min}$  and  $d_{\max}$ , which is given by the capturing setup. To quickly

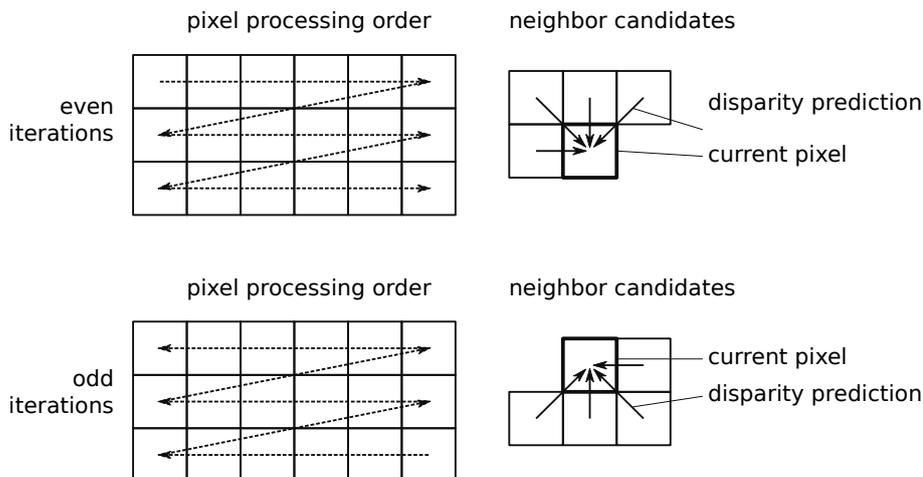


Figure 6.3: To achieve fast propagation of good disparity values over great distances, neighbours are used as hypothesis for the current pixel[50].

propagate correct solutions from neighboring pixels, a PatchMatch approach is used, see Figure 6.3. This means that additionally to a random hypothesis, neighbor pixel solutions are also evaluated, due to their likeliness to provide a good solution. For all even iterations the disparity from the left and top neighbours is evaluated and for all odd iterations the right and bottom pixels are selected.

### 6.2.5 Activation of the BRDF-Surface Normal Optimizer

After some iterations, usually somewhere around 2 to 5, there will still be a high residual for locations where we have specular reflections. Here, we replace the diffuse error metric  $\sigma_{\text{diffuse}}^2(u, d)$  with the residual presented in chapter 5. Additionally we convert the reconstructed surface normals from 3D space to disparity space and use them in analogy to Equation 6.9 to further enhance the reconstruction.

$$\sigma_{\text{spec}}^2 = \sum_j \|L_d(\Theta, \mathbf{n}, \omega_{\mathbf{o}, \mathbf{j}}) - I(\omega_{\mathbf{o}, \mathbf{j}})\|^2. \quad (5.3 \text{ rev.})$$

This way we can reach a low error even for specular surfaces.

## 6.3 Results

In this section we present disparity and normal maps from synthetic and real scenes. For the synthetic scene generated with Blender, we provide and errors for the surface normals and disparities compared to the ground truth. For the real scenes we present comparisons of disparity maps with and without additional BRDF surface normal information.

### 6.3.1 Synthetic

In Figure 6.4 we see the reconstructed disparity and surface normals of the Stanford bunny. In contrast to chapter 5, we were able to give accurate disparity estimates for specular regions. It is interesting to note, that the accuracy in regions where the specular peak is clearly visible in both light field directions, is even higher than for the diffuse parts. This clearly indicates, that additional information from the intensity distribution leads to globally consistent depth estimates. In analogy to chapter 5 we only keep surface normals where the residual of the fit is low enough and the general variability of the intensity signal is high enough. Thus, we can assume that these detected surface normals are a result of the actual surface and not an artifact of the non-linear optimization.

The largest disparity differences, up to  $\approx 0.2$  are at boundary of diffuse and specular regions. Here the signal of the specular part is not strong enough to give reliable information about the surface orientation, but strong enough to disturb the constant intensity assumption.

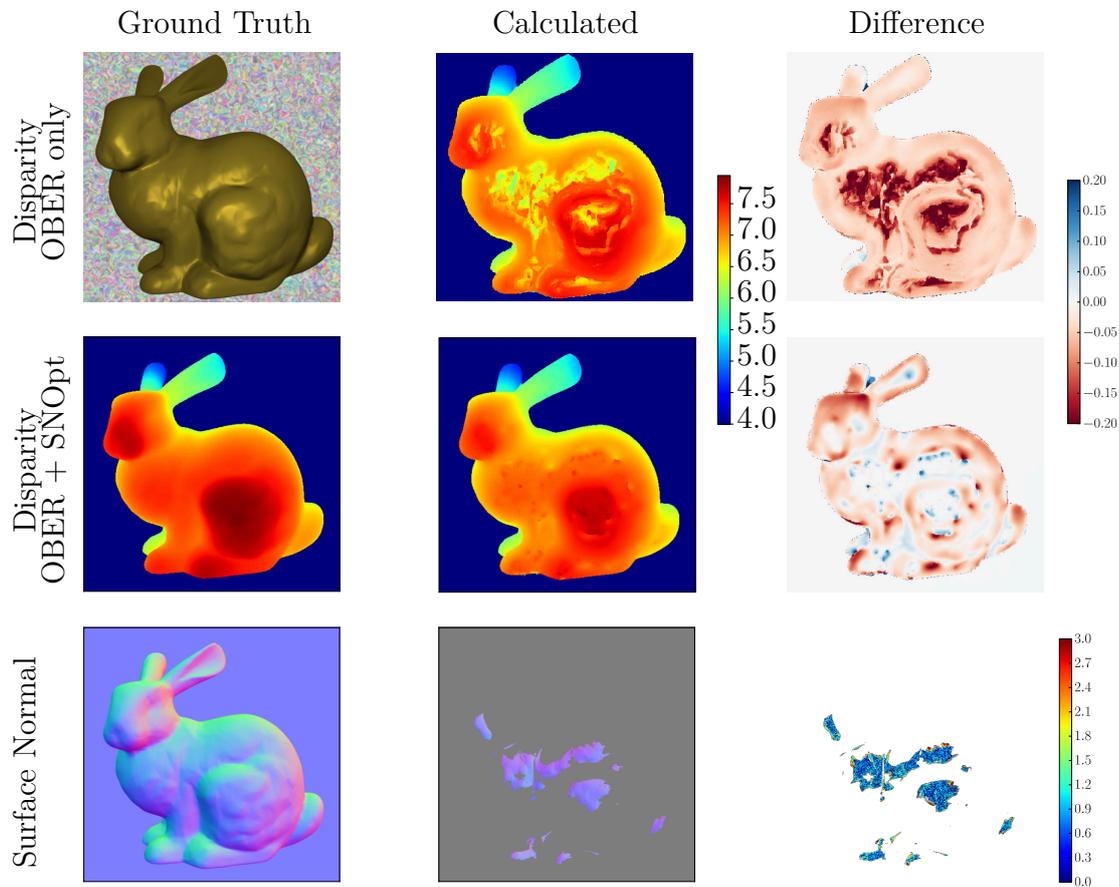


Figure 6.4: Synthetic evaluation of the Stanford bunny. The top row presents the disparities calculated using the OBER method without additional BRDF term for specular surfaces. The middle row shows disparities using OBER with additional surface normal optimisation using a BRDF model. The last row shows the surface normals obtained for specular regions. From left to right: the ground truth, the derived values from the optimization scheme and the difference to the ground truth are presented. The disparity difference is given by  $d_{\text{calc}} - d_{\text{gt}}$ , whereas for the surface normal, the difference is the angular error to the ground truth in degree. The surface normals are converted to RGB images by using the commonly used conversion  $RGB = (\mathbf{n} + 1)/2$ . We notice that the usage of the surface normal optimization greatly improves the accuracy of the disparity estimate and hence the depth. The surface normals achieve an accuracy of less than  $3^\circ$ .

### 6.3.2 Real

Given a measurements geometry described in chapter 3, we are able to observe a large fraction of the specular lobe. This can be mainly achieved by either a large distance between the outermost cameras or small distance between the cameras and the object. Here we want to push the boundaries by explicitly choosing sub-optimal acquisition settings. We used a cross light field setup with a baseline of 2 mm and 41 views per axis. The image sensor has a resolution of  $2160 \times 2560$  pixels. The camera, as well as the light source were carefully calibrated. The distance of the center camera to the objects was roughly 85 cm. The light source was located at  $(-397.0, -83.2, 20.9)$ mm with respect to the camera center. The coordinate system is right hand sided with  $z$ -axis pointing towards the objects and the  $y$  axis pointing upwards.

In Figure 6.5 we can see the resulting disparities and normal maps on real scenes. Only these areas exhibit any information about surface orientation, which specular lobe intersected with the camera geometry. This was also the case for the synthetic images, but is here even more pronounced.



Figure 6.5: Overview of real world results. From left to right: Center view, disparity using OBER only, disparity and surface normals with additional Surface Normal Optimizer. We see a small improvement in specular regions in the disparity map and can reconstruct plausible surface normals for the regions, where we the intensity change along the views is strong enough.



# 7

## Conclusion

In the following, we review the presented insights about disparity estimation and surface normal reconstruction using light fields. Additionally, we provide a brief outlook on possible future research questions.

### 7.1 Summary

After giving a brief overview of the fundamental principals in chapter 2, we showed to optimally acquire a light field for surface normal reconstruction in chapter 3. Here different configurations of a single light source and a moving camera were discussed. In essence, the configuration of light source, surface normal and camera must fulfill the law of reflection to retrieve specular highlights.

In chapter 4 we have derived how surface normals can be recovered based on two visible intensity peaks in a cross light field setup. Geometric considerations imply that the camera position of these peaks and a single 3D point on a surface are sufficient to recover the surface orientation. Only taking this limited amount of information into account has the advantage that the surface normal can be efficiently computed. However, this approach ignores large parts of the light field data, which could further increase the accuracy.

To use the full extend of the light field, the expected reflectance distribution based on BRDF models was considered in chapter 5. This approach offers multiple benefits. First, it allows to recover the surface normals when only parts of the peaks are visible. Second, the shape of the reflected intensity distribution contains information on the material properties. Third, it increases the accuracy compared to using only the maximum peak intensity due to an adequate interpolation.

The problem of simultaneously estimating depth and surface orientation is tackled in chapter 6. Here, we use the surface normal optimization method based on BRDF models presented in chapter 5 in combination with an occlusion aware bilateral filter for improving sparse depth estimates. The BRDF optimization is turned on for image regions where the model of occlusions and a Lambertian reflection disagrees with the data term. This combined approach increases the precision of the disparity estimation and gives precise surface normals.

Recovering the 3D scene together with material properties and illumination still poses a great challenge in computer vision. Many available approaches ignore the additional information specular highlights can provide. This is partly due to saturated images but also to the inherent ambiguity in the image formation process caused by the illumination, material properties and geometry of a scene. Each of these influences the light transport, and they are difficult to separate. To approach this hard problem, we developed new algorithms which use the special properties of structured light fields.

## 7.2 Conclusions and Outlook

Recent progress in the area of computer vision has been aligned with the tremendous improvement of camera sensors and computational capabilities. Despite these advances, analysing light field data poses high demands on computational

hardware, due to its high dimensionality. This calls for the development of efficient algorithms. Additionally, computing power will likely increase further, as can be observed in the trend of ever more powerful graphic cards to support deep learning algorithms. Hence, the capturing and processing of light field data will become more affordable leading to a rise in its use.

The reconstruction of global illumination and shadows is a highly challenging task. Creating more accurate and versatile algorithms depends on increasing computational power. Even the generation of physical realistic scenes is difficult to create in real time without relying on shortcuts, such as ambient lighting. The inverse process is, as we have seen, much harder.

A lot of algorithmic choices still depend on handcrafted thresholds, often derived from trial and error. For example, in chapter 6 we use a threshold to decide between a constant reflection model or a more complex BRDF. In future work, we could use neural networks to learn these thresholds from examples.

Further research is required to develop more tractable lighting models to ease reconstruction processes and use the recovered geometry for a global model to estimated shadowed regions. One step in that direction would be to solve for the surface orientation and depth of a surface in a more tightly joint fashion.

The holy grail of inverse rendering is still not reached, but there is light at the end of the tunnel.



# A

## Appendix

### A.1 Analytical Solution for Geometrical Reconstruction of the Surface Normal

The analytical solution to section 4.2 can be derived by simplifying Equation 4.3

$$\mathbf{q}'_i = \mathbf{p}_i + \frac{(\mathbf{p}_0 - \mathbf{p}_i) \cdot \mathbf{n}_r}{\mathbf{g}_i \cdot \mathbf{n}_r} \mathbf{g}_i. \quad (4.3 \text{ rev.})$$

Setting  $\mathbf{p}_i$  to zero and  $d_{r,z} = 1$ , since the direction of  $\mathbf{d}_r$  depends only on two variables, the third is subject to normalization. Thus, we can setup a system of two equations

$$\mathbf{q}_1 = \frac{\mathbf{p}_0 \cdot \mathbf{n}_{r,1}}{\mathbf{g}_1 \cdot \mathbf{n}_{r,1}} \mathbf{g}_1 \quad (A.1)$$

$$\mathbf{q}_2 = \frac{\mathbf{p}_0 \cdot \mathbf{n}_{r,2}}{\mathbf{g}_2 \cdot \mathbf{n}_{r,2}} \mathbf{g}_2, \quad (A.2)$$

where for the first equation only the first component is relevant and for the second the second component. This leads to

$$(\mathbf{g}_1 \cdot \mathbf{n}_{\mathbf{r},1})q_1 = \mathbf{p}_0 \cdot \mathbf{n}_{\mathbf{r},1} \quad (\text{A.3})$$

$$(\mathbf{g}_2 \cdot \mathbf{n}_{\mathbf{r},2})q_2 = \mathbf{p}_0 \cdot \mathbf{n}_{\mathbf{r},2} \quad (\text{A.4})$$

and with

$$\mathbf{n}_{\mathbf{r}} = \mathbf{d}_{\mathbf{r}} \times (\mathbf{g}_i \times \mathbf{d}_{\mathbf{r}}). \quad (\text{4.4 rev.})$$

to

$$(1 + d_y^2)p_x - d_x d_y p_y - d_x p_z - (1 + d_y^2)q_1 = 0 \quad (\text{A.5})$$

$$(1 + d_x^2)p_y - d_x d_y p_x - d_y p_z - (1 + d_x^2)q_1 = 0. \quad (\text{A.6})$$

This can be solved analytically for  $d_x$  and  $d_y$  with a CAS program, *e.g.* Mathematica. Unfortunately this solution is way to large to be of any use. An iterative approach usually converges in four iterations which is much faster.

# List of Own Publications

- [13] Diebold, M., Blum, O., Gutsche, M., Wanner, S., Garbe, C. S., Baker, H., and Jähne, B. “Light-field camera design for high-accuracy depth estimation”. In: *SPIE Optical Metrology*. International Society for Optics and Photonics. 2015, pp. 952803–952803.
- [19] Gutsche, M. “Surface velocity measurements at the aeolotron by means of active thermography”. MA thesis. Heidelberg University, 2014.
- [20] Gutsche, M., Schilling, H., Diebold, M., and Garbe, C. “Surface Normal Reconstruction From Specular Information in Light Field Data”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. July 2017.
- [26] Johannsen, O. et al. “A Taxonomy and Evaluation of Dense Light Field Depth Estimation Algorithms”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. July 2017.
- [51] Schilling, H., Diebold, M., Gutsche, M., and Jähne, B. “On the design of a fractal calibration pattern for improved camera calibration”. In: *tm-Technisches Messen* (2017).



# List of Figures

1.1	Many technologies today depend on accurate 3D models. . . . .	3
2.1	Light at an interface can be transmitted, absorbed or reflected. The total energy of the light is conserved. Second order effects, such as subsurface scattering, are not shown here. . . . .	10
2.2	There is a plethora of different BRDF models, which can be categorized according to different attributes [37]. . . . .	15
2.3	The halfway vector parameterization allows for a more efficient computation of the Blinn-Phong model and is especially useful for isotropic BRDFs. . . . .	16
2.4	The microfacet model describes reflection properties by modeling a surface with randomly oriented microspheres. The underlying probability distribution, as well as the roughness of the surface dominate the appearance of a surface. . . . .	17
2.5	Scene geometry for estimating depth by triangulation. The different positions $x_l$ and $x_r$ for the projections of the 3D point $\mathbf{x}$ result in the disparity $d = x_l - x_r$ . The depth $z$ can then be inferred by $z = \frac{bf}{d}$ . . . . .	20
2.6	To capture a structured light field, we add more cameras. For the ease of computation we keep the baseline between camera pairs constant. . . . .	22
2.7	Object with a specular highlight. An epipolar plane image was built from different views and the same image row indicated by the red line. The EPI was reshifted for better visibility of the line structures. . . . .	23

2.8 Representation of a straight line in Hough space in the  $yt$ -plane. On the left hand side the line in image space is visualized. The right hand side shows the corresponding hough transformation with intensity values on a logarithmic scale. Image taken from [19]. 26

3.1 Light fields can be captured using a translation stage with a mounted camera. This is a cost-effective setup, but only static scenes can be captured. . . . . 28

3.2 With a linear camera array in a verged setup also dynamic scenes can be captured. . . . . 29

3.3 Consumer cameras such as the Lytro use a microlens array to capture a light field. Proportions are not to scale. . . . . 30

3.4 The surface normal range which can be recovered by a light field setup is limited by the geometry between cameras  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , object and light source. . . . . 31

3.5 Inspecting 2D slices of BRDFs in half-way vector space we can see how the characteristics of different materials differ. For example the fall off rate for aluminum oxide is much larger than for gold paint. The data is scaled logarithmically for better visibility and taken from the MERL database[36]. . . . . 33

3.6 We investigate three different acquisition setups for light field data. 33

3.7 Trajectories through the 2D BRDF space for a straight moving camera and a fixed light source. The camera moves in the x direction perpendicular to its viewing direction anti-parallel to the surface normal (in the blue case). The green trajectory describes a path were the surface normal is tilted by  $22^\circ$  around the y-axis (pointing upwards). The red trajectory is tilted by  $45^\circ$  around the y-axis and small amount around the x-axis facing now upwards. . 34

3.8 Trajectories through the 2D BRDF space for a straight moving camera and a co-moving light source. The different colors encode the same normal displacements as in Figure 3.7. . . . . 35

3.9	Trajectories for circular motion. The distance between camera and light source is given by $d$ . . . . .	36
3.10	BRDF coverage of a circular light field with different light positions. Differently colored lines encode the light position. The light position is also varied in a circular fashion around the object. Zero indicates the start position close to the camera and one the opposite site of the circle. The camera is stationary and the surface normal is in the same plane spanned by the light source and camera. . . . .	37
4.1	EPI with specular highlight. . . . .	40
4.2	Sketch of the cross array setup. $\mathbf{q}_1$ and $\mathbf{q}_2$ are the nearest points to $\mathbf{r}$ . Hence, in the closest view we will see an intensity maximum. . . . .	41
4.3	Accuracy depending on the relative signal noise. The error-bars represent the standard deviation given by $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$ . The angular error does not approach zero due to quantization of the highlight position by the fixed number of cameras. . . . .	45
4.4	Accuracy depending on the number of cameras used. . . . .	46
5.1	An intensity distribution can be obtained using the disparity in the EPI. Given a correct disparity, the intensity distribution depends on the relative orientation of the surface, the material properties and the illumination. . . . .	49
5.2	Process of simultaneously estimating surface normals and BRDF parameters. . . . .	51
5.3	The residual of the objective function as heat maps for different pairs of variables. On the left side all variable are a initialised according to a best guess. On the right hand side we see the residual after optimisation. The minimum is highlighted with a red cross. If we compare the left and right hand side we see that there are possibilities for getting stuck in a local minimum. . . . .	53

5.4	a) The orientation in the horizontal EPI directly relates to the disparity of the input image. b) Reshifting the EPI by the extracted disparity allows to investigate the intensity distribution by moving along the views which are here along the y-direction. c) Similarly the disparity for all views can be viewed in a "disparity EPI". d) Reshifting it leads to a straight forward computation of occlusions by looking at gradients in the y-direction. . . . .	56
5.5	Overview of different intensity distributions. . . . .	57
5.6	Distribution of the angular error in degree before filtering out diffuse regions. The algorithm fails for diffuse regions, where no intensity variation is present. Hence we will suppress the activation on those regions. . . . .	60
5.7	Angular error in degree depending on the residual and the relative intensity difference. The relative intensity difference is defined as $r = \frac{I_{\max} - I_{\min}}{I_{\max}}$ . We see that we can effectively filter out regions where the residual is low, while the intensity difference is only reliable if $r > 0.25$ . . . . .	62
5.8	Distribution of the angular error after excluding diffuse areas and non converged optimizations. . . . .	63
5.9	Comparison of BRDFs of different materials from the MERL database. On the left hand side only the purely specular component is present. On the right hand side Fresnel effects become apparent. For better visibility both axes are scaled logarithmically. . . . .	64
5.10	Analysis of the angular error and the relative error in the BRDF parameters, dependent on the angle error of the light direction. We rotated the light direction around the x and the y-axis up to 90°. We took 10 different samples and took the median for each different light orientation. . . . .	65

5.11 Analysis of the angular error, dependent on the number of views. The precisions increases remarkably up to 30 views. With more than 150 views discretization errors play a significant role and we can't increase the precision anymore. . . . . 66

5.12 Comparison of angular errors for the reconstruction with correct disparity maps (left) and with a constant disparity roughly in the plane of the bunny. . . . . 67

5.13 Comparison of angular and BRDF error improvements for the reconstruction with increasing amount of noise on the intensity values. Since the optimization may get stuck to a local minimum more easily in the presence of noise, we use the improvement yielded by the optimization with respect to the initialization as metric. The improvement is calculated by  $p = \frac{v_{init}}{v_{opt}}$ , where  $v$  stands for the angular error or the BRDF error accordingly. . . . . 68

5.14 *a)* Original center view image. *b)* Color coded ground truth surface normals. *c)* Reconstructed surface normals. White areas were omitted, since there was not enough signal to calculate reliable estimates. *d)* Angular error in degree. . . . . 69

5.15 *a)* Original center view image. *b)* Color coded surface normals. White areas were omitted by our specular detection routine. *c)* Intensity curves for the horizontal and vertical view. The straight lines indicate the estimated model and the dashed line the initialization. . . . . 70

6.1 Centre view, EPI and sampled intensities for two slightly different disparity values. The EPI is taken from the red dotted row, was reshifted by a focus value of 24 and scaled in height, to improve the visibility of the line features. The line profiles are taken from the blue/orange marked lines. . . . . 72

6.2	Left: Intensity with corresponding subpixel accurate second order derivative features (crosses). Right: Simplified EPI with zero crossings and line hypothesis. The solid line corresponds to a correct match, the dashed and dotted lines refer to rejected candidates due to violated constraints. . . . .	75
6.3	To achieve fast propagation of good disparity values over great distances, neighbours are used as hypothesis for the current pixel[50].	81
6.4	Synthetic evaluation of the Stanford bunny. . . . .	83
6.5	. . . . .	85

# Bibliography

- [1] Adato, Y., Vasilyev, Y., Zickler, T., and Ben-Shahar, O. “Shape from Specular Flow”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.11 (Nov. 2010), pp. 2054–2070. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.126.
- [2] Adelson, E. and Bergen, J. “The plenoptic function and the elements of early vision”. In: *Computational models of visual processing 1* (1991).
- [3] Ashikhmin, M. and Shirley, P. “An anisotropic phong BRDF model”. In: *Journal of graphics tools* 5.2 (2000), pp. 25–32.
- [4] Bajcsy, R., Lee, S., and Leonardis, A. “Color image segmentation with detection of highlights and local illumination induced by inter-reflections”. In: *Pattern Recognition, 1990. Proceedings., 10th International Conference on*. Vol. i. June 1990, pp. 785–790. DOI: 10.1109/ICPR.1990.118217.
- [5] Bay, H., Tuytelaars, T., and Van Gool, L. “Surf: Speeded up robust features”. In: *Computer vision—ECCV 2006* (2006), pp. 404–417.
- [6] Beckmann, P. and Spizzichino, A. “The scattering of electromagnetic waves from rough surfaces”. In: *Norwood, MA, Artech House, Inc., 1987, 511 p.* (1987).
- [7] Besse, F., Rother, C., Fitzgibbon, A., and Kautz, J. “PMBP: PatchMatch Belief Propagation for Correspondence Field Estimation”. In: *International Journal of Computer Vision* 110.1 (2014), pp. 2–13. ISSN: 0920-5691. DOI: 10.1007/s11263-013-0653-9.
- [8] Bigun, J. *Optimal orientation detection of linear symmetry*. 1987.
- [9] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Institute, Amsterdam: Blender Foundation, 2016. URL: <http://www.blender.org>.

- [10] Bolles, R., Baker, H., and Marimont, D. “Epipolar-plane image analysis: An approach to determining structure from motion”. In: *International Journal of Computer Vision* 1.1 (1987), pp. 7–55.
- [11] Criminisi, A., Kang, S. B., Swaminathan, R., Szeliski, R., and Anandan, P. “Extracting layers and analyzing their specular properties using epipolar-plane-image analysis”. In: *Computer Vision and Image Understanding* 97.1 (2005), pp. 51–85. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2004.06.001.
- [12] Diebold, M. and Goldluecke, B. “Epipolar Plane Image Refocusing for Improved Depth Estimation and Occlusion Handling”. In: *Vision, Modeling and Visualization Workshop VMV*. 2013.
- [14] *Digital Special Effect Makeup vs. Physical Makeup in Movies - Media-Makeup.Today*. <http://mediamakeup.today/digital-special-effect-makeup-vs-physical-makeup-in-movies/>. (Accessed on 04/30/2018).
- [15] Ganapathi, V., Plagemann, C., Koller, D., and Thrun, S. “Real time motion capture using a single time-of-flight camera”. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, pp. 755–762.
- [16] Gershon, R., Jepson, A. D., and Tsotsos, J. K. “The Use of Color in Highlight Identification.” In: *IJCAI*. 1987, pp. 752–754. URL: <http://ijcai.org/Past%20Proceedings/IJCAI-87-VOL2/PDF/034.pdf>.
- [17] Gortler, S., Grzeszczuk, R., Szeliski, R., and Cohen, M. “The Lumigraph”. In: 1996, pp. 43–54.
- [18] Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F. “The lumigraph”. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 43–54.
- [19] Gutsche, M. “Surface velocity measurements at the aeolotron by means of active thermography”. MA thesis. Heidelberg University, 2014.
- [21] Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision*. second. Cambridge University Press, 2004.

- [22] Holroyd, M., Lawrence, J., Humphreys, G., and Zickler, T. “A photometric approach for estimating normals and tangents”. In: *ACM Transactions on Graphics (TOG)* 27.5 (2008), p. 133.
- [23] Jachnik, J., Newcombe, R. A., and Davison, A. J. “Real-time surface light-field capture for augmentation of planar specular surfaces”. In: *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE. 2012, pp. 91–97.
- [24] Jähne, B. *Digitale Bildverarbeitung, 6th revised and extended edition*. Berlin: Springer, 2005.
- [25] Jin, H., Soatto, S., and Yezzi, A. J. “Multi-view stereo beyond lambert”. In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2003, pp. I–171.
- [27] Kälviäinen, H., Hirvonen, P., Xu, L., and Oja, E. “Probabilistic and non-probabilistic Hough transforms: overview and comparisons”. In: *Image and vision computing* 13.4 (1995), pp. 239–252.
- [28] Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., and Gross, M. H. “Scene reconstruction from high spatio-angular resolution light fields.” In: *ACM Trans. Graph.* 32.4 (2013), pp. 73–1.
- [29] Lee, S. W. and Bajcsy, R. “Detection of specularities using color and multiple views”. In: *Computer Vision ECCV’92*. Ed. by Sandini, G. Vol. 588. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1992, pp. 99–114. ISBN: 978-3-540-55426-4. DOI: 10.1007/3-540-55426-2\_13.
- [30] Levoy, M. and Hanrahan, P. “Light field rendering”. In: 1996, pp. 31–42.
- [31] Levoy, M. and Hanrahan, P. “Light field rendering”. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 31–42.
- [32] Liu, B., Gould, S., and Koller, D. “Single image depth estimation from predicted semantic labels”. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, pp. 1253–1260.

- [33] Liu, F., Shen, C., and Lin, G. “Deep convolutional neural fields for depth estimation from a single image”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 5162–5170.
- [34] Lowe, D. G. “Object recognition from local scale-invariant features”. In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Vol. 2. Ieee. 1999, pp. 1150–1157.
- [35] Lu, F., Matsushita, Y., Sato, I., Okabe, T., and Sato, Y. “Uncalibrated photometric stereo for unknown isotropic reflectances”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 1490–1497.
- [36] Matusik, W. “A data-driven reflectance model”. PhD thesis. Massachusetts Institute of Technology, 2003.
- [37] Montes Soldado, R. and Ureña Almagro, C. “An Overview of BRDF Models”. In: (2012).
- [38] Nair, R., Fitzgibbon, A., Kondermann, D., and Rother, C. “Reflection Modeling for Passive Stereo”. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE. 2015, pp. 2291–2299. DOI: 10.1109/iccv.2015.264.
- [39] Navarro, J. and Buades, A. “Robust and dense depth estimation for light field images”. In: *IEEE Transactions on Image Processing* 26.4 (2017), pp. 1873–1886.
- [40] Norris, T. B., Zhong, Z., Fessler, J. A., Liu, C.-H., and Chang, Y.-C. *Light field imaging with transparent photodetectors*. US Patent App. 15/430,043. Feb. 2017.
- [41] Oren, M. and Nayar, S. K. “Generalization of Lambert’s reflectance model”. In: *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM. 1994, pp. 239–246.

- [42] Oxholm, G. and Nishino, K. “Multiview shape and reflectance from natural illumination”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 2155–2162.
- [43] *Parrot positions its consumer drones for 3D modeling, mapping and agricultural uses | TechCrunch*. <https://techcrunch.com/2017/05/08/parrot-positions-its-consumer-drones-for-3d-modeling-mapping-and-agricultural-uses/>. (Accessed on 04/30/2018).
- [44] Phong, B. T. “Illumination for computer generated pictures”. In: *Communications of the ACM* 18.6 (1975), pp. 311–317.
- [45] Poggio, T., Torre, V., and Koch, C. “Computational Vision and Regularization Theory”. In: *Nature* 317 (1985), pp. 314–319.
- [46] Ramamoorthi, R. and Hanrahan, P. “A signal-processing framework for inverse rendering”. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM. 2001, pp. 117–128. DOI: 10.1145/383259.383271.
- [47] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. “ORB: An efficient alternative to SIFT or SURF”. In: *Computer Vision (ICCV), 2011 IEEE international conference on*. IEEE. 2011, pp. 2564–2571.
- [48] Rusinkiewicz, S. M. “A new change of variables for efficient BRDF representation”. In: *Rendering techniques 98*. Springer, 1998, pp. 11–22.
- [49] S. Wanner and B. Goldluecke. “Variational Light Field Analysis for Disparity Estimation and Super-Resolution”. In: (2013).
- [50] Schilling, H. “OBER - Occluded Bilateral EPI Regularization”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Poster*. 2017.
- [51] Schilling, H., Diebold, M., Gutsche, M., and Jähne, B. “On the design of a fractal calibration pattern for improved camera calibration”. In: *tm-Technisches Messen* (2017).

- [52] Schilling, H., Diebold, M., Rother, C., and Jähne, B. “Trust your Model: Light Field Depth Estimation with inline Occlusion Handling”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.
- [53] Schlick, C. “An Inexpensive BRDF Model for Physically-based Rendering”. In: *Computer graphics forum*. Vol. 13. 3. Wiley Online Library. 1994, pp. 233–246.
- [54] *Sciaky’s First Sale of Big Metal 3D Printer - 3D Printing Industry*. <https://3dprintingindustry.com/news/sciaky-makes-first-official-sale-giant-metal-3d-printer-30952/>. (Accessed on 04/30/2018).
- [55] Seeliger, H. *Zur Theorie der Beleuchtung der grossen Planeten insbesondere des Saturn: Aus den Abhandlungen der k. bayer. Akademie der Wiss. II Cl. 16 Bd. II Abth.* Verlag der k. Akademie in Comm. bei G. Franz, 1887.
- [56] Si, L. and Wang, Q. “Dense Depth-Map Estimation and Geometry Inference from Light Fields via Global Optimization”. In: *Asian Conference on Computer Vision*. 2016.
- [57] Stephens, R. S. “Probabilistic approach to the Hough transform”. In: *Image and vision computing* 9.1 (1991), pp. 66–71.
- [58] Torrance, K. E. and Sparrow, E. M. “Theory for off-specular reflection from roughened surfaces”. In: *Josa* 57.9 (1967), pp. 1105–1114.
- [59] Tsin, Y., Kang, S. B., and Szeliski, R. “Stereo matching with reflections and translucency”. In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. Vol. 1. IEEE. 2003, pp. I–702.
- [60] Wang, T.-C., Chandraker, M., Efros, A. A., and Ramamoorthi, R. “SVBRDF-Invariant Shape and Reflectance Estimation From Light-Field Cameras”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

- [61] Wanner, S. and Goldluecke, B. “Globally consistent depth labeling of 4D light fields”. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE. 2012, pp. 41–48.
- [62] Wanner, S. and Goldluecke, B. “Reconstructing Reflective and Transparent Surfaces from Epipolar Plane Images”. In: *Pattern Recognition*. Ed. by Weickert, J., Hein, M., and Schiele, B. Vol. 8142. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pp. 1–10. ISBN: 978-3-642-40601-0. DOI: 10.1007/978-3-642-40602-7\_1.
- [63] Ward, G. J. “Measuring and Modeling Anisotropic Reflection”. In: *SIGGRAPH Comput. Graph.* 26.2 (July 1992), pp. 265–272. ISSN: 0097-8930. DOI: 10.1145/142920.134078.
- [64] Woodham, R. J. “Photometric method for determining surface orientation from multiple images”. In: *Optical engineering* 19.1 (1980), pp. 191139–191139.
- [65] Zhang, S., Sheng, H., Li, C., Zhang, J., and Xiong, Z. “Robust depth estimation for light field via spinning parallelogram operator”. In: *Computer Vision and Image Understanding* 145 (2016), pp. 148–159.
- [66] Zisserman, A., Giblin, P., and Blake, A. “The information available to a moving observer from specularities”. In: *Image and Vision Computing* 7.1 (1989), pp. 38–42. ISSN: 0262-8856. DOI: [http://dx.doi.org/10.1016/0262-8856\(89\)90018-8](http://dx.doi.org/10.1016/0262-8856(89)90018-8).