

# Bildbasierte Beleuchtung in 3D-Photo-Collections



Vom Promotionsausschuss des Fachbereichs 4: Informatik  
der Universität Koblenz-Landau zur Verleihung des akademi-  
schen Grades Doktor der Naturwissenschaften (Dr. rer. nat.)  
genehmigte

## Dissertation

VON

**M.Sc. Konrad Kölzer**

Datum der wissenschaftlichen Aussprache:

23. Juli 2014

Vorsitzender des Promotionsausschusses:

JProf. Dr. Mario Schaarschmidt

1. Berichterstatter:

Prof. Dr.-Ing. Stefan Müller

2. Berichterstatter:

Prof. Dr.-Ing. Dietrich Paulus

3. Berichterstatter:

Prof. Dr. Paul Grimm



# Danksagung

Ich möchte an dieser Stelle allen Personen danken, die zum Gelingen meiner Doktorarbeit beigetragen haben. Dank gebührt vor allem meinem Doktorvater Prof. Dr. Stephan Müller, der mich nicht nur als Doktorand angenommen, sondern auch über all die Jahre fachlich betreut und damit auf meinem Weg zur Promotion begleitet hat. Bedanken möchte ich mich auch bei meinem Zweitbetreuer Prof. Dr. Paul Grimm, der als mein Projektleiter die Promotion erst ins Rollen gebracht hat.

Bedanken möchte ich mich auch bei meinen Freunden und langjährigen Kollegen Frank Nagl und Jonas Etzold, mit denen die Arbeit als wissenschaftlicher Mitarbeiter viel Spaß gemacht hat und nie in Alltag erstarrt ist. Dank euch war so manches überlanges Projektmeeting erträglich ;-).

Dank gilt auch meinen Eltern für meine Existenz im Allgemeinen sowie Korrekturlesen im Speziellen. Zu guter Letzt möchte ich mich bei Alotta für die willkommene Ablenkung sowie manch geistreiches Bonmot bedanken.



# Kurzfassung

Ziel dieser Arbeit ist die Entwicklung von Verfahren zur realistischen augmentierten Bildsynthese auf Basis von 3D-Photo-Collections. 3D-Photo-Collections sind aus Einzelfotos automatisch erzeugte Repräsentationen einer realen Szene und geben diese als Menge von Bildern mit bekannten Kameraposen sowie einer groben punktbasierten Modellierung der Szenengeometrie wieder. Es wird eine fotorealistische augmentierte Bildsynthese von realen und virtuellen Anteilen in Echtzeit angestrebt, wobei die reale Szene durch 3D-Photo-Collections beschrieben wird. Um dieses Ziel zu erreichen, werden drei Problemfelder bearbeitet.

Da die Fotos unter Umständen in verschiedenen geräteabhängigen RGB-Farbräumen liegen, ist eine Farbcharakterisierung der 3D-Photo-Collections notwendig, um korrekte, der menschlichen Wahrnehmung entsprechende Farbinformationen zu erhalten. Das hierzu entwickelte Verfahren transformiert alle Bilder automatisch in einen gemeinsamen Farbraum und vereinfacht so die Farbcharakterisierung von 3D-Photo-Collections.

Als Hauptproblem der augmentierten Bildsynthese muss die Umgebungsbeleuchtung der realen Szene bekannt sein, um eingefügte virtuelle Anteile konsistent zu den realen Anteilen zu beleuchten. Hierfür wurden zwei neue Verfahren zur Rekonstruktion der Umgebungsbeleuchtung aus den Bildern der 3D-Photo-Collection entwickelt.

Um eine Bildsynthese für beliebige Ansichten auf die Szene durchzuführen, wurde ein neues Verfahren zur bildbasierten Darstellung entwickelt, welches neue Ansichten auf die 3D-Photo-Collection unter direkter Verwendung der Punktwolke erzeugt. Dieser Ansatz erzeugt neue Ansichten in Echtzeit und erlaubt somit eine freie Navigation.

Insgesamt konnte mit den entwickelten Methoden gezeigt werden, dass 3D-Photo-Collections für Augmented-Reality eine geeignete Repräsentation von realen Anteilen sind und eine realistische Synthese mit virtuellen Anteilen durchgeführt werden kann.



# Abstract

The goal of this thesis is the development of methods for augmented image synthesis using 3D photo collections. 3D photo collections are representations of real scenes automatically generated from single photos and describe a scene as a set of images with known camera poses as well as a sparse point-based model of the scene geometry. The main goal is to perform a photo-realistic augmented image synthesis of real and virtual parts, where the real scene is provided as a 3D photo collection. Therefore, three main problems are addressed.

Since the photos may be represented in different device-specific RGB color spaces, a color characterization of the 3D photo collections is necessary to gain correct color information that is consistent with human perception. The proposed novel method automatically transforms all images into a common RGB color space and thereby simplifies color characterization of 3D photo collections.

As a main problem for augmented image synthesis, all environmental lighting has to be known in order to apply illumination to virtual parts that is consistent with the real portions shown in the photos. To solve this problem, two novel methods were developed to reconstruct the lighting from 3D photo collections.

In order to perform image synthesis for arbitrary views on the scene, an image-based approach was developed that generates new views in 3D photo collections making direct use of its point cloud. The novel method creates new views in real-time and allows free-navigation.

In conclusion, the proposed novel methods show that 3D photo collections are a useful representation for real scenes in Augmented Reality and they can be used to perform a realistic image synthesis of real and virtual portions.



# Inhaltsverzeichnis

<b>Tabellenverzeichnis</b>	<b>XIII</b>
----------------------------	-------------

<b>Abbildungsverzeichnis</b>	<b>XV</b>
------------------------------	-----------

<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation und Zielsetzung . . . . .	2
1.1.1 Farbcharakterisierung . . . . .	4
1.1.2 Bildbasierte Beleuchtung . . . . .	4
1.1.3 Darstellung bildbasierter Szenen . . . . .	5
1.2 Wissenschaftlicher Beitrag . . . . .	6
1.2.1 Farbmétrische Charakterisierung . . . . .	6
1.2.2 Bildbasierte Beleuchtung . . . . .	6
1.2.3 Bildbasierte Darstellung . . . . .	7
<b>2 Grundlagen</b>	<b>9</b>
2.1 Physikalische Eigenschaften von Licht . . . . .	9
2.1.1 Spektrum des sichtbaren Lichts . . . . .	9
2.1.2 Raumwinkel . . . . .	10
2.1.3 Radiométrische Größen . . . . .	11
2.1.4 Photométrische Größen . . . . .	12
2.1.5 Photométrische Gesetzmäßigkeiten . . . . .	14
2.1.6 Beschreibung von Reflexionsverhalten . . . . .	17
2.2 Farbmétrische Grundlagen . . . . .	19
2.2.1 CIE-Normspektralraum . . . . .	19
2.2.2 CIE-Normfarbtafel . . . . .	22
2.2.3 CIE-L*a*b*-Farbraum . . . . .	22
2.2.4 Geräteabhängige Farbräume . . . . .	24
2.2.5 Farbmémerie . . . . .	25

2.3	Bildbasierte Repräsentation von Szenen . . . . .	26
2.3.1	Plenoptische Funktion . . . . .	26
2.3.2	Erzeugung von 3D-Photo-Collections . . . . .	27
<b>3</b>	<b>Farbmetrische Charakterisierung von bildbasierten Szenen</b>	<b>33</b>
3.1	Einleitung . . . . .	33
3.1.1	Problemstellung . . . . .	33
3.1.2	Charakterisierung in 3D-Photo-Collections . . . . .	34
3.1.3	Kalibrierung der Kameramodelle . . . . .	36
3.2	Verwandte Arbeiten . . . . .	37
3.2.1	Farbcharakterisierung von digitalen Kameras . . . . .	37
3.2.2	Farbkorrektur für mehrere Kameras . . . . .	39
3.3	Eigener Ansatz . . . . .	40
3.3.1	Suche nach korrespondierenden Merkmalspunkten . . . . .	42
3.3.2	Abtastung von RGB-Werten . . . . .	43
3.3.3	Kompensation von Belichtungszeit und Blende . . . . .	47
3.3.4	Modell für Farbraumabbildung . . . . .	49
3.4	Auswertung . . . . .	51
3.4.1	Ergebnisse . . . . .	51
3.4.2	Fehlerbetrachtung . . . . .	51
3.5	Zusammenfassung und Ausblick . . . . .	56
<b>4</b>	<b>Bildbasierte Beleuchtung in 3D-Photo-Collections</b>	<b>59</b>
4.1	Einleitung . . . . .	59
4.1.1	Anforderungen . . . . .	59
4.1.2	Zielsetzung . . . . .	61
4.2	Verwandte Arbeiten . . . . .	63
4.2.1	Augmentierte Bildsynthese . . . . .	63
4.2.2	Rekonstruktion der Kamerakennkurve aus Belichtungsreihen . . . . .	73
4.3	Light-Probe-Synthese ohne bekannte Szenengeometrie . . . . .	82
4.3.1	Abbildung von Strahldichte auf RGB-Werte . . . . .	82
4.3.2	Approximation der Szenengeometrie . . . . .	84
4.3.3	Akkumulation der Strahldichten . . . . .	86
4.3.4	Generierung direktonaler Lichtquellen . . . . .	88
4.3.5	Ergebnisse . . . . .	89

4.4	Synthese von Light-Probes mit gegebener Szenengeometrie . . .	92
4.4.1	Bereitstellung der Szenengeometrie . . . . .	92
4.4.2	Verwendung eines Texturatlas . . . . .	94
4.4.3	Bestimmung der Strahldichten . . . . .	95
4.4.4	Generierung von Light-Probes . . . . .	97
4.4.5	Ergebnisse . . . . .	98
4.5	Auswertung . . . . .	99
4.5.1	Zusammenfassung und Vergleich beider Verfahren . . .	99
4.5.2	Ausblick . . . . .	104
<b>5</b>	<b>Darstellung bildbasierter Szenen</b>	<b>107</b>
5.1	Einleitung . . . . .	107
5.2	Verwandte Arbeiten . . . . .	109
5.2.1	Plenoptisches Modellieren . . . . .	109
5.2.2	Lumigraphs und Lichtfelder . . . . .	110
5.2.3	View-dependent Texture Mapping . . . . .	112
5.2.4	Unstructured Lumigraphs . . . . .	114
5.2.5	Bildbasierte Darstellung mit Multi-View Stereo . . . . .	115
5.3	Eigener Ansatz . . . . .	118
5.3.1	Daten der 3D-Photo-Collection . . . . .	118
5.3.2	Ziele . . . . .	119
5.3.3	Modellierung von Szenenoberflächen . . . . .	120
5.3.4	Zuordnung von Bildern an Punkte . . . . .	122
5.3.5	Blickwinkelabhängige Projektion der Bildinhalte . . . . .	123
5.3.6	Umsetzung der Darstellung . . . . .	126
5.3.7	Rendering . . . . .	126
5.4	Auswertung . . . . .	128
5.5	Zusammenfassung und Ausblick . . . . .	132
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>133</b>
6.1	Zusammenfassung . . . . .	133
6.1.1	Farbcharakterisierung . . . . .	133
6.1.2	Bildbasierte Beleuchtung . . . . .	134
6.1.3	Darstellung bildbasierter Szenen . . . . .	135
6.2	Zusammenhang zwischen den Verfahren . . . . .	135
6.3	Anforderungen an 3D-Photo-Collections . . . . .	137

## *Inhaltsverzeichnis*

---

6.4	Ausblick . . . . .	139
6.4.1	Abhängigkeiten zu externen Daten . . . . .	139
6.4.2	Überschneidung der Problemfelder . . . . .	140
6.4.3	Fazit . . . . .	141
	<b>Eigene Veröffentlichungen</b>	<b>143</b>
	<b>Literaturverzeichnis</b>	<b>147</b>

# Tabellenverzeichnis

2.1	Photometrische und radiometrische Größen . . . . .	14
3.1	Gemessener Farbfehler für verschiedene Szenen . . . . .	56
4.1	Vergleich beider Verfahren zur Beleuchtungsrekonstruktion . .	103
5.1	Darstellungsleistung für bildbasiertes Rendering . . . . .	130
6.1	Anforderungen der Verfahren . . . . .	137



# Abbildungsverzeichnis

1.1	Aufbau einer 3D-Photo-Collection . . . . .	2
1.2	Grundlegende Problemfelder dieser Arbeit . . . . .	3
2.1	Farbanteile im Spektrum des sichtbaren Lichtes . . . . .	9
2.2	Raumwinkel eines Flächenelements $A_E$ . . . . .	10
2.3	Strahldichte als Quotient aus Strahlungsleistung und dem Produkt aus Raumwinkel $d\omega$ und projizierter Fläche $\cos\theta dA$ . . .	12
2.4	Empfindlichkeitskurve des hell adaptierten Auges $V(\lambda)$ . . . . .	13
2.5	Photometrische Gesetzmäßigkeiten zwischen einer Sender- und Empfängerfläche . . . . .	15
2.6	Oberfläche als Lambertscher Strahler mit konstanter Leuchtdichte . . . . .	16
2.7	Durch Kugelkoordinaten parametrisierte ein- und ausgehende Strahldichten . . . . .	18
2.8	Näherungsweise Wiedergabe der spektralen Empfindlichkeit der L, M, und S-Zapfen des menschlichen Auges . . . . .	20
2.9	CIE-Normspektralwertfunktionen . . . . .	21
2.10	CIE-Normfarbtafel mit monochromatischen Spektren . . . . .	23
2.11	Zwei metamere Lichtspektren mit gleichem XYZ-Wert . . . . .	26
2.12	Darstellung der plenoptischen Funktion . . . . .	27
2.13	Eingabebilder einer 3D-Photo-Collection . . . . .	28
2.14	Durch SIFT erkannte Merkmalspunkte in zwei Bildern . . . . .	29
2.15	Durch Bundler in zwei Bildern gefundene korrespondierende Merkmalspunkte . . . . .	29
2.16	3D-Photo-Collection mit bekannten Kameraposen und 3D-Punktwolke . . . . .	31
3.1	Charakterisierung in einen einheitlichen RGB-Farbraum mit optionaler Charakterisierung in das CIE-Normvalenzsystem im Anschluss . . . . .	35

3.2	Von Bundler erzeugte 3D-Photo-Collection <i>Buchstapel</i> . . . . .	41
3.3	Korrespondierende Merkmalspunkte ermittelt aus dem uncharakterisierten Kameramodell, $D_u$ , (unten) und dem Referenzkameramodell, $D_r$ (oben) . . . . .	43
3.4	Ermittlung des Radius anhand des Strahlensatzes . . . . .	44
3.5	Berechnung der Brennweite in Pixeln aus Öffnungswinkel $\alpha$ und Bildbreite $w$ in Pixeln . . . . .	45
3.6	Berechnete Radien der Kreisausschnitte für beide Bildpaare . . . . .	46
3.7	Über Gewichtungsfunktion $w(x)$ ermittelte Werte für jeden Pixel des Kreisausschnitts mit dem Radius $r$ . . . . .	47
3.8	Bilder der Szene <i>Buchstapel</i> von drei verschiedenen Kameramodellen vor (links) und nach Charakterisierung (rechts) . . . . .	52
3.9	Bilder der Szene <i>Flyerwand</i> von drei verschiedenen Kameramodellen vor (links) und nach Charakterisierung (rechts) . . . . .	53
3.10	Ergebnisse der vorgestellten Charakterisierung in Kombination mit einer bildbasierten Darstellung der 3D-Photo-Collection <i>Flyerwand</i> . . . . .	54
3.11	Korrespondierende RGB-Werte aus der Szene <i>Buchstapel</i> vor und nach Charakterisierung . . . . .	55
4.1	Geänderte Lichtpfade durch Einfügen virtueller Anteile . . . . .	60
4.2	Rückschluss von Bildpixeln auf Strahldichten der Umgebung . . . . .	62
4.3	Synthese nach Nakamae et. al. . . . .	64
4.4	Synthese nach Fournier et. al. . . . .	65
4.5	Synthese nach Drettakis et. al. . . . .	66
4.6	Synthese nach Debevec et. al. . . . .	67
4.7	Synthese nach Sato et. al. . . . .	69
4.8	Synthese nach Loscos et. al. . . . .	70
4.9	Synthese mit differenziellem Photon Mapping nach Grosch [Gro05]. Im oberen Bereich sind virtuelle Objekte eingefügt, während die Objekte im unteren Bildbereich real sind. . . . .	72
4.10	Belichtungsreihe bestehend aus Fotos, welche die gleiche Ansicht auf die Szene zeigen, aber mit unterschiedlichen Belichtungszeiten erstellt wurden . . . . .	73
4.11	Gewichtungsfunktion $w(Z)$ nach Debevec . . . . .	76
4.12	Rekonstruierte Kamerakennkurve nach Debevec . . . . .	77

---

4.13 Auswahl der 201 gemessenen Kamerakennkurven von Grossberg und Nayar . . . . .	78
4.14 Empirisches Modell (EMoR) zur Abschätzung der Kamerakennkurve nach Grossberg und Nayar . . . . .	79
4.15 Gewichtungsfunktion $w(Z)$ nach Robertson et. al. . . . .	80
4.16 Rekonstruktion der Kamerakennkurve und Abschätzung der Gewichtungsfunktion nach Robertson . . . . .	81
4.17 Bildpixel als Bewertung der von der Kamera empfangenen spektralen Strahldichte aus der realen Szene . . . . .	82
4.18 Modell nach [GN04] für die Abbildung der aus der Szene empfangenen Strahldichte $L$ über die Bestrahlungsstärke $E$ des Bildsensorausschnitts auf den Wert $Z$ durch eine Kamera. .	83
4.19 Rückprojektion der vom Punkt $P$ empfangenen Strahldichte (rot) auf die Szenenoberfläche (blau) und anschließende Projektion der Bestrahlungsstärke in Richtung $P_R$ unter Annahme vollständig diffusen Reflexionsverhaltens der Szenenoberfläche	85
4.20 Punktwolke einer 3D-Photo-Collection als grobe und verrauschte Repräsentation der Szenengeometrie . . . . .	85
4.21 Kameras einer 3D-Photo-Collection, die innerhalb des um den Objektmittelpunkt $P_R$ liegenden Radius $r_R$ liegen. . . . .	86
4.22 Synthese von Beleuchtungsdaten aus den Bildern der 3D-Photo-Collection in die HDR-Environment-Map . . . . .	89
4.23 Synthese von 3D-Photo-Collection und virtuellem Modell . . .	90
4.24 Synthese von realen und virtuellen Anteilen unter Nutzung der Beleuchtungsrekonstruktion ohne Szenengeometrie . . . .	90
4.25 Ansichten eines 3D-Modells mit von hinten kommender Beleuchtung, welche aus den Fotos extrahiert wurde . . . . .	91
4.26 Bilder der 3D-Photo-Collection mit registrierter Szenengeometrie	93
4.27 Szenengeometrie mit Texturatlas . . . . .	94
4.28 Werte der Sichtbarkeitsfunktion zwischen dem Patch $p$ und den Pixeln aus den drei Bildern $I_0$ , $I_1$ und $I_2$ . . . . .	96
4.29 Berücksichtigung von Verdeckungen bei der Projektion von Strahldichten auf die Szenengeometrie . . . . .	97
4.30 Virtuelles Objekt, dessen Beleuchtung aus den berechneten Strahldichten der realen Oberflächen erzeugt wurde . . . . .	98
4.31 Eingefügtes virtuelles Objekt mit stark reflektierendem Material	99

4.32	Virtuelles Objekt, beleuchtet durch Light-Probe, welche aus den bekannten Strahldichten der realen Oberflächen erzeugt wurde . . . . .	100
4.33	Generierte Light-Probes in Form einer HDR-Cube-Map für zwei unterschiedliche Positionen in der Szene . . . . .	101
4.34	Erzeugte direktionale Lichtquellen . . . . .	102
4.35	Fehlerhafte Rekonstruktion von Strahldichten auf Grund falsch projizierter Bildpixel an Bereichen mit großen Tiefenunterschieden im Bildraum . . . . .	103
5.1	3D-Photo-Collection mit begrenzter Anzahl von Ansichten . . .	108
5.2	Modellierung der plenoptischen Funktion nach McMillan und Bishop . . . . .	109
5.3	Parametrisierung von Strahlen im Lichtfeld in Form von zwei Schnittpunkten . . . . .	110
5.4	Rendering von Lichtfeldern nach Levoy . . . . .	111
5.5	Rendering von Lichtfeldern nach Gortler . . . . .	112
5.6	View-dependent Texture Mapping nach Debevec . . . . .	113
5.7	Rendering von Unstructured Lumigraphs nach Buehler . . . . .	114
5.8	Darstellung bildbasierter Szenen nach Furukawa . . . . .	116
5.9	Sichtinterpolation bildbasierter Szenen nach Goesele . . . . .	117
5.10	Schematische Darstellung einer 3D-Photo-Collection als bildbasiertes Modell der realen Szene . . . . .	118
5.11	Wahrscheinlichkeitsbasierte Modellierung der Szenengeometrie durch Funktion $f_p(\vec{x})$ . . . . .	121
5.12	Einteilung der Umgebung in $n_c$ Abschnitte . . . . .	122
5.13	Projektion eines Bildausschnitts auf ein Billboard . . . . .	124
5.14	Blickwinkelabhängiger Term $o_v(\alpha)$ . . . . .	125
5.15	Erhöhung des Deckungsgrades als Nachbearbeitungsschritt . .	127
5.16	Verschiedene generierte Ansichten auf die Szene <i>Gassen</i> . . . .	128
5.17	Verschiedene generierte Ansichten auf die Szene <i>Marktplatz</i> . .	129
5.18	Verschiedene Größen der Billboards durch unterschiedliche Werte des Skalierungsfaktors $k_r$ . . . . .	130
5.19	Bildbasierte Darstellung mit eingefügtem virtuellen Modell . .	131
6.1	Prozessablauf der vorgestellten Verfahren . . . . .	136
6.2	Benötigte Eingabedaten der Verfahren . . . . .	139

# Kapitel 1

## Einleitung

Augmented-Reality (AR) stellt einen Teilbereich der Computergrafik dar und befasst sich mit dem Hinzufügen von virtuellen Anteilen zur realen Welt. Im Gegensatz zu virtueller Realität (Virtual-Reality, VR) wird nicht eine vollständig künstliche Welt generiert, sondern lediglich die reale Welt um einige virtuelle Objekte erweitert. Für viele Anwendungen wird dabei eine Synthese von realen und virtuellen Anteilen mit hohem realistischen Anspruch angestrebt, um eine für den menschlichen Betrachter als real wahrgenommene Darstellung zu erhalten. Dies ist vor allem bedeutsam für Anwendungsszenarien, wie zum Beispiel die Visualisierung von Planungen für Architektur oder die Gestaltung von Innenräumen, in denen ein möglichst genauer Eindruck vermittelt werden soll, wie bestimmte Planungen nach Fertigstellung in der realen Umgebung aussehen.

Um diesen Anspruch umzusetzen, müssen für die reale Szene deren räumliche Struktur, sowie das Reflexionsverhalten und die Leuchtdichte ihrer Oberflächen bekannt sein. Für die eigentliche Darstellung der realen Anteile wird meist ein einzelnes Foto oder ein Videostrom verwendet, deren Kameraposen ebenfalls bekannt sein müssen. In bestehenden Ansätzen werden noch viele Informationen über reale Anteile manuell abgeschätzt oder mit relativ hohem Aufwand gemessen. Dazu zählt zum Beispiel die manuelle Konstruktion der Szenengeometrie oder die Messung von Leuchtdichten der Umgebung mit Hilfe von Light-Probes [Deb98].

Stattdessen kann eine reale Szene auch durch eine Menge von Fotos beschrieben werden, welche die Szene aus unterschiedlichen Perspektiven zeigen. Mit Hilfe von bestehenden Verfahren der Computer Vision und der Photogrammetrie wie zum Beispiel Detektion von Bildmerkmalen und Bündelblockausgleichung können daraus automatisch die Kameraposen für jedes

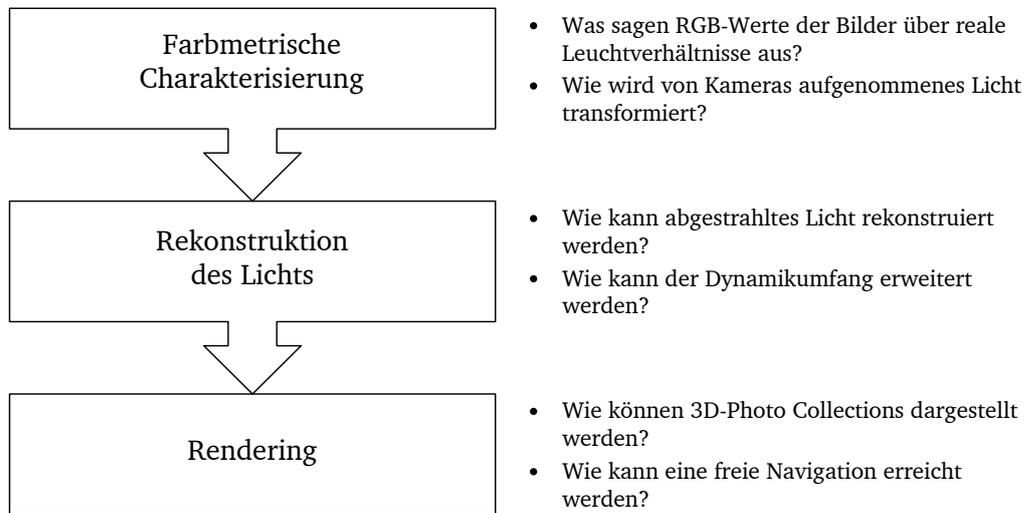


**Abbildung 1.1:** Eine 3D-Photo-Collection besteht aus Bildern mit bekannten Kameraposen sowie einer 3D-Punktwolke, welche eine grobe Rekonstruktion der Szenengeometrie darstellt.

Bild sowie eine sehr grobe Rekonstruktion der Szenengeometrie in Form von Punktwolken gewonnen werden (siehe Abb. 1.1). Auf diese Weise bilden Bilder, Kameraparameter und Punktwolke eine Datenstruktur, welche bereits viele Eigenschaften der realen Szene beschreibt. Nachfolgend soll diese Repräsentation von Szenen als 3D-Photo-Collection bezeichnet werden. In dieser Arbeit soll deren Verwendung in der Augmented-Reality-Synthese von virtuellen und realen Anteilen erforscht werden.

## 1.1 Motivation und Zielsetzung

Durch stetige Weiterentwicklung von Bundle-Adjustment-Verfahren [LA04] [SSS06] [FFGG<sup>+</sup>10] können aus ungeordneten Fotos einer Szene intrinsische und extrinsische Kameraparameter robust und vollständig automatisch rekonstruiert werden. Entsprechende Software Bundle-Adjustment-Tools wie etwa Bundler [Sna] oder Microsoft<sup>®</sup> Photosynth [Mic] sind bereits seit mehreren Jahren frei zugänglich und insbesondere Letzteres wendet sich an Endnutzer ohne technisches Fachwissen. Als Eingabebilder genügen mit einer Consumer-Digitalkamera gemachte Fotos, so dass für die Erstellung von 3D-



**Abbildung 1.2:** Grundlegende Problemfelder dieser Arbeit mit exemplarischen Fragestellungen

Photo-Collections insgesamt keine aufwändige Messtechnik oder spezielle Fachkenntnisse benötigt werden.

Auf Grund ihres einfachen Aufnahmeprozesses, der Fülle rekonstruierter Daten und der allgemeinen Verfügbarkeit entsprechender Software, eignen sich 3D-Photo-Collections gut zur Beschreibung von realen Umgebungen in Form bildbasierter Szenen. Damit stellen sie auch in der Augmented-Reality eine Alternative zu bisherigen Verfahren dar, bei denen reale Umgebungen meist durch Einzelfotos sowie einer Light-Probe beschrieben werden.

Im Gegensatz zu Einzelbildern liefern 3D-Photo-Collections bereits eine grobe Rekonstruktion der Szenengeometrie, was für die Einbettung von virtuellen Anteilen genutzt werden kann. Gegenüber Light-Probes benötigen 3D-Photo-Collections keine aufwändige Messausrüstung (Stativ sowie reflektierende Kugeln oder Fischaugenlinsen). Außerdem liefern sie Bildinformationen für viele verschiedene Positionen der realen Umgebung, während eine Light-Probe nur Umgebungslicht für genau einen Punkt der Szene beschreibt. Auf Grund dieser Eigenschaften lautet das Hauptziel dieser Arbeit geeignete Verfahren zu entwickeln, mit denen eine realistische AR-Synthese von realen und virtuellen Anteilen unter Verwendung von 3D-Photo-Collections durchgeführt werden kann.

Jedoch stellen sich für eine realistische Synthese von realen und virtuellen Anteilen unter Verwendung von 3D-Photo-Collections drei grundlegende Fragestellungen (siehe auch Abbildung 1.2), die im Rahmen dieser Arbeit behandelt werden sollen.

### 1.1.1 Farbcharakterisierung

Jedes Bild einer 3D-Photo-Collection beschreibt das aus den Kamerapositionen empfangene Licht der realen Szene als RGB-Werte. Dabei ist jeder RGB-Wert Folge einer spektralen Bewertung von Licht durch die Kamera. Um aus RGB-Werten der menschlichen Wahrnehmung entsprechende Farbinformationen zu erhalten, muss der Bildsensor die Luther-Ives-Bedingung [Lut27] [Ive15] erfüllen. Das bedeutet, die spektrale Bewertung des Bildsensors muss sich linear aus der des menschlichen Auges ergeben. Dies ist jedoch nicht oder höchstens in grober Näherung der Fall.

Auf Grund dieser Abweichung der spektralen Bewertung, geben Kameras Farben anders wieder als dies der menschlichen Wahrnehmung entspricht. Die Farbwiedergabe ist nicht nur je nach Kamera unterschiedlich, sondern hängt zusätzlich von Kameraparametern wie zum Beispiel Weißabgleich, Belichtungszeit oder Blende ab. Daraus folgt, dass bei einer 3D-Photo-Collection, insbesondere bei der Verwendung von verschiedenen Kameras, viele verschiedene RGB-Farbräume vorliegen.

Sowohl für eine Rekonstruktion der realen Umgebungsbeleuchtung als auch für eine bildbasierte Darstellung der 3D-Photo-Collection ist es jedoch notwendig, dass sich alle Bilder in einem RGB-Farbraum – oder noch besser – im geräteunabhängigen CIE XYZ-Farbraum befinden. Erst dadurch können alle aus den Bildern entnommenen Farbwerte gemeinsam verwendet werden. Im Rahmen dieser Arbeit sollen daher Wege zur möglichst präzisen farbmtrischen Charakterisierung der 3D-Photo-Collection gefunden werden.

### 1.1.2 Bildbasierte Beleuchtung

Um eine realistische Synthese von virtuellen Objekten und realen Anteilen zu erreichen, muss die Beleuchtung virtueller Objekte konsistent zum abgestrahlten Licht realer Oberflächen sein. Diese reale Umgebungsbeleuchtung muss aus den Bildern der 3D-Photo-Collection extrahiert werden. Dabei stellen sich zwei Probleme:

Zum einen muss das ausgestrahlte Licht der realen Szene lokalisiert werden. Das von einer Kamera aufgenommene Bild gibt für jeden Pixel ein aus der Szene in die Kameraöffnung eingestrahktes Lichtbündel wieder, welches vom Objektiv der Kamera auf eine bestimmte Position des Bildsensors abgebildet wurde. Für eine Lokalisierung des ausgestrahlten Lichtes muss eine pixelweise Zurückverfolgung von RGB-Werten auf reale Oberflächen erfolgen und deren Strahldichte ermittelt werden.

Zum anderen unterliegen die Farbwerte der Bildpixel einer zu geringen Quantisierung. In der Regel verwenden die Fotos als Pixelformat 24 Bit, womit die Farbkanäle Rot, Grün und Blau mit 8 Bit, also 256 Werten quantisiert werden. Dies ist aber nicht ausreichend, um den kompletten Dynamikumfang einer realen Szene zu erfassen – das Verhältnis der Leuchtdichten von Licht und Schatten kann bei Tageslichtszenen zwischen direktem Sonnenlicht und Schattenbereichen 10000:1 oder mehr betragen. Jedoch kann auf einem Foto – je nach gewähltem Belichtungswert der Kamera – nur ein kleiner Ausschnitt des realen Dynamikumfangs gemessen werden. Für eine Rekonstruktion der realen Umgebungsbeleuchtung und die anschließende Darstellung der realen Szene muss aber im wesentlichen der gesamte Dynamikumfang abgedeckt werden. Im Rahmen dieser Arbeit soll nach Möglichkeiten gesucht werden, den kompletten Dynamikumfang einer realen Szene zu rekonstruieren, ohne spezielle Hardware wie zum Beispiel HDR-Kameras vorauszusetzen.

### 1.1.3 Darstellung bildbasierter Szenen

Durch den Einsatz von 3D-Photo-Collections liegen Fotos der realen Szene aus mehreren unterschiedlichen Kameraperspektiven vor. Im einfachsten Fall kann die Synthese von realen und virtuellen Anteilen für eine Ansicht der realen Szene aus einer dieser Kameraperspektiven dargestellt werden, so dass das zugehörige Foto bildschirmfüllend verwendet werden kann. Für eine interaktive Verwendung ist jedoch eine freie und interaktive Wahl der Betrachterperspektive auf die Szene erstrebenswert, so dass für eine beliebige Betrachterposition eine neue Ansicht aus den umliegenden Kameraperspektiven interpoliert werden kann. Im Rahmen dieser Arbeit soll ein Verfahren für eine möglichst freie und interaktive Navigation durch die reale Szene entwickelt werden.

## 1.2 Wissenschaftlicher Beitrag

Diese Arbeit präsentiert Verfahren und Ergebnisse aus den oben beschriebenen drei Problemfeldern Farbcharakterisierung, bildbasierte Beleuchtung und Darstellung bildbasierter Szenen. Auf das Problemfeld bildbasierte Beleuchtung wurde besonderer Augenmerk gelegt, da sie für die AR-Synthese von zentraler Bedeutung ist.

### 1.2.1 Farbmétrische Charakterisierung

Es wurde ein neues Verfahren vorgestellt [KMG13], mit dem die Farbcharakterisierung der Bilder von 3D-Photo-Collections vereinfacht wird. Dabei werden Bilder, welche mit unterschiedlichen Kameramodellen oder mit unterschiedlicher Belichtungszeit oder Blendenzahl fotografiert wurden, automatisch in einen gemeinsamen RGB-Farbraum transformiert, welcher durch die Auswahl eines Referenzkameramodells festgelegt wird. Die Charakterisierung in einen geräteunabhängigen Farbraum muss dann im Anschluss nur für das Referenzkameramodell durchgeführt werden, was den Messaufwand mit der Anzahl der verwendeten Kameramodelle erheblich verringert.

### 1.2.2 Bildbasierte Beleuchtung

Für dieses Problemfeld wurden insgesamt zwei neue Verfahren entwickelt. Beide Verfahren setzen die Rekonstruktion der realen Umgebungsbeleuchtung aus den Bildern einer 3D-Photo-Collection um und stellen diese als HDR-Environment-Maps zur Verfügung, womit anschließend virtuelle Anteile konsistent beleuchtet werden. Beide Verfahren extrahieren Beleuchtungsdaten der realen Umgebung aus den gegebenen Bilddaten und erzeugen so eine Environment-Map mit hohem Dynamikumfang.

Im ersten Verfahren [KNBG09] [KFNG09] werden die Bilder der 3D-Photo-Collection auf Ebenen projiziert, welche als grobe, geometrische Repräsentation der Szenenstruktur verwendet werden. Im zweiten Verfahren [KNG11] wird Szenengeometrie in Form eines 3D-Meshes vorausgesetzt, worauf die Bilder unter Annahme von diffusen Szenenoberflächen reprojiziert werden.

Im Vergleich zu bereits existierenden Verfahren, die Beleuchtungsdaten in Form von Light-Probes mittels spezieller Ausrüstung messen, können so Be-

leuchtungsdaten aus Bildern von gewöhnlichen digitalen Kameras gewonnen werden.

### **1.2.3 Bildbasierte Darstellung**

Der vorgestellte Ansatz [KNGM12] generiert beliebige neue Ansichten auf die Szene in Echtzeit ohne vollständige Geometrie vorauszusetzen. Stattdessen werden geometrische Informationen der aus den Bildmerkmalen gewonnenen Punktwolke verwendet. Die Darstellung wird dabei an die Dichte der Punktwolke in unterschiedlichen Bereichen der Szene angepasst.

Im Gegensatz zu vielen bereits existierenden Verfahren im Bereich Image-based Rendering, welche ebenfalls Informationen über die Szenengeometrie verwenden, entfällt im hier vorgestellten Verfahren eine zeitaufwändige Rekonstruktion der Szenengeometrie. Stattdessen wird für die Darstellung die Punktwolke als grobe Repräsentation der Szenengeometrie direkt genutzt.



# Kapitel 2

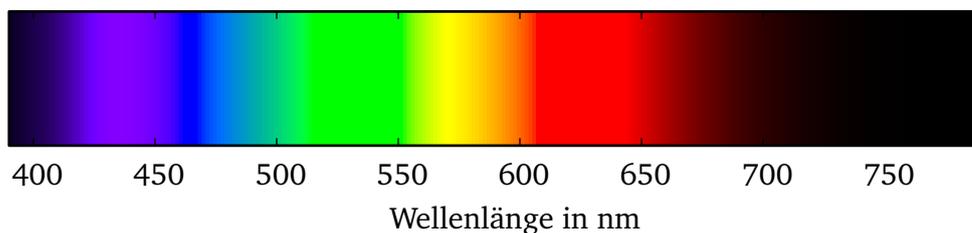
## Grundlagen

### 2.1 Physikalische Eigenschaften von Licht

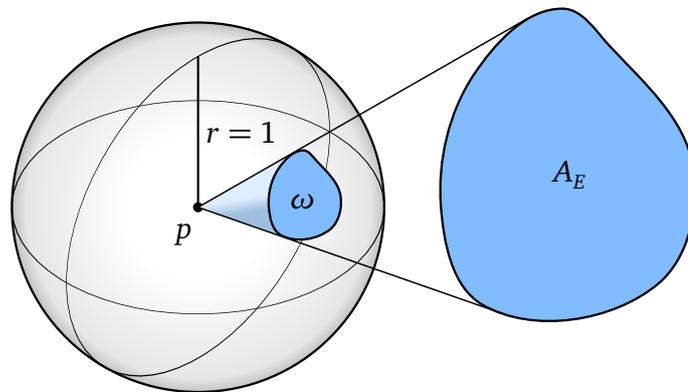
#### 2.1.1 Spektrum des sichtbaren Lichts

Elektromagnetische Wellen im Wellenlängenbereich von ca. 390 bis 790 nm werden vom menschlichen Auge als sichtbares „Licht“ wahrgenommen [DKV06]. Je nach Wellenlänge werden dabei unterschiedliche Farbreize empfunden (siehe auch Abb. 2.1). Violett bis blau werden im Bereich von etwa 390–480 nm empfunden. Wellenlängen von 480–560 nm werden als grün, 560–580 nm als gelb, 580–630 nm als orange und 630–790 nm als rot empfunden.

Das von Lichtquellen oder Körpern abgestrahlte Lichtspektrum ist eine Überlagerung von Licht unterschiedlicher Wellenlängen mit unterschiedlichen Intensitäten. Ein Spektrum, wie zum Beispiel Tageslicht, ist kontinuierlich, falls Licht aller sichtbaren Wellenlängen enthalten ist und keine abrupten Intensitätsabfälle vorliegen. Anderenfalls spricht man – wie etwa im Fall von Gasentladungslampen – von einem Linienspektrum, da das Spektrum nur



**Abbildung 2.1:** Farbanteile im Spektrum des sichtbaren Lichtes



**Abbildung 2.2:** Der Raumwinkel eines Flächenelements  $A_E$  für den Bezugspunkt  $p$  entspricht der auf eine Einheitskugel projizierten Fläche.

Licht weniger Wellenlängen („Intensitätsspitzen“) enthält. Falls sogar nur Licht einer bestimmten Wellenlänge enthalten ist, spricht man von einem monochromatischen Spektrum.

### 2.1.2 Raumwinkel

Der Raumwinkel bezeichnet das Verhältnis eines Flächenausschnitts  $A$  auf der Oberfläche einer Kugel zum Quadrat des Kugelradius  $r$ :

$$\omega = \frac{A}{r^2} \text{ [sr]} \quad (2.1)$$

Jedem Flächenelement  $A_E$  kann für einen Bezugspunkt ein Raumwinkel zugeordnet werden. Bei der Konstruktion einer Einheitskugel um den Bezugspunkt, ergibt sich der Raumwinkel aus dem Flächeninhalt des auf die Kugeloberfläche projizierten Flächenelements (siehe Abb. 2.2).

Trotz ihrer Eigenschaft als einheitenlose Verhältnisgröße wird für Raumwinkel die SI-Einheit Steradian (sr) verwendet, um so in physikalischen Größen die Abhängigkeit vom Raumwinkel auszudrücken. Der maximale Raumwinkel beträgt  $4\pi$  und erstreckt sich vollständig über die Oberfläche der gedachten Kugel.

### 2.1.3 Radiometrische Größen

#### Strahlungsleistung

Die Strahlungsleistung  $\Phi$ , auch als Strahlungsfluss bezeichnet, misst die gesamte pro Zeiteinheit transportierte Strahlungsmenge  $Q$  [ZZ11]:

$$\Phi = \frac{dQ}{dt} [W] \quad (2.2)$$

#### Strahlstärke

Für eine Strahlungsquelle ist nicht nur die insgesamt erzeugte Strahlungsleistung interessant, sondern auch deren Verteilung in verschiedene Ausstrahlrichtungen, welche durch die Strahlstärke beschrieben wird. Betrachtet man für eine als punktförmig angenommene Strahlungsquelle diejenige Strahlungsleistung  $d\Phi$ , die innerhalb eines bestimmten Raumwinkel  $d\omega$  abgestrahlt wird, dann ergibt sich die Strahlstärke  $I$  als deren Quotient [DKV06]:

$$I = \frac{d\Phi}{d\omega} \left[ \frac{W}{sr} \right] \quad (2.3)$$

#### Spezifische Ausstrahlung

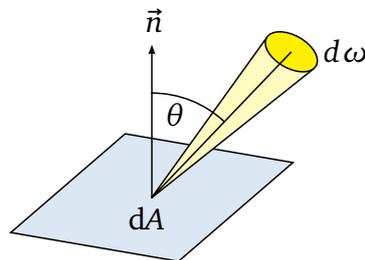
Die spezifische Ausstrahlung misst die von einem ebenen Flächenelement  $dA$  abgestrahlte Strahlungsleistung und ist als deren Quotient definiert:

$$M = \frac{d\Phi}{dA} \left[ \frac{W}{m^2} \right] \quad (2.4)$$

#### Bestrahlungsstärke

Analog zur spezifischen Ausstrahlung misst die Bestrahlungsstärke  $E$  die von einem ebenen Flächenelement  $dA$  empfangene Strahlungsleistung und ist als deren Quotient definiert:

$$E = \frac{d\Phi}{dA} \left[ \frac{W}{m^2} \right] \quad (2.5)$$



**Abbildung 2.3:** Strahldichte als Quotient aus Strahlungsleistung und dem Produkt aus Raumwinkel  $d\omega$  und projizierter Fläche  $\cos \theta dA$

### Strahldichte

Die Strahldichte  $L$  misst die Strahlungsleistung, welche in einem bestimmten Raumwinkel  $d\omega$  auf ein in Richtung des Eingangswinkels  $\theta$  projiziertes Flächenelement  $\cos \theta dA$  eintrifft (siehe Abb. 2.3):

$$L = \frac{d^2\Phi}{d\omega \cos \theta dA} \left[ \frac{W}{sr \cdot m^2} \right] \quad (2.6)$$

### Bestrahlung

Die Bestrahlung  $H$  beschreibt die über einen Zeitraum  $t$  insgesamt für ein Flächenelement  $dA$  empfangene Strahlungsleistung  $\Phi$ :

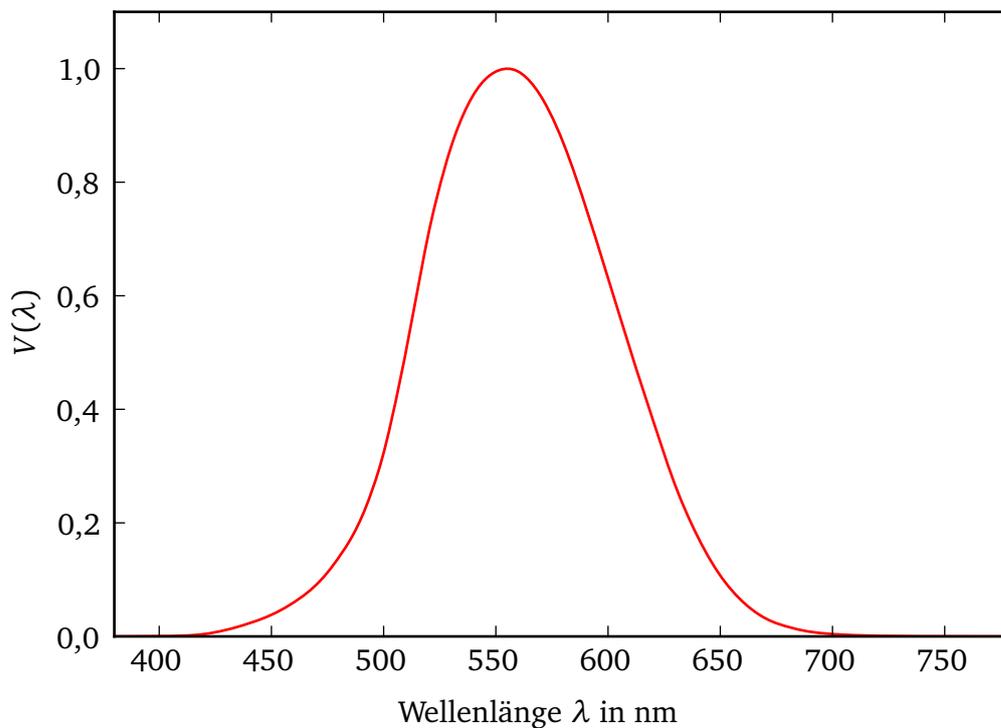
$$H = \int \frac{d\Phi}{dA} dt \left[ \frac{J}{m^2} \right] \quad (2.7)$$

Falls die Bestrahlungsstärke der Fläche über den gesamten Zeitraum konstant ist, vereinfacht sich der Zusammenhang folgendermaßen:

$$H = E t \quad (2.8)$$

## 2.1.4 Photometrische Größen

Für die menschliche Wahrnehmung ist nur die Messung von Licht – also des sichtbaren Spektrums elektromagnetischer Strahlung – interessant. Daher werden unter Verwendung der Empfindlichkeitsfunktion des helladaptierten



**Abbildung 2.4:** Empfindlichkeitskurve des hell adaptierten Auges  $V(\lambda)$

menschlichen Auges  $V(\lambda)$  (siehe Abb. 2.4) aus den radiometrischen Größen deren analoge photometrische Größe abgeleitet [BS04]. Um Verwechslungen bei gemeinsamen Auftreten zu vermeiden, wird dem Symbol radiometrischer Größen ein  $e$  und dem Symbol photometrischer Größen ein  $v$  angehängt.

Für die Ableitung des Lichtstroms  $\Phi_v$  wird die Strahlungsleistung  $\Phi_e$  als spektrale Größe, also als Funktion der Wellenlänge  $\Phi_e(\lambda)$  aufgefasst und mit  $V(\lambda)$  bewertet:

$$\Phi_v = K_m \int V(\lambda) \frac{d\Phi_e(\lambda)}{d\lambda} d\lambda \quad [lm] \quad (2.9)$$

Die Übersetzung von Watt in die Einheit des Lichtstroms, Lumen, ist durch das photometrische Strahlungsäquivalent  $K_m$  definiert:

$$K_m = 683 \frac{lm}{W} \quad (2.10)$$

**Tabelle 2.1:** Photometrische und radiometrische Größen

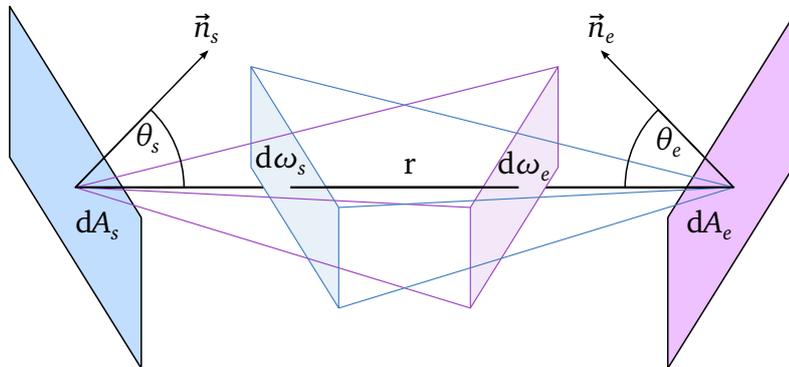
Größe	Symbol	Einheit
Strahlungsenergie	$Q_e$	$J$
Lichtmenge	$Q_v$	$lm \cdot s$
Strahlungsleistung	$\Phi_e$	$W$
Lichtstrom	$\Phi_v$	$lm$
Spezifische Ausstrahlung	$M_e$	$W \cdot m^{-2}$
Spezifische Lichtausstrahlung	$M_v$	$lm \cdot m^{-2}$
Strahlstärke	$I_e$	$W \cdot sr^{-1}$
Lichtstärke	$I_v$	$cd, lm \cdot sr^{-1}$
Strahldichte	$L_e$	$W \cdot m^{-2} \cdot sr^{-1}$
Leuchtdichte	$L_v$	$cd \cdot m^{-2}$
Bestrahlungsstärke	$E_e$	$W \cdot m^{-2}$
Beleuchtungsstärke	$E_v$	$lx, lm \cdot m^{-2}$
Bestrahlung	$H_e$	$J \cdot m^{-2}$
Belichtung	$H_v$	$lx \cdot s$

Wird eine radiometrische Strahlungsgröße anhand des Lichtstroms anstatt der Strahlungsleistung definiert, so ergibt daraus die analoge photometrische Größe (siehe Tabelle 2.1). Neben Lumen wird außerdem die Einheit Candela ( $1 \text{ cd} = 1 \text{ lm} \cdot \text{sr}^{-1}$ ) zur Messung der Lichtstärke sowie die Einheit Lux ( $1 \text{ lx} = 1 \text{ lm} \cdot \text{m}^{-2}$ ) zur Messung der Beleuchtungsstärke und der spezifischen Lichtausstrahlung eingeführt.

### 2.1.5 Photometrische Gesetzmäßigkeiten

Wie in Abbildung 2.5 dargestellt, wird der von einer Senderfläche  $dA_s$  an eine Empfängerfläche  $dA_e$  versandte Lichtstrom untersucht. Zwischen beiden Flächen besteht eine Sichtdistanz  $r$ . Beide Flächen sind nicht senkrecht orientiert, sondern die Senderfläche ist um den Winkel  $\theta_s$  und die Empfängerfläche um den Winkel  $\theta_e$  vom Sichtstrahl abgeneigt.

Aus Sicht einer der Flächen kann der jeweils gegenüber liegenden Fläche ein



**Abbildung 2.5:** Für photometrische Gesetzmäßigkeiten relevante geometrische Größen zwischen einer Sender- und Empfängerfläche

Raumwinkel –  $d\omega_s$  und  $d\omega_e$  – zugeordnet werden. Da differenzielle Flächen betrachtet werden, werden die Raumwinkel auf folgende Weise berechnet:

$$d\omega_s = \frac{\cos \theta_s dA_s}{r^2} \quad (2.11)$$

$$d\omega_e = \frac{\cos \theta_e dA_e}{r^2} \quad (2.12)$$

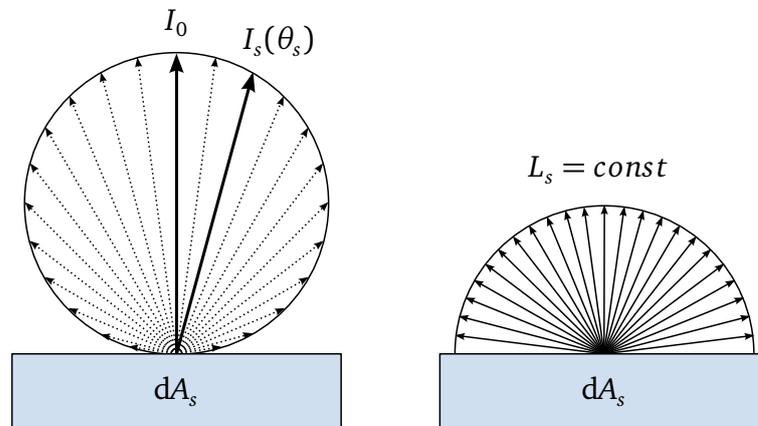
### Photometrisches Grundgesetz

Der von Sender- zur Empfängerfläche versandte Lichtstrom ergibt sich aus dem Produkt von Leuchtdichte der Senderfläche  $L_s$ , der aus Sicht der Empfängerfläche sichtbaren Senderfläche  $\cos \theta_s dA_s$  und dem Raumwinkel zur Empfängerfläche  $d\omega_e$ :

$$\Phi_{se} = L_s \cos \theta_s dA_s d\omega_e \quad (2.13)$$

Daraus leitet sich das *photometrische Grundgesetz* ab:

$$\Phi_{se} = L_s \frac{\cos \theta_s dA_s \cos \theta_e dA_e}{r^2} \quad (2.14)$$



**Abbildung 2.6:** Oberfläche als Lambert'scher Strahler mit einer Lichtstärke, welche ausschließlich von  $\cos \theta_s$  abhängt, und einer konstanten Leuchtdichte

### Photometrisches Entfernungsgesetz

Aus dem photometrischen Grundgesetz folgt, dass sich die Belichtungsstärke der Empfängerfläche in Abhängigkeit der Lichtstärke der Senderfläche und dem Neigungswinkel zur Senderfläche hin ergibt und mit dem Quadrat der Entfernung abnimmt. Dieser Zusammenhang wird als *photometrisches Entfernungsgesetz* bezeichnet.

$$E_e = \frac{dI_s \cos \theta_e}{r^2} \quad (2.15)$$

### Lambertsches Gesetz

Falls die Leuchtdichte der Senderfläche für jede Richtung  $\theta_s$  gleich ist, dann wird diese als *Lambert-Strahler* bezeichnet und es gilt:

$$L_s(\theta_s) = \text{const} \quad (2.16)$$

Zudem gilt der aus dem photometrischen Grundgesetz abgeleitete Zusammenhang zwischen Leuchtdichte und Lichtstärke:

$$L_s = \frac{I_s}{\cos \theta_s dA_s} \quad (2.17)$$

Da sowohl Leuchtdichte  $L_s$  als auch die Fläche  $dA_s$  konstant sind, ist die Lichtstärke  $I_s$  proportional zu  $\cos \theta_s$ . Dieser Zusammenhang wird als Lambertsches Gesetz bezeichnet (siehe Abb. 2.6):

$$I_s(\theta_s) = I_0 \cos \theta_s \quad (2.18)$$

Das Produkt aus beiden Konstanten  $L_s$  und  $dA_s$  ist die maximale, also in senkrechter Richtung abgestrahlte Lichtstärke  $I_0$ .

Ohne Annahme eines Lambertschen Strahlers ergibt sich die spezifische Lichtausstrahlung einer Senderfläche  $M$  allgemein aus dem Integral seiner richtungsabhängigen Leuchtdichte  $L(\phi, \theta)$  über einen Raumwinkel von  $2\pi sr$ , also allen Richtungen, die von der Senderfläche weg zeigen:

$$M = \int_0^{2\pi} \int_0^{\pi/2} L(\phi, \theta) \cos \theta \sin \theta \, d\theta \, d\phi \quad (2.19)$$

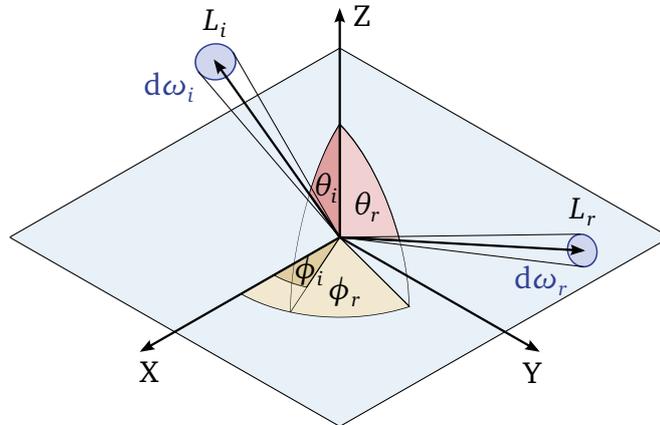
Für einen Lambertschen Strahler ist die Leuchtdichte für alle Richtungen konstant und kann als Konstante vor beide Integrale verschoben werden. Daraus folgt:

$$\begin{aligned} M &= L \int_0^{2\pi} \int_0^{\pi/2} \cos \theta \sin \theta \, d\theta \, d\phi \\ M &= L \pi \end{aligned} \quad (2.20)$$

### 2.1.6 Beschreibung von Reflexionsverhalten

Je nach Beschaffenheit ihrer Oberfläche weisen Materialien wie Holz, Metall oder Plastik ein charakteristisches Reflexionsverhalten auf, das vom menschlichen Auge zum Beispiel als matt oder glänzend wahrgenommen wird. Reflexionsverhalten kann je nach dessen Komplexität durch geeignete Modelle beschrieben werden.

BRDFs (*Bidirectional reflectance distribution functions*) beschreiben das Reflexionsverhalten einer Oberfläche als Quotient aus ausgesendeter Strahldichte  $L_r$  und eingestrahler Bestrahlungsstärke  $E_i$  in Abhängigkeit einer Eingangsrichtung  $(\phi_i, \theta_i)$  und einer Ausgangsrichtung  $(\phi_r, \theta_r)$  (siehe Abb. 2.7):



**Abbildung 2.7:** Durch Kugelkoordinaten parametrisierte ein- und ausgehende Strahldichten

$$f_r(\phi_i, \theta_i, \phi_r, \theta_r) = \frac{L_r(\phi_r, \theta_r)}{E_i(\phi_i, \theta_i)} = \frac{L_r(\phi_r, \theta_r)}{L_i(\phi_i, \theta_i) \cos \theta_i d\omega_i} \left[ \frac{1}{sr} \right] \quad (2.21)$$

Eine BRDF beschreibt demnach, wie eine aus bestimmter Richtung eingestrahlte Leuchtdichte für eine bestimmte Ausgangsrichtung in welchem Maße reflektiert wird.

Für Lambertsche Strahler besteht zwischen ausgesendeter und eingestrahlter Leuchtdichte keine Richtungsabhängigkeit. Stattdessen wird das Reflexionsverhalten ausschließlich durch den *Reflexionsgrad*  $\rho$  bestimmt. Dieser kann als Quotient aus dem aus- und eingestrahlttem Lichtstrom oder der spezifischen Lichtausstrahlung  $M$  und der Bestrahlungsstärke  $E$  für die gleiche betrachtete Fläche  $dA$  definiert werden:

$$\rho = \frac{\Phi_r}{\Phi_i} = \frac{M}{E} \quad (2.22)$$

Da für Lambertsche Strahler  $M = \pi L_r$  gilt (siehe Abschnitt 2.1.5), ist deren BRDF demnach:

$$f_r = \frac{L_r}{E_i} = \frac{M}{\pi E_i} = \frac{\rho}{\pi} \quad (2.23)$$

## 2.2 Farbmétrische Grundlagen

Die menschliche Wahrnehmung von Farben basiert auf der Bewertung der spektralen Zusammensetzung von Licht durch das menschliche Auge. Das eingestrahelte Licht stellt dabei einen Farbreiz dar, welchem durch das menschliche Auge eine Farbvalenz zugeordnet wird. Ein Farbreiz kann durch eine Strahldichte als Funktion der Wellenlänge,  $L(\lambda)$ , im Bereich des sichtbaren Lichtes (ca. 380 bis 780 nm) beschrieben werden:

$$L(\lambda) = \frac{d^2\Phi(\lambda)}{d\omega \cos\theta dA} \left[ \frac{W}{m^2 \cdot sr} \right] \quad (2.24)$$

In der Praxis liegen Lichtspektren meist nicht als analytische Funktionen vor, sondern ein Lichtspektrum wird in einem gleichmäßigen Abstand der Wellenlänge abgetastet. Bei einer Wahl von  $n$  Abtastwerten und einem Abtastintervall von  $[\lambda_{min}, \lambda_{max}]$  beträgt der Abstand zwischen benachbarten Wellenlängen:

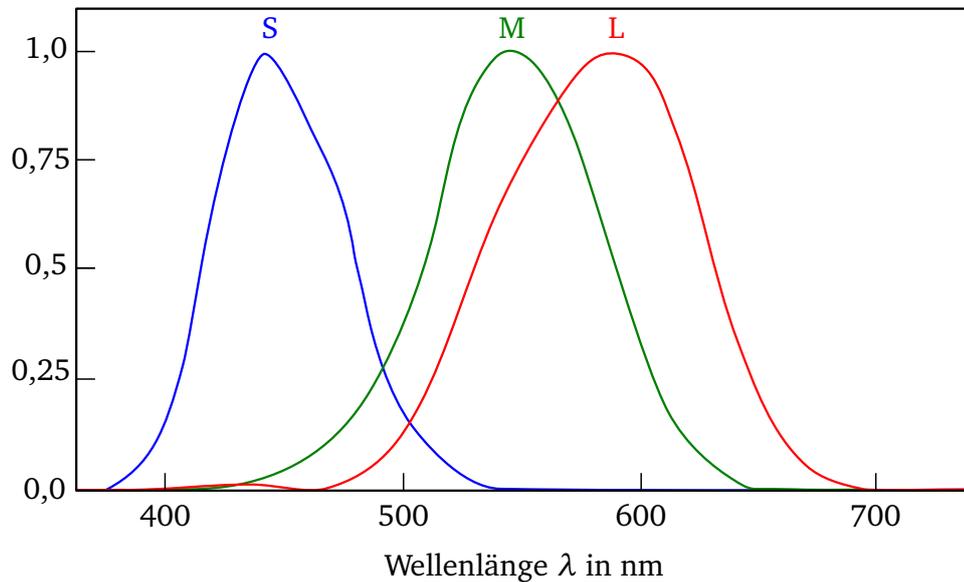
$$\Delta\lambda = \frac{\lambda_{max} - \lambda_{min}}{n - 1} \quad (2.25)$$

Ein so diskretisiertes Lichtspektrum kann als Vektor  $\vec{L}$  eines  $n$ -dimensionalen Vektorraums aufgefasst werden. Dieser wird als Spektralraum bezeichnet, da er sämtliche Lichtspektren enthält und jedem Spektrum ein Ortsvektor zugeordnet wird.

### 2.2.1 CIE-Normspektralraum

Die physiologische Wahrnehmung von Licht wird durch die Photorezeptoren bestimmt, welche sich in der Netzhaut des menschlichen Auges befinden und aus Stäbchen sowie Zapfen bestehen. Stäbchen sind sehr lichtempfindlich und für das unbunte Nachtsehen verantwortlich. Die Farbwahrnehmung wird durch drei verschiedene Typen von Zapfen bestimmt. Die drei Zapfentypen haben jeweils eigene spektrale Absorptionskurven mit einem spezifischen Intensitätsmaximum und werden als L-, M- und S-Zapfen bezeichnet (siehe Abb. 2.8).

Durch die Bewertung eines Lichtspektrums mit den spektralen Absorptionskurven der drei Zapfentypen ergibt sich eine Farbvalenz als Vektor mit drei Komponenten. Diese Abbildung von  $n$ -dimensionalen Lichtspektren, also

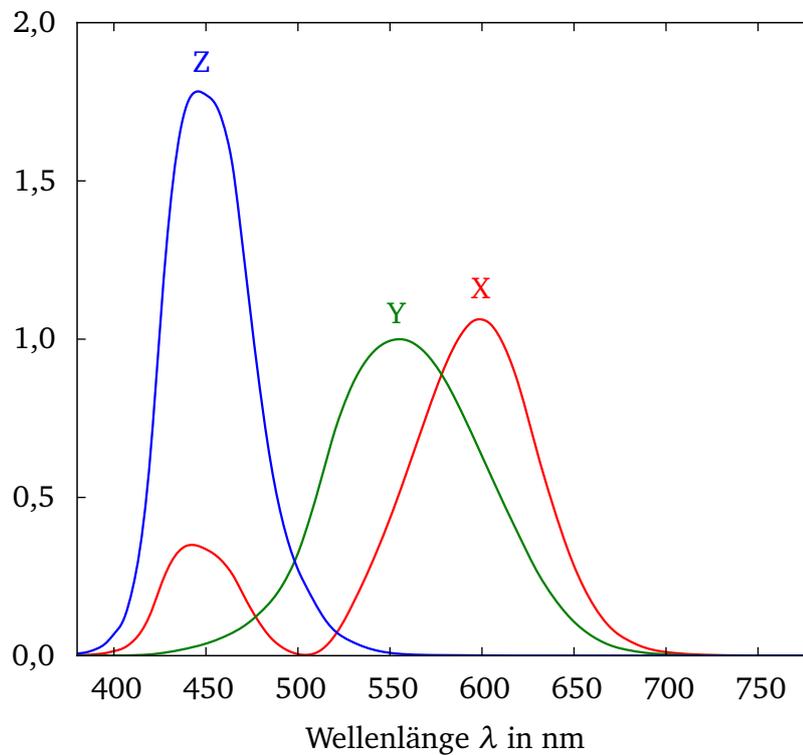


**Abbildung 2.8:** Näherungsweise Wiedergabe der spektralen Empfindlichkeit der L, M, und S-Zapfen des menschlichen Auges [HP11]

den Farbreizen, auf die dreidimensionalen Farbvalenzen wird durch das *CIE-Normvalenzsystem* für einen Normalbeobachter beschrieben. Die Ausführungen im folgenden Abschnitt beziehen sich dabei auf den CIE-Standard von 1931 für ein Sichtfeld von  $2^\circ$ .

Das CIE-Normvalenzsystem kann als dreidimensionaler Unterraum des  $n$ -dimensionalen Spektralraums verstanden werden. Es bildet sämtliche physikalisch möglichen Farbreize auf dreidimensionale Farbvalenzen ab, die somit eindeutig als XYZ-Koordinaten parametrisiert werden. Diese Abbildung wird durch die Normspektralwertfunktionen  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$  und  $\bar{z}(\lambda)$  definiert, welche Linearkombinationen der spektralen Absorptionskurven sind und in Form  $n$  diskretisierter Werte als Basisvektoren des CIE XYZ-Koordinatensystem im Spektralraum dienen (siehe Abb. 2.9).

Die Normspektralwertfunktionen sind so gewählt, dass alle Funktionswerte größer als 0 sind und  $\bar{y}(\lambda)$  der spektralen Empfindlichkeitsfunktion des hell adaptierten Auges,  $V(\lambda)$ , entspricht [Sha03]. Eine Abbildung einer spektralen Strahlendichte  $L(\lambda)$  in XYZ-Koordinaten berechnet sich folgendermaßen:



**Abbildung 2.9:** CIE-Normspektralwertfunktionen

$$\begin{aligned}
 X &= k \int L(\lambda) \bar{x}(\lambda) d\lambda \\
 Y &= k \int L(\lambda) \bar{y}(\lambda) d\lambda \\
 Z &= k \int L(\lambda) \bar{z}(\lambda) d\lambda
 \end{aligned}
 \tag{2.26}$$

Der Faktor  $k$  bestimmt die Skalierung der XYZ-Werte. Im Fall absoluter XYZ-Koordinaten entspricht  $k$  dem photometrischen Strahlungsäquivalent  $K_m$  (683 lm/W). Damit kann an der Y-Koordinate eines absoluten XYZ-Werts die zugehörige Leuchtdichte in Lumen abgelesen werden.

Werden Normspektralwertfunktionen und Lichtspektren als n-dimensionale

Vektoren betrachtet, kann die Ermittlung der XYZ-Koordinaten  $\vec{L}_{XYZ}$  einfach als lineare Abbildung durch eine  $n \times 3$ -Matrix  $B$  formuliert werden:

$$B = (\vec{x} \quad \vec{y} \quad \vec{z}) \quad (2.27)$$

$$\vec{L}_{XYZ} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = B^T \cdot \vec{L} = \begin{pmatrix} \vec{x} \cdot \vec{L} \\ \vec{y} \cdot \vec{L} \\ \vec{z} \cdot \vec{L} \end{pmatrix} = \quad (2.28)$$

Dabei ist zu beachten, dass die Vektoren  $\vec{x}$ ,  $\vec{y}$  und  $\vec{z}$  sich aus den diskretisierten Normspektralwertfunktionen ergeben, aber zusätzlich noch mit  $\Delta\lambda$  multipliziert werden, um mit dem Skalarprodukt eine numerische Integration durchzuführen und so absolute XYZ-Werte zu erhalten.

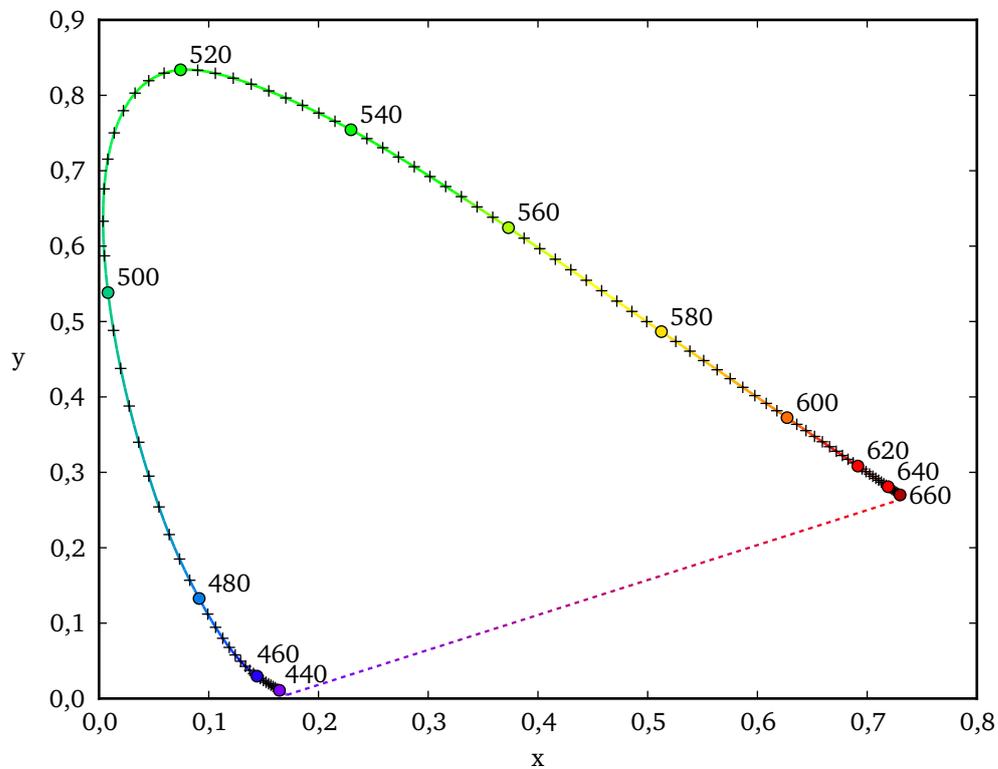
### 2.2.2 CIE-Normfarbtafel

Um die Darstellung von dreidimensionalen CIE-Tristimuli zu vereinfachen, können diese in die zweidimensionale CIE-Normfarbtafel (auch als xy-Chromatizitätsdiagramm bezeichnet) abgebildet werden (siehe Abb. 2.10). Dabei werden die XYZ-Tristimuli so transformiert, dass  $x + y + z = 1$  gilt:

$$\begin{aligned} x &= \frac{X}{X + Y + Z} \\ y &= \frac{Y}{X + Y + Z} \\ z &= \frac{Z}{X + Y + Z} = 1 - x - y \end{aligned} \quad (2.29)$$

### 2.2.3 CIE-L\*a\*b\*-Farbraum

Das CIE-Normvalenzsystem bietet eine eindeutige Repräsentation von Farbvalenzen, da verschiedene Farbvalenzen auch auf unterschiedliche XYZ-Koordinaten abgebildet werden und gleiche Farbvalenzen auf identische XYZ-Koordinaten. Jedoch ist die Distanz zwischen zwei verschiedenen XYZ-Koordinaten kein gutes Maß für den vom Menschen wahrgenommenen Farbumterschied. Ein Paar von Farbvalenzen, die so ähnlich sind, dass sie vom



**Abbildung 2.10:** CIE-Normfarbtafel mit eingetragenen monochromatischen Spektren

menschlichen Auge gerade noch unterschieden werden können, erhalten je nach Lage im CIE-Normvalenzsystem sehr unterschiedliche Distanzen.

Der CIE-L\*a\*b\*-Farbraum baut auf dem CIE-Normvalenzsystem auf, aber bildet unterschiedliche Farbvalenzen so ab, dass ihr Abstand eher dem wahrgenommenen Farbabstand entspricht. Die Farbwerte werden dazu in ein dreidimensionales Koordinatensystem mit den Achsen L\*, a\* und b\* auf folgende Weise transformiert:

$$\begin{aligned}
 L^* &= 116 \cdot \sqrt[3]{\frac{Y}{Y_n} - 16} \\
 a^* &= 500 \cdot \sqrt[3]{\frac{X}{X_n} - \frac{Y}{Y_n}} \\
 b^* &= 200 \cdot \sqrt[3]{\frac{Y}{Y_n} - \frac{Z}{Z_n}}
 \end{aligned} \tag{2.30}$$

Die Transformation erfolgt in Abhängigkeit eines in XYZ-Koordinaten angegebenen Weißpunktes,  $(X_n, Y_n, Z_n)$ . Im  $L^*a^*b^*$ -Farbraum ist der Farbabstand von zwei Farbvalenzen  $\Delta E_{ab}^*$  dann als die euklidische Distanz der  $L^*a^*b^*$ -Koordinaten definiert:

$$\Delta E_{ab}^* = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \tag{2.31}$$

## 2.2.4 Geräteabhängige Farbräume

Die Farbwiedergabe von digitalen Kameras wird durch die spektralen Sensorantwortkurven des verwendeten Bildsensors bestimmt. Ähnlich den Photorezeptoren des menschlichen Auges wird die spektrale Zusammensetzung von Licht, welches auf den Bildsensor trifft, von diesem bewertet und auf RGB-Farbtripel, also ebenfalls auf dreidimensionale Farbkoordinaten, abgebildet. Analog zu den Normspektralwertfunktionen spannen die Sensorantwortkurven einen dreidimensionalen RGB-Farbraum auf. Die Sensorantwortkurven sind allerdings je nach verwendeter Kamera unterschiedlich und können sogar zwischen zwei gleichen Kameramodellen variieren. Daher ist der zugehörige RGB-Farbraum der Kamera *geräteabhängig*.

Wie ähnlich die Farbwiedergabe einer Kamera zur Farbwahrnehmung des menschlichen Auges ist, hängt vom spektralen Verlauf der Sensorantwortkurven ab: Falls die Sensorantwortkurven eine Linearkombination der Normspektralwertfunktionen darstellen, bildet die Kamera Lichtspektren auf den gleichen dreidimensionalen Unterraum wie das CIE-Normvalenzsystem ab und die RGB-Werte können durch eine Transformation mit einer für den RGB-Farbraum spezifischen  $3 \times 3$ -Matrix exakt auf XYZ-Werte abgebildet

werden. Diese Bedingung wird als Luther-Ives-Bedingung [Lut27] [Ive15] bezeichnet und ist eine fundamentale Anforderung, um einen geräteabhängigen RGB-Farbraum ohne Farbfehler in das CIE-Normvalenzsystem abzubilden.

Die Luther-Ives-Bedingung kann auch in Vektornotation ausgedrückt werden:

$$\begin{aligned} B &= (\vec{x} \quad \vec{y} \quad \vec{z}) \\ R &= (\vec{r} \quad \vec{g} \quad \vec{b}) \\ B &= R \cdot M \end{aligned} \tag{2.32}$$

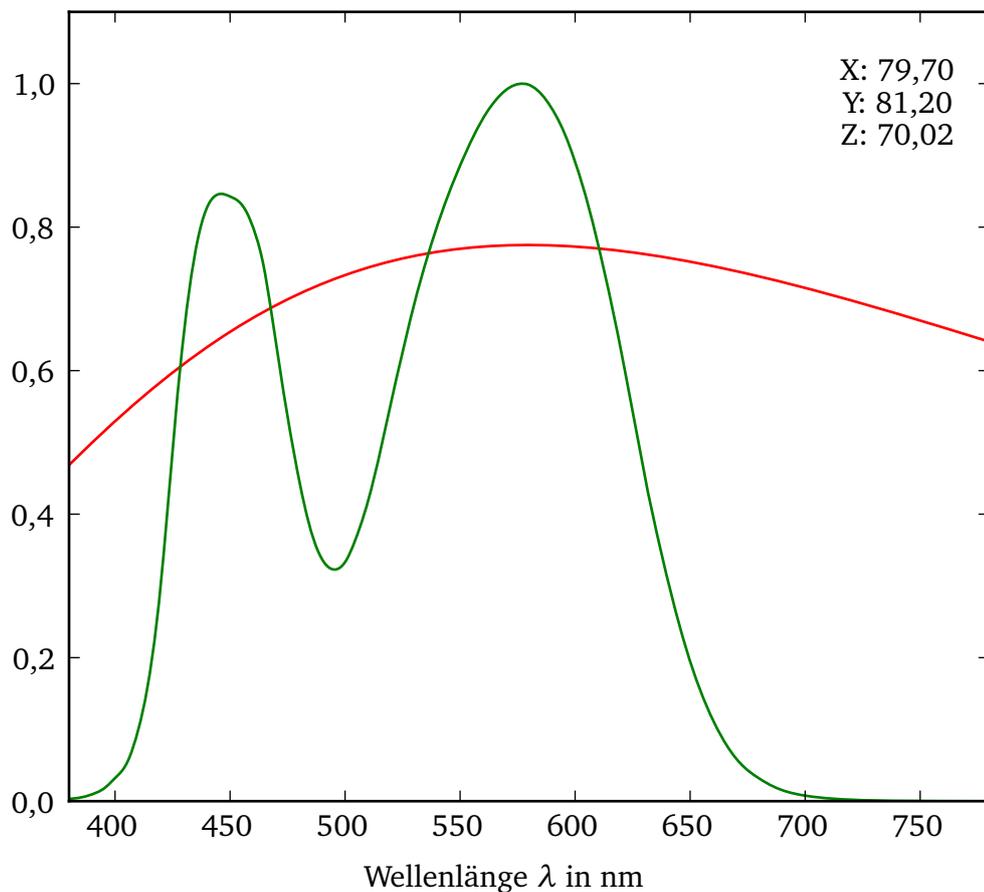
Dabei beschreibt die  $n \times 3$ -Matrix  $R$  die Abbildung in den RGB-Farbraum durch die drei diskretisierten Sensorantwortkurven  $\vec{r}$ ,  $\vec{g}$  und  $\vec{b}$ . Die  $n \times 3$ -Matrix  $B$  beschreibt die Abbildung in den CIE-XYZ-Farbraum. Die Luther-Ives-Bedingung ist erfüllt, falls  $R$  in  $B$  mit Hilfe der  $3 \times 3$ -Matrix  $M$  transformiert werden kann.

Die Luther-Ives-Bedingung wird von den meisten Kameras nicht erfüllt, so dass RGB-Raum und XYZ-Raum nicht den gleichen Unterraum im Spektralraum bilden und ein systematischer Farbfehler vorliegt. Die falsche spektrale Bewertung der Kamera kann zwar nicht abgeändert werden, jedoch kann durch eine Farbcharakterisierung die Abbildung von RGB-Koordinaten auf XYZ-Koordinaten verbessert werden.

### 2.2.5 Farbmémerie

Wenn zwei Lichtspektren auf die gleichen XYZ-Koordinaten abgebildet werden, also auch den gleichen Farbeindruck erzeugen, spricht man von *Metameren* (siehe Abb. 2.11). Mémerie basiert auf der Tatsache, dass Farbvalenzen lediglich dreidimensionale Projektionen wesentlich höher dimensionaler Lichtspektren sind.

Mémerie tritt auch bei geräteabhängigen Farbräumen auf: Falls unterschiedliche Lichtspektren mit unterschiedlichen XYZ-Werten durch eine Kamera mit dem gleichen RGB-Wert wiedergegeben werden, spricht man von *Gerätémémerie*. Da verschiedene XYZ-Werte auf einen RGB-Wert abgebildet werden, kann ohne Kenntnis der ursprünglichen Lichtspektren die Abbildung nicht invertiert werden, es können also nicht die originalen Farbvalenzen rekonstruiert werden.

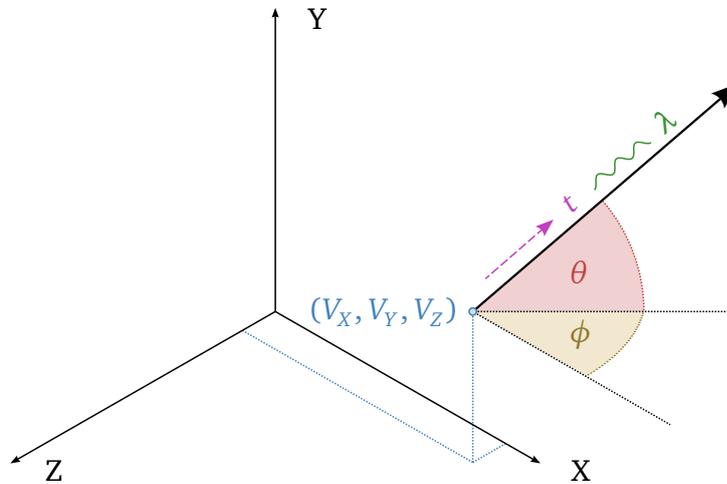


**Abbildung 2.11:** Zwei metamere Lichtspektren mit gleichem XYZ-Wert

## 2.3 Bildbasierte Repräsentation von Szenen

### 2.3.1 Plenoptische Funktion

Die plenoptische Funktion ist ein strahlenoptisches Modell, welches das in einer Szene vorhandene Licht beschreibt. Sie wurde von Adelson und Bergen eingeführt [AB91] und beschreibt die Intensität empfangenen Lichtes für alle Positionen  $(V_x, V_y, V_z)$ , alle Richtungen  $(\phi, \theta)$ , alle Wellenlängen  $\lambda$  und alle Zeitpunkte  $t$  (siehe Abb. 2.12):



**Abbildung 2.12:** Darstellung der Parameter der plenoptischen Funktion: Position  $(V_x, V_y, V_z)$ , Einstrahlrichtung  $(\theta, \phi)$ , Wellenlänge  $(\lambda)$  und Zeit  $(t)$

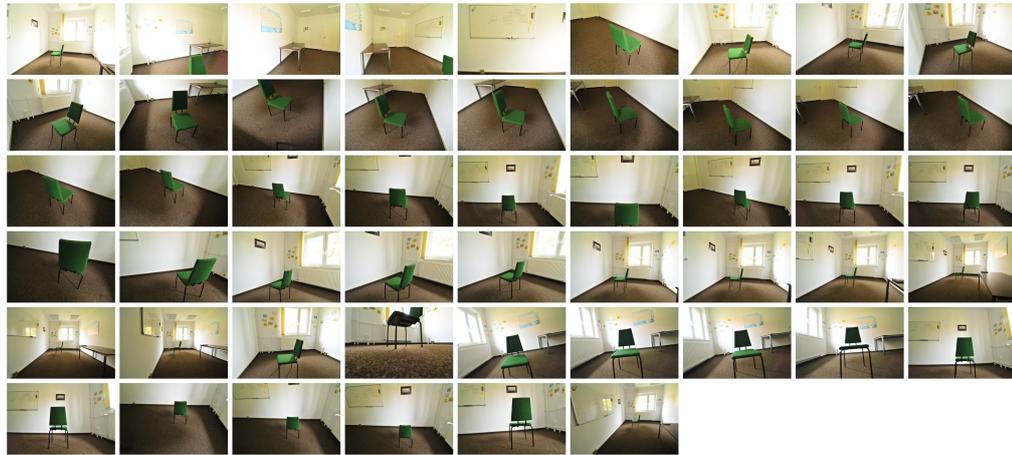
$$P(\theta, \phi, \lambda, V_x, V_y, V_z, t) \quad (2.33)$$

Wie in [SCK07] erläutert, können sämtliche Verfahren des Image-based Rendering als verschiedene Strategien gesehen werden, die plenoptische Funktion einer Szene (egal, ob virtuell oder real) diskret abzutasten und anschließend zur Erzeugung neuer Ansichten eine kontinuierliche Rekonstruktion durchzuführen.

Im Fall eines einzelnen Bildes wird die plenoptische Funktion an einer festen Kameraposition für ein rechteckiges Bündel von Richtungen abgetastet und reduziert sich somit auf die zwei Parameter  $\phi$  und  $\theta$ . Im Fall von 3D-Photo-Collections wird die plenoptische Funktion an den Positionen der Kameras für ein rechteckiges Bündel von Richtungen abgetastet, wobei das empfangene Licht in Form eines RGB-Tripels beschrieben wird.

### 2.3.2 Erzeugung von 3D-Photo-Collections

In dieser Arbeit werden bildbasierte Repräsentationen von realen Szenen verwendet, um Wege zur bildbasierten Beleuchtung aufzuzeigen. Diese bildbasierten Repräsentationen werden unter Verwendung von ungeordneten Fotos



**Abbildung 2.13:** Eingabebilder einer 3D-Photo-Collection

einer Szene vollständig automatisch durch spezielle Verfahren der Photogrammetrie erzeugt und als *3D-Photo-Collections* bezeichnet. Im Rahmen dieser Arbeit wird dazu die Software Bundler [Sna] verwendet. Im folgenden Abschnitt soll kurz auf die wichtigsten Arbeitsschritte von Bundler eingegangen werden, um einen Überblick über die Erstellung von 3D-Photo-Collections zu geben.

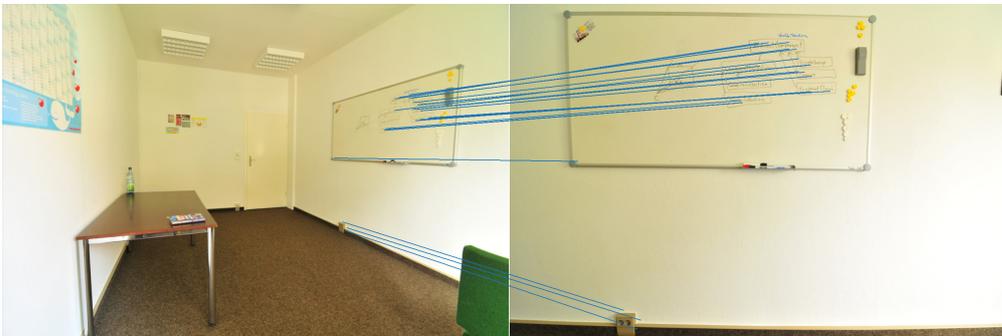
Bundler verwendet zunächst einen Merkmalsdetektor, um Merkmalspunkte für jedes Bild zu generieren. Anschließend werden Bilder paarweise einer Korrespondenzanalyse von Merkmalspunkten unterzogen und anhand der korrespondierenden Merkmalspunkte wird ein iteratives Bündelblockausgleichsverfahren durchgeführt, um Positionen der Merkmalspunkte, sowie in- und extrinsische Kameraparameter für jedes Bild zu rekonstruieren.

### Detektion von Merkmalspunkten

Mit Hilfe des SIFT-Detektors [Low04] werden im ersten Schritt für jedes der Eingabebilder Merkmalspunkte ermittelt (siehe Abb. 2.13 und Abb. 2.14). SIFT liefert dabei Merkmalspunkte, welche bis zu einem gewissen Grad skalen- und rotationsinvariant sind, wodurch auch bei Bildern mit abweichenden Kameraperspektiven auf die Szene eine Schnittmenge von gleichen Merkmalspunkten gefunden wird. Auch in Bildern mit unterschiedlichen Belichtungs- werten oder abweichenden Tonwertkurven können gleiche Merkmalspunkte



**Abbildung 2.14:** Durch SIFT erkannte Merkmalspunkte in zwei Bildern



**Abbildung 2.15:** Durch Bundler in zwei Bildern gefundene korrespondierende Merkmalspunkte

gefunden werden, insofern ähnliche Bereiche der Szene abgebildet sind. Erst diese robuste Merkmalserkennung ermöglicht die Rekonstruktion von Kameraposen aus Bildern, welche von unterschiedlichen Kameras stammen oder etwa mit unterschiedlichen Belichtungszeiten aufgenommen wurden.

### Suche nach korrespondierenden Merkmalspunkten

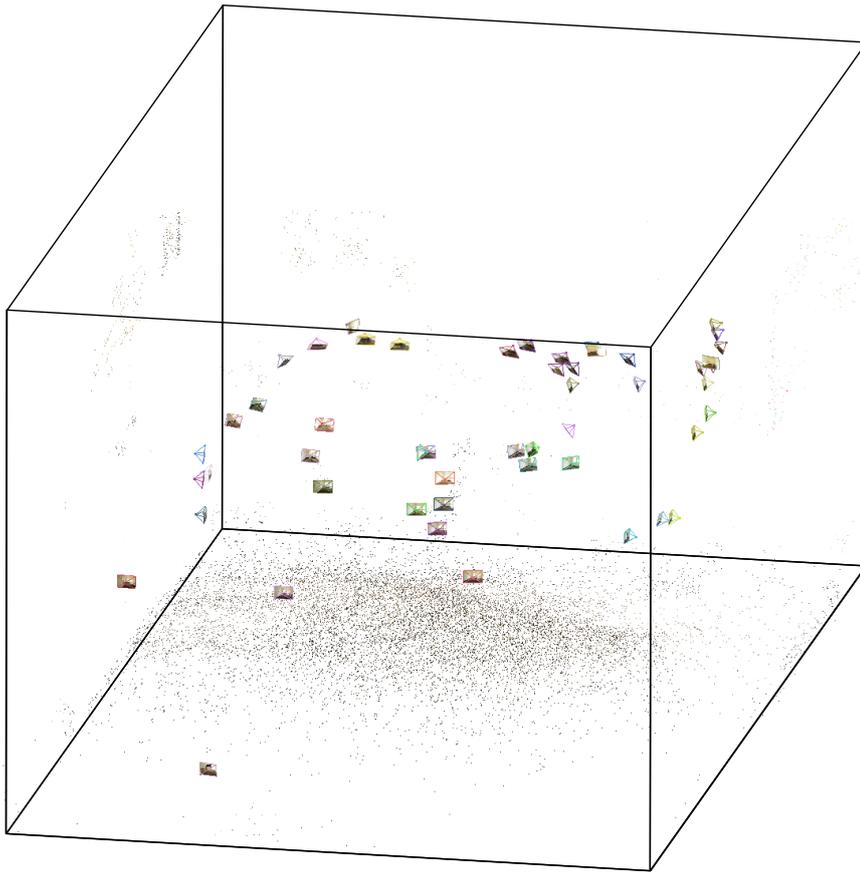
Anschließend werden die gefundenen Merkmalspunkte zwischen zwei Bildern anhand ihrer von SIFT für jeden Merkmalspunkt erzeugte Deskriptoren verglichen. Dazu wird für jeden Merkmalsdeskriptor des einen Bildes der ähnlichste Merkmalsdeskriptor im anderen Bild gesucht, was auf Grund der 128 Dimensionen der Deskriptoren über eine approximative Suche nach dem

nächsten Nachbarn durchgeführt wird.

Aus der Menge gefundener Korrespondenzen wird über RANSAC für jedes Bildpaar geprüft, welche Schnittmenge von Korrespondenzen geometrisch konsistent ist, indem für einzelne RANSAC-Iterationen durch Verwendung des 8-Punkt-Algorithmus die Fundamentalmatrix näherungsweise berechnet wird und Ausreißer entfernt werden. Ergebnis ist schließlich eine Menge von Korrespondenzen zwischen allen Bildpaaren (siehe Abb. 2.15).

### **Bündelblockausgleich**

Im Anschluss erfolgt die Rekonstruktion von Kameraposen und Punktwolke. Dafür wird nach den in- und extrinsischen Kameraparametern sowie den 3D-Positionen der Merkmalspunkte gesucht, welche den geringsten Reprojektionsfehler aufweisen. Es handelt sich dabei um ein als Bündelblockausgleich bezeichnetes nicht-lineares Optimierungsproblem, was von Bundler durch einen iterativen Ansatz unter Verwendung des Levenberg-Marquardt-Algorithmus gelöst wird. Begonnen wird zunächst mit der Rekonstruktion von Kameraparametern für zwei Bilder und anschließend wird mit jedem Iterationsschritt für ein weiteres Bild dessen Kameraparameter abgeschätzt. Nach Abschluss liegen für sämtliche Bilder die rekonstruierten Kameraparameter sowie die 3D-Positionen aller Merkmalspunkte vor (siehe Abb. 2.16).



**Abbildung 2.16:** 3D-Photo-Collection mit bekannten Kameraposen und 3D-Punktwolke



# Kapitel 3

## Farbmetrische Charakterisierung von bildbasierten Szenen

### 3.1 Einleitung

#### 3.1.1 Problemstellung

Die Bilder einer 3D-Photo-Collection liegen als Rastergrafik vor, in denen die Farbwerte von Bildpixeln als RGB-Tripel wiedergegeben werden. Die RGB-Werte sind das Ergebnis der von einer Kamera durchgeführten Messung und Bewertung von einfallenden spektralen Strahldichten. Diese Bewertung von Strahldichten mit RGB-Werten wird durch verschiedene Einflüsse wie die optische Abbildung durch das Kameraobjektiv auf die Bildsensoroberfläche und die spektrale Bewertung durch die Sensorantwortkurven des Bildsensors bestimmt (siehe Abschnitt 2.2). Zusätzlich werden resultierende RGB-Werte noch durch Schritte wie Weißabgleich, Farb- und Tonwertkorrekturen sowie Rauschreduktion beeinflusst. Da je nach verwendetem Bildsensor die Sensorantwortkurven einen unterschiedlichen Kurvenverlauf aufweisen, bilden die jeweiligen Kameramodelle Lichtspektren in unterschiedliche gerätespezifische RGB-Farbräume ab.

Falls für eine 3D-Photo-Collection die Bilder mit mehreren Kameramodellen fotografiert wurden, bedeutet dies, dass die Bilder auch in unterschiedlichen Farbräumen liegen. Erschwerend kommt hinzu, dass die Bilder in der Regel mit unkalibrierten Kameras fotografiert wurden, daher kann je nach gewählter Belichtungszeit und Blendenzahl die Wiedergabe von RGB-Werten auch für Bilder des gleichen Kameramodells unterschiedlich ausfallen.

Alle Bilder der 3D-Photo-Collection müssen jedoch in einem einheitlichen

Farbraum liegen, falls Verfahren zum Einsatz kommen, die korrespondierende Farbinformationen aus unterschiedlichen Bildern verwenden, wie zum Beispiel die Rekonstruktion der realen Beleuchtung in Kapitel 4 oder die bildbasierte Darstellung in Kapitel 5.

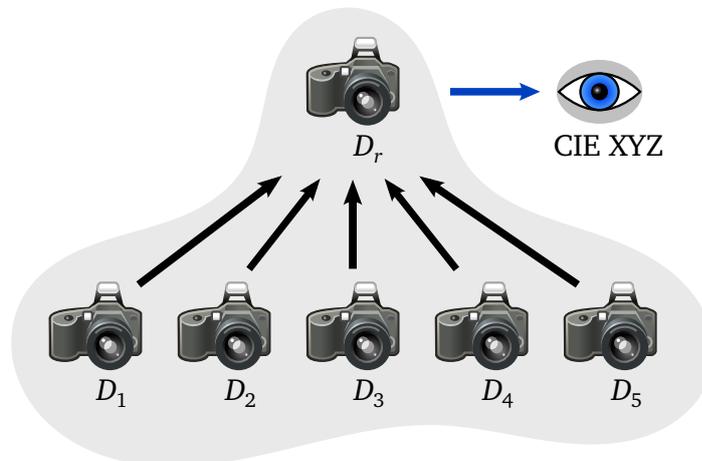
Zudem müssen für eine farbmetrisch korrekte Darstellung sämtliche Bilder mittels einer Charakterisierung der Kameramodelle in das geräteunabhängige CIE-Normvalenzsystem transformiert werden. Für die Darstellung auf einem Bildschirm oder den Ausdruck auf einem Drucker kann dann über ein entsprechendes Farbprofil in den geräteabhängigen Farbraum des Ausgabegeräts farbmetrisch korrekt umgewandelt werden.

Klassische empirische Methoden zur Farbcharakterisierung (siehe Abschnitt 3.2) verwenden Color-Checker, Farbtafeln mit bekanntem XYZ-Tristimuli, welche bei ungefähr bekannter Umgebungsbeleuchtung fotografiert werden. Anschließend werden teilautomatisch die Zusammenhänge zwischen den RGB-Werten der fotografierten Proben und den Tristimuli durch ein geeignetes Modell, im einfachsten Fall eine  $3 \times 3$ -Transformationsmatrix, modelliert.

Dieser Schritt muss allerdings für jede Kamera durchgeführt werden. Falls die Bilder einer 3D-Photo-Collection von vielen verschiedenen Kameras stammen, kann daher eine separate Kameracharakterisierung sehr aufwändig sein. Zudem müssen kalibrierte Kameras vorliegen, womit Bilder mit variierender Belichtungszeit und Blendenzahl nicht genutzt werden können.

### **3.1.2 Charakterisierung in 3D-Photo-Collections**

In diesem Kapitel wird ein Verfahren zur Vereinfachung der Farbcharakterisierung in 3D-Photo-Collections vorgestellt. Anstatt für jedes Kameramodell eine separate Farbcharakterisierung durchzuführen, wird eine Referenzkamera ausgewählt und die RGB-Farbräume aller anderen Kameramodelle werden in den RGB-Farbraum der Referenzkamera transformiert. Die RGB-Transformation in einen einheitlichen Referenz-Farbraum soll automatisch für jedes Kameramodell bestimmt werden. Die eigentliche Farbcharakterisierung in das CIE-Normvalenzsystem muss dann lediglich für die Referenzkamera durchgeführt werden (siehe Abb. 3.1). Die so gewonnenen Farbcharakterisierungsinformationen können nicht nur für die Referenzkamera, sondern in Kombination mit der jeweiligen RGB-Transformation auch für alle anderen Kameramodelle verwendet werden.



**Abbildung 3.1:** Charakterisierung in einen einheitlichen RGB-Farbraum mit optionaler Charakterisierung in das CIE-Normvalenzsystem im Anschluss

Viele Verfahren – darunter auch die bildbasierte Beleuchtung und bildbasierte Darstellung – können bereits gute Ergebnisse erzielen, wenn die Bilder der 3D-Photo-Collection in einem zwar geräteabhängigen, aber gemeinsamen RGB-Farbraum vorliegen. Die Charakterisierung in das CIE-Normvalenzsystem kann also entfallen, falls keine farbmetrisch korrekte Darstellung erforderlich ist.

Für eine empirische Charakterisierung müssen die Bilder der 3D-Photo-Collection eine reale Szene mit konstanter Beleuchtung abbilden, da die Bewertung der Farbwiedergabe unterschiedlicher Kameramodelle nur möglich ist, wenn alle Kameramodelle die gleichen Strahldichten empfangen haben. Bundler liefert eine 3D-Punktwolke, in der jedem Punkt eine Reihe von durch SIFT detektierten Merkmalspunkten aus unterschiedlichen Bildern zugeordnet ist. Auf Grund der Robustheit gegenüber Helligkeits- und Farbtonunterschieden, liefert der SIFT-Detektor auch in Bildern aus verschiedenen RGB-Farbräumen korrespondierende Merkmalspunkte. Wenn an den Merkmalspunkten in mehreren Bildern von verschiedenen Kameramodellen auf geeignete Weise RGB-Werte des Bildes abgetastet werden, dann liegen für den gleichen Oberflächenpunkt RGB-Messwerte aus unterschiedlichen RGB-Farbräumen vor. Unter Annahme von diffusem Reflexionsverhalten der

Oberfläche, liegt zudem eine für alle Richtungen konstante Strahldichte vor, d.h. die erhaltenen RGB-Messwerte sind eine durch unterschiedliche Kameramodelle durchgeführte Bewertung der gleichen Strahldichte. Diese RGB-Werte können daher als Messdaten für eine empirische Charakterisierung zwischen verschiedenen RGB-Farbräumen verwendet werden. Vorteil dieser Vorgehensweise ist, dass Messdaten direkt aus den Bildern der Szene entnommen werden. Dadurch liegen Messdaten in exakt der gleichen Beleuchtungsumgebung der Szene, was für eine empirische Charakterisierung die besten Resultate liefert [Sha03].

Die Forderung nach diffusen Szenenoberflächen wird durch die Funktionsweise von Bundler unterstützt: Trotz der robusten Detektion von SIFT kann Bundler korrespondierende Merkmalspunkte nur auf realen Oberflächen finden, deren Aussehen aus verschiedenen Kameraperspektiven der Bilder gleich bleibt, also ein diffuses Reflexionsverhalten aufweisen. Bundler liefert auf Oberflächen mit spekularem Reflexionsverhalten keine Merkmalspunkte und daher werden diese bei der Charakterisierung nicht berücksichtigt. Daraus folgt, dass nicht alle Oberflächen der Szene diffus sein müssen.

### **3.1.3 Kalibrierung der Kameramodelle**

Vor der eigentlichen Farbcharakterisierung muss eine Kalibrierung der Kamera erfolgen, um diese in einen definierten Zustand der Farbwiedergabe zu versetzen [WR04]. Dazu müssen alle Kameraparameter wie zum Beispiel Blende, Belichtungszeit und Weißabgleich auf feste Werte eingestellt sowie alle automatischen Bildkorrekturen deaktiviert werden.

Anschließend wird die Kamerakennkurve ermittelt, welche auch als optoelektronische Übertragungskurve bezeichnet wird. Diese beschreibt den Zusammenhang zwischen einfallender Strahldichte und resultierendem Wert für jeden Farbkanal. Sie wird der eigentlichen Farbraumtransformation vorgeschaltet.

Die in diesem Kapitel vorgestellte Farbcharakterisierung von 3D-Photo-Collections soll lediglich teilweise kalibrierte Kameras erfordern. Konkret sollen auch Bilder mit unterschiedlichen Belichtungszeiten und Blendenwerten zur Farbcharakterisierung genutzt werden. Die Auswirkungen dieser beiden Kameraparameter auf die Farbwiedergabe sollen daher modelliert werden, was als eine parametrisierte Form der Kamerakalibrierung angesehen werden kann.

## 3.2 Verwandte Arbeiten

### 3.2.1 Farbcharakterisierung von digitalen Kameras

Ziel der Farbcharakterisierung ist die möglichst präzise Abbildung eines geräteabhängigen RGB-Farbraums in das geräteunabhängige CIE Normvalenzsystem. Dabei kann zwischen den zwei grundlegenden Ansätzen von modellbasierter und empirischer Farbcharakterisierung unterschieden werden [Sha03].

#### Modellbasierte Charakterisierung

Die modellbasierte Charakterisierung zielt auf eine Modellierung der Kameraabbildung von Lichtspektren auf RGB-Werte ab, indem die spektralen Sensorantwortkurven der Kamera für alle drei Farbkanäle rekonstruiert werden. Ein naiver Ansatz besteht hier darin, die RGB-Wiedergabe des Bildsensors für eine bestimmte Wellenlänge durch Verwendung einer monochromatischen Lichtquelle zu ermitteln. Dabei muss aber monochromatisches Licht für eine große Anzahl von Wellenlängen des sichtbaren Spektrums erzeugt werden. Ein anderer modellbasierter Ansatz [BF02] verwendet fotografierte Messproben mit unterschiedlichen bekannten Reflexionsspektren, um aus deren RGB-Werten die Sensorantwortkurven abzuleiten.

#### Empirische Charakterisierung

Empirische Charakterisierungsmethoden verwenden unter bekannten Beleuchtungsbedingungen fotografierte Farboberflächen mit ebenfalls bekannten XYZ-Tristimuli. Zu diesem Zweck werden in der Regel Farbtafeln mit genormten Farbproben (Colorchecker) verwendet. In den meisten Fällen wird zur Linearisierung der RGB-Werte die Kamerakennkurve ermittelt [WR04]. Anschließend wird die Relation zwischen den fotografierten RGB-Daten und den bekannten XYZ-Tristimuli der Farbprofilen durch eine geeignete Abbildung beschrieben. Dazu zählen lineare Ausgleichsrechnung, Look-Up-Tabellen, polynomiale Ausgleichskurven sowie künstliche neuronale Netzwerke.

**Lineare Ausgleichsrechnung** Lineare Ausgleichsrechnung erfolgt durch Berechnung der Relation zwischen RGB- und XYZ-Daten als ein überbestimm-

tes lineares Gleichungssystem. Die Lösung dieses Systems nach der Methode der kleinsten Quadrate liefert eine  $3 \times 3$ -Transformationsmatrix, welche die lineare Abbildung mit dem kleinsten Fehler beschreibt [ST97]. Vorteile dieser Methode sind die Robustheit gegenüber verrauschten Daten und die einfache Extrapolation des Zusammenhangs für den kompletten RGB-Farbraum.

**Look-Up-Tabellen** Look-Up-Tabellen (LUT) platzieren gleichmäßig abgetastete RGB-Werte als Punkte auf einem regelmäßigen 3D-Gitter. Für jeden Punkt ist der zu einem RGB-Wert ermittelte XYZ-Wert hinterlegt. Wird ein RGB-Wert abgerufen, der nicht auf einem Gitterpunkt liegt, dann wird der XYZ-Wert aus den XYZ-Werten der umliegenden Gitterpunkte interpoliert. Der Vorteil von 3D-LUTs ist, dass sie keine Annahmen über die Abbildung von RGB- auf XYZ-Werte machen, lediglich die Gitterauflösung muss ausreichend hoch sein. Nachteilig ist jedoch der relativ große Speicherbedarf, der mit der Auflösung kubisch wächst.

**Polynomiale Ausgleichskurven** Mit polynomialen Ausgleichskurven können nicht-lineare Zusammenhänge zwischen RGB- und XYZ-Daten mit wenigen Koeffizienten beschrieben werden [HLR01]. Jedoch müssen die Koeffizienten geeignet gewählt werden, um Überanpassung, also das Mitmodellieren von Rauschen in den Messdaten, zu vermeiden.

**Künstliche neuronale Netzwerke** Ebenfalls können künstliche neuronale Netzwerke verwendet werden, um unbekannte, nicht-lineare Zusammenhänge zwischen den Messdaten zu beschreiben. Mehrlagige Perzeptrons wurden bereits für die Farbcharakterisierung verwendet [CW02], jedoch ist deren Training mit den Messdaten rechenaufwändig und es muss manuell eingegriffen werden, um ähnlich gute Ergebnisse wie mit polynomialen Ausgleichskurven zu erhalten.

### **Charakterisierung durch Ermittlung spektraler Reflexionsgrade**

Die Charakterisierung anhand der Abschätzung von spektralen Reflexionsgraden aus RGB-Bildern von Materialproben kann als Spezialfall der empirischen Charakterisierung angesehen werden. Von Cheung et. al. [CWL<sup>+</sup>05] wurden vier Methoden zur Messung von spektralen Reflexionsgraden mit Hilfe von

trichromatischen Kameras [MW86] [IB99] [LL01] [SH02] auf ihre Eignung für die Charakterisierung geprüft. Die Ermittlung von Reflexionsspektren nur anhand von RGB-Daten geschieht hierbei unter der Annahme von glatten Kurvenverläufen der spektralen Reflexionsgrade. Aus diesen abgeschätzten Spektren können CIE-Tristimuli berechnet werden, welche durch Gegenüberstellung der entsprechenden RGB-Werte eine Farbcharakterisierung ermöglichen. Ergebnis der Bewertung war, dass zwar mit allen vier Methoden eine Charakterisierung durchgeführt werden kann, aber klassische empirische Charakterisierungsansätze zu genaueren Ergebnissen kommen.

### 3.2.2 Farbkorrektur für mehrere Kameras

#### Farbangleichung in 3D-Photo-Collections

Snavely et. al. [SGSS08] führt als Teilschritt für eine Sichtinterpolation in 3D-Photo-Collections eine Farbkorrektur von Bildern durch. Die Bilder der 3D-Photo-Collection stammen aus Internetbilddatenbanken und wurden daher zu verschiedenen Tageszeiten mit unkalibrierten Kameras aufgenommen. Dadurch variiert Beleuchtungsumgebung der Szene und Farbwiedergabe der Kameras je nach Bild sehr stark. Auf Grund dieser sehr heterogenen Bilder führt Snavely keine Farbcharakterisierung im eigentlichen Sinn durch, sondern gleicht nur den Farbton der Bilder ab um für animierte Wechsel von einem zum nächsten Bild sanfte Sichtinterpolationen durchzuführen.

Anhand korrespondierender Merkmalspunkte werden Bilder aus ähnlichen Kameraperspektiven übereinander gelegt und die Unterschiede in Farbe und Belichtung werden durch Berechnung eines Skalierungsfaktors und einem Offset für jeden Farbkanal modelliert. Diese Farbadaption funktioniert für Bilder mit vergleichsweise geringen Farbunterschieden. Falls die Beleuchtung der Bilder jedoch zu starke Unterschiede aufweist, zum Beispiel Tag- und Nachtszenen, dann schlägt diese Adaption fehl und entsprechende Bilder werden bei der Sichtinterpolation nicht verwendet. Insgesamt wird also keine Charakterisierung von Farbräumen durchgeführt, jedoch ermöglicht der Ansatz auch bei Beleuchtungsunterschieden eine homogenere Darstellung von 3D-Photo-Collections, wenn auch nur für die Sichtinterpolation.

## Charakterisierung von Multikamerasystemen

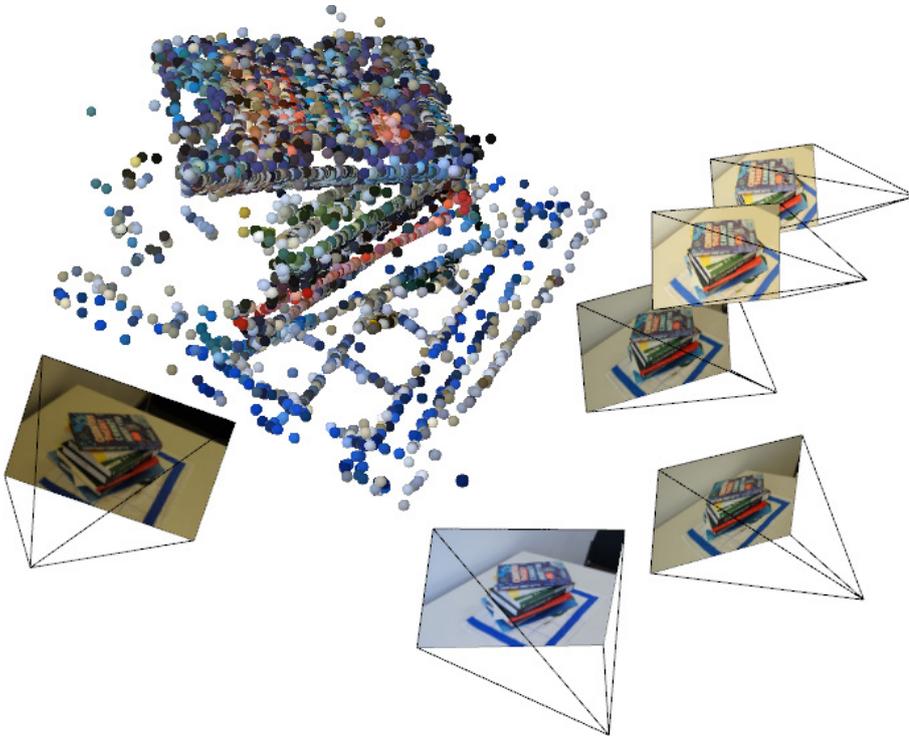
Multikamerasysteme bestehen aus mehreren fest montierten Videokameras, welche eine Szene aus verschiedenen Kameraperspektiven zeigen. Die Charakterisierung dieser Kameras lässt sich bis zu einem gewissen Grad auf 3D-Photo-Collections übertragen.

Ilie und Welch [IW05] führen eine halbautomatische empirische Charakterisierung in zwei Schritten durch. Dabei wird ein Colorchecker eingesetzt und RGB-Werte mit einem Referenzgerät gemessen. In einem iterativen linearen Ausgleichsverfahren werden die Hardwareparameter wie etwa Helligkeit und Kontrast der Kameras so justiert, dass jede Kamera möglichst zum Referenzgerät ähnliche RGB-Werte wiedergibt. In einem zweiten Schritt werden die verbleibenden Unterschiede der Farbräume wahlweise durch lineare oder polynomiale Ausgleichskurven ermittelt. Besonders die polynomialen Ausgleichskurven liefern präzise Abbildungen zwischen den Farbräumen, jedoch muss ein Colorchecker verwendet werden und es müssen für jede Kamera, die Farbflächen des Colorcheckers manuell lokalisiert werden.

Andere Ansätze [PTISK10] [YYF<sup>+</sup>06] verwenden SIFT, um in den Bildern von verschiedenen Kameras korrespondierende Merkmalspunkte zu finden, ähnlich zum in diesem Kapitel vorgestellten Verfahren. Ebenfalls wird eine Referenzkamera ausgewählt, in dessen Farbraum die Farbräume der anderen Kameras abgebildet werden. Da keine Kameraposen rekonstruiert werden, müssen die Kameraperspektiven sehr ähnlich sein. Außerdem ist das Verfahren für vollständig kalibrierte Videokameras konzipiert, daher ist keine Modellierung von variablen Belichtungszeiten oder Blendenwerten vorgesehen.

### 3.3 Eigener Ansatz

Die Bilder der 3D-Photo-Collection zeigen die Szene aus verschiedenen Blickwinkeln und Kamerapositionen und wurden mit unterschiedlichen Kameramodellen fotografiert. Es liegt also eine Menge  $\{D_0, D_1, D_2, \dots, D_{m-1}\}$  von insgesamt  $m$  verwendeten Kameramodellen vor. Es wird vorausgesetzt, dass für jedes Kameramodell der Weißabgleich auf einen konstanten Wert eingestellt ist, also auch kein automatischer Weißabgleich verwendet wird. Die Bilder von gleichen Kameramodellen wurden mit verschiedenen Belichtungs-



**Abbildung 3.2:** Von Bundler erzeugte 3D-Photo-Collection *Buchstapel*

zeiten und Blendenwerten fotografiert.

Wie in Abbildung 3.2 am Beispiel der Szene *Buchstapel* zu sehen, liegt eine 3D-Punktwolke mit einer groben Abtastung der Szenenoberflächen vor. Jedem Punkt sind mindestens zwei korrespondierende Merkmalspunkte aus unterschiedlichen Bildern zugeordnet, welche vom SIFT-Detektor erkannt wurden. Falls diese Bilder von verschiedenen Kameramodellen fotografiert wurden, dann liegen für den gleichen Punkt der realen Oberfläche RGB-Informationen aus den unterschiedlichen geräteeigenen RGB-Farbräumen vor. Auf Grund der großen Anzahl von korrespondierenden Merkmalspunkten können für zwei verschiedene Kameramodelle sehr viele Paare von korrespondierenden RGB-Werten aufgestellt werden, woraus die Abbildung von einem zum anderen Farbraum ermittelt werden kann.

Wie in Abschnitt 3.1 beschrieben, wird ein Referenzkameramodell,  $D_r$ , aus den  $m$  verfügbaren Kameramodellen ausgewählt. Unter Nutzung der oben be-

schriebenen Paare von korrespondierenden RGB-Werten soll für jede der  $m-1$  Kameramodelle die Abbildung deren Farbräume in den Farbraum des Referenzkameramodells modelliert werden. Nachfolgend wird diese Modellierung zwischen dem Referenzkameramodell,  $D_r$ , und einer der uncharakterisierten Kameramodelle,  $D_u$ , beschrieben.

### 3.3.1 Suche nach korrespondierenden Merkmalspunkten

Zunächst werden aus den beiden Kameramodellen  $D_u$  und  $D_r$  deren Bildermengen  $I_u$  und  $I_r$  gebildet. Für jeden Punkt  $p$  der Punktwolke existiert eine Menge von Bildern  $I_p$  in denen dieser Punkt sichtbar, d.h. in dem ein durch SIFT gefundener Merkmalspunkt existiert. Durch Ermittlung der Schnittmengen  $I_p \cap I_u$  und  $I_p \cap I_r$  wird für den untersuchten Punkt  $p$  ermittelt, in welchen Bildern von  $I_u$  und  $I_r$  dieser sichtbar ist. Es liegen korrespondierende Merkmalspunkte vor, falls gilt:

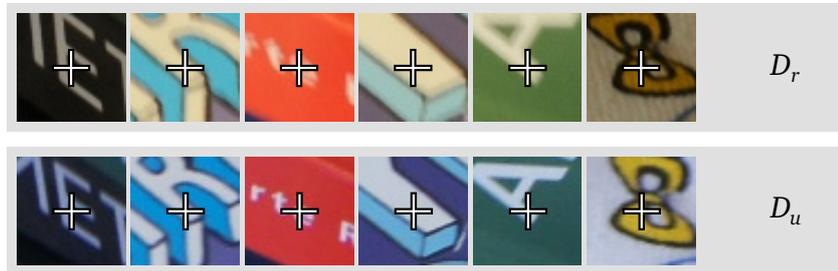
$$\begin{aligned} I_p \cap I_u &\neq \emptyset \\ I_p \cap I_r &\neq \emptyset \end{aligned} \tag{3.1}$$

Beide Schnittmengen haben in der Regel jedoch mehr als ein Bild, da korrespondierende Merkmalspunkte auch in vielen Bildern des gleichen Kameramodells existieren. Daher wird ein Qualitätsmaß benötigt, mit dem aus den Schnittmengen das am besten geeignete Bildpaar ausgewählt werden kann.

Hierfür wird der Winkel  $\alpha$  verwendet, welcher sich aus den Kamerapositionen von zwei Bildern,  $\vec{x}_1$  und  $\vec{x}_2$ , sowie der Position des untersuchten Punktes  $\vec{p}$  folgendermaßen errechnet:

$$\alpha = \arccos \left( \frac{\vec{x}_1 - \vec{p}}{|\vec{x}_1 - \vec{p}|} \cdot \frac{\vec{x}_2 - \vec{p}}{|\vec{x}_2 - \vec{p}|} \right) \tag{3.2}$$

Wenn Bildpaare so ausgewählt werden, dass  $\alpha$  möglichst gering ist, dann liegen RGB-Werte aus sehr ähnlichen Blickwinkeln vor. Bei nicht vollständig diffus reflektierenden Oberflächen werden so abweichende Strahldichten minimiert. Nach diesem Vorgehen liegen nach Auswertung aller Punkte der Punktwolke eine Menge von Bildpaaren vor, die jeweils von  $D_u$  und  $D_r$  fotografierte korrespondierende Merkmalspunkte enthalten (siehe Abb. 3.3).



**Abbildung 3.3:** Korrespondierende Merkmalspunkte ermittelt aus dem uncharakterisierten Kameramodell,  $D_u$ , (unten) und dem Referenzkameramodell,  $D_r$ , (oben)

### 3.3.2 Abtastung von RGB-Werten

RGB-Werte eines Bildes unterliegen einem gewissen Fehler, auf Grund von Artefakten und Rauschen. Artefakte entstehen durch Demosaicing und JPEG-Kompression, Rauschen tritt vor allem in schwach belichteten Bereichen auf. Um RGB-Werte mit höherer Genauigkeit zu erhalten, sollen daher die RGB-Werte von mehreren Pixeln im Umkreis des Merkmalspunktes abgetastet und daraus ein gemittelter RGB-Wert berechnet werden.

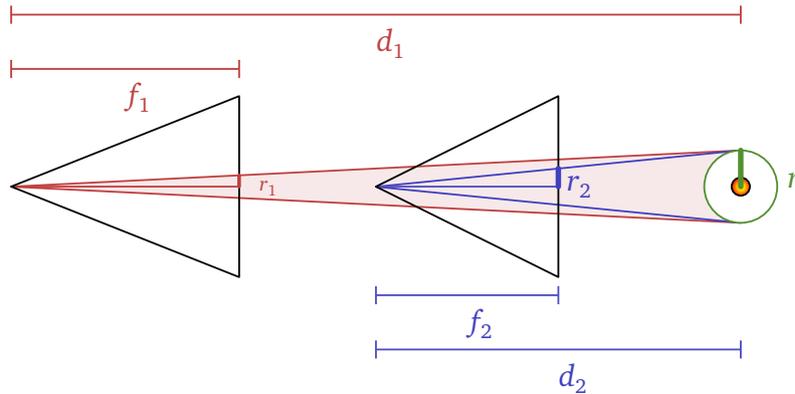
#### Linearisierung der RGB-Werte

Da das Mittel aus mehreren RGB-Werten eine lineare Interpolation darstellt, müssen die RGB-Werte selbst linear zur von der Kamera bewerteten Strahldichte sein. Andernfalls erfolgt eine falsch gewichtete Berechnung des mittleren RGB-Wertes. Im folgendem Abschnitt wird daher ein einfacher Linearisierungsschritt vorgenommen.

Für lineare RGB-Werte gilt folgender Zusammenhang zwischen der Bestrahlung des Bildsensors, ausgedrückt als spektrale Vektorgröße  $\vec{H}$  und dem RGB-Tripel  $\vec{v}$ , welches eine Folge der Bewertung durch die unbekannte Kameraantwortkurve  $f_D$  ist:

$$k \vec{v} = f_D(k \vec{H}) \quad (3.3)$$

Das bedeutet, dass eine Verdopplung der eingehenden Strahldichte auch zu einer Verdopplung des resultierenden RGB-Werts führen muss, also ein linearer Zusammenhang zwischen beiden vektoriellen Größen herrscht.



**Abbildung 3.4:** Anhand des Strahlensatzes unter Nutzung von Kameraabstand ( $d_1, d_2$ ) und Brennweite ( $f_1, f_2$ ) kann der Radius im Bildraum ( $r_1, r_2$ ) ermittelt und zwischen zwei Bildern in Relation gesetzt werden.

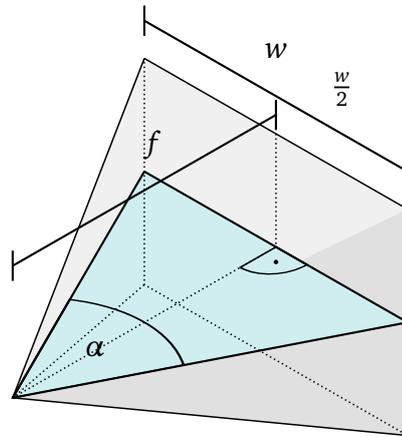
Für die Bilder wird angenommen, dass diese in einem sRGB-ähnlichem Farbraum liegen. Dieser Farbraum ist in Abhängigkeit des CIE-Normvalenzsystems definiert mit einer zusätzlichen Delinearisierungsfunktion zum Zweck der Gammakorrektur [SACM96]. Durch Invertierung dieser Gammakorrektur wird eine Linearisierung eines RGB-Werts  $\vec{v}$  durch die folgende Funktion  $l$  näherungsweise beschrieben:

$$l(\vec{v}) = \begin{pmatrix} l_c(v_r) \\ l_c(v_g) \\ l_c(v_b) \end{pmatrix} \quad (3.4)$$

$$l_c(v_i) = \begin{cases} ((v_i + 0.055)/1.055)^{2.4} & \text{wenn } v_i > 0.03928 \\ v_i/12.92 & \text{wenn } v_i \leq 0.03928 \end{cases} \quad (3.5)$$

### Ermittlung von korrespondierenden Regionen

Der Radius des Kreisausschnitts muss für jedes Bild so gewählt werden, dass ein Ausschnitt mit gleichem Raumwinkel gewählt wird. Da keine Szenengeometrie vorliegt, wird eine Kugel mit dem Radius  $r$  als geometrischer Hilfskörper um den Punkt  $p$  konstruiert, auf den sich beide Merkmalspunk-



**Abbildung 3.5:** Berechnung der Brennweite in Pixeln aus Öffnungswinkel  $\alpha$  und Bildbreite  $w$  in Pixeln

te beziehen. Unter Verwendung des Strahlensatzes kann die Kugel in den Bildraum eines Bildes  $I_1$  projiziert werden und der resultierende Radius  $r_1$  in Pixeln berechnet werden. Dieser errechnet sich aus der Brennweite  $f_1$  und dem Abstand der Kameraposition zum Punkt  $d_1$  (siehe auch Abb. 3.4):

$$\frac{r_1}{f_1} = \frac{r}{d_1} \quad (3.6)$$

Die Brennweite  $f$  ist in Pixeln in Abhängigkeit des horizontalen Öffnungswinkels  $\alpha$  und der Bildbreite  $w$  definiert (siehe Abb. 3.5):

$$\tan \frac{\alpha}{2} = \frac{w}{2f} \quad (3.7)$$

Für die projizierten Kugelradien von Bildpaar  $I_1$  und  $I_2$  gilt dann folgender Zusammenhang:

$$\frac{r_1}{r_2} = \frac{f_1 d_2}{d_1 f_2} \quad (3.8)$$

$$\max(r_1, r_2) = r_{\max} \quad (3.9)$$

Aus den oben beschriebenen Verhältnis wird ein Radius festgelegt, indem der größere der beiden Radien  $r_1$  und  $r_2$  auf den Maximalwert  $r_{\max} = 8$  gesetzt wird.



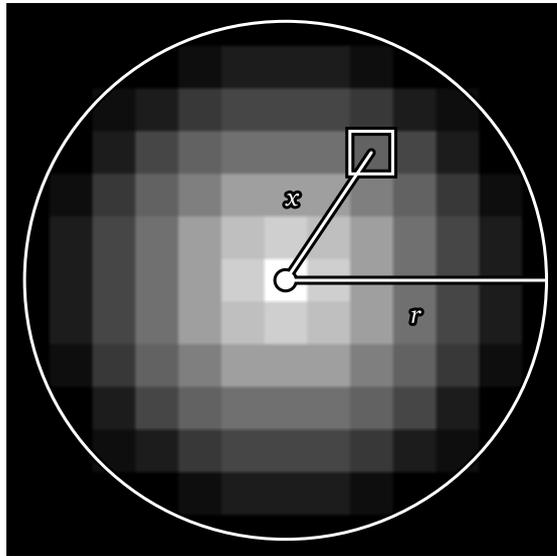
**Abbildung 3.6:** Berechnete Radien der Kreisausschnitte für beide Bildpaare wählen in guter Näherung korrespondierende Regionen auf der realen Oberfläche aus

Die oben beschriebene Projektion einer konstruierten Kugel macht zum einen die vereinfachende Annahme, dass die Kamerapositionen von  $I_1$  und  $I_2$  sowie der untersuchte Punkt  $p$  der Punktwolke auf einer Gerade liegen und zum anderen, dass die reale Oberfläche um  $p$  eben ist. Die erste Annahme wird durch Selektion von Bildpaaren mit ähnlichem Blickwinkel auf  $p$  in guter Näherung erfüllt. Zudem treten durch Wahl von sehr kleinen Radien (maximal 8 Pixel bei Bildgrößen von ca. 2 Megapixel) in beiden Bildpaaren nur geringe Projektionsfehler auf und dies auch nur bei vom Merkmalspunkt entfernten Pixeln (siehe Abb. 3.6).

Für die Berechnung eines RGB-Wertes aus Pixelwerten innerhalb des Kreisausschnitts wird eine Gewichtungsfunktion  $w(x)$  in Abhängigkeit der Entfernung des Pixels zum Merkmalspunkt  $x$  und des oben ermittelten Radius in Pixeln,  $r$ , verwendet (siehe Abb. 3.7):

$$w(x) = 1 - \frac{x}{r} \quad (3.10)$$

Durch die Gewichtung werden Pixel in der Nähe des Merkmalspunktes



**Abbildung 3.7:** Über Gewichtungsfunktion  $w(x)$  ermittelte Werte für jeden Pixel des Kreisabschnitts mit dem Radius  $r$

höher und Pixel am Rand geringer bewertet, was auch dem oben beschriebenen Projektionsfehler entspricht. Der endgültige RGB-Wert  $\vec{v}$  ergibt sich damit als gewichtetes Mittel aller  $N$  linearisierten RGB-Werte innerhalb des Kreisabschnitts:

$$\vec{v} = \sum_{i=0}^N \frac{l(\vec{v}_i)w(x_i)}{w(x_i)} \quad (3.11)$$

Nach diesem Vorgehen wird jedem Bild der im vorherigen Abschnitt erzeugten Menge von Bildpaaren ein RGB-Wert zugeordnet. Aus der Menge der RGB-Paare wird im nächsten Abschnitt eine Abbildungsvorschrift für eine Transformation von einen in den anderen RGB-Farbraum ermittelt.

### 3.3.3 Kompensation von Belichtungszeit und Blende

Um RGB-Werte, welche zwar von Bildern des gleichen Kameramodells stammen, aber mit unterschiedlicher Belichtungszeit oder Blende fotografiert wurden, vergleichbar zu machen, müssen die Auswirkungen von Blende und

Belichtungszeit auf die resultierenden RGB-Werte modelliert werden. Beide Werte sind für jedes fotografierte Bild in Form von EXIF-Daten bekannt.

Belichtungszeit  $t$  als auch Blendenzahl  $N$  beeinflussen, die Bestrahlung  $H$  der Sensoroberfläche. Bei Annahme einer im Vergleich zur Brennweite weit entfernten Szene ist die Bestrahlungsstärke  $E$  proportional zum Quotient aus Strahldichte  $L$  und  $N^2$  [AT10]:

$$E \sim \frac{L}{N^2} \quad (3.12)$$

Der Bestrahlungswert  $H$  resultiert aus der als konstant angenommenen Bestrahlungsstärke  $E$  über die Belichtungszeit  $t$ :

$$H = E t \quad (3.13)$$

Da lineare RGB-Werte vorliegen (siehe Gleichung 3.3), lässt sich der Zusammenhang zwischen Strahldichte  $\vec{L}$  und Bestrahlung  $\vec{H}$  auf den RGB-Wert  $\vec{v}$  übertragen:

$$k \vec{v} = f_D(k \frac{\vec{L} t}{N^2}) \quad (3.14)$$

$$k \frac{\vec{v} t}{N^2} = f_D(k \vec{L}) \quad (3.15)$$

Der Faktor  $k$  drückt die Proportionalität von Gleichung 3.12 aus und ist unbekannt. Für zwei RGB-Werte  $\vec{v}_1$  und  $\vec{v}_2$ , welche mit dem gleichen Kameramodell, aber unter verschiedenen Belichtungszeiten  $t_1$  und  $t_2$  sowie unterschiedlichen Blendenzahlen  $N_1$  und  $N_2$  die gleiche Strahldichte  $\vec{L}$  wiedergeben, gilt dann:

$$k \frac{t_1}{N_1^2} \vec{v}_1 = k \frac{t_2}{N_2^2} \vec{v}_2 = f_D(k \vec{L}) \quad (3.16)$$

Daraus folgt, dass bei bekannten Belichtungszeiten und Blendenzahlen RGB-Wert  $\vec{v}_1$  sich aus  $\vec{v}_2$  auf folgende Weise berechnet:

$$\vec{v}_1 = \frac{t_1 N_2^2}{t_2 N_1^2} \vec{v}_2 \quad (3.17)$$

Damit können die linearen RGB-Werte eines Bildes bei bekannter Belichtungszeit und Blendenzahl für beliebige andere Kombinationen von Belichtungszeit und Blendenzahl näherungsweise umgerechnet werden. Da hierbei nur der Quotient von Blendenzahl und Belichtungszeit relevant ist, wird jedem Bild sein Lichtwert  $LW$  [AT10] zugeordnet:

$$2^{LW} = \frac{N^2}{t} \quad (3.18)$$

Für zwei RGB-Werte aus zwei unterschiedlichen Bildern, welche die gleiche Strahldichte wiedergeben, gilt dann:

$$\vec{v}_1 = 2^{LW_2 - LW_1} \vec{v}_2 \quad (3.19)$$

Die Kompensation der unterschiedlichen Belichtungszeiten erfolgt, indem zunächst für jede Bildmenge eines Kameramodells ein Ziellichtwert  $LW_M$  errechnet wird. Hierfür wird der Median aus den Lichtwerten der Bilder verwendet, um einen Mittelwert zu erhalten, der robust gegenüber einigen wenigen stark abweichenden Lichtwerten ist. Der RGB-Wert  $\vec{v}$  eines Bildes  $I$  mit dem Lichtwert  $LW_I$  wird dann für Ziellichtwert  $LW_M$  auf folgende Weise umgerechnet:

$$\vec{v}' = 2^{LW_I - LW_M} \vec{v} \quad (3.20)$$

### 3.3.4 Modell für Farbraumabbildung

Nach den vorhergehenden Schritten liegen nun für zwei Kameramodelle  $D_u$  und  $D_r$  eine Menge von  $n$  korrespondierenden linearen RGB-Werten vor:

$$\{(\vec{v}_{r,0}, \vec{v}_{u,0}), (\vec{v}_{r,1}, \vec{v}_{u,1}), \dots, (\vec{v}_{r,n-1}, \vec{v}_{u,n-1})\} \quad (3.21)$$

Der Zusammenhang zwischen beiden Farbräumen soll dann als lineare Abbildung in Form einer  $3 \times 3$ -Transformationsmatrix erfolgen:

$$\overset{3 \times 1}{\vec{v}_r} = \overset{3 \times 3}{M} \overset{3 \times 1}{\vec{v}_u} \quad (3.22)$$

Um diese Matrix zu ermitteln werden die RGB-Werte von  $D_u$  und  $D_r$  als zwei  $3 \times n$ -Matrizen  $V_r$  und  $V_u$  dargestellt:

$$V_r = \begin{bmatrix} \vec{v}_{r,0} & \vec{v}_{r,1} & \cdots & \vec{v}_{r,n-1} \end{bmatrix} \quad V_u = \begin{bmatrix} \vec{v}_{u,0} & \vec{v}_{u,1} & \cdots & \vec{v}_{u,n-1} \end{bmatrix} \quad (3.23)$$

Da  $n$  wesentlich größer als 3 ist, liegt ein überbestimmtes Gleichungssystem vor, für das aber keine eindeutige Lösung existiert:

$$V_r = M V_u \quad (3.24)$$

Stattdessen muss  $M$  so gewählt werden, dass  $M V_u$  möglichst ähnlich zu  $V_r$  ist. Über die Methode der kleinsten Quadrate wird eine approximierte Lösung von  $M$  angestrebt, indem folgende Zielfunktion minimiert wird:

$$\min_M \|V_r - M V_u\|_F^2$$

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (3.25)$$

Die Lösung des Gleichungssystems erfolgt durch Berechnung der Pseudoinversen von  $V_u^+$  unter Verwendung der Singulärwertzerlegung [GK65]:

$$V_u = U \Sigma V^T \quad (3.26)$$

$$V_u^+ = V \Sigma^+ U^T \quad (3.27)$$

Die Matrix  $M$  ergibt sich dann folgendermaßen:

$$M = V_r \cdot V_u^+ \quad (3.28)$$

Über die Einzelschritte von Linearisierung, Lichtwertkompensation und Transformationsmatrix kann ein beliebiger nicht-linearer RGB-Wert aus einem Bild des Kameramodells  $C_u$  nun in den RGB-Farbraum des Referenzkameramodells  $C_r$  transformiert werden. Dazu wird zunächst der RGB-Wert  $\vec{v}_u$  linearisiert und an den Ziellichtwert  $LW_M$  angeglichen:

$$\vec{v}'_u = 2^{LW_r - LW_M} l_u(\vec{v}_u) \quad (3.29)$$

Anschließend erfolgt die Umwandlung in den Referenzfarbraum über Transformation mit  $M$ . Ergebnis der Transformation ist allerdings ein linearer

RGB-Wert, der mit der inversen Linearisierungsfunktion  $l_r^{-1}$  delinearisiert wird:

$$\vec{v}_r = l_r^{-1}(M \vec{v}'_u) \quad (3.30)$$

## 3.4 Auswertung

### 3.4.1 Ergebnisse

Für die verwendeten Szenen wurden drei verschiedene Kameramodelle verwendet – eine Nikon D40, eine Nikon D300 sowie eine Sony Alpha 450. Es wurden die zwei Szenen *Buchstapel* und *Flyerwand* fotografiert, wobei für jede Szene Bilder mit allen drei Kameramodellen geschossen wurden. Der Weißabgleich wurde für alle drei Kameras auf einen festen Wert eingestellt. Die abweichende Farbwiedergabe der einzelnen Kameramodelle ist Folge von unterschiedlichen Kameraeinstellungen (z.B. Blendenzahl, Weißabgleich und Belichtungszeit) sowie unterschiedlich verlaufenden spektralen Sensorantwortkurven.

Wie in Abbildung 3.8 und 3.9 zu sehen, kann das vorgestellte Charakterisierungsverfahren die unterschiedlichen RGB-Farbräume sichtbar einander anpassen. Um den Referenz-Farbraum in das CIE-Normvalenzsystem zu überführen, kann eine klassische Charakterisierung mit einem zusätzlich durch das Referenzkameramodell fotografierten Color-Checker durchgeführt werden. Wie in Abbildung 3.10 zu sehen, führt das vorgestellte Verfahren zu sichtbaren Verbesserungen bei der bildbasierten Darstellung von 3D-Photo-Collections.

### 3.4.2 Fehlerbetrachtung

Zur Berechnung des Abbildungsfehlers von  $\vec{v}'_u$  auf  $\vec{v}_r$  wird der Betrag der Differenz aus tatsächlich gemessenem  $\vec{v}_r$  und dem durch  $M$  transformierten  $\vec{v}'_u$  verwendet (siehe Abb. 3.11):

$$\Delta E_{RGB} = |l_r^{-1}(\vec{v}_r) - l_r^{-1}(M \vec{v}'_u)| \quad (3.31)$$

Dieses Fehlermaß wurde allen korrespondierenden RGB-Paaren zugeordnet und daraus ein minimaler, maximaler und mittlerer Fehler sowie die



(a) Nikon D40 (Referenzkameramodell)

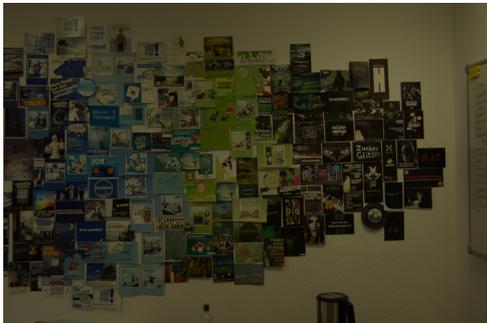


(b) Nikon D300



(c) Sony Alpha 450

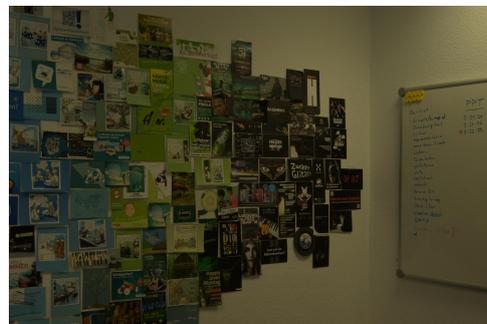
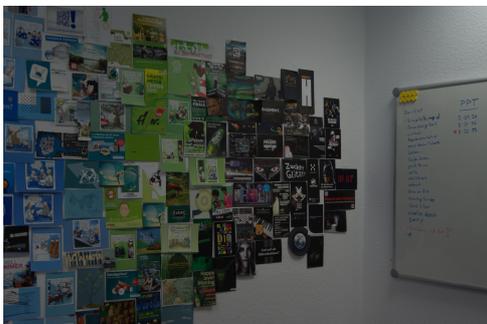
**Abbildung 3.8:** Bilder der Szene *Buchstapel* von drei verschiedenen Kameramodellen vor (links) und nach Charakterisierung (rechts)



(a) Nikon D40 (Referenzkameramodell)



(b) Nikon D300



(c) Sony Alpha 450

**Abbildung 3.9:** Bilder der Szene *Flyerwand* von drei verschiedenen Kamera-  
modellen vor (links) und nach Charakterisierung (rechts)

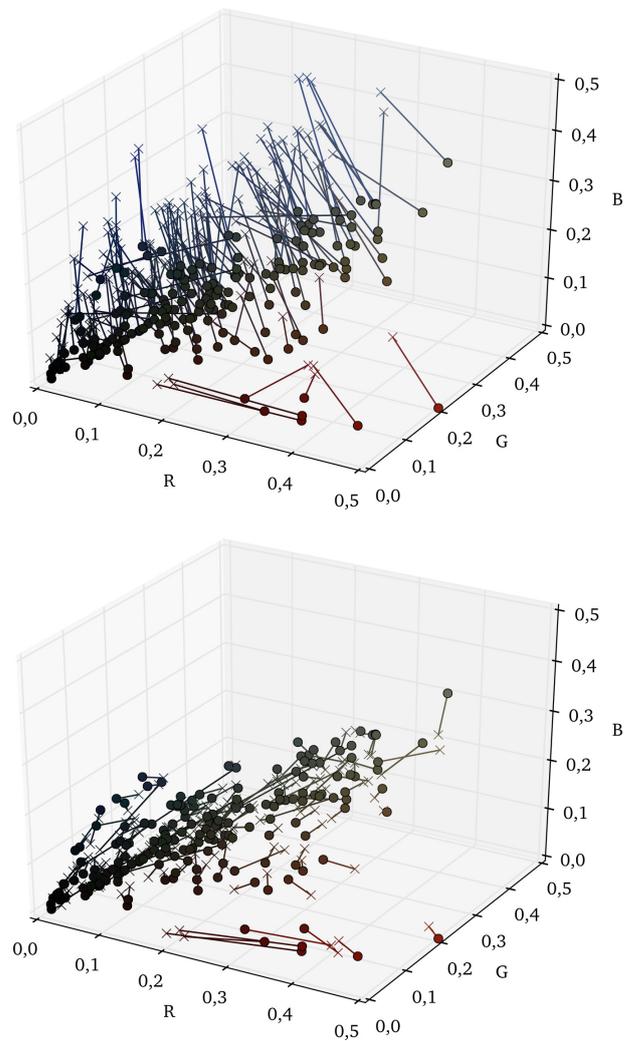


(a) Ohne Charakterisierung



(b) Mit Charakterisierung

**Abbildung 3.10:** Ergebnisse der vorgestellten Charakterisierung in Kombination mit einer bildbasierten Darstellung der 3D-Photo-Collection *Flyerwand*.



**Abbildung 3.11:** Korrespondierende RGB-Werte aus der Szene *Buchstapel* vor (oben) und nach Charakterisierung (unten). Mit • sind RGB-Werte des Referenz-Farbraums (Nikon D40) und mit × RGB-Werte des charakterisierten Farbraums (Sony Alpha 450) markiert.

Standardabweichung berechnet. Wie in Tabelle 3.1 zu sehen, liegt der mittlere Fehler zwischen 3% und 6%. Für diesen Farbfehler existieren mehrere Ursachen. Zum einen trifft die Annahme einer linearen Abbildung zwischen zwei RGB-Farbräumen nicht unbedingt zu, da etwa abweichende Sensorantwortkurven oder automatisch Nachbearbeitungsschritte durch die Kamera zu nicht-linearen Zusammenhängen zwischen verschiedenen RGB-Farbräumen führen. Zum anderen unterliegt die Abtastung von RGB-Werten einem Rauschen und die Linearisierung sowie Lichtwertkompensation ist nicht exakt. Insbesondere durch die vereinfachende Annahme eines sRGB-ähnlichen Farbraums ist die Linearisierungsfunktion  $l$  nur eine Näherung, was wiederum die Lichtwertkompensation negativ beeinflusst. Ebenfalls wurden nicht die Auswirkungen der optischen Abbildung berücksichtigt, wie etwa Vignettierung und Tiefenunschärfe.

Um den Grad des Fehlers zu ermitteln, der bei Abtastung, Linearisierung und Lichtwertkompensation von RGB-Werten entsteht, wurde eine Testszene mit nur einem Kameramodell fotografiert. Die Bilder wurden in zwei gleich große Bildmengen aufgeteilt und als von zwei verschiedenen Kameramodellen stammend behandelt. Da beide Bildmengen den gleichen Farbraum besitzen, fallen Farbraumunterschiede (ob als lineare oder nicht-lineare Abbildung) generell als Ursache für den Farbfehler weg. In der verwendeten Testszene wurde dann ein mittlerer Fehler von ca. 2% gemessen.

### 3.5 Zusammenfassung und Ausblick

Es wurde ein Verfahren zur automatischen Charakterisierung von mehreren Kameras in einen einheitlichen RGB-Farbraum unter Nutzung von 3D-Photo-

**Tabelle 3.1:** Gemessener Farbfehler für verschiedene Szenen

Name	Szene		RGB-Fehler in %		
	Bilder	Kameras	Min	Max	Mittelwert
Buchstapel	35	3	0,009	46,9	$5,6 \pm 5,1$
Flyerwand	33	3	$< 0,001$	18,1	$2,6 \pm 2,2$
Testszene	20	1	0,001	16,1	$1,9 \pm 1,7$

Collections präsentiert. Anhand der bekannten Kameraposen und Merkmalspunkte der 3D-Photo-Collection können für korrespondierende RGB-Werte unterschiedliche Kameramodelle ermittelt werden und daraus eine Transformation von einem zum anderen Farbraum durchgeführt werden. Durch Charakterisierung in einen einheitlichen RGB-Farbraum kann die Abbildung in einen geräteunabhängigen Farbraum vereinfacht werden, da dieser Schritt nur für das Referenzkameramodell und nicht mehr separat für jedes Kameramodell durchgeführt werden muss.

Zukünftige Entwicklungen könnten in die Linearisierung der RGB-Werte fließen. Eine Rekonstruktion der Kamerakennkurve könnte eine wesentlich bessere Linearisierung erzielen, da keine Annahmen über den Farbraum des Kameramodells gemacht werden müssen. Nachteilig ist hier jedoch, dass eine separate Belichtungsreihe fotografiert werden muss, was manuellen Aufwand für jedes zusätzliche Kameramodell bedeutet. Im Kapitel zur Beleuchtungsrekonstruktion mit Geometrie wird in Abschnitt 4.5.2 noch ein Ausblick auf eine Rekonstruktion der Kamerakurve gegeben, welche bei vorhandener Szenengeometrie automatisch durchgeführt werden könnte.

Neben Modellierung von Blendenzahl und Belichtungszeit könnte auch der Einfluss von ISO-Wert und Weißabgleich auf die RGB-Werte näherungsweise bestimmt werden. Beide Einstellungen sind in der Regel als EXIF-Tags hinterlegt. Der ISO-Wert könnte mit in die Lichtwertkompensation aufgenommen werden.

Auch könnte anstatt einer Transformationsmatrix ein komplexeres Abbildungsmodell verwendet werden, welches auch nicht-lineare Zusammenhänge zwischen zwei Farbräumen beschreiben kann. Kurze Tests mit 3D-Look-Up-Tabellen wurden im Rahmen dieses Verfahrens bereits durchgeführt. Allerdings besteht bei 3D-LUTs das Problem der Überanpassung, wenn die Eingabedaten verrauscht sind. Dies könnte vermieden werden, indem nah aneinander liegende RGB-Werte geglättet werden, zum Beispiel durch Einsatz eines gaußschen Konvolutionsfilters. Alternativ könnte auch eine Modellierung durch Polynome erfolgen, wobei hier die Koeffizienten geeignet gewählt müssen, da sonst ebenfalls eine Überanpassung eintritt.



# Kapitel 4

## Bildbasierte Beleuchtung in 3D-Photo-Collections

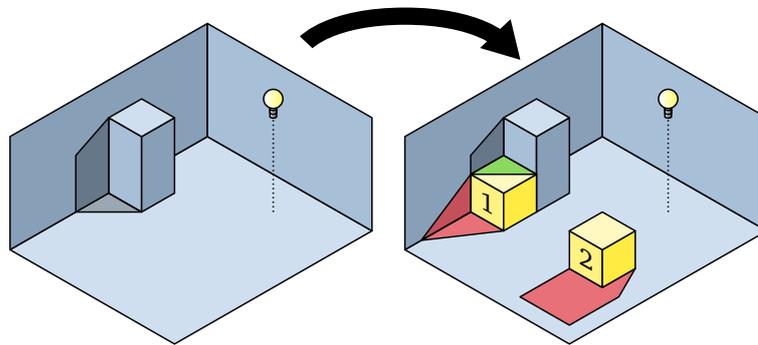
### 4.1 Einleitung

Im vorherigen Kapitel wurde ein Verfahren entwickelt, um die Bilder der 3D-Photo-Collections automatisch in einen einheitlichen Farbraum zu transformieren. In diesem Kapitel soll auf Grundlage der Bilder die realistische Augmentierung von realen Anteilen mit virtuellen Anteilen durchgeführt werden. Dieser Teilbereich der Augmented Reality wird auch als augmentierte Bildsynthese bezeichnet und setzt einen hohen bis fotorealistischen Anspruch an die Bildsynthese.

#### 4.1.1 Anforderungen

Um eine möglichst realistische Augmentierung zu erreichen, muss die Interaktion von Licht und Materie zwischen virtuellen und realen Anteilen simuliert werden. Durch das Erweitern von realen Anteilen, wie etwa Fotos, durch virtuelle Anteile, zum Beispiel 3D-Objekte oder auch Lichtquellen, muss eine Abänderung der realen Lichtpfade simuliert werden.

Einerseits wird das Aussehen des virtuellen Objekts durch aus der realen Szene einstrahlendes Licht in seinem Aussehen bestimmt. Andererseits blockiert ein virtuelles Objekt die ursprünglichen Lichtpfade und verursacht so zusätzliche Schattenwürfe (siehe Abb. 4.1). Zudem strahlen virtuelle Lichtquellen zusätzliches Licht auf die reale Szene und virtuelle Objekte ein. Darüber hinaus haben auch Mehrfachreflexionen von Licht zwischen virtuellen



**Abbildung 4.1:** Mit dem Einfügen von virtuellen Anteilen geänderte Lichtpfade; Die reale Lichtquelle verursacht Schattenwürfe auf die reale Szene durch virtuelle Objekte (rot) und umgekehrt (grün).

und realen Anteilen Einfluss auf das visuelle Ergebnis und müssen für einen fotorealistischen Anspruch ebenfalls simuliert werden.

Falls eine augmentierte Bildsynthese mit fotorealistischen Anspruch angestrebt wird, muss eine vollständige Berücksichtigung von allen geänderten Lichtpfaden durchgeführt werden. Dafür muss reflektiertes Licht und Reflexionsverhalten der Oberflächen jeweils für reale und virtuelle Anteile möglichst präzise bekannt sein. Für virtuelle Anteile sind Geometrie, Materialien und Lichtquellen zum Beispiel in Form von 3D-Modellen und Materialdefinitionen in der Regel bereits definiert. Reale Anteile sind in den meisten Verfahren für augmentierte Bildsynthese (siehe Abschnitt 4.2) in Form von Bildern oder Videostreamen gegeben, so dass Informationen über Geometrie, Reflexionsverhalten oder reflektiertes Licht erst daraus rekonstruiert werden müssen oder als zusätzliche Daten vorausgesetzt werden.

Im Fall von 3D-Photo-Collections liegen reale Anteile jedoch nur als Bildmenge mit bekannten Kameraposen und einer Punktwolke vor, welche die Geometrie der realen Szene nur sehr grob und verrauscht wiedergibt. Um mit Hilfe von 3D-Photo-Collections eine augmentierte Bildsynthese durchzuführen, müssen aus diesen unvollständigen Daten der bildbasierten Szene möglichst genaue Beleuchtungsinformationen der realen Umgebung gewonnen werden.

Daher wurden im Rahmen dieser Arbeit zwei Verfahren entwickelt, welche

aus 3D-Photo-Collections die Umgebungsbeleuchtung rekonstruieren und somit eine augmentierte Bildsynthese ermöglichen. Im folgenden Abschnitt soll auf die genaue Zielsetzung beider Verfahren eingegangen werden.

### 4.1.2 Zielsetzung

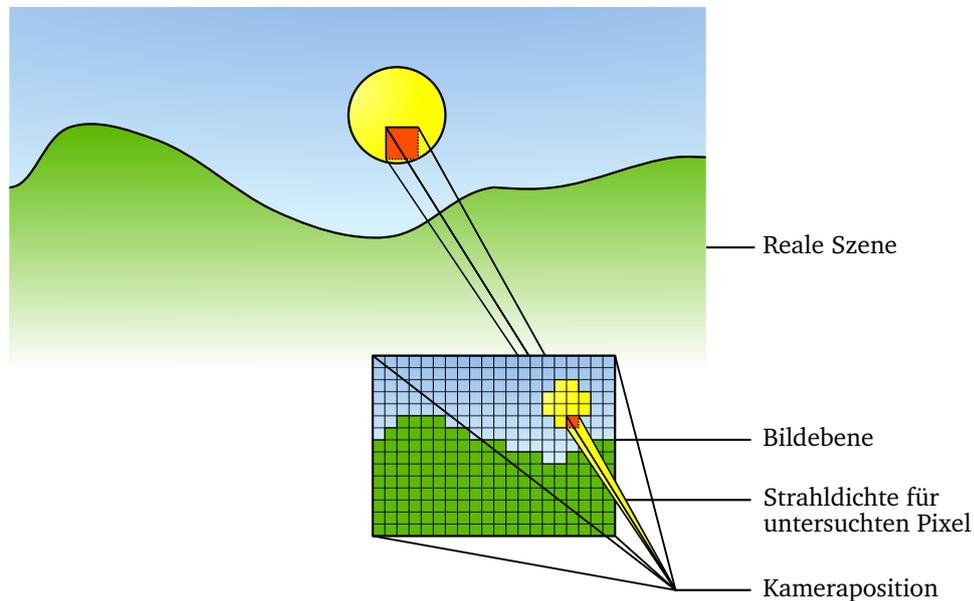
#### Erweiterung des Dynamikumfangs

Die Eingabebilder der 3D-Photo-Collection geben das von der realen Szene in Richtung der Kamera reflektierte Licht im RGB-Format wieder. Mit der vorhergehenden Farbcharakterisierung wurde sichergestellt, dass alle Bilder im gleichen Farbraum liegen. Allerdings liegen die Eingabebilder nur mit einer Quantisierung von 8 Bit je Farbkanal vor, womit ein geringer Dynamikumfang von weniger als 8 Blendstufen erreicht wird. Der Dynamikumfang kann aber in realen Szenen – zum Beispiel in Außenszenen bei Tageslicht – 10000:1 und größer sein [RWPD05]. Zu helle Oberflächen werden dann überbelichtet und zu dunkle Oberflächen der Szene werden durch das Signalrauschen der Kamera überdeckt.

Jedoch kann die große Menge der erstellten Eingabebilder ausgenutzt werden, um den Dynamikumfang der Bilddaten zu erweitern, falls Bilder mit unterschiedlichen Belichtungseinstellungen vorliegen. Durch die große Zahl an Bildern existieren viele ähnliche Ansichten auf die reale Szene. Je nachdem wie viele Bilder mit ähnlichen Ansichten, aber unterschiedlichen Belichtungseinstellungen vorliegen, können daraus Beleuchtungsdaten mit größerem Dynamikumfang, wenn nicht sogar dem gesamten Dynamikumfang der realen Szene gewonnen werden.

#### Verwendung von Szenegeometrie

Da für jedes Bild extrinsische und intrinsische Kameraparameter bekannt sind, kann jedes Eingabebild als ein Bündel von Abtastwerten der plenoptischen Funktion betrachtet werden. Ein einzelner Bildpixel repräsentiert eine bestimmte Strahldichte der realen Szene, welche aus der Aufnahmeposition der Kamera von einer bestimmten Eingangsrichtung aufgenommen wurde (siehe Abb. 4.2). Nur für die Aufnahmepositionen der Kameras, und dort ausschließlich für die innerhalb des Kamerasichtvolumens abgedeckten Eingangswinkel, liegen Abtastwerte der plenoptischen Funktion vor. Durch



**Abbildung 4.2:** Rückschluss von Bildpixeln auf Strahldichten der Umgebung

diese lückenhafte Abtastung der realen Szene sind dementsprechend wenige Informationen über die Lichtverteilung der realen Szene vorhanden.

Um eine bessere Ausnutzung der vorhandenen Abtastwerte zu erreichen, soll für die reale Szene die Annahme getroffen werden, dass alle Oberflächen eingestrahktes Licht vollständig diffus reflektieren. Basierend auf dieser Annahme können die Abtastwerte der plenoptischen Funktion als Abtastwerte der Strahldichte von einzelnen Punkten der Szenenoberfläche angesehen werden, welche für alle Betrachtungswinkel gleich groß sind. Falls die Position dieser Oberflächenpunkte bekannt ist, kann der Abtastwert auch für andere Betrachterpositionen und Eingangswinkel verwendet werden. Dies entspricht einer Reprojektion der Bilddaten unter Nutzung der Szenengeometrie.

Da 3D-Photo-Collections keine vollständige Szenengeometrie definieren, muss entweder eine näherungsweise Rekonstruktion durchgeführt werden oder eine vollständige Szenengeometrie wird zusätzlich vorausgesetzt. Die beiden entwickelten Verfahren (siehe Abschnitt 4.3 und 4.4) verfolgen jeweils eine dieser beiden Strategien.

### **Synthese von Light-Probes**

In existierenden Verfahren des Image-based Lighting wird das reale Umgebungslicht mittels spezieller Messtechnik in Form von Light-Probes gemessen (siehe Abschnitt 4.2). Analog dazu wird in den beiden vorgestellten Verfahren eine Synthese von virtuellen Light-Probes aus den Bildern der 3D-Photo-Collection durchgeführt. Das heißt, für ausgewählte Positionen der realen Szene wird das aus der realen Umgebung eingestrahlte Licht ermittelt und in Form einer HDR-Environment-Map beschrieben.

### **Echtzeitfähige Aktualisierung**

Die Synthese der Light-Probes soll in Echtzeit durchgeführt werden können. Dadurch kann eine interaktive Darstellung von realen und virtuellen Anteilen erreicht werden. Beispielsweise können so virtuelle Objekte in der Szene verschoben werden und die rekonstruierte Umgebungsbeleuchtung wirkt sich dynamisch auf das verschobene Objekt aus.

## **4.2 Verwandte Arbeiten**

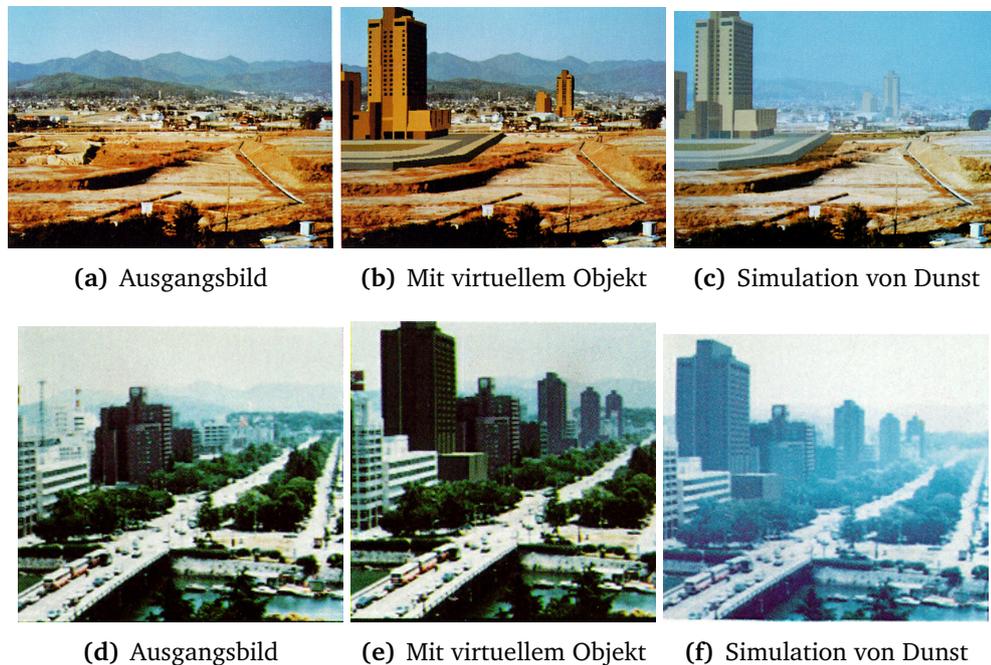
### **4.2.1 Augmentierte Bildsynthese**

#### **Nakamae et. al.**

Ein erster Ansatz zur Einbettung von virtuellen Objekten in reale Fotos unter Berücksichtigung der realen Umgebungsbeleuchtung wurde von Nakamae et. al. entwickelt [NHIN86] (siehe Abb. 4.3). Nakamae beschränkt sich dabei auf virtuelle Modelle von Gebäuden, die in Fotos von Außenszenen eingebettet werden.

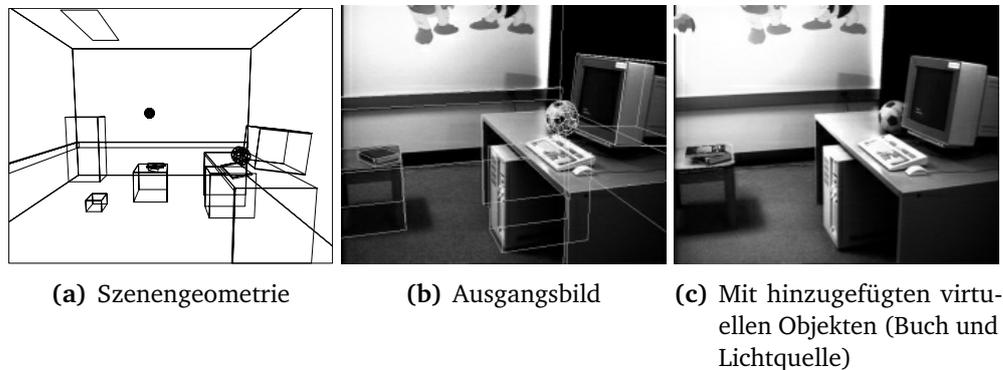
Für die Darstellung der virtuellen Anteile werden zunächst die Kameraparameter wie Position und Brennweite ermittelt. Dies erfolgt durch manuelles Abgleichen von Referenzpunkten zwischen dem gegebenen Eingabebild und einer zusätzlich verwendeten geographischen Karte der Außenszene. Zusätzlich wird noch ein iteratives Ausgleichsverfahren zur Bestimmung der Kameraposition aus den gegebenen Referenzpunkten vorgestellt.

Eine einfache Rekonstruktion des realen Umgebungslichts wird durchgeführt, indem anhand von Aufnahmezeitpunkt und -standort des Fotos der



**Abbildung 4.3:** Die Bildsynthese nach Nakamae et. al. [NHIN86] erlaubt die Einbettung von virtuellen Objekten mit angepasster lokaler Beleuchtung und Simulation der Luftperspektive.

Stand und die Einstrahlrichtung der Sonne berechnet werden. Anschließend wird damit das virtuelle Objekt anhand eines Lambertmodells mit ambienten Term lokal beleuchtet. Der Einfluss des ambienten Terms wird näherungsweise aus den Helligkeiten von zwei im Eingangsbild vorhandenen Oberflächen bestimmt, wobei die eine Oberfläche im Schatten liegt und die andere direktem Sonnenlicht ausgesetzt ist. Zusätzlich wird die im Eingangsbild – zum Beispiel durch Dunst oder Nebel – verursachte Luftperspektive modelliert. Dabei wird die Dichte des Dunstes mit einer parametrisierten Exponentialfunktion in Abhängigkeit der Sichtdistanz beschrieben. Es werden mehrere Farbwerte gleichfarbiger, aber unterschiedlich weit entfernter Objekte aus dem Eingangsbild entnommen und mittels Ausgleichsrechnung daraus die Parameter der Exponentialfunktion entnommen. Neben der Simulation des im Bild vorhandenen Nebels für virtuelle Objekte kann zusätzlich dem Bild Nebel hinzugefügt oder daraus entfernt werden.



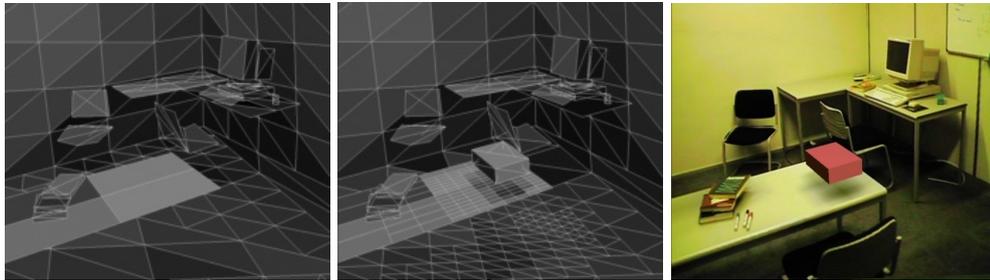
**Abbildung 4.4:** Differenzielles Rendering nach Fournier et. al. [FGR93]

Die Methode von Nakamae berücksichtigt in Ansätzen bereits viele Anforderungen für eine glaubwürdige Synthese von realen und virtuellen Anteilen und auch eine Abänderung der Szene ist bereits durch Hinzufügen oder Entfernen von Dunst möglich. Allerdings müssen noch viele Parameter manuell bestimmt werden. Außerdem findet nur eine rein lokale Beleuchtungssimulation statt, was sich insbesondere durch das Fehlen von Schattenwürfen äußert.

### Fournier et. al.

Das Verfahren von Fournier et. al. [FGR93] stellt einen neuen Ansatz zur globalen Beleuchtung von virtuellen und realen Anteilen vor, welcher auch als differenzielles Rendering bezeichnet wird (siehe Abb. 4.4). In Fourniers Verfahren wird zunächst die Geometrie der realen Szene vereinfacht mit Hilfe von Parallelepipeden rekonstruiert. Unter Annahme von vollständig diffusen Oberflächen werden anschließend in einem iterativen Verfahren auf Grundlage der Farbwerte des Eingangsbildes Strahldichten und Reflexionsgrade der realen Oberflächen näherungsweise bestimmt. Für nicht im Bild sichtbare Oberflächen wird ein ambienter Term verwendet.

Der eigentliche Schritt des differenziellen Renderings erfolgt nun, indem die rekonstruierte Szene einmal ohne ( $L_{Old}$ ) und einmal mit virtuellen Objekten ( $L_{New}$ ) gerendert wird. Aus beiden Bildern wird pixelweise das Differenzbild ( $L_{New} - L_{Old}$ ) berechnet und auf das reale Eingangsbild addiert. Die in  $L_{New}$



(a) Rekonstruierte reale Szene (b) Rekonstruierte Szene mit virtuellem Objekt (c) Reale Anteile mit virtuellem Objekt

**Abbildung 4.5:** Rendering nach Drettakis et. al. [DRB97]

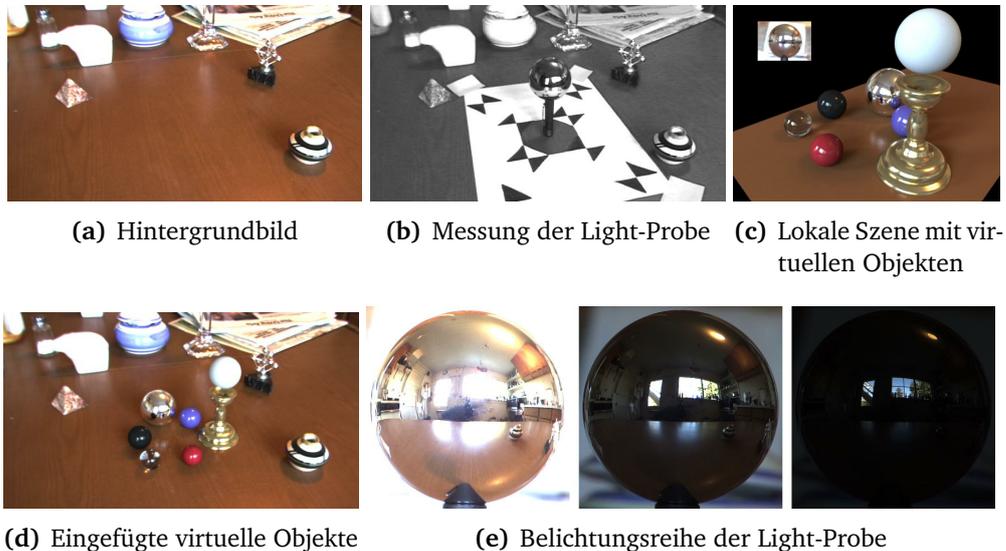
von virtuellen Objekten geworfenen Schatten werden im Differenzbild als negative Werte wiedergegeben, welche in der anschließenden Addition die Bereiche der realen Anteile verdunkeln.

Durch Rekonstruktion von Geometrie, Umgebungsbeleuchtung und Reflexionsverhalten kann Radiosity als globale Beleuchtungssimulation eingesetzt werden. Dadurch kann auch indirekte diffuse Lichtreflexion simuliert werden, was sich in Form von realistisch weichen Schattenverläufen äußert.

#### **Drettakis et. al.**

Das Verfahren von Drettakis [DRB97] baut auf dem Verfahren von Fournier auf, bietet aber eine weniger aufwändige, halbautomatische Rekonstruktion der Szenengeometrie sowie eine interaktive Beleuchtungssimulation (siehe Abb. 4.5). Es liegen mehrere Eingabebilder der realen Szene vor, welche von mehreren Positionen geschossen wurden. Für jede Position liegen mehrere Bilder mit unterschiedlicher Kameraorientierung vor, aus denen ein Panoramabild erzeugt wird. Indem ein Kalibrierungs-Target fotografiert wird, werden die intrinsischen Kameraparameter nach einem Verfahren von Robert [Rob96] ermittelt.

Die Rekonstruktion der realen Szene wurde gegenüber dem Ansatz von Fournier erweitert, indem zunächst auf mehreren Bildern halbautomatisch korrespondierende Bildpunkte ausgewählt werden. Aus diesen epipolargeometrischen Zusammenhängen wird für Bildpaare die Fundamentalmatrix berechnet, aus der mit Hilfe der bekannten intrinsischen Kameraparameter



**Abbildung 4.6:** In der Synthese nach Debevec et. al. [Deb98] wird zwischen einer durch Light-Probes bestimmten entfernten Szene und einer lokalen Szene unterschieden.

die relative Orientierung und Position der Kameras zueinander berechnet wird. Strahldichte und Reflexionsgrade der realen Oberflächen werden identisch zu Fournier berechnet. Die globale Beleuchtungssimulation erfolgt durch ein inkrementelles Radiosity-Verfahren. Damit können durch Hinzufügen, Verschieben oder Entfernen von virtuellen Objekten ausgelöste Änderungen der Lichtwege interaktiv simuliert werden.

### Debevec et. al.

Debevec [Deb98] unterteilt in seinem Ansatz die realen Anteile in eine entfernte und eine lokale Szene (siehe Abb. 4.2.1). Die lokale Szene zeigt nur die realen Objekte, die sich in unmittelbarer Nähe der virtuellen Anteile befinden, also z.B. die Teile der realen Szene, auf die Schatten von virtuellen Objekten fallen können. Die Geometrie der lokalen Szene wird grob manuell rekonstruiert und die Reflexionsparameter werden unter Auswahl eines BRDF-Modells näherungsweise ermittelt. Debevec geht davon aus, dass die entfernte Szene Licht in die lokale Szene und auf die virtuellen Objekte einstrahlt, je-

doch von diesen kein Licht empfängt. Daher werden die Lichtverhältnisse der entfernten Szene als unabhängig von der lokalen Szene angesehen und durch eine „omni-directional Radiance-Map“ genannte HDR-Environment-Map beschrieben. Diese wird durch das – mit Auswahl von unterschiedlichen Belichtungszeiten – mehrmalige Fotografieren von spiegelnden Kugeln (als „Light-Probe“ bezeichnet) gewonnen und beschreibt für einen Punkt im Raum das aus der entfernten Szene aus allen Richtungen eingestrahlte Licht.

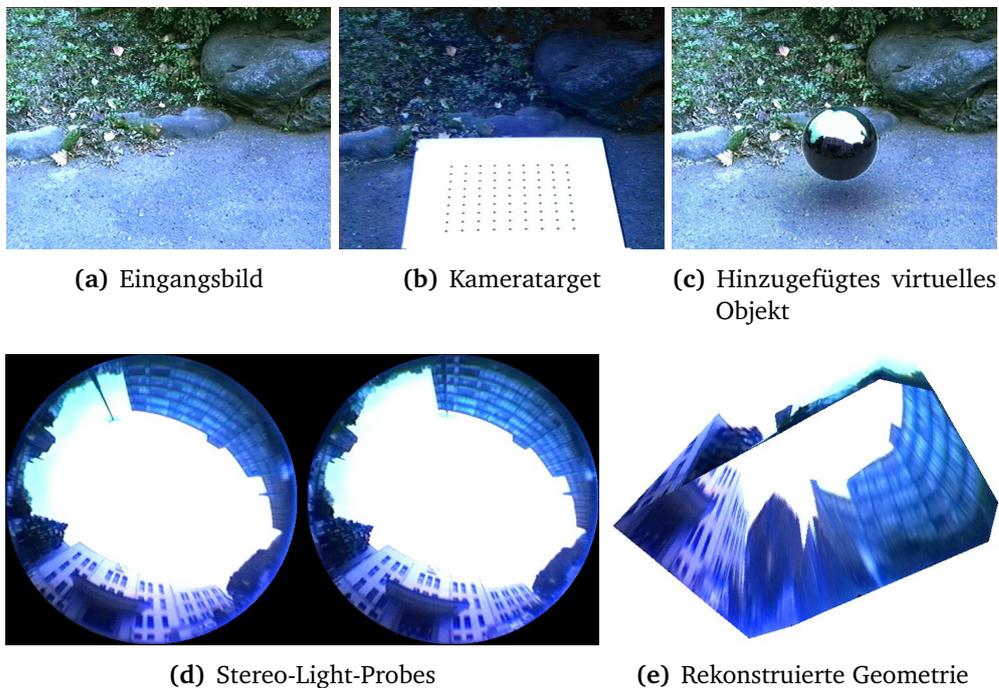
Wie Fournier, verwendet auch Debevec differenzielles Rendering mit der Erweiterung, dass bei Wiedergabe der realen Anteile zwischen entfernter und lokaler Szene unterteilt wird. Durch Fotografieren einer Light-Probe für die entfernte Szene können beliebig komplexe Umgebungsbeleuchtungen verwendet werden. Außerdem muss so nur ein Bruchteil der realen Szene manuell rekonstruiert werden. Allerdings kann für virtuelle Objekte keine Änderung der Lichtwege auf die entfernte Szene simuliert werden, wodurch virtuelle Lichtquellen nur bedingt verwendet werden können.

#### **Sato et. al.**

Sato et. al. [SSI99] verwendet einen Ansatz, bei dem keine Geometrie der realen Szene vorliegen muss (siehe Abb. 4.7). Neben einem Einzelbild der realen Szene werden zusätzlich Bilder von zwei unterschiedlichen Positionen mit einem Fischaugenobjektiv fotografiert. Dabei wird für beide Positionen mit nach oben gerichteter Kamera eine Belichtungsreihe der Umgebung unter variierender Belichtungszeit erstellt, so dass das Umgebungslicht in Form eines omnidirektionalen Stereo-HDR-Bildes vorliegt. In- und extrinsische Kameraparameter werden mit Hilfe eines zusätzlich fotografierten Kameratargets nach der Methode von Tsai [Tsa87] berechnet.

Aus dem Stereo-Bild der Umgebung wird durch ein Stereo-Verfahren automatisch eine einfache Szenengeometrie erstellt. Die Verteilung der Strahldichte in der realen Szene wird durch Projektion des HDR-Bildes auf die Szenengeometrie ermittelt. Per Raycasting werden anschließend die virtuellen Objekte entsprechend beleuchtet und mit weichen Schatten versehen.

Da die Reflexionsgrade realer Oberflächen nicht bekannt sind, werden Schattenwürfe von virtuellen Objekten auf die reale Szene nur sehr grob durch eine Skalierung der Farbwerte des Einzelbildes umgesetzt. Außerdem kann die Verdeckung virtueller Objekte durch reale Anteile nicht dargestellt werden.

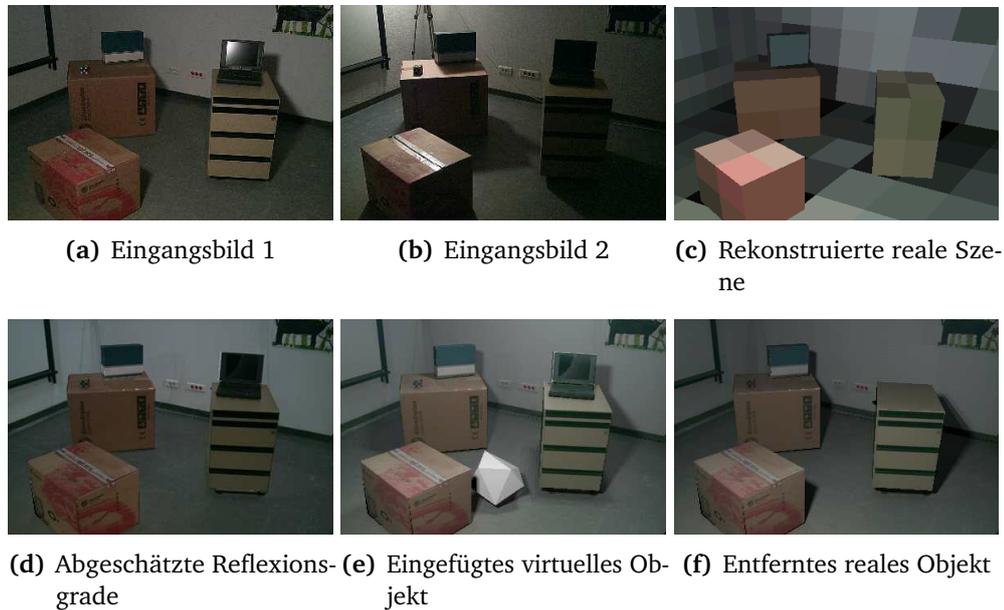


**Abbildung 4.7:** Rendering nach Sato et. al. [SSI99]

#### Loscos et. al.

Das Verfahren von Loscos [LFD<sup>+</sup>99] liefert eine neue Vorgehensweise zur Bestimmung der Reflexionsgrade (siehe Abb. 4.8). Zunächst erfolgt eine Bestimmung der Szenengeometrie, ähnlich zu dem Ansatz von Drettakis, durch halbautomatische Registrierung von korrespondierenden Punkten in mehreren Einzelbildern.

Zusätzlich werden Bilder der Szene mit gleicher Kamerapose fotografiert, in denen eine Lichtquelle bewegt wird. Die Position der Lichtquelle wird für jedes Bild manuell gemessen. Anschließend wird separat für jedes Bild auf Grundlage der gegebenen Geometrie und der Lichtquellen für jeden Pixel der Reflexionsgrad des entsprechenden Oberflächenpunktes durchgeführt. Da für Schatten diese Schätzung ungenau ist, wird durch Filterung und Zusammenführung der Reflexionsgrade aus jedem Bild ein genaueres Ergebnis erzielt.



**Abbildung 4.8:** Rendering nach Loscos et. al. [LFD<sup>+</sup>99]

Unter Nutzung der geschätzten Reflexionsgrade wird eine globale Beleuchtungssimulation in Form eines hierarchischen Radiosity-Ansatzes durchgeführt, welche die interaktive Platzierung von virtuellen Objekten, die Entfernung von realen Objekten sowie die Änderung der Strahlstärke von realen Lichtquellen ermöglicht. Zudem wird die direkte Beleuchtung von virtuellen Objekten durch ein schnelleres Raycasting durchgeführt, während Radiosity lediglich für die Berechnung des indirekten Lichtes verwendet wird.

#### **Gibson et. al.**

Ein neueres Verfahren von Gibson et. al. [GCHH03] wurde mit dem Ziel der echtzeitfähigen Darstellung von weichen Schatten für die Nutzung von Grafikkhardware optimiert (siehe Abb. 4.2.1). Die Geometrie wird über ein halbautomatisches bildbasiertes Modellierungsverfahren [GHCH03] näherungsweise rekonstruiert. Aus einer Light-Probe wird eine HDR-Environment-Map erzeugt. Diese wird auf die Geometrie projiziert und beschreibt so die Beleuchtungsstärke der realen Oberflächen unter der Annahme, dass diese



(a) Eingangsbild mit realen Objekt (b) Referenz-Rendering mit virtuellem Objekt (c) Interaktives Rendering mit virtuellem Objekt

sämtliches Licht vollständig diffus reflektieren. Die Oberflächen der Geometrie werden in Patches aufgeteilt, welche als Sender- und Empfängerpatches klassifiziert werden. Unter Angabe eines Schwellwertes wird dabei eine Menge von Patches mit der größten Bestrahlungsstärke als Senderpatches klassifiziert. Die Senderpatches werden als Repräsentation der realen Lichtquellen verwendet, woraus durch Anwendung eines Inverse-Radiosity-Ansatzes diffuse Reflexionsgrade der Patches bestimmt werden.

Der Lichttransport von jedem Senderpatch zu sämtlichen Empfängerpatches wird vorausberechnet und in einer hierarchischen Datenstruktur abgelegt. Dadurch kann mit dem Einfügen eines virtuellen Objektes sehr schnell ermittelt werden, welche Lichtwege von Sender- zu Empfängerpatches blockiert werden. An eingefügten virtuellen Objekten werden mit Hilfe von Shadow-Mapping Schatten durch die Grafikhardware gerendert, wobei blockierte Senderpatches als Lichtquellenposition für den Schattenwurf verwendet werden. Da für viele verschiedene Senderpatches Schatten generiert werden, entstehen realistisch weiche Schattenkanten. Die Darstellungsgeschwindigkeit wird zusätzlich erhöht, indem die Strahldichte der realen Umgebung an verschiedenen Punkten der Szene vorberechnet und in einem Irradiance-Volume [GSHG98] gespeichert wird. Anhand des Irradiance-Volume erfolgt die Berechnung des diffusen Anteils für die lokale Beleuchtung von virtuellen Objekten. Der spekulare Anteil wird durch Verwendung der Environment-Map in Kombination mit Reflection-Mapping bestimmt. Durch die vorgenommenen Optimierungen kann eine interaktive Darstellung mit mehr als 10 Bildern je Sekunde erreicht werden. Somit können virtuelle Objekte in der Szene bewegt werden und entsprechende Änderungen des Lichteinfalls werden zeitnah dargestellt.

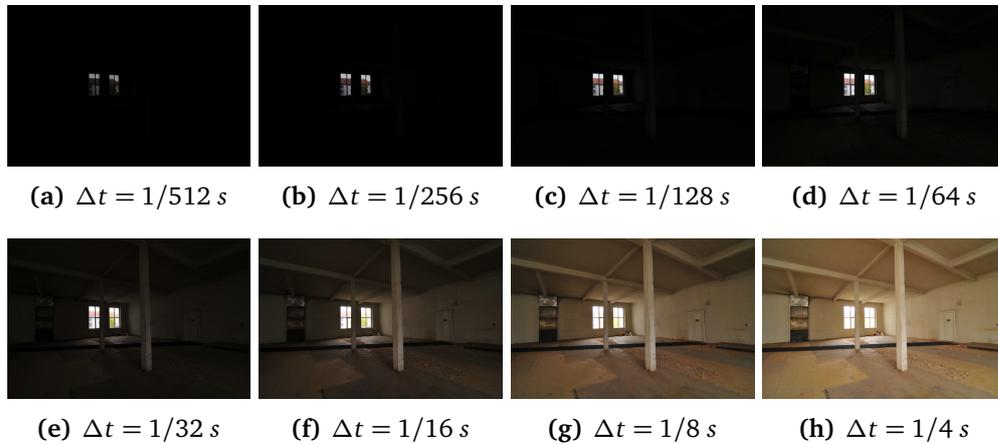


**Abbildung 4.9:** Synthese mit differenziellem Photon Mapping nach Grosch [Gro05]. Im oberen Bereich sind virtuelle Objekte eingefügt, während die Objekte im unteren Bildbereich real sind.

### Grosch

Von Grosch wurde differenzielles Photon-Mapping [Gro05] vorgestellt (siehe Abb. 4.9). Als reines Offlineverfahren für augmentierte Bildsynthese ist es für fotorealistische Ergebnisse ausgelegt und hat gegenüber dem differenziellen Rendering den Vorteil, dass die globale Beleuchtungssimulation nur einmal durchlaufen werden muss. Grundlage ist dabei Photon-Mapping [Jen96] als leistungsfähiges globales Beleuchtungsverfahren.

Wie bei Gibson wird die reale Szene durch eine auf die Szenengeometrie projizierte Light-Probe beschrieben. Es werden Photonen aus der realen Szene ausgesendet und in einer Photon-Map gesammelt. Diese beschreibt die durch Einfügen von virtuellen Objekten verursachte Änderung der Lichtpfade, indem positive Photonen den hinzugefügten Lichtstrom und negative Photonen den



**Abbildung 4.10:** Belichtungsreihe bestehend aus Fotos, welche die gleiche Ansicht auf die Szene zeigen, aber mit unterschiedlichen Belichtungszeiten erstellt wurden

fehlenden Lichtstrom repräsentieren. Für jede Oberfläche kann mit Hilfe der Photon-Map anschließend die resultierende Leuchtdichte und damit das augmentierte Bild berechnet werden.

#### 4.2.2 Rekonstruktion der Kamerakennkurve aus Belichtungsreihen

Die folgenden Verfahren führen eine Rekonstruktion der Kamerakennkurve anhand von Belichtungsreihen durch. Eine Belichtungsreihe besteht aus mehreren Fotos, welche die Szene aus dem gleichen Blickwinkel und in der gleichen Beleuchtungssituation zeigen, aber von der gleichen Kamera mit unterschiedlichen Belichtungszeiten fotografiert wurden.

Die Kamerakennkurve beschreibt die Abbildung der Kamera von auf den Bildsensor einfallenden Bestrahlungswerten  $H$  auf Pixelwerte  $Z$ , wobei der Bestrahlungswert sich aus einer konstanten Bestrahlungsstärke  $E$  über eine gewisse Belichtungszeit  $\Delta t$  ergibt:

$$Z = f(H) = f(E \Delta t) \quad (4.1)$$

Die Rekonstruktion der Kamerakennkurve ist entscheidend, um durch eine

Invertierung dieser Kurve aus bekannten Pixelwerten näherungsweise die ausgesendeten Strahldichten der Szene abzuschätzen. Die Bilder verwenden jedoch nicht einzelne Graustufenwerte, sondern RGB-Tripel. Daher werden die folgenden Verfahren in der Regel für jeden der drei Farbkanäle unabhängig voneinander durchgeführt.

Die Belichtungsreihe beschreibt für mehrere Belichtungszeiten  $\Delta t_j$  und verschiedene Pixelkoordinaten  $i$  eine Menge von Pixelwerten  $Z_{ij}$  (siehe Abb. 4.10). Für gleiche Pixelkoordinaten liegen somit Werte für unterschiedliche Belichtungszeiten vor. Aus diesen Korrespondenzen wird durch die folgenden Verfahren auf unterschiedliche Weise die Kamerakennkurve modelliert und teilweise daraus ein einzelnes Bild mit hohem dynamischem Umfang (HDR) erzeugt. Die dabei rekonstruierten Kamerakennkurven und Bestrahlungsstärken stellen allerdings lediglich relative Messwerte dar. Um Pixelwerten absolute Bestrahlungsstärken in  $W/m^2$  zuzuordnen, muss eine absolute radiometrische Kalibrierung durchgeführt werden.

### Rekonstruktion nach Debevec

Debevec [DM97] beschreibt einen Pixelwert  $Z_{ij}$  als eine durch die Kamerakennkurve bewertete Bestrahlung der Kamerasensoroberfläche  $H_i$ , wobei diese sich aus der Bestrahlungsstärke  $E_i$  über die Dauer der Belichtungszeit  $\Delta t_j$  ergibt:

$$Z_{ij} = f(E_i \Delta t_j) \quad (4.2)$$

Debevec geht von einer monoton steigenden und damit invertierbaren Kamerakennkurve aus und definiert für deren inverse Funktion  $f^{-1}(Z_{ij})$  folgenden Zusammenhang:

$$f^{-1}(Z_{ij}) = E_i \Delta t_j \quad (4.3)$$

Durch Anwendung des natürlichen Logarithmus auf beiden Seiten wird die abgeleitete Bewertungsfunktion  $g(Z_{ij})$  folgendermaßen definiert:

$$\begin{aligned} \ln f^{-1}(Z_{ij}) &= \ln E_i + \ln \Delta t_j \\ g(Z_{ij}) &= \ln E_i + \ln \Delta t_j \end{aligned} \quad (4.4)$$

Die Szene wird als statisch angenommen, daher ist die Bestrahlungsstärke  $E_i$  über alle Bilder konstant für jeden Pixel an der Position  $i$ . Da die Pixelwerte  $Z_{ij}$  mit 8 Bit quantisiert sind, liegt der Wertebereich zwischen  $Z_{min} = 0$  und  $Z_{max} = 255$ . Damit ist auch die Bewertungsfunktion  $g(Z_{ij})$  mit 256 Funktionswerten vollständig beschrieben. Debevec zielt darauf ab,  $g(Z_{ij})$  sowie die Bestrahlungsstärke  $E_i$  für alle Pixelpositionen so zu rekonstruieren, dass Gleichung 4.4 nach der Methode der kleinsten Quadrate für alle  $Z_{ij}$  möglichst gut erfüllt wird. Daraus formuliert Debevec ein lineares Ausgleichsproblem mit folgender zu minimierender Zielfunktion:

$$\min \sum_{i=1}^N \sum_{j=1}^P (w(Z_{ij}) (g(Z_{ij}) - \ln E_i - \ln \Delta t_j))^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} w(z) (g(z-1) - 2g(z) + g(z+1)) \quad (4.5)$$

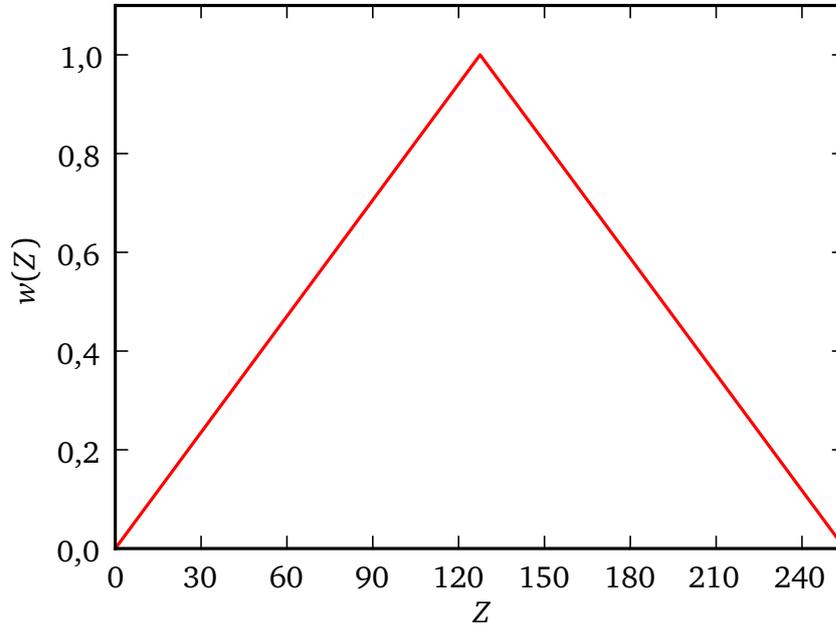
Der erste Term führt eine Minimierung nach der Methode der kleinsten Quadrate von Gleichung 4.4 über alle  $N$  Pixelpositionen und  $P$  Belichtungszeiten durch. Der zweite Term bewirkt zusätzlich eine Glättung des Kurvenverlaufs, indem die zweite Ableitung der Kamerakennkurve über alle Pixelwerte minimiert wird. Die Wahl des Parameters  $\lambda$  beeinflusst die Stärke dieser Glättung. Die Gewichtungsfunktion  $w(Z)$  bewertet den Einfluss von einzelnen Pixelwerten und ist folgendermaßen definiert:

4.11):

$$w(Z) = 1 - \left| \frac{Z}{\frac{1}{2}(Z_{max} - Z_{min})} - 1 \right| \quad (4.6)$$

Zweck dieser Gewichtung ist die Einstufung der Verlässlichkeit von Pixelwerten und damit auch der daraus abgeleiteten Strahldichten. Abgewertet wird die Verlässlichkeit von dunklen Pixelwerten auf Grund des schlechteren Signal-Rausch-Verhältnisses sowie von helleren Pixelwerten auf Grund des möglichen Clippings.

Debevec verwendet für die Ausgleichsrechnung nicht alle  $N$  Pixelpositionen, sondern lediglich eine so große Teilmenge, dass das lineare Gleichungssystem überbestimmt ist. Dies ist mit der Ungleichung  $N(P - 1) > (Z_{max} - Z_{min})$  erfüllt.



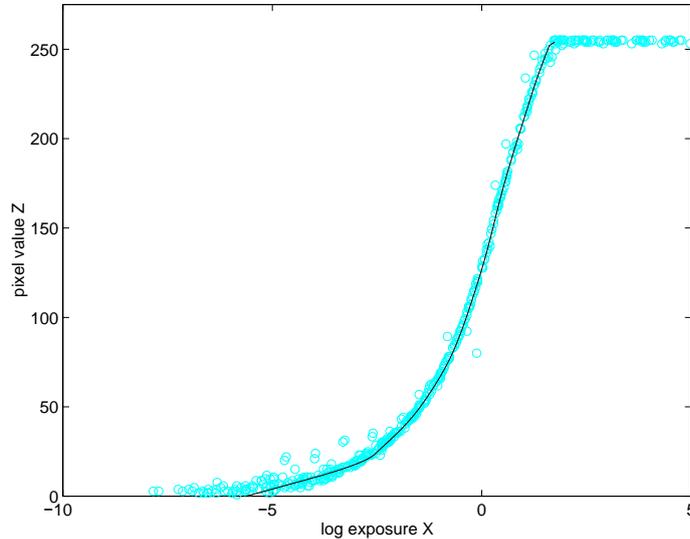
**Abbildung 4.11:** Gewichtungsfunktion  $w(Z)$  nach Debevec [DM97]

Nach erfolgter Rekonstruktion der Kamerakennkurve (siehe Abb. 4.12) kann für beliebige Pixelpositionen  $i$  die Bestrahlungsstärke  $E_i$  berechnet werden. Um dabei Werte mit möglichst hohem dynamischem Umfang (HDR) und geringem Rauschen zu erhalten, wird ein gewichtetes Mittel über alle Belichtungszeiten verwendet:

$$\ln E_i = \frac{\sum_{j=1}^P w(Z_{ij}) (g(Z_{ij}) - \ln \Delta t_j)}{\sum_{j=1}^P w(Z_{ij})} \quad (4.7)$$

### Rekonstruktion nach Grossberg und Nayar

Grossberg und Nayar [GN03] [GN04] führen ein als *Empiric Model of Response* (EMoR) bezeichnetes empirisches Modell von Kamerakennkurven ein, welche auf den Daten von 201 unterschiedlichen gemessenen Kamerakennkurven basiert (siehe Abb. 4.13). Mit Hilfe des empirischen Modells wird eine präzise

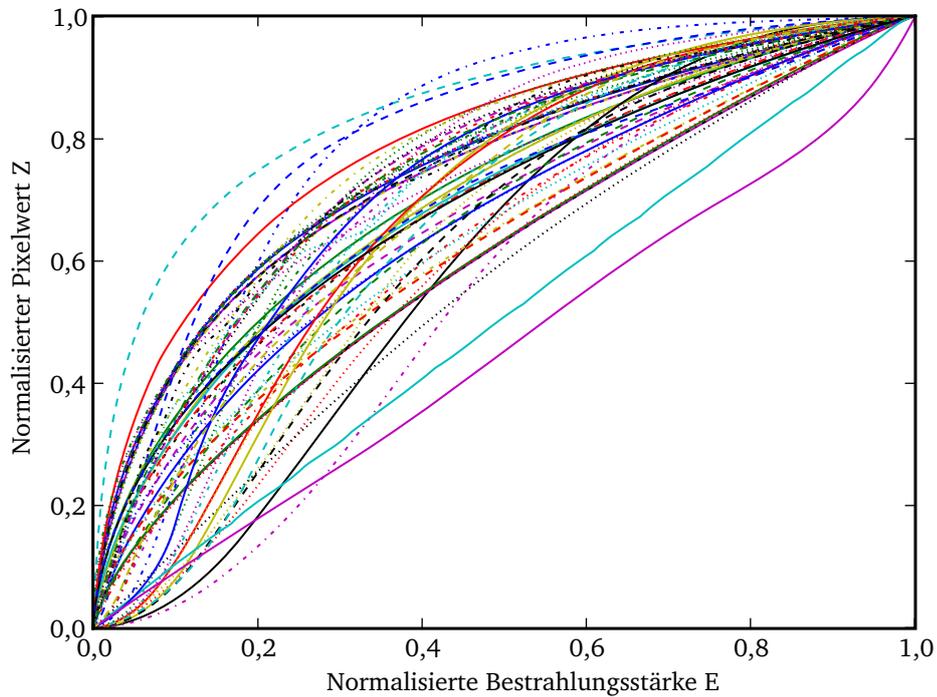


**Abbildung 4.12:** Rekonstruierte Kamerakennkurve nach Debevec [DM97]

Parametrisierung von beliebigen Kamerakennkurven mit möglichst wenig Koeffizienten angestrebt. Die Kamerakennkurven werden dazu als Vektoren in einem endlich dimensionalen Koordinatensystem diskretisiert und einer Hauptkomponentenanalyse unterzogen. Ergebnis ist eine Anzahl von endlich vielen Basisvektoren, welche den Vektorraum aller Kamerakennkurven mit absteigendem Anteil an der Gesamtstreuung aufspannen. Eine beliebige Kamerakennkurve  $f(H)$  wird dann als Summe der Mittelwerte aller Vektoren,  $f_0$ , und einer Linearkombination aus den ersten  $M$  berechneten Basisvektoren  $h_i$  beschrieben:

$$f(H) = f_0(H) + \sum_{i=1}^M c_i h_i(H) \quad (4.8)$$

Eine unbekannte Kamerakennkurve kann in diese Parametrisierung umgewandelt werden, indem als Eingabedaten entweder einige Paare von Pixelwerten mit zugehöriger Bestrahlungsstärke hinzugezogen werden oder eine Belichtungsreihe verwendet wird. Für den ersten Fall kann Grossberg bereits mit 3 Koeffizienten und 6 Messpaaren mit recht großer Genauigkeit die Kamerakennkurve abschätzen (siehe Abb. 4.14).



**Abbildung 4.13:** Auswahl der 201 gemessenen Kamerakennkurven von Grossberg und Nayar [GN03]

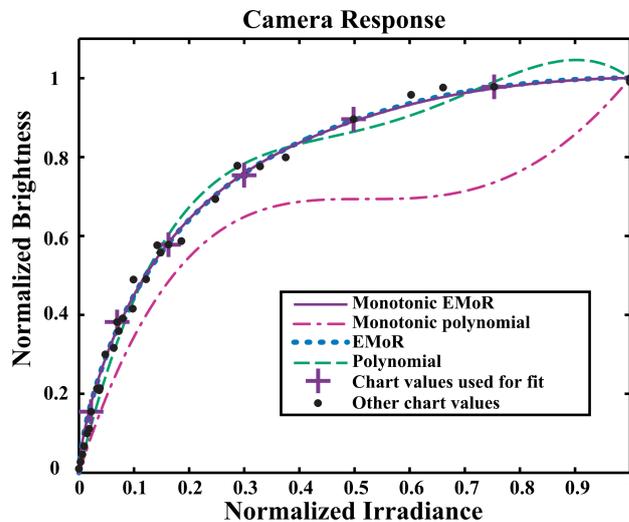
Für eine Parametrisierung anhand einer Belichtungsreihe rekonstruiert Grossberg die inverse Kamerakennkurve  $f^{-1} = g$ . Zunächst werden neue Basisvektoren  $h^{-1}$  durch eine Hauptkomponentenanalyse der inversen 201 Kamerakennkurven aufgestellt und es gilt:

$$g(Z) = g_0(Z) + \sum_{i=1}^M c_i h_i^{-1}(B) \quad (4.9)$$

Für Bilder mit unterschiedlichen Belichtungszeiten  $\Delta t_j$  und  $\Delta t_k$  gilt dabei folgender Zusammenhang für deren Pixelwerte:

$$g(Z_{ij}) - \frac{\Delta t_j}{\Delta t_k} g(Z_{ik}) = 0 \quad (4.10)$$

Aus diesem linearen Zusammenhang ermittelt Grossberg die Koeffizienten der inversen Kamerakennkurve durch lineare Ausgleichsrechnung nach der



**Abbildung 4.14:** Empirisches Modell (EMoR) zur Abschätzung der Kamerakennkurve nach Grossberg und Nayar [GN03]. Zur Abschätzung der Kurve (lila) werden 3 Koeffizienten und 6 Messwerte (mit + dargestellt) verwendet.

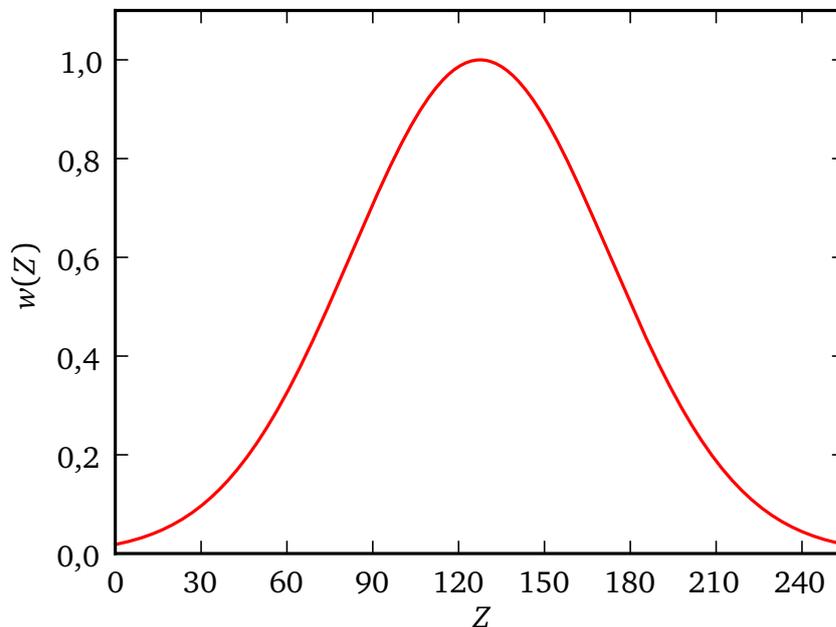
Methode der kleinsten Quadrate. Auch hier können bereits mit 3 Koeffizienten sehr gute Ergebnisse erzielt werden.

### Rekonstruktion nach Robertson

Der Ansatz von Robertson et. al. [RBS99] [RBS03] liefert für eine Belichtungsreihe als Eingabe, ähnlich zum Ansatz von Debevec, die inverse Kamerakennkurve als diskretisierten Vektor und ein HDR-Bild. Robertson rekonstruiert zunächst die Kamerakennkurve durch ein iteratives lineares Lösungsverfahren und leitet daraus eine Gewichtungsfunktion ab, welche Bildpixel auf Grundlage der Wahrscheinlichkeit von Signalrauschen bewertet.

Die inverse Kamerakennkurve  $g(Z) = f^{-1}(Z)$  sowie die Bestrahlungsstärken  $E_i$  werden durch Minimierung der folgenden Zielfunktion mit Hilfe des Gauß-Seidel-Verfahrens bestimmt:

$$\min_{g, E_i} \sum_{i=1}^N \sum_{j=1}^P w(Z_{ij}) (g(Z_{ij}) - \Delta t_j E_i)^2 \quad (4.11)$$



**Abbildung 4.15:** Gewichtungsfunktion  $w(Z)$  nach Robertson et. al. [RBS99]

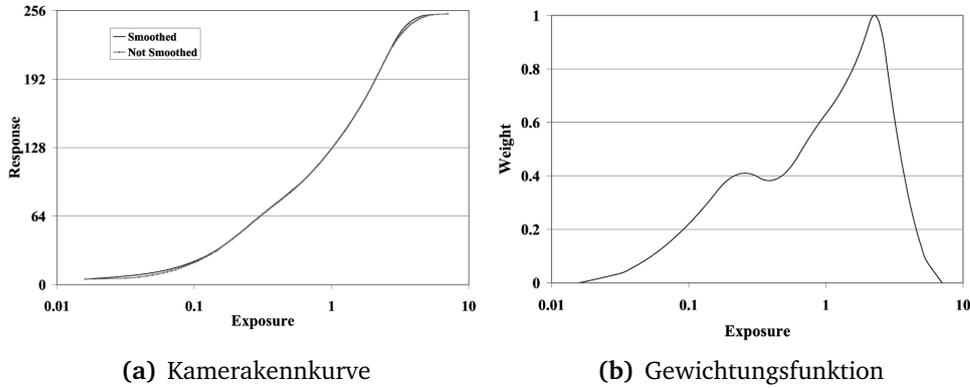
Die Gewichtungsfunktion  $w(Z)$  bewertet mittlere Pixelwerte höher und ist folgendermaßen definiert (siehe Abb. 4.15):

$$w(Z) = \exp\left(-4 \frac{(Z_{ij} - 127,5)^2}{127,5^2}\right) \quad (4.12)$$

Da nur relative Bestrahlungsstärken ermittelt werden können, wird zusätzlich folgende Skalierung festgelegt:

$$g(128) = 1.0 \quad (4.13)$$

Nach der Rekonstruktion der Kamerakurve können aus der Belichtungsreihe die Bestrahlungsstärken  $E_i$  durch ein gewichtetes Mittel über alle Bilder der Belichtungsreihe gewonnen werden:



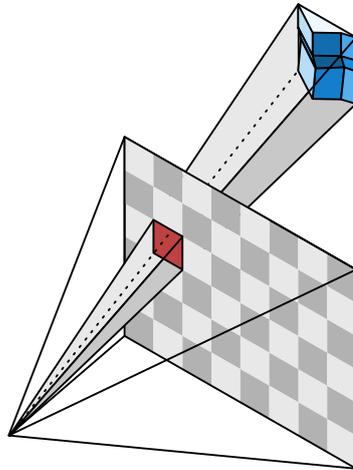
**Abbildung 4.16:** Rekonstruktion der Kamerakennkurve und Abschätzung der Gewichtungsfunktion nach Robertson [RBS03]

$$E_i = \frac{\sum_j^P c(g(Z_{ij})) \Delta t_j g(Z_{ij})}{\sum_j^P c(g(Z_{ij})) \Delta t_j^2} \quad (4.14)$$

Im Gegensatz zu Debevec wird durch das Quadrat der Belichtungszeit im Nenner ein Bildpixel proportional zu der zugehörigen Belichtungszeit gewichtet. Zudem verwendet Robertson in [RBS03] eine separate Gewichtungsfunktion  $c(g(Z_{ij}))$  mit dem Ziel Bildpixel anhand der Auftretungswahrscheinlichkeit von Rauschen zu bewerten (siehe Abb. 4.16). Für einen vorhergesagten Bestrahlungswert  $g(Z_{ij})$  wird das Gewicht aus dem Anstieg der Kamerakennkurve  $f(H)$  bei Verwendung einer logarithmischen Skalierung auf folgende Weise berechnet:

$$\begin{aligned} f_l(x) &= f(\log x) \\ c_l(x) &= \frac{df_l(x)}{dx} \\ c(\log x) &= c_l(x) \end{aligned} \quad (4.15)$$

Da  $f(H)$  lediglich eine diskrete Funktion ist, erfolgt die Berechnung auf Grundlage einer kontinuierlichen Repräsentation durch kubische Splines.



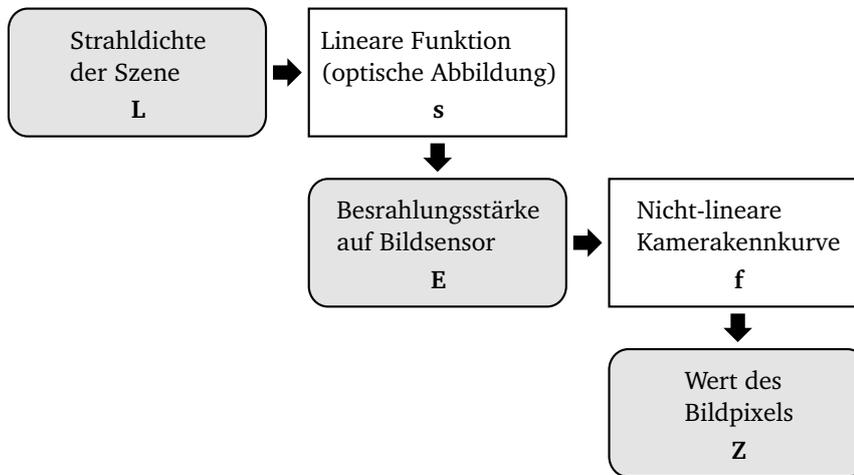
**Abbildung 4.17:** Der RGB-Wert eines Bildpixels ist die Bewertung der von der Kamera empfangenen spektralen Strahldichte aus der realen Szene. Die genaue Geometrie der Szene (blau) und damit die Position der Senderfläche ist jedoch unbekannt.

### 4.3 Light-Probe-Synthese ohne bekannte Szenengeometrie

Nachfolgend wird ein Verfahren zur Rekonstruktion der Beleuchtung einer realen Szene ausschließlich aus den Bildern einer 3D-Photo-Collection vorgestellt. Dabei wird Szenengeometrie weder vorausgesetzt noch manuell modelliert. Ziel des Verfahrens ist die Sammlung von Beleuchtungsinformationen für eine bestimmte Position im Raum. Die Beleuchtungsinformationen werden in Form einer HDR-Environment-Map repräsentiert, mit dem Anspruch den gleichen Informationsgehalt wie eine manuell gemessene Light-Probe zu enthalten. Daher kann dieses Vorgehen auch als Synthese einer Light-Probe angesehen werden.

#### 4.3.1 Abbildung von Strahldichte auf RGB-Werte

Für jedes Foto sind extrinsische und intrinsische Kameraparameter bekannt. Daher kann aus Position, Orientierung und Öffnungswinkel der Kamera ermit-



**Abbildung 4.18:** Modell nach [GN04] für die Abbildung der aus der Szene empfangenen Strahldichte  $L$  über die Bestrahlungsstärke  $E$  des Bildsenausschnitts auf den Wert  $Z$  durch eine Kamera.

telt werden, welchen Ausschnitt der realen Umgebung ein einzelner Bildpixel wiedergibt (siehe Abb. 4.17). Zwar geben die RGB-Werte eines aufgenommenen Bildes die Beleuchtungssituation der fotografierten Szene wieder, allerdings kann nicht ohne weiteres auf die tatsächlich aus der Szene einfallende Leuchtdichte geschlossen werden.

Zunächst sind die RGB-Werte das Ergebnis einer trichromatischen Abbildung des spektralen Umgebungslichtes durch den Kamerasensor, die nicht umkehrbar ist. Genau betrachtet führen die Sensorantwortkurven der Kamera auch keine der Definition der Leuchtdichte äquivalente spektrale Bewertung des Lichtes durch, da sie hierfür als eine beliebige Linearkombination die Empfindlichkeitskurve  $V(\lambda)$  nachbilden müssten (siehe Abschnitt 2.1.4). Daher ist es auch präziser von einer für Rot, Grün und Blau spektral gewichteten Strahldichte zu sprechen.

Selbst wenn die spektrale Bewertung der Kamera für den Rot-, Grün und Blaukanal als gegeben hingenommen wird, ist der Zusammenhang zwischen einfallender Bestrahlungsstärke und resultierenden R, G und B-Werten komplex. Ein häufig verwendetes Kameramodell [GN04] bildet die eingehende Strahldichte der Szene  $L$  unter Verwendung der optischen Abbildungsfunkti-

on  $s$  und der Kamerakennkurve  $f$  auf resultierende RGB-Werte der Kamera  $Z$  (siehe Abb. 4.18) ab. Die Funktion  $s$  modelliert vor allem Auswirkungen des Kameraobjektivs wie etwa Vignettierung. Sie variiert mit der Position auf dem Kamerasensor, ist jedoch stets linear zu Strahldichte der Szene [GN04]. Die Kamerakennkurve  $f$  ist die nicht-lineare Abbildung der über eine bestimmte Belichtungszeit wirksamen Bestrahlungsstärke  $E$  auf den resultierenden Wert  $Z$  des Pixels. Im Gegensatz zu  $s$  kann die Kamerakennkurve als unabhängig von der Position im Bildraum und auch unabhängig von der Wellenlänge angenommen werden [BF02].

Um von Pixelwerten auf die Strahldichte der Szene zu schließen, muss daher eine Rekonstruktion der Kamerakennkurve durchgeführt werden. Im Rahmen des hier vorgestellten Verfahrens wurde dabei die Methode von Robertson et. al. [RBS03] gewählt (siehe Abschnitt 4.2.2).

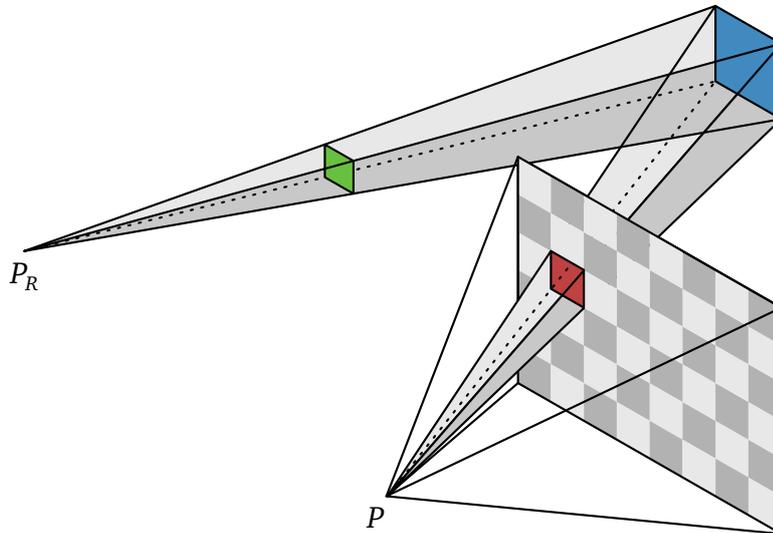
### 4.3.2 Approximation der Szenengeometrie

Informationen über die Strahldichte der Umgebung liegen nur an den Positionen vor, an denen auch Bilder aufgenommen wurden. Wenn die Geometrie der Szene jedoch näherungsweise bekannt ist, dann kann – unter Annahme vollständig diffusen Reflexionsverhaltens – auf die spezifische Ausstrahlung der abgebildeten Szenenoberflächen geschlossen werden. Für beliebige Punkte im Raum kann dann die aus den umliegenden Szenenoberflächen einfallende Strahldichte berechnet werden (siehe Abb. 4.19).

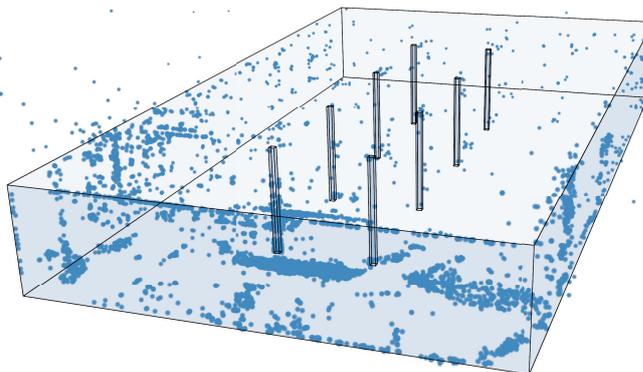
Da keine vollständige Szenengeometrie vorliegt, wird eine Approximation der Szenengeometrie anhand der Punktwolke der 3D-Photo-Collection vorgenommen. Jedem Punkt der Punktwolke sind in den Bildern detektierte Merkmalspunkte zugeordnet. Somit ist auch bekannt, welcher Punkt in welchen Bildern sichtbar ist. Die 3D-Punktwolke ist in der Regel aber nur eine grobe und verrauschte Repräsentation der realen Szenenoberflächen (siehe Abb. 4.20), so dass nur für wenige Bildpixel Tiefeninformationen vorliegen, die zudem auch fehlerhaft sein können.

Es wurde ein Ansatz gewählt, der aus den in einem Bild sichtbaren Punkten den Median ihrer Tiefenwerte verwendet. Der Tiefenwert  $d$  ist die Z-Koordinate des Punktes in Kamerakoordinaten und errechnet sich aus Punktposition  $\vec{x}$ , Kamerastandort  $\vec{p}$  und normierter Blickrichtung  $\vec{v}_0$  wie folgt:

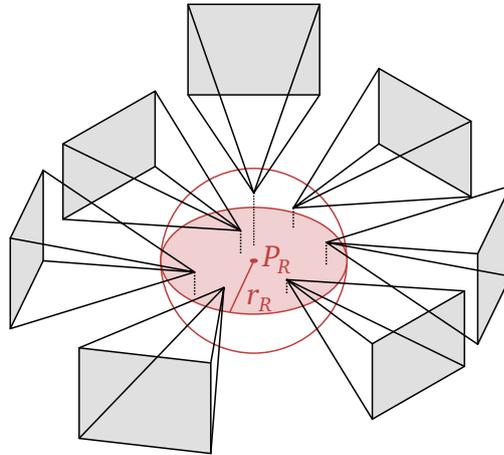
$$d = \vec{v}_0 \cdot (\vec{p} - \vec{x}) \quad (4.16)$$



**Abbildung 4.19:** Rückprojektion der vom Punkt  $P$  empfangenen Strahldichte (rot) auf die Szenenoberfläche (blau) und anschließende Projektion der Bestrahlungsstärke in Richtung  $P_R$  unter Annahme vollständig diffusen Reflexionsverhaltens der Szenenoberfläche



**Abbildung 4.20:** Punktwolke einer 3D-Photo-Collection als grobe und verrauschte Repräsentation der Szenengeometrie



**Abbildung 4.21:** Kameras einer 3D-Photo-Collection, die innerhalb des um den Objektmittelpunkt  $P_R$  liegenden Radius  $r_R$  liegen.

Aus der Menge der nach Größe sortierten  $n$  Tiefenwerte  $d_i$  ( $i \in n$ ) ergibt sich der Median  $\tilde{d}$  wie folgt:

$$\tilde{d} = \begin{cases} d_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2} \left( d_{\frac{n}{2}} + d_{\frac{n+1}{2}} \right), & \text{falls } n \text{ gerade} \end{cases} \quad (4.17)$$

An dem so gewonnenen Tiefenwert  $\tilde{d}$  wird dann als ein geometrisches Hilfsobjekt eine Ebene in das Sichtvolumen der Kamera konstruiert, auf welche das Bild projiziert wird. Durch Verwendung des Medians wird eine Robustheit gegenüber verrauschten Tiefenwerten erzielt.

Der einfache Ansatz der Geometriepäsentation führt in den meisten Szenen zu einem sehr großem Projektionsfehler bei der Reprojektion der Strahldichten. Daher wird um einen Punkt  $P_R$ , für den die Strahldichten der Umgebung akkumuliert werden sollen, ein Suchradius  $r_R$  festgelegt (siehe Abb. 4.21). Nur Bilder, deren Kamerastandorte innerhalb dieses Radius liegen, sollen für die Reprojektion nach  $P_R$  verwendet werden.

### 4.3.3 Akkumulation der Strahldichten

Die Synthese der Light-Probe erfolgt unter Nutzung der Grafikhardware, um eine echtzeitfähige Extraktion der Beleuchtungsdaten durchzuführen.

Die Strahldichten der Umgebung für den Punkt  $P_R$  werden dabei in Form einer Cube-Map repräsentiert. Der Farbkanal eines RGB-Tripels wird dabei im Fließkommaformat mit einfacher Genauigkeit kodiert (32 Bit Float), um Strahldichten mit hohem dynamischen Umfang wiederzugeben.

Für jede Seite der Cube-Map werden die Strahldichten aus den umliegenden Bildern akkumuliert. Dazu wird zunächst für jede Seite ein Kamerasichtvolumen mit einem horizontalen und vertikalen Öffnungswinkel von  $90^\circ$  konstruiert, das unter Verwendung der Kameraposition  $P_R$  die entsprechende Seite der Cube-Map als Bildebene verwendet.

Für die Berechnung der Strahldichte eines Texels werden alle Bilder berücksichtigt, welche im Sichtvolumen des Texels liegen. Bilder mit unterschiedlichen Belichtungszeiten werden so, analog zu Debevec [Deb98], über ein gewichtetes Mittel zur Erweiterung des Dynamikumfangs verwendet.

Die Strahldichte  $L_T$  des Texels wird berechnet, indem ein Strahl entlang des Sichtvolumens in Richtung  $\vec{d}$  geschickt wird und für Schnittpunkte mit allen  $n$  konstruierten Bildebenen deren jeweilige Strahldichte  $L_j$  gewichtet summiert wird:

$$L_T = \frac{\sum_{j=1}^n w(Z_j) \cdot L_j}{\sum_{j=1}^n w(Z_j)} \quad (4.18)$$

Die für jedes Bild getrennt abgeleitete Strahldichte  $L_j$  ergibt sich aus der Bestrahlungsstärke  $E_j$ , welche sich unter Nutzung der inversen Kamerakennkurve  $f^{-1}$  aus dem am Schnittpunkt des Strahls abgetasteten Pixelwert  $Z_j$  und der Belichtungszeit  $\Delta t$  ergibt. Da alle Oberflächen diffus sind, gilt  $L_j = E_j / \pi$ .

$$E_j = \frac{f^{-1}(Z_j)}{\Delta t} \quad (4.19)$$

Die Gewichtungsfunktion  $w(Z)$  ist wie in [Deb98] definiert (siehe Abschnitt 4.2.2) und stuft die Verlässlichkeit von Pixelwerten ein. Abgewertet wird die Verlässlichkeit von dunklen, verrauschten sowie überbelichteten Pixelwerten.

Die Gleichung 4.18 wird für jeden Texel der Cube-Map parallel unter Verwendung von drei Renderdurchgängen ausgewertet. Im ersten Renderdurchgang wird die Summe im Zähler für jeden Texel ausgewertet und das

Ergebnis wird in der temporären Textur  $T_{Den}$  gespeichert, welche die gleiche Seitenlänge wie die Seiten der Cube-Map besitzt. Analog zum ersten Durchlauf, wird im zweiten Durchlauf die Summe im Nenner gebildet und in eine zweite temporäre Textur  $T_{Num}$  je Texel gespeichert. Im letzten Durchlauf wird schließlich je Texel der Quotient aus  $T_{Den}$  und  $T_{Num}$  gebildet und in die Seite der Cube-Map geschrieben. Auf diese Weise wird eine Parallelisierung der Akkumulation von Strahldichten auf der Grafikhardware erreicht.

#### 4.3.4 Generierung direktonaler Lichtquellen

Um eine echtzeitfähige Beleuchtung von virtuellen Objekten zu ermöglichen, werden aus der synthetisierten Light-Probe 16 direktonale Lichtquellen bestimmt. Dazu wird zunächst die Cube-Map in eine einzelne 2D-Textur unter Verwendung einer Zylinderprojektion abgebildet. An die gegebene Texturkoordinaten  $(u; v)$  wird über Berechnung der sphärischen Winkel  $(\phi; \theta)$  ein Texel der Cube-Map mit dem Richtungsvektor  $\vec{d}$  abgebildet:

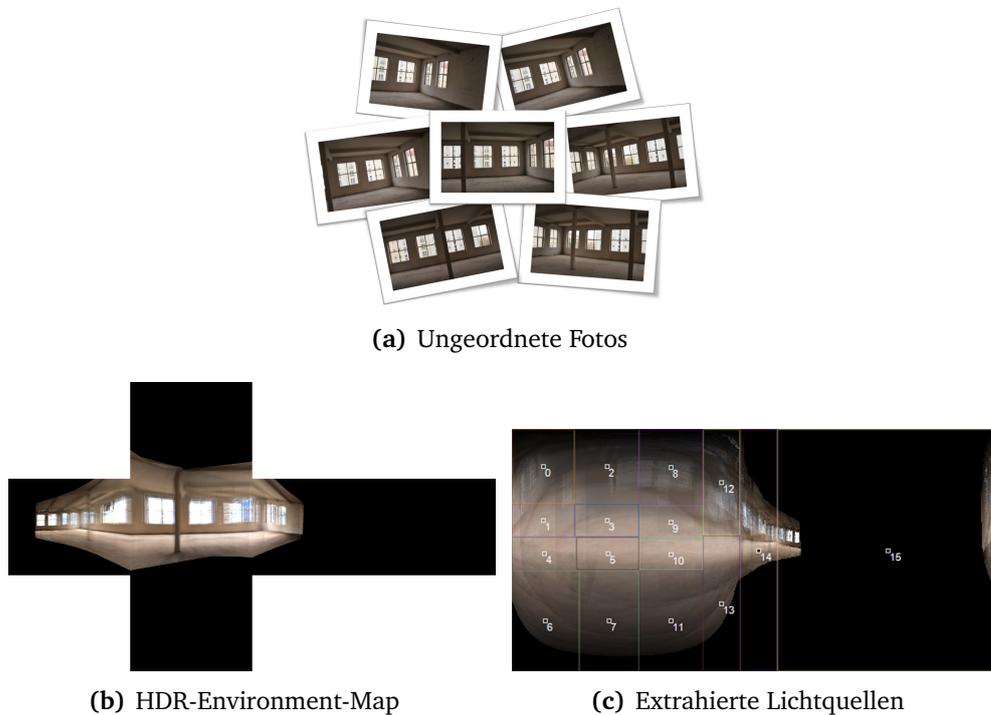
$$\phi = 2\pi u \quad (4.20)$$

$$\theta = \pi(0.5 + v) \quad (4.21)$$

$$\vec{d} = \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix} \quad (4.22)$$

Auf diese Textur wird ein Median-Cut-Algorithmus [Deb05] angewendet, der 16 direktonale Lichtquellen extrahiert (siehe Abb. 4.22(c)). Dies erfolgt, indem die Textur in 16 Segmente mit gleich großer Energie aufgeteilt wird. Aus jedem Segment wird eine direktonale Lichtquelle erzeugt, indem das Segmentzentrum die Richtung liefert und die Energiesumme die Strahldichte. Die erzeugten Lichtquellen werden zur interaktiven Darstellung des virtuellen Objekts verwendet.

Für die Synthese von virtuellen Anteilen und realen Bildern wird anschließend ein Tone-Mapping-Operator auf die virtuellen Anteile angewendet, um die linearen RGB-Werte mit hohem dynamischem Umfang wieder in den RGB-Farbraum der Bilder umzuwandeln. Dies erfolgt mit Hilfe der rekonstruierten Kamerakennkurve.



**Abbildung 4.22:** Synthese von Beleuchtungsdaten aus den Bildern der 3D-Photo-Collection in die HDR-Environment-Map

#### 4.3.5 Ergebnisse

Wie in Abbildung 4.23 und 4.24 zu sehen, wird das Objekt realistisch von aus den Fotos rekonstruiertem Licht beleuchtet. Dadurch wird eine wesentlich glaubwürdigere Synthese von virtuellen und realen Anteilen erreicht. In Abbildung 4.25 wird das Objekt akkurat durch das von hinten eingestrahlte Licht beleuchtet.

Wie erläutert, ist die Light-Probe-Synthese auf der Grafikhardware vollständig parallelisierbar und kann so auch für große Bildmengen in Echtzeit durchgeführt werden. Damit können Operationen wie das Einfügen oder Verschieben von neuen virtuellen Objekten interaktiv umgesetzt werden, indem für neue Objektpositionen eine neue Light-Probe generiert wird. Für die Generierung von direktionalen Lichtquellen aus der Environment-Map muss allerdings noch teilweise auf die CPU zurückgegriffen werden. Durch



**Abbildung 4.23:** Synthese von 3D-Photo-Collection und virtuellem Modell



(a) Virtuelles Modell ohne extrahierte HDR-Beleuchtung (b) Virtuelles Modell mit extrahierter HDR-Beleuchtung

**Abbildung 4.24:** Synthese von realen und virtuellen Anteilen unter Nutzung der Beleuchtungsrekonstruktion ohne Szenengeometrie



(a) Frontansicht



(b) Von-Oben-Ansicht

**Abbildung 4.25:** Ansichten eines 3D-Modells mit von hinten kommender Beleuchtung, welche aus den Fotos extrahiert wurde

Verlagerung auf die Grafikhardware kann hier noch eine wesentlich schnellere Ausführung erreicht werden.

Nachteil am vorgestellten Verfahren ist das Fehlen detaillierter Szenengeometrie. Zwar kann auch ohne vollständige Geometrie eine glaubwürdige Synthese von virtuellen und realen Anteilen erreicht werden (siehe Ansatz von Sato [SSI99]), jedoch müssen dann Einschränkungen bei der Wahl der realen Szene gemacht werden und geänderte Lichtpfade, wie zum Beispiel weiche Schattenwürfe, können nur angenähert wiedergegeben werden.

Für eine augmentierte Bildsynthese mit hoch realistischem Anspruch muss jedoch eine vollständige Szenengeometrie vorhanden sein. Erst dann können Leuchtdichte und Reflexionsgrade realer Oberflächen hinreichend genau berechnet werden. Diese werden von globalen Beleuchtungssimulationen wie

etwa Radiosity oder Photon-Mapping benötigt, um bei geänderten Lichtpfaden auch indirektes Licht und die Farbe von Schatten korrekt wiederzugeben. Im nachfolgenden Abschnitt wird daher ein Verfahren für die Synthese von Light-Probes vorgestellt, welches eine vollständig definierte Szenengeometrie verwendet und damit eine Bildsynthese mit fotorealistischem Anspruch ermöglicht.

## **4.4 Synthese von Light-Probes mit gegebener Szenengeometrie**

Im Gegensatz zum im vorherigen Abschnitt präsentierten Verfahren, soll nun für die Synthese von Light-Probes eine detaillierte Szenengeometrie verwendet werden. Dadurch kann die Bestrahlungsstärke von realen Oberflächenelementen wesentlich präziser rekonstruiert werden, da die Rückverfolgung von Strahldichten einzelner Bildpixel auf reale Oberflächen mit einem wesentlich geringeren Projektionsfehler behaftet ist.

Im Gegensatz zum vorherigen Verfahren sollen die Bilder nicht direkt in die Light-Probe gerendert werden. Stattdessen soll zunächst für sämtliche Oberflächen der realen Szene die spezifische Ausstrahlung ermittelt werden. Anschließend können dann für beliebige Positionen in der Szene sehr schnell Light-Probes generiert werden.

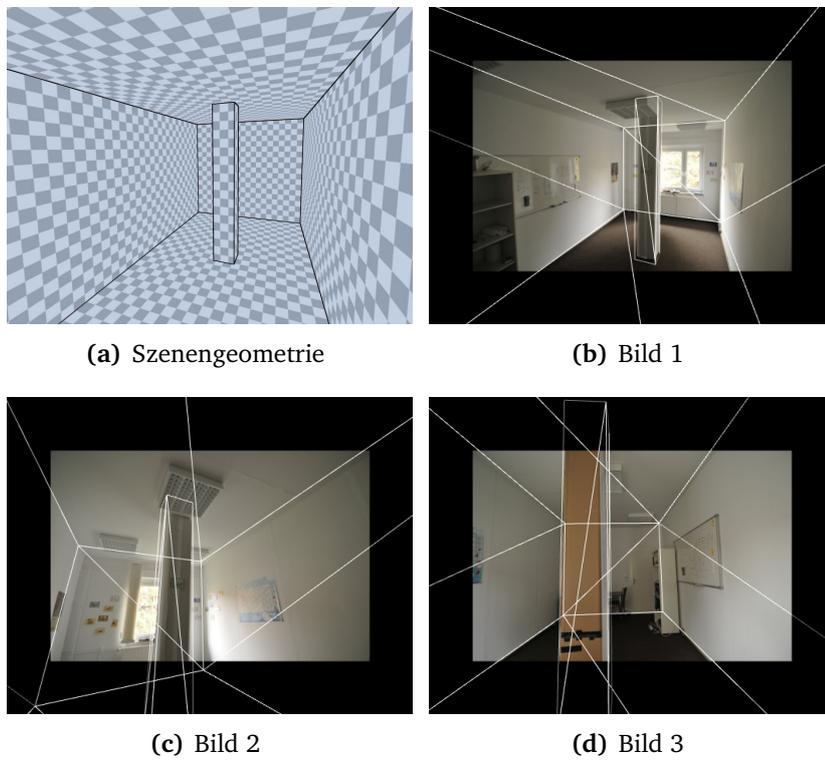
Ein weiterer Vorteil mit explizit definierter Szenengeometrie ist, dass Reflektionsgrade und Lichtquellen abgeschätzt werden, was für eine realistischere Bildsynthese von virtuellen und realen Anteilen genutzt werden kann.

### **4.4.1 Bereitstellung der Szenengeometrie**

Die Geometrie der Szene liegt in Form eines Polygonnetzes vor und ist in der Szene registriert, so dass in die Bilder projizierte Kanten der Geometrie deckungsgleich mit den realen Umrisen der Szene sind (siehe Abb. 4.26). Wie einige der in Abschnitt 4.2 vorgestellten Verfahren (z.B. [DRB97]) kann die reale Szene durch Ausnutzung der Epipolargeometrie modelliert werden. Dazu eignen sich 3D-Photo-Collections gut, da sie viele verschiedene Ansichten auf die Szene bieten und die Kameraparameter bekannt sind.

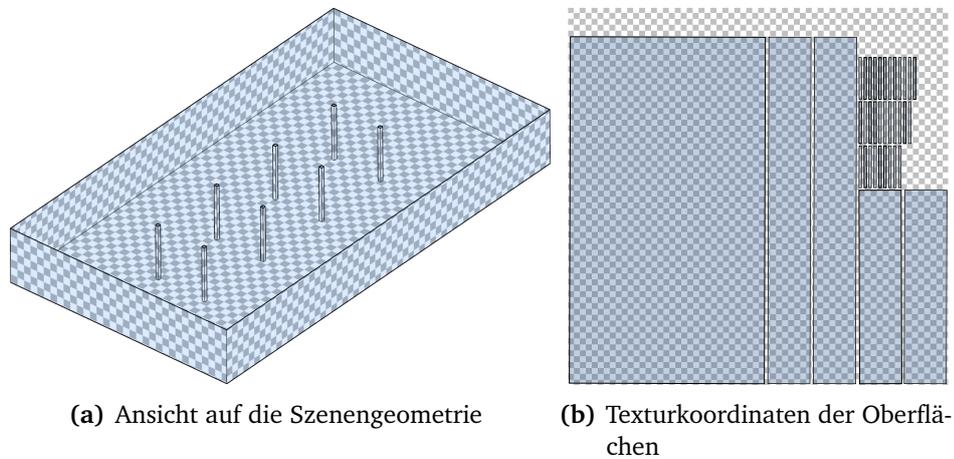
Mit Hilfe von Multi-View-Stereo-Algorithmen kann auch die Geometrie der

#### 4.4 Synthese von Light-Probes mit gegebener Szenengeometrie



**Abbildung 4.26:** Bilder der 3D-Photo-Collection mit registrierter Szenengeometrie

Szene automatisch aus einer 3D-Photo-Collection rekonstruiert werden. In [SCD<sup>+</sup>06] werden Ausführungsdauer und Genauigkeit der Rekonstruktion verschiedener Verfahren miteinander verglichen. Es existieren auch Verfahren, die Annahmen über die Beschaffenheit der realen Szene machen, um robustere Ergebnisse zu erzielen. In einem Verfahren von Furukawa [FCSS09a] [FCSS09b] etwa werden alle Oberflächen als an Achsen ausgerichtete Rechtecke modelliert, was das Verfahren allerdings auf Innenräume und Gebäude beschränkt. Nachteile von Multi-View-Stereo-Verfahren sind neben der langen Ausführungszeit allerdings, dass die rekonstruierte Geometrie meist verrauscht oder zu detailarm ist und bei bestimmten Verfahren auch Löcher aufweisen kann. Dadurch sind die Ergebnisse automatischer Rekonstruktionsverfahren oft nicht präzise genug für die Synthese von Light-Probes.



**Abbildung 4.27:** Szenengeometrie mit Texturatlas

#### 4.4.2 Verwendung eines Texturatlas

Durch die bekannte Szenengeometrie kann jeder Pixel eines Bildes auf einen Ausschnitt einer Oberfläche der Geometrie projiziert werden. Da, wie im vorherigen Verfahren gezeigt, einem Bildpixel eine Strahldichte zugeordnet wird, kann daraus bei der Annahme von vollständig diffus reflektierenden Oberflächen die spezifische Ausstrahlung der Oberfläche berechnet werden. Um jedoch für verschiedene Ausschnitte der Oberfläche das versendete Licht zu erfassen, wird zunächst eine Parametrisierung für alle Punkte der realen Oberflächen benötigt.

Diese Parametrisierung erfolgt, indem ein Texturatlas aufgebaut wird. Dabei muss jedem Eckpunkt aller Polygone der Szenengeometrie 2D-Texturkoordinaten zugewiesen werden, so dass sich kein Polygon im Texturraum mit einem anderen überschneidet und Polygone möglichst flächentreu abgebildet werden (siehe Abb. 4.27). Diese Parametrisierung kann durch Verfahren wie zum Beispiel LSCM (Least Squares Conformal Maps) [LPRMt02] durchgeführt werden. Damit werden die Oberflächen der Geometrie in finite Elemente, auch als Patches bezeichnet, unterteilt. Je nachdem, welche Auflösung für den Texturatlas verwendet wird, kann so eine feinere oder gröbere Aufteilung in Patches vorgenommen werden.

### 4.4.3 Bestimmung der Strahldichten

Die Bestimmung der Strahldichten erfolgt analog zu der Gleichung aus Abschnitt 4.3.1. Ebenfalls gilt, dass lediglich relative Strahldichten berechnet werden. Da jeder Patch als Lambertscher Strahler angenommen wird, ist die Strahldichte des Patches für alle Richtungen gleich. Daher wird die Strahldichte des Patches  $L_p$  berechnet, indem bei allen  $m$  Bildern für alle  $n$  Pixel  $Z_{i,j}$  das gewichtete Mittel aus deren zugeordneter Strahldichte  $L_{i,j}$  gebildet wird:

$$L_p = \frac{\sum_i^n \sum_{j=1}^m w(Z_{i,j}) \cdot V(p, Z_{i,j}) \cdot L_{i,j}}{\sum_i^n \sum_{j=1}^m w(Z_{i,j}) \cdot V(p, Z_{i,j})} \quad (4.23)$$

Die pixelspezifische Strahldichte  $L_{i,j}$  wird aus dem Wert  $Z_{i,j}$ , für Pixel  $j$  in Bild  $i$ , in Abhängigkeit der inversen Kamerakennkurve  $f^{-1}$  und der Belichtungszeit des Bildes  $\Delta t_i$  abgeleitet:

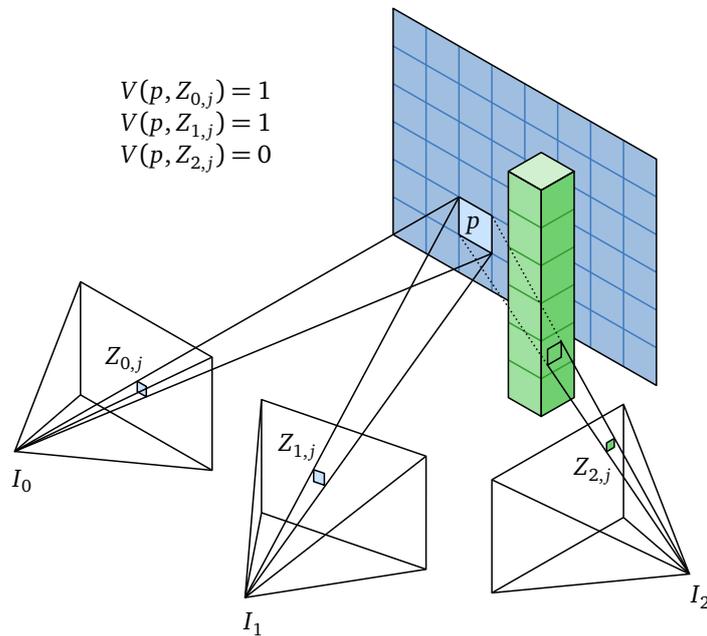
$$L_{i,j} = \frac{f^{-1}(Z_{i,j})}{\Delta t_i} \quad (4.24)$$

Die Sichtbarkeitsfunktion  $V(p, Z)$  bewertet, ob das durch Bildpixel  $Z$  repräsentierte Sichtvolumen, den durch Patch  $p$  repräsentierten Ausschnitt der realen Oberfläche zeigt (siehe Abb. 4.28):

$$V(p, Z) = \begin{cases} 0, & \text{falls der Pixel nicht den Patch zeigt} \\ 1, & \text{falls der Pixel den Patch zeigt} \end{cases} \quad (4.25)$$

Dabei ist es für  $V(p, Z) = 1$  notwendig, dass zum einen der Patch im Sichtvolumen des Bildpixels liegt und zum anderen, dass der Sichtstrahl nicht durch andere Geometrie verdeckt wird (siehe Abb. 4.29). Dieser Verdeckungstest kann mit Verfahren wie Shadow Volumes [Cro77] oder Shadow Mapping [Wil78] durchgeführt werden.

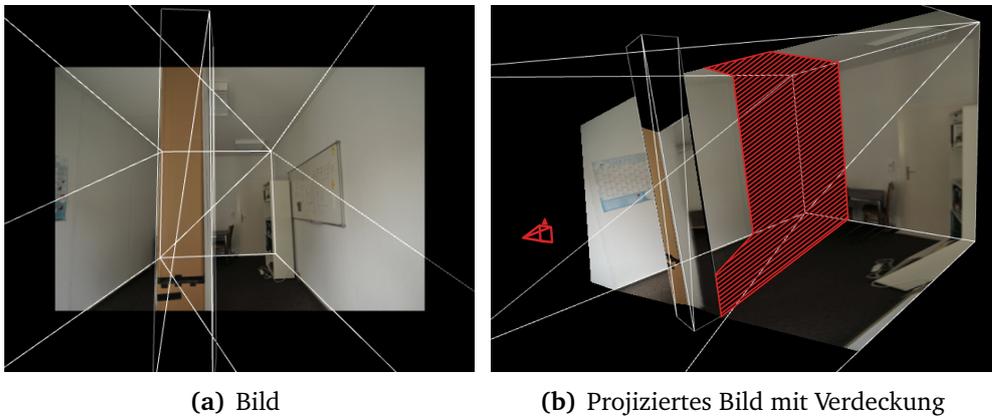
Die Berechnung der Strahldichten aller Patches lässt sich durch die Nutzung von Shadern auf der Grafikhardware parallelisieren. Die Strahldichten der Patches werden nach Gleichung 4.23 für den Rot-, Grün- und Blaukanal der Bilder getrennt berechnet und als Texel im Texturatlas kodiert. Jeder Farbkanal wird dabei im Fließkommaformat mit einfacher Genauigkeit (32



**Abbildung 4.28:** Werte der Sichtbarkeitsfunktion zwischen dem Patch  $p$  und den Pixeln aus den drei Bildern  $I_0$ ,  $I_1$  und  $I_2$

Bit) kodiert, um die Strahldichten mit hohem dynamischen Umfang wiederzugeben. Die Berechnung der Sichtbarkeit wird mit Hilfe von Shadow Mapping durchgeführt. Insgesamt werden folgende drei Renderdurchgänge durchgeführt:

1. Im ersten Renderdurchgang wird in einer separaten Textur  $T_{Den}$  der Zählerterm  $w(Z_{i,j}) \cdot V(p, Z_{i,j}) \cdot L_{i,j}$  für jeden Patch  $p$  über alle Bilder  $i$  summiert. Die Berechnung der Summe erfolgt, indem alle Bilder in den Texturatlas der Geometrie unter Nutzung von additivem Rendering abgebildet werden.
2. Ähnlich zum ersten Renderdurchgang wird in einer separaten Textur  $T_{Num}$  der Nennerterm  $w(Z_{i,j}) \cdot V(p, Z_{i,j})$  für jeden Patch  $p$  über alle Bilder  $i$  summiert.
3. Im dritten Renderdurchgang wird für jeden Patch der Quotient aus Zähler- und Nennerterm gebildet, indem aus den beiden Texturen



**Abbildung 4.29:** Verdeckte Bereiche (schraffiert) müssen bei der Projektion von Strahldichten auf die Oberflächen berücksichtigt werden - zum Beispiel durch Shadow Mapping.

$T_{Den}$  mit  $T_{Num}$  texelweise dividiert wird. Die Strahldichte von Patches, bei denen der Nennerterm auf Grund fehlender Bildabdeckung einen gewissen Schwellwert unterschreitet, wird als unbekannt markiert.

#### 4.4.4 Generierung von Light-Probes

Die berechneten Strahldichten liegen nun im Texturatlas als im Fließkommaformat kodierte RGB-Textel vor. Eine Light-Probe kann nun für einen beliebigen Punkt  $P_R$  in Form einer HDR-Cube-Map erstellt werden. Dazu wird jeweils für jede Seite der Cube-Map eine Kamera an der Position  $P_R$  mit der Seite der Cube-Map als Bildebene konstruiert und die mit dem Texturatlas texturierte Szenengeometrie in die Cube-Map gerendert.

Da die Szenengeometrie in der Regel nur aus wenigen Polygonen besteht, kann das Rendern in die Cube-Map sehr schnell ausgeführt werden, so dass für beliebige  $P_R$  in Echtzeit HDR-Cube-Map generiert werden können. Für die Darstellung von entsprechend beleuchteten virtuellen Objekten können mit dem Verfahren von Debevec [Deb05] wieder direktionale Lichtquellen gewonnen werden (siehe Abschnitt 4.3.4).



**Abbildung 4.30:** Virtuelles Objekt, dessen Beleuchtung aus den berechneten Strahldichten der realen Oberflächen erzeugt wurde

#### 4.4.5 Ergebnisse

Mit dem vorgestellten Verfahren wurde für verschiedene Szenen auf Basis einer 3D-Photo-Collection und gegebener Szenengeometrie die Beleuchtung der realen Szene berechnet. In die Szene wurde ein virtuelles Objekt platziert. Ebenfalls wurde an dieser Position eine Light-Probe generiert und aus dieser wiederum 16 direktionale Lichtquellen, welche zur Beleuchtung des virtuellen Objektes verwendet wurden. Ergebnisse dieser Synthese sind in den Abbildungen 4.32 und 4.30 zu sehen.

Die Ansicht auf die Szene wurde passend zu einer Kamerapose der 3D-Photo-Collection gewählt, so dass das reale Bild im Hintergrund eingeblendet werden kann. Die hellste der 16 direktionalen Lichtquellen wird zur Erzeugung eines Schattens verwendet. Durch mehrfaches Rendern des Schattens mit leichtem Variieren der Lichtquellenposition werden weiche Schattenkonturen realisiert.

Gegenüber der Light-Probe-Synthese ohne Szenengeometrie kann die Be-



**Abbildung 4.31:** Eingefügte virtuelles Objekt mit stark reflektierendem Material

leuchtung der realen Oberflächen nicht nur positionsunabhängig, sondern auch wesentlich genauer rekonstruiert werden. Wenn alle realen Oberflächen fotografiert wurden, kann auch eine vollständige Berechnung der Strahldichten aller Patches erreicht werden. Aus einem vollständigen Texturatlas können für beliebige Positionen innerhalb der realen Szene auch immer vollständige Light-Probes erzeugt werden (siehe Abb. 4.33 und 4.34).

Wie in Abbildung 4.31 zu sehen, können mit vollständigen Light-Probes virtuelle Objekte mit metallisch glänzender Oberfläche dargestellt werden. Auf der Oberfläche des virtuellen Objekts spiegelt sich die Umgebung der realen Szene. Diese Reflexion der Umgebung wird unter Verwendung der erzeugten HDR-Cube-Map mit Reflection-Mapping wiedergegeben.

## 4.5 Auswertung

### 4.5.1 Zusammenfassung und Vergleich beider Verfahren

Die in Abschnitt 4.3 und 4.4 vorgestellten Verfahren führen eine Extraktion der realen Beleuchtungsinformationen unter Nutzung von 3D-Photo-Collections als Eingabedaten durch. Für beide Verfahren werden alle realen Oberflächen als Lambertsche Strahler modelliert, womit die Strahldichten als richtungsunabhängig angesehen werden und so für verschiedene Standorte im Raum als konstant angenommen werden können. Voraussetzung für beide Verfahren ist, dass die Beleuchtung der realen Szene in allen Bildern konstant ist

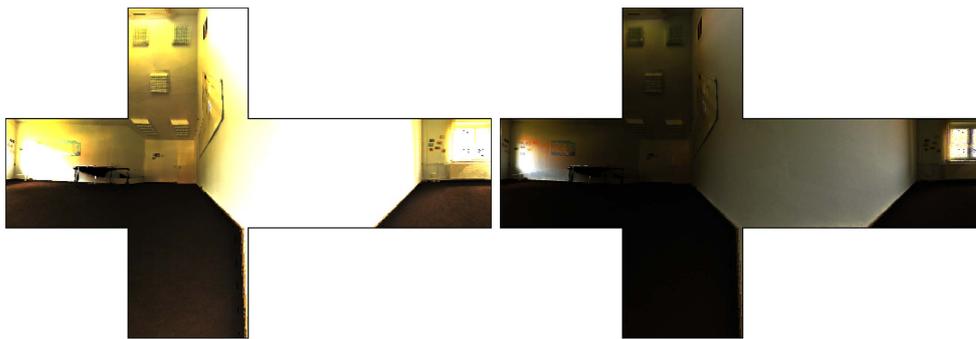


(a) Reale Szene mit eingefügtem virtuellem Objekt ohne Berücksichtigung der realen Beleuchtung



(b) Virtuelles Objekt mit benutzerdefiniertem Lichteinfall

**Abbildung 4.32:** Virtuelles Objekt, beleuchtet durch Light-Probe, welche aus den bekannten Strahldichten der realen Oberflächen erzeugt wurde

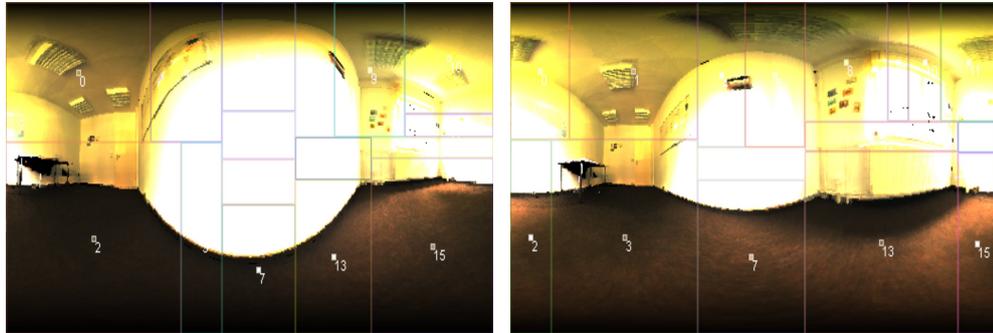


(a) Light-Probe 1



(b) Light-Probe 2

**Abbildung 4.33:** Aus dem Texturatlas generierte Light-Probes in Form einer HDR-Cube-Map für zwei unterschiedliche Positionen in der Szene. Zur Illustration des hohen dynamischen Umfangs ist jede HDR-Cube-Map einmal mit großer und einmal mit kleiner Skalierung der Strahldichten dargestellt.



(a) Direktionale Lichtquellen von Light-Probe 1 (b) Direktionale Lichtquellen von Light-Probe 2

**Abbildung 4.34:** Direktionale Lichtquellen, generiert aus den HDR-Cube-Maps von Abbildung 4.33. Jedes Segment repräsentiert eine direktionale Lichtquelle, deren Richtung durch das Segmentzentrum bestimmt wird. Durch die veränderte Position ergibt sich auch ein veränderter Einfall des Umgebungslichts.

und dass Bilder mit unterschiedlicher Belichtungszeit vorliegen. Ebenfalls benötigen beide Verfahren eine Rekonstruktion der Kamerakennkurve, wofür das Verfahren von Robertson (siehe Abschnitt 4.2.2) verwendet wurde. Beide Verfahren sind in der Lage für beliebige Punkte in der Szene die Synthese einer Light-Probe durchzuführen.

Das Verfahren mit Geometrie kann durch die bekannten Oberflächen die reale Beleuchtung wesentlich genauer konstruieren. Dies ist hauptsächlich darin begründet, dass durch Projektion auf die Geometrie alle Bilder der 3D-Photo-Collection zur Rekonstruktion der Strahldichten beitragen. Der Projektionsfehler ist nur in Bereichen der Szene vorhanden, welche nicht in der Szenengeometrie modelliert wurden, was aber meist nur Objektdetails betrifft. Abtastlücken bei der Rekonstruktion der Strahldichten können nur für Bereiche der Szene entstehen, für die keine Bildinformationen vorliegen. Am häufigsten tritt dieses Problem auf, wenn sehr helle Bereiche zwar fotografiert wurden, aber durch Überbelichtung entsprechenden Bildpixel unbrauchbar sind.

Im Gegensatz dazu muss im Verfahren ohne Geometrie die Nutzung der

**Tabelle 4.1:** Vergleich beider Verfahren zur Beleuchtungsrekonstruktion

Merkmal	Verfahren	
	ohne Geometrie	mit Geometrie
Eingabedaten	Bilder	Bilder & Geometrie
Rekonstruktion	Für Punkt $P_R$	Für gesamte Szene
Projektionsfehler	Mittel bis groß	Sehr gering
Abtastungslücken	Mittel bis groß	Minimal

**Abbildung 4.35:** Fehlerhafte Rekonstruktion von Strahldichten auf Grund falsch projizierter Bildpixel an Bereichen mit großen Tiefenunterschieden im Bildraum

Bilder auf den Radius  $r_R$  beschränkt werden, da sonst die sehr grobe Repräsentation der Geometrie durch einfache Ebenen zu einem großen Projektionsfehler führt. Ein zu klein gewählter Radius  $r_R$  wiederum führt zu Lücken in der generierten Light-Probe.

Im Verfahren mit Geometrie treten zwar vergleichsweise nur geringe Projektionsfehler auf. Wenn jedoch zwei Oberflächen räumlich weit auseinander liegen, in einem Bild aber aneinander grenzende Bildkoordinaten besitzen, können durch kleinere Projektionsfehler Bildpixel der einen Oberfläche zur Berechnung der Strahldichte der anderen Oberfläche beitragen. Dies führt insbesondere dann zu sichtbaren Fehlern, wenn Bildpixel einer signifikanten

Lichtquelle im Raum falschen Patches zugeordnet werden (siehe Abb. 4.35).

Jedoch sind die Projektionsfehler und die Abtastungslücken in der Light-Probe kleiner als beim Verfahren ohne Geometrie. Die wichtigsten Unterschiede beider Verfahren sind in Tabelle 4.1 aufgelistet.

Ein weiterer Vorteil des Verfahrens mit Szenengeometrie ist, dass die Beleuchtungsinformationen für die gesamte Szene rekonstruiert werden. Mit Hilfe von Verfahren wie etwa Inverse Radiosity [YDMH99] können bei vollständig definierter Szenengeometrie die Reflexionsgrade aller Oberflächen näherungsweise bestimmt werden. Verfahren wie zum Beispiel von Gibson [GCHH03] oder Grosch [Gro05] realisieren damit eine glaubwürdige Darstellung von Schattenwürfen durch virtuelle Objekte. Auch andere Synthesetechniken wie zum Beispiel das Einfügen von virtuellen Lichtquellen, das Entfernen von realen Objekten oder das Verändern von realen Materialien benötigen bekannte Reflexionsgrade, um die geänderten Lichtwege zu simulieren. Insgesamt liefert das Verfahren ohne Geometrie eine weniger genaue Rekonstruktion der Beleuchtungsdaten. Jedoch ist es für Anwendungsfälle geeignet, in denen keine Szenengeometrie vorhanden ist oder ein manuelles Modellieren der Geometrie nicht möglich ist.

## 4.5.2 Ausblick

### Projektionsfehler verringern

Zu große Projektionsfehler haben sich im Verfahren ohne Geometrie als größtes Hindernis für eine präzise Bestimmung der Umgebungsbeleuchtung erwiesen. Durch eine aufwändigere Rekonstruktion der Szenengeometrie könnten sie verringert werden. Dazu müssten zunächst komplexere Hilfsobjekte zur Repräsentation der Geometrie gewählt werden. Es existieren bereits Ansätze aus dem Bereich Image-based Rendering, wie z.B. [SSS09], in dem die Geometrie für einzelne Bilder als aus planaren Oberflächen zusammengesetztes 3D-Modell rekonstruiert wird und aus diesem für jedes Bild eine angenäherte Depth-Map erzeugt wird. Durch die genauere Tiefenbestimmung können letztlich Strahldichten aus der Umgebung besser im Raum lokalisiert werden. Dadurch können geringere Projektionsfehler oder durch entsprechende Wahl eines größeren Radius  $r_R$  geringere Abtastungslücken erreicht werden.

Die im Verfahren mit Geometrie erwähnten Fehler durch Projektion von

Strahldichten auf falsche Patches (siehe Abb. 4.35) können durch Einführung eines pixelbasierten Fehlermaßes verringert werden. Da die Chance, einen Bildpixel auf einen falschen Patch zu projizieren dann größer ist, wenn benachbarte Pixel zu entfernten Oberflächen gehören, kann hierfür das Fehlermaß je Pixel als Varianz der Tiefenwerte benachbarter Pixel modelliert werden. Dieses Fehlermaß kann mit Hilfe eines Pixelshaders effizient für alle Bilder der 3D-Photo-Collection aus deren Depth-Maps berechnet werden, um dann in die Gewichtungsfunktion  $w(Z)$  (siehe Abschnitt 4.4.3) mit einzufließen. Die in Abbildung 4.35 fälschlicherweise vom Fenster auf die Säule projizierten Bildpixel würden somit in ihrem Einfluss herabgestuft werden, da zwischen Säulenrand und Fenster große Tiefenunterschiede vorliegen.

### Rekonstruktion der Kamerakennkurve

Für beide Verfahren ist noch eine separate Rekonstruktion der Kamerakennkurve notwendig, um von Pixelwerten auf lineare Bestrahlungsstärken zu schließen. Dabei muss eine separate Belichtungsreihe erstellt werden, was manuellen Aufwand bedeutet. Dieser zusätzliche Schritt kann für das Verfahren mit Geometrie jedoch entfallen, wenn eine Rekonstruktion der Kamerakennkurve ausschließlich auf der Grundlage der Bilder der 3D-Photo-Collection durchgeführt wird. Möglich ist dies zum Beispiel, indem durch eine Projektion der Bilder auf die Geometrie eine Rekonstruktion der Kamerakurve je Patch erfolgt. Analog zur Rekonstruktion mit Belichtungsreihen (siehe Abschnitt 4.2.2) kann für ein Bild mit Belichtungszeit  $\Delta t_j$  eine Menge von sichtbaren Patches  $p_i$  aufgestellt werden. Jedem dieser Patches kann durch Projektion von Bilddaten ein entsprechender Pixelwert  $Z_{ij}$  zugeordnet werden. Da mehrere Pixelwerte mit unterschiedlichen Belichtungszeiten für gleiche Patches vorliegen, kann aus diesen Zusammenhang ein Gleichungssystem aufgestellt werden, dessen Lösung eine Rekonstruktion der Kamerakennkurve liefert.



# Kapitel 5

## Darstellung bildbasierter Szenen

### 5.1 Einleitung

Im vorherigen Kapitel wurden Verfahren zur Bestimmung der realen Beleuchtungsumgebung vorgestellt, mit denen eine glaubwürdige Synthese von realen und virtuellen Anteilen durchgeführt werden kann. Da als reale Anteile 3D-Photo-Collections verwendet werden, liegt nur eine durch die Anzahl der Bilder begrenzte Menge von Ansichten auf die reale Szene vor (siehe Abb. 5.1). Für eine augmentierte Bildsynthese muss daher stets die Kamerapose eines Bildes der 3D-Photo-Collection gewählt werden, um das entsprechende Bild im Hintergrund als realen Anteil darzustellen.

Durch die begrenzte Anzahl an möglichen Ansichten auf die Szene, kann somit auch keine kontinuierliche Bewegung entlang von Kamerapfaden und erst recht keine freie Navigation durch die Szene umgesetzt werden. Diese Beschränkung soll aufgehoben werden, indem aus den vorhandenen Bildern beliebige neue Ansichten auf die Szene generiert werden. Dieses Vorgehen wird auch als *View Synthesis* bezeichnet und ist eines der Hauptziele im Bereich Image-based Rendering, also der Darstellung von bildbasierten Szenen. In diesem Kapitel wird ein Verfahren vorgestellt, welches beliebige Ansichten auf die Szene in Echtzeit generiert unter Nutzung der Daten der 3D-Photo-Collection.

Falls eine vollständige Szenengeometrie vorliegt, wie sie im Verfahren aus Abschnitt 4.4 verwendet wird, dann können unter Nutzung von bestehenden Verfahren, wie z.B. [DYB98], beliebige Ansichten auf die Szene generiert werden (siehe Abschnitt 5.2).

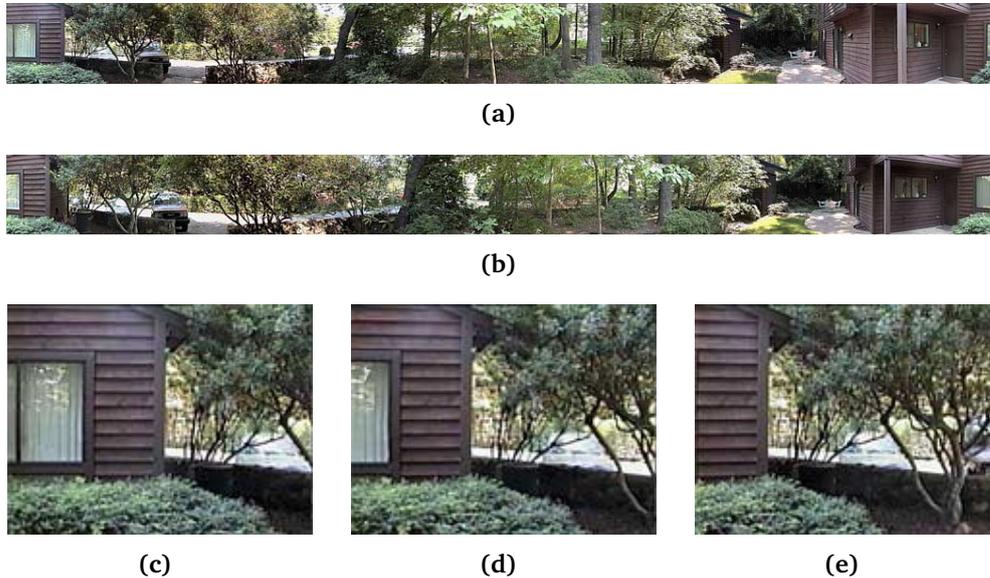
Ohne eine vollständige Szenengeometrie bleibt jedoch nur die aus den Bildmerkmalen gewonnene Punktwolke als grobe Repräsentation der realen



**Abbildung 5.1:** Eine 3D-Photo-Collection bietet nur eine begrenzte Anzahl von Ansichten auf die reale Szene

Oberflächen. Jeder Punkt der 3D-Punktwolke kann als Abtastwert der realen Oberflächen angesehen werden, jedoch sind die Positionen der Punkte verrauscht. Zudem ist die Punktdichte der Wolke bei vielen Szenen großen lokalen Schwankungen unterlegen.

Im vorgestellten Verfahren wird daher auf der Basis der lückenhaften Informationen der Punktwolke ein Vorhersagemodell für die Aufenthaltswahrscheinlichkeit von Szenenoberflächen um einzelne Punkte verwendet. Aus diesem Modell werden Kugeln als geometrische Hilfskörper konstruiert. Die Generierung neuer Ansichten auf die Szene erfolgt dann durch eine blickwinkelabhängige Projektion von Bildinhalten auf die konstruierten Kugeln unter Verwendung von Billboards als Renderprimitive.



**Abbildung 5.2:** Modellierung der plenoptischen Funktion nach McMillan und Bishop durch zylindrische Panoramen (a, b) und anschließender Generierung neuer Ansichten (c–e)

## 5.2 Verwandte Arbeiten

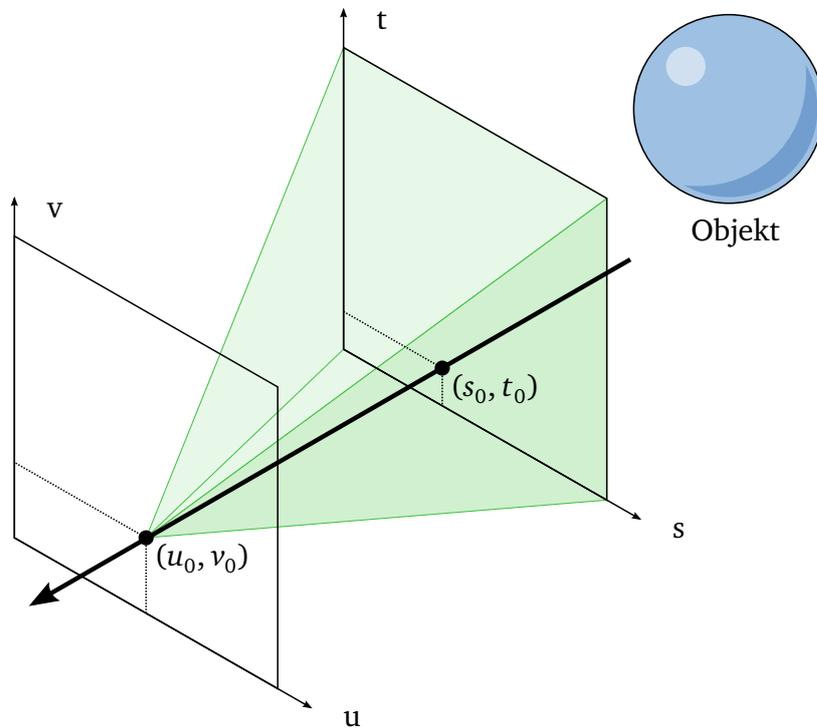
### 5.2.1 Plenoptisches Modellieren

Die Verfahren des Image-based Rendering sind verschiedene Ansätze, die plenoptische Funktion (siehe Abschnitt 2.3.1) einer Szene diskret abzutasten und anschließend zur Erzeugung neuer Ansichten eine kontinuierliche Rekonstruktion durchzuführen [SCK07].

Eines der ersten Verfahren zur Modellierung der plenoptischen Funktion wurde von McMillan und Bishop [MB95] vorgestellt (siehe Abb. 5.2). In diesem wird unter Beschränkung auf statische Szenen und durch Ignorieren der Wellenlänge die plenoptische Funktion auf fünf Parameter reduziert:

$$P(\theta, \phi, V_x, V_y, V_z) \quad (5.1)$$

Das Sampling der Szene erfolgt durch die Generierung von zylindrischen Panoramen an zwei verschiedenen Positionen. Mit Hilfe einer Stereobild-

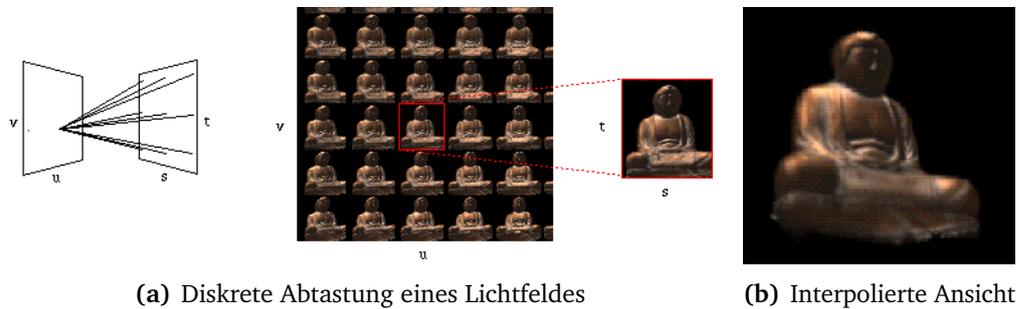


**Abbildung 5.3:** Parametrisierung von Strahlen im Lichtfeld in Form von Schnittpunkten in zwei Ebenen

analyse werden korrespondierende Pixel gefunden. Daraus werden Tiefeninformationen gewonnen, welche für die Generierung neuer Ansichten durch Interpolation beider Bilder verwendet werden.

## 5.2.2 Lumigraphs und Lichtfelder

Lumigraphs[GGSC96] und Lichtfelder[LH96] reduzieren die Komplexität der plenoptischen Funktion, indem beide eine vierdimensionale Parametrisierung des versendeten Lichts verwenden. Im Raum wird eine Kameraebene  $(u, v)$  sowie eine Fokusebene  $(s, t)$  konstruiert und sämtliche von der Fokus- nach der Kameraebene versandte Lichtstrahlen werden durch deren Schnittpunkte mit beiden Ebenen parametrisiert (siehe Abb. 5.3):



**Abbildung 5.4:** Rendering von Lichtfeldern nach Levoy [LH96]

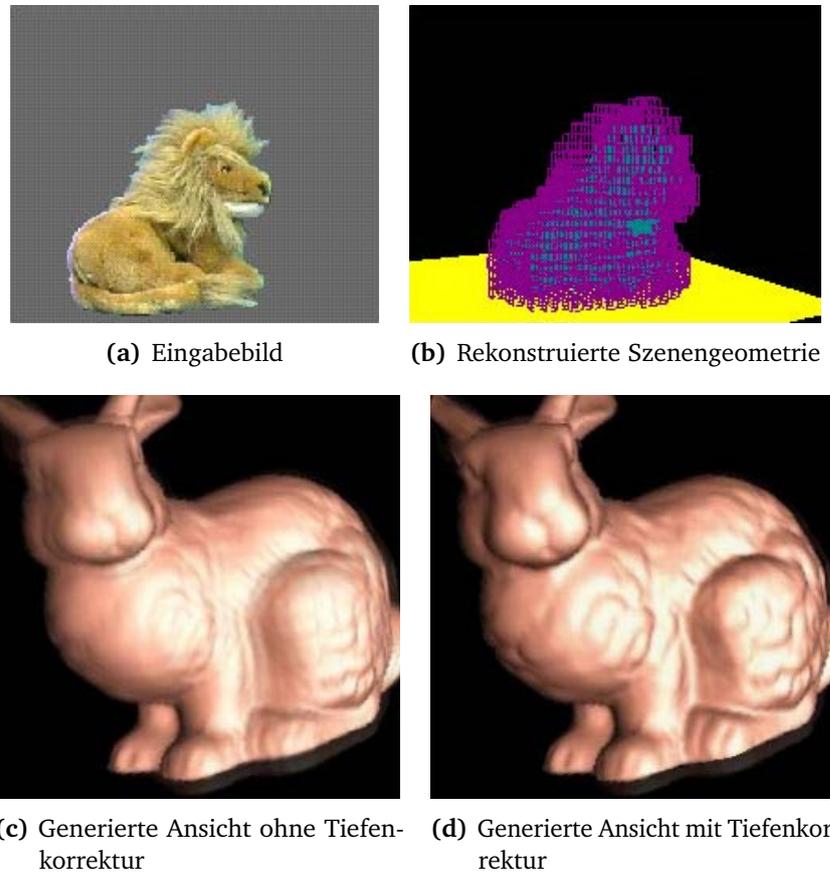
$$P(u, v, s, t) \quad (5.2)$$

In dieser Parametrisierung bilden alle Lichtstrahlen, welche die Fokusebene in einem beliebigen Punkt durchstoßen und auf der Kameraebene in einem Punkt  $(s_0, t_0)$  zusammenfallen, ein Kameravolumen einer perspektivischen Projektion.

Für beide Verfahren wird angenommen, dass sich keine Geometrie zwischen Kamera- und Fokusebene befindet. Das Verfahren von Levoy macht darüber hinaus keine weiteren Annahmen über die Geometrie und generiert neue Ansichten durch quadrilineare Interpolation (siehe Abb. 5.4).

Lumigraphs werden erzeugt, in dem ein reales Objekt vor einem Bluescreen fotografiert wird. Die Kameraposen werden durch markerbasiertes Tracking rekonstruiert. Aus den Objektsilhouetten und bekannten Kameraposen wird über einen Shape-From-Silhouette-Ansatz die Geometrie des Objektes näherungsweise rekonstruiert. Bei der Generierung neuer Ansichten werden diese Informationen für die Tiefenkorrektur bei dem Sampling des Lichtfeldes verwendet, wodurch die Unschärfe verringert wird (siehe Abb. 5.5).

Beide Verfahren liefern eine Generierung von neuen Ansichten innerhalb eines eingeschränkten Sichtbereichs, benötigen dazu aber eine sehr dichte Abtastung der plenoptischen Funktion. Dies gilt insbesondere für Lichtfelder, da hier keinerlei Szenengeometrie vorliegt.



**Abbildung 5.5:** Rendering von Lichtfeldern nach Goertler [GGSC96]

### 5.2.3 View-dependent Texture Mapping

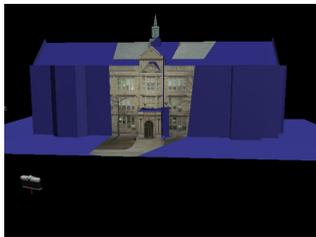
Von Debevec et. al. wurde View-dependent Texture Mapping [DTM96] vorgestellt. View-dependent Texture Mapping (VDTM) verwendet eine geringe Anzahl von Eingabebildern mit bekannten Kameraposen, benötigt jedoch eine explizit definierte Geometrie. Dazu wird ein 3D-Modell der Szene mit Hilfe des Nutzers unter Ausnutzung von epipolargeometrischen Zusammenhängen zwischen den Bildern konstruiert. Für die Generierung neuer Ansichten werden die Bilder unter Nutzung von projektivem Texture-Mapping auf die Geometrie abgebildet. Wenn mehrere Bilder auf den gleichen Punkt der Geometrie abgebildet werden, erfolgt eine Gewichtung der Bilder nach dem



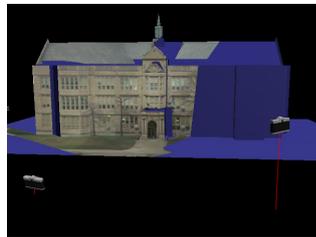
(a) Bild und Geometrie



(b) Generierte Ansicht



(c) Ein projiziertes Bild



(d) Zwei projizierte Bilder

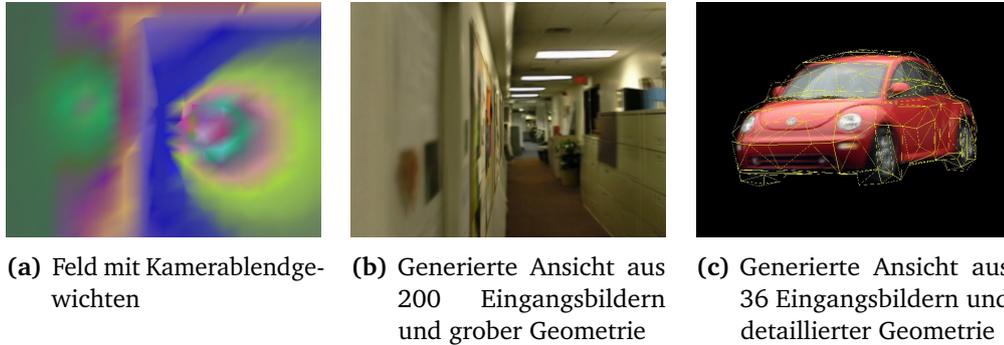


(e) Alle projizierten Bilder

**Abbildung 5.6:** Bei View-dependent Texture Mapping nach Debevec [DTM96] werden die Bilder auf die bekannte Geometrie projiziert, wobei der Einfluss eines Bildes in Abhängigkeit des Winkels von Betrachter- und Kamerasehstrahl gewichtet wird.

Winkel zwischen Betrachter- und Kamerasehstrahl. Dadurch erhalten die Bilder das meiste Gewicht, deren Kameraperspektive der des Betrachters am ähnlichsten ist (siehe Abb. 5.6). Ein Vorteil dieser Gewichtung ist, dass auch glänzende Szenenoberflächen und im 3D-Modell der Szene nicht berücksichtigte feine geometrische Details wiedergegeben werden können, insofern zur Perspektive des Betrachters passende Bilder existieren.

Später wurde View-dependent Texture Mapping für echtzeitfähiges Rendering [DYB98] angepasst, indem die Gewichte von projizierten Bildern für jedes Polygon vorberechnet wurden. Zudem wurde hardwarebeschleunigtes projektives Texture-Mapping verwendet, um eine höhere Darstellungsge-



**Abbildung 5.7:** Rendering von Unstructured Lumigraphs nach Buehler [BBM<sup>+</sup>01]. Grob definierte Szenengeometrie kann durch eine höhere Anzahl von Bildern ausgeglichen werden (b). Umgekehrt werden bei detaillierter Geometrie (c) ähnlich zu VDTM nur wenige Bilder benötigt.

schwindigkeit zu erreichen.

Mit VDTM wurde gezeigt, dass bei vollständig vorhandener Geometrie eine überzeugende Generierung von neuen Ansichten auch mit nur sehr wenigen Abtastwerten der plenoptischen Funktion erreicht werden kann. Zudem können bei entsprechend höherer Anzahl von Bildern auch Oberflächen mit spekularem Reflexionsverhalten oder fein strukturierten geometrischen Details dargestellt werden.

### 5.2.4 Unstructured Lumigraphs

Unstructured Lumigraphs [BBM<sup>+</sup>01] vereinheitlichen die beiden unterschiedlichen Ansätze von VDTM und Lumigraphs. Wie bei VDTM wird eine Menge von Bildern mit bekannten Kameraposen verwendet, jedoch ist die Verwendung einer Szenengeometrie lediglich optional. Für die Erzeugung neuer Ansichten in Echtzeit wird zunächst für den Bildraum des Betrachters ein Gewichtungsfeld erzeugt, welches in Form eines grob aufgelösten Gitters für jeden Gitterpunkt die Blendgewichte der beitragenden Kameras enthält. Die Blendgewichte eines Gitterpunktes werden dann für die Berechnung der umliegenden Pixel verwendet (siehe Abb. 5.7). Die Auflösung des Gitters richtet sich – insofern vorhanden – an Kanten der Geometrie aus, so dass

Artefakte an Objektkanten vermieden werden. Die Berechnung der Blendgewichte berücksichtigt verschiedene Ziele wie Kontinuität, Auflösung und minimaler Winkel (vergleichbar zu VDTM).

Um eine überzeugende Darstellung mit möglichst geringen Artefakten zu erzielen, muss zumindest eine einfache Geometrie vorhanden sein. Bei detaillierten Geometrieinformationen werden ähnlich zu VDTM nur wenige Bilder benötigt. Umgekehrt nähert sich das Verfahren bei nur grob definierter Geometrie der Darstellung von Lumigraphs an und benötigt mehr Bilder, um eine akzeptable Bildqualität zu erreichen.

### 5.2.5 Bildbasierte Darstellung mit Multi-View Stereo

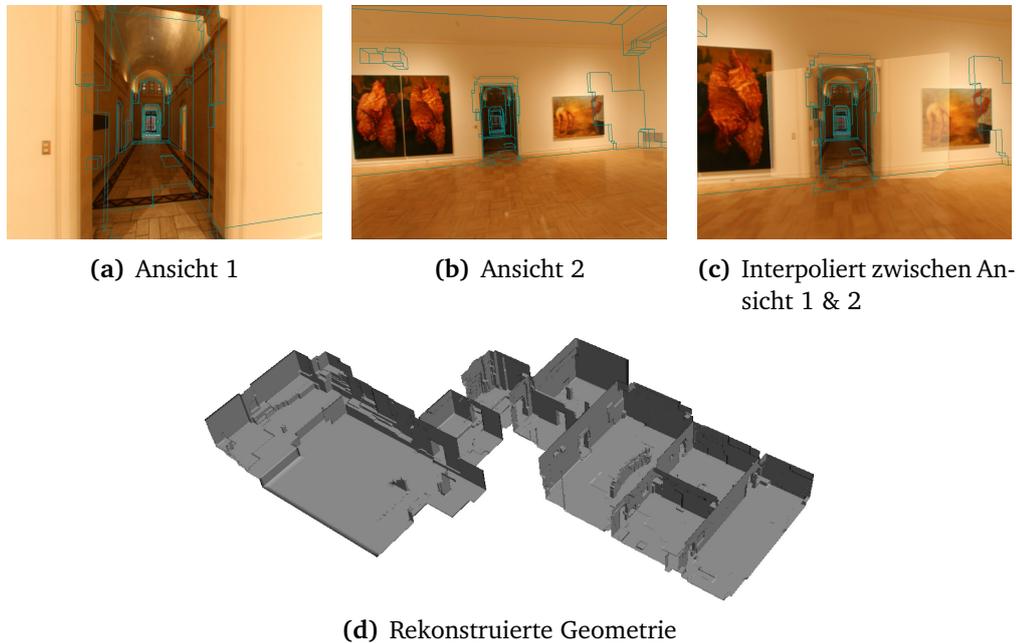
Die Darstellung von bildbasierten Szenen kann auch auf der Basis einer zuvor durchgeführten 3D-Rekonstruktion der Szenengeometrie mittels eines Multi-View-Stereo-Verfahrens erfolgen.

#### **Furukawa et. al.**

Furukawa et. al. [FCSS09b] führen vor der eigentlichen Darstellung der Szene eine mehrstufige 3D-Rekonstruktion durch (siehe Abb. 5.8). Aus den Eingabebildern werden mit Hilfe von Bundler – identisch zum hier vorgestellten Verfahren – Kameraposen registriert sowie die Punktwolke der Szene generiert. Anschließend wird mit einem patch-basiertem Multi-View-Stereo-Verfahren [FP07] [FP10] eine detailliertere Punktwolke mit Normalen für jeden Punkt generiert. Die Szenengeometrie wird als eine „Manhattan“-Welt, d.h. als Menge von an den Koordinatenachsen ausgerichteten Boxen, angenommen und daraus mit einem weiteren Multi-View-Stereo-Verfahren [FCSS09a] für jedes Bild eine Depth-Map erzeugt. Aus der Menge der Depth-Maps wird schließlich wieder unter der Annahme einer „Manhattan“-Welt ein vollständiges 3D-Modell der Szene erzeugt.

Die eigentliche Darstellung der Szene erfolgt über einen Ansatz ähnlich zu VDTM. Neue Ansichten auf die Szene werden generiert, indem zwei Bilder mit zum Betrachter ähnlichen Perspektiven auf die Geometrie projiziert werden. Die Bilder werden dabei nach Distanz des Betrachters zu deren Aufnahmeposition unterschiedlich gewichtet.

Die automatische 3D-Rekonstruktion der Geometrie ist für Architekturszenen robust genug, um neue Ansichten auf die Szene durch Projektion der

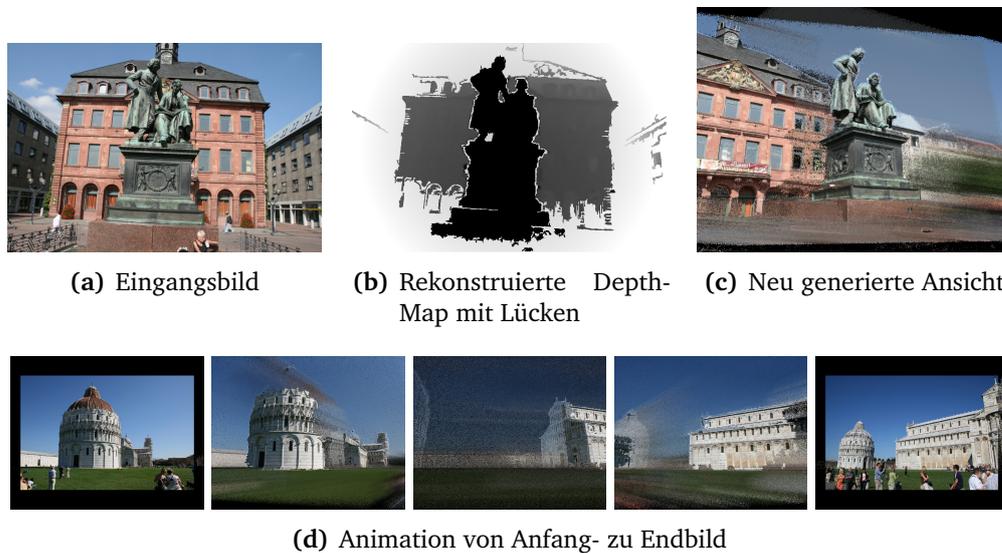


**Abbildung 5.8:** Darstellung bildbasierter Szenen nach Furukawa et. al. [FCSS09b]. Nach automatischer 3D-Rekonstruktion der Geometrie werden, ähnlich zu VDTM, darauf Bilder projiziert, um neue Ansichten zu generieren.

Bilder zu generieren. Jedoch wird noch nicht das gleiche Qualitätsmaß wie bei VDTM mit manuell modellierter Geometrie erreicht. Die Geometrie enthält noch kleinere Löcher oder falsch modellierte Flächen, welche zu fehlerhaft projizierten Bildinhalten führen. Zudem werden für gute Ergebnisse noch wesentlich mehr Eingabebilder benötigt als bei VDTM und die Rekonstruktion der Geometrie kann bei großen Bildmengen mehrere Stunden dauern und ist im Wesentlichen nur für Architekturszenen geeignet.

#### **Goesele et. al.**

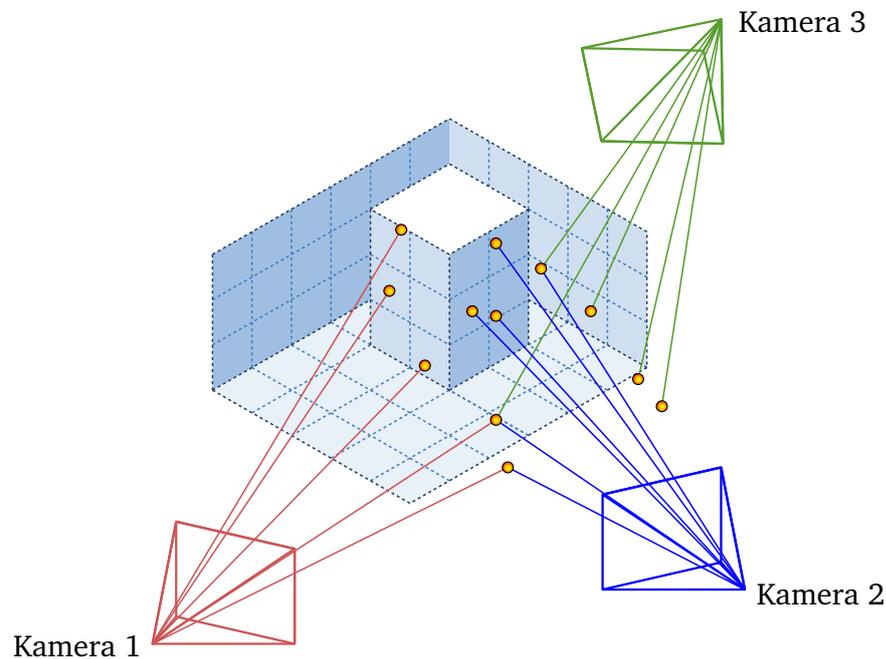
Der Ansatz zur bildbasierten Darstellung von Goesele et. al. [GAF<sup>+</sup>10] verwendet ebenfalls ein vorangehendes Multi-View-Stereo-Verfahren [GSC<sup>+</sup>07], welches jedoch im Gegensatz zu Manhattan-World-Stereo von Furukawa



**Abbildung 5.9:** Sichtinterpolation bildbasierter Szenen nach Goesele [GAF<sup>+</sup>10] unter Nutzung von ambienten Punktwolken zur Überblendung von Bereichen ohne Tiefeninformation

keine Annahmen über die Architektur der Szene macht. Die Rekonstruktion der Geometrie ist zwar präzise, aber unvollständig, so dass das generierte 3D-Modell nur einen Teil der realen Oberflächen beschreibt. Ziel von Goeseles Verfahren ist es, trotz der unvollständigen Geometrieinformationen, visuell überzeugende Animationen zwischen den Kameraposen von zwei Bildern der Szene umzusetzen, d.h. das Verfahren ist für die Interpolation zwischen zwei Ansichten und nicht für eine freie Navigation konzipiert (siehe Abb. 5.9).

Zunächst werden aus jedem Bild Depth-Maps generiert, welche auf Grund des unvollständigen 3D-Modells ebenfalls lückenhaft sind. Für die Generierung einer neuen Ansicht wird zunächst aus der Betrachterperspektive eine globale Punktwolke gerendert, welche aus der rekonstruierten 3D-Geometrie generiert wurde und im Gegensatz zu der von Bundler generierten Punktwolke detaillierter ist. Anschließend werden jeweils für Anfangs- und Endbild „ambiente“ Punktwolken gerendert. Goesele konstruiert diese Punktwolke als eine punktbasierte Repräsentation der Epipolarlinien von Bildpixeln mit unbekanntem Tiefeninformationen. Über die ambienten Punktwolken werden



**Abbildung 5.10:** Schematische Darstellung einer 3D-Photo-Collection als bildbasiertes Modell der realen Szene

zwei 3D-Meshes gerendert, welche jeweils für Anfang- und Endbild aus den Depth-Maps gewonnenen wurden und auf deren Oberfläche diese Bilder projiziert werden. Die Überlappung beider Meshes wird durch Verwendung von Soft Z-Buffering [PCD<sup>+</sup>97] und Alpha-Blending als kontinuierlicher Übergang dargestellt.

## 5.3 Eigener Ansatz

### 5.3.1 Daten der 3D-Photo-Collection

Für die bildbasierte Darstellung der Szene liegen folgende Daten der 3D-Photo-Collection vor (siehe auch Abb. 5.10):

- 3D-Punkte der Punktwolke, welche Abtastwerte der Szenenoberflächen repräsentieren – jeder Punkt wurde aus den Merkmalspunkten rekonstruiert

- Für jedes Bild definierte Kamerapose in Form extrinsischer (Rotation und Position) sowie intrinsischer Kameraparameter (Brennweite, Seitenverhältnis, radiale Verzerrung)
- Liste aller detektierten Merkmalspunkte für jedes Bild – jeder Merkmalspunkt referenziert den korrespondierenden Punkt der 3D-Punktwolke.

Die Dichte der Punktwolke und die Anzahl der detektierten Merkmalspunkte je Bild können stark variieren, je nachdem wie die fotografierten Bildinhalte beschaffen sind. Auf strukturierten Oberflächen mit detailreicher Textur werden mehr Merkmalspunkte detektiert. Ebenso kann die Anzahl der Merkmalspunkte bei realen Oberflächen höher liegen, wenn diese öfter oder aus geringen Entfernungen fotografiert wurden. Beides führt zu lokal höheren Punktdichten.

### 5.3.2 Ziele

Für die Erzeugung neuer Ansichten aus den beschriebenen Daten werden folgende Ziele definiert:

**Geometrische Informationen der Punktwolke nutzen** Die Punktwolke soll als direkte Quelle für geometrische Informationen genutzt werden. Erst mit vorhandenen Informationen über die Szenengeometrie können Bildinhalte im Raum lokalisiert werden und unter Annahme von diffusen Oberflächen, ähnlich zum Ansatz des View-dependent Texture Mapping (siehe Abschnitt 5.2.3), für verschiedene Betrachterpositionen und Blickwinkel dargestellt werden.

**Erzeugung neuer Ansichten in Echtzeit** Es wird eine echtzeitfähige Darstellungsgeschwindigkeit angestrebt, um interaktive Navigation zu ermöglichen. Auch für sehr große Szenen mit mehr als 300 Bildern soll eine flüssige Bildwiederholrate erreicht werden.

**Berücksichtigung von Bildauflösung und Blickwinkel** In den Bildern der 3D-Photo-Collection werden die Oberflächen der realen Szene mit großer Redundanz wiedergegeben, da viele Bilder ähnliche Bereiche der Szene zeigen.

Für die Erzeugung neuer Ansichten bedeutet dies, dass für die Wiedergabe von bestimmten Oberflächen aus der Menge der Bilder eine geeignete Auswahl gemacht werden muss. Zum Beispiel eignet sich für die Darstellung einer bestimmten realen Oberfläche eher ein Bild, welches aus einem zur Betrachterperspektive ähnlichem Blickwinkel fotografiert wurde. Zudem werden Oberflächen detailreicher wiedergegeben, wenn Bilder verwendet werden, deren Aufnahmeposition möglichst nah an der wiedergegebenen Oberfläche liegt.

Im Gegensatz zu Ansätzen wie Unstructured Lumigraphs (siehe Abschnitt 5.2.4) oder View-dependent Texture Mapping (siehe 5.2.3) liegen keine vollständigen Informationen über reale Oberflächen – z.B. in Form von Polygonen – vor, sondern die Punktwolke liefert nur einige Abtastwerte der realen Oberflächen. Daher wird ein Modell benötigt, das näherungsweise Szenenoberflächen beschreibt.

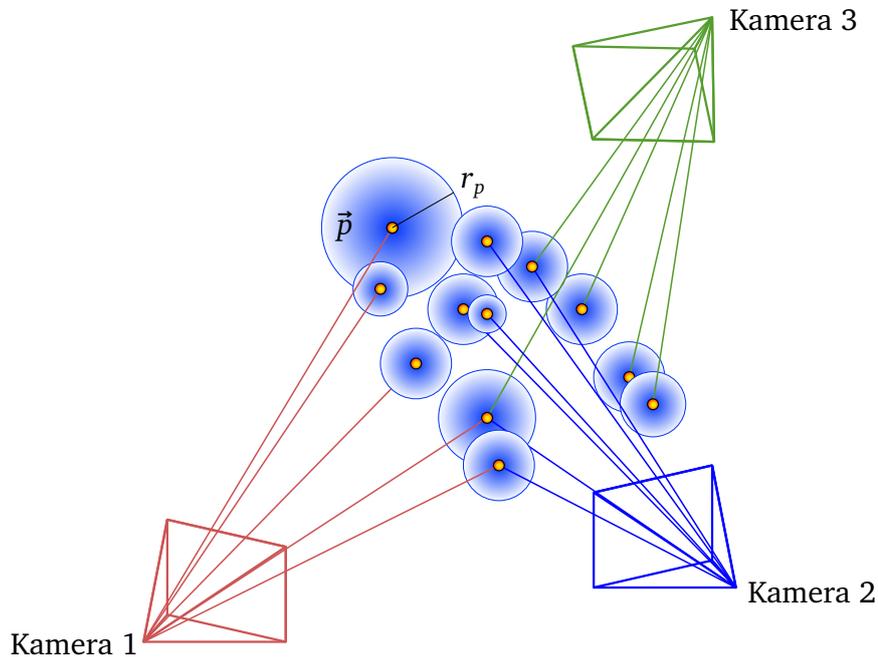
### 5.3.3 Modellierung von Szenenoberflächen

Für die Modellierung der Szenengeometrie wird ein wahrscheinlichkeitsbasierter Ansatz gewählt, bei dem eine Vorhersage von realen Oberflächen auf der Grundlage der Entfernung zu Abtastwerten der Szenengeometrie – also Punkte der Punktwolke – gemacht wird. Dies wird einzeln für jeden Punkt  $p$  der Punktwolke als eine Funktion der euklidischen Distanz zu einer Position  $\vec{x}$  modelliert:

$$f_p(\vec{x}) = \max\left(1 - \frac{|\vec{x} - \vec{p}|}{r_p}, 0\right) \quad (5.3)$$

Dabei wird die Wahrscheinlichkeit einer Szenenoberfläche an Position  $\vec{x}$  beschrieben, welche mit größer werdender Distanz zu Punkt  $\vec{p}$  abnimmt. Am Punkt  $\vec{p}$  selbst ( $\vec{x} = \vec{p}$ ) ist die Wahrscheinlichkeit 1. Der Parameter  $r_p$  beschränkt die Reichweite der Vorhersage auf einen bestimmten Radius um  $\vec{p}$ , so dass für alle  $\vec{x}$  außerhalb von Radius  $r_p$   $f_p(\vec{x}) = 0$  ist. Der Radius  $r_p$  wird für jeden Punkt einzeln berechnet und richtet sich nach der lokalen Punktdichte (siehe Abb. 5.11).

Die Wahl des Radius für einen einzelnen Punkt ist ein Kompromiss zwischen hoher Reichweite und genauer Vorhersage. Zu große Radien führen zu falschen Annahmen über die Geometrie und zu geringe Radien führen zu

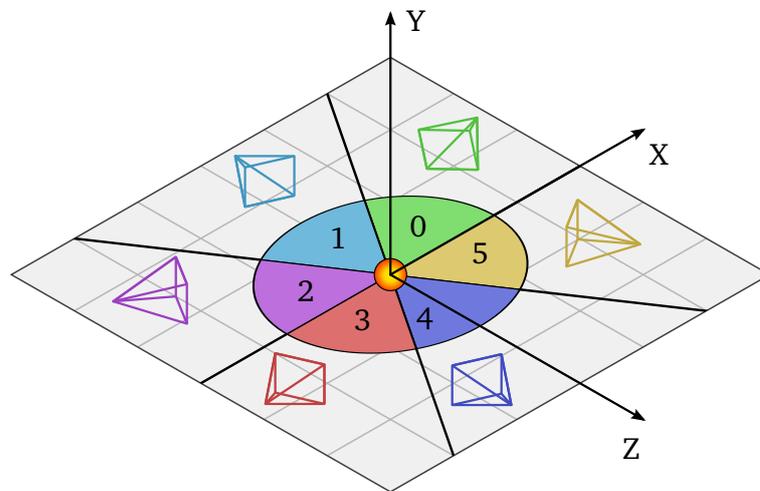


**Abbildung 5.11:** Die Funktion  $f_p(\vec{x})$  (blau) beschreibt die Wahrscheinlichkeit von Szenengeometrie um einzelne Punkte der Punktwolke und nimmt mit größer werdender Entfernung zu einem Punkt  $\vec{p}$  ab. Für jeden Punkt wird die Wahrscheinlichkeit durch den punkteigenen Radius  $r_p$  (blauer Kreis) lokal beschränkt.

großen Löchern. Im vorgestellten Ansatz erfolgt diese Berechnung, indem die Punktwolke in ein reguläres Gitter aus würfelförmigen Zellen mit der Seitenlänge  $s_g$  eingeteilt wird. Jede Gitterzelle besitzt das Volumen  $s_g^3$  was einen Ausschnitt der Punktwolke mit  $n_g$  Punkten enthält. Jeder Punkt der Gitterzelle wird als Kugel behandelt, dessen Radius  $r_p$  berechnet wird, indem das Volumen der Gitterzelle auf alle darin enthaltenen Punkte gleich aufgeteilt wird:

$$r_p = k_r \sqrt[3]{\frac{3 s_g^3}{4 n_g \pi}} \quad (5.4)$$

Der Koeffizient  $k_r$  ist ein globaler Skalierungsfaktor aller Radien  $r_p$  und



**Abbildung 5.12:** Einteilung der Umgebung in  $n_c$  Abschnitte (im Bild  $n_c = 6$ ) entsprechend des Längewinkels des Punktes, jede Kamera wird einem Abschnitt zugeordnet.

justiert in der Darstellung den Kompromiss zwischen Abdeckung und Vorhersagefehler.

### 5.3.4 Zuordnung von Bildern an Punkte

Ähnlich zu View-dependent Texture Mapping sollen auf Oberflächen mehrere Bilder projiziert werden, die je nach Orientierung und Position des Betrachters unterschiedlich stark gewichtet werden. Zunächst sollen für jeden Punkt der Punktwolke Bilder gefunden werden, die dessen zugehörige Szenenoberfläche als Bildinformation enthalten. Die für jeden Punkt aus den einzelnen Bildern detektierten Merkmalspunkte können hierfür als Visibilitätsinformation verwendet werden: Falls ein Punkt  $p$  in einem Bild sichtbar ist, dann wird dieses Bild den durch  $f_p$  vorhergesagten Szenenoberflächen zugeordnet.

Die Anzahl der beitragenden Bilder je Punkt kann sehr hoch sein, so dass ihre Anzahl auf  $n_c$  reduziert werden soll. Generell geben näher gelegene Kameras die Oberfläche mit höherer Auflösung wieder unter der Annahme, dass alle Bilder die gleiche Auflösung besitzen. Außerdem sollten Bilder gewählt werden, welche den Punkt aus unterschiedlichen Richtungen abbilden, da das Aussehen der um den Punkt liegenden Oberfläche unter verschiedenen

Betrachterwinkeln variieren kann – in der Regel durch geometrische Details, die nicht durch die Punktwolke wiedergegeben werden.

Um Kameraposen aus möglichst unterschiedlichen Richtungen zu erhalten, wird die Umgebung des betreffenden Punkts in  $n_c$  gleich große Abschnitte mit gleich großem Raumwinkel  $4\pi/n_c$  unterteilt (siehe Abb. 5.12). Jedes beitragende Bild wird einem der Abschnitte, entsprechend der Lage zum untersuchten Punkt, zugeordnet. Anschließend wird für jeden Abschnitt die Kamera mit der geringsten Distanz zum Punkt gewählt. Daraus resultieren  $n_c$  Bilder, welche den Punkt aus unterschiedlichen Blickwinkeln zeigen.

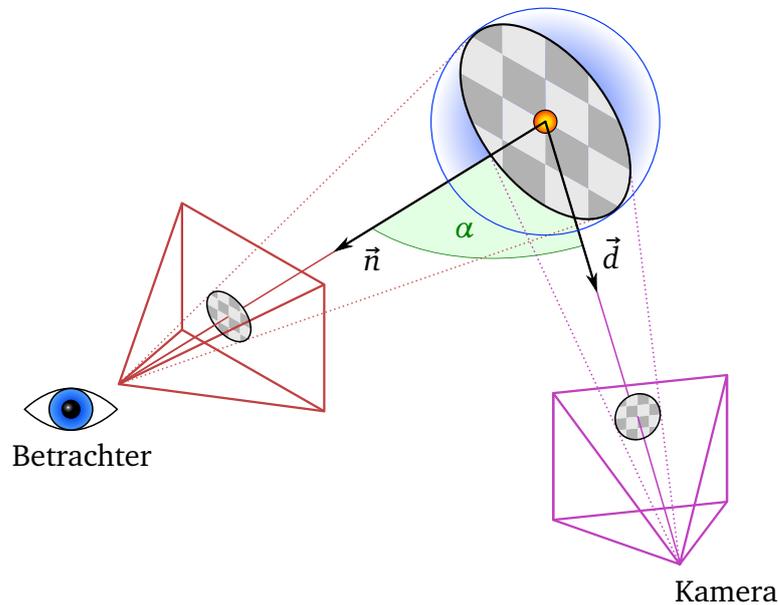
Da die Bilder der meisten Szenen aus ähnlichen Höhen fotografiert wurden, sind die Kameras in der Szene näherungsweise auf der Ebene verteilt. Mit der Verwendung eines Weltkoordinatensystems, bei der die Y-Achse nach oben zeigt, ist dies die XZ-Ebene. Die oben beschriebene Aufteilung in Abschnitte um den Punkt erfolgt nun auf Grundlage des Längewinkels des Vektors  $\vec{d}$  zwischen Punktposition  $\vec{p}$  und Kameraposition  $\vec{c}$  auf der XZ-Ebene. Der zugehörige Abschnitt  $i$  ( $0 \leq i < n_c$ ) berechnet sich daher wie folgt:

$$\vec{d} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix} = \vec{p} - \vec{c} \quad (5.5)$$

$$i = \frac{1}{n_c} \left\lfloor \frac{\text{atan2}(d_y, d_x)}{2\pi} n_c \right\rfloor \quad (5.6)$$

### 5.3.5 Blickwinkelabhängige Projektion der Bildinhalte

Nach den vorhergehenden Schritten wurde für jeden Punkt auf der Grundlage der lokalen Punktdichte ein Radius zugeordnet, innerhalb dessen mit einer gewissen Wahrscheinlichkeit die Anwesenheit von Szenengeometrie vorhergesagt wird. Zudem sind jedem Punkt maximal  $n_c$  beitragende Bilder zugeordnet. Im folgenden Schritt wird aus diesen Informationen für jeden Punkt die Geometrie konstruiert, auf welche die beitragenden Bilder projiziert werden. Auf Grund der großen Anzahl von Punkten werden hierfür Billboards verwendet, die als scheibenförmige Flächen mit dem Radius  $r_p$  um die Punkte konstruiert werden und deren Normale parallel zum Sichtstrahl des Betrachters ist.



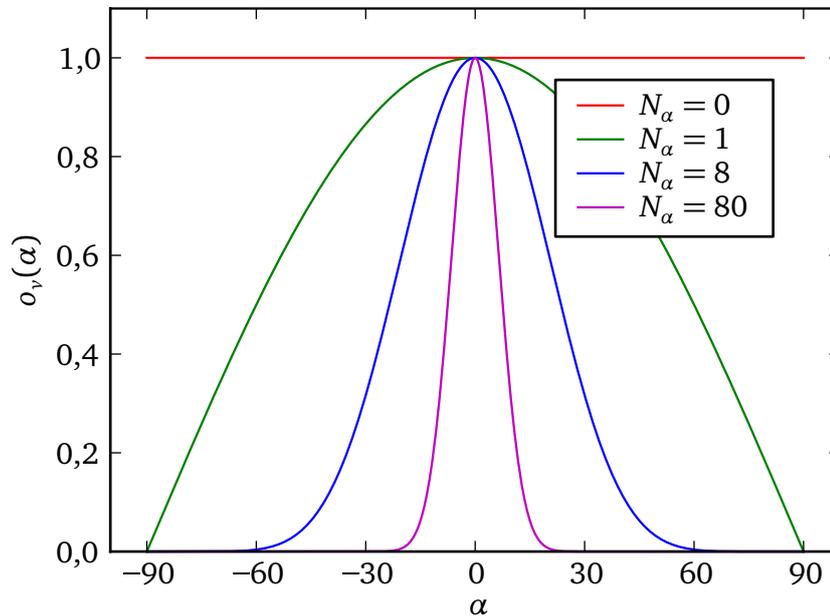
**Abbildung 5.13:** Projektion eines Bildausschnitts auf das um den Punkt  $\vec{p}$  konstruierte Billboard in Richtung  $\vec{d}$  und Reprojektion in Richtung des Betrachters  $\vec{n}$ . Der Reprojektionswinkel  $\alpha$  beschreibt die Größe des Projektionsfehler durch die abweichenden Blickwinkel.

Auf ein Billboard werden sämtliche Bilder projiziert. Dabei wird die durch  $f_p(\vec{x})$  um Punkt  $\vec{p}$  vorhergesagte Wahrscheinlichkeit als Deckungsgrad projizierter Bildpixel durch den folgenden Term  $o_p(\vec{x})$  modelliert:

$$o_p(\vec{x}) = f_p(\vec{x})^{N_f} \quad (5.7)$$

Der Exponent  $N_f$  kontrolliert die Abbildung von Wahrscheinlichkeiten auf Deckungsgrade und erlaubt eine Justierung der Darstellung ähnlich zu  $k_r$  (siehe Gleichung 5.4). Falls  $N_f$  auf 1 gesetzt wird, herrscht ein linearer Zusammenhang. Für größere Werte erhalten die Positionen im Raum mit geringer Wahrscheinlichkeit einen wesentlich geringeren Deckungsgrad.

Zusätzlich wird der blickwinkelabhängige Term  $o_v$  eingeführt, um den Deckungsgrad des Billboards zu steuern. Dies geschieht in Abhängigkeit des Winkels zwischen Normale des Billboards – welche der Blickrichtung des



**Abbildung 5.14:** Der blickwinkelabhängige Term  $o_v(\alpha)$  beschreibt die Transparenz des Billboards als Funktion des Reprojektionswinkels  $\alpha$ . Größere Werte von  $N_\alpha$  schränken die Sichtbarkeit des Billboards auf kleinere Bereiche von  $\alpha$  ein.

Betrachters entspricht – und dem Vektor  $\vec{d}$  zwischen der Kameraposition  $\vec{c}$  und der Punktposition  $\vec{p}$ . Da dieser Winkel gleichzeitig auch den Grad des Abbildungsfehlers bei der Reprojektion von Bildern in Richtung des Betrachters beschreibt, wird er als Reprojektionswinkel  $\alpha$  bezeichnet (siehe Abb. 5.13). Der blickwinkelabhängige Term  $o_v$  ist wie folgt definiert:

$$\cos \alpha = \frac{\vec{n} \cdot \vec{d}}{|\vec{n}| |\vec{d}|} \quad (5.8)$$

$$o_v(\alpha) = \max(0, \cos \alpha)^{N_\alpha} \quad (5.9)$$

In Abhängigkeit der Blickrichtung des Betrachters werden die Bildinhalte der beitragenden Bilder mit dem größten Deckungsgrad dargestellt, welche

den kleinsten Reprojektionswinkel  $\alpha$  aufweisen. Um den Einfluss von  $\alpha$  auf den Deckungsgrad besser zu steuern, wird der Exponent  $N_\alpha$  auf  $\cos \alpha$  angewendet. Größere Werte von  $N_\alpha$  schränken die Sichtbarkeit des Billboards auf kleinere Bereiche von  $\alpha$  ein. Falls  $N_\alpha$  auf 0 gesetzt wird, geht jegliche Blickwinkelabhängigkeit verloren (siehe Abb. 5.14). Der resultierende Deckungsgrad des Billboards wird als das Produkt von beiden Termen berechnet:

$$o(\alpha) = o_v(\alpha) \cdot o_p(\vec{x}) \quad (5.10)$$

Nach diesem Vorgehen werden für jeden Punkt Billboards konstruiert und für alle beitragenden Bilder die entsprechenden Bildausschnitte auf das Billboard projiziert. Auf Grund der großen Anzahl von Billboards wird das Rendering auf die GPU-Hardware ausgelagert.

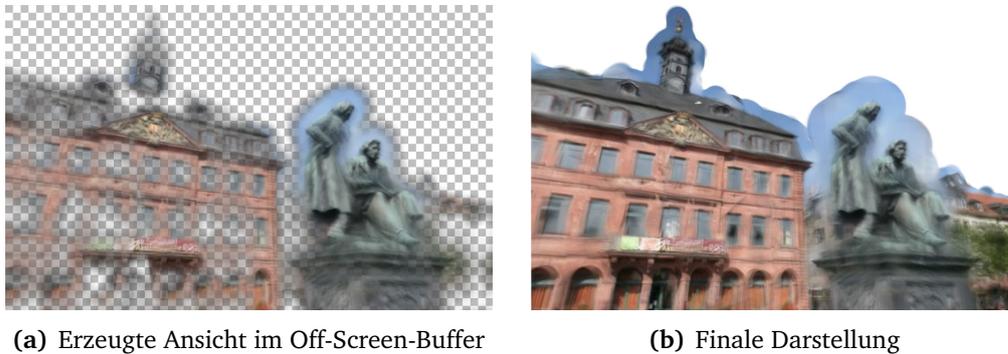
### 5.3.6 Umsetzung der Darstellung

Da der Deckungsgrad der Billboards mit Hilfe von Alpha-Blending umgesetzt wird, müssen die Billboards aus Sicht des Betrachters in der Reihenfolge von hinten nach vorn gerendert werden. Die Anzahl der Billboards kann sehr groß werden und die Sortierung muss für jede veränderte Betrachterposition erneut durchgeführt werden. Daher werden die Billboards in Cluster gruppiert und die Tiefensortierung erfolgt für Cluster anhand derer Mittelpunkte. Für die Tiefensortierung wird Mergesort [Knu98] als stabiles Sortierungsverfahren verwendet, um Popping-Artefakte bei Billboards, welche sich überschneiden und die gleiche Tiefe besitzen, zu vermeiden.

### 5.3.7 Rendering

Die Billboards für jeden Cluster werden durch zwei Dreiecke modelliert und in einem Vertex- sowie Indexbuffer im Videospeicher der Grafikkhardware abgelegt. Für jeden darzustellenden Frame erfolgt zunächst die Tiefensortierung der Cluster für die aktuelle Betrachterposition. Anschließend wird das Rendering mit Hilfe von zwei Durchgängen durchgeführt:

Der erste Durchgang rendert alle Billboards für die aktuelle Ansicht auf die Szene. Dabei werden die Billboards in der durch die Tiefensortierung bestimmten Reihenfolge, wie in Abschnitt 5.3.5 beschrieben, in einen Off-Screen-Buffer gerendert. Der Off-Screen-Buffer besitzt die Auflösung des



**Abbildung 5.15:** Pixel im Off-Screen-Buffer mit großer Transparenz sind eine Folge geringer Wahrscheinlichkeit. In einem Nachbearbeitungsschritt wird pixelweise eine Verstärkung des Deckungsgrades durchgeführt.

Framebuffers und verwendet für jeden Farbkanal (Rot, Grün, Blau und Alpha-Transparenz) eine Fließkommagenauigkeit von 32 Bit. In den Alphakanal wird pixelweise der resultierende Deckungsgrad der gerenderten Billboards gespeichert.

Der zweite Durchgang dient als Nachbearbeitungsschritt, in welchem der Off-Screen-Buffer in den Framebuffer gerendert wird. Dabei wird die Transparenz der gerenderten Pixel reduziert, um für Bereiche mit gerenderten Billboards eine volle Abdeckung zu erhalten (siehe Abb. 5.15). Diese Korrektur wird pixelweise berechnet. Jeder Pixel des Framebuffers  $C'$  wird aus dem entsprechenden Pixel des Off-Screen-Buffers  $C$  für jeden der vier Farbkanäle  $k$  ( $k \in \{R, G, B, A\}$ ) folgendermaßen berechnet:

$$C_k' = \frac{C_k}{\max(C_A, T_A)} \quad (5.11)$$

Da der Off-Screen-Buffer Alpha-Werte auf R, G und B multipliziert (auch als Premultiplied Alpha [PD84] bezeichnet), müssen für die Überführung in den Framebuffer Rot-, Grün-, und Blaukanal ebenfalls moduliert werden. Der Wert  $T_A$  (standardmäßig gesetzt auf 0,01) dient als untere Begrenzung des Alphawerts. Besitzen die Alphawerte von Pixeln mindestens diesen Wert, dann werden diese mit vollem Deckungsgrad wiedergegeben.



**Abbildung 5.16:** Verschiedene generierte Ansichten auf die Szene *Gassen*

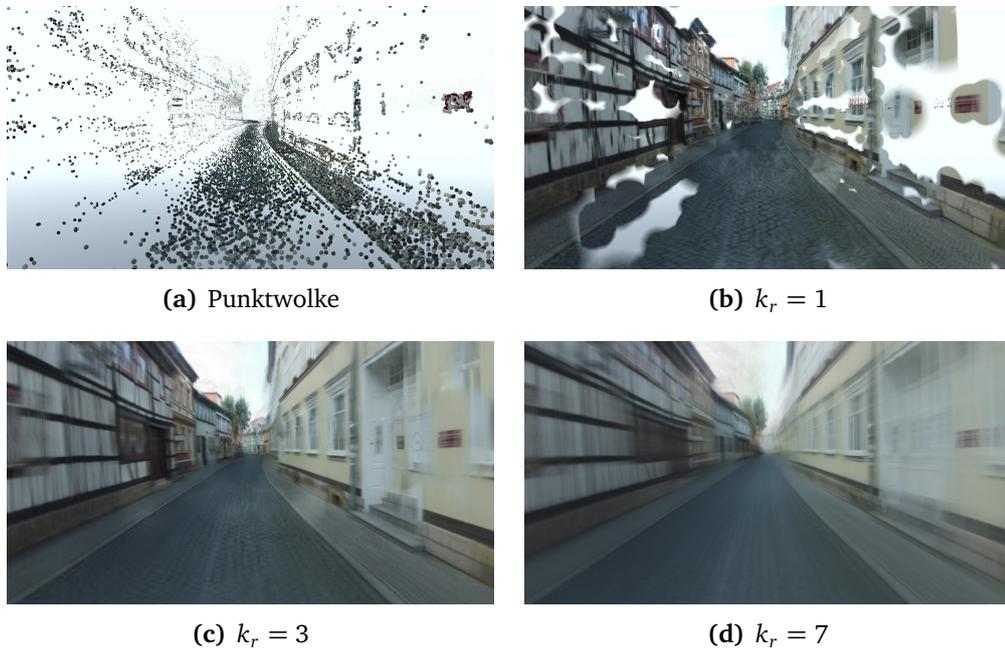
## 5.4 Auswertung

Die Abbildungen 5.16 und 5.17 zeigen generierte Ansichten auf zwei verschiedene Szenen. Für die Darstellung wurde  $N_\alpha$ , der Exponent zur Justierung des blickwinkelabhängigen Terms, auf 8 gesetzt. Der Exponent zur Übersetzung von Wahrscheinlichkeit in Deckungsgrade,  $N_f$ , wurde auf 3 gesetzt. Der Skalierungsfaktor  $k_r$  wurde ebenfalls auf 3 gesetzt. Für beide Szenen kann das vorgestellte Verfahren beliebige Ansichten allein aus den Daten der 3D-Photo-Collection in Echtzeit erzeugen, so dass Nutzer frei und interaktiv durch die Szene navigieren können. In Abbildung 5.19 wird die bildbasierte Darstellung mit der Synthese eines virtuellen Objekts kombiniert. Dabei können beliebige Ansichten in Echtzeit auf die Szene mit eingefügtem virtuellem Objekt erzeugt werden.

Wie in der Szene *Marktplatz* deutlich zu sehen, wird die Projektion der Bilder auf die mit einem begrenzten Radius konstruierten Billboards beschränkt. Einerseits wird hierdurch das Risiko von falsch vorhergesagter Geometrie reduziert, andererseits kann dadurch keine vollständige Abdeckung erreicht werden. Sichtbare Löcher sind vor allem in Bereichen der Szene zu sehen, welche eine zu geringe Punktdichte besitzen und daher nicht flächendeckend



**Abbildung 5.17:** Verschiedene generierte Ansichten auf die Szene *Markt-  
platz*



**Abbildung 5.18:** Verschiedene Größen der Billboards durch unterschiedliche Werte des Skalierungsfaktors  $k_r$  in der Szene *Gassen*.

**Tabelle 5.1:** Darstellungsleistung für die beiden Szenen *Marktplatz* und *Gassen*. Die Angaben für Bildrate und Renderingdauer der Billboards sind gemittelte Werte aus verschiedenen Ansichten.

Merkmal	Szene	
	Marktplatz	Gassen
Bilder	274	383
Billboards	32212	98895
Cluster	1773	2501
Dauer Tiefensortierung (ms)	1,31	2,03
Dauer Rendering der Billboards (ms)	13,25	24,71
Gesamtdauer für Frame (ms)	14,56	26,74
Durchschnittliche Bildrate (FPS)	68,7	37,4



**Abbildung 5.19:** Bildbasierte Darstellung mit eingefügtem virtuellen Modell

durch Billboards abgedeckt werden. Mangelnde Abdeckung ist auch ein Problem bei zu weit entfernten Bereichen (z.B. Himmel), hierfür liegen überhaupt keine Punkte vor, wodurch diese Bereiche in der Szene nicht wiedergegeben werden können.

Die Abdeckung der Billboards wird durch den blickwinkelabhängigen Term  $o_v(\alpha)$  beeinflusst: In Abhängigkeit von  $N_\alpha$  werden Billboards bei einem zu großen Reprojektionswinkel  $\alpha$  ausgeblendet, um Reprojektionsfehler zu verringern. Durch Vergrößerung des Skalierungsfaktors  $k_r$  kann die Abdeckung verbessert werden, jedoch führen zu große Billboards zu sehr unscharfen Bildergebnissen (siehe Abb. 5.18).

Wie in Tabelle 5.1 zu sehen, kann das vorgestellte Verfahren auch für große Szenen mit mehr als 300 Bildern hohe Bildraten erzielen. Die Angaben von Zeitdauer und Bildrate sind gemittelte Werte, da je nach Betrachterposition die Dauer für das Rendern aller Billboards variiert. Die Tiefensortierung macht ca. 8–9 % der Gesamtdauer für die Berechnung eines Frames aus. Alle

Testergebnisse wurden auf einem Rechner mit Intel Core i7 CPU 920 2.67GHz, 6 GB RAM und einer NVIDIA GeForce GTX 285 gemessen.

## **5.5 Zusammenfassung und Ausblick**

Es wurde ein Verfahren zum bildbasierten Rendern vorgestellt, welches die 3D-Punktwolke direkt verwendet und beliebige neue Ansichten in Echtzeit erzeugt. Durch die Modellierung der Geometrie mit Hilfe von Billboards, deren Größe durch die lokale Punktdichte bestimmt wird, können Bereiche der Szene entsprechend der Fülle der geometrischen Informationen mehr oder weniger detailliert dargestellt werden.

In zukünftigen Arbeiten sollen die nicht abgedeckten Bereiche reduziert werden. Eine Möglichkeit hierfür ist zum Beispiel das dynamische Skalieren des Billboardradius in Abhängigkeit der Nähe des Betrachters zur Kamera des beitragenden Bildes. Denn je ähnlicher sich Betrachter- und Kameraposition sind, desto geringer ist auch der Reprojektionsfehler.

Die Geschwindigkeit der Tiefensortierung kann ebenfalls verbessert werden. In der momentanen Implementierung wurde Mergesort von der CPU lediglich als ein Thread ausgeführt. Durch Parallelisierung auf der CPU oder unter Umständen sogar durch eine Verlagerung auf die Grafikhardware kann eine wesentlich bessere Leistung erzielt werden.

# Kapitel 6

## Zusammenfassung und Ausblick

### 6.1 Zusammenfassung

Ziel dieser Arbeit war die Entwicklung von Verfahren, um unter Nutzung von 3D-Photo-Collections eine hoch realistische Synthese von virtuellen und realen Anteilen zu ermöglichen. Dazu wurden die drei verschiedenen Problemfelder Farbcharakterisierung, bildbasierte Beleuchtung und bildbasierte Darstellung identifiziert und geeignete Methoden entwickelt. Mit diesen Verfahren konnte gezeigt werden, dass 3D-Photo-Collections für Augmented-Reality eine geeignete Repräsentation von realen Anteilen sind und mit ihnen eine realistische Synthese mit virtuellen Anteilen durchgeführt werden kann.

Nachfolgend sollen die entwickelten Verfahren noch einmal kurz dargestellt werden. Dabei soll auch der Zusammenhang zwischen den Verfahren und ihre unterschiedlichen Anforderungen an 3D-Photo-Collections erläutert werden. Zuletzt soll ein Ausblick auf mögliche zukünftige Weiterentwicklungen gegeben werden.

#### 6.1.1 Farbcharakterisierung

Es wurde ein neues Verfahren vorgestellt, mit dem die Farbcharakterisierung der Bilder von 3D-Photo-Collections vereinfacht wird. Dabei werden Bilder, welche mit unterschiedlichen Kameramodellen sowie mit unterschiedlicher Belichtungszeit oder Blendenzahl fotografiert wurden, automatisch in einen einzelnen, durch Auswahl eines Referenzkameramodells bestimmten, RGB-Farbraum transformiert.

Diese Abbildung in einen Farbraum wird erreicht, indem nach korrespondierenden Merkmalspunkten in Bildern von unterschiedlichen Kameramodellen

gesucht wird. Es werden RGB-Werte an den Merkmalspunkten entnommen, welche die Relation der RGB-Farbräume der einzelnen Kameramodelle zu einander beschreiben. Durch lineare Ausgleichsrechnung wird aus den korrespondierenden RGB-Werten eine  $3 \times 3$ -Transformationsmatrix berechnet, welche die RGB-Werte des einen Farbraums mit minimalem Fehler auf die RGB-Werte des anderen Farbraum abbildet. Durch diese vollständig automatisch bestimmte Abbildung können alle RGB-Farbräume in den RGB-Farbraum des Referenzkameramodells transformiert werden.

Nach der Transformation in den Referenz-RGB-Farbraum muss die eigentliche Charakterisierung in einen geräteunabhängigen Farbraum ausschließlich für das Referenzkameramodell durchgeführt werden. Die gewonnenen Charakterisierungsinformationen können dann für alle Bilder der 3D-Photo-Collection verwendet werden, da deren RGB-Werte ebenfalls im Referenz-RGB-Farbraum vorliegen.

### **6.1.2 Bildbasierte Beleuchtung**

Die bildbasierte Beleuchtung mit 3D-Photo-Collections ist das Hauptziel dieser Arbeit. Daher wurden zwei unterschiedliche Ansätze für die Rekonstruktion der realen Beleuchtung anhand von 3D-Photo-Collections entwickelt. Das Ziel beider Verfahren ist die Synthese einer Light-Probe, also die Erzeugung einer HDR-Environment-Map, welche für einen Punkt der Szene für alle Richtungen die eingestrahlte Strahldichte der Umgebung beschreibt.

Das erste Verfahren benötigt keine vollständige Szenengeometrie, sondern konstruiert anhand der Merkmalspunkte eines Bildes planare geometrische Hilfskörper. Auf diese werden die Bilder der 3D-Photo-Collection projiziert. Durch Ausnutzung überlappender Bilder mit unterschiedlichen Belichtungszeiten kann eine Erhöhung des dynamischen Umfangs durchgeführt werden, womit sowohl Lichtquellen als auch dunkle Oberflächen wiedergegeben werden können.

Im zweiten Verfahren ist die Szenengeometrie in Form eines 3D-Meshes gegeben, dessen Texturkoordinaten als Texturatlas definiert sind. Die Oberflächen des Meshes werden in Patches aufgeteilt und die Bilder auf die Patches projiziert. Wieder werden überlappende Bilder mit unterschiedlichen Belichtungszeiten dafür genutzt, um Strahldichten mit einem größeren dynamischen Umfang zu erhalten. Im Gegensatz zum ersten Verfahren wird für jeden Patch, und damit für jede reale Oberfläche der Szene, die Strahldichte rekonstruiert.

Aus dieser Repräsentation kann für einen beliebigen Punkt der Szene eine HDR-Environment-Map erzeugt werden, indem der Mesh in eine Cube-Map gerendert wird.

Beide Verfahren ermöglichen die Synthese von Light-Probes anhand der Bilder einer 3D-Photo-Collection für einen bestimmten Punkt im Raum, aus denen dann zum Beispiel direktionale Lichtquellen abgeleitet werden, mit denen eine konsistente Beleuchtung von virtuellen Modellen durchgeführt werden kann.

### 6.1.3 Darstellung bildbasierter Szenen

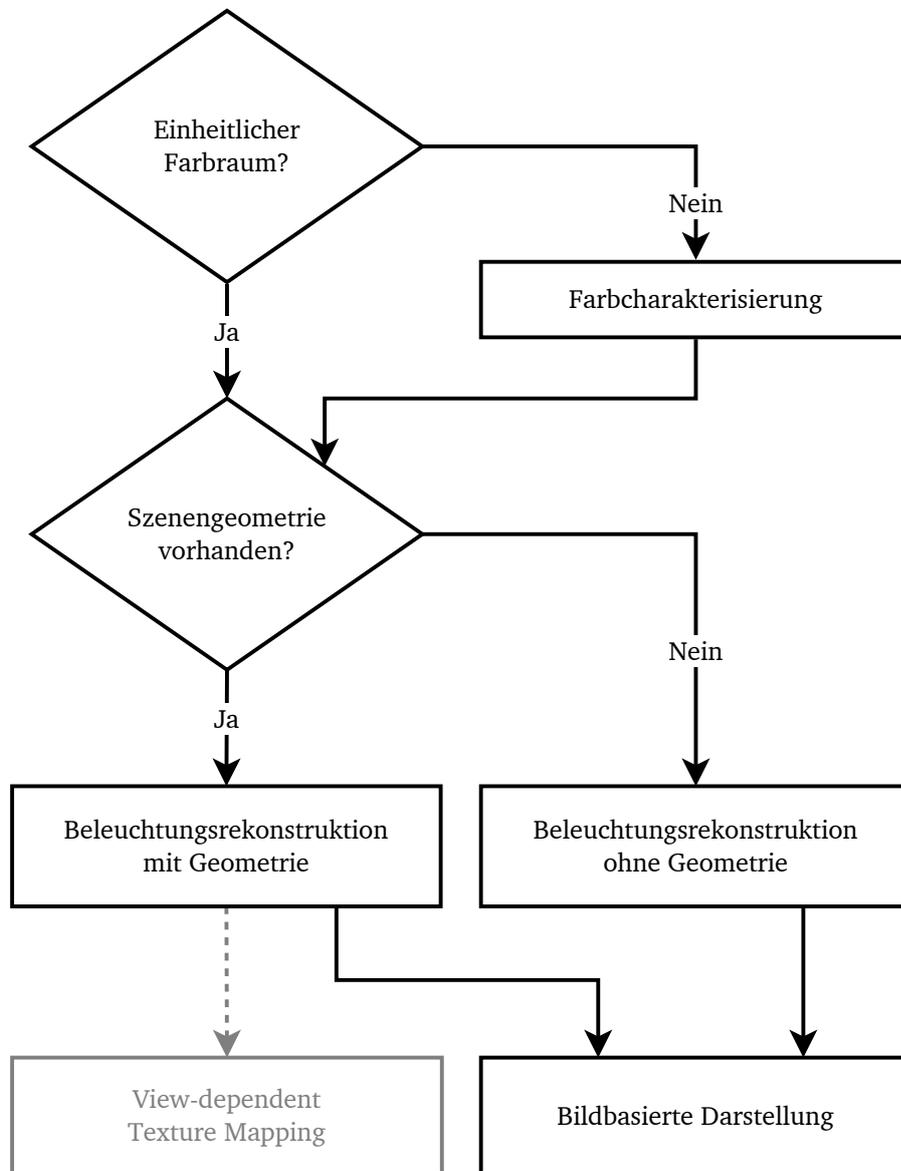
Es wurde ein Verfahren zur bildbasierten Darstellung von 3D-Photo-Collections entwickelt, mit dem beliebige neue Ansichten auf die Szene erzeugt werden können. Die Erzeugung neuer Ansichten basiert auf einer direkten Nutzung der Punktwolke, so dass keine zeitaufwändige 3D-Rekonstruktion durchgeführt werden muss.

Die räumliche Struktur der Szene wird hierbei mit Hilfe von Billboards dargestellt. Der Radius und die Transparenz der Billboards werden aus einer wahrscheinkeitsbasierten Modellierung der Szenengeometrie anhand der Punktwolke berechnet. Dadurch passt sich der Detailgrad der Darstellung an die lokale Dichte der Punktwolke an. Durch die Ausrichtung des Verfahrens auf die Grafikhardware wird eine echtzeitfähige Darstellung auch für sehr große Szenen erreicht.

## 6.2 Zusammenhang zwischen den Verfahren

Jedes der vorgestellten Verfahren trägt einen Teil zur Umsetzung von bildbasierter Beleuchtung mit 3D-Photo-Collections bei (siehe Abb. 6.1). Das Verfahren zur Farbcharakterisierung vereinheitlicht unterschiedliche RGB-Farbräume der Bilder zu einem einzigen RGB-Farbraum. Das ist die Voraussetzung für eine Beleuchtungsrekonstruktion und eine bildbasierte Darstellung, da hier RGB-Informationen aus unterschiedlichen Bildern kombiniert werden.

Für die bildbasierte Beleuchtung wurden 3D-Photo-Collections zur Beschreibung realer Anteile verwendet. Allerdings bieten 3D-Photo-Collections nur eine begrenzte Anzahl von Ansichten auf die reale Szene in Form der Eingabebilder mit bekannten Kameraposen. Das entwickelte Verfahren zur



**Abbildung 6.1:** Prozessablauf der vorgestellten Verfahren

### 6.3 Anforderungen an 3D-Photo-Collections

Anforderung	Farbcharakterisierung	Beleuchtungsrekonstruktion		Bildbasierte Darstellung
		ohne Geometrie	mit Geometrie	
Einheitlicher Farbraum	Nein	Ja	Ja	Ja
Merkmalspunkte	Ja	Ja	Nein	Ja
Diffuse Oberflächen	Teilweise	Ja	Ja	Nein
Kamerakennkurve	Nein	Ja	Ja	Nein
Variable Belichtungszeit	Nein	Ja	Ja	Nein
Vollständige Geometrie	Nein	Nein	Ja	Nein

**Tabelle 6.1:** Anforderungen der Verfahren

bildbasierten Darstellung erweitert die Nutzung von 3D-Photo-Collections für die Synthese von realen und virtuellen Anteilen, indem es die Generierung von beliebigen Ansichten auf die reale Szene ermöglicht. Virtuelle Anteile können aus beliebigen Ansichten dargestellt werden, da sie in Form von 3D-Meshes gegeben sind und im Koordinatensystem der Kameraposen liegen. Insgesamt kann durch die bildbasierte Darstellung eine Synthese von realen Anteilen und konsistent beleuchteten virtuellen Objekten für beliebige Ansichten auf die Szene erzeugt werden.

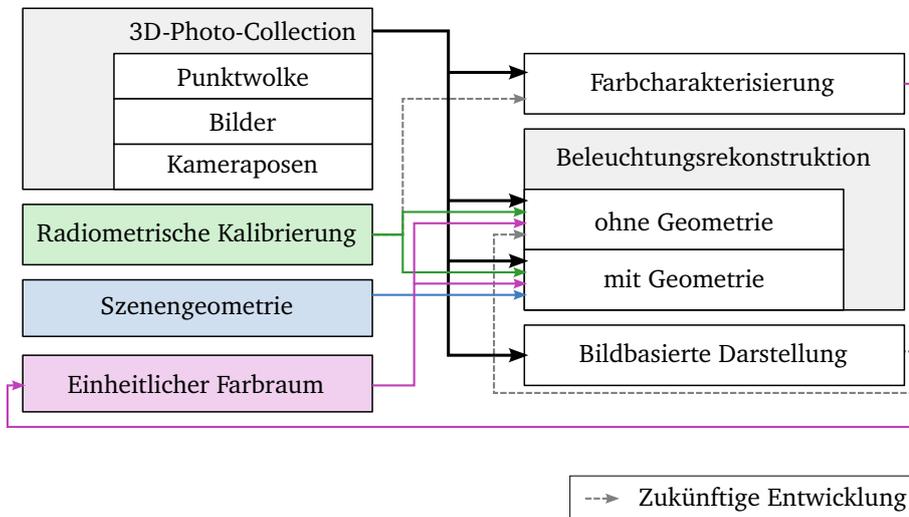
## 6.3 Anforderungen an 3D-Photo-Collections

Jedes der insgesamt vier Verfahren hat verschiedene Anforderungen an die Beschaffenheit der 3D-Photo-Collections (siehe Tab. 6.1). Beleuchtungsrekonstruktion und bildbasierte Darstellung benötigen einen einheitlichen RGB-Farbraum für alle Bilder. Die Farbcharakterisierung führt diese Farbraumvereinheitlichung selbst durch und benötigt daher diese Anforderung nicht.

Punktwolke und Merkmalspunkte sind in jeder 3D-Photo-Collection vorhanden, da sonst auch keine Kameraposen rekonstruiert werden können. Jedoch variiert das Rauschen und die Dichte der Punktwolke je nach Beschaffenheit der realen Szene, was für viele der vorgestellten Verfahren bedeutsam ist. Punktwolke und Merkmalspunkte werden von allen Verfahren – bis auf die Beleuchtungsrekonstruktion mit Geometrie – benötigt. Die Farbcharakterisierung modelliert die Relation zwischen unterschiedlichen Farbräumen anhand

korrespondierender Merkmalspunkte zwischen Bildern von unterschiedlichen Kameramodellen. Je mehr Merkmalspunkte vorhanden sind, desto mehr Korrespondenzen existieren in der Regel, was bis zu einem gewissen Grad zu einer robusteren Farbraumtransformation führt. Für die Beleuchtungsrekonstruktion ohne Geometrie sind die Merkmalspunkte wichtig für die Konstruktion geometrischer Hilfskörper, auf denen die Bilder projiziert werden. Diese Konstruktion ist aber relativ robust gegenüber einer verrauschten oder dünn besetzten Punktwolke, da lediglich der Median aus Tiefenwerten berechnet wird. Für die Beleuchtungsrekonstruktion mit Geometrie werden keine Merkmalspunkte verwendet, da die Bilder direkt auf die bekannten Oberflächen projiziert werden. Die bildbasierte Darstellung erfordert allerdings eine möglichst dichte Punktwolke und kann auch nur die realen Oberflächen wiedergeben, welche ansatzweise in der Punktwolke repräsentiert werden.

Das Reflexionsverhalten der Szenenoberflächen ist ebenfalls für die meisten Verfahren bedeutend. Vollständig diffuses Reflexionsverhalten (Lambertscher Strahler) aller Oberflächen wird für die Beleuchtungsrekonstruktion benötigt, da fotografierte Strahldichten der Oberflächen für alle Richtungen als konstant angenommen werden (siehe Lambertsches Gesetz in Abschnitt 2.1.5). Für die Farbcharakterisierung muss das Reflexionsverhalten an Merkmalspunkten für alle Richtungen stabil sein, um für unterschiedliche Kameramodelle identische Strahldichten an diesen Punkten zu messen und somit vergleichbare RGB-Werte zu erhalten. Für die bildbasierte Darstellung wird kein diffuses Reflexionsverhalten gefordert. Dieses Verfahren ist sogar in der Lage bis zu einem gewissen Grad glänzende Materialien wiederzugeben, da für eine bestimmte Blickrichtung Inhalte aus Bildern mit ähnlicher Kamerarichtung stärker gewichtet werden, womit eine blickwinkelabhängige Darstellung erreicht wird. Allerdings ist hier das Problem, dass Bundler nur für diffuse Oberflächen Merkmalspunkte ausgibt, so dass glänzende Oberflächen in der Regel keine Merkmalspunkte aufweisen und somit bei der bildbasierten Darstellung nicht oder nur teilweise wiedergegeben werden können. Speziell für die beiden Verfahren zur Beleuchtungsrekonstruktion sind eine Rekonstruktion der Kamerakennkurve sowie Bilder mit unterschiedlichen Belichtungszeiten erforderlich. Die Kamerakennkurve wird zur Erweiterung des dynamischen Umfangs anhand der Bilder mit unterschiedlichen Belichtungszeiten verwendet.



**Abbildung 6.2:** Benötigte Eingabedaten der Verfahren

## 6.4 Ausblick

Aus den für die einzelnen Verfahren ausgewiesenen möglichen zukünftigen Entwicklungen (siehe Abschnitt 3.5, 4.5.2 und 5.5) soll in diesem Abschnitt ein Gesamtausblick auf die bildbasierte Beleuchtung mit 3D-Photo-Collections gegeben werden.

### 6.4.1 Abhängigkeiten zu externen Daten

Einige der Verfahren hängen von zusätzlichen Eingabedaten wie Kamerakennkurve oder vollständiger Szenengeometrie ab (siehe Abb. 6.2). Da diese mit erhöhtem Aufwand bereit gestellt werden müssen, soll nach Wegen gesucht werden diese externen Daten automatisch aus der 3D-Photo-Collection abzuleiten.

So benötigt die Beleuchtungsrekonstruktion eine getrennt durchgeführte Rekonstruktion der Kamerakennkurve. Hierfür muss für jedes Kameramodell eine separate Belichtungsreihe von Bildern fotografiert werden. Dieser zusätzliche Aufwand kann entfallen, falls eine Möglichkeit zur automatischen Rekonstruktion der Kamerakennkurve entwickelt wird (siehe Abschnitt 4.5.2).

Eine weitere Abhängigkeit ist die vollständige Szenengeometrie für die Beleuchtungsrekonstruktion mit Geometrie. Die manuelle Erzeugung einer Szenengeometrie könnte entfallen, falls automatische 3D-Rekonstruktionsverfahren, wie etwa Multi-View-Stereo, in Zukunft robust genug sind, so dass die generierten 3D-Meshes für die Beleuchtungsrekonstruktion verwendet werden können.

## **6.4.2 Überschneidung der Problemfelder**

Im Rahmen dieser Arbeit wurden die drei bearbeiteten Problemfelder Farbcharakterisierung, Rekonstruktion der Beleuchtung und bildbasierte Darstellung getrennt voneinander betrachtet. Einige Verfahren der unterschiedlichen Problemfelder verfolgen jedoch gleiche Teilziele.

### **Kamerakennkurve und Linearisierung**

Sowohl der Linearisierungsschritt der Farbcharakterisierung als auch die Kamerakennkurve der Beleuchtungsrekonstruktion zielen darauf ab, den Zusammenhang zwischen fotografierten RGB-Werten und der entsprechenden Bestrahlung der Kamerasensorfläche zu beschreiben.

Die für die Beleuchtungsrekonstruktion gewonnene Kamerakennkurve könnte auch zur Linearisierung der RGB-Werte für die Farbcharakterisierung verwendet werden. Es würden dann gegenüber dem bisherigen Linearisierungsschritt genauere Ergebnisse erzielt, da dieser lediglich auf der Annahme eines sRGB-ähnlichen Farbraums beruht, die Kamerakennkurve jedoch auf einer Messung der Farbwiedergabe der Kamera basiert.

Umgekehrt kann auch auf eine separate Rekonstruktion der Kamerakennkurve verzichtet werden, wenn diese automatisch anhand der Eingabedaten der 3D-Photo-Collection erreicht wird. Dies könnte über zwei verschiedene Ansätze gelingen. Zum einen kann bei vorhandener vollständiger Szenengeometrie eine Rekonstruktion der Kamerakennkurve durch die Projektion der Bilder auf Patches durchgeführt werden (siehe Abschnitt 4.5.2). Zum anderen könnten in der Farbcharakterisierung korrespondierende RGB-Werte von Bildern, welche vom gleichen Kameramodell, aber mit unterschiedlichen Belichtungszeiten fotografiert wurden, dazu genutzt werden, die Kamerakennkurve zu rekonstruieren. Die Herausforderung dabei ist, dass wesentlich

weniger Abtastwerte vorliegen als bei der Rekonstruktion über eine Belichtungsreihe. An dieser Stelle könnte daher das Verfahren von Grossberg und Nayar (siehe Abschnitt 4.2.2) eingesetzt werden, da hier nur wenige Messwerte benötigt werden.

### **Bildbasierte Darstellung für Beleuchtungsrekonstruktion**

Für die Beleuchtungsrekonstruktion ohne Geometrie werden bisher planare geometrische Hilfskörper konstruiert und darauf Bilder projiziert. Ähnlich zum Verfahren der bildbasierten Darstellung kann es als eine grobe Rekonstruktion der plenoptischen Funktion angesehen werden, da es beliebige neue Ansichten auf die 3D-Photo-Collection erzeugt. Demgegenüber bietet das vorgestellte bildbasierte Verfahren bei einer entsprechend dichten Punktwolke jedoch eine genauere Wiedergabe der räumlichen Struktur der Szene. Daher könnte es für geeignete Szenen anstatt planarer Hilfskörper für die Beleuchtungsrekonstruktion ohne Geometrie verwendet werden. Allerdings müsste hierfür die bildbasierte Darstellung so angepasst werden, dass mehrere auf gleiche Billboards projizierte Bilder mit unterschiedlichen Belichtungszeiten zur Erweiterung des dynamischen Umfangs verwendet werden.

### **6.4.3 Fazit**

Insgesamt wurden mehrere Verfahren entwickelt, welche eine realistische Augmented-Reality-Synthese von realen und virtuellen Anteilen unter Verwendung von 3D-Photo-Collections durchführen oder dazu beitragen. Wie bereits erläutert, können für die präsentierten Verfahren noch viele Möglichkeiten zur Weiterentwicklung untersucht werden, womit die in dieser Arbeit thematisierte Problemstellung sicher auch in Zukunft noch Gegenstand der Forschung sein wird.



## Eigene Veröffentlichungen

- [1] BIRNBACH, Bastian ; GRIMM, Paul ; KÖLZER, Konrad ; NAGL, Frank: Prototyping von AR - Präsentationen mit ARBlender. In: 5. *GI-Workshop „Virtuelle und Erweiterte Realität“ der GI-Fachgruppe VR/AR*, 2008
- [2] BIRNBACH, Bastian ; GRIMM, Paul ; KÖLZER, Konrad ; NAGL, Frank: Automatische Bild- und Darstellungsoptimierung für 3D-Photo-Collections. In: 6. *GI-Workshop „Virtuelle und Erweiterte Realität“ der GI-Fachgruppe VR/AR*, 2009
- [3] KÖLZER, Konrad ; FRANK NAGL, Bastian B. ; GRIMM, Paul: Rendering Virtual Objects with High Dynamic Range Lighting Extracted Automatically from Unordered Photo Collections. In: *Lecture Notes in Computer Science - Advances in Visual Computing* Bd. 5876/2009, Springer Berlin / Heidelberg, 2009, S. 992–1001
- [4] KÖLZER, Konrad ; JUNG, Yvonne ; PAUL GRIMM, Bastian B.: Grafikkartenbasierte Simulation von tröpfchenförmigen Flüssigkeiten auf Oberflächen. In: SCHUMANN, M. (Hrsg.): *Virtuelle und Erweiterte Realität : 5. Workshop der GI-Fachgruppe VR/AR*. Aachen : Shaker, 2008
- [5] KÖLZER, Konrad ; MÜLLER, Stefan ; GRIMM, Paul: Automatic color profiling of digital cameras using unordered photos. In: *IbPRIA 2013: 6th Iberian Conference on Pattern Recognition and Image Analysis*, 2013. – Veröffentlichung steht bevor
- [6] KÖLZER, Konrad ; NAGL, Frank ; BIRNBACH, Bastian ; GRIMM, Paul: Automatische Extraktion von HDR-Beleuchtungsdaten aus Fotos. In: 6. *GI-Workshop „Virtuelle und Erweiterte Realität“ der GI-Fachgruppe VR/AR*, 2009
- [7] KÖLZER, Konrad ; NAGL, Frank ; GRIMM, Paul: Synthesizing Relative Radiance for Realistic Rendering of Virtual Objects in 3D Photo Collec-

- tions. In: *Eurographics 2011 - Areas Papers*, Eurographics Association, 2011, 71-72
- [8] KÖLZER, Konrad ; NAGL, Frank ; GRIMM, Paul ; MÜLLER, Stefan: Rendering Image-Based Scenes in Real-Time Utilizing Sparse Scene Geometry. In: *Computer Graphics International (CGI'2012)*, 2012
- [9] NAGL, Frank ; BIRNBACH, Bastian ; KÖLZER, Konrad ; GRIMM, Paul: Perspektivisch korrektes Einblenden virtueller 3D-Objekte in Einzelfotos. In: *6. GI-Workshop "Virtuelle und Erweiterte Realität" der GI-Fachgruppe VR/AR*, 2009
- [10] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul: MOPS - A lightweight software layer for the definition and rendering of materials. In: *Proceedings No.1, Section 1: Information Technologies and Artificial Intelligence MIET - Modern Informations and Electronic Technologies*, 11th International scientific-functional conference., Politechperiodika, 2010
- [11] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul: Occlusion Handling and Image-based Lighting using Sliced Images in 3D Photo Collections. In: *International Conference on Artificial Reality and Telexistence*, 2011
- [12] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul ; BINDEL, Tobias ; ROTHE, Stephan: ConGrap – Contour Detection based on Gradient Map of Images. In: *Computer Technology and Application 2 (2011)*, Nr. 8, S. 628–637
- [13] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul ; BINDEL, Tobias ; ROTHE, Stephan: ConGrap – Contour Detection based on Gradient Map of Images. In: ARABNIA, Hamid R. (Hrsg.) ; DELIGIANNIDIS, Leonidas (Hrsg.) ; SCHAEFER, Gerald (Hrsg.): *Proceedings of the 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 2011)* Bd. 2, CSREA Press, 2011, S. 870–875
- [14] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul ; BIRNBACH, Bastian: Generating Depth Maps From Photo Collections. In: *Proceedings No.1, Section 1: Information Technologies and Artificial Intelligence MIET - Modern Informations and Electronic Technologies*, 10th International scientific-functional conference., Politechperiodika, 2009

- [15] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Paul ; BIRNBACH, Bastian: Verdeckung virtueller Objekte durch reale Bildanteile in Fotos. In: 6. *GI-Workshop „Virtuelle und Erweiterte Realität“ der GI-Fachgruppe VR/AR*, 2009
- [16] NAGL, Frank ; KÖLZER, Konrad ; GRIMM, Prof. Dr. P.: SBIP – A Framework for Real-time Capable Image Processing in Videos. In: *MIET – Modern Informations and Electronic Technologies 2011. 12th International scientific-functional conference*. Odessa : Politechperiodika, 2011 (Proceedings, Section 2: Information Technologies and Artificial Intelligence), S. 90



# Literaturverzeichnis

- [AB91] ADELSON, E.H. ; BERGEN, J.R.: The Plenoptic Function and the Elements of Early Vision. In: *Computational Models of Visual Processing 1* (1991), S. 3–20
- [AT10] ALLEN, E. ; TRIANTAPHILLIDOU, S.: *The Manual of Photography and Digital Imaging*. Elsevier Science, 2010. – ISBN 9780240520377
- [BBM<sup>+</sup>01] BUEHLER, Chris ; BOSSE, Michael ; MCMILLAN, Leonard ; GORTLER, Steven ; COHEN, Michael: Unstructured Lumigraph Rendering. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01* (2001), 425–432. <http://dx.doi.org/10.1145/383259.383309>. – DOI 10.1145/383259.383309. ISBN 158113374X
- [BF02] BARNARD, K. ; FUNT, B.: Camera characterization for color research. In: *COLOR research and application 27* (2002), Nr. 3, S. 153–164
- [BS04] BERGMANN, Ludwig ; SCHAEFER, Clemens ; NIEDRIG, Heinz (Hrsg.): *Lehrbuch der Experimentalphysik Band 3 Optik*. 10. Walter de Gruyter, 2004
- [Cro77] CROW, Franklin C.: Shadow algorithms for computer graphics. In: *Proceedings of the 4th annual conference on Computer graphics and interactive techniques*. New York, NY, USA : ACM, 1977 (SIGGRAPH '77), 242–248
- [CW02] CHEUNG, T. L. V. ; WESTLAND, S.: Color Camera Characterisation Using Artificial Neural Networks. In: *IS&T/SID Tenth Color Imaging Conference*, 2002, S. 117–120

- [CWL<sup>+</sup>05] CHEUNG, V. ; WESTLAND, S. ; LI, C. ; HARDEBERG, J. ; CONNAH, D.: Characterization of trichromatic color cameras by using a new multispectral imaging technique. In: *JOSA A* 22 (2005), Nr. 7, S. 1231–1240
- [Deb98] DEBEVEC, Paul: Rendering synthetic objects into real scenes: Birging traditional and image-based graphics with global illumination and high dynamic reage photograph. In: *SIGGRAPH'98 Conf. Proc.*, 1998, S. 189–198
- [Deb05] DEBEVEC, Paul: A median cut algorithm for light probe sampling. In: *ACM SIGGRAPH 2005 Posters*. New York, NY, USA : ACM, 2005 (SIGGRAPH '05)
- [DKV06] DOBRINSKI, Paul ; KRAKAU, Gunter ; VOGEL, Anselm: *Physik für Ingenieure*. 11. Wiesbaden : B. G. Teubner Verlag, 2006
- [DM97] DEBEVEC, Paul E. ; MALIK, Jitendra: Recovering High Dynamic Range Radiance Maps from Photographs. In: WHITTED, Turner (Hrsg.): *Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, Annual Conference Series*. Los Angeles, California : Addison Wesley, August 1997, S. 369–378. – ISBN 0-89791-896-7
- [DRB97] DRETTAKIS, George ; ROBERT, Luc ; BUGNOUX, Sylvain: Interactive Common Illumination for Computer Augmented Reality. In: *8th Eurographics workshop on Rendering*. Saint Etienne, France, 1997, 45–56
- [DTM96] DEBEVEC, Paul E. ; TAYLOR, Camillo J. ; MALIK, Jitendra: Modeling and Rendering Architecture fromPhotographs: A hybrid geometry- and image-based approach. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, 1996, 11–20
- [DYB98] DEBEVEC, P.E. ; YU, Yizhou ; BORSHUKOV, George: Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping. In: *Eurographics Rendering Workshop Bd. 98*, 1998, 105–116

- [FCSS09a] FURUKAWA, Y. ; CURLESS, B. ; SEITZ, S.M. ; SZELISKI, R.: Manhattan-world stereo. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on IEEE*, 2009, S. 1422–1429
- [FCSS09b] FURUKAWA, Y. ; CURLESS, B. ; SEITZ, S.M. ; SZELISKI, R.: Reconstructing building interiors from images. In: *Computer Vision, 2009 IEEE 12th International Conference on IEEE*, 2009, S. 80–87
- [FFGG<sup>+</sup>10] FRAHM, Jan-Michael ; FITE-GEORGEL, Pierre ; GALLUP, David ; JOHNSON, Tim ; RAGURAM, Rahul ; WU, Changchang ; JEN, Yi-Hung ; DUNN, Enrique ; CLIPP, Brian ; LAZEBNIK, Svetlana ; POLLEFEYS, Marc: Building Rome on a cloudless day. In: *Proceedings of the 11th European conference on Computer vision: Part IV*. Berlin, Heidelberg : Springer-Verlag, 2010 (ECCV'10). – ISBN 3–642–15560–X, 978–3–642–15560–4, 368–381
- [FGR93] FOURNIER, A. ; GUNAWAN, A.S. ; ROMANZIN, C.: Common illumination between real and computer generated scenes. In: *Graphics Interface Canadian Information Processing Society*, 1993, S. 254–262
- [FP07] FURUKAWA, Y. ; PONCE, J.: Accurate, Dense, and Robust Multi-View Stereopsis. In: *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007. – ISSN 1063–6919, S. 1–8
- [FP10] FURUKAWA, Y. ; PONCE, J.: Accurate, Dense, and Robust Multiview Stereopsis. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32 (2010), aug., Nr. 8, S. 1362–1376. <http://dx.doi.org/10.1109/TPAMI.2009.161>. – DOI 10.1109/TPAMI.2009.161. – ISSN 0162–8828
- [GAF<sup>+</sup>10] GOESELE, Michael ; ACKERMANN, Jens ; FUHRMANN, Simon ; HAUBOLD, Carsten ; KLOWSKY, Ronny ; STEEDLY, Drew ; SZELISKI, Richard: Ambient point clouds for view interpolation. In: *ACM SIGGRAPH 2010 papers*. New York, NY, USA : ACM, 2010 (SIGGRAPH '10). – ISBN 978–1–4503–0210–4, 95:1–95:6

- [GCHH03] GIBSON, S. ; COOK, J. ; HOWARD, T. ; HUBBOLD, R.: Rapid shadow generation in real-world lighting environments. In: *Proceedings: EuroGraphics Symposium on Rendering, Leuven, Belgium Citeseer, 2003*
- [GGSC96] GORTLER, Steven J. ; GRZESZCZUK, Radek ; SZELISKI, Richard ; COHEN, Michael F.: The lumigraph. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, NY, USA : ACM, 1996 (SIGGRAPH '96). – ISBN 0-89791-746-4, 43-54
- [GHCH03] GIBSON, S. ; HUBBOLD, R.J. ; COOK, J. ; HOWARD, T.L.J.: Interactive reconstruction of virtual environments from video sequences. In: *Computers & Graphics 27 (2003), Nr. 2, S. 293-301*
- [GK65] GOLUB, G. ; KAHAN, W.: Calculating the singular values and pseudo-inverse of a matrix. In: *Journal of the Society for Industrial & Applied Mathematics, Series B: Numerical Analysis 2 (1965), Nr. 2, S. 205-224*
- [GN03] GROSSBERG, M.D. ; NAYAR, S.K.: What is the space of camera response functions? In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* Bd. 2, 2003. – ISSN 1063-6919, S. II – 602-9 vol.2
- [GN04] GROSSBERG, M.D. ; NAYAR, S.K.: Modeling the space of camera response functions. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on 26 (2004), Nr. 10, S. 1272-1282*
- [Gro05] GROSCH, Thorsten: Differential Photon Mapping - Consistent Augmentation of Photographs with Correction of all Light Paths. In: *EG Short Presentations*. Trinity College, Dublin, Ireland, 2005, 53-56
- [GSC<sup>+</sup>07] GOESELE, Michael ; SNAVELY, Noah ; CURLLESS, Brian ; HOPPE, Hugues ; SEITZ, Steven M.: Multi-View Stereo for Community Photo Collections. In: *2007 IEEE 11th International Conference on Computer Vision, IEEE, 2007*. – ISBN 978-1-4244-1630-1, 1-8

- [GSHG98] GREGER, Gene ; SHIRLEY, Peter ; HUBBARD, Philip M. ; GREENBERG, Donald P.: The Irradiance Volume. In: *IEEE Comput. Graph. Appl.* 18 (1998), März, Nr. 2, 32–43. <http://dx.doi.org/10.1109/38.656788>. – DOI 10.1109/38.656788. – ISSN 0272–1716
- [HLR01] HONG, Guowei ; LUO, M. R. ; RHODES, Peter A.: A study of digital camera colorimetric characterisation based on polynomial modelling. In: *Color Research and Application* 26 (2001), S. 76–84
- [HP11] HUNT, Robert W. G. ; POINTER, Michael R. ; KRIS, Michael A. (Hrsg.): *Measuring Colour*. 4. Wiley, 2011
- [IB99] IMAI, F.H. ; BERNS, R.S.: Spectral estimation using trichromatic digital cameras. In: *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives* Bd. 42, 1999
- [Ive15] IVES, H.E.: The transformation of color-mixture equations from one system to another. In: *Journal of the Franklin Institute* 180 (1915), Nr. 6, S. 673–701
- [IW05] ILIE, Adrian ; WELCH, Greg: Ensuring Color Consistency across Multiple Cameras. In: *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*. Washington, DC, USA : IEEE Computer Society, 2005 (ICCV '05). – ISBN 0–7695–2334–X–02, 1268–1275
- [Jen96] JENSEN, H.W.: Global illumination using photon maps. In: *Rendering Techniques* 96 (1996), S. 21–30
- [KFNG09] KÖLZER, Konrad ; FRANK NAGL, Bastian B. ; GRIMM, Paul: Rendering Virtual Objects with High Dynamic Range Lighting Extracted Automatically from Unordered Photo Collections. In: *Lecture Notes in Computer Science - Advances in Visual Computing* Bd. 5876/2009, Springer Berlin / Heidelberg, 2009, S. 992–1001
- [KMG13] KÖLZER, Konrad ; MÜLLER, Stefan ; GRIMM, Paul: Automatic color profiling of digital cameras using unordered photos. In:

*IbPRIA 2013: 6th Iberian Conference on Pattern Recognition and Image Analysis*, 2013. – Veröffentlichung steht bevor

- [KNBG09] KÖLZER, Konrad ; NAGL, Frank ; BIRNBACH, Bastian ; GRIMM, Paul: Automatische Extraktion von HDR-Beleuchtungsdaten aus Fotos. In: *6. GI-Workshop „Virtuelle und Erweiterte Realität“ der GI-Fachgruppe VR/AR*, 2009
- [KNG11] KÖLZER, Konrad ; NAGL, Frank ; GRIMM, Paul: Synthesizing Relative Radiance for Realistic Rendering of Virtual Objects in 3D Photo Collections. In: *Eurographics 2011 - Areas Papers*, Eurographics Association, 2011, 71-72
- [KNGM12] KÖLZER, Konrad ; NAGL, Frank ; GRIMM, Paul ; MÜLLER, Stefan: Rendering Image-Based Scenes in Real-Time Utilizing Sparse Scene Geometry. In: *Computer Graphics International (CGI'2012)*, 2012
- [Knu98] KNUTH, Donald E.: *The Art of Computer Programming, Volume 3: Sorting and Searching*. 2. Addison-Wesley, 1998
- [LA04] LOURAKIS, Manolis I. A. ; ARGYROS, Antonis A.: The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm / Institute of Computer Science - FORTH. 2004. – Forschungsbericht
- [LFD<sup>+</sup>99] LOSCOS, Céline ; FRASSON, Marie-Claude ; DRETTAKIS, George ; WALTER, Bruce ; GRANIER, Xavier ; POULIN, Pierre: Interactive Virtual Relighting and Remodeling of Real Scenes / INRIA. Version: 1999. <http://hal.inria.fr/inria-00069942>. 1999 (RT-0230). – Rapport de recherche. – 16 S.
- [LH96] LEVOY, Marc ; HANRAHAN, Pat: Light Field Rendering. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, 1996. – ISBN 0897917464, 31–42
- [LL01] LI, Changjun ; LUO, Ronnier: The Estimation of Spectral Reflectances Using the Smoothest Constraint Condition. In: *Ninth Color Imaging Conference: Color Science and Engineering: Systems, Technologies, and Applications* Bd. 9, 2001, S. 62–67

- [Low04] LOWE, David G.: Distinctive image features from scale-invariant keypoints. In: *International Journal of Computer Vision* 60 (2004), Nr. 2, S. 91–110
- [LPRMt02] LÉVY, Bruno ; PETITJEAN, Sylvain ; RAY, Nicolas ; MAILLO T, Jérôme: Least Squares Conformal Maps for Automatic Texture Atlas Generation. In: ACM (Hrsg.): *ACM SIGGRAPH conference proceedings, 2002*
- [Lut27] LUTHER, Robert: Aus dem Gebiet der Farbreizmetrik (On color stimulus metrics). In: *Zeitschrift für technische Physik* 8 (1927), Nr. 1927, S. 540–558
- [MB95] MCMILLAN, Leonard ; BISHOP, Gary: Plenoptic modeling: an image-based rendering system. In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. New York, NY, USA : ACM, 1995 (SIGGRAPH '95). – ISBN 0–89791–701–4, 39–46
- [Mic] MICROSOFT: *Photosynth - Capture your world in 3D*. <http://photosynth.net>
- [MW86] MALONEY, Laurence T. ; WANDELL, Brian A.: Color constancy: a method for recovering surface spectral reflectance. In: *J. Opt. Soc. Am. A* 3 (1986), Jan, Nr. 1, 29–33. <http://dx.doi.org/10.1364/JOSAA.3.000029>. – DOI 10.1364/JOSAA.3.000029
- [NHIN86] NAKAMAE, Eihachiro ; HARADA, Koichi ; ISHIZAKI, Takao ; NISHITA, Tomoyuki: A montage method: the overlaying of the computer generated images onto a background photograph. In: *SIGGRAPH Comput. Graph.* 20 (1986), August, S. 207–214. – ISSN 0097–8930
- [PCD<sup>+</sup>97] PULLI, Kari ; COHEN, Michael ; DUCHAMP, Tom ; HOPPE, Hugues ; SHAPIRO, Linda G. ; STUETZLE, Werner: View-base Rendering: Visualizing Real Objects from Scanned Range and Color Data. In: *Proceedings of the Eurographics Workshop on Rendering Techniques '97*. London, UK, UK : Springer-Verlag, 1997. – ISBN 3–211–83001–4, 23–34

- [PD84] PORTER, T. ; DUFF, T.: Compositing digital images. In: *ACM Siggraph Computer Graphics* 18 (1984), Nr. 3, S. 253–259
- [PTISK10] PANAHOUPUR TEHRANI, Mehrdad ; ISHIKAWA, Akio ; SAKAZAWA, Shigeyuki ; KOIKE, Atsushi: Iterative colour correction of multicamera systems using corresponding feature points. In: *J. Vis. Comun. Image Represent.* 21 (2010), Juli, Nr. 5-6, 377–391. <http://dx.doi.org/10.1016/j.jvcir.2010.03.007>. – DOI 10.1016/j.jvcir.2010.03.007. – ISSN 1047–3203
- [RBS99] ROBERTSON, M.A. ; BORMAN, S. ; STEVENSON, R.L.: Dynamic range improvement through multiple exposures. In: *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* Bd. 3 IEEE, 1999, S. 159–163
- [RBS03] ROBERTSON, M.A. ; BORMAN, S. ; STEVENSON, R.L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. In: *Journal of Electronic Imaging* 12 (2003), S. 219
- [Rob96] ROBERT, Luc: Camera calibration without feature extraction. In: *Comput. Vis. Image Underst.* 63 (1996), März, Nr. 2, 314–325. <http://dx.doi.org/10.1006/cviu.1996.0021>. – DOI 10.1006/cviu.1996.0021. – ISSN 1077–3142
- [RWPD05] REINHARD, Erik ; WARD, Greg ; PATTANAIK, Sumanta ; DEBEVEC, Paul: *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2005. – ISBN 0125852630
- [SACM96] STOKES, Michael ; ANDERSON, Matthew ; CHANDRASEKAR, Srinivasan ; MOTTA, Ricardo: A Standard Default Color Space for the Internet - sRGB. In: *Microsoft and Hewlett-Packard Joint Report* (1996)
- [SCD<sup>+</sup>06] SEITZ, S.M. ; CURLESS, B. ; DIEBEL, J. ; SCHARSTEIN, D. ; SZELISKI, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Computer Vision and Pattern Recognition*,

- 2006 IEEE Computer Society Conference on Bd. 1 IEEE, 2006, S. 519–528
- [SCK07] SHUM, Heung-Yeung ; CHAN, Shing-Chow ; KANG, Sing B.: *Image-Based Rendering*. Springer, 2007
- [SGSS08] SNAVELY, Noah ; GARG, Rahul ; SEITZ, Steven M. ; SZELISKI, Richard: Finding paths through the world's photos. In: *ACM SIGGRAPH 2008 papers*. New York, NY, USA : ACM, 2008 (SIGGRAPH '08). – ISBN 978–1–4503–0112–1, 15:1–15:11
- [SH02] SHI, Miaohong ; HEALEY, Glenn: Using reflectance models for color scanner calibration. In: *J. Opt. Soc. Am. A* 19 (2002), Apr, Nr. 4, 645–656. <http://dx.doi.org/10.1364/JOSAA.19.000645>. – DOI 10.1364/JOSAA.19.000645
- [Sha03] SHARMA, Gaurav: *Digital Color Imaging: Handbook*. CRC Press, 2003 (Electrical Engineering and Applied Signal Processing Series). – ISBN 9780849309007
- [Sna] SNAVELEY, Noah: *Bundler: Structure from Motion (SfM) for Unordered Image Collections*. #<http://phototour.cs.washington.edu/bundler/#>
- [SSI99] SATO, I. ; SATO, Y. ; IKEUCHI, K.: Acquiring a radiance distribution to superimpose virtual objects onto a real scene. In: *Visualization and Computer Graphics, IEEE Transactions on* 5 (1999), Nr. 1, S. 1–12
- [SSS06] SNAVELY, Noah ; SEITZ, Steven M. ; SZELISKI, Richard: Photo tourism: exploring photo collections in 3D. In: *ACM SIGGRAPH 2006 Papers*. New York, NY, USA : ACM, 2006 (SIGGRAPH '06). – ISBN 1–59593–364–6, 835–846
- [SSS09] SINHA, S.N. ; STEEDLY, D. ; SZELISKI, R.: Piecewise planar stereo for image-based rendering. In: *Computer Vision, 2009 IEEE 12th International Conference on IEEE*, 2009, S. 1881–1888
- [ST97] SHARMA, G. ; TRUSSELL, H.J.: Digital color imaging. In: *Image Processing, IEEE Transactions on* 6 (1997), Nr. 7, S. 901–932

- [Tsa87] TSAI, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. In: *Robotics and Automation, IEEE Journal of* 3 (1987), Nr. 4, S. 323–344
- [Wil78] WILLIAMS, Lance: Casting curved shadows on curved surfaces. In: *Proceedings of the 5th annual conference on Computer graphics and interactive techniques*. New York, NY, USA : ACM, 1978 (SIGGRAPH '78), 270–274
- [WR04] WESTLAND, Stephen ; RIPAMONTI, Caterina: *Computational colour science using MATLAB*. Chichester, West Sussex, England, Hoboken, NJ : J. Wiley, 2004 <http://opac.inria.fr/record=b1122104>. – ISBN 0–470–84562–7
- [YDMH99] YU, Y. ; DEBEVEC, P. ; MALIK, J. ; HAWKINS, T.: Inverse global illumination: Recovering reflectance models of real scenes from photographs. In: *International Conference on Computer Graphics and Interactive Techniques: Proceedings of the 26th annual conference on Computer graphics and interactive techniques* Bd. 1999, 1999, S. 215–224
- [YYF<sup>+</sup>06] YAMAMOTO, Kenji ; YENDO, Tomohiro ; FUJII, Toshiaki ; TANIMOTO, Masayuki ; SUTER, David: Color correction for multi-camera system by using correspondences. In: *ACM SIGGRAPH 2006 Research posters*. New York, NY, USA : ACM, 2006 (SIGGRAPH '06). – ISBN 1–59593–364–6
- [ZZ11] ZINTH, Wolfgang ; ZINTH, Ursula: *Optik: Lichtstrahlen - Wellen - Photonen*. 3. Oldenbourg Verlag München, 2011



# Lebenslauf

## Persönliche Daten

Name	Konrad Kölzer, Master Of Science
Geburtsdatum/ -ort	16.03.1984 in Rudolstadt

## Schulische und berufliche Ausbildung

März 1994 – Juni 2002	Heinrich-Böll-Gymnasium Saalfeld, Abitur
Oktober 2002 – Juni 2003	Wehrdienst in Ellwangen, Tätigkeit im Stabsdienst
Oktober 2003 – August 2006	Bachelorstudium Angewandte Informatik an Fachhochschule Erfurt, Bachelor Of Science
Oktober 2006 – August 2008	Masterstudium Angewandte Informatik an Fachhochschule Erfurt, Master Of Science
Oktober 2008 – Mai 2009	Hilfswissenschaftlicher Mitarbeiter der Fachhochschule Erfurt
Juni 2009 – Juli 2011	Wissenschaftlicher Mitarbeiter der Fachhochschule Erfurt im Forschungsprojekt „PoP-EYE“
Seit 2009	Promotion zum Thema „Bildbasierte Beleuchtung in 3D-Photo-Collections“ unter Betreuung durch Prof. Dr. Stefan Müller
Juli 2011 – Dezember 2012	Wissenschaftlicher Mitarbeiter der Hochschule Fulda im Forschungsprojekt „PoP-EYE“ und „kARbon“