

Thomas Nürnberg

# Entwurf von Computational-Imaging-Systemen am Beispiel der monokularen Tiefenschätzung



Thomas Nürnberg

**Entwurf von Computational-Imaging-Systemen  
am Beispiel der monokularen Tiefenschätzung**

**Forschungsberichte aus der Industriellen Informationstechnik**  
**Band 21**

Institut für Industrielle Informationstechnik  
Karlsruher Institut für Technologie  
Hrsg. Prof. Dr.-Ing. Fernando Puente León

Eine Übersicht aller bisher in dieser Schriftenreihe erschienenen Bände  
finden Sie am Ende des Buchs.



# Entwurf von Computational-Imaging-Systemen am Beispiel der monokularen Tiefenschätzung

von  
Thomas Nürnberg

Karlsruher Institut für Technologie  
Institut für Industrielle Informationstechnik

Entwurf von Computational-Imaging-Systemen am  
Beispiel der monokularen Tiefenschätzung

Zur Erlangung des akademischen Grades eines Doktor-Ingenieurs  
von der KIT-Fakultät für Elektrotechnik und Informationstechnik des  
Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Thomas Nürnberg, M. Sc., geboren in Karlsruhe

Tag der mündlichen Prüfung: 14. Mai 2019

Hauptreferent: Prof. Dr.-Ing. Fernando Puente León, KIT

Korreferent: Prof. Dr.-Ing. habil. Gert F. Trommer, KIT

#### Impressum



Karlsruher Institut für Technologie (KIT)  
KIT Scientific Publishing  
Straße am Forum 2  
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark  
of Karlsruhe Institute of Technology.

Reprint using the book cover is not allowed.

[www.ksp.kit.edu](http://www.ksp.kit.edu)



*This document – excluding the cover, pictures and graphs – is licensed  
under a Creative Commons Attribution-Share Alike 4.0 International License  
(CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>*



*The cover page is licensed under a Creative Commons  
Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0):  
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2019 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 2190-6629

ISBN 978-3-7315-0941-7

DOI 10.5445/KSP/1000095736





# **Entwurf von Computational-Imaging-Systemen am Beispiel der monokularen Tiefenschätzung**

Zur Erlangung des akademischen Grades eines

**DOKTOR-INGENIEURS**

von der KIT-Fakultät für  
Elektrotechnik und Informationstechnik  
des Karlsruher Instituts für Technologie (KIT)  
genehmigte

**DISSERTATION**

von

**Thomas Nürnberg, M. Sc.**

geb. in Karlsruhe

Tag der mündl. Prüfung: 14. Mai 2019  
Hauptreferent: Prof. Dr.-Ing. Fernando Puente León, KIT  
Korreferent: Prof. Dr.-Ing. habil. Gert F. Trommer, KIT



# Vorwort

Diese Arbeit wäre nicht möglich ohne die Hilfe und Unterstützung vieler Kollegen. Allen sei an dieser Stelle gedankt.

Prof. Dr.-Ing. Fernando Puente León danke ich für das Ermöglichen meiner Forschungen als wissenschaftlicher Mitarbeiter am Institut für Industrielle Informationstechnik (IIIT). Dankbar bin ich auch Prof. Dr.-Ing. habil. Gert F. Trommer für die Übernahme des Korreferats und dafür, dass er mich am IIIT empfohlen hat, nachdem ich meine Bachelor- und Masterarbeit an seinem Institut anfertigen durfte.

Mein Dank gilt außerdem allen festangestellten Kolleginnen und Kollegen des Instituts, für die Unterstützung in bürokratischen Angelegenheiten und beim mechanischen und elektrischen Aufbau der während der Arbeit entstandenen Prototypen.

Danken möchte ich auch den vielen Studenten, die ich während meiner Zeit am Institut betreuen durfte. In zahlreichen fachlichen Diskussionen sind wertvolle Beiträge und Ideen für diese Arbeit entstanden.

Bei allen wissenschaftlichen Mitarbeitern möchte ich mich für die gemeinsame Zeit am und außerhalb des Instituts bedanken. Ohne die unkomplizierte Zusammenarbeit und den fachlichen Austausch auch über Themengrenzen hinweg wäre diese Arbeit nicht möglich gewesen. Ich danke den vielen Kollegen, mit denen ich während meiner Zeit das Büro teilen durfte, für die lockere Atmosphäre und insbesondere Sebastian Bauer, Wolfgang Krippner, Maximilian Schambach und Sebastian Vater für das Korrekturlesen der Arbeit.

Zu guter Letzt bin ich meinen Eltern Barbara und Hans Georg und meinem Bruder Andreas zutiefst dankbar für die anhaltende Unterstützung.

Karlsruhe, im Mai 2019

Thomas Nürnberg





# Inhaltsverzeichnis

<b>Symbolverzeichnis</b> . . . . .	<b>v</b>
<b>1 Einleitung</b> . . . . .	<b>1</b>
1.1 Problemstellung . . . . .	3
1.2 Eigener Beitrag . . . . .	5
1.3 Gliederung der Arbeit . . . . .	6
<b>2 Simulativer Kameraentwurf</b> . . . . .	<b>9</b>
2.1 Stand der Technik . . . . .	9
2.2 Formulierung der Simulationsgrößen . . . . .	13
2.3 Bildsynthese nach dem Raytracing-Verfahren . . . . .	15
2.3.1 Kameramodellierung . . . . .	16
2.3.2 Umgebungsmodellierung . . . . .	20
2.3.3 Stärken und Grenzen des Raytracing-Verfahrens . . . . .	22
2.4 Simulation einer Computational Camera . . . . .	24
2.4.1 Rauschquellen . . . . .	25
2.4.2 Geometrische Abbildungsfehler . . . . .	38
2.5 Referenzdaten . . . . .	48
2.6 Kameraoptimierung . . . . .	50
<b>3 Kamera mit programmierbarer Apertur</b> . . . . .	<b>53</b>
3.1 Stand der Technik . . . . .	54
3.1.1 Adaptive Verfahren . . . . .	54
3.1.2 Adaptive Blendencodierung . . . . .	56
3.2 Objektivumbau . . . . .	59
3.3 Objektivcharakterisierung . . . . .	63
3.3.1 Blendengeometrie . . . . .	67
3.3.2 Transmissionseigenschaften . . . . .	69

<b>4</b>	<b>Monokulare Tiefenschätzung</b>	<b>73</b>
4.1	Stand der Technik	75
4.1.1	Allgemeine kamerabasierte Verfahren	75
4.1.2	<i>Depth-from-Defocus</i> -Verfahren	77
4.2	Messprinzip	81
4.3	Algorithmus	85
4.3.1	Signalmodell	87
4.3.2	Identifikation der Skalierung	94
4.3.3	Nachbearbeitung	105
4.4	Blendenoptimierung	111
<b>5</b>	<b>Ergebnisse</b>	<b>119</b>
5.1	Versuchsdurchführung	119
5.2	Ergebnisse realer Aufnahmen	127
5.3	Ergebnisse simulierter Aufnahmen	135
5.4	Fazit	147
<b>6</b>	<b>Zusammenfassung</b>	<b>151</b>
6.1	Ausblick	153
<b>A</b>	<b>Übersicht Raytracing-Simulationsumgebung</b>	<b>157</b>
A.1	Frontend	157
A.2	Kamera	159
A.3	Tracer	160
A.4	Objekt	161
A.5	Material	162
A.6	BRDF/BTDF	163
A.7	Textur	163
A.8	Lichtquelle	165
A.9	Spektrum	166
<b>B</b>	<b>Modifiziertes Objektiv</b>	<b>167</b>
	<b>Literaturverzeichnis</b>	<b>173</b>
	Eigene Veröffentlichungen	183
	Betreute studentische Arbeiten	183

# Symbolverzeichnis

## Allgemeine Abkürzungen

Abkürzung	Bedeutung
bzw.	beziehungsweise
ca.	circa
d. h.	das heißt
engl.	englisch
frz.	französisch
lat.	lateinisch
u. a.	unter anderem
u. U.	unter Umständen
vgl.	vergleiche
z. B.	zum Beispiel
BRDF	bidirektionale Reflektanzverteilungsfunktion, engl. <i>bidirectional reflectance distribution function</i>
BTDF	bidirektionale Transmissionsverteilungsfunktion, engl. <i>bidirectional transmittance distribution function</i>
CCD	engl. <i>charge-coupled device</i>
CIE	Internationale Beleuchtungskommission, frz. <i>Commission Internationale de l'Éclairage</i>
DFT	diskrete Fouriertransformation
EMVA	engl. <i>European Machine Vision Association</i>
ESD	Energiedichtespektrum, engl. <i>energy spectral density</i>
HDR	Hochkontrastbild, engl. <i>high dynamic range</i>
LC	Flüssigkristall, engl. <i>liquid crystal</i>
LCD	Flüssigkristalldisplay, engl. <i>liquid crystal display</i>
LED	Leuchtdiode, engl. <i>light-emitting diode</i>
MAE	mittlerer absoluter Fehler, engl. <i>mean absolute error</i>

Abkürzung	Bedeutung
MAPE	mittlerer absoluter prozentualer Fehler, engl. <i>mean absolute percentage error</i>
OTF	optische Transferfunktion, engl. <i>optical transfer function</i>
PSF	Punktverschmierungsfunktion, engl. <i>point spread function</i>
RDF	engl. <i>Random Decision Forest</i>
SNR	Signal-Rausch-Verhältnis, engl. <i>signal-to-noise ratio</i>
SPI	engl. <i>serial peripheral interface</i>
SSE	Summe der quadrierten Abweichungen, engl. <i>sum of squared errors</i>
SV	Signalverarbeitung
TN	nematische Drehzelle, engl. <i>twisted nematic</i>
USB	engl. <i>universal serial bus</i>

## Symbole

### Lateinische Buchstaben

Symbol	Bedeutung	Einheit
$\mathbf{d}$	Richtungsvektor	
$d_i$	Bildweite, engl. <i>image distance</i>	m
$d_o$	Gegenstandsweite, engl. <i>object distance</i>	m
$E$	Bestrahlungsstärke	$\text{W}/\text{m}^2$
$f$	Brennweite	m
$g(\mathbf{p})$	Skalarwertiges Bild, z. B. Grauwertbild	
$\mathbf{g}(\mathbf{p})$	Mehrkanalbild, z. B. Farbbild	
$H$	Bestrahlung	$\text{J}/\text{m}^2$
$k$	Blendenzahl	
$K$	Systemverstärkung einer Kamera	
$L$	Strahldichte	$\text{W}/\text{m}^2/\text{sr}$
$n$	Brechungsindex	
$\mathbf{n}$	Normalenvektor einer Oberfläche	
$n(\mathbf{p})$	Störbild	
$\mathbf{o}$	Ortsvektor	
$s$	Skalierung	px

Symbol	Bedeutung	Einheit
$S$	Menge der untersuchten Skalierungen	
$S^E$	Energiedichte	
$S$	Empfindlichkeit	
$T$	Belichtungszeit	s
$V_{s,exp}$	Auf Signalmodell basierende Vergleichsgröße	
$V_{obs}$	Auf Beobachtung basierende Vergleichsgröße	
$w_p$	Pixelabmessung	m/px

## Griechische Buchstaben

Symbol	Bedeutung	Einheit
$\eta$	Quantenausbeute	
$\theta_{in}$	Einfallswinkel	
$\theta_{out}$	Ausfallswinkel	
$\theta_r$	Ausfallswinkel bei Reflexion	
$\theta_{tr}$	Transmissionswinkel	
$\lambda$	Wellenlänge	nm
$\mu$	Mittelwert	
$\mu_I$	Dunkelstrom	$e^-/s$
$\sigma^2$	Varianz	
$\Omega_{\frac{1}{2}}$	Raumwinkel der Halbkugel	sr

## Indizes

Index	Bedeutung
$(\bullet)_i$	bildseitig, engl. <i>image side</i>
$(\bullet)_{in}$	einfallend, engl. <i>incident</i>
$(\bullet)_o$	objektseitig, engl. <i>object side</i>
$(\bullet)_{out}$	ausfallend, engl. <i>outgoing</i>
$(\bullet)_{tr}$	transmittiert, engl. <i>transmitted</i>

## Mathematische Operatoren

Operator	Bedeutung
$\langle \bullet, \bullet \rangle$	Allgemeines Distanzmaß
$\langle \bullet, \bullet \rangle_{JS}$	Jensen-Shannon-Divergenz
$\langle \bullet, \bullet \rangle_{KL}$	Kullback-Leibler-Divergenz

# 1 Einleitung

Die Wahrnehmung des Menschen und zahlreicher Tiere basiert häufig in hohem Maße auf dem Sehsinn, mit dem über die Aufnahme von Licht der Umgebung Informationen über das Umfeld gewonnen werden. Schon früh wurde versucht, die im Licht enthaltene Information in Form von beständigen Abbildern festzuhalten und zu konservieren, woraus im 19. Jahrhundert die Fotografie entstand [71]. Die in der Fotografie eingesetzten abbildenden Systeme orientieren sich am Aufbau des menschlichen Auges, wodurch die aufgenommenen Bilder dem menschlichen Seheindruck nahe kommen und unmittelbar interpretierbar sind. Aufgrund der vielfältigen im Licht enthaltenen Information ist dessen Nutzung auch als Informationsquelle eines technischen Systems von hohem Interesse. In der zweiten Hälfte des 20. Jahrhunderts entstand mit der ersten digitalen Kamera [54] die Möglichkeit, Bilddaten in einer maschinennahen Form aufzunehmen und damit für technische Systeme nutzbar zu machen.

Mit der Weiterentwicklung digitaler Rechentechnologien kann heutzutage immer mehr Information in kürzerer Zeit verarbeitet werden. Damit halten Bilddaten als wichtige Informationsquelle immer mehr Einzug in vielfältigen technischen Anwendungsgebieten, wie der Robotik, der Automatisierungstechnik oder der Mensch-Maschine-Interaktion. Dabei ergibt sich häufig die Herausforderung, dass die mit einer Kamera aufgenommenen Bilddaten nicht genügend Informationen zur Lösung der betrachteten Problemstellung enthalten.

Im Licht ist viel mehr Information enthalten als mit einer konventionellen Kamera erfasst werden kann. In der Strahlenoptik kann das Licht mathematisch durch die plenoptische Funktion beschrieben werden, welche die Lichtintensität in Abhängigkeit von sieben Dimensionen angibt: drei Ortsdimensionen, zwei Richtungsdimensionen, der Wellenlänge und der Zeit [112]. Bei der Aufnahme mit einer Kamera wird die plenoptische Funktion auf einen zweidimensionalen Film oder Sensor projiziert, was mit einem beträchtlichen Informationsverlust verbunden ist. Durch einen

Trade-Off zwischen Farb- und Ortsinformation können heute Farbkameras zweidimensionale Abbilder mit drei Farbkanälen zu einem Zeitpunkt aufnehmen. Für viele technische Anwendungen ist dies jedoch nicht ausreichend. Teilweise können Modellannahmen über die beobachteten Szenen getroffen werden, was als zusätzliche Informationsquelle zur Lösung eines Problems beitragen kann. Der Verlust eines Großteils der grundsätzlich im Licht enthaltenen Information kann damit allerdings nicht vollständig kompensiert werden. Dem Informationsverlust kann nur durch Maßnahmen vor der eigentlichen Bildgewinnung begegnet werden.

Mit dem Ziel, bei der Bildgewinnung einen größeren Anteil oder andere Dimensionen der im Licht enthaltenen Information erfassen zu können, sind in den letzten Jahren zahlreiche neuartige Kameraansätze entstanden [32, 63, 64, 81, 92, 105, 112], die unter dem Oberbegriff *Computational Imaging* zusammengefasst werden. Abbildung 1.1 stellt den üblichen Signalfluss dieser Systeme dar. Das Lichtfeld der Szene, symbolisiert durch die Strahldichte  $L$ , wird durch die Kameraoptik verändert und anschließend vom Bildsensor als Bildsignal  $g$  erfasst. Die weitere Signalverarbeitung (SV) findet danach ausschließlich im Digitalen statt. Der digitale Teil der Signalverarbeitungskette ist dabei immer auf den physikalischen Teil angepasst. Auch wenn sich der Begriff *Computational Imaging* erst in der jüngeren Vergangenheit etabliert hat, können zahlreiche schon früher etablierte Verfahren der Bildgewinnung, z. B. die Nutzung einer Bayer-Matrix mit anschließendem Demosaicing heutiger Farbsensoren [10], zu dieser Kategorie gezählt werden.

Durch die grundlegende Änderung des Aufbaus einer Kamera können andere Informationsanteile als die reine Ortsabhängigkeit der Lichtintensität in das erfasste Bild codiert werden. Der Aufbau dieser Systeme orientiert sich somit nicht mehr zwangsläufig an dem eines Auges. Die aufgenommenen Bilder entsprechen daher oftmals nicht mehr dem menschlichen Seheindruck, weshalb sie nicht immer unmittelbar interpretierbar sind. Gerade in technischen Anwendungen ist dies jedoch nicht erforderlich. Hier steht vielmehr der Informationsgehalt der aufgenommenen Bilddaten im Vordergrund. Mit den Möglichkeiten der digitalen Bildverarbeitung und dem Wissen über den genauen Aufbau des abbildenden Systems kann die in den aufgenommenen Bildern codierte Information anschließend extrahiert und nutzbar gemacht werden.



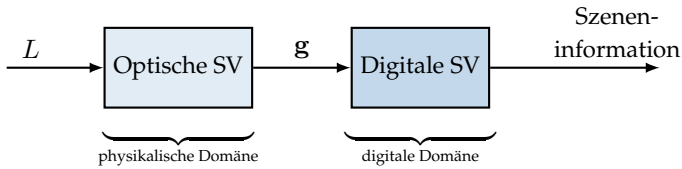


Abbildung 1.1 Signalfluss des *Computational Imaging*.

## 1.1 Problemstellung

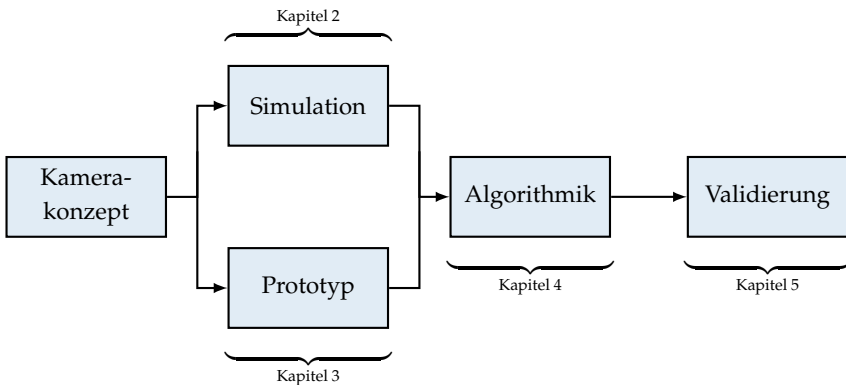
Mit dem Denkmuster des *Computational Imaging* hat sich mit der physikalischen Domäne ein Entwurfsspielraum bei der Konzeption neuartiger Kamerasysteme eröffnet. Die Möglichkeiten der Modifikation klassischer Kamerakonfigurationen reichen von Eingriffen auf Sensorebene über die Beeinflussung der optischen Komponenten einer Kamera bis zu intelligenten Beleuchtungsschemata und Multi-Kamera-Systemen [112]. Unterschiedliche Maßnahmen zur Beeinflussung des physikalischen Prozesses der Bildgewinnung erlauben es, jeweils unterschiedliche Anteile der im Licht enthaltenen Information in das erfasste Bild zu codieren und somit für spezifische Anwendungen nutzbar zu machen.

Aus der beschriebenen Verzahnung zwischen spezifischer Problemstellung sowie optischer und digitaler Signalverarbeitung ergeben sich neue Herausforderungen für den Entwurf neuartiger Kamerasysteme. Eine separate Betrachtung der einzelnen Teile der Signalverarbeitungskette ist nur begrenzt möglich. Vielmehr wird eine ganzheitliche und domänenübergreifende Behandlung der Problemstellung benötigt. Nur so können alle Teile der Kette optimal an das konkret vorliegende Problem angepasst werden.

Durch die Hinzunahme der optischen Signalverarbeitung in den Systementwurf wird auch die prototypische Umsetzung eines Bildverarbeitungssystems erschwert. Während Methoden und Parameter in der digitalen Domäne schnell und mit geringem Aufwand angepasst werden können, erfordern Anpassungen des Aufbaus der Kamera u. U. einen kompletten Neuaufbau eines Prototyps, was mit sehr hohem Zeit- und Arbeitsaufwand verbunden sein kann. Doch gerade das Entwurfsstadium ist oftmals ein iterativer Prozess, bei dem immer wieder Veränderungen

am entworfenen System vorgenommen werden müssen. Darüber hinaus kann es je nach Anwendung sogar notwendig sein, das System adaptiv auf vorherrschende Umgebungsbedingungen anzupassen. Für den digitalen Teil der Kette ist dies wiederum einfach realisierbar. Für den physikalischen Teil werden jedoch flexible Lösungen zur Anpassung ausgewählter Systemparameter benötigt.

Mit der vorliegenden Arbeit sollen neue Entwurfsmethoden für Systeme des Gebiets des *Computational Imaging* erarbeitet werden. Diese Methoden sollen den domänenübergreifenden Charakter dieser Systeme berücksichtigen und deren ganzheitliche Betrachtung ermöglichen. Die Methoden sollen einerseits dabei helfen, auf einfache Weise ein neuartiges Kamerakzept hinsichtlich der Erfüllung einer gewünschten Funktionalität untersuchen zu können, und andererseits Strategien zur konkreten Auslegung und Wahl der Parameter eines prototypischen Aufbaus eröffnen. Darüber hinaus sollen die Methoden beim konkreten Entwurf eines neuartigen Kamerasystems angewendet und validiert werden. Ein Fokus bei der Konzeption des neuen Systems sollen außerdem Methoden der adaptiven oder programmierbaren Gestaltung des physikalischen Teils der Signalverarbeitungskette sein. Sowohl das entworfene Kamerasystem als auch die Entwurfsmethoden sollen anhand einer Analyse der Leistungsfähigkeit des Systems validiert werden.



**Abbildung 1.2** Übersicht über den Entwurfsprozess einer *Computational Camera*.

## 1.2 Eigener Beitrag

Der in dieser Arbeit vorgestellte Entwurfsprozess einer *Computational Camera*, also einer Kamera des Themenfelds *Computational Imaging*, ist in Abbildung 1.2 dargestellt. Vor oder parallel zum Aufbau eines Prototyps kann der physikalische Teil der Signalverarbeitungskette durch eine Simulation ersetzt werden. Anhand der Ausgangsdaten der Simulation kann sowohl der konkrete Aufbau eines Prototyps als auch die nachfolgende digitale Signalverarbeitung entworfen, erprobt und optimiert werden. Die in den einzelnen Teilschritten dieses Entwurfsprozesses erzielten Verbesserungen gegenüber dem Stand der Technik stellen weitere Beiträge der vorliegenden Arbeit dar. Diese werden nachfolgend kurz zusammengefasst:

- Es wurde eine Raytracing-Simulationsumgebung mit dem Fokus auf der strahlenoptisch korrekten Simulation neuartiger Kamerasysteme und der für die Leistungsfähigkeit des Systems relevanten Einflussgrößen implementiert. Dadurch wurde eine ganzheitliche Beschreibung einer *Computational Camera* in der digitalen Domäne ermöglicht. Somit verkürzt sich die Zeit zur prototypischen Umsetzung neuartiger Kamerakonzepte erheblich und es wird die Möglichkeit zur objektiven Bewertung und Optimierung der konkreten Ausführung eines Kamerasystems eröffnet.
- Parallel dazu wurde ein Kameraprototyp mit programmierbarer Apertur durch Nutzung eines transmissiven Flüssigkristalldisplays aufgebaut. Die Kamera zeichnet sich durch die Möglichkeit der berührungslosen Änderung der Aperturform, gute Abbildungseigenschaften, eine kompakte Bauform und flexibel einstellbare Objektivparameter aus.
- Entsprechend dem vorgeschlagenen simulativen Entwurfsprozess wurde ein Algorithmus zur Tiefenschätzung auf Grundlage eines Einzelbildes nach dem *Depth-from-Defocus*-Prinzip anhand simulierter Aufnahmen entwickelt. Weiter wurden damit bezüglich der Unterscheidbarkeit verschiedener Tiefen optimale Blendenformen bestimmt.

- Der vorgeschlagene simulative Entwurfsprozess wurde durch einen Vergleich der qualitativen Ergebnisse realer und simulierter Aufnahmen verschiedener Testszenarien validiert. Bei diesen Untersuchungen hat sich gezeigt, dass in beiden Fällen die Qualität der Ergebnisse stark vom vorherrschenden Bildrauschen beeinflusst wird. Außerdem konnte, übereinstimmend für simulierte und tatsächliche Aufnahmen, eine Abhängigkeit der Ergebnisse von der verwendeten Blendenform nachgewiesen werden. Darüber hinaus wurde eine quantitative Auswertung der Tiefenschätzergebnisse eines auf Einzelbildern basierenden *Depth-from-Defocus*-Algorithmus mit codierten Blenden durchgeführt.

### 1.3 Gliederung der Arbeit

Die vorliegende Arbeit behandelt, abgesehen von der Generierung von Ideen für neuartige Kamerakonzepte, sämtliche Schritte des vorgeschlagenen Entwurfsprozesses einer *Computational Camera* aus Abbildung 1.2. Als Fallbeispiel dient dabei eine Kamera mit programmierbarer Apertur zur monokularen Tiefenschätzung.

Kapitel 2 behandelt die Simulation des Bildgewinnungsprozesses einer Kamera. Neben dem grundlegenden Vorgehen wird dabei insbesondere darauf eingegangen, welche physikalischen Aspekte der Bildgewinnung im Allgemeinen und der digitalen Bildgewinnung im Speziellen berücksichtigt werden müssen und wie diese akkurat simuliert werden können. Außerdem wird in diesem Kapitel darauf eingegangen, wie per Simulation Referenzdaten generiert und wie diese zur Optimierung der Parameter der *Computational Camera* genutzt werden können.

Eine rein simulative Betrachtung ist für den Nachweis der Leistungsfähigkeit eines Bildverarbeitungssystems nicht ausreichend. Die Erkenntnisse einer Simulation können bei der Konzeption und beim Aufbau eines konkreten Prototyps helfen, sie können diesen jedoch nicht ersetzen. In Kapitel 3 wird der im Rahmen dieser Arbeit aufgebaute Prototyp eines Objektivs mit programmierbarer Apertur vorgestellt und charakterisiert.

Nach dem Bildgewinnungsprozess, bei dem beim *Computational Imaging* bereits zusätzliche Information in das Bild codiert wird, muss diese Information in einer digitalen Signalverarbeitung aus dem aufgenommenen

Bild extrahiert werden. Ein Algorithmus zur Extraktion von Tiefeninformation aus einem mit einer codierten Blende aufgenommenen Bild wird in Kapitel 4 vorgestellt.

In Kapitel 5 wird einerseits die Leistungsfähigkeit der entworfenen Signalverarbeitungskette, bestehend aus Kamera und digitaler Signalverarbeitung, anhand verschiedener Testszenarien untersucht. Andererseits wird darauf eingegangen, inwiefern die mit simulierten Aufnahmen erzielten Ergebnisse denen realer Aufnahmen entsprechen. Dieses Kapitel stellt eine Validierung sowohl der in dieser Arbeit vorgestellten Kamera mit programmierbarer Apertur zur monokularen Tiefenschätzung als auch des gesamten simulativen und prototypischen Entwurfsprozesses dar.

Abschließend fasst Kapitel 6 die Ergebnisse der vorliegenden Arbeit zusammen und gibt einen Ausblick auf fortführende Arbeiten.



## 2 Simulativer Kameraentwurf

Beim Entwurf und der Optimierung eines neuartigen Kamerasystems ist es wünschenswert, die Parameter der einzelnen Komponenten der Signalverarbeitungskette zu variieren. Die Anpassung der Komponenten der physikalischen Domäne ist jedoch oftmals mit einem hohen Aufwand verbunden, wenn etwa neue Bauteile angefertigt werden müssen, die den geänderten Anforderungen entsprechen. Insbesondere eine programmierbare und automatisierbare Anpassung ist bei komplexen Optiken praktisch unmöglich. Daher wird in diesem Kapitel vorgestellt, wie die physikalische Seite beim *Computational Imaging* simuliert werden kann. Damit lässt sich, wie in Abbildung 2.1 dargestellt, die gesamte Signalverarbeitungskette ganzheitlich in einer Domäne betrachten. Vorausgesetzt, dass die synthetisch erzeugten Aufnahmen die für die konkrete Anwendung benötigten Charakteristika erfüllen, ist es für den digitalen Teil der Signalverarbeitungskette unerheblich, ob das Eingangsbild einer Simulation oder einer Kamera entstammt. Dadurch kann das gesamte Verfahren vorab erprobt und bewertet werden, ohne dass bereits eine prototypische Implementierung nötig ist. Die in einer Simulation verfügbaren Ground-Truth-Daten können zur Bewertung und Optimierung, insbesondere des physikalischen Teils der Kette, genutzt werden. Dies lässt sich beim Entwurfsprozess einer *Computational Camera* einsetzen.

### 2.1 Stand der Technik

Bereits 1989 simulierten Raczkowsky und Mittenbuehler [80] das Kamerabild eines industriellen Roboters zur Unterstützung des Programmierers bei der Erstellung eines Aktionsplans. Die Autoren nutzen das Verfahren der Rückprojektion und Rasterung zur Erstellung eines Abbilds der modellierten Szene. Zwei Jahre später nutzten Ikeuchi und Robert [35] ebenfalls das Verfahren der Rückprojektion und Rasterung zur Simulation eines

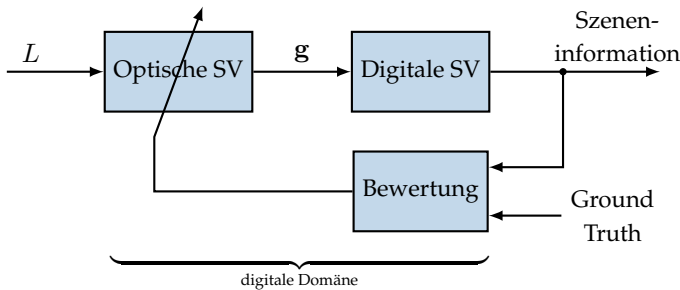


Abbildung 2.1 Signalfluss bei der Simulation einer *Computational Camera*.

Sensorbilds, das bei der Auslegung eines visuellen Inspektionssystems genutzt werden kann. Der Fokus dieser beiden frühen Arbeiten liegt auf der Simulation eines Kamerabilds und nicht auf dessen weiterer Nutzung.

Brady und Gehm [14] simulieren die Bildgewinnung einer Hyperspektralkamera nach dem *Compressive-Sensing*-Prinzip durch idealisierte Faltung eines Eingangsdatensatzes mit einer codierten Blende entlang der Spektralrichtung und anschließender Superposition auf dem angenommenen Sensor. Durch die rein idealisierte Betrachtung der Auswirkung einer codierten Blende auf das Lichtfeld der Kamera können keine strahlenoptischen Störeffekte des abbildenden Systems simuliert werden.

Reiner [83] nutzt die quelloffene Raytracing-Software *Blender* zur Simulation eines Inspektionssystems von Lagern. Anhand der simulierten Aufnahmen kann die Fähigkeit des Inspektionssystems zur Defekterkennung untersucht werden und so beim Entwurf eines realen Systems helfen. Ein Vergleich der Simulationsergebnisse mit einem realen System wurde in der genannten Arbeit jedoch nicht durchgeführt. Diesen Vergleich erbringen Nilsson et al. [69] am Beispiel der visuellen Erkennung des Schweißstoßes zweier Metallplatten. Die Autoren erzielen bei einer einfachen Kantendetektion vergleichbare Ergebnisse mit simulierten und realen Aufnahmen. Zur Simulation wurde eine kommerzielle Software zur Roboterprogrammierung eingesetzt.

Butler et al. [15] verwenden die Rohdaten eines animierten quelloffenen Kurzfilms, um Test-, Trainings- und Referenzdaten für Algorithmen zur Bestimmung des optischen Flusses zu erstellen. Die künstlichen Bilder



wurden nach dem Raytracing-Verfahren mit *Blender* erstellt. Die Verfügbarkeit sämtlicher Ausgangsdaten zur Erstellung des Kurzfilms eröffnet die Möglichkeit der Variation der Simulationskomplexität, um die Robustheit der Algorithmen hinsichtlich einzelner Störeffekte, wie z. B. Defokus, Bewegungsunschärfe oder atmosphärische Effekte, zu beurteilen. Zum Nachweis der praktischen Relevanz der synthetischen Daten vergleichen die Autoren die ersten statistischen Momente der Bilder und Bewegungen mit denen realer Bildsequenzen.

Glatard et al. [28] fassen mehrere Simulationsmethoden medizinischer Bildgebungsverfahren, wie Magnetresonanztomographie, Computertomographie oder Sonographie, in einem einheitlichen Framework zusammen. Zur Simulation von auf Röntgenstrahlung basierenden Verfahren kommen u. a. Raytracing-Ansätze zum Einsatz [94]. Die Herausforderung in diesem Gebiet liegt vor allem in der realitätsnahen Modellierung der zu untersuchenden Objekte und Szenen.

Irgenfried et al. [36] erzeugen synthetische Bild- und Ground-Truth-Daten mit dem Raytracing-Verfahren. Für zwei Beispielanwendungen werden mehrere Verfahren zur Bildsegmentierung sowohl auf den synthetischen Bilddaten als auch auf reale Aufnahmen angewendet. Das Training eines *Random Decision Forests* (RDF) wird ausschließlich mit synthetischen Aufnahmen durchgeführt. Die Fehlerraten der untersuchten Verfahren sind ähnlich für synthetische und reale Aufnahmen. Vor allem bei der Segmentierung mit dem RDF werden mit den simulierten Aufnahmen bessere Ergebnisse erzielt. Die Autoren weisen auf die Wichtigkeit der Nutzung eines physikalisch begründeten Verfahrens zur Bildsynthese hin, gehen allerdings nicht auf die Bildcharakteristiken ein, die mit derartigen Verfahren erzeugt werden können und in welcher Weise die untersuchten Segmentierungsverfahren darauf beruhen. Darüber hinaus verwenden die Autoren selbst kein physikalisch begründetes Störmodell zur Erhöhung der Varianz der synthetisch erzeugten Bilddaten.

Die Auswirkung des Realitätsgrads bei der Erzeugung synthetischer Aufnahmen auf Verfahren der Bildverarbeitung untersucht Meister [61] in seiner Dissertation anhand einer realen und detailgetreu nachgestellten synthetischen Szene am Beispiel der Bestimmung des optischen Flusses. Ein Ergebnis ist, dass die Eignung synthetischer Daten nicht unbedingt vom Realitätsgrad der Szene und der Methoden zur Erzeugung abhängt,

sondern vor allem von der Plausibilität der erzeugten Daten. Es ist wichtig, dass die Phänomene, auf denen ein Verfahren maßgeblich beruht, möglichst realitätsnah simuliert werden. Andere Effekte, die zwar für einen höheren Realismus der erzeugten Bilder sorgen könnten, haben u. U. nur einen geringen Einfluss auf die Ergebnisse der untersuchten Algorithmen. Weiter stellt Meister die hohe Flexibilität und Kontrolle bei der Erzeugung synthetischer Aufnahmen hervor, weshalb mit derartigen Daten gut die Robustheit eines Verfahrens gegenüber Störeinflüssen untersucht werden kann.

Retzlaff et al. [84] erzeugen synthetische Aufnahmen von farbigen Glascherben, die als Trainingsdaten einer automatischen Sortieranlage genutzt werden. Obwohl die Merkmale zur Klassifikation vorab definiert wurden, legen die Autoren hohen Wert auf eine möglichst realitätsnahe Simulation mit dem Raytracing-Verfahren. Angewandt auf Aufnahmen realer Glascherben erreichen die Autoren nahezu gleich hohe Klassifikationsraten mit einem auf synthetischen Aufnahmen trainierten Klassifikator wie mit einem auf realen Aufnahmen trainierten Klassifikator, obwohl sich die Farbstatistiken der synthetisch erzeugten Bilder von denen der realen Aufnahmen unterscheiden.

Birch et al. [12] nutzen ein Raytracing-Framework zur Simulation und Kalibrierung eines linsenlosen Kamerasystems mit zufällig streuendem Element. Die zufällige Streuung der auftreffenden Wellenfronten dient dabei als analoge *Compressive-Sensing*-Codierungsmethode, womit das Bild aus wenigen einzelnen Intensitätsmessungen, also bei einer Abtaststrategie weit unterhalb der Nyquist-Rate, wieder vollständig rekonstruiert werden kann [53].

In den aufgeführten Arbeiten werden simulierte Aufnahmen einerseits zur Unterstützung der Wahl der Szenen- und Kameraparameter genutzt. Andererseits dienen die Aufnahmen zur Verbesserung der Bildverarbeitungsschritte, sei es durch Bereitstellung von Ground-Truth-Daten zur objektiven Bewertung der Leistungsfähigkeit der entworfenen Algorithmen oder zur Erzeugung einer großen Menge annotierter Daten zum Trainieren eines Klassifikators. Der Aufbau der simulierten Kamera selbst ist hingegen selten Gegenstand der Untersuchungen. Bei einer *Computational Camera* ist der konkrete Kameraaufbau jedoch wichtiger Bestandteil der Signalverarbeitungskette und muss daher beim Entwurf eines

Bildverarbeitungssystemen einbezogen werden. Dies erfordert die Erweiterung bestehender Simulationswerkzeuge um neue Kameramodule oder, wie in den folgenden Abschnitten beschrieben, die Implementierung eigener Simulationsumgebungen von Grund auf.

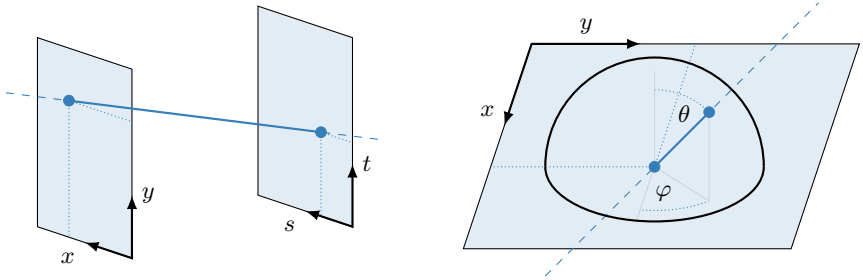
## 2.2 Formulierung der Simulationsgrößen

Die grundlegende physikalische Größe, die im Rahmen der geometrischen Optik den Lichttransport im Raum beschreibt, ist die Strahldichte  $L$ , oftmals auch als Lichtfeld bezeichnet. Sie beschreibt die Strahlungsleistung an einem beliebigen Punkt im Raum, die aus einer bestimmten Einfallrichtung auf ein dazu senkrecht stehendes Flächenelement einfällt. Im Freiraum, ohne dämpfende oder absorbierende Objekte, ist die Strahldichte konstant entlang einer Ausbreitungsrichtung. Daher lässt sich die Geometrie der Strahldichte, die im Allgemeinen fünfdimensional ist (3 Ortsdimensionen, 2 Richtungsdimensionen), im Freiraum über vier Dimensionen beschreiben (2 Ortsdimensionen, 2 Richtungsdimensionen). Die Zeit- und Wellenlängenabhängigkeit der Strahldichte und der damit verwandten Größen werden im Folgenden nur dann explizit angegeben, sofern diese für die jeweiligen Betrachtungen wichtig sind.

In der Literatur werden zahlreiche Parametrisierungen für die 4D-Strahldichte  $L$  verwendet [112], um die Ausbreitungsrichtung zu definieren. Im Rahmen dieser Arbeit werden die folgenden Parametrisierungen verwendet:

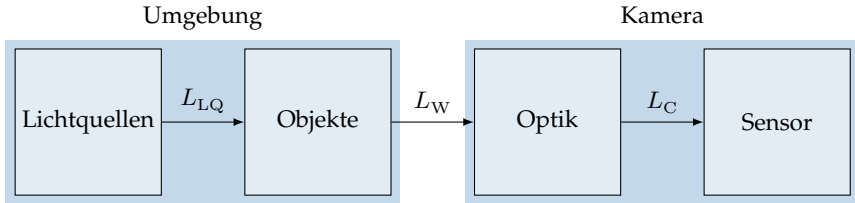
**Ebene-Ebene-Parametrisierung:** Die Strahldichte  $L(x, y, s, t) = L(\mathbf{x}, \mathbf{s})$ , mit  $\mathbf{x} = (x, y)^T$  und  $\mathbf{s} = (s, t)^T$ , gibt die Ausbreitungsrichtung anhand zweier Punkte auf zwei parallel zueinander stehenden Ebenen im Einheitsabstand an (vgl. Abbildung 2.2(a)).

**Ebene-Halbkugel-Parametrisierung:** In der Strahldichte  $L(x, y, \theta, \varphi)$  wird die Ausbreitungsrichtung ausgehend von einem Punkt  $(x, y)$  über die beiden Winkel  $\theta$  und  $\varphi$ , bzw. über  $\boldsymbol{\omega} = (\theta, \varphi)^T$  angegeben (vgl. Abbildung 2.2(b)). In dieser Parametrisierung wird häufig über die Halbkugel über dem betrachteten Punkt integriert, was im Folgenden über den Raumwinkel  $\Omega_{\frac{1}{2}}$  ausgedrückt wird.



(a) Darstellung der Strahldichte  $L(x, y, s, t)$  in Ebene-Ebenen-Parametrisierung. (b) Darstellung der Strahldichte  $L(x, y, \theta, \varphi)$  in Ebene-Halbkugel-Parametrisierung.

**Abbildung 2.2** Unterschiedliche Parametrisierungen der Strahldichte  $L$ .



**Abbildung 2.3** Signalfluss der Strahldichte  $L$  bei der Aufnahme einer Szene mit einer Kamera.

Abbildung 2.3 zeigt schematisch, durch welche Einflüsse die Strahldichte verändert wird. Die Lichtquellen einer Szene erzeugen die Strahldichte  $L_{LQ}$ . An Objektflächen wird das einfallende Licht u. U. mehrfach reflektiert und transmittiert. Aus der Überlagerung der reflektierten und transmittierten Lichtanteile mit der Strahldichte der Lichtquellen resultiert die Strahldichte  $L_W$  der Szene. Die Strahldichte  $L_W$  an der Eintrittspupille der Kamera wird von der Optik zur Strahldichte  $L_C$  innerhalb der Kamera transformiert und dort schließlich vom Sensor bzw. Film<sup>1</sup> erfasst. Die Herausforderung für ein Verfahren zur Simulation der Bildgewinnung ist es, die Abfolge aus Abbildung 2.3 möglichst exakt nachzubilden.

<sup>1</sup> Im Folgenden wird zur besseren Übersichtlichkeit lediglich von einem Digitalsensor ausgegangen. Alle optischen Betrachtungen gelten analog auch bei der Verwendung eines Films.

## 2.3 Bildsynthese nach dem Raytracing-Verfahren

Raytracing ist ein Verfahren der Computergrafik zur Bildsynthese. Der Prozess der Bildberechnung wird auch als *rendern* (engl. *to render*) bezeichnet. Im Gegensatz zum Echtzeitrendern mittels Rückprojektion und Rasterisierung [34], wo es auf hohe Bildraten ankommt und wofür heutige Grafikkarten optimiert sind, bildet Raytracing die physikalischen Prinzipien des Lichttransports aus Abbildung 2.3 ab. Das Verfahren ist damit in der Lage, fotorealistische Bilder zu erzeugen [74]. Die Berechnung eines Bildes kann jedoch, abhängig von der verfügbaren Rechenleistung, eine beträchtliche Zeit von mehreren Tagen in Anspruch nehmen.

Raytracing beruht auf dem Prinzip, dass der Lichttransport im Freiraum konstant bezüglich der Ausbreitungsrichtung ist und dass er sich bei der Reflexion und der Brechung an Grenzflächen rückwärts berechnen lässt. Daher kann die Richtung der Lichtausbreitung nach Abbildung 2.3 umgekehrt werden: Anstatt ausgehend von den Lichtquellen der Szene den Lichteinfall zur Kamera zu betrachten, wird beim Raytracing die Lichtausbreitung rückwärts, ausgehend vom Sensor der Kamera, untersucht. Es wird also nur das Licht betrachtet, das auch tatsächlich auf die Kamera trifft. Anstatt der Berechnung der kontinuierlichen Strahldichte  $L$  werden beim Raytracing lediglich einzelne Strahlen betrachtet. Die Strahldichte wird also durch diskrete Stichproben approximiert. Aus den beiden genannten Eigenschaften leitet sich auch der Name des Verfahrens ab (engl. *ray*, Strahl, engl. *to trace*, verfolgen).

Im Kontext des Kameraentwurfs kann Raytracing grob in die zwei Module, Kamera und Umgebung, aus Abbildung 2.3 unterteilt werden. Dabei wird vom Kameramodul eine Repräsentierung der Strahldichte in der Kamera bestimmt und zur Eintrittspupille zurückverfolgt. Das Modul der Umgebung übernimmt dann die Berechnung der Wechselwirkung der Strahldichte mit der simulierten Szene und verfolgt den Lichtweg bis zu den Lichtquellen zurück. Die resultierende Bestrahlung wird schließlich an das Kameramodul zurückgegeben und dort dem Ursprungspixel beaufschlagt.

Bis auf diesen Datenaustausch sind die Module der Kamera und der Umgebung vollständig unabhängig voneinander. Daher kann auch der

Grad des Realismus für beide Module separat gewählt werden. Da der Fokus dieser Arbeit auf dem Kameraentwurf liegt, wird ein höherer Wert auf die physikalisch korrekte Simulation im Kameramodul gelegt.

### 2.3.1 Kameramodellierung

Bei der Aufnahme mit einer Kamera wird die Bestrahlungsstärke  $E$ , die auf die Sensorebene trifft, ermittelt. Die Bestrahlungsstärke  $E$  am Ort eines Pixels  $\mathbf{p} = (u, v)^T$  ergibt sich aus der Integration der Strahldichte  $L_C$  über alle Raumwinkel  $\Omega_{\frac{1}{2}}$  und der aktiven Pixelfläche  $A_p$ :

$$E_{\mathbf{p}} = \int_{A_p} \int_{\Omega_{\frac{1}{2}}} L_C(\mathbf{x}, \boldsymbol{\omega}) \cos^4(\theta) \, d\boldsymbol{\omega} \, d\mathbf{x}. \quad (2.1)$$

Dabei bezeichnet  $L_C(\mathbf{x}, \boldsymbol{\omega})$  die Strahldichte entlang des Lichtstrahls zu Punkt  $\mathbf{x}$  auf der Sensorebene aus der Raumrichtung  $\boldsymbol{\omega}$ . Der Winkel  $\theta$  entspricht dem Winkel zwischen diesem Strahl und der Normalen der Sensorebene. Am Sensor werden also die Lichtbeiträge aus allen Raumrichtungen  $(\theta, \varphi)$  akkumuliert.

Die Integration über die Belichtungszeit  $T$  liefert schließlich die Bestrahlung  $H_{\mathbf{p}}$  eines Pixels:

$$H_{\mathbf{p}} = \int_0^T E_{\mathbf{p}}(t) \, dt. \quad (2.2)$$

Diese Bestrahlung wird vom Sensor erfasst und weiterverarbeitet. Im Folgenden wird die Szenen als statisch angenommen, d. h. die zeitliche Abhängigkeit in (2.2) kann entfallen und die zeitliche Integration vereinfacht sich zu einer Gewichtung mit der Belichtungszeit  $T$ .

Die Gleichungen (2.1) und (2.2) beschreiben die tatsächlichen physikalischen Vorgänge bei einer fotografischen Aufnahme im vereinfachten Rahmen der geometrischen Optik. Die Schwierigkeit einer Simulation besteht darin, die kontinuierliche Strahldichte  $L_C$  zu bestimmen. Der Ansatz des Raytracing-Verfahrens, Strahldichten durch diskrete Stichproben – den Strahlen – zu approximieren, entspricht der näherungsweise Lösung

der Gleichung (2.1) durch Monte-Carlo-Integration. Nach Übergang in Kugelkoordinaten  $\theta$  und  $\varphi$  ergibt sich mit  $d\omega = \sin(\theta) d\theta d\varphi$ :

$$\hat{E}_{\mathbf{p}} = A_p 2\pi \frac{1}{N} \sum_{i=1}^N L_C(x_i, y_i, \theta_i, \varphi_i) \cos^4(\theta_i) \sin(\theta_i). \quad (2.3)$$

Aufgrund der Eigenschaft des Monte-Carlo-Verfahrens, dass die Konvergenzgeschwindigkeit unabhängig von der Dimension der Integration ist, ist das Verfahren gut geeignet, Gleichung (2.1) numerisch zu lösen [47]. Die Stichproben  $(x_i, y_i)$  und  $(\theta_i, \varphi_i)$  werden dabei zufällig, entsprechend einer Gleichverteilung auf der Pixelfläche  $A_p$  und auf der Halbkugel  $\Omega_{\frac{1}{2}}$  über dem Punkt  $(x_i, y_i)$  gezogen. Mit steigender Anzahl der Stichproben  $N$  konvergiert (2.3) gegen die tatsächliche Bestrahlungsstärke (2.1) [91].

Wird die Strahldichte anstelle der Ebene-Halbkugel-Parametrisierung  $L_C(x_i, y_i, \theta_i, \varphi_i)$  in der Ebene-Ebene-Parametrisierung  $L_C(x_i, y_i, s_i, t_i)$  angegeben, wobei die  $x$ - $y$ -Ebene der Sensorebene und die  $s$ - $t$ -Ebene der Linsenebene entspricht, ergibt sich

$$\hat{E}_{\mathbf{p}} = \frac{A_p \pi D^2}{4} \frac{1}{d_i^2} \frac{1}{N} \sum_{i=1}^N L_C(x_i, y_i, s_i, t_i) \cos^4(\theta_i), \quad (2.4)$$

mit

$$\theta_i = \arctan\left(\frac{\sqrt{s_i^2 + t_i^2}}{d_i}\right) \quad (2.5)$$

und der Bildweite  $d_i$ , wobei

$$\|\mathbf{s}_i\| = \|(s_i, t_i)^T\| < \frac{D}{2} \quad (2.6)$$

gilt. Dabei wurde ausgenutzt, dass die Stichproben  $(s_i, t_i)$  nicht aus der unendlichen Linsenebenen gezogen werden müssen, sondern lediglich aus der endlichen Eintrittspupille der Kamera, die hier als kreisförmig mit Durchmesser  $D$  angenommen wird. Der Vorfaktor  $d_i^{-2}$  folgt aus dem radiometrischen Grundgesetz [10] und ist notwendig, da die beiden Referenzebenen nicht im Abstand eins, sondern in der Bildweite  $d_i$  voneinander entfernt liegen.

Im Kameramodul werden zunächst, ausgehend von der simulierten aktiven Sensorfläche, Strahlen generiert. Hauptaufgabe des Moduls ist dann, ausgehend von der aktiven Sensorfläche den Verlauf der Strahlen, also z. B. Brechung an Linsen oder Abdunkelung an Blenden, bis zum Austritt in die Kamera zu berechnen. Die Strahlen werden dabei durchgängig durch ein Tupel  $(\mathbf{o}, \mathbf{d})$  beschrieben, das eine Geradengleichung der Form

$$\mathbf{x} = \mathbf{o} + t \mathbf{d}, \quad (2.7)$$

mit dem Ortsvektor  $\mathbf{o}$  und dem Richtungsvektor  $\mathbf{d}$ , beschreibt. Die Berechnung kann vollständig in kamerafesten Koordinaten erfolgen. Nachdem feststeht, dass ein Strahl die Kamera verlässt, wird er nach Transformation in das Weltkoordinatensystem an das Umgebungsmodul übergeben.

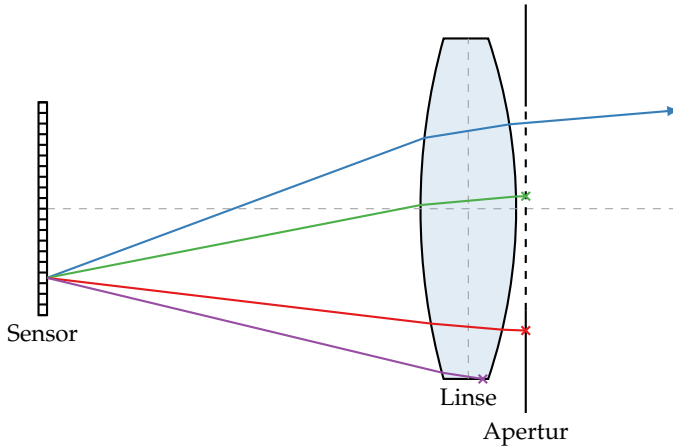
Das Vorgehen zur Berechnung eines Lichtstrahls im Kameramodul wird im Folgenden exemplarisch anhand einer Kamera mit dicker sphärischer Linse und einer codierten Blende, die objektseitig vor der Linse platziert ist, vorgestellt. Abbildung 2.4 zeigt einen Querschnitt der Anordnung. Bei der Wahl der Simulationskomplexität ist es wichtig, dass die für die spezifische Anwendung nötigen Merkmale simuliert werden [61]. So ist es beispielsweise zur Erprobung des Verfahrens zur Vergrößerung der Schärfentiefe nach Cossairt und Nayar [21] unumgänglich, chromatische Aberrationen zu simulieren.

Für eine dicke Linse nach Abbildung 2.4 findet die Berechnung eines Strahls schrittweise statt:

1. Bildseitigen Anfangspunkt  $\mathbf{o}_i$  auf aktiver Pixelfläche entsprechend einer Gleichverteilung zufällig wählen.
2. Richtungsvektor  $\mathbf{d}_i$  des bildseitigen Strahls durch zufällige Wahl eines Punktes in Linsenebene bestimmen.
3. Schnittpunkt des bildseitigen Strahls mit bildseitiger Linsenoberfläche berechnen.
4. Richtungsvektor  $\mathbf{d}_L$  innerhalb der Linse nach Brechung an bildseitiger Linsenoberfläche gemäß

$$\mathbf{d}_L = \frac{1}{n_L} \mathbf{d}_i - \left( \cos(\theta_{tr}) - \frac{1}{n_L} \cos(\theta_{in}) \right) \mathbf{n}_i \quad (2.8)$$





**Abbildung 2.4** Mögliche Strahlen in einer Kamera am Beispiel eines Objektivs mit einer einzelnen dicken sphärischen Linse.

mit

$$\cos(\theta_{\text{in}}) = \mathbf{n}_i \cdot (-\mathbf{d}_i) \quad (2.9)$$

und

$$\cos(\theta_{\text{tr}}) = \sqrt{1 - \frac{1}{n_L^2} (1 - \cos^2(\theta_{\text{in}}))} \quad (2.10)$$

berechnen [27]. Dabei bezeichnet  $\mathbf{n}_i$  die Normale der bildseitigen Linsenoberfläche am Schnittpunkt mit dem Strahl,  $n_L$  den Brechungsindex des Linsenmaterials und  $\theta_{\text{tr}}$  den Transmissionswinkel.

5. Schnittpunkt des Strahls innerhalb der Linse mit objektseitiger Linsenoberfläche berechnen.
6. Richtungsvektor  $\mathbf{d}_o$  nach Brechung an objektseitiger Linsenoberfläche analog zu (2.8)–(2.10) berechnen, wobei  $n_L^{-1}$  anstelle von  $n_L$  und die Normale  $\mathbf{n}_o$  der objektseitigen Linsenoberfläche eingesetzt wird.
7. Schnittpunkt des objektseitigen Strahls mit Aperturebene bestimmen.

8. Strahl entsprechend der Aperturfunktion  $m(\mathbf{x}) \in [0, 1]$  dämpfen.
9. Resultierendes objektseitiges Strahl tupel  $(\mathbf{o}_o, \mathbf{d}_o)$  in Weltkoordinaten umrechnen und an Umgebungsmodul übergeben.

Auf diese Weise wird im Zuge der Simulation mindestens ein, zumeist jedoch deutlich mehr Strahlen pro Pixel berechnet.

Wie in Abbildung 2.4 dargestellt, kann es vorkommen, dass ein Strahl die Kamera nicht verlässt, weil er entweder nicht auf die objektseitige Linsenoberfläche oder einen durchlässigen Bereich der Blende trifft. Diese Fälle müssen bei der Durchführung der Schritte 1–9 erkannt und nicht dem Umgebungsmodul übergeben werden.

### 2.3.2 Umgebungsmodellierung

Bei der Modellierung der Strahldichte in der Szene ist vor allem die Wechselwirkung des Lichts mit Objektflächen von Interesse. Die sogenannte Rendergleichung

$$L_{\text{out}}(\mathbf{x}, \boldsymbol{\omega}_{\text{out}}) = L_{\text{em}}(\mathbf{x}, \boldsymbol{\omega}_{\text{out}}) + \int_{\Omega_{\frac{1}{2}}} BRDF(\mathbf{x}, \boldsymbol{\omega}_{\text{in}}, \boldsymbol{\omega}_{\text{out}}) L_{\text{in}}(\mathbf{x}, \boldsymbol{\omega}_{\text{in}}) \cos(\theta_{\text{in}}) d\boldsymbol{\omega}_{\text{in}} \quad (2.11)$$

beschreibt die Strahldichte  $L_{\text{out}}$  eines von einer Oberfläche ausgehenden Lichtstrahls. Diese setzt sich aus der vom Material emittierten Strahldichte  $L_{\text{em}}$  und den in die betrachtete Richtung reflektierten Lichtanteilen zusammen. Den Zusammenhang zwischen einer einfallenden Strahldichte  $L_{\text{in}}$  und der reflektierten Strahldichte stellt die bidirektionale Reflektanzverteilungsfunktion (engl. *bidirectional reflectance distribution function*, kurz BRDF) her. Durch Integration über die Hemisphäre  $\Omega_{\frac{1}{2}}$  über der Oberfläche werden alle einfallenden Lichtbeiträge berücksichtigt.

Das an einem Punkt  $\mathbf{x}$  einer Oberfläche einfallende Licht, in (2.11) beschrieben durch die Strahldichte  $L_{\text{in}}$ , kann nicht aus dem Nichts entstehen, sondern entstammt anderen Objekten der Szene. Von dort breitet es sich im Freiraum zwischen den Objekten aus und fällt auf den betrachteten Punkt  $\mathbf{x}$ . Die einfallende Strahldichte  $L_{\text{in}}$  ist also nach dem Zurückverfolgen durch den Freiraum gleichzeitig die ausfallende Strahldichte  $L_{\text{out}}$

eines anderen Punktes. Diese ausfallende Strahldichte  $L_{\text{out}}$  kann ebenfalls mit der Rendergleichung bestimmt werden. Durch rekursives Anwenden der Rendergleichung lässt sich so die Lichtausbreitung in der Szene bis zu den Lichtquellen zurückverfolgen bzw. bis die Lichtbeiträge aufgrund von Mehrfachreflexion vernachlässigbar klein sind.

Analog zu Abschnitt 2.3.1 wird Gleichung (2.11) beim Raytracing-Verfahren durch Monte-Carlo-Integration gelöst: Die Integration über die Hemisphäre  $\Omega$  an der Oberfläche wird durch Einsetzen zufälliger Stichproben  $\omega_{\text{in},i}$  approximiert. Aufgrund des rekursiven Charakters stellt die Lösung der Rendergleichung für eine Szene auch mit dieser Vereinfachung weiterhin eine große Herausforderung dar und erfordert einen hohen Berechnungsaufwand. Daher sind oftmals weitere Annahmen und Vereinfachungen nötig, u. a. [91]:

- Die Rekursionstiefe bei der Anwendung der Rendergleichung (2.11) kann beschränkt werden. Insbesondere bei der Beschränkung auf nur eine Reflexion müssen lediglich Einfallrichtungen betrachtet werden, in denen Lichtquellen vorhanden sind. Dieser Anteil an der Beleuchtung wird *direkte* Beleuchtung genannt. Lichtanteile, die nicht direkt von Lichtquellen auf eine Oberfläche fallen, sondern erst von anderen Oberflächen reflektiert werden, werden als *indirekte* Beleuchtung bezeichnet.
- Durch Annahme einer allgegenwärtigen, in allen Raumrichtungen konstanten Ausleuchtung ist bei der Berechnung der direkten Beleuchtung die einfallende Strahldichte  $L_{\text{in}}(\mathbf{x}, \omega_{\text{in}})$  in Gleichung (2.11) konstant und kann vor das Integral gezogen werden.
- Bei Annahme von Punktlichtquellen ohne Ausdehnung und einer allgegenwärtigen Beleuchtung in lediglich einer Ausbreitungsrichtung vereinfachen sich bei der Berechnung der direkten Beleuchtung die einfallenden Strahldichten  $L_{\text{in}}(\mathbf{x}, \omega_{\text{in}})$  in (2.11) zu  $\delta$ -Distributionen, was das unmittelbare Lösen des Integrals ermöglicht.
- Die BRDF eines Lambert'schen Strahlers ist unabhängig von der Ein- und Ausfallsrichtung  $\omega_{\text{in}}$  und  $\omega_{\text{out}}$ . Somit ist die BRDF bezüglich des Integrals konstant.

Die genannten Vereinfachungen beeinflussen sowohl den Realitätsgrad der Umgebungssimulation als auch den Rechenaufwand maßgeblich. Sofern eine hohe Rekursionstiefe bei der Anwendung der Rendergleichung und ansonsten keine der oben genannten Vereinfachungen angenommen wird, spricht man von der Simulation einer *globalen* Beleuchtung. Diese Art der Beleuchtungssimulation wird bei der Erzeugung fotorealistischer Bilder verwendet, der Berechnungsaufwand kann jedoch enorm hoch sein. Es muss analog zur Kameramodellierung sichergestellt sein, dass die für die jeweilige Anwendung notwendigen Umgebungsmerkmale durch die Vereinfachungen nicht verloren gehen [61]. Die in Kapitel 4 vorgestellte monokulare Tiefenschätzung ist beispielsweise auf ein breitbandiges Anregungssignal angewiesen, das durch texturierte Objekte erzeugt wird. Die Simulation einer einfachen allgegenwärtigen und ausschließlich direkten Beleuchtung ist dafür ausreichend. Werden spiegelnde Oberflächen untersucht, ist im Gegensatz dazu die Simulation der indirekten Beleuchtung unbedingt erforderlich.

Eine weitere Maßnahme zur Reduktion des Rechenaufwands bei der Monte-Carlo-Integration der Rendergleichung (2.11) ist das Ziehen der Stichproben aus einer Kosinus-Verteilung anstelle einer Gleichverteilung. Da der Integrand in (2.11) proportional zu  $\cos(\theta_{\text{in}})$  ist, kann dadurch, sofern die korrekte Normierung sichergestellt ist, die Konvergenzgeschwindigkeit der Approximation erhöht werden. Dies entspricht dem *Importance Sampling* bei der Monte-Carlo-Integration [47].

### 2.3.3 Stärken und Grenzen des Raytracing-Verfahrens

Aufgrund des Ursprungs in den physikalischen Gesetzmäßigkeiten des Lichttransports ist das Raytracing-Verfahren in der Lage, Licht in einer Szene oder einer Kamera im Sinne der geometrischen Optik detailliert zu simulieren. Die Ergebnisse sind bei einer detaillierten Modellierung der Szene für menschliche Betrachter oftmals nur schwer von realen Aufnahmen zu unterscheiden. Daher kommt Raytracing heute beispielsweise auch häufig im Bereich der Architektur und dem industriellen Lichtdesign zum Einsatz [11, 103, 106].

Raytracing ist als strahlbasiertes Verfahren insbesondere zur Simulation strahlenoptischer Effekte geeignet, z. B. der Brechung des Lichts an Linsen und Prismen oder der Reflexion an spiegelnden Oberflächen. Eine Vielzahl

der Ansätze zur Konstruktion neuartiger *Computational Cameras* basieren auf derartigen Komponenten und Effekten [1, 3, 49, 65, 68, 112].

Trotzdem können mit Raytracing auch wellenoptische Phänomene simuliert werden, wodurch der Berechnungsaufwand jedoch nochmals deutlich ansteigen kann. Durch die Approximation des kontinuierlichen Lichtspektrums durch Stichproben in der Monte-Carlo-Integration können neben Effekten, die sich in einer reinen Wellenlängenabhängigkeit der Intensität der Strahldichte äußern, auch geometrische Wellenlängenabhängigkeiten wie Dispersion modelliert werden. Der Polarisationszustand eines Lichtstrahls, also die Ausrichtung der Schwingungsebene des elektrischen und magnetischen Felds, kann beispielsweise mit einem Stokes-Vektor [10] repräsentiert werden. Damit wird es insbesondere möglich, das Reflexionsverhalten dielektrischer Materialien oder Polarisationskameras [99] zu simulieren. Auch Beugung und Interferenz können in eine Raytracing-Simulation einbezogen werden [4, 70]. Trotz dieser Möglichkeiten zählt die Simulation wellenoptischer Phänomene nicht zu den Stärken des Raytracing-Verfahrens. Sofern sich bei der konkret betrachteten Kamera und Sichtprüfaufgabe die Abmessungen nicht in der Größenordnung der Wellenlänge befinden, ist die Vernachlässigung wellenoptischer Phänomene gerechtfertigt.

Auch wenn Raytracing die strahlenoptischen Gesetzmäßigkeiten korrekt nachbildet, bleibt das Ergebnis aufgrund der Diskretisierung der kontinuierlichen Strahldichten stets eine Approximation. Der Erwartungswert des Betrags des Approximationsfehlers der klassischen Monte-Carlo-Integration ist mit

$$E \left\{ \left| E - \hat{E} \right| \right\} \leq \frac{\sigma_E}{\sqrt{N}} \quad (2.12)$$

unabhängig von der Dimension des Integrals [47]. Trotzdem erfordern Szenen mit komplexer Beleuchtung aufgrund des langsamen Anwachsens der Wurzelfunktion in (2.12) eine sehr hohe Anzahl  $N$  an Stichproben. Der Approximationsfehler äußert sich in Bildrauschen oder Artefakten, die in Abschnitt 2.4.1 näher betrachtet werden.

Insbesondere bei sehr kleinen Skalen können die Simulationsergebnisse neben der Vernachlässigung wellenoptischer Phänomene auch durch Fehler aufgrund einer unzureichenden Präzision digital gespeicherter Zahlenwerte von der Realität abweichen. Dies kann selbst bei Simulation

in 64-Bit-Genauigkeit auftreten. Geringe Quantisierungsfehler, insbesondere im Richtungsvektor  $d$  eines Strahls, können beim Verfolgen über längere Distanzen dafür sorgen, dass sichtbare Abweichungen entstehen. Quantisierungsfehler der Intensität der Strahldichten können hingegen vernachlässigt werden, da die Quantisierung eines Digitalsensors um ein Vielfaches gröber als eine simulierte Intensität mit Gleitkommazahlen ist.

## 2.4 Simulation einer Computational Camera

Bei der im Rahmen der vorliegenden Arbeit entstandenen Raytracing-Simulationsumgebung liegt der Fokus auf der realitätsnahen und aussagekräftigen Simulation von *Computational Cameras*. Aus diesem Grund wurde sich auf eine einfache Umgebungsmodellierung bestehend aus simplen geometrischen Formen beschränkt. Sofern für eine Anwendung komplexe Szenengeometrien benötigt werden, können diese als Polygonmodelle erstellt und geladen werden. Es können zahlreiche verschiedene Kameras modelliert und simuliert werden. Diese reichen von einfachen Lochkameras oder einer Kamera mit idealisierter Zentralprojektion über Kameras mit einer Linse und beliebig platzierten Blenden bis zu einer Lichtfeldkamera mit Mikrolinsenarray nach Ng et al. [68]. Aufgrund ihres modularen Aufbaus kann die Simulationsumgebung mit geringem Aufwand um neue Kameratypen erweitert werden. Dazu muss der Strahlengang innerhalb der Kamera definiert werden, indem ein Methode implementiert wird, die die Parameter  $(o, d)$  der die Kamera verlassenden Strahlen sowie deren Ursprung auf dem Sensor berechnet. Für vom Sensor ausgehende Strahlen, die die Kamera nicht verlassen, muss die Methode explizit den Lichtbeitrag Null zurückgeben.

Die grundlegende C++-Implementierung der Simulationsumgebung orientiert sich an den Büchern von Glassner [27], Hughes et al. [34], Saffern [91] sowie Pharr und Humphreys [74]. Eine Übersicht über die Programmstruktur und die verfügbaren Module ist in Anhang A angegeben. Der Fokus dieser Bücher liegt auf einem Einstieg in das Thema Computergrafik und Raytracing oder auf der Erzeugung fotorealistischer Bilder und nicht auf der Nutzung von Raytracing als Entwurfs- und Simulationswerkzeug einer *Computational Camera*. Daher wird im Folgenden beschrieben, welche weiteren Maßnahmen notwendig sind und im Zuge der Entwicklung der

Simulationsumgebung ergriffen wurden, um mittels Raytracing belastbare Aussagen für den Entwurf einer neuartigen *Computational Camera* treffen zu können.

Cossairt et al. [20] haben für *Defocus Deblurring* mit örtlich codierten Blenden, für *Motion Deblurring* mit zeitlich codierten Blenden und für blendenbasierte Lichtfeldaufnahmen gezeigt, dass das Verbesserungspotenzial von *Computational Cameras* gegenüber konventionellen Maßnahmen, z. B. der Verwendung einer kleineren Blende oder einer kürzeren Belichtung, auch von den äußeren Bedingungen und damit vom Störszenario abhängt. In der Entwurfsphase kann durchaus zum Nachweis der grundlegenden Funktion bei der Simulation eines neuen Kamerakonzepts auf derartige Störeinflüsse verzichtet werden. Aber um später einen quantitativen Vergleich der Leistungsfähigkeit mit bestehenden Ansätzen zu ermöglichen, müssen zur Sicherstellung der Aussagekräftigkeit der Ergebnisse negative Störeinflüsse bei einer Simulation mit berücksichtigt werden.

Bei der Aufnahme mit einem realen Kamerasystem wirken stets verschiedene Störeinflüsse auf das Bild ein. Gleichzeitig entstehen in der Simulation Approximationsfehler, die ebenfalls als Störung im Bild modelliert werden können. Diese simulationsbedingten Störungen dürfen gegenüber den natürlichen Störquellen nicht ins Gewicht fallen.

Die Störeinflüsse lassen sich grob in Rauschen, dessen Ursprung im Allgemeinen am Bildsensor liegt, und Abbildungsfehler, die durch die Optik vor dem Sensor entstehen, unterteilen. In den nachfolgenden Abschnitten wird betrachtet, wie sich die einzelnen Einflüsse auf das Bild auswirken und wie sie in der Raytracing-Simulationsumgebung modelliert werden.

### 2.4.1 Rauschquellen

Zunächst werden im folgenden Abschnitt die unterschiedlichen Rauschquellen einer realen Kamera modelliert. Außerdem wird betrachtet, wie sich diese möglichst präzise in einer Simulation nachbilden lassen. Danach wird auf die zusätzlich durch die Approximation der Simulation erzeugten Störungen eingegangen und es werden Maßnahmen betrachtet, deren Auswirkung zu verringern. Anschließend werden die Rauscheigenschaften realer und simulierter Aufnahmen miteinander verglichen und bewertet.

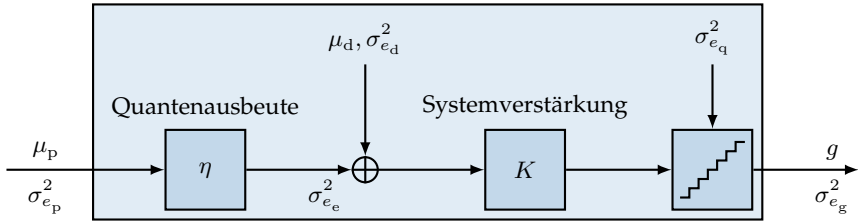


Abbildung 2.5 Kameramodell des EMVA-1288-Standards (Bild nach Jähne [39]).

### 2.4.1.1 Reale Rauschquellen

Einer realen Aufnahme ist stets Rauschen unterschiedlicher Quellen überlagert. Diese Quellen liegen an unterschiedlichen Schritten des Bildgewinnungsprozesses. Das vorhandene Bildrauschen hängt somit auch vom konkret verwendeten Sensor ab. Je nach nötigem Detailgrad und je nach Sensortyp wurden unterschiedliche Modelle zur Charakterisierung des Bildrauschens entwickelt [33, 97, 102]. Im Folgenden werden die Rauschquellen anhand des Kameramodells des EMVA-1288-Standards [24] aus Abbildung 2.5 beschrieben.

Zunächst weist das zu erfassende Lichtsignal selbst Rauschcharakter auf, da die Anzahl der am Sensorelement eintreffenden Photonen näherungsweise einer Poisson-Verteilung folgt [10]. Die Dichtefunktion, die angibt, mit welcher Wahrscheinlichkeit genau  $k$  Photonen auf ein Sensorpixel treffen, lautet daher

$$f_p(k) = \frac{\mu_p^k}{k!} e^{-\mu_p} \quad (2.13)$$

mit dem Erwartungswert  $\mu_p$ . Die als Rauschen wahrnehmbare Varianz  $\sigma_{e_p}^2$  entspricht bei der Poisson-Verteilung genau dem Erwartungswert  $\mu_p$ . Die vom Sensor absorbierten Photonen regen mit der Wahrscheinlichkeit  $\eta$ , die als Quantenausbeute bezeichnet wird, Elektronen an. Auch die Anzahl der angeregten Elektronen folgt einer Poisson-Verteilung mit Mittelwert  $\mu_e$  und Varianz  $\sigma_{e_e}^2$ . Damit ist die Rauschleistung proportional zur Lichtintensität bzw. zum Bildwert. Aufgrund des Ursprungs in der zufälligen Anzahl an Photonen wird dieses Rauschen als Photonengeräusch bezeichnet.



Eine weitere Rauschquelle ist die Kameraelektronik. Dieses ist grundsätzlich unabhängig von der Signalintensität und wird daher auch als Dunkelrauschen bezeichnet. Damit wirkt sie additiv auf das Bildsignal. Die Varianz  $\sigma_{e_d}^2$  ist somit konstant bezüglich der Signalintensität. Das Dunkelrauschen ist insbesondere bei schwachen Lichtintensitäten sichtbar, wenn das Photonenrauschen gering ist. Der Erwartungswert der zusätzlich angeregten Elektronen  $\mu_d$  setzt sich zusammen aus einem konstanten Rauschanteil der Kameraelektronik (A/D-Umsetzer, Verstärker) und einem bezüglich der Belichtungszeit  $T$  linearen Anteil [24]:

$$\mu_d = \mu_{d0} + \mu_I T. \quad (2.14)$$

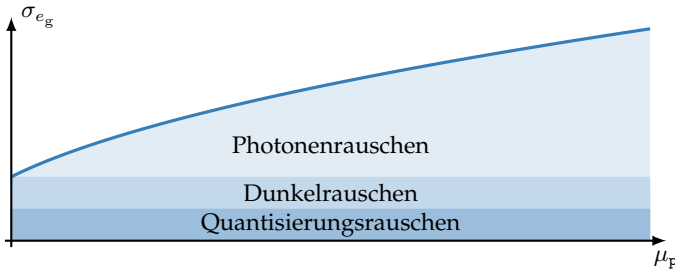
Der konstante Anteil  $\mu_{d0}$  wird als Ausleserauschen (engl. *read noise*) bezeichnet. Die Größe  $\mu_I$  beschreibt den Dunkelstrom, der entsteht, da durch thermische Anregung auch ohne Belichtung freie Ladungsträger im Sensor erzeugt werden. Dieser Prozess kann wiederum durch einen Poisson-Prozess modelliert werden. Insgesamt ergibt sich die Varianz des Dunkelrauschens damit zu

$$\sigma_{e_d}^2 = \sigma_{e_{d0}}^2 + \mu_I T. \quad (2.15)$$

Die Ladung der angeregten Elektronen bewirkt eine Spannung  $U$  über einem Kondensator, die verstärkt und digitalisiert wird, um den digitalen Grauwert  $g$  zu erhalten. Um den Dynamikbereich des Eingangssignals möglichst gut auf den verfügbaren Wertebereich bei der Speicherung des Bildes, also 0 bis  $2^N - 1$  bei einer Auflösung von  $N$  bit, abzubilden, bieten heutige Digitalkameras die Möglichkeit, das Signal nach dieser A/D-Umsetzung zu verstärken und so die Empfindlichkeit zu verändern<sup>2</sup>. Dies findet jedoch ausschließlich im Digitalen statt und hat daher keine Auswirkung auf die Rauscheigenschaften der Kamera. Da diese Verstärkung jedoch in die Systemverstärkung  $K$  eingeht, wird im Folgenden stets davon ausgegangen, dass das Signal im Digitalen nicht weiter verstärkt wird.

Der bei der Digitalisierung entstehende Quantisierungsfehler  $e_q$  kann als weitere Störquelle modelliert werden [78]. Unter Annahme eines linearen

<sup>2</sup> Bei der analogen Fotografie entspricht dies der Empfindlichkeit des Films.



**Abbildung 2.6** Exemplarischer Verlauf der Standardabweichung des Rauschens einer Kamera.

Quantisierungsmodells und gleichverteilter Signalamplituden beträgt die Varianz des Rauschprozesses

$$\sigma_{e_q}^2 = \frac{q^2}{12}, \quad (2.16)$$

mit der Breite der Quantisierungsstufen  $q$ . Damit ist es ebenfalls unabhängig von der Signalintensität.

Da die Rauschprozesse voneinander unabhängig sind, ergibt sich, nach der Umrechnung in die digitale Domäne, die Varianz der resultierenden Störung des Grauwerts  $\sigma_{e_g}^2$  aus der Summe der Varianzen der einzelnen Rauschprozesse:

$$\sigma_{e_g}^2 = K^2 (\sigma_{e_e}^2 + \sigma_{e_d}^2) + \sigma_{e_q}^2. \quad (2.17)$$

Die physikalisch begründeten Rauschanteile werden also bei der A/D-Umsetzung mit dem Faktor  $K^2$  verstärkt. Daher ist bei der klassischen Fotografie das Rauschen insbesondere bei schwachen Lichtbedingungen wahrnehmbar, wenn das Bild zur Verbesserung des Kontrasts nachträglich verstärkt wird. Abbildung 2.6 veranschaulicht die Standardabweichung des resultierenden Rauschens in Abhängigkeit von der Lichtintensität.

In der im Rahmen dieser Arbeit aufgebauten Raytracing-Simulationsumgebung stehen hauptsächlich strahlenoptische Aspekte und weniger die Radiometrie im Vordergrund. Aus diesem Grund werden Intensitäten nicht in radiometrischen Größen, sondern direkt in Bildintensitäten

definiert. Damit müssen keine geeignete Belichtungszeit und Systemverstärkung  $K$  gewählt werden, um ein unmittelbar interpretierbares Bild zu erzeugen. Allerdings wird so die Domäne, auf die sich das Photonen- und das Dunkelrauschen unmittelbar auswirken, gar nicht simuliert. Gleichung (2.17) liefert jedoch eine Vorschrift, wie nach der Wahl einer Verstärkung  $K$  die Abweichungen in der Einheit der digitalen Grauwerte umgerechnet werden können. Diese Verstärkung wirkt sich dann nur auf die Störung und nicht auf das ideale Bildsignal aus.

Für das Photonen-, das Dunkel- und das Quantisierungsrauschen ist die Störgröße eines Pixels unabhängig von den Nachbarpixeln. Daher können diese Störungen in der Simulation als örtlich unkorreliert angenommen werden und damit den Pixeln einzeln beaufschlagt werden.

Die Anzahl der Photonen bzw. der angeregten Elektronen, also das Eingangssignal, folgt der Poisson-Verteilung (2.13). Für ausreichend große  $\mu_p$  ( $\mu_p > 25$  [23]) konvergiert diese Verteilung gegen eine Normalverteilung mit Mittelwert  $\mu_p$  und Varianz  $\mu_p$ . Da zur Aussteuerung einer Kamera, sofern diese nicht speziell zur Messung extrem niedriger Intensitäten ausgelegt ist, stets eine Vielzahl von Photonen nötig ist, ist eine Approximation der Poisson-Verteilung durch eine Normalverteilung meist gerechtfertigt. Daher kann zur Simulation des Photonenrauschens in der digitalen Domäne der Störterm  $n_p$  aus einer Normalverteilung gezogen werden:

$$n_p \sim \mathcal{N}\left(0, K^2 g\right). \quad (2.18)$$

Das Dunkelrauschen wird in der Simulation als mittelwertbehaftetes weißes Gauß'sches Rauschen angenommen, mit dem Erwartungswert des Dunkelrauschens  $\mu_d$  als Mittelwert, umgerechnet in Grauwertintensitäten. Die Amplitudenvarianz des Rauschens wird anhand der EMVA-Eigenschaften des zu simulierenden Sensors und der Belichtungszeit  $T$  gewählt:

$$n_d \sim \mathcal{N}\left(K\mu_I T, K^2\left(\sigma_{e_{d0}}^2 + \mu_I T\right)\right), \quad (2.19)$$

wobei  $\mu_{d0}$  aus (2.14) zu null angenommen wird. Hierbei wird die Poisson-Verteilung der durch den Dunkelstrom vorhandenen Ladungsträger wieder durch eine Normalverteilung approximiert.

Zur Simulation der Quantisierung werden die Intensitäten  $g$  als Gleitkommazahlen geführt. Zur Ausgabe des Bildes wird dann die Quantisierung in der gewünschten Bitauflösung durchgeführt. Damit wird dem Intensitätswert  $g$  implizit das Quantisierungsrauschen mit der Varianz  $\sigma_q^2$  hinzugefügt. Selbst bei Simulation der Intensitätswerte in einfacher Genauigkeit (32 bit) ist damit die Genauigkeit um ein Vielfaches höher als die Bitauflösung gängiger Digitalsensoren (8 bis 16 bit). Daher wird das Quantisierungsrauschen durch dieses Vorgehen sehr gut approximiert.

Für den unwahrscheinlichen Fall, dass der simulierte Grauwert  $g$  nach dem Hinzufügen des Photonen- und des Dunkelrauschens negativ ist, wird dieser bei der Quantisierung auf 0 gesetzt.

Örtlich korrelierte Störungen des EMVA-1288-Standards [24], wie Inhomogenitäten des Dunkelrauschens (engl. *dark signal nonuniformity*, kurz: DSNU) oder Inhomogenitäten der Empfindlichkeit (engl. *photo response nonuniformity*, kurz: PRNU), werden in der Simulation nicht betrachtet.

### 2.4.1.2 Approximationsfehler bei der Simulation

Die Aufnahme einer Szene mit einem Digitalsensor entspricht dem örtlichen Abtasten des Lichtfelds  $L_C$  in der Kamera. Bei dieser Abtastung tritt daher Aliasing auf, sofern das Abtasttheorem nicht erfüllt ist, also wenn die örtliche Abtastfrequenz nicht mindestens doppelt so groß wie die Bandbreite des Lichtfelds  $L_C$  ist. So kann bei der scharfen Abbildung periodischen Strukturen in Abhängigkeit der Entfernung Aliasing auftreten. Darüber hinaus treten z. B. an scharf abgebildeten Objektkanten, insbesondere bei der Beschränkung auf die geometrische Optik, Signalanteile sehr hoher Ortsfrequenz auf. Derartige Szenen verursachen sowohl bei realen als auch bei simulierten Aufnahmen Aliasing, weshalb hier zur realitätsnahen Simulation keine weiteren Maßnahmen ergriffen werden müssen.

Ein wichtiger Unterschied zwischen der Aufnahme mit einer realen Kamera und einer simulierten Aufnahme mit dem Raytracing-Verfahren ist jedoch, dass die Integration der Strahldichte  $L$  über die Pixelfläche und die Raumrichtung im realen Fall kontinuierlich stattfindet, wohingegen das Integral beim Raytracing durch Monte-Carlo-Integration gelöst wird, also ebenfalls abgetastet wird. Bei dieser zusätzlichen Abtastung können

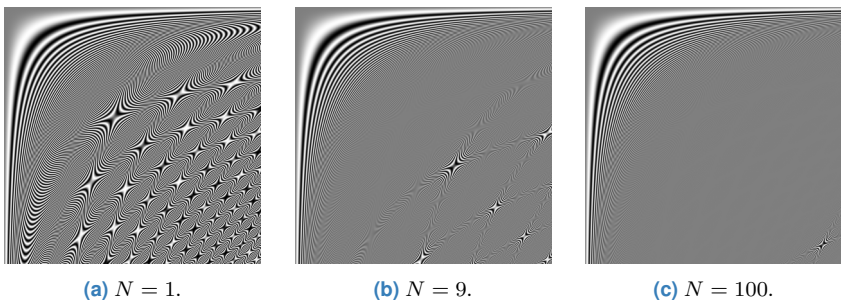
Aliasing-Fehler entstehen, die sich in realen Aufnahmen nicht ergeben. In Abbildung 2.7(a) ist die simulierte Aufnahme einer gemäß [91]

$$g(x, y) = \frac{1}{2} \left( 1 + \sin(x^2 y^2) \right) \quad (2.20)$$

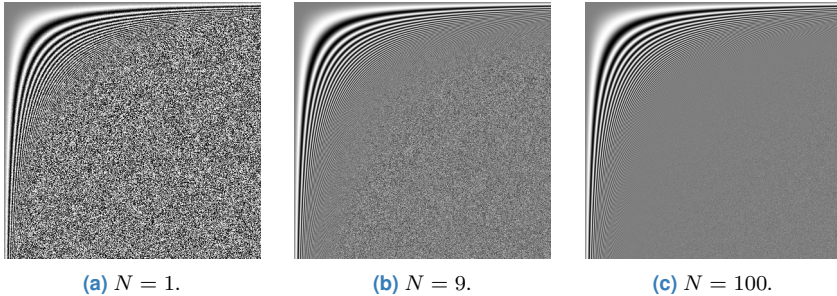
texturierten Ebene mit einer Lochkamera dargestellt. Dabei wurde kein Rauschen simuliert. Durch die hohen Frequenzen entsteht Aliasing, das als Moiré sichtbar wird. Die Erhöhung der Anzahl  $N$  simulierter Strahlen pro Pixel bringt nur bedingt Abhilfe. In Abbildung 2.7(c) wurden  $N = 100$  Strahlen pro Pixel simuliert, trotzdem ist in der Abbildung unten rechts noch ein deutliches Aliasing-Artefakt erkennbar.

Eine zur Erhöhung der Abtastfrequenz alternative Strategie, diesem Problem zu begegnen, ist die Pixelfläche und die Raumrichtungen nicht regelmäßig, sondern zufällig gleichverteilt abzutasten. Abbildung 2.8 zeigt dieselbe Szene für unterschiedliche  $N$  mit zufälligen Abtastpunkten nach dem *Multi-Jittered*-Verfahren [91], wiederum ohne simuliertes Photon- oder Dunkelrauschen. Anstelle des Moirés tritt hier Rauschen auf, das umso stärker ist, je höher die Frequenz der Szene wird. Analog zur regelmäßigen Abtastung können durch Erhöhung der Abtastfrequenz die Approximationsfehler reduziert werden.

Diese Beobachtungen entsprechen dem Verhalten des Approximationsfehlers nach (2.12) aus Abschnitt 2.3.3. Je größer die Variation der Szene ist, desto größer ist der Fehler aus der Monte-Carlo-Approximation. Diese können durch Erhöhung der Stichprobenanzahl  $N$ , also der Anzahl simulierter



**Abbildung 2.7** Simulierte Aufnahme einer texturierten Ebene für unterschiedliche Anzahlen  $N$  simulierter Strahlen pro Pixel mit regelmäßiger Abtastung.



**Abbildung 2.8** Simulierte Aufnahme einer texturierten Ebene für unterschiedliche Anzahlen  $N$  simulierter Strahlen pro Pixel mit zufälliger Abtastung.

Strahlen pro Pixel, reduziert werden, allerdings wird bei höheren  $N$  die Zunahme der Approximationsqualität aufgrund der Wurzelabhängigkeit immer geringer.

Damit die Auswirkungen dieser Approximation die Aussagekraft der per Simulation erzielten Ergebnisse nicht beeinflussen, muss sichergestellt werden, dass die Störszenarios realer Aufnahmen vergleichbar mit denen simulierter Aufnahmen sind. Sowohl bei regelmäßiger als auch bei zufälliger Abtastung werden die Aliasing-Fehler durch Erhöhung der Anzahl simulierter Strahlen  $N$  verringert, was allerdings mit einer Zunahme der Berechnungszeit verbunden ist. Ganz verhindern lassen sich die Approximationsfehler nicht, da die Bandbreite der Szenen, bei der Betrachtung der geometrischen Optik, im Allgemeinen nicht beschränkt ist und daher das Abtasttheorem nicht eingehalten werden kann. Vor diesem Hintergrund ist es vorzuziehen, eine zufällig gleichverteilte Abtastung vorzunehmen. Bei der regelmäßigen Abtastung sind die Approximationsfehler örtlich stark konzentriert. Die Lage der Artefakte ist jedoch nicht vorhersehbar, weshalb eine Kompensation nicht möglich ist. Dies kann eine weitere Signalverarbeitung in diesen Bildbereichen erschweren oder sogar unmöglich machen. Mit der zufällig gleichverteilten Abtaststrategie führen die Approximationsfehler zu zusätzlichem Bildrauschen. Damit entsprechen sie im Charakter der in der Realität auftretenden Störungen. Durch die Wahl der Anzahl simulierter Strahlen  $N$  kann die Rauschleistungsdichte der Approximationsfehler so gewählt werden, dass diese gegenüber dem Photonen- und dem Dunkelrauschen nicht ins Gewicht fallen.

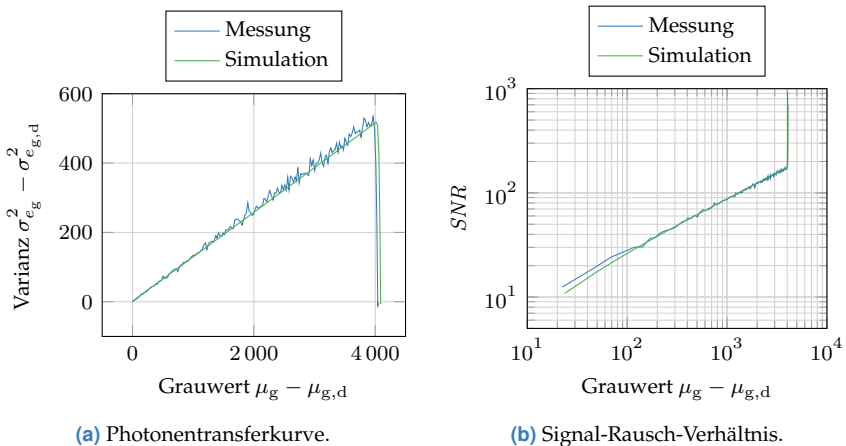
**Tabelle 2.1** Rauschparameter der *Prosilica GT4907C* CCD-Kamera ([\*]: aus Messungen nach der Photonentransfermethode [24, 40] bestimmter Parameter).

Ausleserauschen	$\sigma_{e_{d0}}$	$12 e^-$
Dunkelstrom	$\mu_I$	$145 e^-/s$
Systemverstärkung*	$K$	$3,243 \cdot 10^{-5} \text{ DN} / e^-$

### 2.4.1.3 Vergleich der Eigenschaften realer und simulierter Aufnahmen

Im Folgenden sollen die Eigenschaften des Bildrauschens der Aufnahmen der Raytracing-Simulation mit den Eigenschaften einer realen CCD-Kamera verglichen werden. Dazu wird eine *Prosilica GT4907C* Kamera von *Allied Vision* verwendet. Die relevanten Parameter des Rauschverhaltens der Kamera sind in Tabelle 2.1 angegeben. Diese werden entsprechend auch in der Simulation verwendet.

In Abbildung 2.9 werden die Rauscheigenschaften der Aufnahmen der realen CCD-Kamera mit den Eigenschaften simulierter Aufnahmen verglichen. In beiden Fällen wurde eine homogene Szene abgebildet, um Aliasing zu vermeiden. Abbildung 2.9(a) zeigt die Photonentransferkurve,



**Abbildung 2.9** Vergleich der Rauscheigenschaften realer und simulierter Aufnahmen anhand einer homogenen Szene.

anhand derer sich die Systemverstärkung  $K$  der Kamera bestimmen lässt [40]. Im Diagramm ist die Varianz des Bildsignals, kompensiert um die Varianz des Dunkelrauschens, über dem um den Mittelwert des Dunkelsignals bereinigten Mittelwert des Signals aufgetragen. Zur Bestimmung des Mittelwerts und der Varianz des Dunkelrauschens wurden für jede Signalintensität zusätzlich ein Dunkelbild bei identischer Belichtungszeit aufgenommen bzw. simuliert.

Bei fast allen Verhältnissen zeigt sich der lineare Zusammenhang zwischen der Signalintensität und der Varianz des Photonenrauschens. Die Sättigung ist als das Maximum der Photonentransferkurve definiert. Ab dieser Intensität werden Rauschanteile durch den begrenzten Wertebereich des A/D-Umsetzers abgeschnitten, weshalb die Varianz des Bildes abnimmt. Bei noch stärkerer Beleuchtung fällt die Varianz zu Null ab, da der A/D-Umsetzer dann immer voll angesteuert wird.

Abbildung 2.9(b) zeigt den zugehörigen Verlauf des Signal-Rausch-Verhältnisses (SNR), das sich aus dem Quotient des Mittelwerts der am Sensorpixel einfallenden Photonen und der Standardabweichung der verschiedenen Rauschquellen ergibt:

$$SNR = \frac{\eta\mu_p}{\sqrt{\sigma_{e_d}^2 + \sigma_{e_q}^2/K^2 + \eta\mu_p}}. \quad (2.21)$$

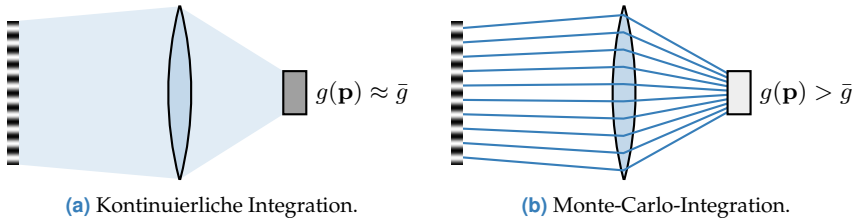
Mit der bestimmten Systemverstärkung  $K$  stimmen in beiden Diagrammen sowohl die Werte als auch die Form der gemessenen und simulierten Kurven sehr gut überein. Das angenommene Modell aus Abbildung 2.5 ist also bei den beschriebenen Bedingungen sehr gut zur Beschreibung des Störverhaltens einer realen Kamera und damit für die realitätsnahe Simulation des Bildrauschens einer Kamera geeignet.

Als Nächstes soll das Störverhalten anhand einer Szene untersucht werden, bei der sich mit einer Raytracing-Simulation Approximationsfehler durch die Monte-Carlo-Integration über die Strahldichte ergeben. Dazu wird eine Ebene mit einer hochfrequenten Sinustextur gemäß

$$g(x, y) = \frac{1}{2} (1 + \sin(2\pi f x)), \quad \text{mit } f \gg \frac{1}{2} f_A, \quad (2.22)$$

abgebildet. Der Mittelwert der Textur liegt somit bei  $\bar{g} = \frac{1}{2}$ . Bei der simulierten Aufnahme wird der Wertebereich der Grauwerte eines Pixels





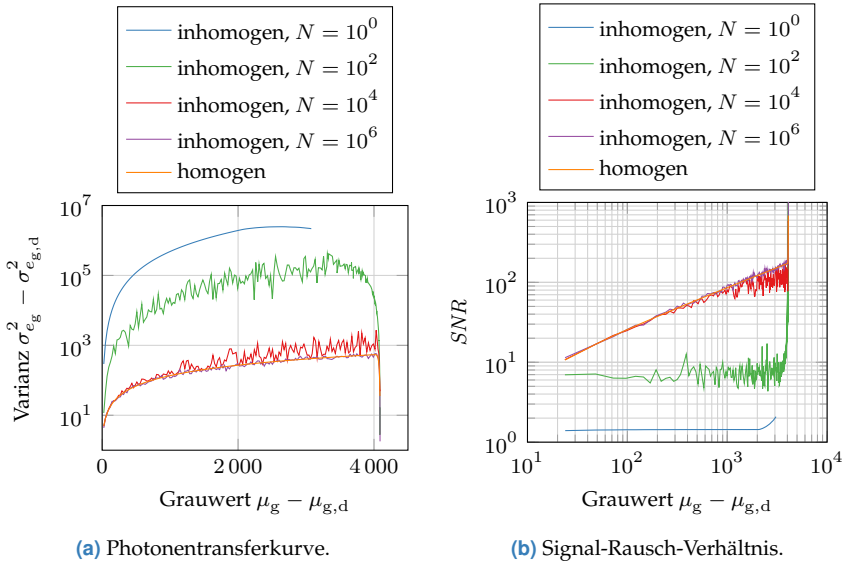
**Abbildung 2.10** Szenenkonfiguration zur Untersuchung der Approximationsfehler einer Raytracing-Simulation.

$g(\mathbf{p}) \in [0, 1]$  dann auf den Wertebereich der realen Kamera  $g(\mathbf{p}) \in [0, 4095]$  skaliert.

Sowohl mit einer realen Kamera als auch in der Simulation ist bei einer scharfen Abbildung mit Aliasing zu rechnen. Wird die Szene jedoch unscharf abgebildet, wird die Auswirkung der Monte-Carlo-Integration bei der Simulation deutlich. In diesem Fall sind mit einer realen Kamera Grauwerte nahe des Mittelwerts der Sinustextur zu erwarten, da, wie in Abbildung 2.10(a) dargestellt, pro Pixel über die Lichtbeiträge einer Vielzahl von Sinusperioden integriert wird und unvollständig beobachtete Perioden daher nicht ins Gewicht fallen. Das zu erwartende Störverhalten realer Aufnahmen dieser Szene entspricht somit dem der homogenen Szene aus Abbildung 2.9.

Bei der Monte-Carlo-Integration mit einer regulären Abtastung können dagegen trotz der unscharfen Abbildung Moiré-Artefakte aus der Interferenz der regelmäßigen Abtastpunkte mit der regelmäßigen Sinustextur entstehen, wie Abbildung 2.10(b) verdeutlicht. Der ermittelte Grauwert  $g(\mathbf{p})$  kann daher stark vom zu erwartenden mittleren Grauwert  $\bar{g}$  abweichen. Diese Abweichungen sind die Approximationsfehler der Raytracing-Simulation und werden im Folgenden als Störgröße angesehen.

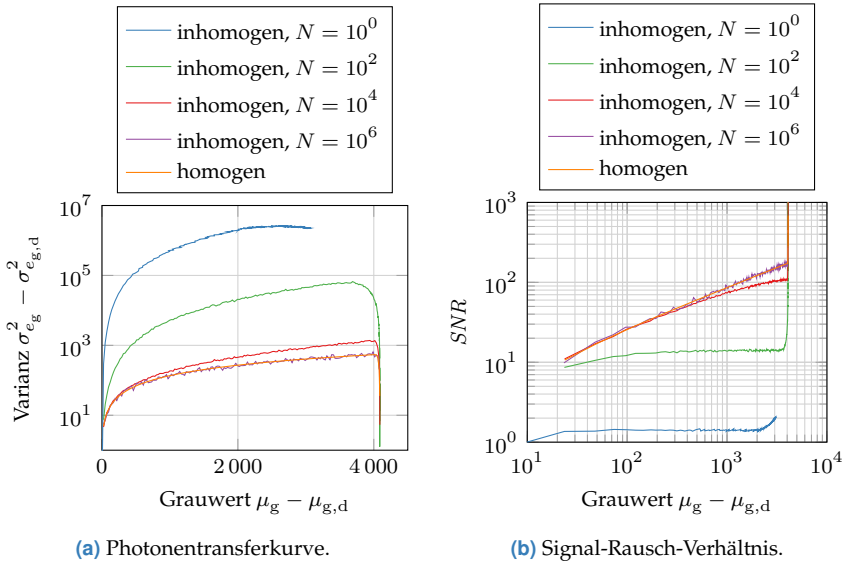
In der Photonentransferkurve aus Abbildung 2.11(a), die mit der beschriebenen Szene mit regelmäßiger Abtastung simuliert wurde, äußern sich die Approximationsfehler der Monte-Carlo-Integration in einer im Vergleich zur homogenen Szene teilweise deutlich höheren Varianz. Es wurde eine logarithmische Skala gewählt, um die Varianzen für verschiedene Anzahlen simulierter Strahlen  $N$  überhaupt gemeinsam darstellen zu können. Darüber hinaus wird bei niedrigen  $N$  die Sättigung früher



**Abbildung 2.11** Vergleich der Rauscheigenschaften simulierter Aufnahmen mit regelmäßiger Abtastung anhand einer unscharfen Abbildung einer Szene mit hoher Bandbreite.

erreicht, da Teile der entstehenden Moiré-Artefakte schon früh abgeschnitten werden. Aufgrund der dunklen Bildbereiche in den Moiré-Artefakten erreicht der mittlere Grauwert auch nicht den maximalen Grauwert der simulierten Kamera. Durch die Erhöhung der Strahlenanzahl pro Pixel auf  $N = 10^6$  können diese Effekte in diesem Fall zwar vollständig kompensiert werden, es kann jedoch bei der regelmäßigen Abtastung grundsätzlich immer zu Interferenzen kommen. Gleichzeitig ist der SNR-Gewinn durch ein höheres  $N$  sehr gering (vgl. Abbildung 2.11(b)).

In Abbildung 2.12 sind die Ergebnisse derselben breitbandigen Szene dargestellt, jedoch mit einer zufälligen Abtastung nach dem *Multi-jittered*-Verfahren. Bei geringen Strahlenzahlen  $N$  ergibt sich ebenfalls eine erhöhte Varianz in der Photonentransferkurve. Die Approximationsfehler äußern sich hier jedoch nicht als Moiré, sondern als zusätzliches Bildrauschen. Dieses Bildrauschen kann durch Simulation einer größeren Anzahl Strahlen pro Pixel  $N$  ebenfalls reduziert werden, allerdings konvergiert das SNR bei der zufälligen Abtastung schneller gegen den Wert der homogenen



**Abbildung 2.12** Vergleich der Rauscheigenschaften simulierter Aufnahmen mit zufälliger Abtastung anhand einer unscharfen Abbildung einer Szene mit hoher Bandbreite.

Szene, was besonders an den Kurven für  $N = 10^2$  deutlich wird. Die Übereinstimmung mit den Eigenschaften der homogenen Szene wurde, wie im Fall der regelmäßigen Abtastung, für  $N = 10^6$  erreicht. Darüber hinaus ist zu beobachten, dass die Störeeigenschaften bei zufälliger Abtastung wesentlich weniger über die korrigierte Signalintensität  $\mu_g - \mu_{g,d}$  variieren. Das bedeutet, dass ein bestimmtes Störszenario mit einer zufälligen Abtastung wesentlich präziser eingestellt werden kann als mit einer regelmäßigen Abtastung. Dies spiegelt die Unvorhersehbarkeit der Moiré-Artefakte bei regelmäßiger Abtastung wider.

Aus den obigen Untersuchungen lässt sich das Fazit ziehen, dass sich eine Raytracing-Simulation so erweitern lässt, dass die Rauscheigenschaften realer Kameras sehr gut simuliert werden können. Die Störungen durch simulationsbedingte Approximationsfehler lassen sich durch die Wahl der Abtaststrategie und die Erhöhung der Anzahl simulierter Strahlen vermindern bzw. sogar vollständig kompensieren. Allerdings nimmt dabei der Berechnungsaufwand deutlich zu. Die Approximationsfehler einer

zufälligen Abtastung weisen Rauschcharakter auf und nehmen mit ansteigender Anzahl Strahlen schneller ab als bei einer regelmäßigen Abtastung. Aus diesen Gründen wird im Folgenden die zufällige Abtastung nach dem *Multi-Jittered*-Verfahren bevorzugt.

Da die Stärke der Approximationsfehler in hohem Maße von der simulierten Szene abhängen, lässt sich keine feste Aussage über die notwendige Anzahl zu simulierender Strahlen treffen. Zwar nehmen die Approximationsfehler mit steigender Anzahl Strahlen ab, die Berechnungsdauer wächst dagegen jedoch überproportional an. Die Simulation von  $N = 10^6$  Strahlen pro Pixel wie in den Abbildungen 2.11 und 2.12 kann schon bei relativ kleinen Bildern von z. B.  $200 \times 200$  Pixeln und lediglich direkter Beleuchtung auf einem üblichen Desktop-PC bereits mehrere Stunden in Anspruch nehmen. Jedoch erfordern nur ausgewählte Szenen, wie die untersuchte Sinustextur oder beispielsweise auch die Impulsantworten des folgenden Abschnitts, derartig hohe Anzahlen von Strahlen. Abhängig von der Szene sind häufig wesentlich weniger Strahlen ausreichend und nur in Einzelfällen muss ein Kompromiss aus Simulationsgenauigkeit und Berechnungsdauer gefunden werden.

### 2.4.2 Geometrische Abbildungsfehler

Die Aufgabe des Objektivs einer Kamera ist es, das Licht der Szene zu sammeln und diese auf den Sensor abzubilden. Diese optische Abbildung kann jedoch zahlreichen Abbildungsfehlern unterliegen, welche deren Qualität maßgeblich beeinflussen können. Bei der Herstellung eines Objektivs muss daher ein hoher Aufwand betrieben werden, z. B. durch Verwendung mehrerer Linsengruppen oder Achromaten, um in möglichst allen Fällen eine gute Abbildung der Szene zu ermöglichen.

Sphärische Linsen sind einfacher als asphärische Linsen zu fertigen [10] und kommen daher oft zum Einsatz. Eine Vielzahl der Abbildungsfehler lassen sich auf die sphärische Form von Linsen zurückführen. Im Folgenden wird eine Übersicht über die Ursache und die Auswirkungen einiger Abbildungsfehler gegeben. Anschließend wird darauf eingegangen, wie diese Effekte in einer Raytracing-Simulation erzeugt werden können.

### 2.4.2.1 Abbildungsfehler sphärischer Linsen

Die sogenannten sphärischen Aberrationen sind darauf zurückzuführen, dass sich, wie in Abbildung 2.13 dargestellt, auf eine sphärische Linse parallel einfallende Strahlen nicht exakt in einem Brennpunkt schneiden. Insbesondere Strahlen, die weit von der optischen Achse entfernt auf die Linse treffen, werden stärker gebrochen. Die sphärische Aberration äußert sich dadurch, dass keine perfekt scharfe Abbildung möglich ist, da ein Teil des eintreffenden Lichts immer unscharf abgebildet wird. Die Punktverschmierungsfunktion (PSF) ist aufgrund der sphärischen Aberration nicht impulsförmig wie in Abbildung 2.17 für  $\Delta d_i = 0$ , sondern hat stets, unabhängig von der Fokussierung, eine deutliche kreisförmige Ausdehnung.

Der Effekt der Koma tritt ebenfalls bei sphärischen Linsen auf. Die Koma bezeichnet das Phänomen, dass bei schrägem Lichteinfall die PSF keine symmetrische Intensitätsverteilung mehr aufweist, sondern, wie in Abbildung 2.14, asymmetrisch verzerrt wird. Je größer der Winkel zwischen dem einfallenden Strahlbündel und der optischen Achse ist, desto stärker ausgebildet ist die Koma. Die Form der PSF hängt damit also auch von der Position im Bild ab, am Bildrand ist der Effekt der Koma am stärksten.

Astigmatismus bezeichnet den Effekt, dass sich bei schiefem Lichteinfall für Strahlbündel unterschiedlicher Schnittebenen unterschiedliche Brennweiten ergeben. Dies liegt daran, dass sich bei schrägem Lichteinfall die Krümmungsradien der Linsenoberfläche verschiedener Schnittebenen unterscheiden. Wie in Abbildung 2.15 dargestellt, schneiden sich die Strahlen

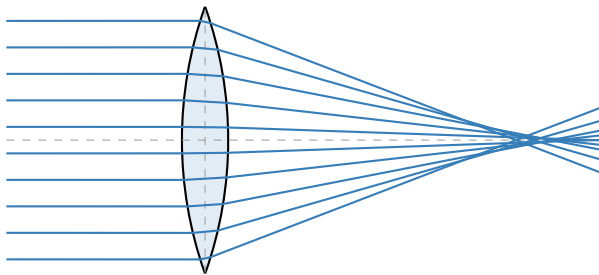


Abbildung 2.13 Sphärische Aberration.

in der Meridionalebene (grün) vor den Strahlen in der Sagittalebene (rot). Die Form der PSF variiert also am Bildrand auch in Abhängigkeit der Bildweite.

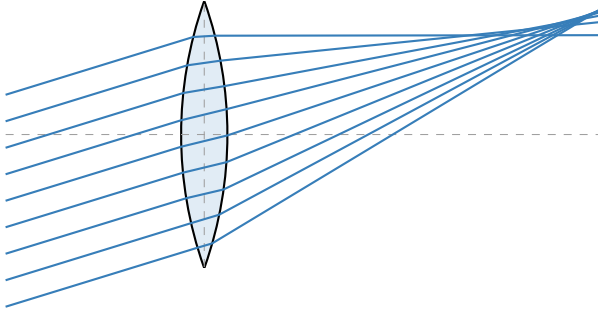


Abbildung 2.14 Koma.

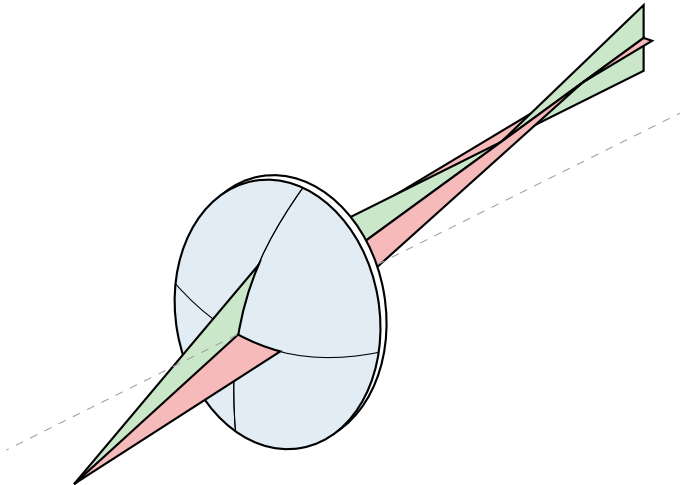


Abbildung 2.15 Astigmatismus.

Der Brechungsindex optischer Materialien hängt aufgrund der Dispersion im Allgemeinen von der Wellenlänge des Lichts ab ( $n_L = n_L(\lambda)$ ). Somit werden Strahlen unterschiedlicher Wellenlängen an einer Linse unterschiedlich stark gebrochen. Daher wird einfallendes Licht mehrerer Wellenlängen durch eine Linse, wie in Abbildung 2.16 dargestellt, in die einzelnen spektralen Anteile aufgetrennt, was in einer Aufnahme zu Farbfehlern, insbesondere an Objektkanten, führt. Diese Art von Abbildungsfehlern nennt man chromatische Aberration. Bezogen auf die Impulsantwort einer Kamera bewirkt sie, dass die PSF zusätzlich von der Wellenlänge  $\lambda$  abhängt bzw. dass sich bei einer Beleuchtung mit mehreren Wellenlängen die PSFs der einzelnen Wellenlängen überlagern.

### 2.4.2.2 Vignettierung

Neben den Abbildungsfehlern einer Linse treten in realen Objektiven weitere störende Effekte auf. Die optimale Position der Aperturblende liegt in der Linsenebene, was sich bei einem System mit nur einer Linse nicht realisieren lässt. Daher muss die Aperturblende stets vor oder hinter der Linse platziert werden. Eine objektseitig vor der Linse platzierte Aperturblende wirkt als Eintrittspupille. Die zugehörige Austrittspupille ergibt sich aus der Abbildung der Eintrittspupille auf die Bildseite. Entsprechend lässt sich eine bildseitig platzierte Aperturblende auf die Objektseite transformieren [10].

Durch die Ein- und Austrittspupille werden die Öffnungswinkel der ein- und ausfallenden Strahlen beschränkt. Licht, das unter einem zu steilen Winkel auf ein Objektiv trifft, wird von einer der Pupillen abgeschattet.

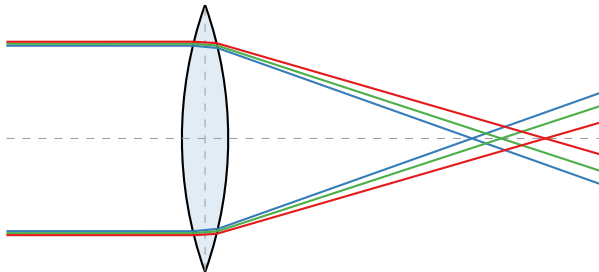


Abbildung 2.16 Chromatische Aberration.

Randbereiche des Sensors erfassen das schräg einfallende Licht, das von der Eintritts- bzw. Austrittspupille teilweise blockiert wird. Im Bild ist dieser Effekt daher als Intensitätsabfall am Bildrand erkennbar, der als Vignettierung bezeichnet wird.

Die PSF einer Kamera mit Eintritts- bzw. Austrittspupille weist am Bildrand nicht die ideale Form der Apertur der Linse auf, sondern wird durch die Form der Pupillen begrenzt. Neben der Koma stellt die Vignettierung also eine weitere winkelabhängige Beeinflussung der PSF dar.

### 2.4.2.3 Randlichtabfall

Ein bezüglich der Auswirkung zur Vignettierung ähnlicher Effekt ist der Randlichtabfall aufgrund von schrägem Lichteinfall auf den Sensor. Die effektive Fläche eines auf den Sensor treffenden Strahlbündels nimmt mit zunehmendem Winkel zur optischen Achse ab. Dies bewirkt ebenfalls einen Intensitätsabfall am Bildrand. Der Randlichtabfall wird mathematisch beschrieben durch den  $\cos^4(\theta)$ -Term in (2.1), der in einer Simulation unmittelbar berücksichtigt werden kann. Da die Ursache dieses Effekt jedoch am Sensor und nicht am Objektiv liegt, hat er keine Auswirkung auf die Form der PSF.

### 2.4.2.4 Simulation geometrischer Abbildungsfehler

Um die Auswirkung der unterschiedlichen Abbildungsfehler auf die Signalverarbeitungskette einer *Computational Camera* untersuchen zu können, wird im Folgenden betrachtet, wie diese in einer Raytracing-Simulation nachgebildet werden können. Dabei stellt sich die Herausforderung, dass der genaue Aufbau kommerziell erhältlicher Objektive nicht verfügbar ist und sich daher nur schwer exakt in einer Simulation modellieren lässt. In der im Rahmen dieser Arbeit entstanden Raytracing-Simulationsumgebung wird daher ein Objektiv mit Abbildungsfehlern lediglich mit einer sphärischen dicken Linse simuliert, dessen Parameter sich zur Gewichtung der Abbildungsfehler flexibel einstellen lassen.

Das Objektiv ist wie in Abschnitt 2.3.1 beschrieben implementiert: Ausgehend vom simulierten Sensorpixel werden Strahlen in Richtung der Linse erzeugt, der jeweilige Schnittpunkt mit der Linsenoberfläche bestimmt und die Brechung an der Grenzfläche berechnet. Da lediglich Strahlen,



die die Linse treffen, erzeugt werden, wird implizit eine Kreisblende simuliert. Zur Simulation der chromatischen Aberration wird die Sellmeier-Gleichung [90] genutzt, die eine empirische Näherung der Wellenlängenabhängigkeit des Brechungsindex  $n_L(\lambda)$  für optisch transparente Medien bereitstellt:

$$n_L^2(\lambda) = 1 + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} + \frac{B_3\lambda^2}{\lambda^2 - C_3}. \quad (2.23)$$

Über die Sellmeier-Koeffizienten  $B_1, B_2, B_3, C_1, C_2$  und  $C_3$  können gängige optische Materialien simuliert werden. Da sich die Strahlgeometrien für unterschiedliche Wellenlängen ändern, muss für einen simulierten bildseitigen Strahl jeweils ein objektseitiger Strahl pro simulierter diskreter Wellenlänge berechnet werden. Die Berechnungsdauer eines Bildes steigt daher linear mit der Anzahl simulierter Wellenlängen an.

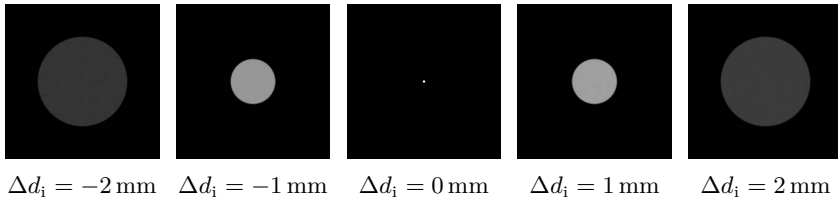
Zum Vergleich kann auch ein ideales Objektiv mit einer dünnen Linse in paraxialer Näherung simuliert werden, das die beschriebenen Abbildungsfehler nicht aufweist. Abbildung 2.17 zeigt die simulierten Impulsantworten dieses idealen Objektivs mit dünner Linse. Dabei wird schrittweise die Bildweite  $d_i$  variiert. Die Szene enthält lediglich eine kleine Kugel auf der optischen Achse. Bei der Bildweite  $d_{i,\text{fok}}$  mit

$$d_{i,\text{fok}} = \left( \frac{1}{f} - \frac{1}{d_o} \right)^{-1} \quad (2.24)$$

wird die Kugel scharf abgebildet. Bei anderen Bildweiten  $d_i = d_{i,\text{fok}} + \Delta d_i$  mit  $\Delta d_i \neq 0$  wird die Form der PSF deutlicher sichtbar, die in diesem Fall, aufgrund der angenommenen Blende und nicht aufgrund der Kugel, kreisförmig ist. Da Blenden häufig kreisförmig sind, spricht man auch vom Unschärfescheibchen. Das Licht der Szene ist homogen über der Fläche der PSF verteilt.

In Abbildung 2.18 sind zur Veranschaulichung der sphärischen Aberration die simulierten Impulsantworten eines Objektivs mit einer dicken sphärischen Linse dargestellt. Die Parameter der Linse werden so gewählt, dass sich nach der Linsenschleiferformel [29]

$$\frac{1}{f} = (n_L - 1) \left( \frac{1}{R_1} - \frac{1}{R_2} + \frac{(n_L - 1) d_L}{n_L R_1 R_2} \right) \quad (2.25)$$

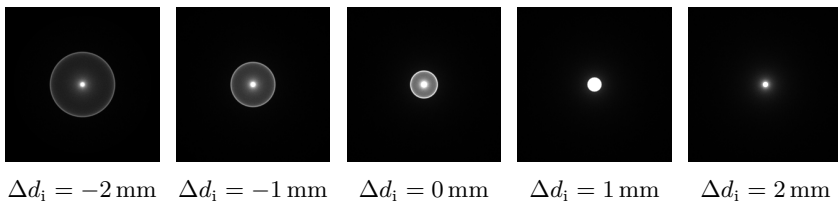


**Abbildung 2.17** Simulierte Impulsantwort einer idealen Kamera mit dünner Linse für unterschiedliche Bildweiten. Gegenstandsweite  $d_o = 1,3$  m, Brennweite  $f = 50$  mm, Blendenzahl  $k = 1,8$  und Bildweite  $d_i = d_{i, \text{fok}} + \Delta d_i$ .

dieselbe Brennweite  $f = 50$  mm wie in Abbildung 2.17 ergibt. Es wird zunächst keine zusätzliche Aperturblende simuliert. Die Apertur entspricht damit der Ausdehnung der Linse. Die Parameter sind in Tabelle 2.2 angegeben.

Im Vergleich zur idealen dünnen Linse fällt zunächst auf, dass die schärfste Abbildung nun bei einer Bildweite von  $d_i = d_{i, \text{fok}} + 2$  mm erreicht wird. Die Schärfe der idealen Linse wird jedoch nicht erreicht. Die Ausdehnung der PSF ist stets größer als beim idealen Objektiv mit  $\Delta d_i = 0$ . Außerdem ist die Intensität der PSF nicht homogen, wie im idealen Fall.

Zur Simulation der Koma wird die Szene weiter modifiziert, indem die abgebildete Kugel von der optischen Achse weg bewegt wird. Damit fallen die von der Kugel ausgehenden Strahlen nicht mehr parallel sondern schräg unter einem Winkel  $\alpha$  zur optischen Achse auf die Linse. Die Kugel wird nicht mehr in der Mitte sondern am Bildrand abgebildet.



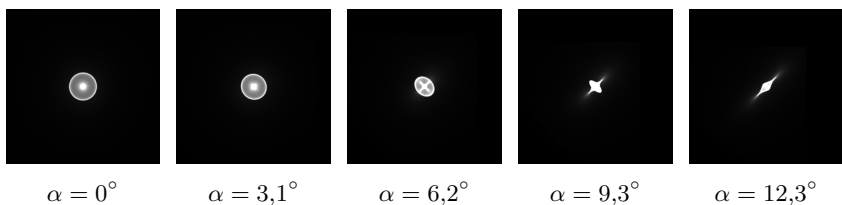
**Abbildung 2.18** Simulierte Impulsantwort einer Kamera mit sphärischer dicker Linse für unterschiedliche Bildweiten. Gegenstandsweite  $d_o = 1,3$  m, Bildweite  $d_i = d_{i, \text{fok}} + \Delta d_i$ . Weitere Parameter nach Tabelle 2.2.

**Tabelle 2.2** Parameter der simulierten Kamera mit dicker sphärischer Linse.

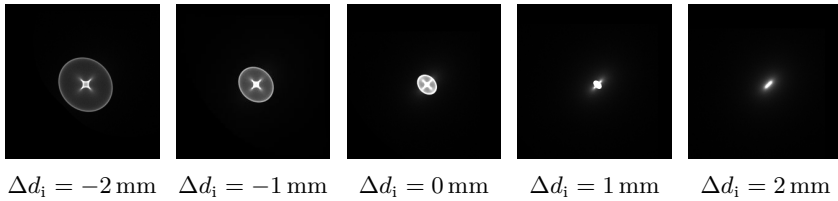
Blendenzahl	$k$	1,8
Brechungsindex	$n_L$	1,46
objektseitiger Krümmungsradius	$R_1$	4,68 cm
bildseitiger Krümmungsradius	$R_2$	-4,68 cm
Linsendicke	$d_L$	5 mm

Abbildung 2.19 zeigt die jeweils zentriert ausgeschnittenen, simulierten Impulsantworten, wenn die Kugel schrittweise nach oben rechts bewegt wird. Die Kugel befindet sich stets in einer Ebene im Abstand  $d_o = 1,3$  m von der Kamera mit der Bildweite  $d_i = d_{i,\text{fok}}$ , so dass mit einer idealen Kamera stets eine scharfe Abbildung vorliegt. Die Form der simulierten PSF wird, wie erwartet, mit wachsender Entfernung zur optischen Achse zunehmend verzerrt, und es entsteht der namensgebende Schweif (lat. *coma*).

Um den Effekt des Astigmatismus auf die PSF sichtbar zu machen, wird in der vorigen Szene mit der um  $\alpha = 6,2^\circ$  von der optischen Achse verschobenen Kugel zusätzlich die Bildweite variiert. Es wurde wieder eine Kamera mit dicker Linse ohne Eintrittspupille simuliert. Die Ergebnisse dieser simulierten Szene sind in Abbildung 2.20 dargestellt. Neben der Verzerrung der PSF aufgrund der Koma ändert sich die Form zusätzlich mit der Bildweite  $d_i$  aufgrund der unterschiedlichen Brennweiten der Meridional- und Sagittalbündel.



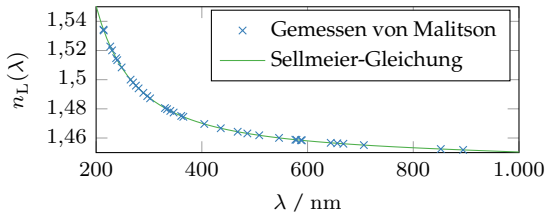
**Abbildung 2.19** Simulierte Impulsantwort einer Kamera mit sphärischer dicker Linse aus unterschiedlichen Winkeln  $\alpha$  zur optischen Achse. Auf optische Achse projizierte Objektentfernung  $d_o = 1,3$  m, Bildweite  $d_i = d_{i,\text{fok}}$ . Weitere Parameter nach Tabelle 2.2.



**Abbildung 2.20** Simulierte Impulsantwort einer Kamera mit sphärischer dicker Linse aus dem Winkel  $\alpha = 6,2^\circ$  zur optischen Achse für unterschiedliche Bildweiten. Gegenstandsweite  $d_o = 1,3 \text{ m}$ , Bildweite  $d_i = d_{i,\text{fok}} + \Delta d_i$ . Weitere Parameter nach Tabelle 2.2.

In Abbildung 2.21(b) ist schließlich die simulierte PSF einer dicken sphärischen Linse mit chromatischer Aberration dargestellt. Dabei wurde, durch die Wahl der Sellmeier-Koeffizienten, Quarzglas als Linsenmaterial simuliert. In Abbildung 2.21(a) ist sowohl der von Malitson [58] gemessene Verlauf des Brechungsindex über der Wellenlänge von Quarzglas als auch der mithilfe der Sellmeier-Gleichung (2.23) für dieses Material approximierte und zur Simulation verwendete Verlauf angegeben. Es wurden äquidistant 30 Wellenlängen zwischen 400 nm und 700 nm simuliert. Die Umrechnung von Wellenlängen in den RGB-Farbraum erfolgt nach Pharr und Humphreys [74] über das CIE-Normvalenzsystem. Durch die stärkere Brechung des kurzwelligeren Lichts sind die blauen Anteile der PSF örtlich stärker konzentriert als die roten Anteile. Im Zentrum überlagern sich die einzelnen Lichtanteile zum ursprünglichen weißen Spektrum.

Die Auswirkung der Vignettierung auf die PSF lässt sich auf einfache Weise durch Hinzufügen einer Aperturblende außerhalb der Linsenebene simulieren. Abbildung 2.22 zeigt die simulierten PSFs einer idealen Kamera mit dünner Linse und objektseitiger kreisförmiger Aperturblende als Eintrittspupille. Die Kugel wurde dabei zusätzlich unscharf abgebildet, um die Auswirkungen der resultierenden Pupillen besser sichtbar zu machen.

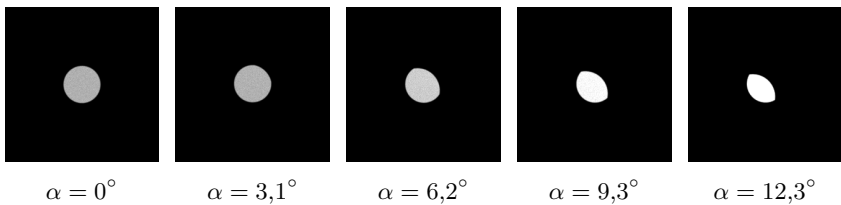


(a) Vergleich der gemessenen Wellenlängenabhängigkeit des Brechungsindex von Quarzglas mit der Approximation aus der Sellmeier-Gleichung.

(b) Simulierte chromatische Aberration einer Impulsantwort.

**Abbildung 2.21** Simulation der chromatischen Aberration von Quarzglas.

Die Parameter des Objektivs der obigen Untersuchungen wurden beliebig so gewählt, dass die grundlegenden Abbildungseigenschaften denen gängiger Normalobjektive entsprechen. Da heutige Objektive zur Reduktion von Abbildungsfehlern deutlich aufwendiger als mit nur einer einzelnen sphärischen Linse konstruiert sind, weisen sie jedoch wesentlich geringere Abbildungsfehler als in den oben beschriebenen Simulationen auf. In der Simulation lassen sich die Abbildungsfehler durch eine flexible Anpassung der Linsenparameter reduzieren, z. B. indem der Betrag der Krümmungsradien  $R_1$  und  $R_2$  erhöht, die Linsendicke  $d_L$  reduziert oder ein höherer Brechungsindex  $n_L$  gewählt wird. Weil reale Objektive ohnehin anders aufgebaut sind, muss dabei die tatsächliche Realisierbarkeit des simulierten Objektivs nicht im Vordergrund stehen.



**Abbildung 2.22** Simulierte Impulsantwort einer Kamera mit ideal dünner Linse aus unterschiedlichen Winkeln  $\alpha$  zur optischen Achse mit zusätzlicher Eintrittspupille. Auf optische Achse projizierte Objektentfernung  $d_o = 1,3$  m, Bildweite  $d_i = d_{i,\text{fok}} - 1$  mm. Weitere Parameter nach Tabelle 2.2.

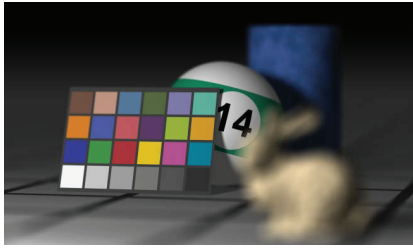
## 2.5 Referenzdaten

Referenzdaten erlauben den quantitativen Vergleich einer *Computational Camera* zu anderen Messverfahren und ermöglichen so eine objektive Bewertung der Leistungsfähigkeit. Darüber hinaus können unterschiedliche Konfigurationen der Kamera oder die Leistungsfähigkeit in verschiedenen Szenarien objektiv bewertet werden. Zur Charakterisierung einer prototypischen Realisierung werden geeignete Referenzmesssysteme benötigt, deren Genauigkeit die des Prototyps übertrifft. Ein Vorteil einer Simulation ist, dass Referenzdaten implizit zur Verfügung stehen, da die Versuchsbedingungen, im Kontext einer Kamera also z. B. die Konfiguration der Szene und die Position der Kamera, exakt bekannt sind. Darüber hinaus lassen sich zur Gewinnung von Referenzdaten ideale Kameras simulieren.

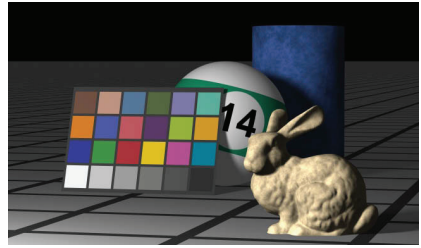
Abbildung 2.23(a) zeigt zur Veranschaulichung der Referenzdatenerzeugung mit dem Raytracing-Verfahren eine simulierte Aufnahme einer Szene. Als Kamera wurde eine Kamera mit einer dünnen Linse gewählt, weshalb die Aufnahme eine endliche Schärfentiefe aufweist. In Abbildung 2.23(b) ist die Aufnahme mit einer idealen Lochkamera derselben Szene dargestellt. Aufgrund der unendlichen Schärfentiefe kann diese Aufnahme als Referenz für Verfahren zur Reduktion von Unschärfe [21, 37, 49, 50, 111] dienen. Außerdem können derartige Aufnahmen auch als Referenz für aus Lichtfeldern rekonstruierte Aufnahmen verwendet werden [44, 51, 68, 100], indem die Position der Kamera und die Parameter entsprechend der rekonstruierten Aufnahme gewählt werden.

Das Raytracing-Verfahren als Simulationswerkzeug einer *Computational Camera* hat zusätzlich den Vorteil, dass zahlreiche geometrische Referenzdaten ohnehin beim Rendern anfallen und auf einfache Weise nutzbar gemacht werden können. Dazu muss lediglich das Modul, das den Datenaustausch zwischen Umgebungs- und Kamerasimulation aus Abbildung 2.3 implementiert, angepasst werden.

Zur Approximation der Strahldichte muss in der simulierten Szene geprüft werden, welches Objekt vom gerade zu untersuchenden Strahl getroffen wird. Schneidet die Geradengleichung des Strahls mehrere Objekte, muss das Objekt mit der geringsten Entfernung zum Strahlursprung ausgewählt werden. Diese Entfernung muss also bereits zur Simulation für jedes getroffene Objekt berechnet werden. Anstelle des vom Objekt reflektierten Farbwerts kann auch unmittelbar die Objektentfernung zurückgegeben



(a) Originalaufnahme



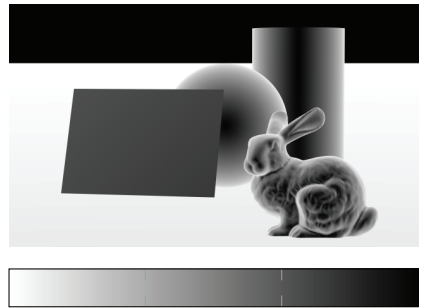
(b) Unendliche Schärfentiefe.



5 10 20 40

Entfernung / m

(c) Referenztiefe.



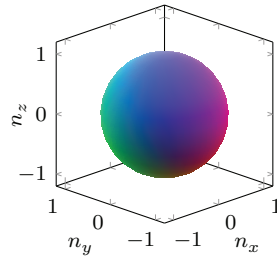
30 60

Ausfallswinkel  $\theta_{\text{out}} / ^\circ$ 

(d) Referenzoberfläche.



(e) 3D-Normalenrichtung.



(f) Codierung der 3D-Normalen.

**Abbildung 2.23** Referenzdatenerzeugung per Raytracing-Simulation.

werden, um ein Referenz-Tiefenbild zu erzeugen. Ein Beispiel dafür ist in Abbildung 2.23(c) angegeben. Diese Referenzaufnahme lässt sich unmittelbar zur Evaluierung von Algorithmen zur dichten Tiefenextraktion nutzen [8, 48, 49, 100, 109, 119, 120].

Bei der Berechnung des reflektierten Farbwertes beim Schnitt eines Strahls mit einem Objekt wird nach (2.11) der Winkel  $\theta_{\text{in}}$  zwischen einfallendem Strahl und der Oberflächennormalen benötigt. Dafür muss an diesen Schnittpunkten die Oberflächennormale aus der Objektgeometrie berechnet werden. Die Oberflächennormale kann unmittelbar in Weltkoordinaten, anstelle des reflektierten Farbwertes, an die Kamera zurückgegeben werden, um die Referenzdaten der Oberflächenorientierung zu erhalten. In Abbildung 2.23(e) sind die Oberflächennormalen der sichtbaren Objekte entsprechend des um eine Einheitskugel gelegten RGB-Farbraums codiert, wobei die  $x$ -Komponente durch den roten, die  $y$ -Komponente durch den grünen und die  $z$ -Komponente durch den blauen Farbkanal, wie in Abbildung 2.23(f) dargestellt, angegeben sind. Wahlweise lässt sich auch die Oberflächennormale bezüglich der Position der Kamera aus der einfallenden Strahlrichtung berechnen und angeben. Abbildung 2.23(d) zeigt entsprechend den als Grauwert codierten Betrag des Winkels  $\theta_{\text{out}}$  zwischen dem zur Kamera ausfallenden Strahl und der Oberflächennormalen. Derartige Referenzdaten lassen sich zur Evaluierung von Verfahren in der Oberflächenmesstechnik, z. B. der Deflektometrie [104], einsetzen.

## 2.6 Kameraoptimierung

Die Simulation des physikalischen Teils der Signalverarbeitungskette ermöglicht eine ganzheitliche Beschreibung und Betrachtung einer *Computational Camera* in der digitalen Domäne. Dadurch kann der Aufwand für die Durchführung von Experimenten oftmals beträchtlich reduziert werden. Darüber hinaus sind Experimente auf Grundlage einer Simulation einfacher automatisierbar. Dadurch lässt sich auf einfache Weise eine Vielzahl von Versuchen, in geringerer Zeit als dies mit einem realen Aufbau möglich wäre, durchführen. Mit der im Rahmen dieser Arbeit entstanden Raytracing-Simulationsumgebung können beispielsweise problemlos mehrere Tausend Szenen pro Tag gerendert werden. Sofern, wie im hier beschriebenen Fall, bei der Simulation gleichzeitig die Möglichkeit



besteht, Szenen bzw. Elemente der Szenen prozedural und/oder zufällig zu erzeugen, kann das Kamerasystem anhand einer großen Vielfalt stochastisch unabhängiger Szenen simulativ erprobt werden. Wie im vorigen Abschnitt beschrieben, können in einer Simulation darüber hinaus auch die unterschiedlichen Störeinflüsse flexibel variiert werden. Im Rahmen der durch die Simulation erzeugbaren Szenen und Szenarios lässt sich so ein breites Spektrum an Eingangsdaten erzeugen, womit sich die erzielten Ergebnisse hinsichtlich ihrer Robustheit bewerten lassen.

Sofern sich der Prozess der Bildgewinnung mit ausreichender Genauigkeit mathematisch und physikalisch modellieren lässt, können die Parameter des Kamerasystems anhand dieses theoretisch Modells bewertet werden, wie dies beispielsweise Levin et al. [49] für codierte Blenden zur Tiefenextraktion durchgeführt haben. Zhou und Nayar [111] nutzen zur Bewertung von Blendenpaaren für die Tiefenextraktion mithilfe zweier Aufnahmen ebenfalls ein mathematisches Modell, das explizit rauschartige Störeinflüsse berücksichtigt.

Oftmals lässt sich jedoch kein ausreichend präzises mathematisches Modell des Kamerasystems aufstellen, das eine aussagekräftige Bewertung ermöglichen würde. Beispielsweise können mit den Modellen der oben genannten Arbeiten keine Abbildungsfehler der Kameraoptik beschrieben und bewertet werden. Die Auswirkungen derartiger Effekte lassen sich jedoch, wie in den vorigen Abschnitten gezeigt, mit dem Raytracing-Verfahren numerisch simulieren. Daher bietet eine Raytracing-Simulation die Möglichkeit der Berücksichtigung derartiger Phänomene bei der Bewertung der Parameter eines Kamerasystems, indem die Leistungsfähigkeit anhand von Simulationsergebnissen anstelle eines analytischen mathematischen Modells beurteilt wird.

Da ohne ein analytisches Modell keine geschlossene Formulierung des Problems existiert, müssen zur Optimierung der Kameraparameter numerische Verfahren eingesetzt werden, die jedoch nur eine Approximation des Optimums liefern können. Als Verfahren können beispielsweise gerichtete Verfahren wie das Gradientenabstiegsverfahren, stochastische Verfahren wie genetische Algorithmen oder *Simulated Annealing* oder auch eine vollständige Suche eingesetzt werden. Die Konvergenzgeschwindigkeit der Verfahren kann je nach Dimension des zu optimierenden Systemparameters stark variieren.



## 3 Kamera mit programmierbarer Apertur

Durch die Erweiterung der Bildsignalverarbeitung auf die physikalische Domäne können durch neuartige *Computational Cameras* auch andere im Licht enthaltenen Informationen als örtliche Intensitäten erfasst werden. Damit können derartige Systeme in immer mehr Anwendungsgebieten eingesetzt werden. Aufgrund des problembezogenen Entwurfs beim *Computational Imaging* ist ein konkretes Kamerasystem jedoch nur für eine spezifische Anwendung konzipiert. Die Umsetzung der optischen Signalverarbeitung in konkrete Hardware erschwert eine flexible Anpassung oder Änderung der Parameter des Kamerasystems. Insbesondere in einem prototypischen Entwicklungsstadium liegt das Hauptaugenmerk der Implementierung üblicherweise auf der Erfüllung einer bestimmten geforderten Funktion und nicht auf Flexibilität.

In der Einteilung nach Zhou und Nayar [112] kann sich sowohl der Ort der zusätzlichen optischen Komponenten als auch deren Typ je nach Anwendung ändern. Der Umbau einer *Computational Camera* auf eine gänzlich andere Problemstellung ist daher nur schwer möglich. Beschränkt man sich auf eine bestimmte optische Komponente, können immer noch mehrere unterschiedliche Anwendungen realisiert werden. Zum Beispiel können codierte Blenden zur Tiefenextraktion [8, 46, 49, 100, 110], zur Verbesserung der Schärfentiefe [45, 49, 56] oder zur Aufnahme eines 4D-Lichtfelds [44, 51, 100] verwendet werden. Die Ansätze unterscheiden sich dabei in der Form und der Position der Blende. Doch auch in diesem eingeschränkten Anwendungsfeld ist eine Änderung dieser Blendenparameter nicht immer einfach möglich und erfordert zumeist einen manuellen Eingriff, der eine erneute Kalibrierung des Systems zur Folge haben kann.

Selbst bei der Beschränkung auf ein konkretes Anwendungsgebiet kann eine flexible Implementierung der optischen Signalverarbeitung hilfreich

und notwendig sein. Wiederum am Beispiel der Tiefenextraktion mit codierten Blendenpaaren hat Levin [48] gezeigt, dass die optimalen Systemparameter einer *Computational Camera* sich bezüglich der äußeren Bedingungen ändern können. Wie bereits in Kapitel 2 beschrieben, liegen in Abhängigkeit der absoluten Intensität des einfallenden Lichts unterschiedliche Rauschbedingungen vor. Bei dunklen Bedingungen liegt vornehmlich additives Bildrauschen der Kameraelektronik vor. Dagegen ist bei hellen Bedingungen das multiplikative Photonenrauschen dominant. Da die Anzahl der einfallenden Photonen auch von der Form der Blende selbst abhängt, ergeben sich abhängig vom vorliegenden Störszenario unterschiedliche optimale Blendenformen.

Um neuartige Kamerasysteme des *Computational Imaging* trotzdem innerhalb des durch die verwendeten optischen Komponenten ermöglichten Rahmens breit anwenden und adaptiv auf vorliegende äußere Bedingungen anpassen zu können, müssen die zusätzlichen Komponenten im Lichtweg flexibel einstellbar oder sogar programmierbar implementiert werden. Nach einer Übersicht über derartige Ansätze wird in diesem Kapitel eine im Rahmen dieser Arbeit aufgebaute Kamera mit programmierbarer Apertur beschrieben und charakterisiert.

## 3.1 Stand der Technik

Zunächst werden flexible Ansätze und programmierbare Systeme aus der Literatur, die allgemein unter den Oberbegriff *Computational Imaging* fallen, vorgestellt. Anschließend werden bestehende Ansätze des in dieser Arbeit behandelten Falls einer programmierbaren Blende sowie damit verwandte Arbeiten näher betrachtet.

### 3.1.1 Adaptive Verfahren

Adams et al. [2] haben mit der *Frankencamera* eine flexibel einstellbares und programmierbares Framework, bestehend aus Kamerahardware und zugehöriger Software, geschaffen. Neben der Kontrolle der kameraeigenen Parameter wie Fokus oder Belichtung bei der Aufnahme lassen sich auch externe Peripheriegeräte programmieren und mit der Aufnahme synchronisieren. Auf diese Weise kann beispielsweise eine in Zeitrichtung

flexibel codierbare Beleuchtung während der Aufnahme realisiert und damit Bewegungsunschärfe mit einer Entfaltung reduziert werden.

Auf ähnliche Weise funktioniert die von Raskar et al. [82] konstruierte Kamera, die mit einem ferroelektrischen *Shutter* ausgestattet ist. Damit lässt sich die Belichtung in Zeitrichtung codieren. Bei der klassischen Belichtung wird das Bildsignal in Zeitrichtung mit einem Rechteckfenster gefaltet, wodurch Signalanteile unwiederbringlich verloren gehen. Durch die Belichtung gemäß einer binären Pseudozufallsfolge, die ein breitbandiges Spektrum aufweist, können mehr Signalfrequenzen erfasst werden. Anschließend kann durch eine Entfaltung das ursprüngliche Bildsignal ohne Bewegungsunschärfe rekonstruiert werden.

Kameras mit mechanisch bewegbarem Bildsensor während der Belichtung können, ähnlich wie bei codierten Blenden, je nach Art der Bewegung unterschiedlich eingesetzt werden. Cho et al. [19] bewegen während der Belichtung den Sensor mithilfe zweier aufeinander platzierter Stellische. Die Stellische können entlang einer Raumrichtung programmierbar bewegt werden. Sie sind so ausgerichtet, dass diese beiden Richtungen orthogonal zueinander und orthogonal zur optischen Achse stehen. Sofern die relative Sensorbewegung mit der Bewegung eines Objektes übereinstimmt, kann dieses scharf abgebildet werden. Um auch beliebig bewegte Objekte in der Szene ohne Informationsverlust erfassen zu können, nehmen die Autoren zwei Bilder auf, bei denen jeweils ein Stellisch eine bezüglich der Zeit parabelförmige Bewegung durchführt. Damit wird pro Richtung die Geschwindigkeit des Sensors linear variiert. Mit der bekannten Sensorbewegung wird anschließend mit einer lokalen Entfaltung die auftretende Bewegungsunschärfe kompensiert.

Ben-Ezra et al. [9] verwenden zwei Gleichstrommotoren, um den Bildsensor einer Videokamera in der Zeit zwischen den Frames senkrecht zur optischen Achse zu verschieben. Die Autoren nennen diesen Aufbau *Jitter-Kamera*. Damit kann das Ort-Zeit-Volumen besser abgetastet werden und es lässt sich nachträglich die örtliche Auflösung der Aufnahmen mit Ansätzen zur *Super-Resolution*-Rekonstruktion verbessern. Die Firma *Pixerra* realisiert diesen Ansatz durch ein verkippbare Glasplättchen vor dem Sensor, das das einfallende Lichtfeld relativ zum Sensor verschiebt [76].

Kuthirummal et al. [43] nutzen einen Linearstellisch, um den Bildsensor während der Belichtung entlang der optischen Achse zu bewegen. Auf

diese Weise werden unterschiedliche Tiefen der Szene während der Aufnahme scharf abgebildet. Durch die bekannte Bewegung des Sensors lässt sich durch eine Entfaltung die Schärfentiefe flexibel einstellen. Eine lineare Bewegung des Sensors ermöglicht die Erweiterung der Schärfentiefe, wohingegen mit einer nichtlinearen Bewegung sogar mehrere, nicht zusammenhängende Entfernungsbereiche scharf abgebildet werden können. In Kombination mit einem *Rolling Shutter* können darüber hinaus entsprechend der Scheimpflug-Bedingung verkippte oder sogar nicht-planare Fokusebenen realisiert werden

Eine räumlich flexible Beleuchtung wurde von Gruna [31] in einem koaxialen Kamera-Projektor-System realisiert. Mit diesem System kann durch Lichttransportinvertierung ein Prüfobjekt derart beleuchtet werden, dass lediglich Abweichungen von der Referenzszene als Kontrastmerkmale sichtbar sind. Durch die Änderung des projizierten Lichtmusters kann das System adaptiv auf die jeweilige Referenzszene angepasst werden. In derselben Arbeit entwickelt Gruna ein katadioptrisches Projektor-Kamera-System, das es ermöglicht, durch gewichtete Superposition von Basisbeleuchtungsmustern aufgabenspezifische Reflektanzmerkmale optisch aufzunehmen. In beiden Fällen ist die Beleuchtung durch die Nutzung eines handelsüblichen Projektors räumlich und zeitlich programmierbar und kann so adaptiv auf die konkrete vorliegende Szene angepasst werden.

#### 3.1.2 Adaptive Blendencodierung

Adaptive Elemente im Lichtweg einer Kamera wurden schon früh für verschiedene Anwendungen vorgeschlagen. Fujikake et al. [26] platzieren ein programmierbares Polarisationsfilter vor einer Videokamera. Das programmierbare Polarisationsfilter wird durch ein festes Polarisationsfilter mit zwei davor platzierten *twisted nematic* Flüssigkristallzellen (TN-LC) realisiert, die abhängig von den angelegten Spannungen den Polarisationswinkel des durchfallenden Lichts drehen. Auf diese Weise lassen sich linear polarisierte Anteile des Lichts, die beispielsweise bei der Reflexion an spiegelnden Oberflächen gemäß der Fresnel'schen Formeln [38] entstehen, unterdrücken. Die Autoren entfernen so das reflektierte Licht an Glasscheiben und Wasseroberflächen, um dahinter liegende Objekte sichtbar zu machen. Durch die frei einstellbare Filterrichtung des Polarisationsfilters

kann die Kamera adaptiv auf den Polarisationszustand der gerade betrachteten Szene angepasst werden. Da das adaptive Polarisationsfilter keine örtliche Variation erlaubt, wirkt das Filter grundsätzlich homogen auf das gesamte Bild, weshalb immer nur eine Polarisationsrichtung vollständig herausgefiltert werden kann.

Nayar und Branzoi [65] montieren ein transmissives Flüssigkristalldisplay (LCD) vor einer Kamera. LCDs sind durch Platzierung von Polarisationsfiltern vor und hinter dem Flüssigkristall in der Lage, die Lichtintensität unabhängig des Polarisationszustands zu dämpfen. Darüber hinaus ist das LCD örtlich aufgelöst, d. h. Licht kann in Abhängigkeit des Ortes gedämpft werden. Die Autoren nutzen diesen Aufbau zur Erweiterung des dynamischen Helligkeitsbereichs der Kamera. An Stellen hoher Lichtintensität wird die Dämpfung des LCDs entsprechend hoch gewählt, damit der Sensor nicht voll ausgesteuert wird und keine Information verloren geht. Da sich das LCD nahe der Aperturebene befindet, müssen die LCD-Pixel mithilfe einer Homographie-Matrix den Sensorpixeln geometrisch zugeordnet werden. Mit der am Sensor gemessenen Intensität und dem Wissen, wie stark das Licht durch das LCD gedämpft wurde, lässt sich die ursprüngliche Intensität bestimmen und ein sogenanntes *high dynamic range*-Bild (HRD) erstellen. Die Autoren nutzen einen regelbasierten Algorithmus, um während der Videoaufnahme kontinuierlich die optimalen Dämpfungsparameter des LCDs für die aktuelle Szene derart zu bestimmen, dass weder eine Sättigung des Sensors noch große Quantisierungsfehler aufgrund einer zu geringen Lichtintensität auftreten.

Mannami et al. [59] nutzen ebenfalls ein LCD zur Aufnahme eines HDR-Bildes. Im Gegensatz zum vorigen Fall verwenden die Autoren jedoch ein reflektives LCD (LCoS, *Liquid Crystal on Silicon*), das anstelle eines zweiten Polarisationsfilters eine reflektive Schicht besitzt. Da bei diesem LCD-Typ die Verdrahtung der Pixel hinter der reflektiven Schicht platziert ist und damit nicht durchleuchtet wird, besitzt es eine höhere Lichteffizienz als ein transmissives LCD, bei dem sich die Verdrahtung negativ auf den Füllfaktor auswirkt. Allerdings muss der Lichtweg in der Kamera über einen Strahlteiler umgelenkt und über Relaisoptiken verlängert werden, damit das einfallende Licht am LCD reflektiert und anschließend auf den Sensor fallen kann. Somit kann kein Standardobjektiv verwendet werden. Dafür kann das LCD in einer Ebene platziert werden, die scharf

auf den Sensor abgebildet wird, weshalb die LCD-Pixel unmittelbar den Sensorpixeln zugeordnet werden können.

Nagahara et al. [62] verwenden ebenfalls einen Strahlteiler mit Relaisoptik, um eine reflektierende Apertur mit einem LCoS-Element zu realisieren. Diesen Aufbau nutzen die Autoren für die sequentielle Lichtfeldaufnahme und zur Bildrestauration bei Defokussierung.

Zomet und Nayar [116] realisieren eine programmierbare Apertur für ein linsenloses Abbildungssystem mit einem transmissiven LCD. Um ohne Linse eine scharfe Abbildung zu ermöglichen, muss, entsprechend dem Modell einer Lochkamera, der durchlässige Bereich des LCDs sehr klein gewählt werden. Durch Änderung der Position dieses Bereichs kann der Blickwinkel der Kamera verändert werden, ohne das System tatsächlich zu bewegen. Wird die Apertur durch eine oder mehrere weitere Schichten realisiert, lassen sich beispielsweise mehrere Bildausschnitte optisch auswählen und nebeneinander auf dem Sensor arrangieren, mit jeweils unabhängig wählbarer Vergrößerung. Für die Implementierung dieser komplexeren geometrischen Operationen verwenden die Autoren jedoch keine zweite programmierbare Schicht, sondern eine feste physikalische Apertur. Durch die fehlende Sammellinse erfordert dieses System längere Belichtungszeiten als klassische Kameras.

Liang et al. [51] verwenden eine Kamera mit einer programmierbaren Apertur zur sequentiellen Aufnahme eines 4D-Lichtfelds mit voller Ortsauflösung. Dabei werden mit jeder Aufnahme mehrere Ansichten des Lichtfelds gleichzeitig, entsprechend einem zuvor durch Optimierung ermittelten Multiplexing-Muster, erfasst und anschließend in einem Demultiplexing-Schritt voneinander getrennt. Durch die gleichzeitige Aufnahme mehrerer Ansichten wird pro Ansicht und pro Aufnahme mehr Licht gesammelt und damit das SNR gegenüber der Aufnahme der Einzelansichten verbessert. Für den experimentellen Nachweis verwenden die Autoren zwei Prototypen mit programmierbarer Apertur. Beim ersten Prototyp ist die Apertur durch ein bedrucktes Folienband realisiert, das mechanisch durch einen Spalt zwischen Kamera und Objektiv bewegt wird. Auf dem Band sind die unterschiedlichen Muster zur Lichtfeldaufnahme gedruckt. Der zweite Prototyp verwendet ein transmissives LCD mit einer Auflösung von lediglich  $7 \times 7$  Pixeln, das ebenfalls zwischen Kamera und Objektiv platziert ist. In beiden Fällen entspricht die Position



der Blende nur näherungsweise der Aperturebene. Durch den Freiraum zwischen Kamera und Objektiv wird darüber hinaus die Bildweite  $d_i$  des Systems vergrößert, weshalb nicht mehr auf Unendlich fokussiert werden kann. Durch die Öffnung wird das System auch anfällig gegenüber Fremdlicht und Verschmutzungen.

## 3.2 Objektivumbau

Zum Vergleich mit dem in Kapitel 2 dargestellten Simulationsverfahren und zum Nachweis der Leistungsfähigkeit eines damit simulativ entworfenen und optimierten Ansatzes des *Computational Imaging* ist ein prototypischer Aufbau des Kamerasystems notwendig. Dazu wurde im Rahmen dieser Arbeit ein Kamerasystem mit programmierbarer Apertur aufgebaut. Damit lassen sich sämtliche Anwendungen der Blendencodierung erproben.

Die Anforderungen an den Aufbau folgen direkt aus den obigen Betrachtungen des Stands der Technik mit dem Ziel, diesen zu verbessern. Die Anforderungen können wie folgt zusammengefasst werden:

1. Eine Änderung der Blendenform soll programmierbar, also ohne manuellen Eingriff, möglich sein. Damit soll die Anpassung der Blendenparameter im laufenden Betrieb ermöglicht und eine potenzielle Änderung der Positionen der Komponenten bei einem solchen Eingriff vermieden werden, welcher eine erneute Kalibrierung erforderlich machen würde.
2. Das programmierbare Element soll möglichst exakt in der Aperturebene platziert werden. In diesem Fall kann die Auswirkung der Blende als eine Faltung des ideal scharfen Abbilds der Szene mit der PSF angesehen werden. Bei abweichender Blendenposition muss zur Betrachtung der Auswirkung der Blende das 4D-Lichtfeld im Objektiv herangezogen werden [44, 67].
3. Die Abbildungsqualität soll vergleichbar mit der Qualität einfacher kommerziell verfügbarer Objektive sein.
4. Trotz der programmierbaren Komponente in der Aperturebene sollen die grundlegenden Funktionalitäten eines Objektivs ermöglicht

werden. Insbesondere die Einstellung der Fokusweite bis auf Unendlich soll möglich sein.

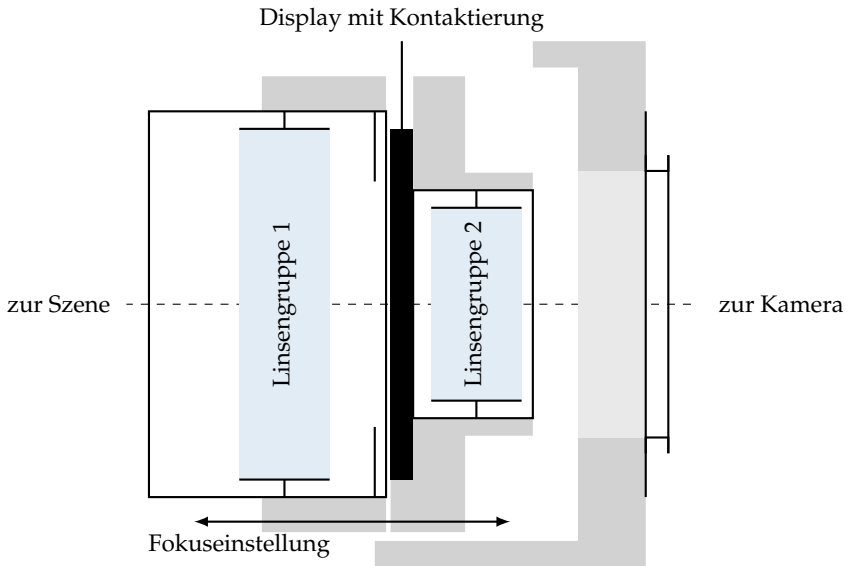
5. Der Aufbau soll möglichst robust gegenüber Fremdlichteinflüssen und Verschmutzungen sein.

In dieser Arbeit wurde ein kommerziell verfügbares Objektiv als Grundlage verwendet und entsprechend den Anforderungen modifiziert. Dies hat den Vorteil, dass aufgrund der unterschiedlichen Korrekturmaßnahmen heutiger Objektive eine sehr gute Abbildungsqualität erreicht werden kann. Eine komplette Neugestaltung eines Objektivs, beispielsweise mit einer einfachen Sammellinse, würde zu stärkeren Abbildungsfehlern führen. Ferner ist das System bei der Verwendung weit verbreiteter Objektivbajonettssysteme nicht auf die Verwendung einer bestimmten Kamera angewiesen, sondern kann mit verschiedenen Kameras genutzt werden.

Um die Abbildungseigenschaften des verwendeten Objektivs beibehalten zu können, sollte das Objektiv möglichst wenig verändert werden. Daher wird ein Aufbau ohne Strahlteiler und zusätzliche Relaisoptik, wie in den Aufbauten von Mannami et al. [60] und Nagahara et al. [62], bevorzugt. Es wird also als programmierbare Apertur ein transparentes Element benötigt. In dieser Arbeit wird ein transmissives LCD verwendet, das eine berührungslose Änderung der Blendenform ermöglicht.

Für die optimale Wirkung der Blende muss das LCD in der Apertur-ebene des Objektivs platziert werden. Da Objektive heute bereits mit einer mechanisch verstellbaren Blende ausgestattet sind, kann die Position dieser Blende als Näherung der Apertur-ebene angesehen werden. Eine Platzierung des LCDs dort hat auch den Vorteil, dass das Objektiv an die Kamera angeschlossen werden kann. Somit wird das Auflagemaß des Objektivs nicht verändert wie beispielsweise in der Arbeit von Liang et al. [51]. Damit kann das Objektiv auf alle zuvor möglichen Entfernungen fokussiert werden. Es entsteht keine große Öffnung, durch die Fremdlicht oder Schmutz in das System gelangen kann.

Als Objektiv wurde in dieser Arbeit ein *AF Nikkor 50 mm f/1,8* Normalobjektiv von *Nikon* verwendet. Durch die feste Brennweite des Objektivs gibt es vergleichsweise wenig bewegliche Teile im Inneren, was die Integration des LCDs erleichtert. Das Objektiv ist aus zwei Linsengruppen aufgebaut, zwischen denen die Lamellen der mechanischen Blende platziert sind. Mit der Blendenzahl von 1,8 ist das Objektiv außerdem lichtstark. Damit kann

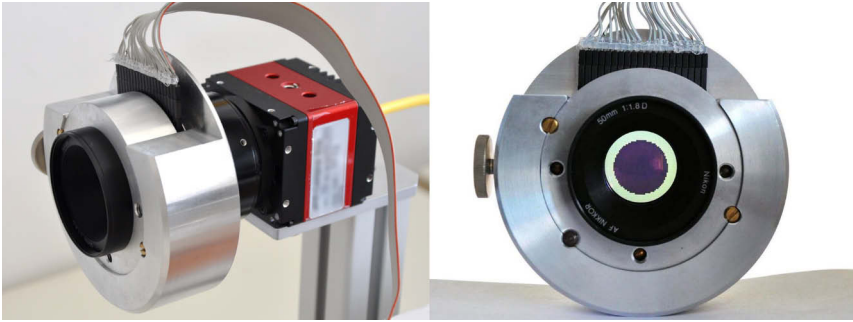


**Abbildung 3.1** Aufbau des Objektivs mit programmierbarer Apertur.

einerseits viel Licht erfasst werden, was sich positiv auf das SNR auswirkt, und andererseits kann eine große Fläche für die programmierbare Blende genutzt werden. Mit dem weit verbreiteten F-Bajonett-Anschluss kann das Objektiv unmittelbar an zahlreiche Kameras angeschlossen werden.

Als programmierbare Blende dient ein grafisches *DOGM 128S-6* Display von *Electronic Assembly* mit einer Pixelauflösung von  $128 \times 64$ . Die aktive Fläche dieses transmissiven LCDs reicht gerade aus, um die Blendenöffnung des Objektivs vollständig zu überdecken. Die Pixelverdrahtung nimmt 7% der aktiven Fläche ein. Die Pixel können lediglich binär geschaltet werden.

In Abbildung 3.1 ist der schematische Aufbau des Objektivs mit programmierbarer Apertur dargestellt. Aufgrund der beweglichen Lagerung der Linsengruppen innerhalb des ursprünglichen Objektivgehäuses ist es nicht möglich, das LCD im Gehäuse zu platzieren und gleichzeitig die Möglichkeit der Fokussierung zu bewahren. Daher wurde das äußere Gehäuse des Objektivs vollständig entfernt und der innere Teil zwischen den



**Abbildung 3.2** Ansichten des Objektivs mit programmierbarer Apertur.

beiden Linsengruppen aufgeteilt. Die ursprüngliche mechanische Blende wurde im vollständig geöffneten Zustand fixiert. Die beiden Objektivteile werden von zwei Fassungen, die fest miteinander verschraubt werden können, gehalten. Dabei wurde besonders darauf geachtet, den ursprünglichen Abstand der beiden Linsengruppen und die Ausrichtung an der optischen Achse beizubehalten, um die Abbildungseigenschaften nicht negativ zu beeinflussen. Zwischen den beiden Fassungen kann das LCD platziert werden. Die beiden verschraubten Fassungen mit LCD werden von einer dritten Fassung gehalten, bleiben jedoch entlang der optischen Achse verschiebbar. An dieser dritten Fassung wird das Objektivbajonett für die Verbindung zur Kamera befestigt. Über die Verschiebung der beiden Linsengruppen in der dritten Fassung kann der Fokus des Objektivs eingestellt werden, wobei eine Verschiebung so weit wie möglich zur Kamera hin einer Fokussierung auf Unendlich entspricht. Die Konstruktionszeichnungen der drei Fassungen sind in Anhang B beigefügt.

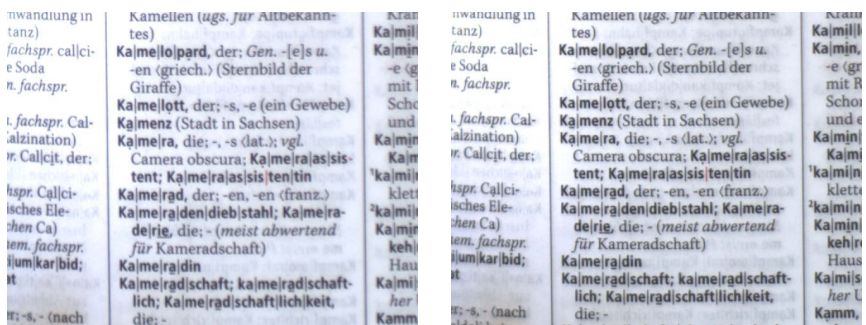
In Abbildung 3.2 sind zwei Ansichten des modifizierten Objektivs dargestellt. Die Fassungen wurden für eine geringe Verformbarkeit und eine hohe Stabilität aus Aluminium angefertigt. Die gewünschten Blendenmuster können von einem PC aus per USB-Schnittstelle an ein *Arduino* Mikrocontroller-Board übermittelt werden, das dann das LCD per SPI-Schnittstelle ansteuert. Im Rahmen dieser Arbeit wird das Objektiv stets in Verbindung mit einer *Prosilica GT4907C* CCD-Farbkamera von *Allied Vision* aus Abbildung 3.2 verwendet, die über denselben PC wie das LCD angesteuert wird.

### 3.3 Objektivcharakterisierung

In diesem Abschnitt werden die Eigenschaften des modifizierten Objektivs mit programmierbarer Apertur untersucht. Zunächst wird überprüft, inwiefern sich die grundlegenden Objektiveneigenschaften durch die Modifikation geändert haben. Anschließend werden die Eigenschaften der programmierbaren Apertur näher betrachtet.

In Abbildung 3.3 werden zwei Aufnahmen eines 70 cm von der Kamera entfernten Wörterbuches gegenübergestellt. Die Aufnahmen wurden lediglich aus den Originalaufnahmen ausgeschnitten, die Auflösung wurde nicht verringert. Eine Aufnahme wurde mit dem modifizierten Objektiv, bei dem das LCD nicht eingesetzt wurde, erstellt. Die andere mit einem Objektiv im Originalzustand, bei dem die mechanische Blende vollständig geöffnet wurde. Beide Objektive wurden manuell so eingestellt, dass die Szene so scharf wie möglich abgebildet wird.

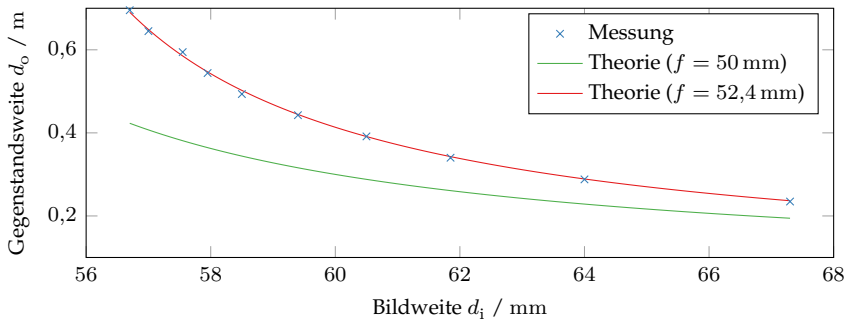
Bezüglich der erreichten Schärfe weisen die beiden Aufnahmen keine signifikanten Unterschiede auf, d. h. die Abbildungsqualität des Objektivs wurde durch die De- und Neumontage nicht maßgeblich beeinflusst. Es fällt allerdings eine zwischen den beiden Aufnahmen leicht unterschiedliche Vergrößerung auf, was auf eine Änderung der Brennweite  $f$  beim Einbau des LCDs hindeutet.



(a) Modifiziertes Objektiv.

(b) Unmodifiziertes Objektiv.

Abbildung 3.3 Aufnahmen zur Beurteilung der Abbildungsqualität.



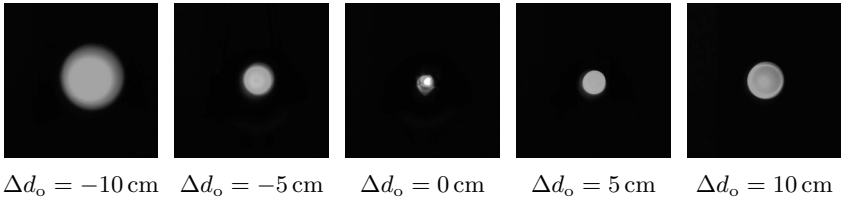
**Abbildung 3.4** Bestimmung der Brennweite  $f$  des modifizierten Objektivs.

Zur Bestimmung der Brennweite des modifizierten Objektivs wurde eine Testszene aus unterschiedlichen Entfernungen scharf abgebildet und dabei die Bildweite  $d_i$  und die Gegenstandsweite  $d_o$  erfasst. Als Apertur-ebene wurde die Ebene definiert, in der das Display eingebaut werden kann. Die ermittelten Messpunkte sind in Abbildung 3.4 dargestellt. Nach der Abbildungsgleichung ergibt sich bei bekannter Brennweite  $f$  eine scharfe Abbildung von Objekten in der Gegenstandsweite  $d_o$ , in Abhängigkeit der Bildweite  $d_i$ , gemäß:

$$d_o = \frac{f d_i}{d_i - f}. \quad (3.1)$$

Mithilfe dieses Zusammenhangs wurde die Brennweite des modifizierten Objektivs zu  $f = 52,4$  mm aus den Messpunkten mittels Regressionsanalyse bestimmt. Mit dieser Brennweite stimmt, im Gegensatz zur ursprünglichen Brennweite von  $f = 50$  mm, der theoretische Verlauf der scharf abgebildeten Gegenstandsweite in Abbildung 3.4 sehr gut mit dem gemessenen Verlauf überein, weshalb im Folgenden stets von der korrigierten Brennweite ausgegangen wird.

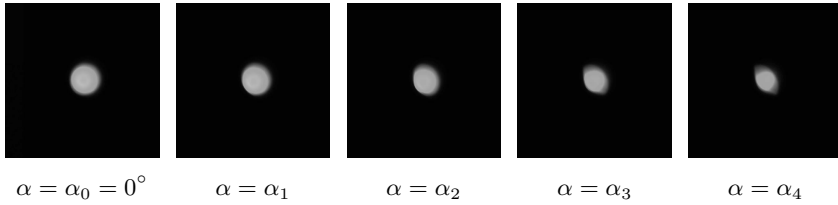
Die Fassungen des modifizierten Objektivs erlauben die Wahl der Bildweite  $d_i$  im Bereich von 52 bis 73,2 mm. Dies entspricht einer Fokussierung auf Gegenstandsweiten  $d_o$  von 18,4 cm bis auf Unendlich. Im kleinen Bereich der Bildweite mit  $d_i < f$  ist keine reelle Abbildung möglich, weshalb dieser nicht genutzt wird.



**Abbildung 3.5** Aufgenommene Impulsantworten des modifizierten Objektivs für unterschiedliche Gegenstandsweiten. Brennweite  $f = 52,4 \text{ mm}$ , Blendenzahl  $k = 1,8$ , Bildweite  $d_i = 60,3 \text{ mm}$  und Gegenstandsweite  $d_o = d_{o,\text{fok}} + \Delta d_o$ .

Die Beurteilung der Abbildungsfehler des modifizierten Objektivs geschieht anhand der PSFs, die durch Abbildung einer kleinen LED aufgenommen wurden. Abbildung 3.5 zeigt die Aufnahmen der LED aus verschiedenen Entfernungen  $d_o = d_{o,\text{fok}} + \Delta d_o$ , mit  $d_{o,\text{fok}} = 40 \text{ cm}$ . Bei den dargestellten PSFs befand sich die LED stets auf der optischen Achse, um zunächst nur den Effekt der sphärischen Aberration zu beurteilen. Außerdem wurden die Aufnahmen in Graustufenbilder umgewandelt. Insgesamt weist das Objektiv keine nennenswerten sphärischen Aberrationen auf. Inhomogenitäten in den abgebildeten PSFs können auch auf die nichtideale und inhomogene Anregung durch die LED zurückgeführt werden. Die prinzipielle Gestalt der aufgenommenen PSFs ähnelt eher den simulierten idealen PSFs aus Abbildung 2.17 als denen einer simulierten dicken Linse aus Abbildung 2.18.

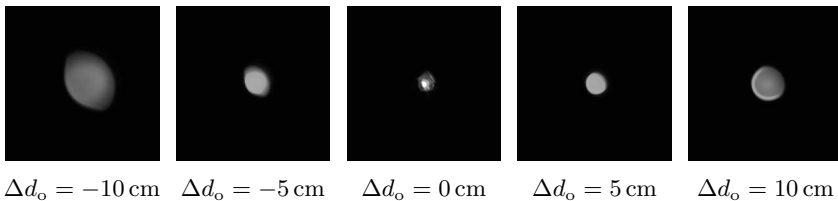
Die Abbildungseigenschaften bei schrägem Lichteinfall werden anhand der in Abbildung 3.6 dargestellten Aufnahmen untersucht. Darin wurde die LED in der senkrecht zur optischen Achse stehenden Ebene verschoben und somit aus unterschiedlichen Winkeln zur optischen Achse betrachtet. Es ist erkennbar, dass die Form mit zunehmendem Beobachtungswinkel immer stärker von der idealen Kreisform abweicht. Die Art der Verformung und die weiterhin nahezu homogene Intensitätsverteilung der PSFs entsprechen in der Simulation der Auswirkung einer Aperturblende wie in Abbildung 2.22. Der Effekt der Koma, wie in der simulierten Abbildung 2.19, ist nicht zu beobachten und kann daher beim modifizierten Objektiv vernachlässigt werden.



**Abbildung 3.6** Aufgenommene Impulsantwort des modifizierten Objektivs aus unterschiedlichen Winkeln  $\alpha$  zur optischen Achse. Auf optische Achse projizierte Objektentfernung  $d_o = 35$  cm, Bildweite  $d_i = 60,3$  mm,  $\alpha_i < \alpha_j$  für  $i < j$ , mit  $i, j = 0, \dots, 4$ .

In Abbildung 3.7 wurde zur Beurteilung des auftretenden Astigmatismus zusätzlich aus schräger Sicht die Gegenstandsweite variiert. Die grundlegende Form der PSF verändert sich nur geringfügig in Abhängigkeit der Gegenstandsweite, was die guten Abbildungseigenschaften des Objektivs unterstreicht. Wie auch schon bei Abbildung 3.5 weist die aufgenommene Intensität für  $\Delta d_o > 0$  eine größere Varianz auf als für  $\Delta d_o < 0$ , was wiederum aus der inhomogenen Anregung folgen kann.

Abbildung 3.8 zeigt schließlich eine unscharfe Farbaufnahme einer weißen LED. Diese besitzt zwar kein kontinuierliches Spektrum, regt aber je nach Bauart zwei oder mehr unterschiedliche Wellenlängen an. Die PSF weist leichte Farbsäume in Gelb und Blau auf, also gerade in den spektralen Bereichen, die durch handelsübliche weiße LEDs angeregt werden. Die Farbsäume fallen im Vergleich zur Simulation aus Abbildung 2.21 äußerst gering aus. Die chromatische Aberration des modifizierten Objektivs kann daher im Folgenden vernachlässigt werden.



**Abbildung 3.7** Aufgenommene Impulsantwort des modifizierten Objektivs aus schräger Sicht für unterschiedliche Gegenstandsweiten. Auf optische Achse projizierte Objektentfernung  $d_o = d_{o, \text{fok}} + \Delta d_o$ , Bildweite  $d_i = 60,3$  mm.

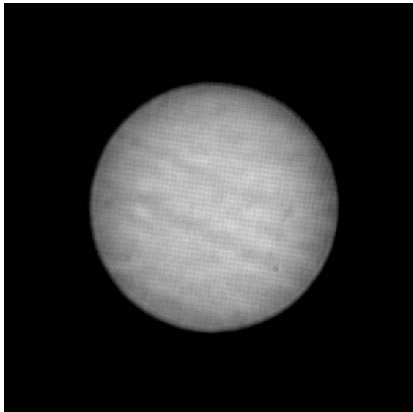




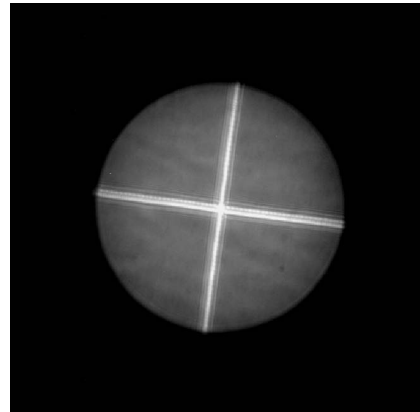
**Abbildung 3.8** Farbaufnahme der Impulsantwort des modifizierten Objektivs. Bildweite  $d_i = 60,3$  mm, Brennweite  $f = 52,4$  mm, Blendenzahl  $k = 1,8$  und Gegenstandsweite  $d_o = d_{o, \text{fok}} - 5$  cm.

### 3.3.1 Blendengeometrie

Für die folgenden Untersuchungen bezüglich der programmierbaren Apertur wurde das modifizierte Objektiv in Verbindung mit dem LCD verwendet. Abbildung 3.9 zeigt zwei damit aufgenommene PSFs. Bereits anhand der in Abbildung 3.9(a) sichtbaren Verdrahtung des LCDs, aber vor allem anhand der Orientierung des Streifenmusters aus Abbildung 3.9(b) lässt sich der Winkel zwischen den Pixelkoordinatensystemen des LCDs und des Bildsensors bestimmen. Dieser beträgt im vorliegenden Fall  $7,3^\circ$ , was später bei der in Kapitel 4 diskutierten Anwendung berücksichtigt werden muss.

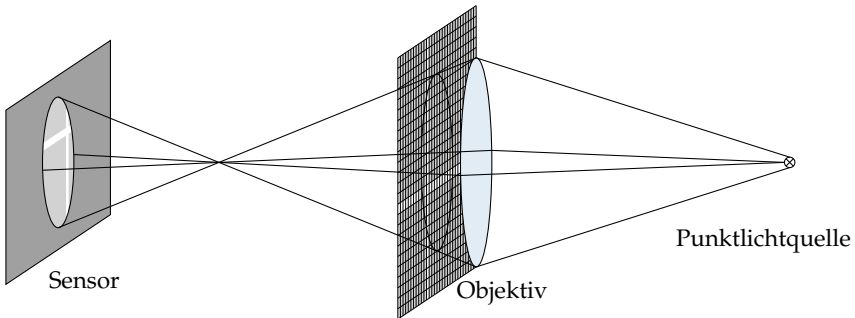


(a) Durchlässig geschaltetes LCD.



(b) LCD mit Streifenmuster.

**Abbildung 3.9** PSFs des modifizierten Objektivs mit unterschiedlichen LCD-Mustern.



**Abbildung 3.10** Schema der Anordnung zur Bestimmung der nutzbaren Displayfläche und der PSF.

Nur ein Teil der aktiven Fläche des LCDs befindet sich tatsächlich im Lichtweg des Objektivs. Zur Bestimmung dieser Fläche wird der in Abbildung 3.10 dargestellte Aufbau verwendet. Zur Aufnahme der PSF wird eine Punktlichtquelle unscharf abgebildet. Das auf dem LCD dargestellte Muster wird so unmittelbar im Bild sichtbar. Durch das Verschieben von horizontalen und vertikalen lichtdurchlässigen Streifenmustern auf dem LCD kann die sichtbare aktive Displayfläche in der Genauigkeit eines Pixels bestimmt werden (vgl. Abbildung 3.9(b)). Mit dem Aufbau sind in horizontaler Richtung 56 Pixel und in vertikaler Richtung 48 Pixel, zumindest teilweise, sichtbar. Die Pixel des verwendeten LCDs sind nicht quadratisch, sondern haben, exklusive Verdrahtung, eine Abmessung von  $360\ \mu\text{m} \times 420\ \mu\text{m}$ . Die Verdrahtung ist in beide Richtungen mit einer Stärke von  $15\ \mu\text{m}$  angegeben. Da beim verwendeten kommerziellen Objektiv von einer perfekt kreisförmigen Apertur ausgegangen werden kann, entspricht die nutzbare Fläche des LCDs einem Kreis mit einem Durchmesser von  $20,565\ \text{mm} \pm 0,3\ \text{mm}$ . Die Unsicherheit von  $0,3\ \text{mm}$  ergibt sich, da in der Aufnahme der PSF die Längen nur in ganzen Pixeln und nicht in Anteilen eines Pixels abgelesen wurden.

Die minimale Blendenzahl des ursprünglichen Objektivs wird mit  $f/1,8$  angegeben. Mit der Brennweite von  $f = 50\ \text{mm}$  ergibt sich daraus ein maximaler Durchmesser der Öffnung des ursprünglichen Objektivs von  $D = 27,8\ \text{mm}$ , was signifikant größer als die ermittelte nutzbare Displayfläche ist. Dies legt die Vermutung nahe, dass die mechanischen

Blendenlamellen, an deren Position das LCD eingebaut wurde, nicht in der effektiven Aperturbene platziert waren. Nichtsdestotrotz wird zur Modellierung der Kamera im Folgenden ein Modell mit einer einzelnen Linse verwendet, bei dem sich die codierte Apertur exakt in der Aperturbene, also der Linsenebene, befindet. Die Auswirkung der Apertur auf das ideale Eingangsbild entspricht dann mathematisch einer Faltung. Die hohe Abbildungsqualität der PSFs aus Abbildung 3.9, z. B. die Tatsache, dass die Pixelverdrahtung mit einer angegebenen Stärke von  $15\ \mu\text{m}$  gut sichtbar ist, rechtfertigen diese Vereinfachung.

### 3.3.2 Transmissionseigenschaften

Die Transmissionseigenschaften des verwendeten LCDs werden anhand der in Abbildung 3.11 dargestellten Aufnahmen des durchleuchteten LCDs in abgedunkelter Umgebung charakterisiert. Durch eine Normalisierung mit einer Referenzaufnahme wurden die Abbildungen homogenisiert, damit die Ergebnisse unabhängig von möglichen örtlichen Beleuchtungsvariationen sind.

Das transmissive LCD dämpft auch im lichtdurchlässigen Zustand (Abbildung 3.11(c)–(d)) einen Großteil des einfallenden Lichts. Hierfür sind drei Einflüsse maßgeblich verantwortlich. Durch das erste Polarisationsfilter wird nur der entsprechend der Filterrichtung polarisierte Anteil des Lichts ungedämpft durchgelassen. Mit einem idealen Polarisationsfilter würde bei unpolarisiertem Lichteinfall die Intensität um exakt 50 % sinken. Weiterhin ist aufgrund der Verdrahtung der Pixel die LCD-Fläche nicht vollständig lichtdurchlässig, was man direkt in der Aufnahme und an den periodischen Abfällen im entsprechenden Intensitätsverlauf entlang einer Bildzeile erkennen kann. Beim verwendeten LCD beträgt der Füllfaktor 93 %. Und schließlich wird vom Flüssigkristall selbst und durch das anschließende Glas Licht absorbiert. All diese Einflüsse wirken unabhängig voneinander, so dass sich die Lichteffizienz verringert:

$$\eta_{\text{hell}} = \eta_{\text{Pol}} \cdot \eta_{\text{Füll}} \cdot \eta_{\text{LC}} \quad (3.2)$$

Gemittelt über alle Pixel und Farbkanäle der Abbildungen 3.11(a) und (c) verringert sich die Intensität durch das hell geschaltete LCD auf 54,6 %. Damit ist unmittelbar ersichtlich, dass die Polarisationsfilter kein ideales Dämpfungsverhalten aufweisen.

Zum Weißabgleich wurden lediglich die einzelnen Farbkanäle mit für die unterschiedlichen Abbildungen konstanten Faktoren gewichtet. Somit lässt sich anhand des Intensitätsverlaufs aus Abbildung 3.11(d) folgern, dass die Dämpfung zusätzlich von der Wellenlänge abhängt, insbesondere der Grünanteil wird durch das LCD schwächer gedämpft als beispielsweise der Blauanteil. Daher weist die Aufnahme 3.11(c) einen leichten Grünstich auf. Dieser kann beim Betrieb des LCDs als Blende jedoch durch Anpassung der Parameter des Weißabgleichs kompensiert werden.

Umgekehrt wird durch das LCD im lichtundurchlässigen Zustand (Abbildung 3.11(e)–(f)) das einfallende Licht nicht vollständig gedämpft. Der durchgelassene Anteil kann auf drei Ursachen zurückgeführt werden:

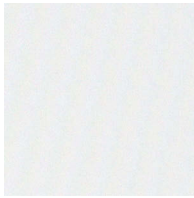
- Aufgrund des nichtidealen Verhaltens der Polarisationsfilter werden Lichtanteile entgegen der Filterrichtung nicht vollständig gedämpft.
- Einfache TN-Zellen bewirken keine ideale Drehung des Polarisationswinkels des einfallenden linear polarisierten Lichts, sondern erzeugen zusätzlich einen zirkular polarisierten Anteil, der z. B. mit der Methode nach Amaya et al. [6] bestimmt werden kann. Dieser Anteil kann teilweise durch das zweite Polarisationsfilter dringen.
- An der Pixelverdrahtung können störende Beugungen und Reflexionen auftreten, die den Polarisationszustand im LCD negativ beeinflussen.

Beim verwendeten LCD tritt im dunklen Zustand ca. 10,5 % des einfallenden Lichts durch das LCD. Dieses zusätzlich auf den Sensor fallende Licht ist dem eigentlichen Nutzsignal stets überlagert:

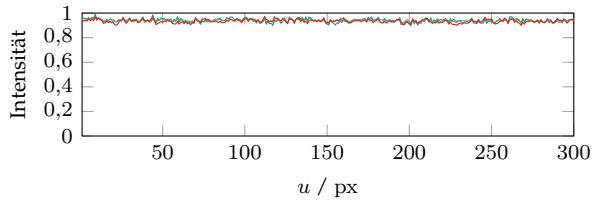
$$\mathbf{g}(\mathbf{p}) = \mathbf{g}_{\text{Nutz}}(\mathbf{p}) + \mathbf{g}_{\text{Stör}}(\mathbf{p}). \quad (3.3)$$

Dies ist für die späteren Anwendungen oftmals unerwünscht, weshalb der Störanteil kompensiert werden muss. Durch zusätzliche Aufnahme eines Störbilds der Szene bei vollständig dunklem LCD kann der szenenspezifische Störanteil erfasst werden. Unter Annahme einer linearen Kamerakennlinie kann dann das Störbild  $\mathbf{g}_{\text{Stör}}(\mathbf{p})$  von der Aufnahme mit der gewünschten Apertur  $\mathbf{g}(\mathbf{p})$  subtrahiert werden, um das Nutzsignal  $\mathbf{g}_{\text{Nutz}}(\mathbf{p})$  zu erhalten:

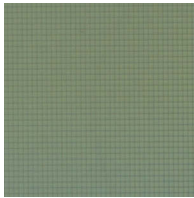
$$\mathbf{g}_{\text{Nutz}}(\mathbf{p}) = \mathbf{g}(\mathbf{p}) - \mathbf{g}_{\text{Stör}}(\mathbf{p}). \quad (3.4)$$



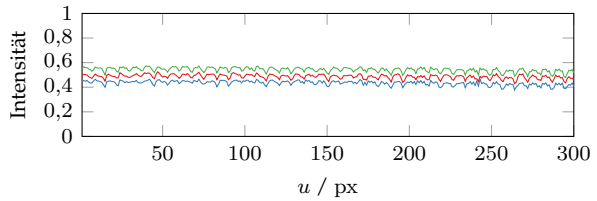
(a) Ohne LCD.



(b) Intensitätsverlauf ohne LCD.



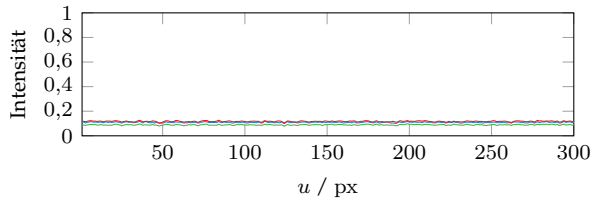
(c) Helles LCD.



(d) Intensitätsverlauf mit hellem LCD.



(e) Dunkles LCD.



(f) Intensitätsverlauf mit dunklem LCD.

**Abbildung 3.11** Aufnahmen und kanalweise Intensitätsverläufe zur Untersuchung des Transmissionsverhaltens des verwendeten LCDs.

Auf diese Weise wird außerdem die Intensitätsabweichung aufgrund des Dunkelstroms  $\mu_I T$  kompensiert.

Ein Nachteil dieser notwendigen Kompensation ist, dass sich die Kameraparameter und die Szene zwischen der eigentlichen Aufnahme und der Aufnahme des Störanteils nicht ändern dürfen, weshalb der aufgebaute Prototyp nur für statische Szenen angewendet werden kann. Da die Linearität der Kamerakennlinie sichergestellt sein muss, darf der Sensor

bei der Aufnahme nicht in die Sättigung gebracht werden. In übersättigten Bildbereichen ist keine zuverlässige Kompensation des Störanteils möglich. Aus diesem Grund und weil durch die Subtraktion des Störanteils alle Intensitäten gedämpft werden, kann mit dem Prototyp der volle Dynamikumfang des Sensors nicht ausgenutzt werden.

Des Weiteren sind, wie in Kapitel 2.4.1 diskutiert, die gemessenen Intensitäten beider Bilder  $g(\mathbf{p})$  und  $g_{\text{Stör}}(\mathbf{p})$  die Realisierungen von Zufallsprozessen, deren Varianzen zu Bildrauschen führen. Die Varianz eines Pixels des nach (3.4) bestimmten Nutzsignals entspricht der Summe der Varianzen beider Aufnahmen:

$$\sigma_{e_{g,\text{Nutz}}}^2 = \sigma_{e_g}^2 + \sigma_{e_{g,\text{Stör}}}^2. \quad (3.5)$$

Das Bildrauschen der Aufnahmen setzt sich aus dem Photonenrauschen und dem Dunkelrauschen der Kamera zusammen. Aufgrund der hohen Dämpfung des LCDs bei der Aufnahme des Störanteils  $g_{\text{Stör}}(\mathbf{p})$  ist hier die Störleistung des Photonenrauschens wesentlich geringer als bei der ursprünglichen Aufnahme  $g(\mathbf{p})$ . Die Leistung des Dunkelrauschens ist jedoch identisch, da dieselbe Belichtungszeit zur Aufnahme des Störanteils gewählt werden muss. Insgesamt wird bei Szenen mit gewöhnlicher Beleuchtung durch das Entfernen des Störanteils  $g_{\text{Stör}}(\mathbf{p})$  das SNR trotz der Verstärkung des Bildrauschens deutlich verbessert, da die Leistung des zusätzlichen Bildrauschens wesentlich geringer ist als die Leistung der Störaufnahme.

## 4 Monokulare Tiefenschätzung

Der Mensch ist in der Lage, mit dem Sehsinn neben einem zweidimensionalen Bild seiner Umgebung auch Informationen über Objektentfernungen, also in der dritten Dimension, zu sammeln. Diese Information spielt eine essentielle Rolle bei der Interaktion mit Objekten, denn sie erlaubt die präzise Abstimmung eigener Bewegungen mit der unmittelbaren Umgebung. Auch für technische Systeme ist die Kenntnis der Entfernung zu Objekten in der Umgebung von hoher Wichtigkeit.

Im Kontext des automatisierten Fahrens kommen zahlreiche Sensoren zur 3D-Erfassung der Umgebung zum Einsatz. Diese tragen maßgeblich zur Umfeldwahrnehmung bei, um Hindernisse zu erkennen, andere Verkehrsteilnehmer wie Fahrzeuge oder Fußgänger zu verfolgen und um deren zukünftiges Verhalten vorhersagen zu können [113]. Auch im Bereich der Robotik werden Entfernungsdaten zur sicheren Interaktion mit der Umgebung oder bei der gleichzeitigen Lokalisierung und Kartierung (engl. *Simultaneous Localization and Mapping*, SLAM) benötigt [16].

Bei der Mensch-Maschine-Interaktion, beispielsweise bei der Gestenerkennung, ist Tiefeninformation von hohem Wert [18, 98]. Ausschließlich erscheinungsbasierte Ansätze leiden unter dem Aperturproblem, das dazu führt, dass bestimmte Bewegungen nur schwer voneinander zu unterscheiden sind. Mithilfe von Entfernungsdaten können derartige Doppeldeutigkeiten aufgelöst werden [107].

Die unterschiedlichen Merkmale zur Tiefenwahrnehmung des Menschen können nach MacAllister [55] in physische und psychische Merkmale eingeteilt werden. Psychische Merkmale liefern Hinweise auf Entfernungen anhand der Erscheinung von Objekten im Bild. Sie sind damit unabhängig vom eigentlichen Sehprozess und beruhen auf Erfahrung. Psychische Merkmale für (relative) Entfernungen sind z. B. das Verdecken von fernen Objekten durch näher gelegene, die vertikale Abfolge von Objekten im Bild oder die Kenntnis der groben absoluten Abmessung von bekannten Objekten aus Erfahrung und Relation mit der Größe der

Erscheinung im Bild. Im Kontext eines automatischen Sichtprüfsystems fällt die Nutzung psychischer Merkmale in den Bereich der klassischen Bildverarbeitung.

Die vier physischen Merkmale der Tiefenwahrnehmung sind:

**Akkommodation:** Damit sich die von einem Objektpunkt ausgehenden Lichtstrahlen in einem Punkt auf der Netzhaut treffen, muss im Auge die Brechkraft der Linse durch Verformung entsprechend der Objektentfernung angepasst werden. Diese Anpassung der Brechkraft wird Akkommodation genannt.

**Konvergenz:** Das nach Innen Ausrichten der Augen, je näher sich ein Objekt beim Betrachter befindet, wird als Konvergenz bezeichnet.

**Binokularsehen:** Die beiden unterschiedlichen Perspektiven, aus denen ein nahes Objekt durch die zwei Augen beobachtet wird, werden im Gehirn zu einem räumlichen Seheindruck kombiniert.

**Bewegungsparallaxe:** Die Bewegungsparallaxe bezeichnet das Phänomen, dass sich bei einer Änderung der Beobachtungsposition nahe Objekte weiter verschieben als ferne.

Diese physischen Merkmale hängen direkt mit dem Sehprozess zusammen. Damit sind sie auch messtechnisch erfassbar und können durch ein geeignetes System, z. B. aus dem Gebiet des *Computational Imaging*, auch zur Tiefenextraktion genutzt werden. So beruhen Stereokamerasysteme auf dem Prinzip des Binokularsehens oder *Depth-from-Focus*- bzw. *Depth-from-Defocus*-Ansätze auf der Akkommodation.

Es existieren weitere Messprinzipien, die aufgrund fehlender Sinnesorgane, zu kurzer Laufzeiten oder fehlender Anregungssignale dem Menschen nicht zugänglich sind. Darunter fallen alle laufzeitbasierten Entfernungsmessverfahren, wie Radar (engl. *radio detection and ranging*), Lidar (engl. *light detection and ranging*) oder ToF-Kameras (engl. *time of flight*), sowie andere aktive Verfahren wie strukturierte Beleuchtung.

Aufgrund der unterschiedlichen Prinzipien wurde eine Vielzahl von Methoden zur Tiefenmessung entwickelt [38]. Abbildung 4.1 zeigt zur Übersicht eine mögliche Einteilung verschiedener optischer Tiefenmessverfahren.



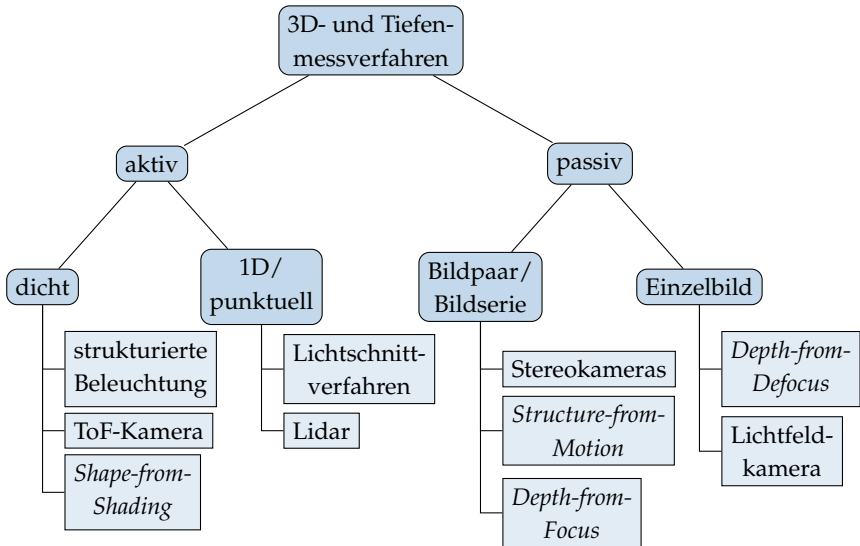


Abbildung 4.1 Übersicht über einige optische 3D- und Tiefenmessverfahren.

## 4.1 Stand der Technik

Aus den zahlreichen Prinzipien zur Tiefenmessung werden im Folgenden zunächst allgemeine kamerabasierte Verfahren kurz vorgestellt und anschließend wird genauer auf Verfahren, die den *Depth-from-Defocus*-Ansatz verfolgen, eingegangen. In beiden Fällen werden lediglich Verfahren zur dichten Tiefenmessung, also Verfahren, welche die Tiefe einer gesamten Szene ermitteln können und nicht nur an einem einzelnen Punkt, betrachtet.

### 4.1.1 Allgemeine kamerabasierte Verfahren

Stereokameras sind weit verbreitete Systeme zur dichten Tiefenmessung. Aus der Verschiebung eines Objekts in den Bildern zweier Kameras lässt sich bei bekanntem Kameraabstand, der sogenannten Basisweite, die Entfernung des Objekts durch eine Triangulation berechnen. Die Herausforderung besteht im Finden der Korrespondenzen, also der Bereiche der beiden

Bilder, die dasselbe Objekt darstellen. Hierfür wurden in der Vergangenheit zahlreiche lokale und globale Verfahren vorgestellt [25]. Insbesondere einige lokale Methoden können in Echtzeit berechnet werden. Eine aktuelle Übersicht über die Leistungsfähigkeit von Stereokorrespondenzalgorithmen bietet die Middlebury Stereo Evaluation [87]. Stereokamerasysteme zeichnen sich durch ihren relativ einfachen Aufbau mit zwei Kameras im starren Abstand der Basisweite aus. Durch die Wahl der Basisweite lässt sich die Messempfindlichkeit des Systems beeinflussen. Für eine zuverlässige Korrespondenzsuche in den Bildpaaren muss ausreichend Struktur in den Szenen vorhanden sein. Bei periodischen oder konstanten Strukturen parallel zur Verbindungslinie der Kameras ist mit Stereokameras keine robuste Tiefenschätzung möglich.

Dem Problem einer für das Messverfahren ungeeigneten Anregung durch die Szene begegnen Verfahren mit strukturierter Beleuchtung durch eine aktive Beleuchtung der Szene mit einem geeigneten Muster. Das Muster in der Szene wird von einer Kamera beobachtet, die sich versetzt zur projizierenden Lichtquelle befindet. Durch Identifikation der Merkmale des Musters im beobachteten Bild kann aus deren Position über einer Triangulation die Tiefe der Szene bestimmt werden. Der Abstand zwischen Kamera und Projektor beeinflusst die Messempfindlichkeit. Mit der aktiven Beleuchtung sind Verfahren mit strukturierter Beleuchtung weitestgehend unabhängig von der Textur und der sonstigen Beleuchtung der Szene. Daher eignen sie sich zur Erstellung hochpräziser dichter Tiefenkarten [88]. Auch bei Anwendungen, bei denen keine kontrollierte Umgebung vorliegt, aber trotzdem eine robuste Tiefenschätzung erforderlich ist, kommen häufig Verfahren mit strukturierter Beleuchtung zum Einsatz, wie beim *Microsoft Kinect* Sensor [108] oder bei der *Apple TrueDepth* Kamera [7]. Bei diesen Sensoren findet die Beleuchtung darüber hinaus im Infrarotbereich statt, weshalb sie die menschliche Wahrnehmung der Szene nicht beeinflussen.

Für *Depth-from-Focus*-Ansätze werden neben einer einzigen Kamera keine weiteren Komponenten benötigt. Die Tiefeninformation wird hier aus einer Bildserie mit unterschiedlichen Fokuseinstellungen extrahiert. Darrell und Wohn [22] wählen pro Bildregion die Fokusweite des Einzelbildes der Bildserie, welches in der betrachteten Region die meisten hochfrequenten Signalanteile aufweist, als Tiefe der Szene aus. Da klassische

Aperturen Tiefpasscharakter besitzen, kann darauf geschlossen werden, dass die betrachtete Bildregion in den anderen Aufnahmen unscharf abgebildet wurde. Alternative Gütemaße zur Bewertung der relativen Schärfe eines Bildes einer Serie, wie beispielsweise ein gradientenbasiertes Maß nach Krotkov [41], wurden auch im Kontext der automatischen Fokussierung einer Kamera entwickelt. Malik und Choi [57] falten das Spektrum der Aufnahme eines Kamerasystems mit zwei nebeneinander gelegenen Eintrittspupillen mit der optischen Transferfunktion (OTF) des Systems und leiten daraus ein Maß für die lokale Schärfe einer Aufnahme ab. Suwajanakorn et al. [93] kombinieren den *Depth-from-Focus*-Ansatz mit einem auf dem optischen Fluss basierten Verfahren, um die Bildserie einer mit der Hand gehaltenen Kamera auszurichten. Ausgehend von dieser ausgerichteten Bildserie extrahieren die Autoren sowohl ein vollständig fokussiertes Bild als auch eine Tiefenkarte der Szene.

### 4.1.2 *Depth-from-Defocus*-Verfahren

Im Gegensatz zu einigen der oben genannten Verfahren zur Tiefenmessung kommen *Depth-from-Defocus*-Verfahren mit nur einer Aufnahme einer einzigen Kamera aus, womit sie einfach zu realisieren sind. Dabei wird der Grad der Unschärfe im Bild bestimmt und damit direkt auf die Objektentfernung geschlossen. Dies ist eng verwandt mit dem Korrespondenzproblem bei Stereosystemen: Bei einem unscharf abgebildeten Objektpunkt wird ein ganzer Bildbereich belichtet. Die Anteile dieses zweidimensionalen Bereichs werden bei der Bestimmung des Grads der Unschärfe implizit dem ursprünglichen Objektpunkt zugeordnet. Diese Zuordnung geschieht bei *Depth-from-Defocus*-Verfahren in zwei Dimensionen. Damit kann mit derartigen Verfahren eine hohe Robustheit gegenüber Störungen erreicht werden und sie sind weniger anfällig bei horizontal periodischen Strukturen als beispielsweise Stereokamerasysteme, bei denen das Korrespondenzproblem nur in einer Dimension gelöst wird [89]. Aufgrund der Beschränkung auf nur eine Beobachtungsposition treten keine Verdeckungen auf, welche die Korrespondenzsuche erschweren.

Pentland [73] zeigte bereits im Jahr 1987 die Möglichkeit auf, aus dem Grad der Unschärfe in einem Bild mittels bekannter Kameraparameter auf die Tiefe der Szene zu schließen. Dabei tritt folgendes Problem auf: Es kann nicht unterschieden werden, ob Unschärfe im Bild von einer

unscharfen optischen Abbildung hervorgerufen wird oder ob diese Unschärfe von einem – möglicherweise scharf abgebildeten – Bereich der Szene stammt, der nur niedrige Ortsfrequenzen aufweist. Diesem Problem begegnet der Autor auf eine von zwei Arten. Unter der Annahme, dass sich in der ideal scharf abgebildeten Szene Intensitätssprünge an Objektkanten ergeben, lassen sich die Intensitätsverläufe in Bildregionen mit hohem Gradienten auswerten, um daraus auf die Ausdehnung der PSF zu schließen. Die PSF wird dabei als gaußförmig angenommen, die lokale Varianz ist damit ein Maß für die Unschärfe und so für den Abstand zur Fokusebene. Alternativ wird vorgeschlagen, den Grad der Unschärfe über zwei Aufnahmen derselben Szene mit unterschiedlicher Blendenzahl zu ermitteln. Durch Quotientenbildung der Fouriertransformierten zweier Bildausschnitte lässt sich der Einfluss der Szene eliminieren und die Varianz der zugrundeliegenden, wiederum als gaußförmig angenommenen PSFs, bestimmen. Zur Sicherstellung der Aufnahme derselben Szene schlägt der Autor vor, das in das Objektiv fallende Licht aufzuteilen und mit zwei separaten Bildsensoren aufzunehmen.

Das Vorgehen zur Bestimmung der Tiefe einer Szene mit zwei Aufnahmen mit unterschiedlichem Blendendurchmesser erweitern Zhou et al. [110], indem sie anstatt kreisförmiger Blenden codierte Blenden verwenden. Mithilfe eines genetischen Algorithmus bestimmen die Autoren ein Blendenpaar, das für ein angenommenes Signalmodell optimal bezüglich der Unterscheidbarkeit verschiedener Tiefen der Szene ist. Zur Bestimmung der Tiefe der Szene wird pixelweise eine Wiener-Entfaltung beider Bilder für eine angenommene Tiefe durchgeführt. Bei der korrekten Tiefe ergibt sich eine minimale Abweichung der beiden Entfaltungsergebnisse. Damit rekonstruieren die Autoren neben der Tiefenkarte der Szene auch ein vollständig fokussiertes Bild.

Levin [48] beschäftigt sich mit der Suche nach geeigneten Blendenformen zur Tiefenextraktion aus Bildpaaren. Im Gegensatz zu Zhou et al. [110] berücksichtigt Levin explizit das Niveau und die Art des Bildrauschens und zeigt eine Abhängigkeit der Leistungsfähigkeit eines Blendenpaares vom vorherrschenden Störszenario auf. Zur Bewertung wird ebenfalls die Unterscheidbarkeit verschiedener Tiefen herangezogen. Bei mehrheitlich additiver Störung, also eher dunklen Bedingungen, erweisen sich Blendenpaare mit großflächigen Öffnungen als am besten. Dagegen schneiden

bei mehrheitlich multiplikativer Störung, also bei vergleichsweise hellen Bedingungen, vor allem Blendenpaare mit räumlich weit getrennten durchlässigen Regionen am besten ab.

Durch die Aufnahme eines oder mehrerer Bilder mit unterschiedlichen Blendenformen kann ausreichend Information über die beobachtete Szene gesammelt werden, um sowohl die Tiefe als auch ein vollständig fokussiertes Abbild der Szene zu extrahieren. Die folgenden Arbeiten beschäftigen sich damit, eine oder beide dieser Informationen aus einem Einzelbild zu extrahieren.

Zhuo und Sim [115] falten das aufgenommene Bild mit einem Gaußfilter bekannter Größe. Aus dem Verhältnis der Gradienten an Kanten des gefilterten und ungefilterten Bildes kann, unter der Annahme, dass es sich bei der betrachteten Kante tatsächlich um eine abrupte Kante der Szene handelt, der Grad der Unschärfe berechnet werden. Mithilfe eines *Markov Random Field* [114] oder mithilfe eines *Joint Bilateral Filter* in Verbindung mit einem Matting-Verfahren [115] kann dann aus der dünnbesetzten Tiefenkarte eine dichte Tiefenkarte geschätzt werden. Das vorgeschlagene Verfahren ist auf abrupte Intensitätssprünge in der Szene angewiesen und beruht stark auf der Annahme, dass diese Sprünge mit Sprüngen in der Tiefe der Szene zusammenfallen.

Levin et al. [49] stellen eine Methode vor, um die Tiefe und ein vollständig scharfes Abbild der Szene aus nur einer Aufnahme mit einer codierten Blende zu extrahieren. Da aufgrund der Verluste bei einer unscharfen Abbildung die Information der Szene mit einem Einzelbild nicht vollständig erfasst werden kann, nutzen die Autoren ein probabilistisches Modell zur Beschreibung der Eigenschaften des idealen Abbilds der Szene. Anhand dieses Modells wird die Skalierung der PSF identifiziert, die das beobachtete Bild am besten erklärt. Aus dieser Skalierung kann nachfolgend die Tiefe des betrachteten Objektpunkts bestimmt und eine Entfaltung zur Rekonstruktion des scharfen Abbilds der Szene durchgeführt werden. Das vorgeschlagene Kriterium zur Auswahl der Blendenform basiert auf der anhand des probabilistischen Modells der Szene berechneten Kullback-Leibler-Divergenz der Wahrscheinlichkeitsverteilungen zweier unscharfer Abbilder. Eine hohe Divergenz unterschiedlich stark defokussierter Abbilder bedeutet, dass die verschiedenen Tiefen gut voneinander unterscheidbar sind. Um Randverzerrungen der PSF zu vermeiden, nutzen die

Autoren nicht die volle zur Verfügung stehende Fläche der Apertur zur Codierung aus, weshalb das vorgestellte System empfindlich auf Fehler bei der Bestimmung der Skalierung der PSF reagiert. Eine Analyse des erreichten Schätzfehlers führen die Autoren nicht durch.

Veeraraghavan et al. [100] nutzen ebenfalls eine codierte Blende, um aus Einzelbildern auf die Tiefe der Szene zu schließen. Die Blendenform wurde bezüglich des Frequenzgehalts bei hohen Frequenzen optimiert. Damit liegt hier der Fokus auf der Konservierung hochfrequenter Signalanteile anstatt der besseren Differenzierbarkeit von Tiefen wie beim oben beschriebenen Ansatz von Levin et al. [49]. Veeraraghavan et al. [100] verwenden ein kombiniertes Gütemaß, bestehend aus einer Likelihood-Funktion und der Krümmung der Gradientenverteilung der betrachteten Bildregion, um die wahrscheinlichste Skalierung der PSF in der Aufnahme zu identifizieren und ein scharfes Abbild zu rekonstruieren. Auch in dieser Arbeit wird die Güte der Tiefenschätzung nicht bewertet.

Neben der örtlichen Codierung der Blende ist auch mithilfe von Farbfiltern, üblicherweise in den drei Grundfarben Rot, Grün und Blau, eine spektrale Codierung möglich [5, 8, 46, 56]. Auf diese Weise werden die spektralen Anteile des von einem nicht im Fokus liegenden Punkt ausgehenden Lichts auf unterschiedliche Bereiche des Sensors abgebildet. Aus dem Abstand der Abbilder des Objektpunkts kann wiederum dessen Entfernung zur Fokusebene berechnet werden. Dies entspricht somit der klassischen Korrespondenzsuche bei Stereokameras, wobei anstelle der Korrespondenz in zwei Kamerabildern die Korrespondenz zwischen den Farbkanälen eines Farbbildes gesucht wird. Bei entsprechender Anregung erfolgt diese Korrespondenzsuche in zwei Dimensionen, da jeder Farbkanal mit den anderen beiden in Beziehung gesetzt werden kann. Werden jedoch nur zwei oder gar nur ein Farbkanal angeregt, ist keine robuste Korrespondenzsuche möglich, weshalb dieses Verfahren stark abhängig von den in der Szene vorkommenden Farben ist. Ein weiterer Nachteil ist, dass die Farbfilter einfallendes Licht und somit Signalintensität dämpfen. Weiterhin beruht das Verfahren bei klassischen Farbsensoren in hohem Maße auf der Interpolation fehlender Farbinformation durch das Demosaicing.

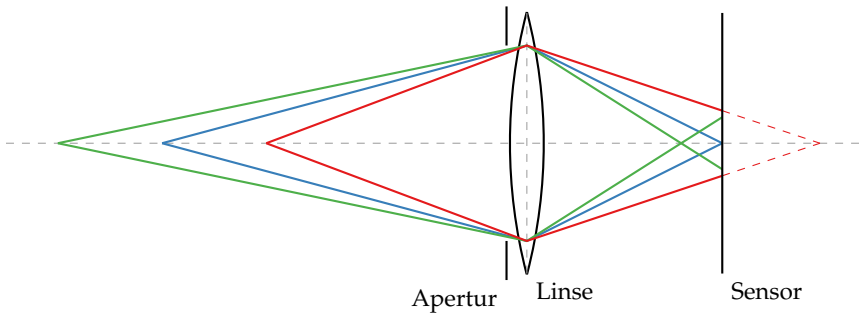
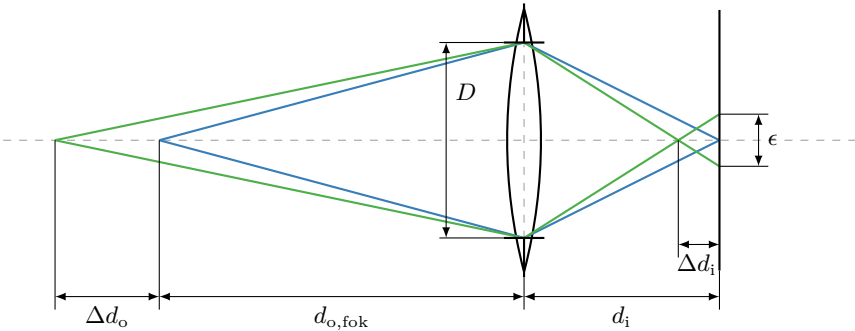


Abbildung 4.2 Kameramodell der *Depth-from-Defocus*-Ansätze.

## 4.2 Messprinzip

Das Messprinzip aller *Depth-from-Defocus*-Ansätze basiert auf dem vereinfachten Kameramodell, das in Abbildung 4.2 dargestellt ist. Nur Objekte im Abstand der Fokusweite  $d_{o,\text{fok}}$  werden scharf auf dem Sensor abgebildet (blau). Die von Objekten geringeren Abstands ausgehenden Strahlen treffen sich hinter der Sensorebene (rot), wohingegen sich die Strahlen von fernerer Objekten vor der Sensorebene treffen (grün). Im unscharfen Fall wird bei Verwendung einer klassischen Blende ein kreisförmiger Bereich, auch Unschärfescheibchen oder engl. *Circle of Confusion*, des Sensors belichtet. Je weiter sich ein Objekt von der Fokusebene entfernt befindet, desto größer ist das Unschärfescheibchen. Bei Nutzung einer codierten Blende entspricht der belichtete Bereich der gewählten Form der Blende. Im Folgenden wird zunächst der Fall einer kreisförmigen Blende betrachtet.

Zur Herleitung des Zusammenhangs zwischen der Größe des Unschärfescheibchens und der Objektentfernung muss unterschieden werden, ob sich das Objekt vor oder hinter der Fokusebene befindet. Abbildung 4.3 zeigt die relevanten Größen für den Fernfall mit  $d_o = d_{o,\text{fok}} + \Delta D_o > d_{o,\text{fok}}$ , wobei die Blende zur Vereinfachung exakt in der Aperturebene liegt. Die grün eingezeichneten ähnlichen Dreiecke auf der Bildseite liefern den Zusammenhang



**Abbildung 4.3** Größen zur Herleitung des Fernfalls mit  $d_o > d_{o,\text{fok}}$ .

$$\frac{d_i - \Delta d_i}{D} = \frac{\Delta d_i}{\epsilon} = \frac{d_i - (d_i - \Delta d_i)}{\epsilon} \quad (4.1)$$

$$\frac{1}{d_i - \Delta d_i} = \frac{1}{d_i} \frac{D + \epsilon}{D}. \quad (4.2)$$

Mit der Abbildungsgleichung der dünnen Linse

$$\frac{1}{f} = \frac{1}{d_o(\epsilon)} + \frac{1}{d_i(\epsilon)} = \frac{1}{d_{o,\text{fok}} + \Delta d_o} + \frac{1}{d_i - \Delta d_i} \quad (4.3)$$

folgt daraus

$$\frac{1}{d_o(\epsilon)} = \frac{1}{f} - \frac{1}{d_i - \Delta d_i} \quad (4.4)$$

$$= \frac{1}{f} - \frac{1}{d_i} \frac{D + \epsilon}{D}. \quad (4.5)$$

Drückt man nun in (4.5) die Bildweite  $d_i$  durch die Abbildungsgleichung aus und löst nach  $d_o(\epsilon)$  auf, erhält man

$$d_o(\epsilon) = \frac{fDd_{o,\text{fok}}}{fD - \epsilon(d_{o,\text{fok}} - f)}. \quad (4.6)$$

Die örtliche Verteilung des Lichts wird durch den Sensor diskret abgetastet. Das kleinste erfassbare Element ist ein Pixel. Mit der Pixelabmessung  $w_p$



und mit  $\epsilon = s w_p, s \in \mathbb{N}_0$ , folgt schließlich die Messgleichung für den Fernfall:

$$d_o(s) = \frac{fDd_{o,\text{fok}}}{fD - s w_p (d_{o,\text{fok}} - f)}, \quad (4.7)$$

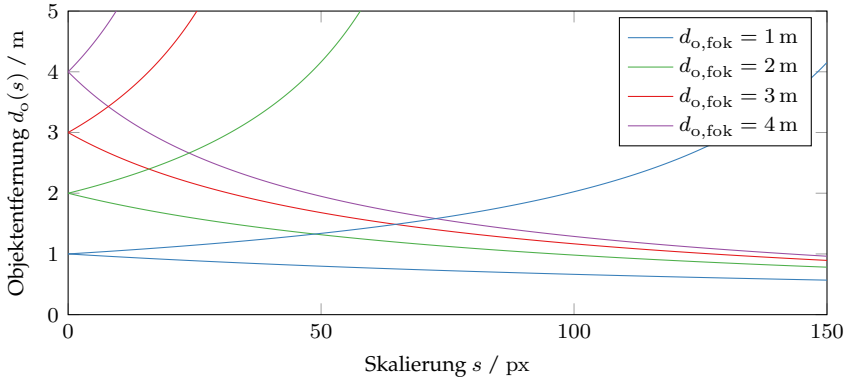
mit der Skalierung  $s$  des Unschärfescheibchens in der Einheit Pixel. Analog lässt sich die Messgleichung für den Nahfall  $d_o \leq d_{o,\text{fok}}$  aufstellen. Insgesamt ergibt sich

$$d_o(s) = \begin{cases} \frac{fDd_{o,\text{fok}}}{fD - s w_p (d_{o,\text{fok}} - f)} & \text{für } d_o > d_{o,\text{fok}}, \\ \frac{fDd_{o,\text{fok}}}{fD + s w_p (d_{o,\text{fok}} - f)} & \text{für } d_o \leq d_{o,\text{fok}}. \end{cases} \quad (4.8)$$

Bei einer Skalierung von  $s = 0$  ergeben beide Fälle der Gleichung mit  $d_o(0) = d_{o,\text{fok}}$  dieselbe Objektentfernung.

Sofern also in einer Aufnahme die Skalierung  $s$  des Unschärfescheibchens bestimmt werden kann, kann bei bekannten Objektiv- und Sensorparametern mithilfe der Kennlinien (4.8) die Entfernung am betrachteten Bildpunkt ermittelt werden. Ein Problem dabei ist, dass ohne Wissen über die grobe Szenenkonfiguration nicht bekannt ist, ob sich Objekte vor oder hinter der Fokusebene befinden. Einer bestimmten Skalierung  $s$  können daher immer zwei Objektentfernungen  $d_o$  zugeordnet werden. Falls in der konkreten Messaufgabe der Messraum begrenzt ist, sollte die Fokusweite  $d_{o,\text{fok}}$  des Objektivs so gewählt werden, dass sie am Rand dieses Messraums liegt. Auf diese Weise wird sichergestellt, dass sich Objekte immer nur vor bzw. hinter der Fokusweite befinden und dass aufgrund dieses Vorwissens immer der zutreffende Teil der Messkennlinie (4.8) genutzt wird.

In Abbildung 4.4 sind die Kennlinien für verschiedene Fokusweiten aufgetragen. Die Mehrdeutigkeit äußert sich dadurch, dass für eine feste Fokusweite eine mit der Skalierung  $s$  ansteigende und eine abfallende Kennlinie existiert. Die Kennlinien für unterschiedliche Fokusweiten verdeutlichen aber, dass beim Entwurf eines 3D-Bildverarbeitungssystems durch Änderung dieses Parameters eine flexible Anpassung der Kennlinien an einen gewünschten Messraum nach [30] möglich ist. Gerade die Fokusweite ist eine Größe, die sich, im Gegensatz z. B. zur Pixelabmessung  $w_p$ , auf einfache Weise anpassen lässt.



**Abbildung 4.4** Objektentfernung  $d_o$  in Abhängigkeit des Durchmessers  $s$  des Unschärfescheibchens für unterschiedliche Fokuseinstellungen  $d_{o,\text{fok}}$  ( $f = 50 \text{ mm}$ ,  $D = f/1,8$ ,  $w_p = 7,4 \mu\text{m}$ ).

Die Messempfindlichkeit folgt durch Ableitung der Messkennlinien nach der Skalierung  $s$ :

$$S(s) = \frac{\partial d_o(s)}{\partial s} = \begin{cases} \frac{fDg_{\text{fok}}^2}{(fD-s w_p(d_{o,\text{fok}}-f))^2} & \text{für } d_o > d_{o,\text{fok}} , \\ \frac{-fDg_{\text{fok}}^2}{(fD+s w_p(d_{o,\text{fok}}-f))^2} & \text{für } d_o \leq d_{o,\text{fok}} . \end{cases} \quad (4.9)$$

Intuitiver interpretierbar ist die Messempfindlichkeit in Abhängigkeit der Objektentfernung  $d_o$ . Der Ausdruck dafür ergibt sich nach Auflösen von (4.8) nach der Skalierung  $s$  und Einsetzen in (4.9) zu

$$S(d_o) = \begin{cases} \frac{d_o^2}{fD} & \text{für } d_o > d_{o,\text{fok}} , \\ \frac{-d_o^2}{fD} & \text{für } d_o \leq d_{o,\text{fok}} . \end{cases} \quad (4.10)$$

Somit hängt nur das Vorzeichen der Empfindlichkeit bezüglich der Objektentfernung von der Fokusweite  $d_{o,\text{fok}}$  ab. In Abbildung 4.4 äußert sich dies dadurch, dass für eine feste Objektentfernung  $d_o$  der Betrag der Steigung für alle Kennlinien gleich ist. Das Verschieben der Fokusweite zur Auflösung der Mehrdeutigkeit hat also keinen Einfluss auf die Tiefenempfindlichkeit des Systems.

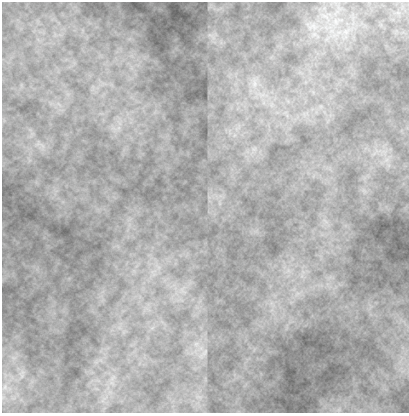
Wie auch bei anderen auf der Triangulation basierenden Verfahren nimmt die Messempfindlichkeit quadratisch mit der Entfernung zu. Fehler bei der Bestimmung der Skalierung  $s$  führen also zu immer größeren Entfernungsfehlern, je weiter das Objekt entfernt ist. Dem kann entsprechend (4.10) entgegengewirkt werden, indem die Brennweite  $f$  und/oder der Durchmesser der Apertur  $D$  vergrößert wird. Aus diesem Grund sollte die zur Verfügung stehende Apertur eines Objektivs durch Öffnen der Blende voll ausgenutzt werden.

Die Herausforderung der *Depth-from-Defocus*-Ansätze besteht in der Bestimmung der lokalen Skalierung  $s$  anhand eines aufgenommenen Bildes. Dem neuartigen Denkmuster des *Computational Imaging* folgend ist der nächste logische Schritt, durch eine geeignete Codierung der Blendenform eine robuste Bestimmung der Skalierung zu unterstützen, beispielsweise durch die Nutzung einer Blendenform, die eine optimale Unterscheidung unterschiedlicher Entfernungen ermöglicht. Die optimalen Formen hängen in hohem Maße vom Algorithmus zur Identifikation der Skalierung, insbesondere von den herangezogenen Bildmerkmalen sowie von den äußeren Beleuchtungsbedingungen [48] ab.

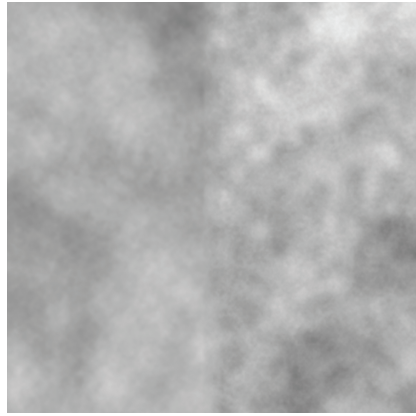
Auf diese Weise kann auch teilweise dem Problem begegnet werden, dass nicht unterschieden werden kann, ob Bildregionen mit wenig hochfrequenten Signalanteilen aufgrund einer unscharfen Abbildung oder aufgrund unzureichender Anregung der Szene entstehen. Sofern eine Blendenform gefunden werden kann, deren Auftreten in der beobachteten Szene unwahrscheinlich ist, treten die Charakteristika der verwendeten Form nur bei einer unscharfen Abbildung auf. Somit lässt sich dieser Fall von einem niederfrequenten Bereich der Szene unterscheiden.

## 4.3 Algorithmus

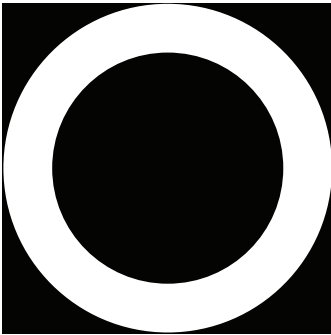
In diesem Abschnitt wird der im Rahmen dieser Arbeit entworfene Algorithmus zur Bestimmung der Skalierung erläutert. Der Algorithmus wurde anhand simulierter Eingangsdaten entworfen und soll daher auch anhand des synthetischen Beispielszenarios aus Abbildung 4.5, bestehend aus zwei texturierten Ebenen, die in unterschiedlichem Abstand senkrecht zur Kamera platziert sind, vorgestellt werden. Die Aufnahmen und die Ground-Truth-Tiefendaten wurden mit der in Kapitel 2 vorgestellten



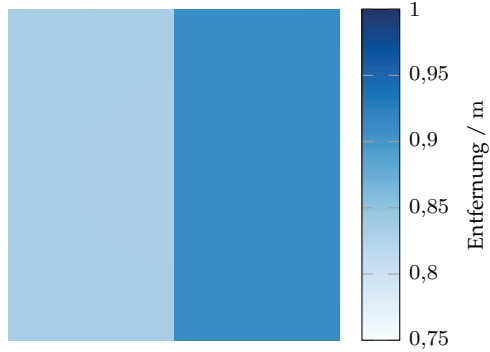
(a) Ideale Lochkameraaufnahme.



(b) Aufnahme mit dünner Linse.



(c) Verwendete Aperturform.



(d) Ground-Truth-Tiefe.

**Abbildung 4.5** Simuliertes Beispielszenario zur Erprobung des Algorithmus zur Tiefenschätzung. Die für (b) simulierte Kamera ist auf eine Entfernung von  $d_{o,foK} = 1$  m fokussiert. Die Ebenen befinden sich in den Entfernungen  $d_{o1} = 0,908$  m und  $d_{o2} = 0,832$  m von der Kamera, was Skalierungen von  $s_1 = 20$  px und  $s_2 = 40$  px entspricht (Brennweite  $f = 50$  mm, Blendenzahl  $k = 1,8$ , Pixelabmessung  $w_p = 7,4 \mu\text{m}$ ).

Raytracing-Simulationsumgebung erzeugt. Die verwendeten Texturen sorgen für eine breitbandige Anregung. Damit in den folgenden Abschnitten die Zwischenschritte leichter visualisiert werden können, wird eine rotationssymmetrische Apertur gewählt. Das Szenario ist bewusst einfach gehalten, um zunächst die grundlegende Funktion des Algorithmus anhand dieser Szene untersuchen zu können.

Der Algorithmus läuft zunächst unabhängig für jedes Pixel der Aufnahme ab. Die Identifikation der Skalierung  $s$  der PSF wird auf den Vergleich zweier Spektren, die auf Grundlage des im folgenden Abschnitt 4.3.1 beschriebenen Signalmodells bestimmt werden, zurückgeführt. In Abschnitt 4.3.2 wird auf den eigentlichen Vergleich und die dabei verwendeten Gütemaße eingegangen. In Abschnitt 4.3.3 wird schließlich aufgezeigt, wie das Schätzergebnis durch eine Verknüpfung der zuvor unabhängig voneinander bestimmten Skalierungen verbessert werden kann.

### 4.3.1 Signalmodell

Das geometrische Modell aus Abbildung 4.2 beschreibt den Verlauf einzelner Lichtstrahlen, die vom Sensor erfasst werden und so das Bild formen. Wird als grundlegende Größe das Bildsignal  $g$  anstelle der Strahlen gewählt, lässt sich das geometrische Modell in das Signalmodell aus Abbildung 4.6 überführen.

Das ideal scharfe Bildsignal  $g_{\text{ideal}}(\mathbf{p})$ , mit  $\mathbf{p} = (u, v)^T$ , wird mit der orts- und skalierungsabhängigen Punktverschmierungsfunktion  $PSF_s(\mathbf{p})$  des abbildenden Systems gefaltet. Mit der Skalierung  $s$  als Index soll ausgedrückt werden, dass die grundlegende Form der PSF unabhängig von der Skalierung stets dieselbe ist. Mit dem Index  $s$  wird lediglich die Skalierung dieser grundlegenden Form in der Einheit Pixel angegeben.

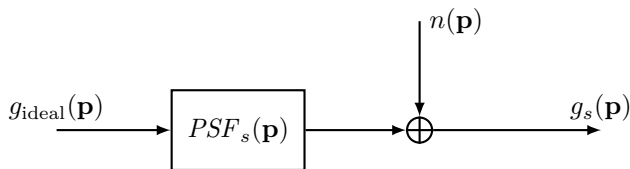


Abbildung 4.6 Signalflussgraph des *Depth-from-Defocus*-Ansatzes.

Dem beobachteten Bildsignal  $g_s(\mathbf{p})$  ist zusätzlich die zum Eingang als unkorreliert angenommene Störgröße  $n$  überlagert. Insgesamt also

$$g_s(\mathbf{p}) = g_{\text{ideal}}(\mathbf{p}) * PSF_s(\mathbf{p}) + n(\mathbf{p}) \quad (4.11)$$

$$\begin{array}{c} \circ \\ | \\ \bullet \end{array} \quad G_s(\mathbf{f}) = G_{\text{ideal}}(\mathbf{f}) \cdot OTF_s(\mathbf{f}) + N(\mathbf{f}), \quad (4.12)$$

mit der Ortsfrequenz  $\mathbf{f} = (f_u, f_v)^T$ . In diesen Gleichungen ist lediglich das beobachtete Bild  $g_s(\mathbf{p})$  und die grundlegende Form der PSF bzw. deren Fouriertransformierten bekannt. Für lineare Koordinatentransformationen  $\mathbf{A}$  im Ortsbereich folgt im Ortsfrequenzbereich [10]:

$$\mathcal{F}\{g(\mathbf{A}\mathbf{p})\} = |\det \mathbf{A}|^{-1} G\left(\left(\mathbf{A}^{-1}\right)^T \mathbf{f}\right). \quad (4.13)$$

Für den Fall einer gleichförmigen Skalierung entlang beider Achsen um einen Faktor  $a$  gilt

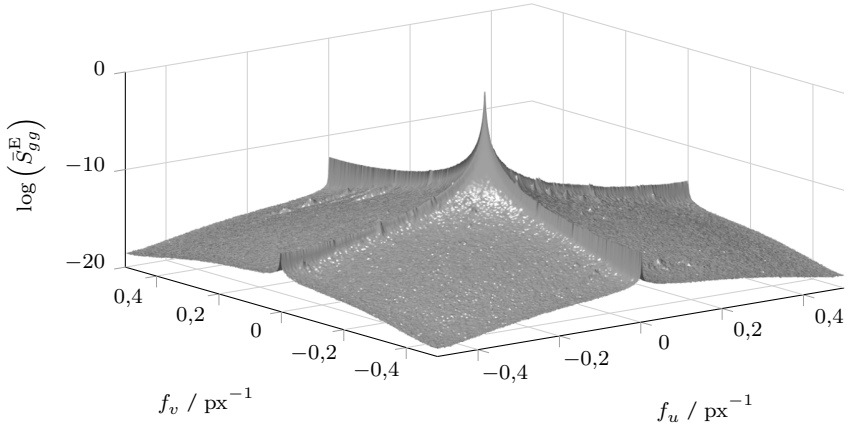
$$\mathbf{A} = a \cdot \mathbf{I}, \quad \left(\mathbf{A}^{-1}\right)^T = \frac{1}{a} \cdot \mathbf{I}, \quad |\det \mathbf{A}|^{-1} = \left|\frac{1}{a^2}\right|. \quad (4.14)$$

Die Skalierungseigenschaft der Fouriertransformation im Eindimensionalen lässt sich also unmittelbar ins Zweidimensionale übertragen. Die lokale Bestimmung der Skalierung  $s$  der PSF, die die Tiefe der Szene am Ort  $\mathbf{p}$  beschreibt, kann somit auch im Ortsfrequenzbereich erfolgen. Dazu müssen, aufgrund der Unterbestimmtheit der Gleichungen (4.11) und (4.12), weitere Modellannahmen getroffen werden, die im Folgenden beschrieben werden.

Die Beträge der Spektren der Abbilder natürlicher bzw. realer Szenen weisen einen  $1/f$ -förmigen Abfall hin zu hohen Frequenzen auf [85, 96]. Damit in Einklang liefert das Riemann-Lebesgue'sche Lemma für stückweise stetige Funktionen eine obere Schranke des Spektrums der Form

$$|G(\mathbf{f})| \leq \frac{M}{|\mathbf{f}|^{K+1}}, \quad (4.15)$$

mit der Anzahl  $K$  der existierenden stückweise stetigen Ableitungen von  $g$  und einer Konstanten  $M \geq 0$  [72, 79]. Es lässt sich keine Aussage über die



**Abbildung 4.7** Mittleres Energiedichtespektrum des Bilddatensatzes ohne zusätzliche Fensterung.

Phase des Spektrums eines Abbilds einer realen Szene treffen, die angeben würde, wo im Bild bestimmte Signalanteile lokalisiert sind [79]. Daher ist es sinnvoll, Gleichung (4.12) in Abhängigkeit der Energiedichtespektren aufzustellen:

$$S_{g_s g_s}^E(\mathbf{f}) = S_{g_{\text{ideal}} g_{\text{ideal}}}^E(\mathbf{f}) \cdot |OTF_s(\mathbf{f})|^2 + S_{nn}^E(\mathbf{f}). \quad (4.16)$$

Damit müssen die Spektren der Abbilder lediglich betragsmäßig betrachtet werden. Die Energiedichten können durch die Periodogramme der beobachteten Bilder approximiert werden.

Um die angenommene Energiedichte  $S_{g_{\text{ideal}} g_{\text{ideal}}}^E(\mathbf{f})$  konkret angeben zu können, wurde anhand eines Datensatzes aus 2500 Aufnahmen des Datensatzes von Vogel und Schiele [101] sowie privaten Aufnahmen das mittlere Energiedichtespektrum  $\bar{S}_{gg}^E(\mathbf{f})$  von Abbildern realer Szenen ermittelt. Abbildung 4.7 zeigt das resultierende mittlere Energiedichtespektrum in logarithmischer Darstellung. Das Energiedichtespektrum wurde dabei unter Vernachlässigung des Gleichanteils auf eine Signalenergie von eins normiert.

Eine Konzentration der Signalenergie entlang der  $f_u$ - und  $f_v$ -Achse tritt, aufgrund häufig in der Natur vorkommenden vertikalen und horizontalen Strukturen, beispielsweise dem Horizont oder vertikal wachsenden

Bäumen [96] auch in Abbildern natürlicher Szenen, wie sie hier zum Großteil verwendet wurden, auf. Die hier zu beobachtenden Flanken entlang der Achsen sind allerdings hauptsächlich auf den Leck-Effekt aufgrund der Fensterung und der impliziten periodischen Fortsetzung bei der Berechnung der diskreten Fouriertransformation (DFT) zurückzuführen [77, 79]. Diese Signalanteile entstammen nicht dem eigentlichen Inhalt der Szenen und müssen daher als Störungen aufgefasst werden, die die Identifikation der Skalierung der PSF negativ beeinflussen können. Um diesen Effekt zu reduzieren wurde jedes Bild vor der Transformation mit einem diskreten Hann-Fenster  $w_{N,k}$  nach

$$w_{N,k} = \frac{1}{2} \left( 1 - \cos \left( 2\pi \frac{k}{N} \right) \right), \quad (4.17)$$

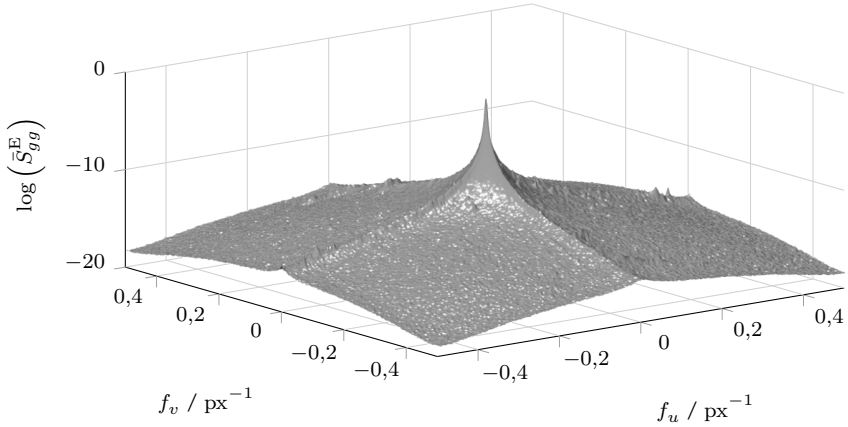
mit der Fensterbreite  $N$ , gewichtet. Das Fenster ist rotationssymmetrisch konstruiert und unterdrückt so die Randeffekte bei der periodischen Fortsetzung der DFT. Aus den damit berechneten Spektren wurde dann wiederum der Mittelwert über alle Bilder des Datensatzes ermittelt. Abbildung 4.8 zeigt das entsprechend korrigierte mittlere Energiedichtespektrum  $\bar{S}_{gg}^E(\mathbf{f})$  in logarithmischer Darstellung, das im Folgenden als Annahme für das Energiedichtespektrum des ideal scharfen Bildes  $S_{g_{\text{ideal}}g_{\text{ideal}}}^E(\mathbf{f})$  dient.

Das Betragsquadrat der OTF verhält sich invariant gegenüber einer Invertierung der Ortsachsen der PSF, was beim Übergang zwischen Fern- und Nahfall geschieht (vgl. Abschnitt 4.2). Nach (4.13) und (4.14) ergibt eine Invertierung der Achsen im Ortsbereich ( $\mathbf{A} = -\mathbf{I}$ ) zwar eine Invertierung der Ortsfrequenzachsen, diese kann jedoch nicht identifiziert werden, da das Betragsquadrat der OTF, also der Fouriertransformierten der reellwertigen PSF, stets eine gerade Funktion ist [79]:

$$g(\mathbf{p}) \in \mathbb{R} \quad \rightarrow \quad G^*(\mathbf{f}) = G(-\mathbf{f}) \quad \rightarrow \quad |G(\mathbf{f})| = |G(-\mathbf{f})|. \quad (4.18)$$

Daher kann anhand der Energiedichtespektren, auch bei Nutzung einer asymmetrischen Aperturform, nicht unterschieden werden, ob sich ein Objekt vor oder hinter der Fokusebene befindet, obwohl sich die zugehörigen PSFs durch eine Invertierung der Ortsachsen unterscheiden. Hierzu würde mehr Information über die Phase der ideal abgebildeten Szene benötigt, die mit einer Einzelaufnahme jedoch nicht erfasst werden kann.





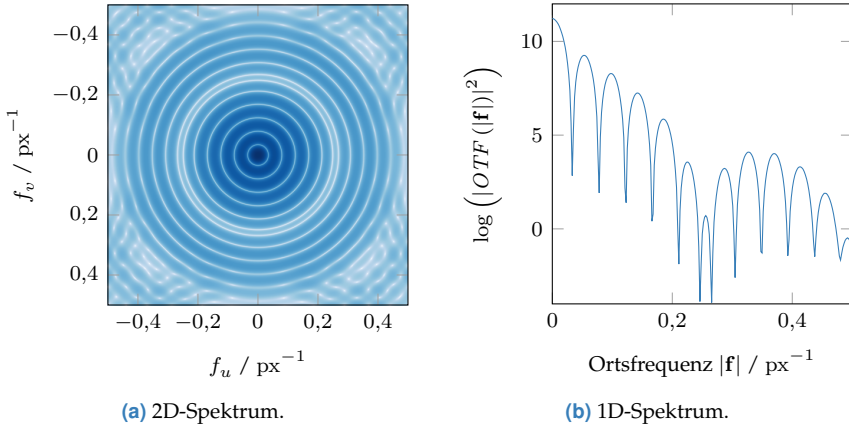
**Abbildung 4.8** Mittleres Energiedichtespektrum des Bilddatensatzes mit zusätzlicher Fensterung.

Im Folgenden werden Energiedichtespektren zur besseren Interpretierbarkeit als eindimensionale Projektionen, berechnet durch Integration in Tangentialrichtung, dargestellt. Daher wurde im Beispielszenario aus Abbildung 4.5 eine rotationssymmetrische Aperturform gewählt, die aufgrund von (4.13) ebenfalls ein rotationssymmetrisches Spektrum aufweist [10], welches somit ohne Informationsverlust eindimensional dargestellt werden kann. Das logarithmierte Betragsquadrat der verwendeten Blendenform ist in Abbildung 4.9 zur Verdeutlichung dieser Eigenschaft in zwei- und eindimensionaler Darstellung für eine Skalierung von  $s = 32 \text{ px}$  angegeben.

Die Texturen der Ebenen aus Abbildung 4.5 wurden zufällig nach dem Verfahren der *Fractional Brownian Motion* [91] erzeugt: Der Grauwert an einem Ort  $\mathbf{x}$  der Textur ergibt sich aus der Überlagerung von  $K$  skalierten und amplitudengedämpften Kopien einer Zufallsfolge  $n(\mathbf{x})$  entsprechend

$$g_{\text{fBm}}(\mathbf{x}) = \sum_{i=0}^{K-1} a^i \cdot n(b^i \cdot \mathbf{x}). \quad (4.19)$$

Mithilfe des Dämpfungsparameters  $a > 0$  und des Skalierungsparameters  $b > 0$  kann der spektrale Verlauf der Textur flexibel eingestellt werden.

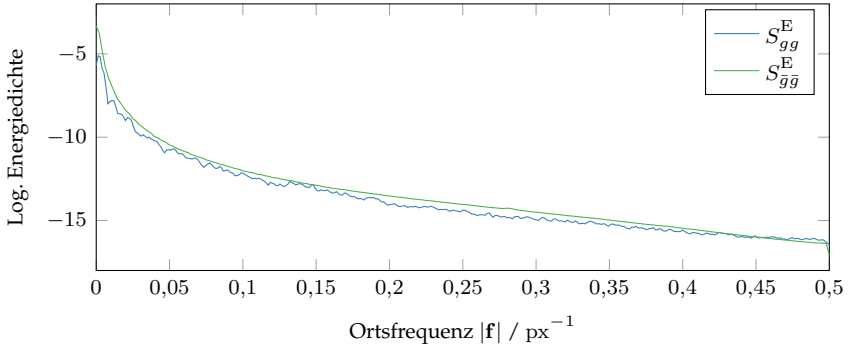


**Abbildung 4.9** Logarithmiertes Betragsquadrat des Spektrums der im Beispielszenario verwendeten Aperturform der Skalierung  $s = 32$  px.

Für  $a < 1$  und  $b > 1$  ergibt sich eine Textur, deren Energiedichtespektrum zu hohen Frequenzen hin abfällt. Im eindimensionalen Vergleich in Abbildung 4.10 weist das Energiedichtespektrum der Lochkameraaufnahme aus Abbildung 4.5(a) einen sehr ähnlichen Verlauf zum angenommenen idealen Energiedichtespektrum des idealen Bildes  $\bar{S}_{gg}^E(\mathbf{f})$  auf.

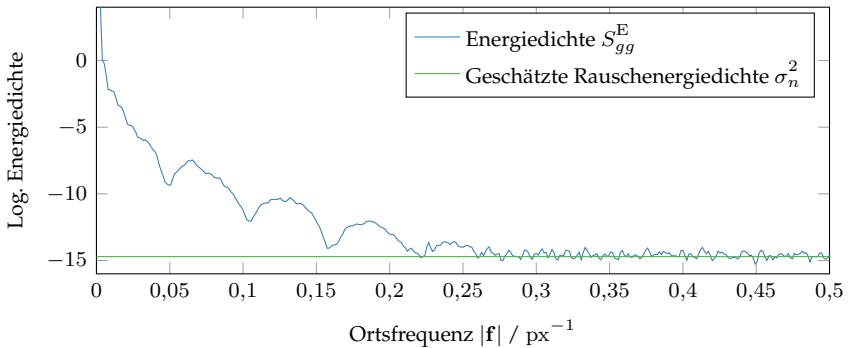
Die additive Störgröße  $n(\mathbf{p})$  wird als mittelwertfreies Gauß'sches weißes Bildrauschen modelliert. Darüber hinaus wird das Bildrauschen als stationär angenommen. Damit kann die Störgröße über die bezüglich der Frequenz konstanten Rauschenergiedichte  $\sigma_n^2$  charakterisiert werden. Aufgrund des stetigen Abfalls des angenommenen idealen Energiedichtespektrums kann abhängig vom konkreten Rauschniveau nur ein Bereich der Ortsfrequenzebene bis zu einer bestimmten Frequenz zur Bestimmung der Skalierung  $s$  genutzt werden. Bei höheren Ortsfrequenzen geht das Nutzsignal im Bildrauschen unter.

Die Energiedichte des Bildrauschens kann nicht wie die Energiedichte der Szene vorab abgeschätzt werden, da sie von den externen Beleuchtungsbedingungen abhängt. Daher muss die Rauschenergiedichte  $\sigma_n^2$  anhand des aufgenommenen Bildes ermittelt werden. Die von Chen et al. [17] vorgestellte, effiziente Methode, die auch in dieser Arbeit angewendet wird, identifiziert die Rauschenergiedichte anhand der Verteilung



**Abbildung 4.10** Vergleich des Spektrums der verwendeten Rauschtextur mit dem ermittelten mittleren Spektrum realer Szenen.

der Eigenwerte der Kovarianzmatrix des Bildes. Abbildung 4.11 zeigt das Ergebnis der Schätzung der Rauschenergiedichte anhand der Aufnahme 4.5(b) des Beispielszenarios. Es kann eine hohe Übereinstimmung zwischen dem Schätzergebnis und dem konstanten Anteil im Energiedichtespektrum der Aufnahme beobachtet werden. In diesem Beispiel ist das Nutzsinal nur bis zu Ortsfrequenzen von ca.  $0,25 \text{ px}^{-1}$  hin im Energiedichtespektrum erkennbar, bei höheren Ortsfrequenzen werden die charakteristischen Merkmale des Nutzsinals zunehmend vom Störsignal überdeckt.



**Abbildung 4.11** Mit dem Verfahren nach Chen et al. [17] ermittelte Rauschenergiedichteschätzung.

### 4.3.2 Identifikation der Skalierung

Die Identifikation der Skalierung der PSF im Bild beruht auf dem Vergleich einer aus dem beobachteten Bild berechneten Vergleichsgröße  $V_{\text{obs}}(\mathbf{f})$  mit einer für eine spezifische Skalierung  $s$  erwarteten Vergleichsgröße  $V_{s,\text{exp}}(\mathbf{f})$ . Zum Vergleich muss ein geeignetes Distanzmaß  $\langle \bullet, \bullet \rangle$  definiert werden. Das Signalmodell (4.16) bietet dabei die Grundlage zur Berechnung der Vergleichsgrößen  $V_{\text{obs}}(\mathbf{f})$  und  $V_{s,\text{exp}}(\mathbf{f})$ , deren Berechnung noch zu definieren ist. Die Skalierung  $\hat{s}$  aus der Menge der untersuchten Skalierungen  $\mathcal{S}$ , die bei diesem Vergleich am besten abschneidet, wird als plausibelste Skalierung als Schätzwert ausgewählt:

$$\hat{s} = \arg \min_{s \in \mathcal{S}} \langle V_{\text{obs}}(\mathbf{f}), V_{s,\text{exp}}(\mathbf{f}) \rangle. \quad (4.20)$$

Im Folgenden werden Ansätze vorgestellt, um die Vergleichsgrößen  $V_{\text{obs}}(\mathbf{f})$  und  $V_{s,\text{exp}}(\mathbf{f})$  zu berechnen. Anschließend wird auf verschiedene mögliche Distanzmaße  $\langle \bullet, \bullet \rangle$  eingegangen.

#### 4.3.2.1 Berechnung der Vergleichsgrößen

Da die grundlegende Form der verwendeten Apertur und damit der OTF bekannt ist, lässt sich mit dem angenommenen Energiedichtespektrum des idealen Abbilds der Szene  $\tilde{S}_{gg}^E(\mathbf{f})$ , einer angenommenen Skalierung  $s$  und der aus der Beobachtung ermittelten Rauschenergiegedichte  $\sigma_n^2$  ein erwartetes Energiedichtespektrum (engl. *energy spectral density*, ESD) des beobachteten Bildes nach (4.16) berechnen:

$$V_{s,\text{exp}}^{\text{ESD}}(\mathbf{f}) = \tilde{S}_{gg}^E(\mathbf{f}) \cdot |OTF_s(\mathbf{f})|^2 + \sigma_n^2. \quad (4.21)$$

Dies lässt sich dann unmittelbar mit dem Energiedichtespektrum des beobachteten Bildes vergleichen:

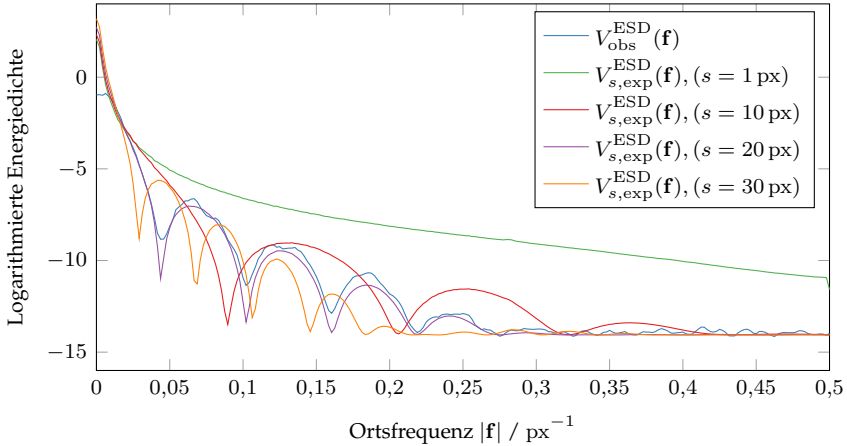
$$V_{\text{obs}}^{\text{ESD}}(\mathbf{f}) = S_{gg}^E(\mathbf{f}). \quad (4.22)$$

Der betrachtete Bildausschnitt muss mithilfe eines Hann-Fensters selektiert werden, da auch bei der Bestimmung des mittleren Energiedichtespektrums  $\hat{S}_{gg}^E(\mathbf{f})$  ein Hann-Fenster verwendet wurde. Das Analysefenster sollte mindestens so groß wie die größte erwartete Skalierung in  $\mathcal{S}$  sein, damit die PSF vollständig in das Beobachtungsfenster passt. Andernfalls würde die PSF nicht vollständig in das Fenster fallen und es könnten somit nur die Auswirkungen des Betragsquadrats des Spektrums einer abgeschnittenen Version der PSF im Energiedichtespektrum des Bildausschnitts beobachtet werden. Darüber hinaus steigt die Auflösung der berechneten Energiedichtespektren, je größer das Analysefenster gewählt wird. Bei einer hohen Auflösung der Energiedichtespektren können feine Merkmale der OTF präzise dargestellt und erkannt werden.

Sofern die Tiefe innerhalb und in der direkten Umgebung dieses Szenenausschnitts homogen ist, enthält der Ausschnitt ausschließlich Signalanteile, die bei der optischen Abbildung mit der OTF derselben Skalierung gefiltert wurden. In diesem Fall sind die Merkmale der OTF im Energiedichtespektrum des beobachteten Bildes ausgeprägt, was eine Identifizierung vereinfacht. Im Fall, dass die Szene in der betrachteten Bildregion mehrere Tiefen aufweist, überlagern sich im Spektrum Signalanteile, die mit unterschiedlichen Skalierungen der OTF gefiltert wurden. Die Merkmale der OTF des zentralen Pixels sind dann u. U. nicht eindeutig im Spektrum identifizierbar. Aus diesem Grund sollte das Analysefenster nicht zu groß gewählt werden. Damit wird einerseits die Wahrscheinlichkeit unterschiedlicher Tiefen im Analysefenster verringert. Andererseits werden im Fall unterschiedlicher Tiefen die PSFs der benachbarten Pixel eher abgeschnitten, weshalb deren Auswirkungen weniger dominant im Energiedichtespektrum des Analysefensters sichtbar sind.

Als Kompromiss dieser komplementären Anforderungen wird in diesem Kapitel als Fenstergröße, zur effizienten Berechnung mit dem FFT-Algorithmus, stets die nächste Zweierpotenz ausgehend vom Doppelten der größten Skalierung in  $\mathcal{S}$  gewählt.

Abbildung 4.12 stellt für ein Pixel der rechten Bildhälfte des Beispielszenarios aus Abbildung 4.5 die beiden Vergleichsgrößen für unterschiedliche Skalierungen  $s$  gegenüber, die, unter Vernachlässigung des Gleichanteils, jeweils auf die Signalenergie eins normiert wurden. Die tatsächliche Entfernung der abgebildeten Ebene entspricht einer Skalierung der PSF von



**Abbildung 4.12** Gegenüberstellung der Vergleichsgrößen  $V_{\text{obs}}^{\text{ESD}}(\mathbf{f})$  und  $V_{s,\text{exp}}^{\text{ESD}}(\mathbf{f})$  beim Vergleich der Energiedichtespektren der Aufnahme für verschiedene Skalierungen  $s$ . Die Skalierung  $s = 20$  px entspricht der tatsächlichen Objektentfernung der Szene.

$s = 20$  px. Vor allem anhand der Lage der Minima der berechneten Vergleichsgrößen lassen sich bei dieser Konfiguration die Effekte der Variation der Skalierung und die Übereinstimmung der Energiedichtespektren beurteilen. Bei der für eine Skalierung von  $s = 20$  px berechneten Vergleichsgröße  $V_{s,\text{exp}}^{\text{ESD}}$  stimmen die Positionen der Minima mit denen des beobachteten Bildes am besten überein.

Weiterhin fällt auf, dass der Bereich der Ortsfrequenzen mit schlechtem SNR (im Beispiel ab ca.  $0,3 \text{ px}^{-1}$ ), also in dem Bereich, in dem das Nutzsignal von der Störgröße dominiert wird, trotzdem zur Beurteilung der Übereinstimmung der beiden Vergleichsgrößen herangezogen werden kann. Signalanteile der Modellgröße, die nicht in der Beobachtung vorkommen, weisen auf eine falsch gewählte Skalierung  $s$  hin (z. B. Abbildung 4.12, rote Kurve für  $s = 10$  px). Dagegen wirken sich Signalanteile der Modellgröße, die unterhalb der Störgröße liegen (z. B. Abbildung 4.12, violette Kurve für  $s = 20$  px und orangefarbene Kurve für  $s = 30$  px), bei denen also erwartet wird, dass sie im aufgenommenen Signal nicht beobachtbar sind, nicht auf den Vergleich aus, da in jedem Fall die dominante Störgröße aus Modell und Beobachtung miteinander verglichen wird. Der

Vergleich ist in diesem Fall in diesem Frequenzband also lediglich von der vorab ermittelten Rauschenergiedichte abhängig.

Alternativ zum Vergleich der Energiedichtespektren der Aufnahmen kann auf Grundlage des Signalmodells auch die OTF des abbildenden Systems geschätzt und mit der angenommenen OTF einer bestimmten Skalierung  $s$  verglichen werden. So lässt sich (4.12) unmittelbar nach der gesuchten OTF auflösen und somit, mit dem mittleren Energiedichtespektrum  $\bar{S}_{gg}^E(\mathbf{f})$  als Annahme für das Energiedichtespektrum des idealen Bildes und der aus der Aufnahme geschätzten Rauschenergiedichte  $\sigma_n^2$ , aus dem aufgenommenen Bild bestimmen. Die rekonstruierte OTF kann dann als die auf dem beobachteten Bild basierende Vergleichsgröße  $V_{\text{obs}}^{\text{OTF}}(\mathbf{f})$  dienen:

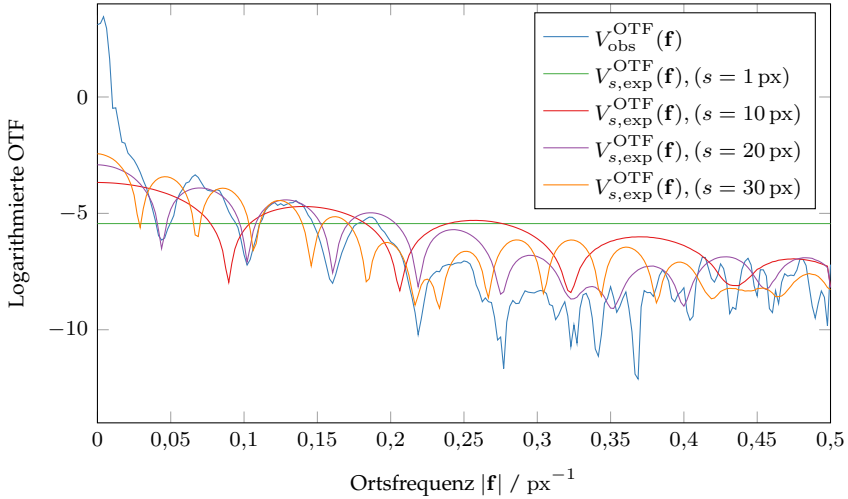
$$V_{\text{obs}}^{\text{OTF}}(\mathbf{f}) = \widehat{OTF}(\mathbf{f}) = \frac{S_{gg}^E(\mathbf{f}) - \sigma_n^2}{\bar{S}_{gg}^E(\mathbf{f})}. \quad (4.23)$$

Aus der gewählten Form der Apertur lässt sich bis auf die unbekannte Skalierung die OTF des Systems berechnen. Als modellbasierte Vergleichsgröße kann daher die eingestellte OTF des Systems für verschiedene Skalierungen  $s$  herangezogen werden:

$$V_{s,\text{exp}}^{\text{OTF}}(\mathbf{f}) = OTF_s(\mathbf{f}). \quad (4.24)$$

Da das angenommene mittlere Energiedichtespektrum  $\bar{S}_{gg}^E(\mathbf{f})$ , wie oben diskutiert, keine Nullstellen aufweist, kann bereits mit diesem naiven Ansatz die OTF gut rekonstruiert werden, wie in Abbildung 4.13 für verschiedene Skalierungen  $s$  zu sehen ist. Wie zuvor weist die Kurve für die tatsächlich vorliegende Skalierung von  $s = 20$  px die größte Übereinstimmung mit der Beobachtung auf. Es fällt jedoch auf, dass ab Ortsfrequenzen von ca.  $0,25 \text{ px}^{-1}$  keine zuverlässige Schätzung der OTF mehr möglich ist, was auf das schlechte SNR in diesem Bereich zurückzuführen ist. Im Gegensatz zum Vergleich des beobachteten mit dem erwarteten Energiedichtespektrum muss dieser Bereich für einen quantitativen Vergleich ausgeschlossen oder gesondert betrachtet werden.

Die vorgestellten Wege zur Berechnung der Vergleichsgrößen  $V_{\text{obs}}$  und  $V_{s,\text{exp}}$  beruhen beide auf demselben Signalmodell aus (4.12). In Bereichen der Ortsfrequenz mit niedrigem SNR erweist sich jedoch der Vergleich der Energiedichtespektren als überlegen gegenüber dem Vergleich der



**Abbildung 4.13** Gegenüberstellung der Vergleichsgrößen  $V_{\text{obs}}^{\text{OTF}}(\mathbf{f})$  und  $V_{s,\text{exp}}^{\text{OTF}}(\mathbf{f})$  beim Vergleich der OTFs für verschiedene Skalierungen  $s$ . Die Skalierung  $s = 20 \text{ px}$  entspricht der tatsächlichen Objektentfernung der Szene.

OTFs, weshalb im Folgenden lediglich der Ansatz über die Energiedichtespektren weiter betrachtet wird. Zur Verbesserung des Ansatzes des Vergleichs der OTFs könnten fortgeschrittenere Verfahren zur Systemidentifikation untersucht werden, die beispielweise das angenommene SNR besser berücksichtigen, um aus dem beobachteten Bild die OTF zu rekonstruieren.

### 4.3.2.2 Berechnung des Distanzmaßes

Zur Bestimmung der Skalierung  $\hat{s}$ , für welche die aus der Beobachtung abgeleitete Vergleichsgröße  $V_{\text{obs}}$  am besten mit der erwarteten Vergleichsgröße  $V_{s,\text{exp}}$  übereinstimmt, wird die Distanz zwischen den beiden Größen über ein Distanzmaß gemäß (4.20) bestimmt:

$$\hat{s} = \arg \min_{s \in \mathcal{S}} \langle V_{\text{obs}}(\mathbf{f}), V_{s,\text{exp}}(\mathbf{f}) \rangle. \tag{4.25}$$

Bei den betrachteten Spektren ist der Großteil der Energie bei kleinen Ortsfrequenzen konzentriert. Um hochfrequente Signalanteile überhaupt



sichtbar darstellen zu können, wurde beispielsweise Abbildung 4.12 in logarithmischer Darstellung angegeben. Damit bei Distanzmaßen, die auf der Bewertung der Energiedifferenz beruhen, hochfrequente Signalanteile auch gegenüber niederfrequenten Anteilen ins Gewicht fallen, wird im Folgenden oftmals vorab der Logarithmus der Spektren gebildet.

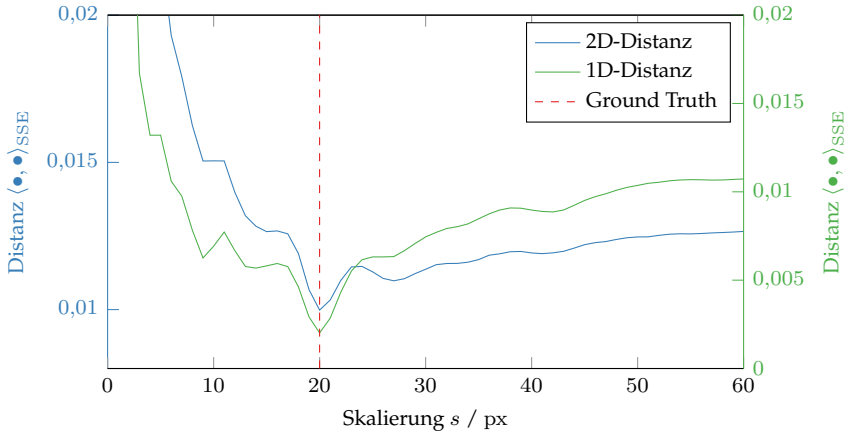
Ein einfaches Maß zur Bewertung der Distanz zweier Energiedichtespektren ist die Summe der quadrierten Abweichung (engl. *sum of squared errors*, SSE):

$$\langle V_{\text{obs}}(\mathbf{f}), V_{s,\text{exp}}(\mathbf{f}) \rangle_{\text{SSE}} = \sum_{\mathbf{f}} (\log(V_{\text{obs}}(\mathbf{f})) - \log(V_{s,\text{exp}}(\mathbf{f})))^2. \quad (4.26)$$

Das so definierte Distanzmaß lässt sich unmittelbar auf eindimensionale Energiedichtespektren, wie in Abbildung 4.12, aber auch auf zweidimensionale Spektren anwenden. Bei rotationssymmetrischen Aperturformen werden durch die Reduktion auf eine Dimension die Energiedichtespektren zusätzlich geglättet, was die allgemeine Qualität der Kurven verbessern kann. Bei nicht rotationssymmetrischen Formen gehen dabei allerdings für den Vergleich potentiell wertvolle spektrale Merkmale verloren, weshalb dieses Distanzmaß im Allgemeinen auf zweidimensionale Energiedichtespektren angewendet werden sollte.

In Abbildung 4.14 ist für das Beispiel aus Abbildung 4.5 und die auf den Energiedichtespektren basierenden Vergleichsgrößen  $V_{\text{obs}}^{\text{ESD}}(\mathbf{f})$  und  $V_{s,\text{exp}}^{\text{ESD}}(\mathbf{f})$  die auf die Fenstergröße normierte Distanz nach (4.26) in Abhängigkeit der betrachteten Skalierung  $s$  dargestellt. Sowohl bei Betrachtung der zweidimensionalen Spektren als auch bei der vorigen tangentialen Projektion der Spektren in eine Dimension ergibt sich im gewählten Szenario für eine Skalierung von  $s = 20$  px, die auch der tatsächlichen Objektentfernung entspricht, eine minimale Distanz der beiden Vergleichsgrößen.

Bei Skalierungen von  $s = 9$  px,  $s = 15$  px und  $s = 27$  px sind deutliche Nebenminima zu beobachten. Diese sind darauf zurückzuführen, dass die OTF der rotationssymmetrischen Blende, wie in Abbildung 4.9(a) dargestellt, konzentrische Kreise mit starker Dämpfung besitzt und dass einige dieser Minima für diese Skalierungen ebenfalls grob mit den Minima des tatsächlich beobachteten Energiedichtespektrums zusammenfallen. Dieser Effekt kann durch die Wahl einer Aperturform, deren OTF weniger regelmäßige Dämpfungsmuster aufweist, vermieden werden.



**Abbildung 4.14** Auf Fenstergröße normierte Distanz nach (4.26) eines Pixels der rechten Bildhälfte des Beispielszenarios.

Mit dem Distanzmaß  $\langle \bullet, \bullet \rangle_{SSE}$  lässt sich besonders dann die Skalierung zuverlässig identifizieren, wenn die spektrale Charakteristik der Szene gut mit dem angenommenen mittleren Energiedichtespektrum  $\tilde{S}_{gg}(\mathbf{f})$  übereinstimmt. Doch auch bei geringen Abweichungen von dieser idealen Anregung, wenn beispielsweise einzelne Ortsfrequenzbereiche nicht angeregt werden, kann mit guten Ergebnissen gerechnet werden. In diesen Situationen kann grundsätzlich nicht unterschieden werden, ob die beobachteten niedrigen Signalanteile dieser Bereiche von einer unzureichenden Anregung oder von der Abbildung herrühren. Da das Distanzmaß aber Ganzheitlich über die gesamte Ortsfrequenzebene gebildet wird und es nicht nur auf der Auswertung einzelner Merkmale beruht, fallen derartige Effekte wenig ins Gewicht.

Die markantesten Merkmale, anhand derer sich die Skalierung der PSF abschätzen lässt, sind die ausgeprägten Minima, die entstehen, da aufgrund der binären Aperturform und der unscharfen Abbildung dort keine Signale durch das System abgebildet werden können. Daher lassen sich Distanzmaße definieren, die die Lage der Minima im Ortsfrequenzbereich bewerten. Die Minima sind vor allem bei niedrigen Ortsfrequenzen ausgeprägt, da hier die Anregung am stärksten ist.

Ein einfaches Distanzmaß für eindimensionale Vergleichsgrößen lässt sich wie folgt definieren: Sei  $\mathcal{F}$  der Bereich der beobachteten Ortsfrequenzen

$$\mathcal{F} = \left[ 0 \text{ px}^{-1}; 0,5 \text{ px}^{-1} \right] \quad (4.27)$$

und  $\mathcal{F}^\circ \subset \mathcal{F}$  eine zusammenhängende Teilmenge von  $\mathcal{F}$ . Weiter bezeichne  $\mathcal{F}_{\text{obs}}^{\min}$  mit

$$\mathcal{F}_{\text{obs}}^{\min} = \left\{ f_{\text{obs}}^{\min} \in \mathcal{F} : f_{\text{obs}}^{\min} = \arg \min_{f \in \mathcal{F}^\circ} (V_{\text{obs}}(|\mathbf{f}|)) \right\} \quad (4.28)$$

die Menge der lokalen Minima der auf eine Dimension projizierten Vergleichsgröße  $V_{\text{obs}}(|\mathbf{f}|)$  und  $\mathcal{F}_{s,\text{exp}}^{\min}$  mit

$$\mathcal{F}_{s,\text{exp}}^{\min} = \left\{ f_{s,\text{exp}}^{\min} \in \mathcal{F} : f_{s,\text{exp}}^{\min} = \arg \min_{f \in \mathcal{F}^\circ} (V_{s,\text{exp}}(|\mathbf{f}|)) \right\} \quad (4.29)$$

entsprechend die lokalen Minima der eindimensionalen Vergleichsgröße  $V_{s,\text{exp}}(|\mathbf{f}|)$ . Ein mögliches Distanzmaß lautet dann:

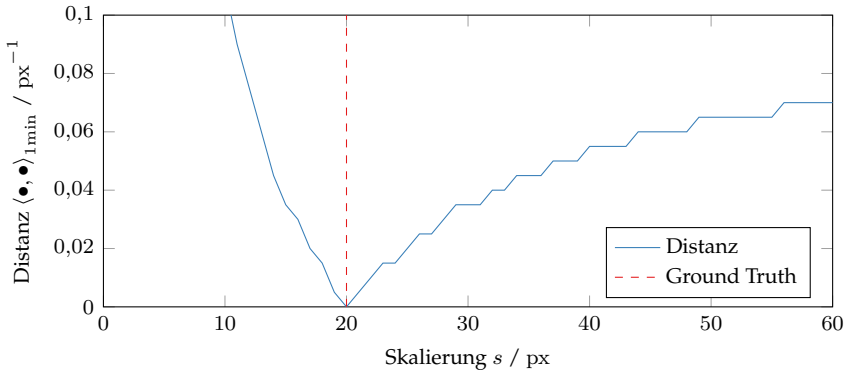
$$\langle V_{\text{obs}}(|\mathbf{f}|), V_{s,\text{exp}}(|\mathbf{f}|) \rangle_{1\text{min}} = \left| f_{0,\text{obs}}^{\min} - f_{0,s,\text{exp}}^{\min} \right|, \quad (4.30)$$

wobei  $f_0^{\min}$ , mit

$$f_0^{\min} \in \mathcal{F}^{\min} : f_0^{\min} \leq f_i^{\min}, \quad \forall f_i^{\min} \in \mathcal{F}^{\min}, \quad (4.31)$$

das jeweils erste lokale Minimum einer Vergleichsgröße bezeichnet. Je nach verwendeter Aperturform kann aus der Ortsfrequenz des ersten Minimums des beobachteten Bildes  $f_{0,\text{obs}}^{\min}$  die Skalierung  $\hat{s}$  auch direkt analytisch berechnet werden. Voraussetzung ist eine analytische Beschreibung der Lage der ersten Nullstelle in Abhängigkeit der Skalierung  $s$ . In diesem Fall braucht die Vergleichsgröße  $V_{s,\text{exp}}$  also gar nicht herangezogen werden.

In Abbildung 4.15 ist der Verlauf der Distanz nach (4.30) über der Skalierung für das betrachtete Beispielszenario dargestellt. Auch hier kann die tatsächlich vorliegende Skalierung korrekt durch das Distanzmaß bestimmt werden.



**Abbildung 4.15** Distanz nach (4.30) eines Pixels der rechten Bildhälfte des Beispielszenarios.

Aufgrund der ausschließlichen Auswertung der Lage der Minima der Energiedichtespektren ist das  $\langle \bullet, \bullet \rangle_{1\min}$ -Distanzmaß robust gegenüber leichten Formvariationen des Anregungsspektrums. Durch die Projektion auf eine Dimension werden die betrachteten Spektren außerdem geglättet, so dass sich bei rotationssymmetrischen Szenen und Aperturformen die abbildungsbedingten Minima deutlich von durch Rauschen verursachten Variationen abheben. Gleichzeitig ist das Distanzmaß allerdings auch anfällig gegenüber Szenen mit unzureichender Anregung bei niedrigen Ortsfrequenzen. Eine schwache Anregung kann fälschlicherweise als durch die unscharfe Abbildung hervorgerufenen Minimum im beobachteten Energiedichtespektrum interpretiert werden und somit stets zu einer falsch geschätzten Skalierung führen.

Um der möglichen Verwechslung eines aufgrund einer unzureichenden Anregung vorhandenen Minimums mit einem durch die unscharfe Abbildung verursachten Minimum im Energiedichtespektrum vorzubeugen, kann das  $\langle \bullet, \bullet \rangle_{1\min}$  Distanzmaß auf den Vergleich der Lage von  $N$  Minima erweitert werden:

$$\langle V_{\text{obs}}(|\mathbf{f}|), V_{s,\text{exp}}(|\mathbf{f}|) \rangle_{N\text{min}} = \sum_{i=0}^{N-1} \left| f_{a(i),\text{obs}}^{\min} - f_{i,s,\text{exp}}^{\min} \right|. \quad (4.32)$$

Über  $a(i)$  werden die Minima im erwarteten Energiedichtespektrum denen der Beobachtung zugeordnet. Dabei muss beachtet werden, dass die

Positionen der Minima korreliert sind, weshalb eine einfache Nächster-Nachbar-Zuordnung leicht zu Fehlzuordnungen führen kann. Die Korrelation der Positionen der Minima im Spektrum kann beispielsweise mit dem *Joint-Compatibility-Branch-and-Bound*-Verfahren bei der Assoziation berücksichtigt werden [66]. Minima, die nicht zugeordnet werden, z. B. weil im beobachteten Energiedichtespektrum nicht genügend Minima vorhanden sind, sollten mit einem konstanten Malus in die Distanz einfließen, da dies auf eine falsch angenommene Skalierung hindeutet.

Levin et al. [49] nutzen die Kullback-Leibler-Divergenz zur Bewertung der Distanz von Wahrscheinlichkeitsdichten der Gradienten in Bildern. Die Kullback-Leibler-Divergenz für zwei (diskrete) Wahrscheinlichkeitsdichten  $p$  und  $q$  lautet [42]

$$\langle p, q \rangle_{\text{KL}} = \sum_n p_n \log \left( \frac{p_n}{q_n} \right). \quad (4.33)$$

Die Kullback-Leibler-Divergenz ist nicht kommutativ ( $\langle p, q \rangle_{\text{KL}} \neq \langle q, p \rangle_{\text{KL}}$ ), was ihren unmittelbaren Einsatz als Distanzmaß in dieser Arbeit verhindert. Es lässt sich nicht plausibel begründen, in welcher Reihenfolge zwei grundsätzlich gleichwertigen Verteilungen  $p$  und  $q$  in die Definitionsgleichung (4.33) eingesetzt werden müssen. Folglich ist nicht sichergestellt, dass Ergebnisse der Kullback-Leibler-Divergenz stets vergleichbar sind.

Davon inspiriert wurde im Rahmen dieser Arbeit als weiteres Distanzmaß, das die Ähnlichkeit der Vergleichsgrößen anhand der gesamten Spektren bestimmt, neben der Summe der quadrierten Abweichungen auch die Jensen-Shannon-Divergenz untersucht. Die Jensen-Shannon-Divergenz wurde von Lin [52] vorgestellt. Sie basiert auf der Kullback-Leibler-Divergenz, besitzt aber die Eigenschaft der Kommutativität:

$$\langle p, q \rangle_{\text{JS}} = \frac{1}{2} (\langle p, m \rangle_{\text{KL}} + \langle q, m \rangle_{\text{KL}}), \quad (4.34)$$

mit

$$m = \frac{1}{2} (p + q). \quad (4.35)$$

Die obigen Gleichungen sind unabhängig von der Dimension von  $p$  und  $q$ , solange diese dieselbe Dimension besitzen. Damit kann die Jensen-Shannon-Divergenz unmittelbar für den Vergleich von ein- oder zweidimensionalen Verteilungen genutzt werden.

Um die Jensen-Shannon-Divergenz zum Vergleich von Energiedichtespektren anwenden zu können, müssen die Spektren normiert werden. Aufgrund der starken Konzentration der Signalenergie bei niedrigen Ortsfrequenzen werden wieder die logarithmierten Spektren verwendet, so dass hochfrequente Signalanteile stärker berücksichtigt werden können. So ergibt sich das Distanzmaß

$$\langle V_{\text{obs}}(\mathbf{f}), V_{s,\text{exp}}(\mathbf{f}) \rangle_{\text{JS}} = \frac{1}{2} (\langle \tilde{V}_{\text{obs}}(\mathbf{f}), \tilde{M}_s(\mathbf{f}) \rangle_{\text{KL}} + \langle \tilde{M}_s(\mathbf{f}), \tilde{V}_{s,\text{exp}}(\mathbf{f}) \rangle_{\text{KL}}), \quad (4.36)$$

mit

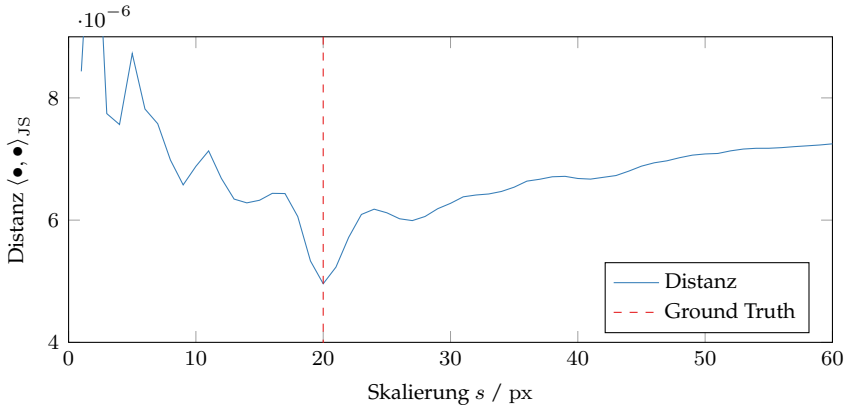
$$\tilde{V}_{\text{obs}}(\mathbf{f}) = \frac{\log(V_{\text{obs}}(\mathbf{f}))}{\sum_{\mathbf{f}} \log(V_{\text{obs}}(\mathbf{f}))} \quad (4.37)$$

$$\tilde{V}_{s,\text{exp}}(\mathbf{f}) = \frac{\log(V_{s,\text{exp}}(\mathbf{f}))}{\sum_{\mathbf{f}} \log(V_{s,\text{exp}}(\mathbf{f}))} \quad (4.38)$$

$$\tilde{M}_s(\mathbf{f}) = \frac{1}{2} (\tilde{V}_{\text{obs}}(\mathbf{f}) + \tilde{V}_{s,\text{exp}}(\mathbf{f})) . \quad (4.39)$$

Abbildung 4.16 zeigt den Verlauf der Jensen-Shannon-Divergenz für das Beispielszenario. Auch mit diesem Distanzmaß kann die vorliegende Skalierung von  $s = 20$  px korrekt erkannt werden. Auffällig sind die im Vergleich zu Abbildung 4.14 stärker ausgeprägten Nebenminima aufgrund der regelmäßigen konzentrischen Kreise im Spektrum der verwendeten Aperturform (vgl. Abbildung 4.9(a)). Aufgrund des abermaligen Logarithmierens der Energiedichtespektren bei der Berechnung der Kullback-Leibler-Divergenz werden die Minima der Spektren stärker gegenüber den Ortsfrequenzbereichen mit hoher Signalamplitude gewichtet.

Die in Abbildung 4.14 und 4.16 beobachtbare Eigenschaft des SSE-Distanzmaßes und der Jensen-Shannon-Divergenz, Nebenminima im Verlauf des Distanzmaßes über der Skalierung zu erzeugen, stellt auf den ersten Blick die Gefahr einer Fehlzuordnung dar. Diese Minima der Distanzverläufe repräsentieren jedoch Skalierungen, die ebenfalls plausibel den Verlauf des beobachteten Energiedichtespektrums erklären. Im folgenden Abschnitt wird anhand der quantitativen Ergebnisse der Distanzmaße bei der Auswertung des Beispielszenarios gezeigt, wie diese Nebenminima helfen können, lokale Fehlzuordnungen der Skalierung zu korrigieren.



**Abbildung 4.16** Distanzmaß nach (4.36) eines Pixels der rechten Bildhälfte des Beispielszenarios.

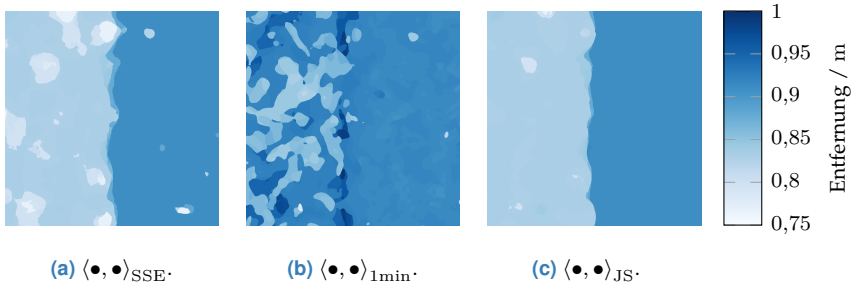
### 4.3.3 Nachbearbeitung

Mit den Methoden der vorigen Abschnitte wird die Skalierung der PSF eines Pixels einer unscharfen Aufnahme anhand der spektralen Eigenschaften in der Pixelumgebung geschätzt. In Abbildung 4.17 sind die Ergebnisse des Beispielszenarios aus Abbildung 4.5, jeweils einzeln für die im vorigen Abschnitt diskutierten Distanzmaße, dargestellt. In Tabelle 4.1 sind die quantitativen Abweichungen von der tatsächlichen Szene anhand des mittleren absoluten Fehlers (engl. *mean absolute error*, MAE), bezüglich der Skalierung in der Einheit Pixel und bezüglich der Entfernung in Metern, angegeben:

$$MAE = \frac{1}{N} \sum_{n=1}^N |\hat{x}_n - x_n|. \quad (4.40)$$

Zur Veranschaulichung der Größenordnung des Schätzfehlers wird außerdem der mittlere absolute prozentuale Fehler (engl. *mean absolute percentage error*, MAPE) bezüglich der Entfernung in Prozent angegeben:

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{x}_n - x_n}{x_n} \right|. \quad (4.41)$$



**Abbildung 4.17** Ermittelte Tiefenkarten des Beispielszenarios für unterschiedliche Distanzmaße ohne Nachbearbeitung.

Sowohl mit dem SSE-Distanzmaß als auch mit der Jensen-Shannon-Divergenz konnte die Skalierung der PSF mehrheitlich korrekt identifiziert werden, wobei letztere am besten abschneidet. Positiv fällt hier auf, dass der Tiefensprung der Szene als solcher erkannt wurde. Es werden ausschließlich Skalierungen nahe den tatsächlichen Skalierungen geschätzt, so dass kein kontinuierlicher Tiefenübergang resultiert. Auch die Position des Tiefensprungs wurde bis auf wenige Pixel gut geschätzt.

Beim Vergleich der Lage des jeweils ersten Minimums der Energiedichtespektren konnte nur in der rechten Bildhälfte ein zufriedenstellendes Ergebnis erzielt werden. In der linken Bildhälfte, in der die tatsächliche Skalierung  $s = 40$  px beträgt, liegen die Minima der Energiedichtespektren bei der gewählten Fenstergröße von 128 px bereits sehr dicht beieinander. Die Frequenzauflösung ist hier nicht mehr ausreichend, um die jeweils ersten Minima zuverlässig zu detektieren.

**Tabelle 4.1** Schätzfehler des Algorithmus zur Tiefenschätzung für das Beispielszenario ohne Nachbearbeitung.

Distanzmaß	Skalierung	Entfernung	
	MAE/px	MAE/cm	MAPE/%
$\langle \bullet, \bullet \rangle_{\text{SSE}}$	1,77	0,62	0,73
$\langle \bullet, \bullet \rangle_{1\text{min}}$	9,65	3,79	4,52
$\langle \bullet, \bullet \rangle_{\text{JS}}$	<b>0,62</b>	<b>0,26</b>	<b>0,31</b>



Für viele Szenarien kann angenommen werden, dass die Szene aus stückweise zusammenhängenden Oberflächen besteht, die sich dadurch auszeichnen, dass sich die Szenentiefe an Objekten nur geringfügig und kontinuierlich ändert. Lediglich an Objektkanten sind große Tiefensprünge zu erwarten. Dies ist eine gängige Annahme globaler Stereokorrespondenzalgorithmen [86, 95].

Auch wenn sich die zur Skalierungsschätzung betrachteten Umgebungen benachbarter Pixel nur geringfügig unterscheiden, läuft die eigentliche Schätzung grundsätzlich unabhängig von den Schätzungen benachbarter Pixel ab. Abhängig von der Szene können sich daher hochfrequente Variationen der ermittelten Skalierung ergeben, die der angenommenen Gestalt der Szene widersprechen. In Abbildung 4.17 sind beispielsweise vereinzelte Regionen mit einer von der Umgebung abweichender geschätzter Entfernung zu beobachten. Mit dem Wissen über die genannten Eigenschaften einer Szene kann eine separate Nachbearbeitung die Qualität der ermittelten Skalierungskarte in diesen Regionen verbessern.

Ein naiver Ansatz zur Unterdrückung hochfrequenter Tiefenvariationen ist eine Tiefpassfilterung der ermittelten Skalierungskarte, beispielsweise mit einem Gauß-Filter. Dadurch wird jedoch die örtliche Auflösung der Tiefenkarte reduziert. Außerdem werden scharfe Objektkanten zu kontinuierlichen Übergängen geglättet, die Tiefenwerte enthalten, auf die die aufgenommenen Daten nicht hinweisen.

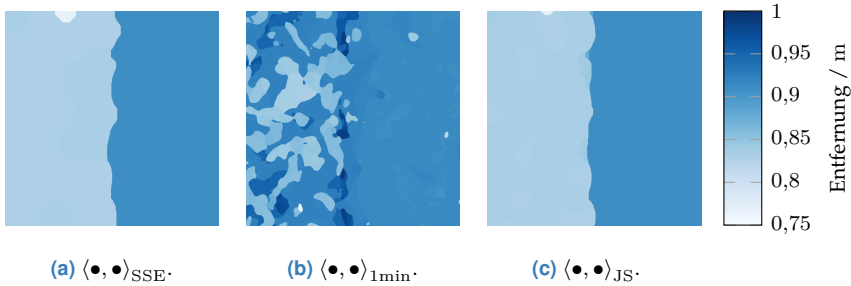
Die Ermittlung einer Skalierungskarte mit sowohl einer geringen globalen Distanz als auch einem stückweise glatten Verlauf wird als Energieminimierung formuliert:

$$\hat{s}(\mathbf{p}) = \arg \min_{s(\mathbf{p})} E_d(s(\mathbf{p})) + E_s(s(\mathbf{p})) , \quad (4.42)$$

mit den Datenkosten  $E_d$  und den Glattheitskosten  $E_s$ . Als Datenkosten kann unmittelbar die Summe der Distanzen der einzelnen Pixel dienen:

$$E_d(s(\mathbf{p})) = \sum_{\mathbf{p}} \langle V_{\text{obs}}, V_{s(\mathbf{p}), \text{exp}} \rangle . \quad (4.43)$$

Die Differenz der Skalierungsschätzungen des Pixel  $\mathbf{p}$  und des Pixels  $\mathbf{q}$  aus der Nachbarschaft  $\mathcal{N}(\mathbf{p})$  von  $\mathbf{p}$  geht bis zu einer oberen Schranke  $E_{s, \text{max}}$  linear in die Glattheitskosten ein. Die Summe dieser Skalierungsdifferenzen



**Abbildung 4.18** Ermittelte Tiefenkarten des Beispielszenarios für unterschiedliche Distanzmaße mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

über die Nachbarschaft  $\mathcal{N}(\mathbf{p})$  ergibt die Kosten eines einzelnen Pixels  $\mathbf{p}$ . Die Glattheitskosten  $E_s(s(\mathbf{p}))$  einer Skalierungskarte folgen aus der Summe der Kosten aller Pixel:

$$E_s(s(\mathbf{p})) = \sum_{\mathbf{p}} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \min(E_{s,\max}, a |s(\mathbf{p}) - s(\mathbf{q})|) . \quad (4.44)$$

Über den Faktor  $a$  und die obere Schranke  $E_{s,\max}$  kann eingestellt werden, ab welcher Differenz Skalierungssprünge erhalten bleiben sollen und wie die Glattheitskosten gegenüber den Datenkosten gewichtet werden.

Das Finden der minimalen Lösung von (4.42) ist ein NP-schweres Problem. Daher wird in dieser Arbeit eine Näherungslösung für (4.42) mit dem *Graph-Cut*-Algorithmus bestimmt [13].

In Abbildung 4.18 und Tabelle 4.2 sind die Ergebnisse mit erfolgter Glättung mit dem *Graph-Cut*-Algorithmus angegeben. Für alle drei Distanzmaße konnte durch die Glättung das Ergebnis verbessert werden. Vor allem die mit dem SSE-Distanzmaß bestimmte Lösung konnte deutlich verbessert werden. Die Jensen-Shannon-Divergenz schneidet auch in Kombination mit der Glättung am besten ab.

Aufgrund der kombinierten Bewertung der Datentreue und der Glattheit der Lösung durch (4.42) kann es in bestimmten Bildregionen durchaus günstiger sein, zur Erlangung einer glatteren Lösung von der Skalierung mit minimaler Distanz abzuweichen. Wenn die Distanz  $\langle \bullet, \bullet \rangle$  eines Pixels mehrere lokale Minima aufweist, deutet dies darauf hin, dass mehrere Skalierungen der PSF die Beobachtung plausibel erklären können. Sofern

**Tabelle 4.2** Schätzfehler des Algorithmus zur Tiefenschätzung für das Beispielszenario mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

Distanzmaß	Skalierung	Entfernung	
	$MAE/px$	$MAE/cm$	$MAPE/\%$
$\langle \bullet, \bullet \rangle_{SSE}$	0,73	0,26	0,30
$\langle \bullet, \bullet \rangle_{1min}$	9,32	3,66	4,37
$\langle \bullet, \bullet \rangle_{JS}$	<b>0,49</b>	<b>0,22</b>	<b>0,26</b>

die Bildregion nicht zu groß ist, kann es daher insgesamt günstiger sein, die Skalierung mit suboptimaler Datentreue aber einer höheren Glattheit zu wählen.

Damit die beschriebene Glättungswirkung ihr Potenzial voll ausschöpfen kann, ist es hilfreich, wenn die berechnete Distanz für ein Pixel in uneindeutigen Szenarien plausible Skalierungen zu lokalen Minima der berechneten Distanz führen. Das SSE-Distanzmaß und die Jensen-Shannon-Divergenz besitzen diese Eigenschaft, wie an den lokalen Minima in Abbildung 4.14 bzw. 4.16 beobachtet werden kann. Dort treten auch für von der tatsächlichen Skalierung abweichende Skalierungen immer dann lokale Minima in der berechneten Distanz auf, wenn Minima im erwarteten Energiedichtespektrum  $V_{s,exp}$  mit Minima in der Beobachtung  $V_{obs}$  zusammenfallen. Da die Minima, wie oben diskutiert, markante Merkmale der Energiedichtespektren sind, ist es vorteilhaft, dass deren Übereinstimmung zu einem lokalen Minimum in der berechneten Distanz führt. So können durch die nachträgliche Glättung lokal begrenzte Fehler korrigiert werden, wie Abbildung 4.18(a) und 4.18(c) verdeutlichen.

Das  $\langle \bullet, \bullet \rangle_{1min}$ -Distanzmaß weist ausgehend vom absoluten Minimum per Definition stets einen ansteigenden Verlauf auf (vgl. Abbildung 4.15). Das Potenzial zur Korrektur von Regionen mit falsch geschätzter Skalierung ist daher bei Verwendung dieses Distanzmaßes sehr gering. Im Beispielszenario wurde durch die Glättung zwar eine geringfügige Verbesserung des Schätzergebnisses erzielt, hauptsächlich jedoch in der rechten Bildhälfte, wo der ursprüngliche Schätzung schon sehr präzise war. Das Schätzergebnis der linken Bildhälfte konnte durch die Glättung nicht signifikant verbessert werden.

**Tabelle 4.3** Schätzfehler des Algorithmus zur Tiefenschätzung für das Beispielszenario mit Nachbearbeitung mit einer Parabelanpassung.

Distanzmaß	Skalierung	Entfernung	
		MAE/px	MAE/cm
$\langle \bullet, \bullet \rangle_{\text{SSE}}$	1,87	0,62	0,73
$\langle \bullet, \bullet \rangle_{1\text{min}}$	9,41	3,69	4,41
$\langle \bullet, \bullet \rangle_{\text{JS}}$	<b>0,70</b>	<b>0,25</b>	<b>0,30</b>

Eine alternative Möglichkeit zur Nachbearbeitung des Schätzergebnisses beruht auf der Verfügbarkeit einer Distanz  $\langle \bullet, \bullet \rangle$  für jede diskrete Skalierung  $s \in \mathcal{S}$ . Unabhängig für jedes Pixel können diese Distanzen genutzt werden, um den Distanzverlauf durch ein geeignetes Polynom zu approximieren, um damit die Skalierung mit minimaler Distanz mit höherer Präzision bestimmen zu können. Zum Beispiel lassen sich auf Grundlage der Skalierung mit minimaler Distanz und den Distanzen der beiden benachbarten Skalierungen in  $\mathcal{S}$  die Parameter einer Parabel bestimmen. Das Minimum dieser allgemeinen Parabel kann dann als Schätzwert  $\hat{s}$  dienen. Auf diese Weise kann die Skalierung wahlweise mit Subpixelgenauigkeit bestimmt werden oder es kann eine kleine Menge an Skalierungen in  $\mathcal{S}$  gewählt werden, um Berechnungszeit einzusparen.

Die Schätzfehler nach einer Bestimmung der geschätzten Skalierung  $\hat{s}$  über eine Parabelanpassung sind in Tabelle 4.3 angegeben. Die Werte unterscheiden sich nur marginal von den ursprünglichen Werten aus Tabelle 4.1, weshalb auf eine visuelle Darstellung der Schätzergebnisse verzichtet wird. Bezüglich des Entfernungfehlers konnte in allen Fällen eine Verbesserung erreicht werden, obwohl sich der Skalierungsfehler mit dem SSE-Distanzmaß und der Jenson-Shannon-Divergenz erhöht hat. Dies ist auf die in Abschnitt 4.2 diskutierte nichtlineare Kennlinie des Messverfahrens und die Tatsache, dass im Beispiel der Nahfall betrachtet wird, zurückzuführen.

Bei Pixeln, für die die geschätzte Skalierung  $\hat{s}$  nach der Glättung mit dem *Graph-Cut*-Algorithmus einem (lokalen) Minimum der berechneten Distanz entspricht, lassen sich die beiden beschriebenen Verfahren kombinieren.

## 4.4 Blendenoptimierung

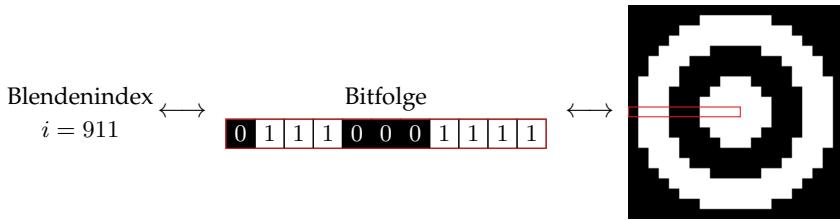
Eine Kamera mit programmierbarer Apertur, wie sie im Rahmen dieser Arbeit aufgebaut wurde, eröffnet die Möglichkeit, die Aperturform so zu wählen, dass sich die Kamera optimal für eine gewünschte Anwendung eignet. Im hier betrachteten Kontext der monokularen Tiefenschätzung nach dem *Depth-from-Defocus*-Prinzip sollte die gewählte Aperturform also eine präzise und robuste Identifizierung der Skalierung der PSF unterstützen. Aufgrund der Implementierung der Apertur mit einem transmissiven LCD müssen keine Bedingungen an die Blendenform, wie beispielsweise die Vermeidung schwebender Teile im Aperturzentrum wie bei Levin et al. [49], gestellt werden.

Im Laufe der Arbeit wurde mithilfe einer vollständigen Suche nach geeigneten Blenden zur Verwendung mit dem  $\langle \bullet, \bullet \rangle_{1\min}$ -Distanzmaß gesucht [120]. Da dieses Distanzmaß speziell für den Vergleich eindimensionaler Energiedichtespektren konzipiert ist, wurde ausschließlich nach rotationssymmetrischen Blenden gesucht, deren Energiedichtespektren sich ohne Informationsverlust mit einer Projektion in Tangentialrichtung in eine eindimensionale Form überführen lassen. Um den Suchraum weiter zu einguzugrenzen, wurde sich auf binäre Formen mit einer diskreten Auflösung von  $21 \times 21$  px beschränkt. Zulässige Blendenformen lassen sich somit durch Bitfolgen der Länge 11 beschreiben, die der Form der Blende entlang einer radialen Linie entsprechen. Wie in Abbildung 4.19 dargestellt, stehen dabei Bits mit Wert 1 für lichtdurchlässige und Bits mit Wert 0 für lichtundurchlässige Bereiche. Die Bitfolgen werden im Folgenden über den Blendenindex  $i$  bezeichnet, der dem Wert der Bitfolge in dezimaler Darstellung entspricht.

Als Gütemaß  $J(i)$  wurde für jede tatsächlich vorliegende Skalierung  $s_t \in \mathcal{S}$  die minimale Distanz zu den Vergleichsgrößen der davon verschiedenen Skalierungen  $s \neq s_t, s \in \mathcal{S}$  ermittelt und darüber der Mittelwert gebildet:

$$J(i) = \frac{1}{|\mathcal{S}|} \sum_{s_t \in \mathcal{S}} \min_{s \neq s_t} \langle V_{\text{obs}}, V_{s,\text{exp}} \rangle_{1\min}. \quad (4.45)$$

Somit sagt das Gütemaß aus, wie gut sich die tatsächlich vorliegende Skalierung von anderen Skalierungen unterscheiden lässt.



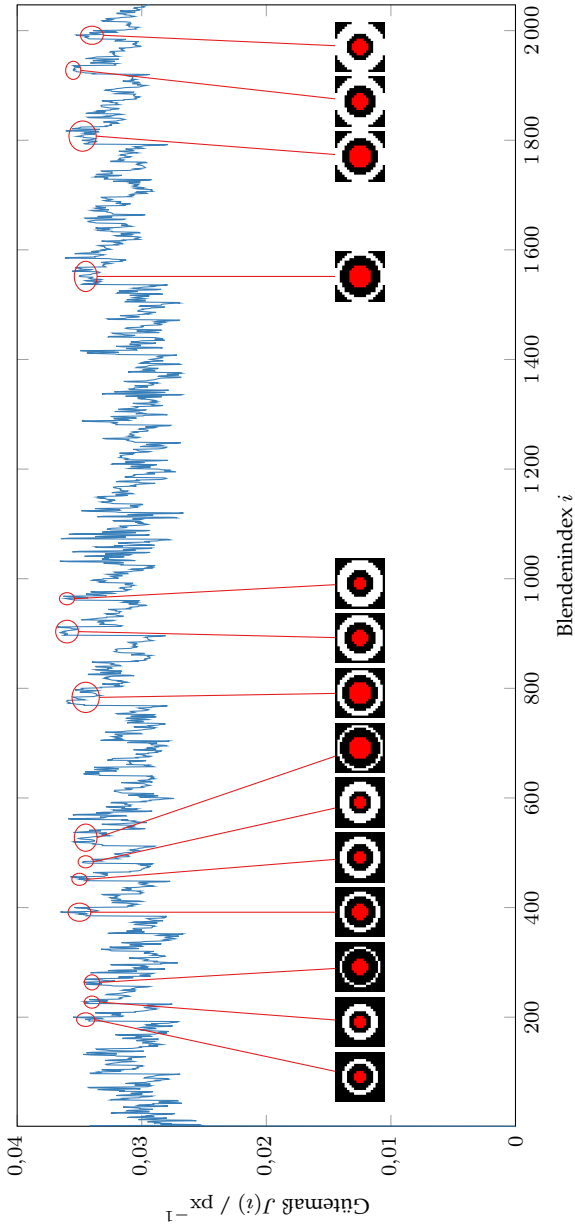
**Abbildung 4.19** Zusammenhang von Blendenindex  $i$  und Aperturform am Beispiel  $i = 911$ .

Die Vergleichsgröße  $V_{\text{obs}}$  der Beobachtung wurde aus einer per Ray-tracing mit der betrachteten Aperturform simulierten Aufnahme einer nach (4.19) texturierten Ebene berechnet. Damit kann die Leistungsfähigkeit einer Blendenform anhand der gesamten Signalverarbeitungskette, also von der Beeinflussung der einfallenden Strahldichte durch die codierte Apertur bis zur Identifikation der Skalierung mit dem vorgestellten Algorithmus, bewertet werden.

In Abbildung 4.20 ist für jede Blende des Suchraums das berechnete Gütemaß  $J(i)$  angegeben. Auffällig am Verlauf sind die markierten und wiederholt auftretenden Bereiche hoher Güte. Diese treten immer dann auf, wenn ein hochwertiges Bit der Bitfolge von 0 auf 1 gesetzt wird, wenn also ein lichtdurchlässiger Ring am äußeren Rand der Apertur hinzukommt. Im Kurvenverlauf ergeben sich an diesen Stellen abrupte Anstiege des Gütemaßes. Direkt nach diesen Anstiegen verbleibt das Gütemaß auf hohem Niveau. Dies deutet darauf hin, dass der innere Bereich der Apertur nur einen geringen Einfluss auf die Leistungsfähigkeit zur Tiefenschätzung besitzt, da in diesen Bereichen des Blendenindex die niederwertigen Bits der Bitfolge variiert werden. In Abbildung 4.20 sind die entsprechenden Bereiche der Aperturen in den Miniaturen rot gekennzeichnet.

Das Maximum des Gütemaßes liegt bei  $i = 911$  ( $J(911) = 0,0365 \text{ px}^{-1}$ ). Die entsprechende Form ist in Abbildung 4.19 angegeben. Die Leistungsfähigkeit der Kreisblende ( $J(2047) = 0,0294 \text{ px}^{-1}$ ) liegt sogar unter dem Durchschnitt aller hier betrachteter Blenden ( $\bar{J} = 0,0313 \text{ px}^{-1}$ ).

Da das  $\langle \bullet, \bullet \rangle_{1\text{min}}$ -Distanzmaß lediglich zum Vergleich eindimensionaler Energiedichtespektren genutzt werden kann, ist es nur für rotationssymmetrische Aperturformen und Szenen geeignet. Vor allem aufgrund der



**Abbildung 4.20** Ergebnis der vollständigen Suche nach optimalen rotationssymmetrischen Blenden für das  $(\bullet, \bullet)_{1\text{min}}$ -Distanzmaß. In den Miniaturen der Blendenformen stehen weiße Pixel für lichtdurchlässige und schwarze Pixel für lichtundurchlässige Bereiche der Aperturblende. Die roten Pixel geben diejenigen Bereiche der Blendenform an, die im jeweiligen Intervall des Blendenindex  $i$  variiert werden.

letzten Bedingung ist keine breite Anwendbarkeit dieses Distanzmaßes möglich. Bei der Suche nach allgemeinen Blendenformen ist der Suchraum, selbst bei geringen Pixelauflösungen, so groß, dass er nicht mehr vollständig durchsucht werden kann. Daher müssen stochastische Verfahren zum Einsatz kommen, um eine Approximationslösung zu bestimmen.

Es wurde neben der vollständigen Suche nach rotationssymmetrischen Blenden auch eine Suche nach allgemeinen Blendenformen durchgeführt. Als Randbedingungen wird weiterhin verlangt, dass die Blende binärwertig ist. Sie soll durch ein quadratisches Pixelbild dargestellt werden können. Die so definierte Form wird so innerhalb der kreisförmigen Apertur platziert, dass sich jedes Pixel im Lichtweg befindet.

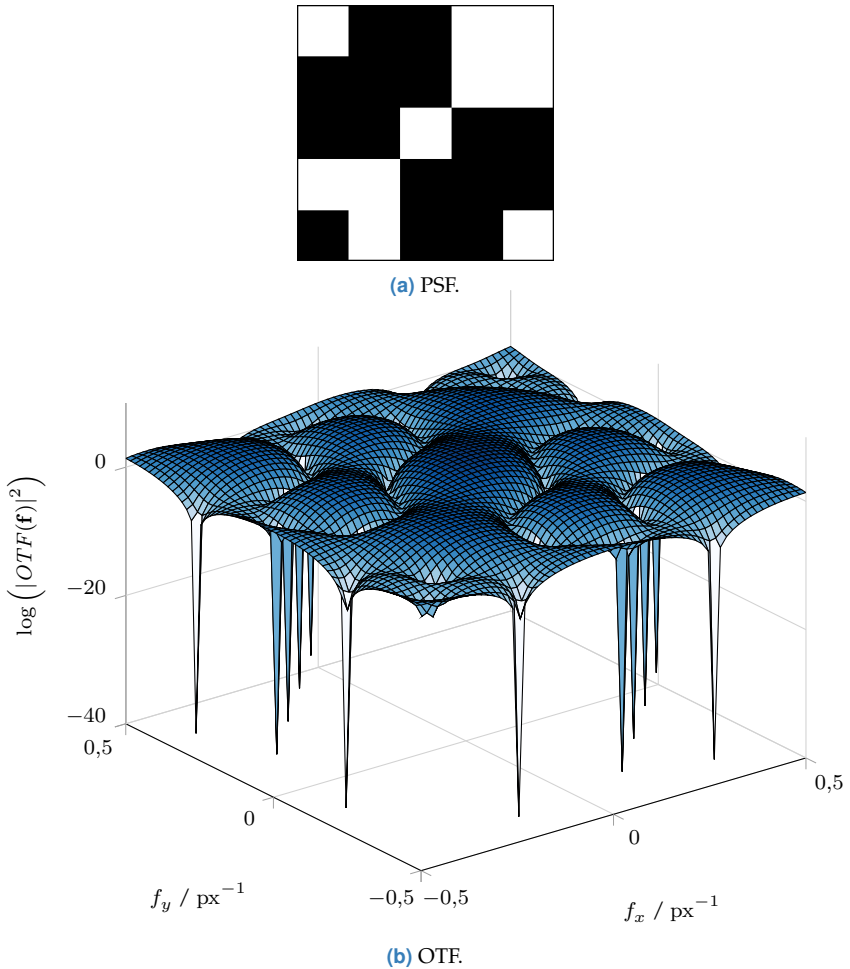
Die Anzahl möglicher Blendenformen mit einer Auflösung von  $N \times N$  px ist durch  $2^{N^2}$  gegeben. Im Gegensatz zu den ca.  $2 \cdot 10^3$  möglichen rotationssymmetrischen Formen der Dimension  $21 \times 21$  px kann dieser Lösungsraum schon für niedrige  $N$  nicht mehr vollständig durchsucht werden. Für  $N = 21$  ergeben sich mögliche Blendenformen in der Größenordnung  $10^{132}$ . Daher wurde ein genetischer Algorithmus eingesetzt.

Zunächst wurde eine Optimierung für Pixelauflösungen von  $5 \times 5$  px durchgeführt. Um sicherzustellen, dass ausreichend Licht durch die Apertur fallen kann, wird außerdem die Randbedingung aufgestellt, dass maximal 75 % der quadratischen Aperturfläche lichtundurchlässig sein dürfen.

Als Gütekriterium wurde, aufgrund des besseren Abschneidens im oben diskutierten Beispielszenario, die Unterscheidbarkeit unterschiedlicher Skalierungen mit der Jensen-Shannon-Divergenz verwendet. Dazu wurde wieder für jede tatsächlich vorliegende Skalierung die minimale Distanz zu allen anderen Skalierungen mit anschließender Mittelwertbildung berechnet. Zur Reduktion der Berechnungsdauer pro Individuum  $i$  wurde die auf der Beobachtung basierende Vergleichsgröße nicht per Raytracing simuliert, sondern auch aus dem Signalmodell berechnet. Bei der Berechnung aus dem Signalmodell wurde zunächst keine zusätzliche Störung angenommen ( $\sigma_n = 0$ ). Insgesamt ergibt sich so als Gütemaß eines Individuums  $i$ :

$$J(i) = \frac{1}{|\mathcal{S}|} \sum_{s_t \in \mathcal{S}} \min_{s \neq s_t} (V_{s_t, \text{exp}}, V_{s, \text{exp}})_{\text{JS}}. \quad (4.46)$$





**Abbildung 4.21** Ergebnis des genetischen Algorithmus zur Suche nach optimalen Blendenformen zur Tiefenextraktion der Größe  $5 \times 5$  px ohne Modellierung einer Störgröße.

In Abbildung 4.21(a) ist das Ergebnis der Optimierung dargestellt. Auffällig an der ermittelten Form ist, dass diese annähernd punktsymmetrisch bezüglich des zentralen Pixels ist. In der berechneten OTF fallen vor allem die einzelnen Punkte mit sehr hoher Dämpfung auf, die in der

dreidimensionalen Darstellung des Betragsquadrats in Abbildung 4.21(b) zu sehen sind. Tatsächlich sind diese Punkte derart ausgeprägt, dass sie maßgeblich zum guten Abschneiden der Blendenform beitragen.

An diesem Beispiel zeigt sich abermals, wie wichtig die Berücksichtigung einer Störgröße bei der Berechnung des Signalmodells ist. Aufgrund von Störeinflüssen wie Bildrauschen oder geringfügigen Abweichungen von der Blendenform, beispielsweise durch Interpolationsartefakte bei der Darstellung der Form auf einem LCD, ist es nicht zu erwarten, dass diese markanten Punkte derart ausgeprägt im Energiedichtespektrum des aufgenommenen Bildes zu beobachten sein werden. Vielmehr werden diese Minima von den Störgrößen des Systems überlagert werden. Daher ist die dargestellte Blendenform für den praktischen Einsatz zu hoch bewertet.

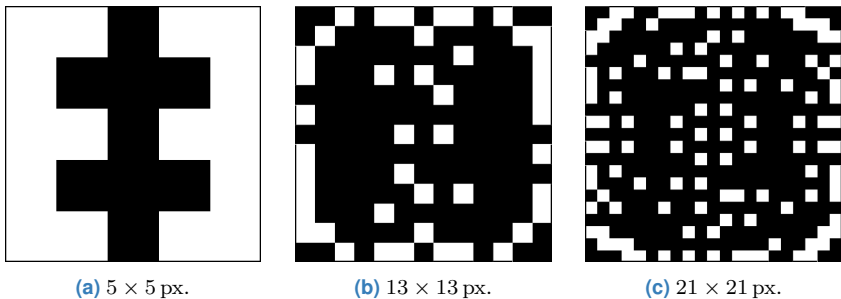
Mit einer erneuten Optimierung mit einem genetischen Algorithmus, bei der diesmal bei der Berechnung des Gütemaßes (4.46) eine additive Störgröße einbezogen wurde, wurde die in Abbildung 4.22(a) dargestellte Form ermittelt. Weiterhin wird vorausgesetzt, dass mindestens 25 % der Fläche lichtdurchlässig ist. Die ermittelte Form ist in diesem Fall tatsächlich punktsymmetrisch, obwohl dies bei der Optimierung nicht vorausgesetzt wurde. Dies deckt sich mit den Beobachtungen von Levin et al. [49], dass symmetrische Formen ausgeprägte Strukturen minimaler Frequenzübertragung aufweisen, die zu einer guten Unterscheidbarkeit verschiedener Skalierungen der PSF mit der Kullback-Leibler-Divergenz führen. Daher ist es zu erwarten, dass mit der hier verwendeten Jensen-Shannon-Divergenz als kommutative Erweiterung der Kullback-Leibler-Divergenz ähnliche Beobachtungen gemacht werden können.

Mit diesem ersten Ergebnis wurde nach weiteren Blendenformen der Größe  $13 \times 13$  px und  $21 \times 21$  px gesucht. Zur Reduktion des Suchaufwands wurde in diesen beiden Fällen die Suche auf punktsymmetrische Formen beschränkt, womit nur noch  $2^{\lceil N^2/2 \rceil}$  mögliche Formen existieren, für  $N = 21$  also nur noch eine Größenordnung von  $10^{66}$ . Die Ergebnisse dieser Optimierungen sind in Abbildung 4.22(b) und 4.22(c) dargestellt.

In allen drei Blendenformen sind vor allem die Bereiche am Blendenrand lichtdurchlässig. Diese Beobachtung deckt sich mit dem Ergebnis der Suche nach rotationssymmetrischen Blenden aus Abbildung 4.20. Obwohl bei der Suche nach Blendenformen keine Bedingung bezüglich schwebender

Teile im Aperturzentrum wie bei Levin et al. [49] einbezogen wurden, weisen die ermittelten Formen keine derartigen Strukturen auf. Die Bereiche der Ecken der Blenden in 4.22(b) und (c) schließen am Bereich außerhalb der optimierten Fläche an, die lichtundurchlässig sein muss. Die Einschränkung von Levin et al. scheint sich also nicht negativ auf die Eignung einer Blendenform zur Tiefenextraktion auszuwirken.

Beim enormen Umfang des zu durchsuchenden Lösungsraums ist es sehr wahrscheinlich, dass es sich bei den ermittelten Formen um lokale Maxima und nicht um das globale Maximum des definierten Gütemaßes handelt. Aber auch dieses Kenntnis ist in diesem Kontext äußerst wertvoll, da davon ausgegangen werden kann, dass die ermittelten Formen im Vergleich zu der enormen Menge an möglichen Formen überdurchschnittlich gut zur Tiefenextraktion geeignet sind. Die tatsächliche Leistungsfähigkeit der ermittelten Formen wird im folgenden Kapitel anhand simulierter und realer Testszenarien untersucht.



**Abbildung 4.22** Ergebnis des genetischen Algorithmus zur Suche nach optimalen Blendenformen zur Tiefenextraktion mit Modellierung einer Störgröße.



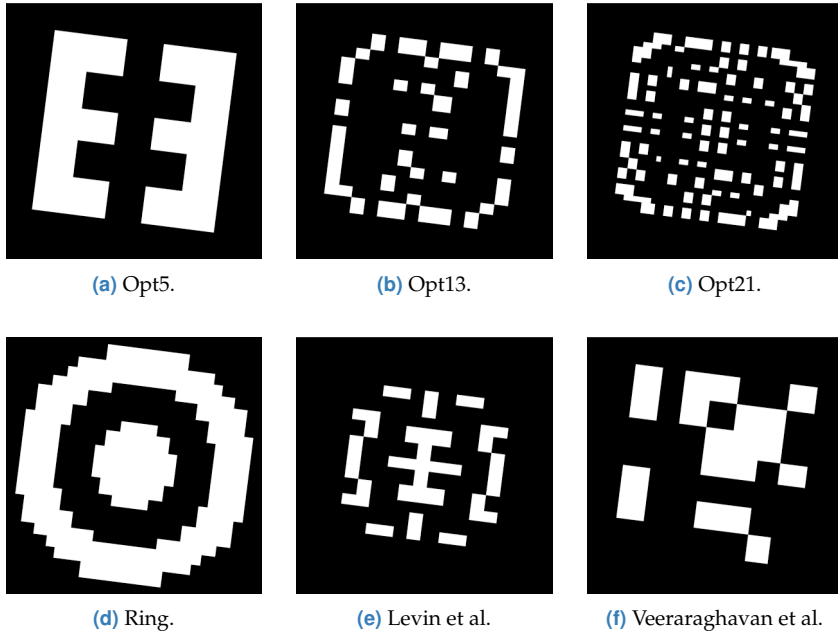
## 5 Ergebnisse

In diesem Kapitel sollen die simulative Entwurfsmethode einer *Computational Camera* mit dem Raytracing-Verfahren aus Kapitel 2 und der entwickelte Prototyp aus Kapitel 3 anhand der Ergebnisse der monokularen Tiefenschätzung nach dem *Depth-from-Defocus*-Prinzip validiert werden. Der Nachweis der grundlegenden Funktion des entwickelten Algorithmus zur Tiefenschätzung in Verbindung mit einer codierten Apertur wurde bereits in Kapitel 4 anhand eines Beispielszenarios erbracht. In diesem Kapitel sollen nun die Ergebnisse realer Aufnahmen mit denen simulierter Aufnahmen verglichen und die Leistungsfähigkeit des Algorithmus anhand anspruchsvollerer Szenen untersucht werden.

Zunächst werden in Abschnitt 5.1 die untersuchten Testszene vorgestellt, die sowohl simulativ als auch anhand eines realen Messaufbaus untersucht werden. Außerdem werden in diesem Abschnitt die Bedingungen, unter denen die Messungen aufgenommen bzw. simuliert wurden, detailliert beschrieben. In den folgenden Abschnitten 5.2 und 5.3 werden die Tiefenschätzergebnisse aus realen bzw. simulierten Aufnahmen vorgestellt und bewertet. Abschließend werden in Abschnitt 5.4 die Erkenntnisse der Untersuchungen zusammengefasst und ein Fazit gezogen.

### 5.1 Versuchsdurchführung

Als einfachstes Testszenario wird eine Szene konstanter Tiefe betrachtet. Um eine optimale Frequenzanregung durch die Szene zu gewährleisten, wird eine Ebene mit einer Rauschtextur nach der *Fractional Brownian Motion* versehen. Die Entfernungen der Szene werden dabei im gesamten Messraum variiert, um die Verfahren für unterschiedliche Skalierungen der PSF zu erproben. Insgesamt werden für diese Szene pro untersuchter Blendenform 41 Bilder mit Skalierungen der PSF von 0 bis 160 px aufgenommen und simuliert.



**Abbildung 5.1** Verwendete Aperturformen, angepasst an die Darstellungsmöglichkeiten des im Objektiv platzierten LCDs.

In Abbildung 5.1 sind die untersuchten Blendenformen und deren in der folgenden Auswertung verwendeten Bezeichnungen dargestellt. In den Abbildungen wurde bereits die Rotation, die begrenzte Auflösung und das Seitenverhältnis der Pixel des verwendeten LCDs innerhalb des Objektivprototyps berücksichtigt. Diese transformierten Formen werden so auch in der Simulation verwendet. Die Formen 5.1(a)–(d) entsprechen den Ergebnissen des genetischen Algorithmus bzw. der vollständigen Suche nach optimalen rotationssymmetrischen Formen, wie sie im vorigen Kapitel erarbeitet wurden. Die Formen (e) und (f) werden in den Arbeiten von Levin et al. [49] bzw. Veeraraghavan et al. [100] vorgeschlagen.

Zur Durchführung der realen Messungen wird der Aufbau aus Abbildung 5.2 verwendet. Auf der horizontal verschiebbaren Halterung wird ein Ausdruck einer Rauschtextur befestigt. Dem geringen Dynamikumfang eines herkömmlichen Ausdrucks auf Papier wird durch eine



Abbildung 5.2 Messaufbau.

geeignete Beleuchtung entgegengewirkt. Die aktive Darstellung der Textur mit einem LCD-Monitor hat sich als ungeeignet erwiesen, da dessen Helligkeit nicht konstant bezüglich der Zeit ist. Dadurch kann sich aufgrund des intensitätsabhängigen Photonenrauschens das Störzenario pro Aufnahme ändern, was die Sicherstellung der Vergleichbarkeit der realen und simulierten Aufnahmen erschwert. Darüber hinaus wird von LCD-Monitoren polarisiertes Licht abgegeben, das, je nach Orientierung, nicht mit dem mit einem LCD versehenen Objektivprototyp wahrgenommen werden kann. Die Kamera ist vor der dargestellten Halterung fixiert und genau senkrecht zu dieser ausgerichtet. In der in der Abbildung dargestellten Position befindet sich die texturierte Ebene 0,85 m von der Kamera entfernt und kann bis unmittelbar vor die Kamera verschoben werden. Bei zu geringem Abstand der Ebene zur Kamera wird die Beleuchtung allerdings durch die Kamera selbst erschwert, da deren Schatten auf die texturierte Ebene fällt. Mithilfe einer Skala an der Führung der Halterung kann die Entfernung der texturierten Ebene zur Kamera millimetergenau eingestellt werden. Diese Entfernung dient in der Auswertung dieser Szene als Ground-Truth-Entfernung.

Aufgrund der nichtidealen Transmissionseigenschaften des verwendeten transmissiven LCDs wird bei den realen Aufnahmen, wie in Abschnitt 3.3.2 beschrieben, eine Kompensation des Störanteils mit einer Aufnahme mit lichtundurchlässigem LCD durchgeführt. Dabei wird eine

lineare Kamerakennlinie vorausgesetzt. Sofern der Sensor nicht in die Sättigung gebracht wird, ist diese Voraussetzung bei der verwendeten Kamera erfüllt. Die Belichtungszeit bei der Aufnahme des Störanteils muss dabei stets der Belichtungszeit der eigentlichen Aufnahme entsprechen.

Um die korrekte Störanteilkompensation sicherzustellen und damit keine Information bei den Aufnahmen abgeschnitten wird, darf der Sensor nicht in die Sättigung gebracht werden. Für eine Messreihe mit unterschiedlichen Entfernungen der texturierten Ebene zur Kamera ist beim Aufbau aus Abbildung 5.2 die Aufnahme mit der geringsten Entfernung am kritischsten, da in diesem Fall der Abstand der Ebene zur Beleuchtung am kleinsten ist. Damit die Aufnahmen ein möglichst hohes SNR aufweisen, sollte der verfügbare Dynamikbereich möglichst gut ausgenutzt werden. Es sollte also möglichst lange belichtet werden. Um diesen beiden komplementären Anforderungen gerecht zu werden, wird eine geeignete Belichtungszeit in der am nächsten zur Kamera gelegenen Einstellung der Ebene bestimmt. Da die untersuchten Blenden unterschiedliche große lichtdurchlässige Flächen aufweisen, muss die Belichtungszeit für jede Blendenform separat gewählt werden. Es werden Belichtungszeiten zwischen 40 und 80 ms verwendet.

In der Raytracing-Simulation wird die Szene konstanter Tiefe durch ein einziges texturiertes Ebenen-Objekt mit einem Lambert'schen Material erzeugt. Als Lichtquelle ist eine allgegenwärtige, homogene Beleuchtung ausreichend. Die Simulationsparameter werden entsprechend der bei den realen Messungen verwendeten Kamera gewählt. Die Szenen werden durchweg mit  $N = 256$  Strahlen pro Pixel gerendert. Bei der vollen Sensorauflösung von  $4864 \times 3232$  px entspricht dies der Berechnung von über 4 Milliarden Strahlen pro Bild.

Damit das Störszenario der simulierten Aufnahmen vergleichbar mit den realen Aufnahmen ist, müssen die Belichtungszeit und die EMVA-Parameter der verwendeten Kamera bekannt sein, woraus sich die Rauschleistung des Dunkelrauschens ergibt. Weiter muss in der Simulation und bei den realen Aufnahmen dieselbe Bitauflösung verwendet werden, damit das Quantisierungsrauschen vergleichbar ist. Zuletzt müssen die Simulationsparameter, beispielsweise durch Anpassung der Beleuchtung oder des Reflexionsgrads des Ebenenmaterials, so gewählt werden, dass sich in der simulierten Aufnahme ähnliche Intensitäten wie bei der realen



Aufnahme ergeben. Damit wird die Vergleichbarkeit des Photonenrauschens sichergestellt.

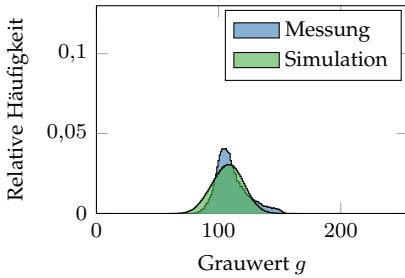
Wie in Abschnitt 3.3.2 beschrieben, wirkt sich die nötige Kompensation des Störanteils bei Verwendung des LCDs auf das Bildrauschen aus. Daher wird sie analog auch bei der Simulation durchgeführt, indem zu jeder Aufnahme eine Aufnahme des Störanteils mit einer homogenen kreisförmigen Blende mit entsprechender Dämpfung simuliert wird.

Der Abgleich der simulierten Grauwertintensitäten mit denen der realen Aufnahmen wird manuell anhand der Histogramme – sowohl für die Aufnahmen mit der untersuchten Apertur als auch für die Aufnahmen der Störanteile – durchgeführt. Da sich beim Aufbau aus Abbildung 5.2 mit dem Verschieben der texturierten Ebene auch der Abstand zur Szenenbeleuchtung ändert, ergeben sich in Abhängigkeit der untersuchten Entfernung auch unterschiedliche Helligkeiten. Daher wird der Abgleich für die Aufnahmen mit der geringsten und der größten Entfernung durchgeführt. Zur Simulation der Aufnahmen, bei denen die Ebene zwischen diesen beiden Entfernungen liegt, werden die Parameter der Beleuchtung und der Reflexionsgrad linear zwischen den beiden bestimmten Parametersätzen interpoliert

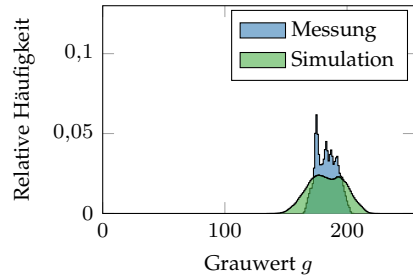
Abbildung 5.3 stellt exemplarisch die Histogramme der simulierten und der realen Aufnahmen für die Blende „Opt5“ gegenüber. Mit der guten Übereinstimmung der Mittelwerte der Verteilungen und dem in Abschnitt 2.4.1.3 erbrachten Nachweis der Vergleichbarkeit der Rauscheigenschaften der realen und simulierten Kamera kann für die untersuchte Szene von näherungsweise gleichen Rauscheigenschaften ausgegangen werden.

Ein erstes Indiz für eine gute Übereinstimmung des realen und des simulierten Bildgewinnungsprozesses zeigt sich bereits an den dargestellten Histogrammen. Auch wenn die Grauwertverteilungen der simulierten Aufnahmen breiter als bei den realen Aufnahmen sind, ergeben sich in beiden Fällen multimodale Grauwertverteilungen bei unscharfer Abbildung (5.3(b) & (d)), obwohl die jeweils scharfen Aufnahmen unimodal verteilt sind (5.3(a) & (c)).

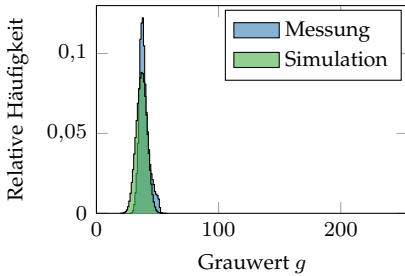
Die Untersuchungen aus Abschnitt 3.3 haben gezeigt, dass das modifizierte Objektiv sehr gute Abbildungseigenschaften aufweist. Daher wird die reale Kamera in der Simulation durch eine Kamera mit dünner Linse



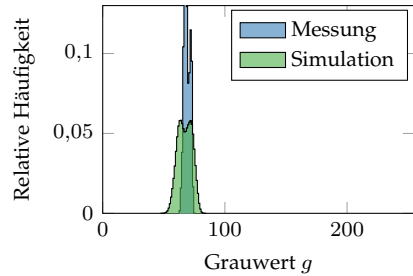
(a) Histogramm von  $g(\mathbf{p})$  bei einer Skalierung von  $s = 0$  px.



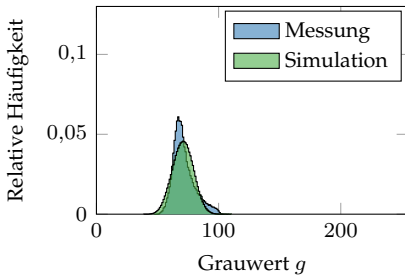
(b) Histogramm von  $g(\mathbf{p})$  bei einer Skalierung von  $s = 160$  px.



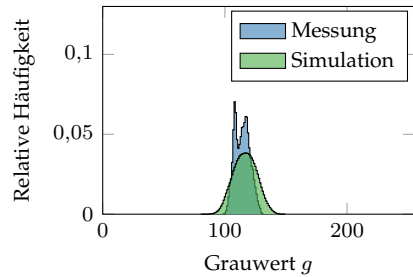
(c) Histogramm von  $g_{\text{Stör}}(\mathbf{p})$  bei einer Skalierung von  $s = 0$  px.



(d) Histogramm von  $g_{\text{Stör}}(\mathbf{p})$  bei einer Skalierung von  $s = 160$  px.



(e) Histogramm von  $g_{\text{Nutz}}(\mathbf{p})$  bei einer Skalierung von  $s = 0$  px.



(f) Histogramm von  $g_{\text{Nutz}}(\mathbf{p})$  bei einer Skalierung von  $s = 160$  px.

**Abbildung 5.3** Vergleich der Grauerthistogramme der realen und simulierten Aufnahmen der Blende „Opt5“ bei Skalierungen von  $s = 0$  px und  $s = 160$  px.

**Tabelle 5.1** Parameter der simulierten Kamera mit dicker sphärischer Linse.

Blendenzahl	$k$	1,8
Brechungsindex	$n_L$	1,44
objektseitiger Krümmungsradius	$R_1$	4,68 cm
bildseitiger Krümmungsradius	$R_2$	-4,68 cm
Linsendicke	$d_L$	5 mm

in paraxialer Näherung modelliert. Aufgrund der detailgetreuen Wiedergabe des eingestellten LCD-Musters in Abbildung 3.9, wird die simulierte Apertur exakt in der Linsenebene platziert. Außerdem wird zusätzlich eine kreisförmige Eintrittspupille simuliert, die, wie in Abbildung 2.22 dargestellt, zu einem Abschneiden der PSF am Bildrand führt. Damit weisen die simulierten PSFs eine hohe Ähnlichkeit zu den tatsächlich aufgenommenen PSFs aus Abbildung 3.6 auf.

Gleichzeitig ist es von Interesse, inwiefern Abbildungsfehler die Tiefenschätzung beeinflussen. Anhand der Szene der texturierten Ebene wird untersucht, welche Tiefenschätzergebnisse bei einer optischen Abbildung mit Abbildungsfehlern erzielt werden können. Der im Rahmen dieser Arbeit aufgebaute Objektivprototyp weist nur geringe Abbildungsfehler auf, wie die Untersuchungen in Kapitel 3 gezeigt haben. Da sich der Grad der Abbildungsfehler des Prototyps nicht beeinflussen lässt, kann lediglich die Raytracing-Simulationsumgebung zur variablen Erzeugung von Abbildungsfehlern eingesetzt werden. Zur Modellierung einer Linse mit Abbildungsfehlern werden die Parameter aus Tabelle 5.1 verwendet. Bis auf den Brechungsindex entsprechen diese Werte den Simulationsparametern aus Abschnitt 2.4.2.4. Der Brechungsindex wurde so gewählt, dass die Brennweite der Linse mit  $f = 52,4$  mm der Brennweite des Objektivprototyps entspricht. Die Ergebnisse dieser Simulationen werden in Abschnitt 5.3 vorgestellt und diskutiert.

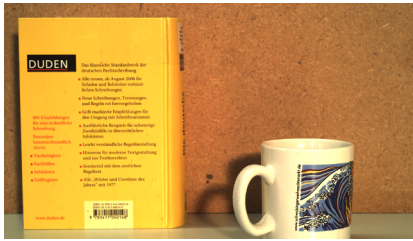
Um die Leistungsfähigkeit der entworfenen *Computational Camera* auch in realistischeren Szenarien einschätzen zu können, werden außerdem die Szenen aus Abbildung 5.4 untersucht. Die Aufnahme in Abbildung 5.4(a) wurde mit einem unmodifizierten Objektiv desselben Typs wie in Kapitel 3 bei einer Blendenöffnung von  $f/22$  aufgenommen. Für die Simulation wurde die Szene möglichst detailgetreu in der Simulationsumgebung

nachgebildet. Die Entfernungen der dargestellten Objekte wurden manuell nachgemessen und in der Simulationsumgebung übernommen. Die Texturen, mit denen die Objekte in der Simulation versehen sind, wurden separat in der realen Szene aufgenommen. Die Kamera ist in der Anordnung auf eine Entfernung von  $d_{o,fok} = 1$  m fokussiert. Die Rückwand befindet sich 80 cm ( $s = 50$  px) von der Kamera entfernt. Die Rückseite des Buches ist im Abstand 73 cm ( $s = 75$  px) platziert und das Zentrum der Tasse befindet sich im Abstand 67 cm ( $s = 100$  px). Als Beleuchtung wurden zwei Punktlichtquellen, deren Positionen denen der Lampen in der realen Szene entsprechen, und eine allgegenwärtige Beleuchtung verwendet. Die Tasse wurde in der Simulation mit dem Material nach Phong [75] versehen, das Glanzlichter erzeugt. Die restlichen Objekte wurden mit Lambert'schem Reflexionseigenschaften simuliert. Die Ground-Truth-Daten aus Abbildung 5.4(c) stehen in der Simulation unmittelbar zur Verfügung. Diese dienen im Folgenden auch als eine Orientierung zur Beurteilung der Ergebnisse der realen Aufnahmen.

Wie auch bei den Aufnahmen der Rauschtextur wird sowohl für die realen Aufnahmen als auch in der Simulation eine Kompensation des Störanteils durchgeführt. Die Anpassung der simulierten Bildintensitäten erfolgt ebenfalls über einen manuellen Abgleich der Histogramme.

In Kapitel 4 wurden verschiedene Möglichkeiten des Vergleichs von beobachteten Merkmalen mit denen eines Signalmodells vorgestellt. Aufgrund des besten Abschneidens in den dort ausgeführten Untersuchungen anhand eines Testszenarios wird im Folgenden die Tiefenextraktion stets auf Grundlage der Energiedichtespektren mit der Jensen-Shannon-Divergenz durchgeführt.

Zur quantitativen Bewertung der Ergebnisse werden, sofern Ground-Truth-Daten über die Szene zur Verfügung stehen, wie zuvor in Kapitel 4, der mittlere absolute Schätzfehler der Skalierung in der Einheit Pixel und der Entfernung in Zentimeter nach (4.40) angegeben. Um die Entfernungsfehler in Relation zur absoluten Entfernung zu setzen, wird darüber hinaus ebenfalls der mittlere prozentuale Entfernungsfehler nach (4.41) herangezogen. Es muss beachtet werden, dass die Umrechnung der Skalierung in eine Entfernung nach (4.8) von den konkreten Objektiv- und Sensorparametern abhängt. Da diese jeweils auf die Gegebenheiten der untersuchten Szene hin ausgerichtet werden, ist eine szenenübergreifende



(a) Reale Aufnahme.



(b) Simulierte Aufnahme.



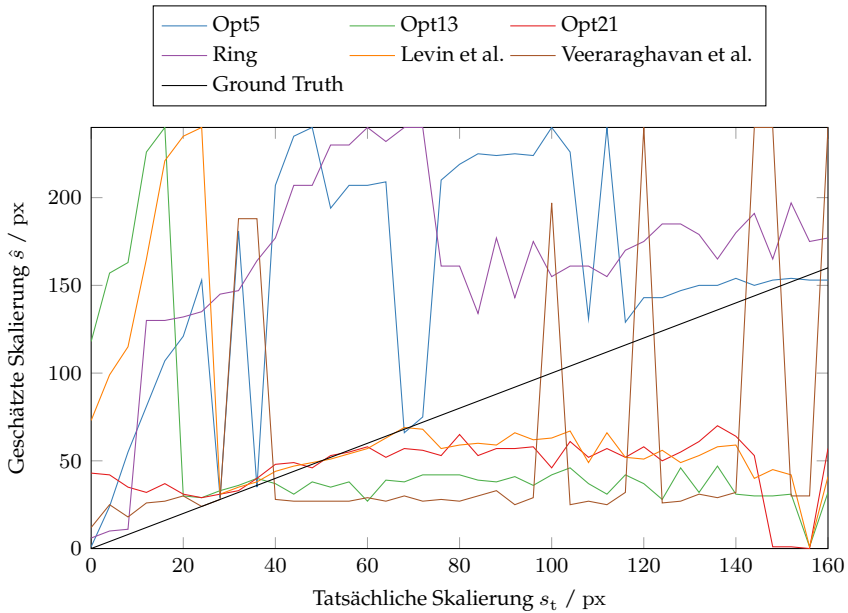
(c) Simulierte Ground Truth.

Abbildung 5.4 Realitätsnahe Szene.

und entfernungsunabhängige Bewertung nur anhand der auftretenden Skalierungsfehler möglich.

## 5.2 Ergebnisse realer Aufnahmen

In Abbildung 5.5 sind die geschätzten Skalierungen  $\hat{s}$  der PSF gegenüber den tatsächlich vorliegenden Skalierungen  $s_t$  dargestellt. Es wurde lediglich das zentrale Pixel einer Aufnahme untersucht, um Auswirkungen der in Abbildung 3.6 festgestellten Vignettierung zu minimieren. Die Aufnahmen wurden auf Skalierungen der Menge  $\mathcal{S} = \{0 \text{ px}, 1 \text{ px}, \dots, 240 \text{ px}\}$  hin untersucht.



**Abbildung 5.5** Ergebnisse der Tiefenschätzung des zentralen Pixels realer Aufnahmen einer Rauschtextur.

Es wird festgestellt, dass unter den vorliegenden Bedingungen die Tiefenschätzung nicht zuverlässig funktioniert. Lediglich mit der Blende „Opt21“ und der Blende nach Levin et al. konnte die Skalierung im Bereich von ca. 20 bis 60 px näherungsweise bestimmt werden. Bei Skalierungen ab 120 px erreicht die Blende „Opt5“ verglichen mit den anderen Formen die besten Ergebnisse. Interessant dabei ist, dass diese Form einen vergleichsweise hohen lichtdurchlässigen Flächenanteil besitzt und die zugehörigen Aufnahmen daher mit geringen Belichtungszeiten aufgenommen wurden. Damit ist bei dieser Aufnahme das Dunkelrauschen verglichen mit den Aufnahmen der anderen Blenden niedrig.

Für dieses schlechte Abschneiden der Tiefenextraktion gibt es mehrere mögliche Ursachen:

- Das Bildrauschen ist aufgrund der hohen Dämpfung des LCDs und trotz der externen Beleuchtung zu stark, sodass charakteristische

Merkmale der OTF in den Energiedichtespektren der Aufnahmen vom Rauschen dominiert und daher nicht erkannt werden können. Maßnahmen zur Reduktion des Bildrauschens sollten derartigen Problemen entgegenwirken. Denkbar ist beispielsweise das mehrmalige Aufnehmen der Bilder mit einer anschließenden Mittelwertbildung [78] oder ein Zusammenfassen mehrerer Pixel, was mit einer Reduktion der Ortsauflösung einhergeht.

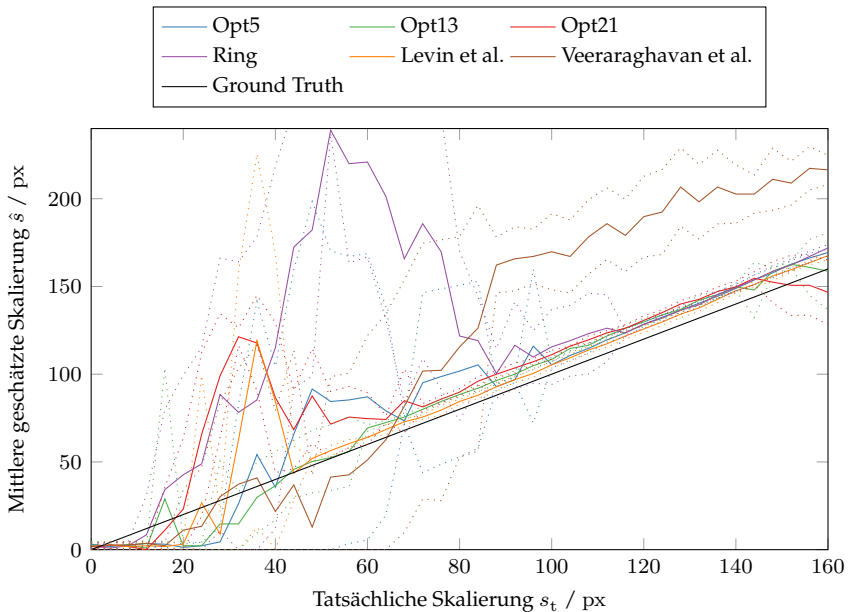
- Obwohl die Auflösung der Rauschtextur der Ebene der angegebenen Auflösung des verwendeten Druckers entspricht, konnten sichtbar im Ausdruck nicht alle Merkmale wiedergegeben werden. Dies entspricht einer Tiefpassfilterung des Anregungssignals. Auch im Fall, dass dies die Ursache für die Qualität der Ergebnisse ist, wäre mit einer Reduktion der Ortsauflösung durch Zusammenfassen mehrerer Pixel mit einer Verbesserung zu rechnen, da so die maximal erfassbare Ortsfrequenz verringert wird.

Die Ergebnisse werden hier trotzdem aufgeführt, um diese in Abschnitt 5.3 mit den Ergebnissen simulierter Daten vergleichen zu können.

Abbildung 5.6 zeigt die Ergebnisse der Tiefenschätzung der realen Aufnahmen, wenn das Eingangsbild durch Zusammenfassen mehrerer Pixel um den Faktor 4 pro Dimension verkleinert wird. Dementsprechend wird auch nur nach Skalierungen in  $\mathcal{S} = \{0 \text{ px}, 1 \text{ px}, \dots, 60 \text{ px}\}$  gesucht. Trotzdem werden die ermittelten Skalierungen zur Visualisierung wieder entsprechend der ursprünglichen Auflösung des Bildes umgerechnet. Um die Anfälligkeit bezüglich Ausreißern zu reduzieren, wurde die Szene an insgesamt neun Pixeln ausgewertet und über die Ergebnisse gemittelt. Die gepunkteten Linien geben die einfache Standardabweichung an. Da sich die ausgewerteten Analysefenster leicht überlappen, können die Einzelergebnisse nicht als stochastisch unabhängig angenommen werden. Die äußeren der ausgewerteten Pixel liegen näher am Bildrand, wo sich potentielle Abbildungsfehler stärker auswirken.

Mit dem Zusammenfassen mehrerer Pixel kann das Ergebnis der Tiefenschätzung deutlich verbessert werden. Jeweils ab einer gewissen Skalierung kann mit allen Formen außer der Form nach Veeraraghavan et al. die Skalierung der PSF zuverlässig geschätzt werden. Der Skalierungsbereich mit geringer Abweichung vom tatsächliche Wert ist für die Form

„Opt13“ am größten (ab ca. 35 px). Mit der Ringblende werden geringe Abweichungen am spätesten erreicht (ab ca. 90 px). Ab diesen Skalierungen weisen die Schätzergebnisse auch eine niedrige Standardabweichung vom Mittelwert auf, was für eine hohe Robustheit spricht. Sofern die Skalierung mit einer Blendenform näherungsweise geschätzt werden kann, ergibt sich für alle diese Formen mehrheitlich eine geringfügig zu groß eingeschätzte Skalierung. Diese Abweichung nimmt mit ansteigender Skalierung  $s_t$  zu. Mögliche Gründe hierfür können einerseits unpräzise Ground-Truth-Daten sein, beispielsweise verursacht durch unpräzises Einstellen der Entfernung der texturierten Ebene oder durch unpräzise Kameraparameter bei der Umrechnung der gewünschten vorliegenden Skalierung in eine Entfernung. Andererseits kann dies auch vom Algorithmus zur Identifikation der Skalierung selbst herrühren.



**Abbildung 5.6** Mittelwert der Ergebnisse der Tiefenschätzung über neun Pixel realer Aufnahmen einer Rauschtextur mit reduzierter Pixelauflösung. Die gepunkteten Linien geben die einfache Standardabweichung vom Mittelwert der Schätzergebnisse an.



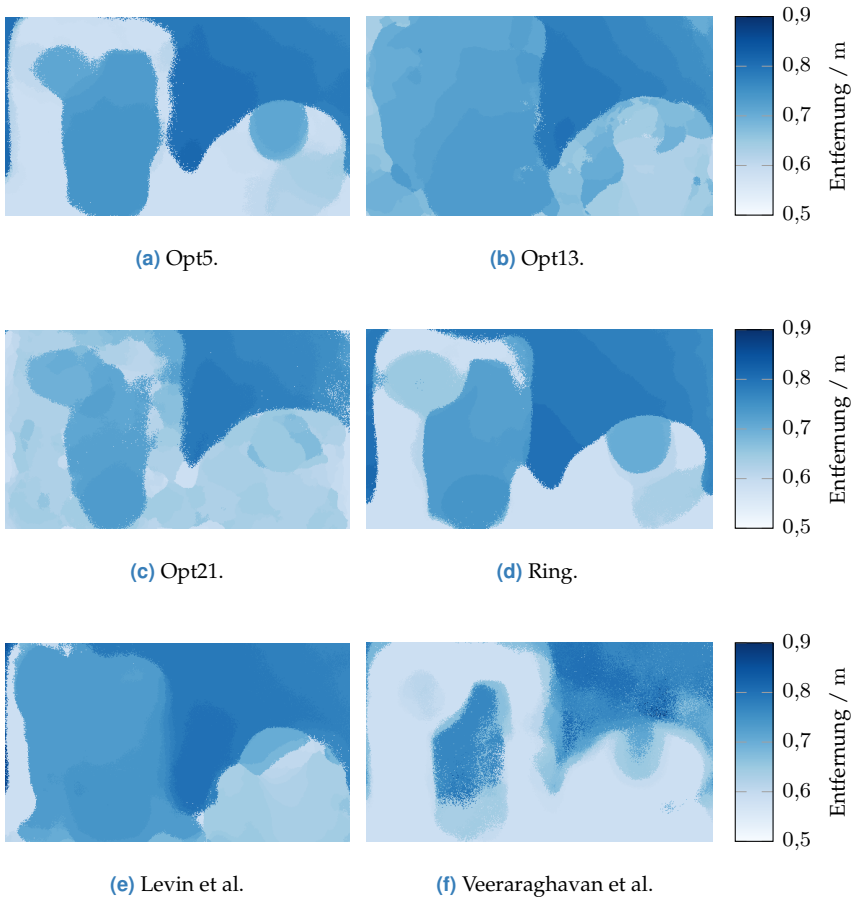
In Tabelle 5.2 sind für die jeweiligen Blendenformen die mittleren Abweichungen bezüglich der Skalierung und der Entfernung sowie die mittlere relative Abweichung bezüglich der Entfernung angegeben. Bezogen auf Skalierungen wird mit der Blende „Opt13“ das beste Ergebnis erreicht, bezüglich der Entfernung schneidet die Blende nach Levin et al. am besten ab. Dies liegt am nichtlinearen Zusammenhang zwischen Skalierung und Entfernung sowie der Betrachtung des Nahfalls aus Abschnitt 4.2. Insgesamt erweisen sich die für die Jensen-Shannon-Divergenz als Distanzmaß optimierten Blendenformen „Opt5“, „Opt13“ und „Opt21“ als geeignet. Lediglich die Blende nach Levin et al. erzielt ähnlich gute Ergebnisse. Die Ringblende ist das Ergebnis einer Optimierung mit anderen Randbedingungen und einer anderen Parametrisierung des Algorithmus zur Tiefenschätzung und erreicht daher nicht das Niveau der anderen ermittelten Formen.

Die bisher vorgestellten Ergebnisse basieren auf den Schätzungen der Tiefe an einem oder einzelnen Punkten des aufgenommenen Bildes, an denen durch die Verwendung der Rauschtextur eine ausreichend breitbandige Anregung durch die Szene sichergestellt ist. In der realitätsnahen Szene aus Abbildung 5.4 ist dies nicht überall gegeben. Die Ergebnisse der dichten Tiefenschätzung für diese Szene sind in Abbildung 5.7 dargestellt. Aufgrund der fehlenden Ground-Truth-Daten kann die Beurteilung dieser Ergebnisse hier nur qualitativ erfolgen.

**Tabelle 5.2** Schätzfehler des Algorithmus zur Tiefenschätzung mit realen Aufnahmen der Szene konstanter Tiefe.

Blendenform	Skalierung	Entfernung	
	<i>MAE</i> /px	<i>MAE</i> /cm	<i>MAPE</i> /%
Opt5	5,17	4,03	5,80
Opt13	<b>2,97</b>	2,57	3,77
Opt21	5,12	4,07	5,85
Ring	12,57	8,20	11,79
Levin et al.	3,03	<b>2,42</b>	<b>3,41</b>
Veeraraghavan et al.	13,96	9,66	15,10

Die visuell besten Ergebnisse können wie zuvor mit der Blende „Opt13“ (Abbildung 5.7(b)) und der Form nach Levin et al. (Abbildung 5.7(e)) erreicht werden. Sowohl die Entfernungen der Rückwand als auch der in der Szene platzierten Objekte kann mit den genannten Blenden korrekt geschätzt werden. Mit der Blende nach Levin et al. können die Abmessungen der Tasse am besten bestimmt werden, wohingegen die Blende „Opt13“ am linken Bildrand die geringsten Fehler aufweist.



**Abbildung 5.7** Tiefenschätzergebnisse realer Aufnahmen der realitätsnahen Szene.

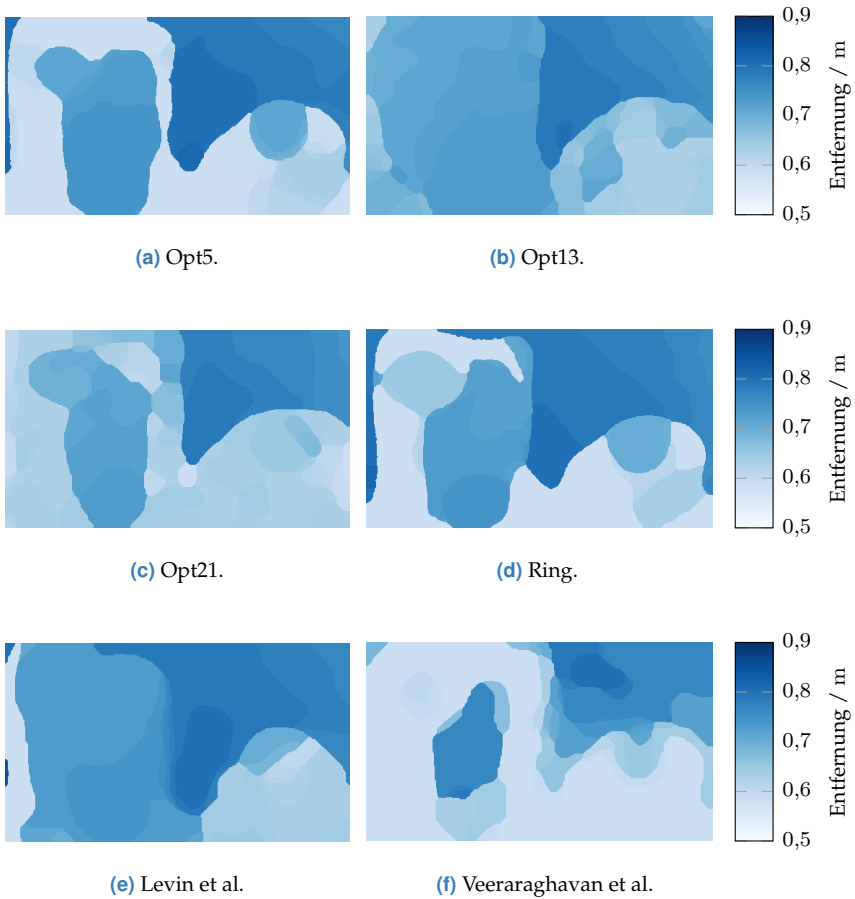
In den Ergebnissen aller Blenden ist zu beobachten, dass in Bereichen der Szene mit breitbandiger Anregung die Tiefenschätzung besser funktioniert als in örtlich homogenen Bereichen. Dies wird vor allem an den Bereichen des Buches mit Schrift, beispielsweise in Abbildung 5.7(a) und (c), deutlich. Auf der Rückwand der Tasse entsteht durch die Beleuchtung ein punktförmiger Lichtreflex, an dem sich die Form und die Skalierung der verwendeten Blende leicht erkennen lässt. Daher kann in diesem begrenzten Bildbereich die Tiefe ebenfalls gut ermittelt werden, was vor allem an den Ergebnissen der Blende „Opt5“ und der Ringblende (Abbildungen 5.7(a) und (d)) sichtbar ist.

In den homogenen Bildbereichen mit wenig Struktur, z. B. in den äußeren Bereichen des Buchrückens, kann die Tiefe der Szene nicht zuverlässig geschätzt werden. Dies liegt daran, dass aufgrund der unzureichenden Anregung die markanten Merkmale der OTF nicht im Energiedichtespektrum zu beobachten sind. Die ausschließliche Anregung niedriger Ortsfrequenzen wird als äußerst unscharfe Abbildung interpretiert, weshalb mit den meisten Blendenformen die größte Skalierung aus der Menge der untersuchten Skalierungen  $\mathcal{S}$  angenommen wird. Lediglich mit den Blenden „Opt13“ und „Opt21“ (Abbildungen 5.7(b) und (c)) wird eine davon verschiedene und damit näher am tatsächlichen Wert liegende Skalierung bestimmt, was darauf hindeutet, dass mit diesen Formen die Skalierung bei schmalbandiger Anregung besser geschätzt werden kann.

In den ermittelten Tiefenkarten sind teilweise einzelne Pixel mit von der Umgebung deutlich abweichenden Tiefenwerten zu beobachten. Dies widerspricht der gängigen Annahme stückweise zusammenhängender Objektflächen. Die Hinzunahme der Nebenbedingung einer stückweise glatten Tiefenkarte und Lösung des resultierenden Minimierungsproblems mit dem *Graph-Cut*-Algorithmus führt auf die in Abbildung 5.8 dargestellten Ergebnisse.

Die beschriebenen einzelnen Pixel mit abweichender Tiefe können durch diese Nachbearbeitung zuverlässig beseitigt werden, bei gleichzeitiger Erhaltung der Objektgrenzen. Große Bildbereiche mit falsch geschätzter Skalierung können allerdings auch mit der Hinzunahme dieser Nebenbedingung nicht korrigiert werden. Daher sind die bei einer visuellen Betrachtung größten Verbesserungen auch bei den Blendenformen zu verzeichnen, die zuvor rauschähnliche Bildbereiche in der Tiefenkarte

aufwiesen, insbesondere also bei den Blendenformen „Opt5“, „Opt21“, „Ring“ und der Form nach Veeraraghavan et al.



**Abbildung 5.8** Tiefenschätzergebnisse realer Aufnahmen der realitätsnahen Szene mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

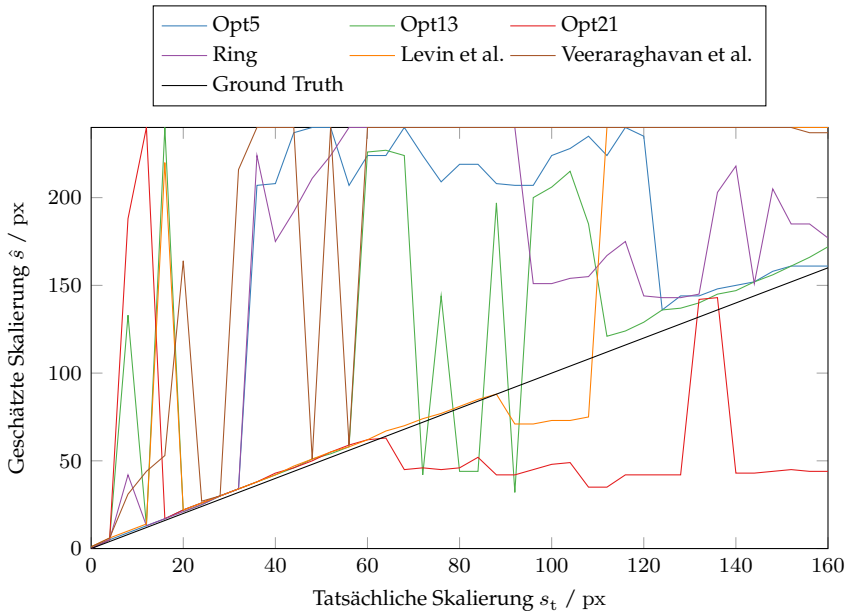
## 5.3 Ergebnisse simulierter Aufnahmen

Bei der Auswertung simulierter Szenen steht neben der erreichbaren Qualität, die anhand der verfügbaren Ground-Truth-Daten beurteilt wird, vor allem die Frage im Vordergrund, inwieweit die Ergebnisse der Simulation denen der realen Aufnahmen entsprechen. Außerdem werden die Auswirkungen verschiedener Störeffekte auf die Qualität der Tiefenschätzung untersucht.

In Abbildung 5.9 sind zunächst die Ergebnisse der Skalierungsschätzung des jeweils zentralen Pixels der simulierten Aufnahmen der Rauschtexturebene bei voller Sensorauflösung dargestellt. Im Gegensatz zu den realen Aufnahmen, bei denen das Anregungssignal aufgrund des Ausdrucks der Rauschtextur sichtbar tiefpassgefiltert ist, kann in der Simulation von einem breitbandigen Anregungsspektrum ausgegangen werden. Trotzdem wird die Skalierung nur stellenweise korrekt geschätzt. Ähnlich wie bei den realen Aufnahmen stimmen die geschätzten Skalierungen lediglich bei einigen Blendenformen im Bereich bis  $s = 60$  px teilweise mit der Erwartung überein. Mit der Blendenform nach Levin et al. ist eine korrekte Schätzung bis ca. 90 px möglich.

Der Vergleich der Ergebnisse dieser Aufnahmen deutet also einerseits darauf hin, dass das vorherrschende Bildrauschen, verursacht durch die hohe Intensitätsdämpfung des verwendeten LCDs und die Kompensation des Störanteils, für die nicht zufriedenstellenden Tiefenschätzergebnisse bei voller Sensorauflösung verantwortlich sind. Andererseits reagiert der Algorithmus ähnlich empfindlich auf dieses Bildrauschen, unabhängig davon, ob reale oder simulierte Aufnahmen untersucht werden.

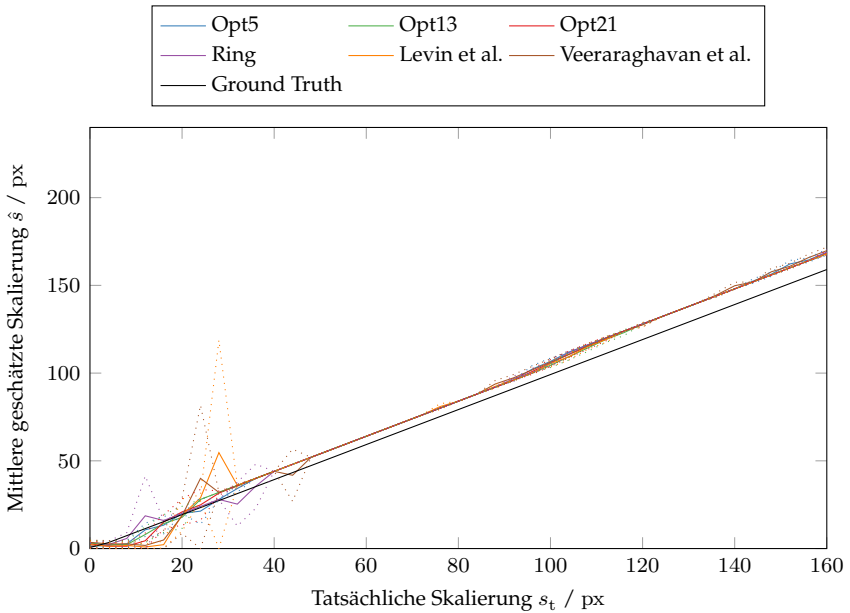
Zur Verbesserung der Qualität der Schätzergebnisse wird analog zu den realen Aufnahmen die Auflösung der simulierten Aufnahmen durch Zusammenfassen benachbarter Pixel verkleinert, wodurch das Bildrauschen deutlich reduziert wird. Im Gegensatz zu den realen Aufnahmen lassen sich in der Simulation neue Texturen einfach prozedural generieren, um so neue Szenen zu erzeugen. Um auch die Aussagekraft der simulativ ermittelten Ergebnisse zu steigern, werden daher mehrere Abbildungen pro untersuchter Skalierung erzeugt und jeweils am zentralen Pixel ausgewertet. Abbildung 5.10 zeigt die gemittelten Ergebnisse für neun simulierte Aufnahmen, wobei die ermittelten Skalierungen wieder auf die volle



**Abbildung 5.9** Ergebnisse der Tiefenschätzung der zentralen Pixel simulierter Aufnahmen einer Rauschtextur.

Sensorauflösung umgerechnet wurden. Die gepunkteten Linien stellen wiederum die einfache Standardabweichung der Schätzergebnisse dar. Da die Texturen bei jeder Aufnahme zufällig neu erzeugt werden, sind die Ergebnisse der einzelnen Aufnahmen stochastisch unabhängig.

Wie zuvor in Abschnitt 5.2 kann mit dem reduzierten Bildrauschen die Qualität der Schätzergebnisse für alle Blendenformen deutlich verbessert werden. Ab Skalierungen von ca. 35 px kann mit allen Blendenformen die Skalierung der PSF korrekt und robust identifiziert werden. Mit den in dieser Arbeit ermittelten Formen kann die Skalierung teilweise schon bei deutlich kleineren Skalierungen präzise geschätzt werden. Auch hier ist wie zuvor bei den realen Aufnahmen zu beobachten, dass mit zunehmender Skalierung die Skalierung tendenziell zu groß geschätzt wird. Dieses Phänomen ist also nicht auf unzureichend bestimmte Kameraparameter oder unpräzise Ground-Truth-Daten zurückzuführen, da diese in der



**Abbildung 5.10** Mittelwert der Ergebnisse der Tiefenschätzung der zentralen Pixel simulierter Aufnahmen einer Rauschtextur mit reduzierter Pixelauflösung. Die gepunkteten Linien geben die einfache Standardabweichung vom Mittelwert der Schätzergebnisse an.

Simulation exakt vorliegen. Da die beobachtete Tendenz sowohl in den Auswertungen realer als auch simulierter Aufnahmen auftritt, kann gefolgert werden, dass die Ursache hierfür im Algorithmus zur Bestimmung der Skalierung der PSF liegt.

In Tabelle 5.3 sind die Schätzfehler in Abhängigkeit der Blendenform angegeben. Auch hier schneidet die Blende „Opt13“ bezüglich der Skalierungsabweichung am besten ab. Insgesamt sind die Unterschiede der Blendenformen aber sehr gering.

**Tabelle 5.3** Schätzfehler des Algorithmus zur Tiefenschätzung mit simulierten Aufnahmen der Szene konstanter Tiefe.

Blendenform	Skalierung	Entfernung	
	<i>MAE</i> /px	<i>MAE</i> /cm	<i>MAPE</i> /%
Opt5	1,49	<b>1,06</b>	<b>1,66</b>
Opt13	<b>1,47</b>	1,08	1,67
Opt21	1,57	1,17	1,80
Ring	1,53	1,11	1,73
Levin et al.	1,78	1,36	2,02
Veeraraghavan et al.	1,83	1,40	2,11

Der größte Unterschied zu den Ergebnissen der realen Aufnahmen aus Abbildung 5.6 ist das bessere Schätzergebnis, insbesondere bei kleinen Skalierungen. Dies kann auf die beschriebenen unterschiedlichen Szenenanregungen aufgrund des Drucks der Rauschtextur zurückgeführt werden. Eine Tiefpassfilterung des Anregungssignals beim Ausdrucken bedeutet, dass im Energiedichtespektrum das Nutzsignal bei hohen Ortsfrequenzen einen geringeren Abstand zum Rauschniveau aufweist. Insbesondere bei kleinen Skalierungen der PSF können die charakteristischen Merkmale der OTF in diesen Bereich der Ortsfrequenz fallen. Erst bei größeren Skalierungen nehmen die Ortsfrequenzen, bei denen diese Merkmale beobachtet werden können, ab. Daher bewirkt eine Tiefpassfilterung des Anregungssignals, dass die Skalierung der PSF erst bei größeren Skalierungen identifiziert werden kann.

Zur Untersuchung der Auswirkungen geometrischer Abbildungsfehler wird dieselbe Szene der texturierten Ebene genutzt. Im Gegensatz zu den obigen Betrachtungen wird jedoch eine dicke sphärische Linse simuliert. Die Entfernung der Ebene wird in diesem Fall nicht variiert. Die Entfernung der Ebene zur Kamera beträgt  $d_o = 0,690$  m, was einer Skalierung der PSF von  $s = 60$  px entspricht. Bei dieser Entfernung kann ohne explizite Hinzunahme von Abbildungsfehlern sowohl mit simulierten als auch mit realen Aufnahmen die Tiefe gut ermittelt werden. Anstatt der Entfernung wird der Winkel  $\alpha$  zur optischen Achse, unter dem das ausgewertete Pixel beobachtet wird, variiert. Bei  $\alpha = 0^\circ$  liegt das ausgewertete

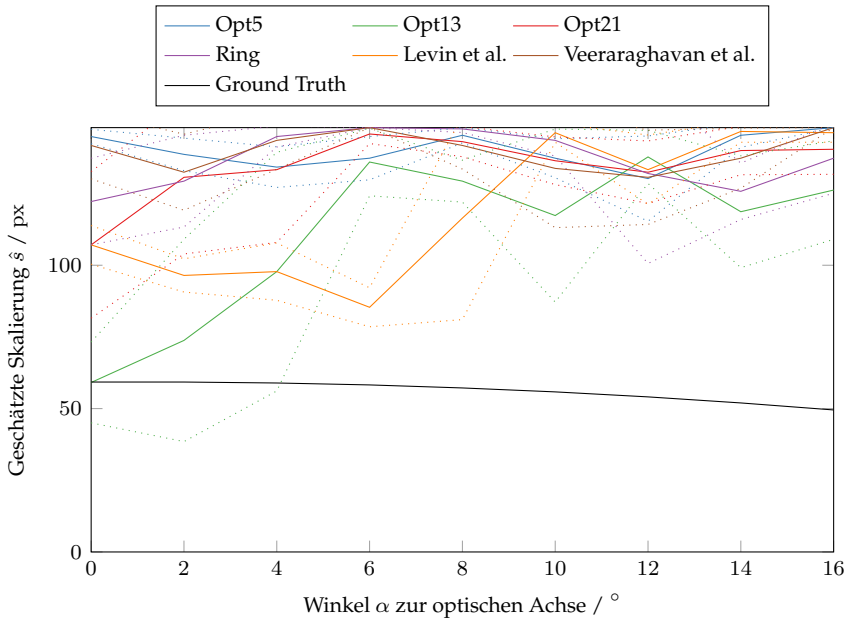


Pixel auf der optischen Achse. Nach Kapitel 2 wirkt sich hier lediglich die sphärische Aberration auf die Form der PSF aus. Mit zunehmendem Winkel  $\alpha$  bewirken auch die Abbildungsfehler Astigmatismus und Koma eine stärkere Formänderung der PSF. Pro Blendenform und Beobachtungswinkel wurden neun Aufnahmen mit zufällig generierter Rauschtextur simuliert.

Abbildung 5.11 zeigt die Mittelwerte der aus den Aufnahmen ermittelten Skalierungen in Abhängigkeit des Beobachtungswinkels  $\alpha$ . Die tatsächlich vorliegende Skalierung nimmt unter größerem Winkel ab, weil mit zunehmendem Beobachtungswinkel die Entfernung zur texturierten Ebene zunimmt. Somit liegt die Objektentfernung näher an der eingestellten Fokusweite  $d_{o,\text{fok}}$ .

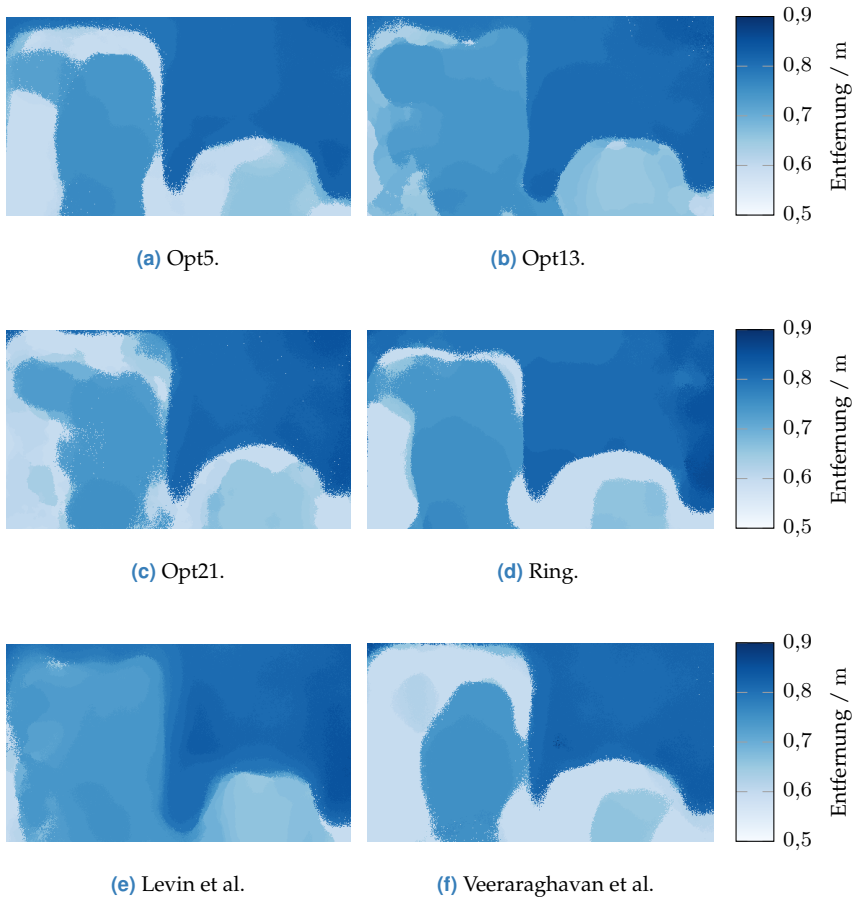
Mit der Blende „Opt13“ kann die Skalierung der PSF bei einem Winkel von  $\alpha = 0^\circ$  am besten geschätzt werden. Die durch die gepunkteten Linien dargestellte einfache Standardabweichung der Schätzergebnisse ist jedoch deutlich höher als bei der Simulation ohne Abbildungsfehler. Mit den übrigen Blenden wird die Skalierung durchweg zu groß geschätzt. Bei der Simulation der PSF einer dicken sphärischen Linse in Abschnitt 2.4.2.4 kann in Abbildung 2.18 beobachtet werden, dass bei Objektentfernungen kleiner als die Fokusweite ( $\Delta d_i = -2 \text{ mm}$ ) der Durchmesser der PSF eher kleiner als bei der Simulation einer idealen dünnen Linse aus Abbildung 2.17 ist. Vor diesem Hintergrund ist zu erwarten, dass die Skalierungen hier ebenfalls zu klein geschätzt werden. Die Tatsache, dass dies nicht der Fall ist, deutet darauf hin, dass durch die sphärischen Aberrationen die PSF so stark verzerrt wurde, dass die Merkmale der OTF im Energiedichtespektrum nicht mehr zuverlässig erkannt werden können.

Mit zunehmendem Beobachtungswinkel  $\alpha$  kann schließlich auch mit der Blende „Opt13“ die Skalierung nicht mehr geschätzt werden. Dies ist auf die sich nun zusätzlich auswirkenden Abbildungsfehler Astigmatismus und Koma zurückzuführen. Wie zu erwarten kann bei Winkeln von  $\alpha > 0$  auch mit den anderen Blenden die Skalierung nicht zuverlässig geschätzt werden.



**Abbildung 5.11** Mittelwert der Ergebnisse der Tiefenschätzung der zentralen Pixel simulierter Aufnahmen einer Rauschtextur mit Abbildungsfehlern und reduzierter Pixelauflösung in Abhängigkeit des Beobachtungswinkels. Die gepunkteten Linien geben die einfache Standardabweichung der Schätzergebnisse an.

Abschließend werden die Ergebnisse der realitätsnahen Szene aus Abbildung 5.4(b) betrachtet. Die ermittelten Tiefenkarten der simulierten Aufnahmen der Szene sind in Abbildung 5.12 für die verschiedenen Blenden dargestellt. Bei visueller Betrachtung schneidet der Algorithmus mit den simulierten Aufnahmen leicht besser als mit den realen Aufnahmen ab. Gleichzeitig lassen sich sowohl in den Ergebnissen der simulierten als auch der realen Aufnahmen zahlreiche Gemeinsamkeiten feststellen: In den Bereichen, in denen die Szene Strukturen aufweist, kann wie im Fall der realen Aufnahmen die Tiefe der Szene besser als in homogenen Bereichen geschätzt werden. In diesen homogenen Bereichen wird zumeist die größte Skalierung in  $S$  und damit die geringste Entfernung zur Kamera geschätzt.



**Abbildung 5.12** Tiefenschätzergebnisse simulierter Aufnahmen der realitätsnahen Szene.

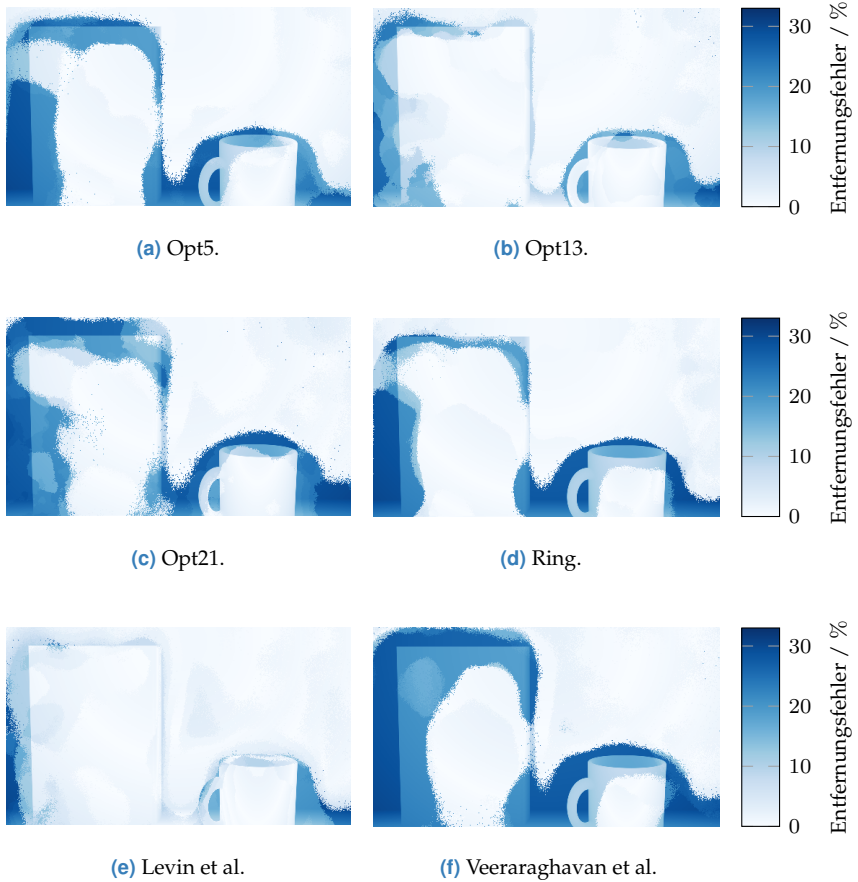
Auch der Vergleich der Blendenformen untereinander fällt ähnlich wie bei den realen Aufnahmen aus. Die Form „Opt13“ und die Form nach Levin et al. erzielen im Vergleich zu den anderen Formen die besten Ergebnisse. Die Tiefenkarte der Form nach Veeraraghavan et al. weist sichtbar die größten Bildregionen mit minimaler Entfernung auf, was der maximalen untersuchten Skalierung entspricht – eine Entfernung, die in der Szene nicht auftritt. Der quantitative Vergleich der Schätzergebnisse mit den

simulierten Ground-Truth-Daten bestätigen diesen Eindruck. Die ermittelten Schätzfehler sind in Tabelle 5.4 angegeben. Für alle Blendenformen fallen die Schätzfehler größer als bei der Analyse der texturierten Ebene aus. Dies liegt an der hohen Komplexität der realitätsnahen Szene, sowohl bezüglich der Varianz der Tiefenkarte als auch bezüglich der Bandbreite der beobachteten Texturen.

Mit den in der Simulation vorliegenden Ground-Truth-Daten der Szene kann genauer untersucht werden, in welchen Bildregionen die größten Schätzfehler auftreten. Abbildung 5.13 zeigt die relativen Entfernungsfehler der ermittelten Tiefenkarten für die jeweiligen Blendenformen. Es zeigt sich, dass neben den Bereichen mit geringer Struktur große Schätzfehler vor allem an Objektgrenzen auftreten. Abrupte Kanten der Intensität an Objektgrenzen sorgen zwar für hochfrequente Anteile im Energiedichtespektrum, allerdings nur entlang einer schmalen Linie. Diese Anregung reicht nicht aus, um die Skalierung der PSF in diesen Bereichen zuverlässig zu schätzen.

**Tabelle 5.4** Schätzfehler des Algorithmus zur Tiefenschätzung bei einer simulierten Aufnahme der realitätsnahen Szene.

Blendenform	Skalierung	Entfernung	
	<i>MAE/px</i>	<i>MAE/cm</i>	<i>MAPE/%</i>
Opt5	23,21	5,20	6,61
Opt13	12,28	3,09	3,85
Opt21	26,95	6,13	7,71
Ring	21,12	4,74	6,00
Levin et al.	<b>9,87</b>	<b>2,54</b>	<b>3,16</b>
Veeraraghavan et al.	33,92	7,46	9,48



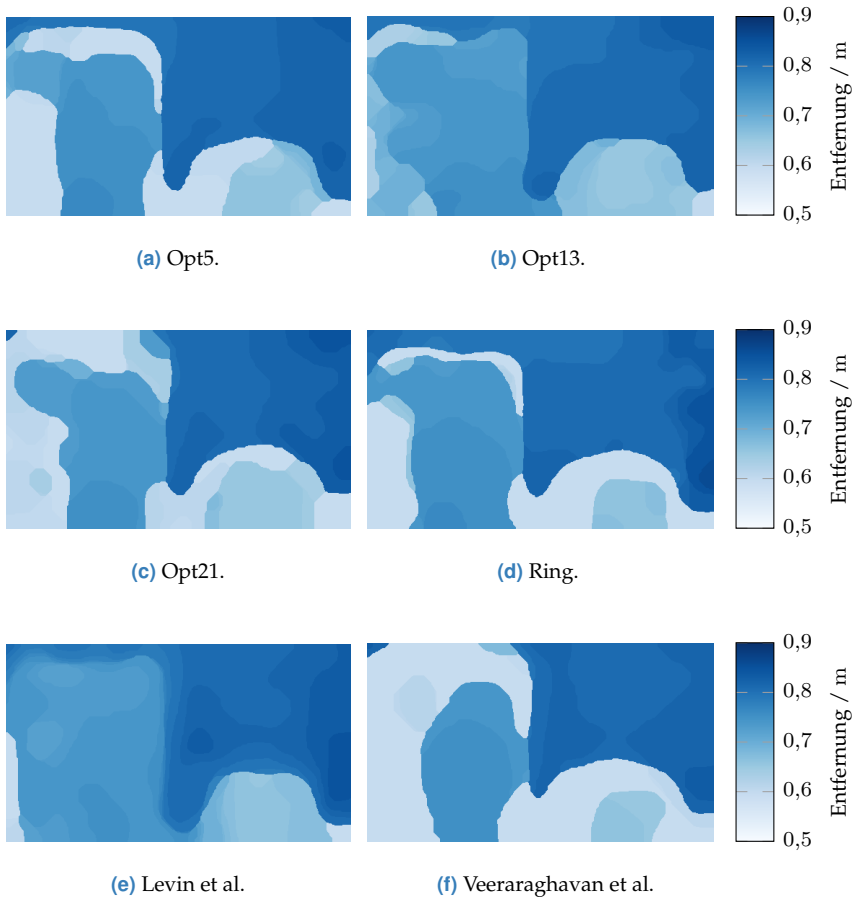
**Abbildung 5.13** Relative Entfernungsfehler der Tiefenkarten der simulierten Aufnahmen der realitätsnahen Szene.

**Tabelle 5.5** Schätzfehler des Algorithmus zur Tiefenschätzung bei einer simulierten Aufnahme der realitätsnahen Szene mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

Blendenform	Skalierung	Entfernung	
	MAE/px	MAE/cm	MAPE/%
Opt5	23,29	5,22	6,61
Opt13	11,88	3,00	3,73
Opt21	27,38	6,21	7,81
Ring	21,11	4,72	5,99
Levin et al.	<b>9,60</b>	<b>2,47</b>	<b>3,07</b>
Veeraraghavan et al.	34,47	7,55	9,59

Die rauschartigen Tiefenwerte in Abbildung 5.12, insbesondere an Objektgrenzen, entstehen aufgrund der unabhängigen Betrachtung der einzelnen Pixel bei der Identifikation der Skalierung. Durch eine Nachbearbeitung mit dem *Graph-Cut*-Algorithmus können derartige Artefakte unterdrückt werden. Abbildung 5.14 zeigt die entsprechend nachbearbeiteten Tiefenkarten. Einzelnen Pixel mit von der Umgebung abweichender Tiefe treten in diesen Tiefenkarten nicht mehr auf. Durch das Zuordnen der Tiefe von Bildregionen geringer Ausdehnung zur in der Umgebung vorherrschenden Tiefe werden Kanten in der Tiefenkarte schärfer. Die Lage der Regionen großer Schätzfehler hat sich durch die Nachbearbeitung nicht grundlegend verändert. Wie an den relativen Entfernungsehlern aus Abbildung 5.15 ersichtlich ist, entstehen die größten Schätzfehler weiterhin in Bildregionen mit wenig Struktur und an Objektkanten.

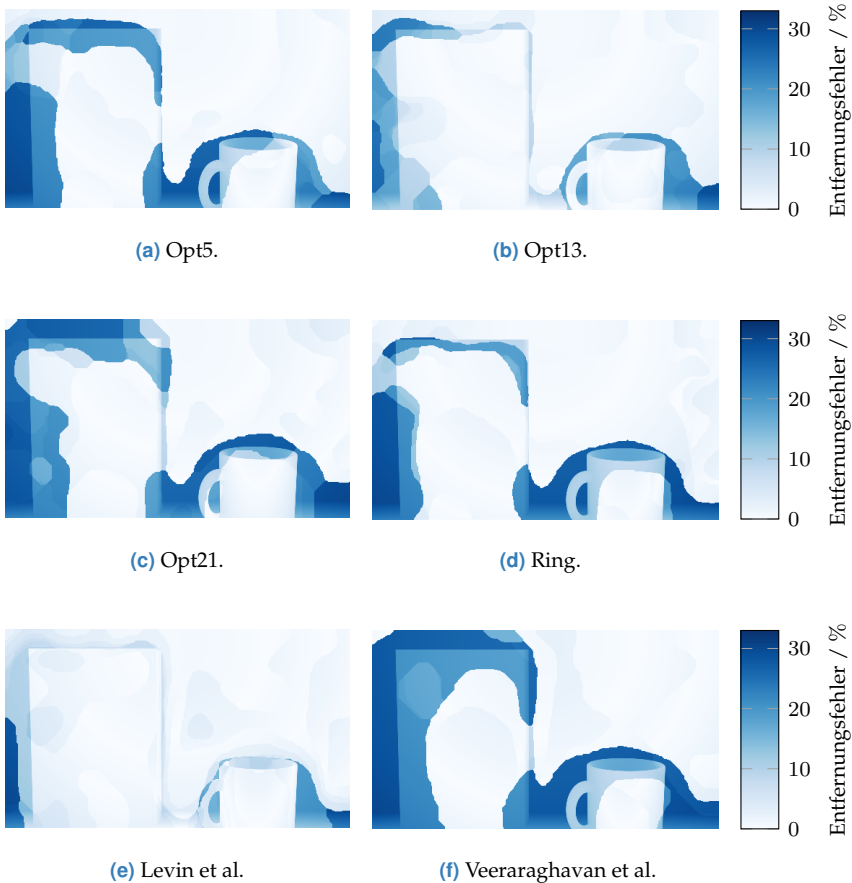
Beim quantitativen Vergleich der Ergebnisse in Tabelle 5.5 mit den Ergebnissen aus Tabelle 5.4 muss allerdings festgestellt werden, dass nicht immer eine Verbesserung durch die Nachbearbeitung erreicht wird. Die Qualität der ermittelten Tiefenkarten nimmt bei den Formen „Opt5“ und „Opt21“ sowie der Form nach Veeraraghavan et al. geringfügig ab. Bei den anderen Blendenformen zeigt sich eine geringfügige Verbesserung des Schätzfehlers. Insgesamt ändern sich bei allen Blendenformen die Schätzfehler durch die Nachbearbeitung nur sehr wenig.



**Abbildung 5.14** Tiefenschätzergebnisse simulierter Aufnahmen der realitätsnahen Szene mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

Es fällt auf, dass gerade bei den drei Blendenformen, die verglichen mit den anderen drei Formen geringere Schätzfehler aufweisen, eine Verbesserung durch die Nachbearbeitung zu beobachten ist. Bei den anderen drei Formen dagegen eine Verschlechterung. Eine mögliche Ursache dafür ist, dass je besser die Qualität der unbearbeiteten Tiefenkarte ist, desto wahrscheinlicher ist es auch, dass durch die Korrektur der Skalierung eines

Pixels der Schätzfehler verringert wird. Wenn bereits großen Bildregionen die korrekte Skalierung zugeordnet werden konnte, werden einzelne Pixel durch Korrektur der Skalierung eher zu diesen Bildregionen hinzugefügt. Die Hinzunahme eines Pixels zu einer Bildregion mit stark abweichender Skalierung ist unwahrscheinlich, da bei einer Tiefenkarte mit geringem Schätzfehler nur wenige dieser Regionen vorhanden sind.



**Abbildung 5.15** Relative Entfernungsfehler der Tiefenkarten der simulierten Aufnahmen der realitätsnahen Szene mit Nachbearbeitung mit dem *Graph-Cut*-Algorithmus.

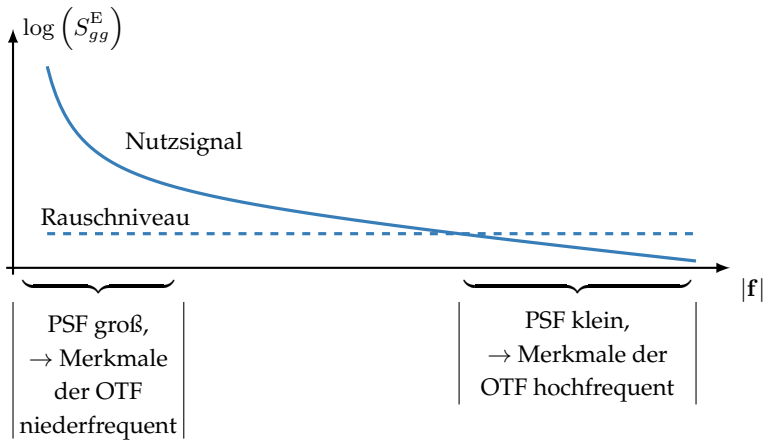


## 5.4 Fazit

In den vorigen Abschnitten wurde der in Kapitel 4 vorgestellte Algorithmus zur monokularen Tiefenschätzung ausgiebig untersucht. Dabei wurde der Algorithmus stets mit denselben Parametern betrieben, unabhängig davon, ob simulierte Aufnahmen der in Kapitel 2 vorgestellten Simulationsumgebung oder reale Aufnahmen des Prototyps aus Kapitel 3 als Eingangsdaten verwendet werden.

Es zeigt sich qualitativ eine hohe Übereinstimmung der Ergebnisse simulierter und realer Aufnahmen. In beiden Fällen kann aufgrund des vorherrschenden Bildrauschens die Tiefe der Szene nicht in der vollen Auflösung des Sensors bestimmt werden. Abhilfe schafft jeweils die Reduktion der Auflösung durch Zusammenfassen mehrerer Pixel, wodurch das Bildrauschen deutlich reduziert wird. Unter diesen Bedingungen kann die Tiefe der Szene mit dem Algorithmus abhängig von der Blendenform über einen großen Entfernungsbereich zuverlässig und robust geschätzt werden.

Bei kleinen Skalierungen der PSF – bei simulierten Aufnahmen bis ca. 20 px, bei realen bis ca. 35 px – kann auch bei reduzierter Auflösung die Tiefe mit keiner der untersuchten Blendenformen präzise geschätzt werden. Eine mögliche Ursache hierfür lässt sich anhand des skizzierten Energiedichtespektrums in Abbildung 5.16 erläutern. Bei sehr kleinen Skalierungen  $s$  liegen die markanten Merkmale der OTF bei hohen Ortsfrequenzen. In diesem Frequenzband ist jedoch üblicherweise der Abstand des Nutzsignals zum Bildrauschen gering oder das Nutzsignal liegt sogar unterhalb des Rauschniveaus. Daher können die Merkmale der OTF in diesem Fall im Bildrauschen untergehen. Die PSF kann also erst ab einer bestimmten Ausdehnung zuverlässig identifiziert werden. Das bezüglich des erreichten Schätzfehlers bessere Abschneiden der simulierten Aufnahmen kann somit auf den im Vergleich zu den realen Aufnahmen höheren Nutzsignalanteil bei hohen Frequenzen zurückgeführt werden. Bei den simulierten Aufnahmen heben sich die Merkmale der OTF schon bei niedrigeren Skalierungen vom Bildrauschen ab.



**Abbildung 5.16** Qualitative Einteilung der Skalierung anhand der Lage der Merkmale der OTF im beobachteten Energiedichtespektrum.

Umgekehrt konzentrieren sich bei großen Skalierungen  $s$  die markanten Merkmale der OTF in einem niedrigen Frequenzband. Je kleiner dieses Band ist, desto schlechter können die Merkmale aufgrund der endlichen Ortsfrequenzauflösung identifiziert werden. Dadurch wird auch bei, verglichen mit der Fenstergröße, großen Skalierungen die Tiefenschätzung erschwert.

Sofern der zu beobachtende Messraum vorab bekannt ist, kann durch eine geeignete Wahl der Fokusweite  $d_{o,\text{fok}}$  sichergestellt werden, dass die auftretenden Skalierungen robust mit dem System ermittelt werden können. Durch eine geeignete Wahl der Fokusweite am Rand des Messraums sollte außerdem festgelegt werden, ob sich Objekte nur ausschließlich vor oder hinter der Fokusebene befinden. Dies ist notwendig, damit eine ermittelte Skalierung der PSF eindeutig in eine Entfernung umgerechnet werden kann. Beim im Rahmen dieser Arbeit aufgebauten Objektivprototyp wurden die grundlegenden Eigenschaften des unmodifizierten Objektivs beibehalten. Somit sind mit diesem Aufbau Fokusweiten von wenigen Zentimeter bis ins Unendliche einstellbar.

Die Auswertungen der vorigen Abschnitte bestätigen, dass die Form der Blende die Leistungsfähigkeit der Kamera zur Tiefenschätzung nach

dem *Depth-from-Defocus*-Prinzip beeinflusst. Es zeigen sich sowohl in der Simulation als auch bei den realen Aufnahmen ähnliche Ergebnisse beim Vergleich der Formen untereinander. Die bei einer Optimierung im Rahmen dieser Arbeit ermittelte Form „Opt13“ und die Form, die von Levin et al. [49] vorgeschlagen wird, erzielen in allen Untersuchungen die geringsten Schätzfehler im Vergleich zu den anderen untersuchten Formen. Die Blendenform nach Veeraraghavan et al. [100] weist im Vergleich die größten Schätzfehler auf. Dabei muss allerdings beachtet werden, dass die Autoren die Form hinsichtlich der Auswertung anderer Merkmalstypen als in dieser Arbeit optimiert haben.

Mit der Simulationsumgebung konnten darüber hinaus die Auswirkungen geometrischer Abbildungsfehler auf die Tiefenschätzung untersucht werden. Dabei zeigt sich, dass Abbildungsfehler eine Identifizierung der Skalierung deutlich erschweren. Umso wichtiger ist es, dass das Objektiv mit codierter Blende gute Abbildungseigenschaften aufweist, was in dieser Arbeit eines der Ziele beim Aufbau des Objektivprototyps war.

Die Schätzergebnisse der realitätsnahen Szene zeigen, dass mit der Kamera mit programmierbarer Apertur auch in komplexen Szenarien eine Tiefenschätzung möglich ist. In Bildregionen mit ausreichend Struktur kann die Tiefe der Szene mit geringen Schätzfehlern bestimmt werden. Mit den simulierten Aufnahmen, die einen direkten Vergleich mit der Ground-Truth-Tiefe ermöglichen, konnte im besten Fall ein mittlerer prozentualer Entfernungsfehler von ca. 3 % erreicht werden.



## 6 Zusammenfassung

Mit dem neuen Denkmuster des *Computational Imaging* ergeben sich vielfältige Möglichkeiten für den Entwurf und Aufbau von Kamerasysteme. In dieser Arbeit wurde ein Entwurfsprozess neuartiger *Computational Camera* vorgestellt, der anhand der monokularen Tiefenschätzung auf Grundlage des Einzelbildes einer Kamera mit codierter Blende angewendet wurde. Durch die vorgeschlagene Kamerasimulation kann bereits vor oder während des Aufbaus eines Kameraprototyps die konkrete hardwareseitige Auslegung sowie die digitale Signalverarbeitung des Systems erprobt, bewertet und verbessert werden.

Eine im Laufe der Arbeit entwickelte Simulationsumgebung nach dem Raytracing-Verfahren ermöglicht eine im Rahmen der geometrischen Optik akkurate Simulation des Bildgewinnungsprozesses einer Kamera. Weil Effekte wie Bildrauschen und geometrische Abbildungsfehler starke Auswirkungen auf die Leistungsfähigkeit einer *Computational Cameras* haben können, wurde bei der Implementierung besonders Wert auf die mit realen Kameras vergleichbare Simulation von Störeinflüssen gelegt. Dies wurde u. a. durch eine Implementierung des EMVA-Störmodells und die Simulation dicker sphärischer Linsen erreicht. Damit wurde die Aussagekräftigkeit der simulierten Aufnahmen sichergestellt.

Parallel zur Kamerasimulation wurde ein Prototyp einer Kamera mit programmierbarer Apertur aufgebaut. Dazu wurde ein transmissives Flüssigkristalldisplay in die Apertur ebene eines handelsüblichen Objektivs eingebaut. Der Prototyp erlaubt die berührungslose Änderung der Blendenform und damit der Übertragungsfunktion der Kamera. Die Eigenschaften des Prototyps wurden ausgiebig untersucht. Es konnten sowohl die Einstellmöglichkeiten der Fokussierung als auch die hohe Abbildungsqualität des ursprünglichen Objektivs erhalten werden.

Die mit einer codierten Blende aufgenommenen Bilder ermöglichen die Identifikation des Grads der Unschärfe im Bild und damit eine Schätzung der Tiefe der betrachteten Szene. Der im Rahmen der Arbeit entworfene

*Depth-from-Defocus*-Algorithmus vergleicht das Energiedichtespektrum eines Bildausschnitts mit einem erwarteten Energiedichtespektrum eines linearen Signalmodells für verschiedene angenommene Unschärfegrade. Der Grad der Unschärfe wird durch die Skalierung der Punktverschmierungsfunktion des Systems ausgedrückt. Als Vergleichsmaß wird u. a. die Jensen-Shannon-Divergenz verwendet. Damit kann eine dichte Tiefenkarte der Szene geschätzt werden. Durch Formulierung der Schätzaufgabe als Energieminimierung kann mit einer optionalen Nachbearbeitung durch Voraussetzen eines stückweise stetigen Verlaufs die Qualität der Tiefenkarte verbessert werden. Anhand des entworfenen Algorithmus wurden darüber hinaus mit einer vollständigen Suche nach rotationssymmetrischen Formen und mit einem genetischen Algorithmus allgemeine Blendenformen ermittelt, die besonders für die monokulare Tiefenschätzung geeignet sind.

Das entwickelte Gesamtsystem, bestehend aus Kamera mit programmierbarer Apertur und nachfolgendem Algorithmus zur Schätzung der Skalierung der Punktverschmierungsfunktion, wurde anhand zweier Test-szenarien unterschiedlicher Komplexität erprobt. Sofern das Bildrauschen nicht zu groß ist, kann in beiden Szenarien sowohl mit simulierten als auch mit realen Aufnahmen die Tiefe der Szene mit Abweichungen im Bereich weniger Prozent geschätzt werden. Insbesondere mit einer der in dieser Arbeit ermittelten Blendenformen und einer in der Literatur vorgeschlagenen Form konnten die besten Ergebnisse erzielt werden.

Mit den durchgeführten Untersuchungen konnte die Aussagekraft der anhand simulierter Aufnahmen erzielten Ergebnisse nachgewiesen werden. Die Ergebnisse des entworfenen Algorithmus zur Tiefenschätzung nach dem *Depth-from-Defocus*-Prinzip zeigen übereinstimmend sowohl in der Simulation als auch mit realen Aufnahmen eine Abhängigkeit vom vorherrschenden Störszenario und von der verwendeten Blendenform. Somit können durch eine Implementierung und Simulation eines Kamera-konzepts wertvolle Erkenntnisse für den Entwurf, den Aufbau und die Optimierung einer neuartigen *Computational Camera* gewonnen werden.

## 6.1 Ausblick

Mit der programmierbaren Apertur kann der aufgebaute Objektivprototyp für verschiedene Problemstellungen angewandt werden. Unmittelbar denkbare Anwendungsgebiete, die in dieser Arbeit nicht untersucht wurden, sind Bildrestauration bei Defokussierung und eine sequentielle Lichtfeldaufnahme. In fortführenden Arbeiten sollte die Leistungsfähigkeit des Prototyps in diesen Problemstellungen untersucht werden, um zu bewerten, wie flexibel der Aufbau eingesetzt werden kann.

Das verwendete Flüssigkristalldisplay in der Aperturebene des Objektivprototyps dämpft einfallendes unpolarisiertes Licht entsprechend der gewünschten Durchlässigkeit. Mit der Blende wird also die Intensität des einfallenden Lichts codiert. Durch Entfernen des objektseitigen Polarisationsfilters lässt sich das Flüssigkristalldisplay als örtlich programmierbares Polarisationsfilter nutzen. Somit könnte die Polarisationsrichtung des einfallenden Lichts zur Codierung genutzt werden. Es wäre zu untersuchen, inwiefern sich ein derartiges System in Anwendungen, bei denen polarisiertes Licht auftritt, wie beispielsweise der Oberflächeninspektion, einsetzen lässt.

Durch die Eigenschaft, dass das abbildende System eine Entwurfsgröße der Signalverarbeitungskette ist, wird der Vergleich unterschiedlicher Systeme des *Computational Imaging* erschwert. In der klassischen Bildverarbeitung können zum Vergleich verschiedener Lösungsansätze einer Problemstellung Eingangs- und Referenzdaten zwischen Forschergruppen ausgetauscht werden. Somit besteht für alle Ansätze eine objektive Bewertungsgrundlage. Beim *Computational Imaging* ist die Eingangsgröße die von einer Szene ausgesendete Strahldichte. Um zwei Systeme objektiv miteinander vergleichen zu können, müssen sie anhand derselben Strahldichte untersucht werden, was nur mit derselben Szene möglich ist. Eine Simulationsumgebung, in der neue Ansätze implementiert werden können, könnte helfen, diese Lücke zu schließen. Mit standardisierten Simulationsparametern und Szenenkonfigurationen könnte sichergestellt werden, dass Kamerasysteme in der Simulation mit identischen Eingangssignalen erprobt werden. Damit wäre eine Möglichkeit gegeben, unterschiedliche *Computational Cameras* objektiv untereinander zu vergleichen.





# Anhang



# A Übersicht Raytracing-Simulationsumgebung

Die im Rahmen dieser Arbeit entstandene Raytracing-Simulationsumgebung besteht derzeit aus ca. 130 C++-Klassen. In Abbildung A.1 ist der grobe strukturelle Ablauf einer Simulation dargestellt. Größen in Schreibmaschinenschrift bezeichnen dabei Datentypen oder Klassen. Die restlichen Größen stellen abstrakte Klassen und damit Platzhalter dar, für die vom Anwender eine konkrete Klasse ausgewählt werden muss. In den folgenden Abschnitten werden die Funktionen der Module und deren Auswahlmöglichkeiten kurz beschrieben.

## A.1 Frontend

Das Frontend implementiert die Nutzerschnittstelle der Simulationsumgebung. Es lädt eine textbasierte Beschreibung der zu simulierenden Szene und ist nach erfolgreicher Simulation zuständig für das Abspeichern der Ergebnisse. In Tabelle A.1 ist eine Übersicht über die Frontends angegeben.

**Tabelle A.1** Übersicht über Frontends.

<b>Klasse</b>	<b>Beschreibung</b>
<code>BasicFrontend</code>	Kommandozeilenbasierte Nutzerschnittstelle. Besonders geeignet für automatische oder gekriptete Aufrufe der Simulationsumgebung.
<code>GuiFrontend</code>	Grafische Benutzeroberfläche.

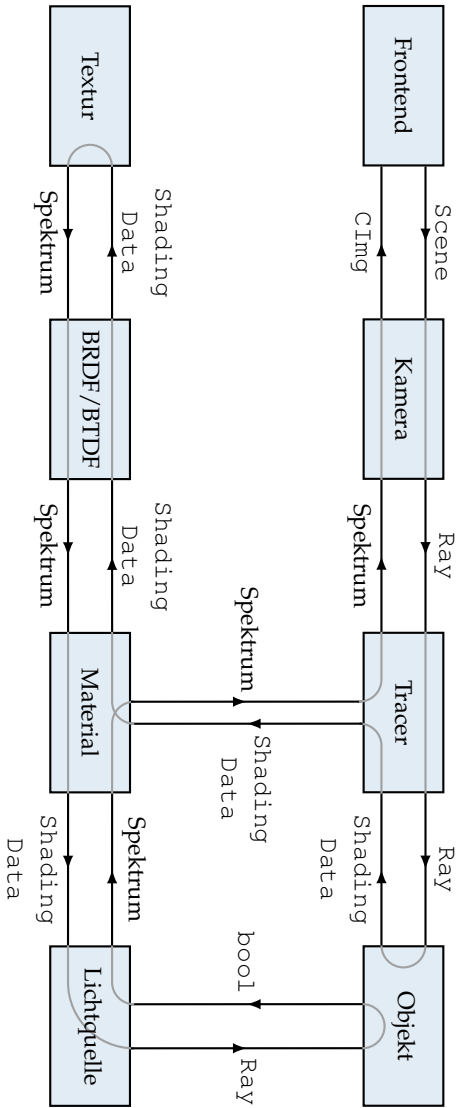


Abbildung A.1 Struktur der Raytracing-Simulationsumgebung.

## A.2 Kamera

Das Kameramodul definiert die Position und die Eigenschaften des Bildsensors und erzeugt die Lichtstrahlen zur Approximation der kontinuierlichen Strahldichte  $L$ . Die Beiträge der Strahlen werden pro Pixel aufsummiert und ggf. mit Rauschen beaufschlagt, um so das Bild zu erzeugen. Jedem Kameratyp können zusätzlich beliebig viele Aperturen an beliebiger Position hinzugefügt werden, deren Form sich frei über ein Pixelbild beliebiger Auflösung einstellen lässt. In Tabelle A.2 sind die derzeit implementierten Kameras angegeben.

**Tabelle A.2** Übersicht über Kameras.

Klasse	Beschreibung
Orthographic	Kamera mit ausschließlich ideal orthogonaler Projektion des Lichts auf den Sensor.
Pinhole	Ideale Lochkamera.
Thinlens	Kamera mit einfacher dünnen Linse. Die Brechung an der Linse wird in paraxialer Näherung nach dem idealen Abbildungsgesetz berechnet.
RealLensCamera	Kamera mit einfacher dicker sphärischer Linse. Die Brechung an der Linse wird anhand des Snellius'schen Brechungsgesetzes an den Linsenoberflächen berechnet. Damit weist diese Kamera geometrische Abbildungsfehler auf.
ChromaticRealLensCamera	Wie RealLensCamera mit wellenlängenabhängigem Brechungsindex des Linsenmaterials. Damit weist diese Kamera zusätzlich zu obiger Kamera auch chromatische Aberrationen auf.
LightFieldCamera	Lichtfeldkamera mit Mikrolinsenarray nach Ng et al. [68].

## A.3 Tracer

Tracer-Objekte bewerkstelligen die Wechselwirkung zwischen Kamera- und Umgebungssimulation. Sie bekommen die Lichtstrahlen des Kameramoduls übergeben und verfolgen diese in der simulierten Szene. Über die Wahl des Tracers kann die Komplexität der Beleuchtungssimulation eingestellt werden, beispielsweise über die Beschränkung der Rekursionstiefe bei reflektierenden Materialien oder die Auswahl zugelassener Beleuchtungsquellen. Nach der Berechnung der Wechselwirkung des Strahls mit der Szene wird die resultierende Farbe an die Kamera zurückgegeben. Durch Rückgabe einer anderen als Farbwert codierten Größe können auch auf einfache Weise Referenzdaten ausgegeben werden. Dazu stehen dem Tracer mit einem `ShadingData`-Objekt alle zur Berechnung des Lichttransports an Oberflächen nötigen Informationen zur Verfügung. In Tabelle A.3 sind die verfügbaren Tracer aufgeführt.

**Tabelle A.3** Übersicht über Tracer.

---

<b>Klasse</b>	<b>Beschreibung</b>
<code>RayCast</code>	Tracer für Punkt- und Umgebungslichtquellen. Es werden lediglich die direkte Beleuchtung und nichtreflektierende Materialien simuliert.
<code>AreaLighting</code>	Tracer zur Simulation räumlich ausgedehnter Lichtquellen.
<code>Whitted</code>	Tracer zur rekursiven Simulation reflektierender Materialien.
<code>WhittedAreaLighting</code>	Kombination des <code>AreaLighting</code> und des <code>Whitted</code> Tracers.
<code>PathTracer</code>	Tracer zur Simulation von globaler, also direkter und indirekter, Beleuchtung.
<code>DepthTracer</code>	Erzeugt Tiefenreferenzdaten durch Rückgabe der Objektentfernung anstelle des reflektierten Farbwerts.

---

Klasse	Beschreibung
SurfaceTracer	Tracer gibt Betrag des Winkels zwischen Strahl und Oberflächennormale codiert als Grauwert an Kamera zurück.
NormalTracer	Tracer gibt Orientierung der Oberflächennormale codiert als RGB-Farbwert als Referenz an Kamera zurück.

## A.4 Objekt

Die zu simulierende Szene wird durch die Menge der in der Welt platzierten Objekte definiert. Jedes Objekt muss bei der Raytracing-Simulation prüfen, ob es vom gerade betrachteten Strahl getroffen wird oder nicht. Im Fall eines Schnittpunkts werden dann die zur Berechnung des Lichttransports an Oberflächen nötigen Informationen in ein `ShadingData`-Objekt geschrieben und an den Tracer zurückgegeben. Tabelle A.4 führt die implementierten Objektklassen sowie die Klassen für Verbundobjekte aus mehreren einzelnen Objekten auf.

**Tabelle A.4** Übersicht über Objekte.

Klasse	Beschreibung
Sphere	Kugelobjekt.
Plane	Unendliche Ebene.
Box	Quaderobjekt.
Square	Ebenes endliches Quadrat.
Disk	Ebene endliche Kreisscheibe.
Triangle	Ebenes endliches Dreieck, das auch zum Aufbau von Polygonmodellen verwendet wird.
Cylinder	Zylinderoberfläche.
Cone	Allgemeinen Kegeloberfläche.

Klasse	Beschreibung
Torus	Torusobjekt.
TriangleMesh	Verbundobjekt aus einzelnen Dreiecken.

## A.5 Material

Jedem Objekt ist eines der in Tabelle A.5 aufgeführten Materialien zugeordnet, das festlegt, welches grundlegende Reflexions- und Transmissionsverhalten, repräsentiert durch verschiedene bidirektionale Reflexanzverteilungsfunktionen (BRDF, engl. *bidirectional reflectance distribution function*) und bidirektionale Transmissionsverteilungsfunktionen (BTDF, engl. *bidirectional transmittance distribution function*), ein Objekt besitzt.

**Tabelle A.5** Übersicht über Materialien.

Klasse	Beschreibung
Matte	Mattes Material mit Lambert'scher Reflexion.
Phong	Empirisches Material mit diffuser und spiegelnder Reflexion der direkten Beleuchtung.
Reflective	Material mit perfekt spiegelnder Reflexion der indirekten Beleuchtung.
Glossy Reflective	Material mit diffus spiegelnder Reflexion der indirekten Beleuchtung.
Transparent	Transparentes Material mit konstanten Reflexions- und Transmissionskoeffizienten.
Dielectric	Dielektrisches Material mit Reflexions- und Transmissionskoeffizienten gemäß den Fresnel'schen Formeln.
Emissive	Licht emittierendes Material für räumlich ausgedehnte Lichtquellen.



## A.6 BRDF/BTDF

Die Klassen der BRDFs definieren allgemein den Zusammenhang zwischen einfallender Strahldichte  $L_i$  und ausfallender Strahldichte  $L_o$  und die BTDFs zwischen einfallender Strahldichte  $L_i$  und transmittierter Strahldichte  $L_t$ . Die implementierten BRDFs sind in Tabelle A.6 und die BTDFs in Tabelle A.7 aufgelistet.

**Tabelle A.6** Übersicht über BRDFs.

Klasse	Beschreibung
Lambertian	Lambert'sche Reflexion.
GlossySpecular	Diffus spiegelnde Reflexion.
PerfectSpecular	Perfekt spiegelnde Reflexion.
FresnelReflector	Reflexion gemäß Fresnel'schen Formeln.

**Tabelle A.7** Übersicht über BTDFs.

Klasse	Beschreibung
Perfect Transmitter	BTDF mit konstantem Transmissionskoeffizienten.
Fresnel Transmitter	Transmission gemäß den Fresnel'schen Formeln.

## A.7 Textur

Texturen implementieren lokale Variationen des von einem Objekt reflektierten Spektrums. Damit verleihen Sie Objektflächen Farben und Strukturen. Tabelle A.8 listet die wählbaren Texturen auf.

Tabelle A.8 Übersicht über Texturen.

Klasse	Beschreibung
ConstantColor	Einfarbige Textur.
Checker3D	Schachbretttextur, ausgerichtet an globalen 3D-Weltkoordinaten.
Checker2D	Schachbretttextur, ausgerichtet an lokalen 2D-Objektkoordinaten.
Grid3D	Gittertextur, ausgerichtet an globalen 3D-Weltkoordinaten.
Grid2D	Gittertextur, ausgerichtet an lokalen 2D-Objektkoordinaten.
Sine3D	Sinustextur, ausgerichtet an globalen 3D-Weltkoordinaten.
Sine2D	Sinustextur, ausgerichtet an lokalen 2D-Objektkoordinaten.
Nearest Neighbour Noise	Prozedural nach der <i>Fractional Brownian Motion</i> erzeugte zufällige Rauschtextur. Farben zwischen den Stützstellen der Rauschfunktion werden mithilfe der Nächster-Nachbar-Interpolation bestimmt.
LinearNoise	Prozedural nach der <i>Fractional Brownian Motion</i> erzeugte zufällige Rauschtextur. Farben zwischen den Stützstellen der Rauschfunktion werden linear interpoliert.
CubicNoise	Prozedural nach der <i>Fractional Brownian Motion</i> erzeugte zufällige Rauschtextur. Farben zwischen den Stützstellen der Rauschfunktion werden kubisch interpoliert.
ImageTexture	Textur aus Pixelbild.

---

## A.8 Lichtquelle

Nachdem das Reflexions- und Transmissionsverhalten an den von der Kamera sichtbaren Objektoberflächen, u. U. rekursiv, berechnet wurde, wird noch die von den Lichtquellen emittierte Strahldichte  $L_Q$  benötigt, um letztlich die Strahldichte  $L_C$  in der Kamera zu bestimmen. Bei der Simulation der Beleuchtung muss geprüft werden, ob eine Lichtquelle ausgehend vom gerade betrachteten Punkt auf der Objektoberfläche überhaupt sichtbar oder durch ein anderes Objekt abgeschattet ist. Daher müssen, wie schon bei der Bestimmung der von der Kamera aus sichtbaren Objekte, alle Objekte mit einem Strahl auf Sichtbarkeit geprüft werden. Es stehen die in Tabelle A.9 aufgelisteten Lichtquellen zur Verfügung.

**Tabelle A.9** Übersicht über Lichtquellen.

Klasse	Beschreibung
Ambient	Global vorhandene, richtungsunabhängige Beleuchtung.
Directional	Global vorhandene, richtungsabhängige Beleuchtung.
Ambient Occluder	Aus dem Unendlichen emittierte, von Objekten abgeschattete, richtungsunabhängige Beleuchtung.
PointLight	Ideale Punktlichtquelle.
AreaLight	Räumlich ausgedehnte Lichtquelle.

## A.9 Spektrum

Die Repräsentation von Farbe in der Simulation kann frei gewählt bzw. sogar bei der Definition einer Szene vermischt werden. Nötige Umrechnungen zwischen den verschiedenen Repräsentationen erfolgen stets über das CIE-Normvalenzsystem. Es können die in Tabelle A.10 aufgeführten spektralen Repräsentierungen ausgewählt werden.

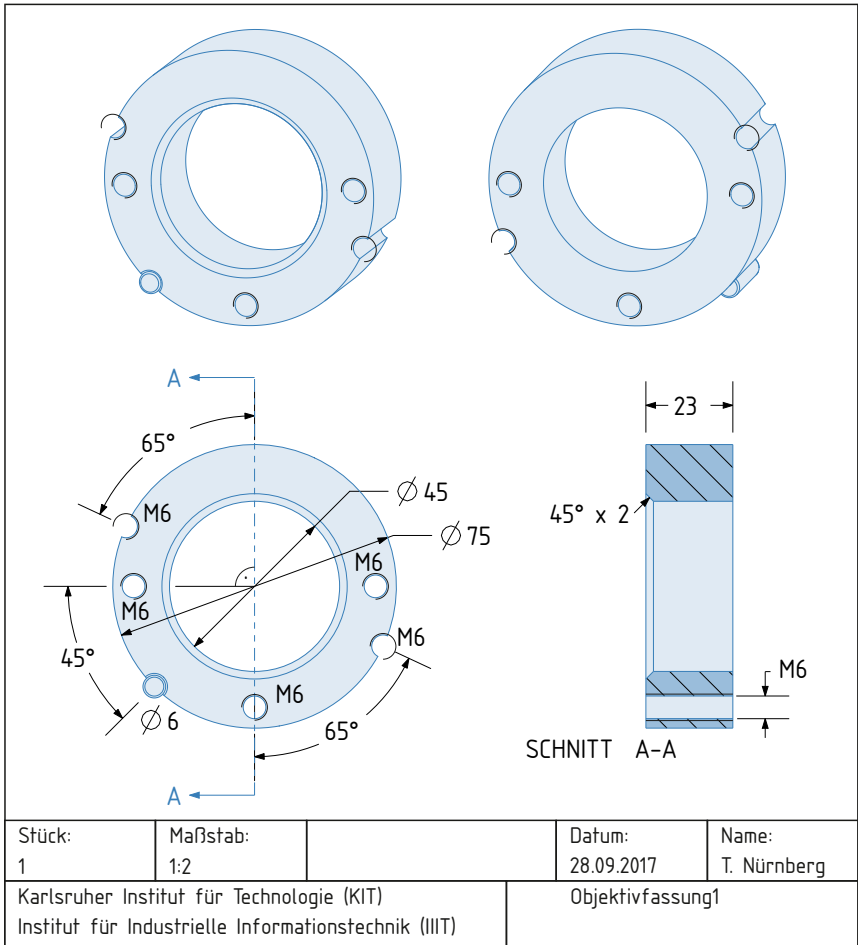
**Tabelle A.10** Übersicht über spektrale Repräsentierungen.

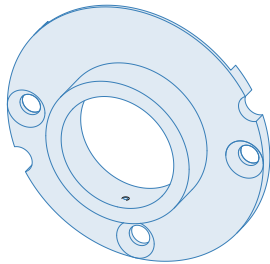
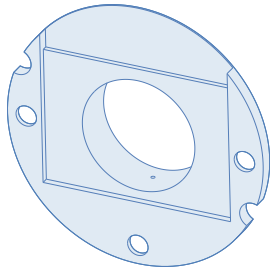
---

<b>Klasse</b>	<b>Beschreibung</b>
Monochromatic	Monochromatisches Licht.
RgbSpectrum	Repräsentation von Farben im RGB-Farbraum.
Sampled Spectrum	Repräsentation von Farben über ein mit 30 Stützstellen abgetastetes Spektrum des sichtbaren Wellenlängenbereichs von 400 nm bis 700 nm. Bereich und Auflösung durch Änderung von Template-Parametern anpassbar.

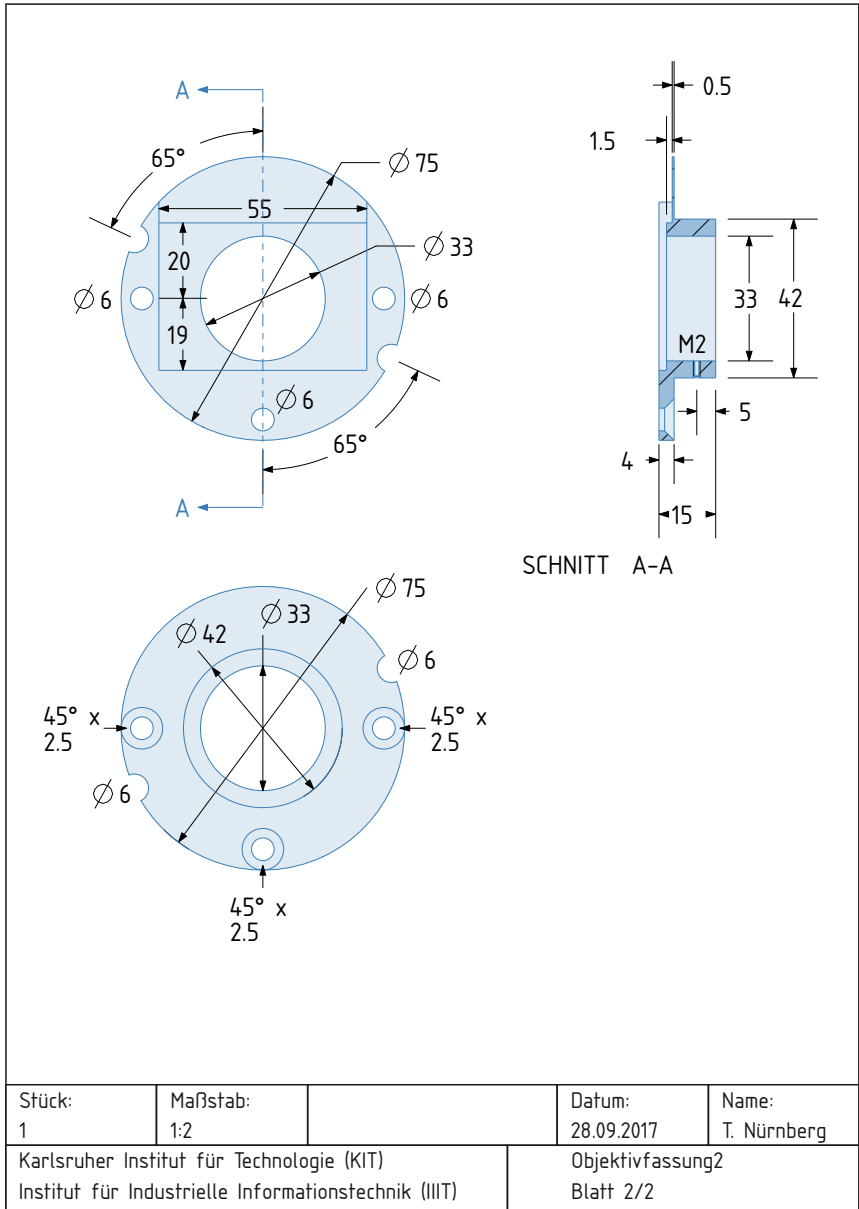
---

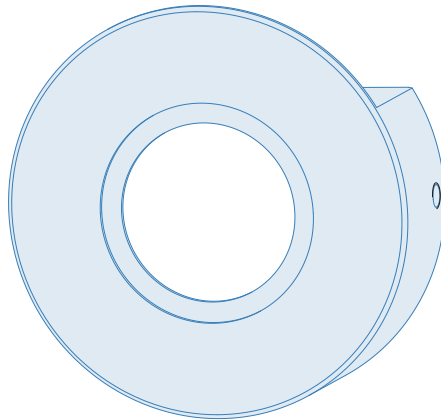
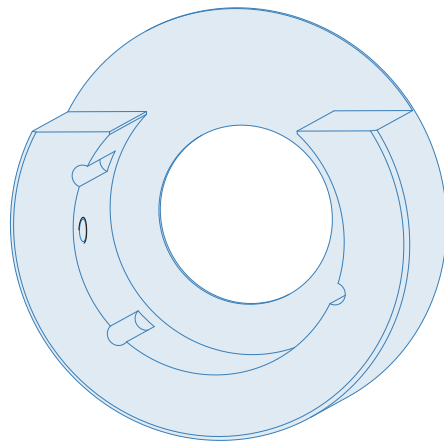
## B Modifiziertes Objektiv





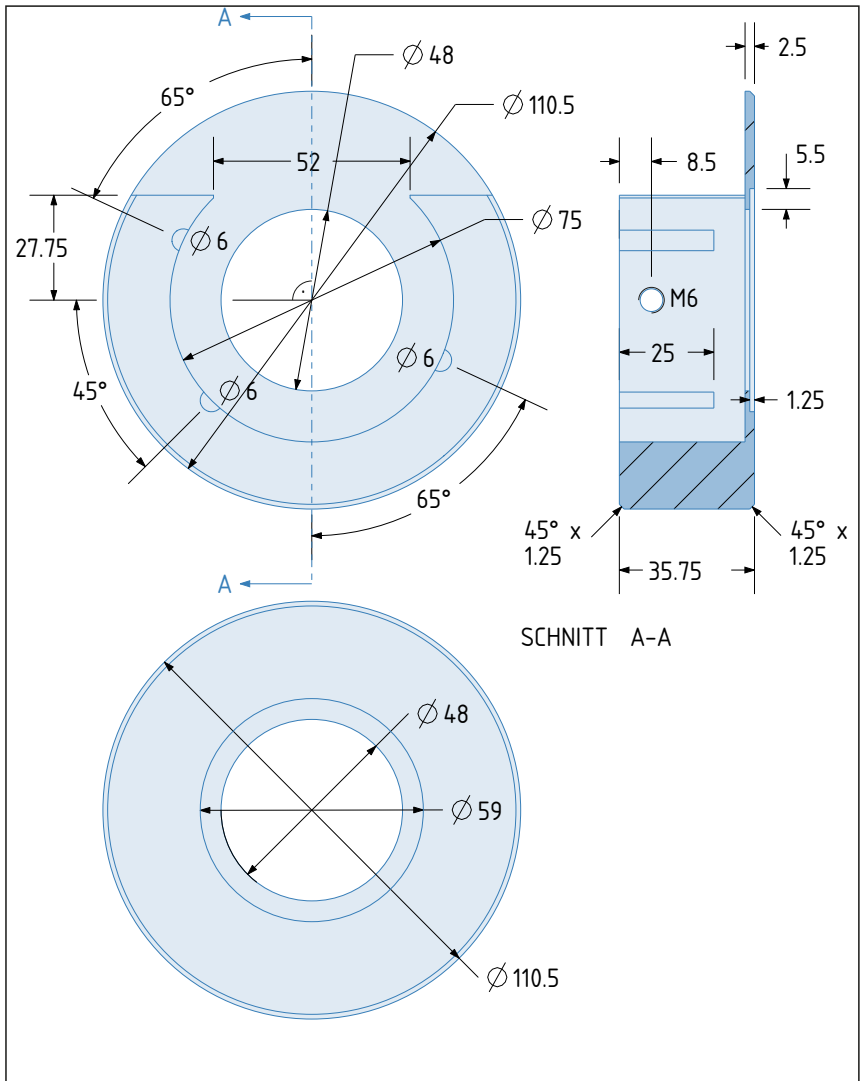
Stück: 1	Maßstab: 1:2		Datum: 28.09.2017	Name: T. Nürnberg
Karlsruher Institut für Technologie (KIT) Institut für Industrielle Informationstechnik (IIT)			Objektivfassung2 Blatt 1/2	





Stück: 1	Maßstab: 1:2		Datum: 28.09.2017	Name: T. Nürnberg
Karlsruher Institut für Technologie (KIT) Institut für Industrielle Informationstechnik (IIT)			Objektivfassung3 Blatt 1/2	





SCHNITT A-A

Stück: 1	Maßstab: 1:2		Datum: 28.09.2017	Name: T. Nürnberg
Karlsruher Institut für Technologie (KIT) Institut für Industrielle Informationstechnik (IIIT)			Objektivfassung3 Blatt 2/2	



# Literaturverzeichnis

- [1] **Accorsi, Roberto** und **Richard C. Lanza**. *High-Sensitivity Dynamic Coded Aperture Imaging*. In: *IEEE Nuclear Science Symposium Conference Record*. Bd. 3. 2003, S. 1833–1837.
- [2] **Adams, Andrew, Eino-Ville Talvala, Sung Hee Park, David E. Jacobs, Boris Ajdin, Natasha Gelfand, Jennifer Dolson, Daniel Vaquero, Jongmin Baek, Marius Tico, Hendrik P. A. Lensch, Wojciech Matusik, Kari Pulli, Mark Horowitz** und **Marc Levoy**. *The Frankencamera: An Experimental Platform for Computational Photography*. In: *ACM Transactions on Graphics* 29.4 (2010), 29:1–29:12.
- [3] **Adelson, Edward H.** und **John Y. A. Wang**. *Single Lens Stereo with a Plenoptic Camera*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), S. 99–106.
- [4] **Agu, Emmanuel** und **Francis S. Hill**. *A Simple Method for Ray Tracing Diffraction*. In: *International Conference on Computational Science and Its Applications (ICCSA 2003)*. Hrsg. von **Kumar, Vipin, Marina L. Gavrilova, Chih Jeng Kenneth Tan** und **Pierre L'Ecuyer**. Montreal, Canada, Springer Berlin Heidelberg, 2003, S. 336–345.
- [5] **Amari, Yasufumi** und **Edward H. Adelson**. *Single-eye range estimation by using displaced apertures with color filters*. In: *Proceedings of the International Conference on Industrial Electronics, Control, Instrumentation, and Automation*. Bd. 3. 1992, S. 1588–1592.
- [6] **Amaya, Dafne, Daniel Actis, Gonzalo Rumi** und **Alberto Lencina**. *Least squares method for liquid crystal display characterization*. In: *Applied Optics* 56.5 (2017), S. 1438–1446.
- [7] **Apple Inc.**, *Face ID Security*. Techn. Ber. 2017.
- [8] **Bando, Yosuke, Bing-Yu Chen** und **Tomoyuki Nishita**. *Extracting Depth and Matte using a Color-Filtered Aperture*. In: *ACM SIGGRAPH Asia 2008 papers*. Singapore, ACM, 2008, 134:1–134:9.

- [9] **Ben-Ezra, Moshe, Assaf Zomet** und **Shree K. Nayar**. *Video Super-Resolution Using Controlled Subpixel Detector Shifts*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.6 (2005), S. 977–987.
- [10] **Beyerer, Jürgen, Fernando Puente León** und **Christian Frese**. *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. 2. Aufl. Berlin Heidelberg, Springer, 2016.
- [11] **Bickford, Jen**. *Lighting Software Tools: The latest features from calculation and rendering programs*. [http://www.archlighting.com/technology/lighting-software-tools\\_o](http://www.archlighting.com/technology/lighting-software-tools_o). Architectural Lighting Magazine, 2008, Abrufdatum: Februar 2018.
- [12] **Birch, Gabriel C., Charles F. LaCasse, Jaclynn J. Stubbs, Amber L. Dagel** und **Jon Bradley**. *Tolerance Analysis Through Computational Imaging Simulations*. In: *Optical Design and Fabrication (Freeform, IODC, OFT)*. OSA Technical Digest (online). Denver, CO, USA, Optical Society of America, 2017.
- [13] **Boykov, Yuri, Olga Veksler** und **Ramin Zabih**. *Fast Approximate Energy Minimization via Graph Cuts*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.11 (2001), S. 1222–1239.
- [14] **Brady, D. J.** und **M. E. Gehm**. *Compressive imaging spectrometers using coded apertures*. In: *Defense and Security Symposium*. Bd. 6264. SPIE, 2006, 62460A-1–62460A-9.
- [15] **Butler, Daniel J., Jonas Wulff, Garrett B. Stanley** und **Michael J. Black**. *A Naturalistic Open Source Movie for Optical Flow Evaluation*. In: *European Conference on Computer Vision (ECCV)*. Florence, Italy, Springer Berlin Heidelberg, 2012, S. 611–625.
- [16] **Cadena, Cesar, Luca Carlone, Henry Carrillo, Latif. Yasir, Davide Scaramuzza, José Neira, Ian Reid** und **John J. Leonard**. *Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age*. In: *IEEE Transactions on Robotics* 32.6 (2016), S. 1309–1332.
- [17] **Chen, Guangyong, Fengyuan Zhu** und **Pheng Ann Heng**. *An Efficient Statistical Method for Image Noise Level Estimation*. In: *IEEE International Conference on Computer Vision (ICCV)*. 2015, S. 477–485.
- [18] **Cheng, Hong, Lu Yang** und **Zicheng Liu**. *Survey on 3D Hand Gesture Recognition*. In: *IEEE Transactions on Circuits and Systems for Video Technology* 26.9 (2016), S. 1659–1673.

- [19] **Cho, Taeg Sang, Anat Levin, Frédo Durand** und **William T. Freeman**. *Motion blur removal with orthogonal parabolic exposures*. In: *IEEE International Conference on Computational Photography (ICCP)*. Cambridge, MA, USA, 2010, S. 1–8.
- [20] **Cossairt, Oliver, Mohit Gupta** und **Shree K. Nayar**. *When Does Computational Imaging Improve Performance?* In: *IEEE Transactions on Image Processing* 22.2 (2013), S. 447–458.
- [21] **Cossairt, Oliver** und **Shree K. Nayar**. *Spectral Focal Sweep: Extended Depth of Field from Chromatic Aberrations*. In: *IEEE International Conference on Computational Photography (ICCP)*. 2010, S. 1–8.
- [22] **Darrell, Trevor** und **Kwangyoen Wohn**. *Pyramid Based Depth from Focus*. In: *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '88)*. 1988, S. 504–509.
- [23] **Durrett, Rick**. *Elementary Probability for Applications*. Cambridge, Cambridge University Press, 2009.
- [24] **European Machine Vision Association, EMVA Standard 1288 – Standard for Characterization of Image Sensors and Cameras**. [www.emva.org](http://www.emva.org). 2010.
- [25] **Farinella, Giovanni Maria, Sebastiano Battiato** und **Roberto Cipolla**, Hrsg. *Advanced Topics on Computer Vision*. Advances in Computer Vision and Pattern Recognition. Springer-Verlag London, 2013.
- [26] **Fujikake, Hideo, Kuniharu Takizawa, Tahito Aida, Toshihiro Negishi** und **Michio Kobayashi**. *Video Camera System Using Liquid-Crystal Polarizing Filter to Reduce Reflected Light*. In: *IEEE Transactions on Broadcasting* 44.4 (1998), S. 419–426.
- [27] **Glassner, Andrew S.** *An Introduction to Ray tracing*. London, Morgan Kaufmann, 1989.
- [28] **Glatard, Tristan et al.** *A Virtual Imaging Platform for Multi-Modality Medical Image Simulation*. In: *IEEE Transactions on Medical Imaging* 32.1 (2013), S. 110–118.
- [29] **Greivenkamp, John E.** *Field Guide to Geometrical Optics*. Bellingham, WA, SPIE Optical Engineering Press, 2004.
- [30] **Grote, Alexander** und **Erwin Schwab**. *Entwurfsmethode für 3D-Bildverarbeitungssysteme in der Anlagenautomatisierung*. In: *tm – Technisches Messen* 82.9 (2015), S. 451–462.
- [31] **Gruna, Robin**. *Beleuchtungsverfahren zur problemspezifischen Bildgewinnung für die automatische Sichtprüfung*. Dissertation. KIT Scientific Publishing, 2013.

- [32] **Hayes, Brian.** *Computational Photography*. In: *American Scientist* 96.2 (2008), S. 94–99.
- [33] **Healey, Glenn E. und Raghava Kondepudy.** *Radiometric CCD Camera Calibration and Noise Estimation*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.3 (1994), S. 267–276.
- [34] **Hughes, John F., Andries van Dam, Morgan McGuire, David F. Sklar, James D. Foley, Steven K. Feiner und Kurt Akeley.** *Computer graphics: principles and practice*. 3. Aufl. Boston, MA, USA, Addison-Wesley Professional, 2013.
- [35] **Ikeuchi, Katsushi und Jean-Christophe Robert.** *Modeling Sensor Detectability with the VANTAGE Geometric/Sensor Modeler*. In: *IEEE Transactions on Robotics and Automation* 7.6 (1991), S. 771–784.
- [36] **Irgenfried, Stephan, Frank Dittrich und Heinz Wörn.** *Realization and evaluation of image processing tasks based on synthetic sensor data: 2 use cases*. In: *Forum Bildverarbeitung*. Regensburg, 2014, S. 35–46.
- [37] **Jacobs, David E., Jongmin Baek und Marc Levoy.** *Focal Stack Compositing for Depth of Field Control*. Techn. Ber. Stanford Computer Graphics Laboratory, 2012.
- [38] **Jähne, Bernd.** *Digitale Bildverarbeitung und Bildgewinnung*. 7. Aufl. Springer-Verlag Berlin Heidelberg, 2012.
- [39] **Jähne, Bernd.** *Der Standard EMVA 1288 zur Charakterisierung von Kameras und Bildsensoren*. In: *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik. Beiträge der Oldenburger 3D-Tage*. Oldenburg, 2013, S. 388–399.
- [40] **Jannesick, James R., Kenneth P. Klaasen und Thom Elliott.** *Charge-coupled-device charge-collection efficiency and the photon-transfer technique*. In: *Optical Engineering* 26.10 (1987), S. 972–980.
- [41] **Krotkov, Eric.** *Focusing*. In: *International Journal of Computer Vision* 1.3 (1988), S. 223–237.
- [42] **Kullback, S. und R. A. Leibler.** *On Information and Sufficiency*. In: *The Annals of Mathematical Statistics* 22.1 (1951), S. 79–86.
- [43] **Kuthirummal, Sujit, Hajime Nagahara, Changyin Zhou und Shree K. Nayar.** *Flexible Depth of Field Photography*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.1 (2011), S. 58.
- [44] **Landman, Douglas.** *Mask-based Light Field Capture and Display*. Dissertation. Brown University, Providence, RI, USA, 2011.

- [45] **Lee, Eunsung, Wonseok Kang, Sangjin Kim und Joonki Paik.** *Color shift model-based image enhancement for digital multifocusing based on a multiple color-filter aperture camera.* In: *IEEE Transactions on Consumer Electronics* 56.2 (2010), S. 317–323.
- [46] **Lee, Seungwon, Junghyun Lee, Monson H. Hayes, Aggelos K. Katsaggelos und Paik Joonki.** *Single camera-based full depth map estimation using color shifting property of a multiple color-filter aperture.* In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2012, S. 801–804.
- [47] **Leobacher, Gunther und Friedrich Pillichshammer.** *Introduction to Quasi-Monte Carlo Integration and Applications.* Springer International Publishing, 2014.
- [48] **Levin, Anat.** *Analyzing Depth from Coded Aperture sets.* In: *Proceedings of the 11th European conference on Computer vision: Part I. ECCV'10.* Berlin, Heidelberg, Springer-Verlag, 2010, S. 214–227.
- [49] **Levin, Anat, Rob Fergus, Frédo Durand und William T. Freeman.** *Image and Depth from a Conventional Camera with a Coded Aperture.* In: *ACM Transactions on Graphics* 26.3 (2007).
- [50] **Levin, Anat, Samuel W. Hasinoff, Paul Green, Frédo Durand und William T. Freeman.** *4D frequency analysis of computational cameras for depth of field extension.* In: *ACM SIGGRAPH 2009 papers. SIGGRAPH '09.* New York, NY, USA, ACM, 2009, 97:1–97:14.
- [51] **Liang, Chia-Kai, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu und Homer H. Chen.** *Programmable Aperture Photography: Multiplexed Light Field Acquisition.* In: *ACM Transactions on Graphics* 27.3 (2008), 55:1–55:10.
- [52] **Lin, Jianhua.** *Divergence Measures Based on the Shannon Entropy.* In: *IEEE Transactions on Information Theory* 37.1 (1991), S. 145–151.
- [53] **Liutkus, Antoine, David Martina, Sébastien Popoff, Gilles Chardon, Ori Katz, Geoffroy Lerosey, Sylvain Gigan, Laurent Daudet und Igor Carron.** *Imaging With Nature: Compressive Imaging Using a Multiply Scattering Medium.* In: *Scientific Reports* 4 (2014), S. 5552-1–5552-7.
- [54] **Lloyd, Gareth A. und Steven J. Sasson.** *Electronic still camera.* United States Patent 4,131,919. 1978.
- [55] **MacAllister, David F.,** Hrsg. *Stereo Computer Graphics and Other True 3D Technologies.* 1. print. Princeton series in computer science. Princeton, NJ, Princeton Univ. Press, 1993.

- [56] **Maik, Vivek, Cho Dohee, Kim Sangjin, Donghwan Har und Paik Joonki.** *Computational filter-aperture approach for single-view multi-focusing.* In: *16th IEEE International Conference on Image Processing (ICIP).* 2009, S. 1541–1544.
- [57] **Malik, Aamir Saeed und Tae-Sun Choi.** *A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise.* In: *Pattern Recognition* 41.7 (2008), S. 2200–2225.
- [58] **Malitson, I. H.** *Interspecimen Comparison of the Refractive Index of Fused Silica.* In: *Journal of the Optical Society of America* 55.10 (1965), S. 1205–1209.
- [59] **Mannami, Hidetoshi, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo und Yasushi Yagi.** *Adaptive dynamic range camera with reflective liquid crystal.* In: *Journal of Visual Communication and Image Representation* 18.5 (2007), S. 359–365.
- [60] **Mannami, Hidetoshi, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo und Yasushi Yagi.** *High Dynamic Range Camera using Reflective Liquid Crystal.* In: *IEEE 11th International Conference on Computer Vision (ICCV).* 2007, S. 1–8.
- [61] **Meister, Stephan Nicolas Robert.** *On Creating Reference Data for Performance Analysis in Image Processing.* Dissertation. IWR, Fakultät für Physik und Astronomie, Universität Heidelberg, 2014.
- [62] **Nagahara, Hajime, Changyin Zhou, Takuya Watanabe, Hiroshi Ishiguro und Shree K. Nayar.** *Programmable Aperture Camera Using LCoS.* In: *Computer Vision – ECCV 2010.* Hrsg. von **Daniilidis, Kostas, Petros Maragos und Nikos Paragios.** Bd. 6316. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010. Kap. 25, S. 337–350.
- [63] **Nayar, Shree K.** *Computational Cameras: Redefining the Image.* In: *Computer* 39.8 (2006), S. 30–38.
- [64] **Nayar, Shree K.** *Computational Cameras Approaches, Benefits and Limits.* Technical reports. Computer Science, 2011.
- [65] **Nayar, Shree K. und Vlad Branzoi.** *Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time.* In: *IEEE International Conference on Computer Vision (ICCV).* Bd. 2. 2003, S. 1168–1175.
- [66] **Neira, José und Juan D. Tardós.** *Data Association in Stochastic Mapping Using the Joint Compatibility Test.* In: *IEEE Transactions on Robotics and Automation* 17.6 (2001), S. 890–897.
- [67] **Ng, Ren.** *Fourier slice photography.* In: *ACM SIGGRAPH 2005 Papers.* Los Angeles, California, ACM, 2005, S. 735–744.



- [68] **Ng, Ren, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz** und **Pat Hanrahan**. *Light Field Photography with a Hand-Held Plenoptic Camera*. In: *Stanford University Computer Science Tech Report CSTR 2005-02 2* (2005), S. 1–11.
- [69] **Nilsson, Jim, Mikael Ericsson** und **Fredrik Danielsson**. *Virtual Machine Vision in Computer Aided Robotics*. In: *IEEE Conference on Emerging Technologies & Factory Automation*. Mallorca, Spain, 2009, S. 1–8.
- [70] **Oh, Se Baek, Sriram Kashyap, Rohit Garg, Sharat Chandran** und **Ramesh Raskar**. *Rendering Wave Effects with Augmented Light Field*. In: *Computer Graphics Forum* 29.2 (2010), S. 507–516.
- [71] **Osterman, Mark**. *The Technical Evolution of Photography in the 19th Century*. In: *The Focal Encyclopedia of Photography (Fourth Edition)*. Boston, Focal Press, 2007, S. 27–36.
- [72] **Papoulis, Athanasios**. *Signal Analysis*. McGraw-Hill, Inc., 1977.
- [73] **Pentland, Alex Paul**. *A New Sense for Depth of Field*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9.4* (1987), S. 523–531.
- [74] **Pharr, Matt** und **Greg Humphreys**. *Physically Based Rendering: From Theory to Implementation*. 2. Aufl. Morgan Kaufmann Publishers Inc., 2010.
- [75] **Phong, Bui Tuong**. *Illumination for Computer Generated Pictures*. In: *Communications of the ACM* 18.6 (1975), S. 311–317.
- [76] **Pixera Corporation**, *Diractor<sup>TM</sup> Technology*. <http://www.pixera.com/technology/diractor.html>. Abrufdatum: Oktober 2017.
- [77] **Press, William H., Saul A. Teukolsky, William T. Vetterling** und **Brian P. Flannery**. *Numerical Recipes: The Art of Scientific Computing*. 3. Aufl. New York, NY, USA, Cambridge University Press, 2007.
- [78] **Puente León, Fernando**. *Messtechnik: Systemtheorie für Ingenieure und Informatiker*. 10. Aufl. Springer Berlin Heidelberg, 2015.
- [79] **Puente León, Fernando** und **Holger Jäkel**. *Signale und Systeme*. 6. Aufl. Berlin, Boston, De Gruyter Oldenbourg, 2015.
- [80] **Raczkowsky, J.** und **K. H. Mittenbuehler**. *Simulation of Cameras in Robot Applications*. In: *IEEE Computer Graphics and Applications* 9.1 (1989), S. 16–25.
- [81] **Raskar, Ramesh**. *Computational Photography: Epsilon to Coded Photography*. In: *Emerging Trends in Visual Computing*. Hrsg. von **Nielsen, Frank**. Bd. 5416. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009. Kap. 10, S. 238–253.

- [82] **Raskar, Ramesh, Amit Agrawal** und **Jack Tumblin**. *Coded Exposure Photography: Motion Deblurring using Fluttered Shutter*. In: *ACM Transactions on Graphics* 25.3 (2006), S. 795–804.
- [83] **Reiner, Jacek**. *Rendering for machine vision prototyping*. In: *Optical Systems Design*. Bd. 7100. SPIE, 2008, S. 1–8.
- [84] **Retzlaff, Max-Gerd, Matthias Richter, Thomas Längle, Jürgen Beyerer** und **Carsten Dachsbacher**. *Combining synthetic image acquisition and machine learning: accelerated design and deployment of sorting systems*. In: *Forum Bildverarbeitung*. Karlsruhe, 2016, S. 49–61.
- [85] **Schaaf, A. van der** und **J. H. van Hateren**. *Modelling the Power Spectra of Natural Images: Statistics and Information*. In: *Vision Research* 36.17 (1996), S. 2759–2770.
- [86] **Scharstein, Daniel** und **Richard Szeliski**. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. In: *International Journal of Computer Vision* 47.1-3 (2002), S. 7–42.
- [87] **Scharstein, Daniel, Richard Szeliski** und **Heiko Hirschmüller**. *Middlebury Stereo Vision Page*. <http://vision.middlebury.edu/stereo>. Middlebury College, Vermont, USA, Abrufdatum: November 2017.
- [88] **Scharstein, Daniel, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang** und **Porter Westling**. *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*. In: *36th German Conference on Pattern Recognition (GCPR)*. Münster, 2014, S. 31–42.
- [89] **Schechner, Yaov Y.** und **Nahum Kiryati**. *Depth from Defocus vs. Stereo: How Different Really are They?* In: *Fourteenth International Conference on Pattern Recognition. Proceedings*. Bd. 2. 1998, S. 1784–1786.
- [90] **Sellmeier, Wolfgang von**. *Zur Erklärung der abnormen Farbenfolge im Spectrum einiger Substanzen*. In: *Annalen der Physik* 219.6 (1871), S. 272–282.
- [91] **Suffern, Kevin G.** *Ray Tracing from the Ground Up*. Wellesley, Massachusetts, A K Peters, Ltd., 2007.
- [92] **Suo, JinLi, XiangYang Ji** und **QiongHai Dai**. *An overview of computational photography*. In: *Science China Information Sciences* 55.6 (2012), S. 1229–1248.
- [93] **Suwajanakorn, Supasorn, Carlos Hernandez** und **Steven M. Seitz**. *Depth from Focus with Your Mobile Phone*. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, S. 3497–3506.

- [94] **Tabary, Joachim, Simon Marache-Francisco, Sébastien Valette, W. Paul Segars und Lartizien Carole.** *Realistic X-Ray CT Simulation of the XCAT Phantom with SINDBAD.* In: *IEEE Nuclear Science Symposium Conference Record (NSS/MIC).* Orlando, FL, USA, 2009, S. 3980–3983.
- [95] **Terzopoulos, Demetri.** *Regularization of Inverse Visual Problems Involving Discontinuities.* In: *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8.4* (1986), S. 413–424.
- [96] **Torralba, Antonio und Aude Oliva.** *Statistics of natural image categories.* In: *Network: Computation in Neural Systems* 14.3 (2003), S. 391–412.
- [97] **Treibitz, Tali und Yoav Y. Schechner.** *Polarization: Beneficial for Visibility Enhancement?* In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2009, S. 525–532.
- [98] **Vater, Sebastian und Fernando Puente León.** *Monokulare Kopfposenschätzung basierend auf dem optischen Fluss und der Verfolgung stabiler Merkmale.* In: *tm – Technisches Messen* 84.7-8 (2017), S. 525–534.
- [99] **Vedel, Mathieu, Sebastien Breugnot und Nick Lechocinski.** *Full Stokes polarization imaging camera.* In: *Proceedings of SPIE 8160 Polarization Science and Remote Sensing V.* Bd. 8160. San Diego, California, USA, 2011, S. 1–13.
- [100] **Veeraraghavan, Ashok, Ramesh Raskar, Amit Agrawal, Ankit Mohan und Jack Tumblin.** *Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing.* In: *ACM Transactions on Graphics* 26.3 (2007), S. 69-1–69-12.
- [101] **Vogel, Julia und Bernt Schiele.** *Semantic Modeling of Natural Scenes for Content-Based Image Retrieval.* In: *International Journal of Computer Vision* 72.2 (2007), S. 133–157.
- [102] **Wach, Hans und Edward R. Dowski.** *Noise modeling for design and simulation of computational imaging systems.* In: *Defense and Security.* Bd. 5438. SPIE, 2004, S. 159–170.
- [103] **Wald, Ingo, Andreas Dietrich, Carsten Benthin, Alexander Efremov, Tim Dahmen, Johannes Günther, Vlastimil Havran, Hans-Peter Seidel und Philipp Slusallek.** *Applying Ray Tracing for Virtual Reality and Industrial Design.* In: *IEEE Symposium on Interactive Ray Tracing.* Salt Lake City, UT, USA, 2006, S. 177–185.
- [104] **Werling, Stefan Bruno.** *Deflektometrie zur automatischen Sichtprüfung und Rekonstruktion spiegelnder Oberflächen.* Dissertation. KIT Scientific Publishing, 2011.

- [105] **Wetzstein, Gordon, Ivo Ihrke, Douglas Lanman und Wolfgang Heidrich.** *State of the Art in Computational Plenoptic Imaging*. In: *STAR Proceedings of Eurographics*. 2011, S. 25–48.
- [106] **Yang, Ding und Liu Cheng.** *Design of light utilization model in interior design*. In: *International Conference on Education, Management and Computing Technology (ICEMCT)*. Tianjin, China, 2015, S. 1234–1237.
- [107] **Ye, Mao, Qing Zhang, Liang Wang, Jiejie Zhu, Ruigang Yang und Jürgen Gall.** *A Survey on Human Motion Analysis from Depth Data*. In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications: Dagstuhl 2012 Seminar on Time-of-Flight Imaging and GCPR 2013 Workshop on Imaging New Modalities*. Hrsg. von **Grzegorzek, Marcin, Christian Theobalt, Reinhard Koch und Andreas Kolb**. Springer Berlin Heidelberg, 2013, S. 149–187.
- [108] **Zhang, Zhengyou.** *Microsoft Kinect Sensor and Its Effect*. In: *IEEE MultiMedia* 19.2 (2012), S. 4–10.
- [109] **Zhou, Changyin, Oliver Cossairt und Shree K. Nayar.** *Depth from Diffusion*. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2010, S. 1110–1117.
- [110] **Zhou, Changyin, Stephen Lin und Shree K. Nayar.** *Coded aperture pairs for depth from defocus*. In: *IEEE 12th International Conference on Computer Vision*. 2009, S. 325–332.
- [111] **Zhou, Changyin und Shree K. Nayar.** *What are good apertures for defocus deblurring?* In: *IEEE International Conference on Computational Photography (ICCP)*. 2009, S. 1–8.
- [112] **Zhou, Changyin und Shree K. Nayar.** *Computational Cameras: Convergence of Optics and Processing*. In: *IEEE Transactions on Image Processing* 20.12 (2011), S. 3322–3340.
- [113] **Zhu, Hao, Ka-Veng Yuen, Lyudmila Mihaylova und Henry Leung.** *Overview of Environment Perception for Intelligent Vehicles*. In: *IEEE Transactions on Intelligent Transportation Systems* 18.10 (2017), S. 2584–2601.
- [114] **Zhuo, Shaojie und Terence Sim.** *On the Recovery of Depth from a Single Defocused Image*. In: *Computer Analysis of Images and Patterns*. Hrsg. von **Jiang, Xiaoyi und Nicolai Petkov**. Bd. 5702. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009. Kap. 108, S. 889–897.
- [115] **Zhuo, Shaojie und Terence Sim.** *Defocus Map Estimation from a Single Image*. In: *Pattern Recognition* 44.9 (2011), S. 1852–1858.

- [116] **Zomet, Assaf** und **Shree K. Nayar**. *Lensless Imaging with a Controllable Aperture*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Bd. 1. 2006, S. 339–346.

## Eigene Veröffentlichungen

- [117] **Nürnberg, Thomas** und **Fernando Puente León**. *Das Raytracing-Verfahren als Simulations- und Entwurfswerkzeug einer Computational Camera*. In: *XXIX Messtechnisches Symposium Arbeitskreis der Hochschullehrer für Messtechnik*. Ilmenau, De Gruyter Oldenbourg, Berlin, 2015, S. 123–129.
- [118] **Nürnberg, Thomas** und **Fernando Puente León**. *Aufbau einer Kamera mit programmierbarer Apertur durch Nutzung eines transmissiven Flüssigkristalldisplays*. In: *Forum Bildverarbeitung 2016*. Karlsruhe, KIT Scientific Publishing, 2016, S. 13–24.
- [119] **Nürnberg, Thomas**, **Hannes Weinreuter** und **Fernando Puente León**. *Tiefenmessung mithilfe einer Kamera mit programmierbarer Apertur*. In: *tm – Technisches Messen* 84.S1 (2017), S52–S59.
- [120] **Nürnberg, Thomas**, **Christian Zimmermann** und **Fernando Puente León**. *Simulationsgestützte Optimierung einer Computational-Kamera zur dichten Tiefenschätzung*. In: *tm – Technisches Messen* 83.9 (2016), S. 511–520.
- [121] **Nürnberg, Thomas**, **Thorsten Beuth**, **Jürgen Becker** und **Fernando Puente León**. *An introductory microcontroller programming laboratory course for first-year students*. In: *International Journal of Electrical Engineering Education* 53.2 (2016), S. 99–113.
- [122] **Nürnberg, Thomas**, **Maximilian Schambach**, **David Uhlig**, **Michael Heizmann** und **Fernando Puente León**. *A simulation framework for the design and evaluation of computational cameras*. In: *Automated Visual Inspection and Machine Vision*. Bd. 11061. Proceedings of SPIE, 2019.

## Betreute studentische Arbeiten

- [123] **Ben Hamadi, Chihabeddine**. *Erweiterung einer Kamera-Simulationsumgebung um Störeffekte der digitalen Bildgewinnung*. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.

- [124] **Braun, Sandro.** *Simulation einer Lichtfeldkamera mit Blendenmodulation.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.
- [125] **Durst, Julius Moritz Stephan.** *Modellierung eines Polarization Field Displays und Implementierung einer Ansteuerung zur Darstellung vierdimensionaler Lichtfelder.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2017.
- [126] **Gutierrez Hermosillo Muriedas, Juan Pedro.** *Erweiterung eines Algorithmus zur Tiefenextraktion auf allgemeine zweidimensionale Blendenformen.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.
- [127] **Koubaa, Mohamed Salem.** *Simulation einer Kamera mit variabler Blendenform mit dem Raytracing-Verfahren.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2015.
- [128] **Mezlini, Ramzi.** *Aufbau eines Prototyps einer Programmable Aperture Camera mit einem transparenten Flüssigkristalldisplay.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.
- [129] **Schmitt, Eva.** *Simulation einer Polarisationskamera.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.
- [130] **Smardanski, Bilian Momchilov.** *Modellierung und Simulation von Farbkameras auf Grundlage hyperspektraler VIS-Aufnahmen.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), Fraunhofer ISOB, Karlsruhe, 2016.
- [131] **Weinreuter, Hannes.** *Tiefenmessung mithilfe einer Programmable Aperture Camera mit einem transparenten Flüssigkristalldisplay.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2017.
- [132] **Wolf, Florian.** *Characterisation and analysis of aeroacoustic signals: application to speech production.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), GIPSA-lab, Grenoble, 2014.
- [133] **Zimmermann, Christian.** *Untersuchung von Algorithmen zur Tiefenextraktion mittels einer Kamera mit kodierter Blendenform.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2015.







# **Forschungsberichte aus der Industriellen Informationstechnik (ISSN 2190-6629)**

**Institut für Industrielle Informationstechnik  
Karlsruher Institut für Technologie (KIT)**

---

Hrsg.: Prof. Dr.-Ing. Fernando Puente León

- Band 1 Pérez Grassi, Ana  
**Variable illumination and invariant features for detecting and classifying varnish defects.** (2010)  
ISBN 978-3-86644-537-6
- Band 2 Christ, Konrad  
**Kalibrierung von Magnet-Injektoren für Benzin-Direkteinspritzsysteme mittels Körperschall.** (2011)  
ISBN 978-3-86644-718-9
- Band 3 Sandmair, Andreas  
**Konzepte zur Trennung von Sprachsignalen in unterbestimmten Szenarien.** (2011)  
ISBN 978-3-86644-744-8
- Band 4 Bauer, Michael  
**Vergleich von Mehrträger-Übertragungsverfahren und Entwurfskriterien für neuartige Powerline-Kommunikationssysteme zur Realisierung von Smart Grids.** (2012)  
ISBN 978-3-86644-779-0
- Band 5 Kruse, Marco  
**Mehrobjekt-Zustandsschätzung mit verteilten Sensorträgern am Beispiel der Umfeldwahrnehmung im Straßenverkehr.** (2013)  
ISBN 978-3-86644-982-4
- Band 6 Dudeck, Sven  
**Kamerabasierte In-situ-Überwachung gepulster Laserschweißprozesse.** (2013)  
ISBN 978-3-7315-0019-3
- Band 7 Liu, Wenqing  
**Emulation of Narrowband Powerline Data Transmission Channels and Evaluation of PLC Systems.** (2013)  
ISBN 978-3-7315-0071-1

- Band 8 Otto, Carola  
**Fusion of Data from Heterogeneous Sensors with Distributed Fields of View and Situation Evaluation for Advanced Driver Assistance Systems.** (2013)  
ISBN 978-3-7315-0073-5
- Band 9 Wang, Limeng  
**Image Analysis and Evaluation of Cylinder Bore Surfaces in Micrographs.** (2014)  
ISBN 978-3-7315-0239-5
- Band 10 Michelsburg, Matthias  
**Materialklassifikation in optischen Inspektionssystemen mithilfe hyperspektraler Daten.** (2014)  
ISBN 978-3-7315-0273-9
- Band 11 Pallauf, Johannes  
**Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern.** (2016)  
ISBN 978-3-7315-0529-7
- Band 12 Sigle, Martin  
**Robuste Schmalband-Powerline-Kommunikation für Niederspannungsverteilernetze.** (2016)  
ISBN 978-3-7315-0539-6
- Band 13 Opalko, Oliver  
**Powerline-Kommunikation für Batteriemangement-Systeme in Elektro- und Hybridfahrzeugen.** (2017)  
ISBN 978-3-7315-0647-8
- Band 14 Han, Bin  
**Characterization and Emulation of Low-Voltage Power Line Channels for Narrowband and Broadband Communication.** (2017)  
ISBN 978-3-7315-0654-6
- Band 15 Alonso, Damián Ezequiel  
**Wireless Data Transmission for the Battery Management System of Electric and Hybrid Vehicles.** (2017)  
ISBN 978-3-7315-0670-6

- Band 16 Hernández Mesa, Pilar  
**Design and analysis of a content-based image retrieval system.** (2017)  
ISBN 978-3-7315-0692-8
- Band 17 Suchanek, André  
**Energiemanagement-Strategien für batterieelektrische Fahrzeuge.** (2018)  
ISBN 978-3-7315-0773-4
- Band 18 Bauer, Sebastian  
**Hyperspectral Image Unmixing Incorporating Adjacency Information.** (2018)  
ISBN 978-3-7315-0788-8
- Band 19 Vater, Sebastian  
**Monokulare Blickrichtungsschätzung zur berührungslosen Mensch-Maschine-Interaktion.** (2019)  
ISBN 978-3-7315-0789-5
- Band 20 Back, Kristine  
**Erkennung menschlicher Aktivitäten durch Erfassung und Analyse von Bewegungstrajektorien.** (2019)  
ISBN 978-3-7315-0909-7
- Band 21 Nürnberg, Thomas  
**Entwurf von Computational-Imaging-Systemen am Beispiel der monokularen Tiefenschätzung.** (2019)  
ISBN 978-3-7315-0941-7

ISSN 2190-6629  
ISBN 978-3-7315-0941-7

Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-7315-0941-7



9 783731 509417 >