

# **Aufbau und Test katadioptrischer Systeme zur Rekonstruktion von 3D**

## **Diplomarbeit**

### **zur Erlangung des Grades eines/r Diplom-Informatikers / Diplom-Informatikerin im Studiengang Computervisualistik**

vorgelegt von

Leif Baierl

Betreuer: Prof. Dr. Dietrich Paulus, Institut für Computervisualistik, Fachbereich Informatik  
Erstgutachter: Prof. Dr. Dietrich Paulus, Institut für Computervisualistik, Fachbereich Informatik  
Zweitgutachter: Dipl.-Inf. Peter Decker, Institut für Computervisualistik, Fachbereich Informatik

Koblenz, im September 2008



## Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien der Arbeitsgruppe für Studien- und Diplomarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja  nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja  nein

Koblenz, den .....

Unterschrift

» *'You don't know how to manage Looking-glass cakes,' the Unicorn remarked. 'Hand it round first, and cut it afterwards.'*«

*Lewis Carroll - Through the Looking-Glass*

## **Danksagung**

Ich bedanke mich bei allen, die mir während der Diplomarbeit und über das ganze Studium hinweg geholfen haben. Mein ganz besonderer Dank gilt Professor Dietrich Paulus, der mir dieses interessante Projekt überhaupt erst ermöglicht hat. Auch großen Dank geht an Detlev Droege und Peter Decker, die mich bei der Beschaffung der Komponenten tatkräftig unterstützt haben. Mein Dank gilt auch dem Labor für Mikrozerspanung der Universität Bremen, die für die Herstellung des konischen Spiegels verantwortlich zeichneten. Vielen Dank an Bine, die mir während der Diplomarbeit, aber auch über das ganze Studium hinweg immer mit Rat und Tat zur Seite gestanden hat, und mir durch das hervorragende Korrekturlesen viele wertvolle Hinweise und Denkanstöße gegeben hat. Der größte Dank geht aber natürlich an meine Eltern, ohne die das alles nicht möglich gewesen wäre.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>11</b>
1.1	Motivation . . . . .	11
1.2	Zielsetzung . . . . .	14
1.3	Gliederung . . . . .	15
<b>2</b>	<b>Grundlagen</b>	<b>17</b>
2.1	Optik . . . . .	17
2.1.1	Brennweite . . . . .	17
2.1.2	Blende . . . . .	19
2.1.3	Fokus . . . . .	19
2.1.4	Bildwinkel . . . . .	20
2.1.5	Schärfentiefe . . . . .	21
2.1.6	Fisheye-Abbildungsfunktion . . . . .	22
2.2	Korrelationsverfahren . . . . .	25
2.2.1	Allgemeine Definition . . . . .	27
2.2.2	Nicht-Normalisierende Korrelationsverfahren . . . . .	27
2.2.3	Normalisierende Korrelationsverfahren . . . . .	28
2.3	Disparitätsschätzung . . . . .	30

<b>3</b>	<b>Katadioptrische Stereokamera</b>	<b>33</b>
3.1	Überblick . . . . .	33
3.2	Computational Camera . . . . .	34
3.3	Katadioptrische Stereokamera . . . . .	39
3.4	Katadioptrische Planspiegel-Stereosysteme . . . . .	40
3.4.1	Konfigurationen . . . . .	42
3.4.2	Eigenschaften . . . . .	43
3.4.3	Rektifizierte Systeme . . . . .	49
3.5	Radiale katadioptrische Stereosysteme . . . . .	53
3.5.1	Eigenschaften . . . . .	56
3.5.2	Zylindrischer Spiegel $\beta = 0$ . . . . .	65
3.5.3	Konischer Spiegel mit $\beta \neq 0$ . . . . .	67
3.6	Andere Formen . . . . .	71
<b>4</b>	<b>Bau und Implementierung</b>	<b>73</b>
4.1	Übersicht . . . . .	73
4.2	Realisierte System . . . . .	74
4.2.1	System mit Planspiegel . . . . .	74
4.2.2	System mit gekrümmtem Spiegel . . . . .	74
4.3	Rektifiziertes Planspiegel-Stereosystem . . . . .	76
4.3.1	Extraktion der Ansichten . . . . .	79
4.3.2	Stereo Matching . . . . .	80
4.4	Radiales katadioptrisches Stereosystem . . . . .	81
4.4.1	Extraktion der relevanten Bereiche . . . . .	85
4.4.2	Fisheye-Entzerrung . . . . .	86



<i>INHALTSVERZEICHNIS</i>	9
4.4.3 Aufteilen in verschiedene Ansichten . . . . .	89
4.4.4 Rektifizierung des nicht-planaren Spiegels . . . . .	92
4.4.5 Dense Stereo Matching . . . . .	97
4.4.6 Aufnahme . . . . .	99
<b>5 Experimente und Ergebnisse</b>	<b>101</b>
5.1 Erstellung synthetischer Testbilder mit POV-Ray . . . . .	101
5.2 Planspiegel-Stereosystem . . . . .	104
5.2.1 Unschärfeausgleich . . . . .	104
5.2.2 Genauigkeit . . . . .	106
5.2.3 Ergebnisse . . . . .	108
5.3 Radiales katadioptrisches Stereosystem . . . . .	110
5.3.1 Genauigkeit . . . . .	110
5.3.2 Ergebnisse . . . . .	112
<b>6 Zusammenfassung und Ausblick</b>	<b>115</b>
<b>A Softwaredesign</b>	<b>121</b>
A.1 Radiale katadioptrische Stereokamera . . . . .	121
<b>B Quelltexte</b>	<b>123</b>
B.1 PostScript . . . . .	123
B.1.1 Kalibrierungsmuster für das Gesamtsystem . . . . .	123
B.2 Equisolid-Angle-Abbildungsfunktion für POV-Ray . . . . .	124
<b>C Konstruktionszeichnung</b>	<b>129</b>



# Kapitel 1

## Einleitung

### 1.1 Motivation

Die Rekonstruktion von 3D aus Aufnahmen ist ein klassisches Problem des Rechnersehens. Die Bedeutung dieses Forschungsgebietes ist in den letzten Jahren aufgrund einer steigenden Anzahl von Anwendungen in der Bildverarbeitung und der Computergraphik stärker in den Fokus der Aufmerksamkeit gerückt. Hierfür werden Scanner benötigt, die in der Lage sind, die dreidimensionale Form und eventuell die Erscheinung (Textur) von Objekten oder der Umgebung zu erfassen.

Es existieren zwei Kategorien von 3D-Scannern: zum einen Systeme, die das zu untersuchende Objekt berühren und zum anderen kontaktfrei arbeitende Systeme. Die kontaktfreien Scanner können noch weiter in aktive und passive Systeme unterteilt werden. In dieser Diplomarbeit sollen zwei Nahbereichs-3D-Scanner entwickelt werden, die sich unter anderem für Gesichter eignen. Aufgrund der hohen Empfindlichkeit von Gesichtern, kommen Systeme für dieses Aufgabengebiet eher aus der Kategorie der kontaktfreien Scanner.

*3D-Laserscanner* sind die wohl bedeutendsten Vertreter der Kategorie der *aktiven Scanner*. Die zwei hierbei wichtigsten Ansätze sind das Prinzip der *time-of-flight* und das Triangulationsverfahren. Bei den *time-of-flight* Scannern wird die Zeit gemessen, die ein Laserstrahl braucht, um von dem Objekt reflektiert zu werden und wieder auf den Sensor zu

treffen. Diese *round-trip-time* des Laserstrahls in Kombination mit dem Wissen über die Lichtgeschwindigkeit ermöglicht es, die Entfernung zwischen Kamera und Objekt zu messen. Dieses Verfahren eignet sich besonders für große und weit entfernte Objekte, da es von der Präzision der Zeiterfassung abhängig ist und längere Intervalle leichter zu messen sind. Die Präzision dieser Systeme ist für Gesichter zumindest zurzeit noch zu gering, da bei einer Entfernung des Gesichts von 30 cm die *round-trip-time* des Laserstrahls unter 2 Nanosekunden<sup>1</sup> betragen würde. Bei angenommenen Höhenunterschieden eines Gesichts von zirka 10 cm und einer gewünschten Höhenauflösung von 100 Stufen müssen Zeitunterschiede von 3,3 Picosekunden<sup>2</sup> gemessen werden. Um solch kleine Zeitunterschiede zu messen, bedarf es hochspezialisierter und äußerst teurer Geräte.

Bei dem Triangulationsverfahren wird ein Laserpunkt oder eine Laserlinie auf das Objekt ausgesendet, welches wiederum von einer Kamera aufgenommen wird. Laser, Kamera und Objekt bilden ein Dreieck. Da der Abstand zwischen Laser und Kamera sowie der Winkel des Lasers bekannt sind und außerdem der Winkel der Kamera mit Hilfe des im Bild gefundenen Laserpunktes ermittelt werden kann, ist es möglich die Tiefe dieses Punktes zu errechnen. Wird der Laser durch einen Projektor ersetzt, der ein Muster auf das Objekt projiziert, so wird daraus ein *Structured-Light-Scanner*.

Die bedeutendsten Vertreter der *passiven 3D-Scanner* basieren auf dem Verfahren der Stereoskopie. Die Funktionsweise entspricht dem Binokularsehen in der Biologie, kann aber unter Umständen deutlich mehr als zwei Aufnahmen der Szene zu deren Rekonstruktion verwenden. Hauptsächlich zwei Vorgehensweisen werden genutzt, um an mehrere Perspektiven der Szene zu gelangen. Die Szene wird hierbei entweder mit mehreren Kameras aufgenommen oder die Kamera bewegt sich in einer statischen Szene. In diese Kategorie der passiven 3D-Scanner reihen sich auch die in der vorliegenden Diplomarbeit vorgestellten Systeme ein. Sie ersetzen die zusätzlichen Kameras durch Spiegel und vereinen dadurch die Vorteile der traditionellen Verfahren ohne deren Nachteile zu erben. Für eine detaillierte Analyse der Vor- und Nachteile stereoskopischer Verfahren siehe Abschnitt 3.1.

Die Biometrie hat in den letzten Jahren vor allem aufgrund der gestiegenen Angst vor

---

<sup>1</sup> 1 Nanosekunde =  $10^{-9}$  Sekunden

<sup>2</sup> 1 Picosekunde =  $10^{-12}$  Sekunden

Terroranschlägen stark an Bedeutung gewonnen. Aufgrund dessen ist der Bedarf an 3D-Gesichtsscannern in den letzten Jahren stark gestiegen. Vorhandene Lösungen sind meist teure Spezialanfertigungen, die nicht für jedermann zugänglich sind. Erschwingliche 3D-Scanner in Form von Stereokameras, die leicht zu bedienen sind, würden aber sicher auch einen Bedarf für 3D-Fotografie außerhalb der Biometrie wecken. Im Hinblick auf die in dieser Diplomarbeit vorgestellten relativ kostengünstigen Stereosysteme, wäre die 3D-Portrait-Fotografie sicherlich ein spannendes Betätigungsfeld für Hobby- und Profifotografen. Es bestünde gar die Möglichkeit, mit Hilfe von 3D-Druckern, aus den Aufnahmen 3D-Skulpturen „auszudrucken“.

Voraussetzung für die Rekonstruktion von Tiefe ist die Aufnahme einer Szene aus mehreren unterschiedlichen Perspektiven, um diese anschließend zu einem Bild mit Tiefeninformationen zu vereinen. Dies ähnelt dem biologischen Vorgang der *Fusion* beim stereoskopischen Sehen, wobei die Bilder beider Augen zu einem verschmolzen werden. Um diese Voraussetzung zu erfüllen, bestehen traditionelle Verfahren zur Gewinnung von Tiefeninformationen einer Szene meist aus einer sich bewegenden Kamera oder mehreren stationären Kameras. Der Vorteil von *Ein-Kamera-Systemen* liegt darin, dass eine Kamera leichter zu kalibrieren ist als mehrere Kameras aufeinander abzustimmen. Jedoch haben sie den Nachteil, dass die Szene statisch sein muss, da sich die Kamera zwischen den Aufnahmen bewegt, um mehreren Perspektiven der Szene zu erhalten. *Mehr-Kamera-Systeme* können zwar dynamische Szenen einfangen, müssen aber aufwendig aufeinander abgestimmt und synchronisiert werden. Außerdem stellen mehrere möglichst identische Kameras einen großen zusätzlichen Kostenfaktor dar.

Der hier vorgestellte Ansatz vereint die Vorteile beider Herangehensweisen, indem eine weitere Optik vor dem Objektiv der Kamera platziert wird. So ist es möglich, eine handelsübliche Kamera<sup>3</sup> mit der Fähigkeit auszustatten dreidimensionale Bilder aufzunehmen. Diese zusätzliche Optik besteht aus einem oder mehreren Spiegeln und ersetzt somit weitere Kameras, die ansonsten nötig wären, um die Szene aus mehreren Perspektiven aufnehmen zu können.

Durch die weitere Optik wird die Szene in einer ungewöhnlichen Weise kodiert auf dem

---

<sup>3</sup>Nahezu alle Klasse von Kameras kommen hier in Frage, von Kompaktkameras über Spiegelreflexkameras bis hin zu Videokameras

Sensor abgebildet, wodurch sich die so entstandenen Bilder in ihrer Rohform nicht direkt nutzen lassen, sondern erst nachbearbeitet werden müssen. Mit Methoden der Bildverarbeitung, die Informationen über das Gesamtsystem nutzen, lassen sich diese kodierten Aufnahmen wieder dekodieren. Im Fall von Stereosystemen kann so die Raumtiefe der Szene rekonstruiert werden. Kameras, deren Bilder erst durch Berechnungen „entwickelt“ werden müssen, werden *Computational Cameras* genannt.

Je nachdem welche Art von Spiegel sie einsetzen, lassen sich zwei Gruppen von katadioptrischen Stereosystemen unterscheiden. Die erste Gruppe von Systemen nutzt Planspiegel, also Spiegel deren reflektierende Oberfläche eben ist. Die zweite große Gruppe verwendet gekrümmte Spiegel, die je nach Einsatzgebiet die unterschiedlichsten Formen aufweisen können. Jeweils ein Stereosystem aus diesen zwei Gruppen soll in dieser Diplomarbeit näher untersucht werden. Der Fokus liegt auf dem System mit gekrümmtem beziehungsweise radialem Spiegel, es soll aber ebenfalls ein System mit einem Planspiegel aufgebaut werden. Bei diesem System handelt es sich um ein einfaches katadioptrisches Stereosystem, das direkt rektifizierte Ergebnisse liefert und deshalb nur minimaler Nachbearbeitung bedarf. Außerdem soll es mit möglichst geringen Kosten und einfachsten Mitteln realisierbar sein.

## 1.2 Zielsetzung

Ziel dieser Diplomarbeit ist es, katadioptrische Stereosysteme aufzubauen und zu testen. Hierfür gilt es, die Komponenten auszuwählen und zu beschaffen. Es soll ein System mit einem Planspiegel und eines mit einem gekrümmten Spiegel aufgebaut werden. Die Vorgehensweise in beiden Fällen unterscheidet sich jedoch deutlich voneinander. Das Planspiegel-System wird unter den Voraussetzungen angefertigt, dass es möglichst kostengünstig ist und bietet sich vor allem deshalb für eine solches Vorgehen an, da es mit handelsüblichen Komponenten realisierbar ist. Die Verwendung dieser Komponenten bringt allerdings Einschränkungen in der Bildqualität mit sich. Hinzu kommt noch die Tatsache, dass es sich um eine Computational Camera mit geringer Komplexität handelt, da es direkt rektifizierte Ergebnisse liefert und aufgrund dessen leicht zu implementieren ist. Diese Kombination von Eigenschaften macht es zu einem geeigneten Kandidaten für

eine solche Herangehensweise. Das radiale System wird hingegen unter der Prämisse maximaler Qualität gebaut. Der verwendete Spiegel ist eine Einzelanfertigung und zusätzlich wird ein zirkuläres Fisheye-Objektiv benötigt, was die Kosten für ein solches System relativ hoch werden lässt. Des Weiteren ist die Dekodierung der Rohaufnahmen deutlich komplexer, als bei den bereits rektifizierten Aufnahmen des Planspiegel-Systems.

Nach erfolgreichem Aufbau der beiden Gesamtsysteme sollen diese unter qualitativen Aspekten miteinander verglichen werden. Das radiale System hat durch seine theoretisch unendlich vielen zirkulär angeordneten virtuellen Viewpoints sicherlich das höhere Potenzial, muss aber im Vergleich zum Planspiegel-System die Verzerrungen des Fisheye-Objektiv rückgängig machen, die Ansichten rektifizieren und mit den Verzerrungen des gekrümmten Spiegels zurecht kommen. Die Schwäche des Planspiegel-Systems liegt zum einen in ihrer Simplität, da das Objekt nur aus einem weiteren virtuellen Viewpoint aufgenommen wird und zum anderen an der minderwertigen Qualität eines Haushaltsspiegels. Ein industrieller Planspiegel aus Duran 50 mit einem Durchmesser von 300mm würde jedoch bei mittlerer Flächenabweichung<sup>4</sup> zirka 3800 € kosten.

Als Testobjekt sollen Gesichter gescannt werden, da sie sich von ihrer Größe und Oberflächenbeschaffenheit gut für die geplanten Systeme eignen. Einen weiteren Grund stellt aber auch das hohe wissenschaftliche Interesse dar, das von dreidimensionalen Gesichtsdaten ausgeht. Es bleibt abzuwarten mit welchem System sich die besten Ergebnisse erzielen lassen.

## 1.3 Gliederung

Zunächst werden in Kapitel 2 die wichtigsten Grundlagen erläutert, danach in Kapitel 3 der Stand der Wissenschaft und das theoretische Fundament für katadioptrische Stereokameras gelegt. In Kapitel 4 wird der Bau und die Implementierung der beiden zu realisierenden Stereokameras beschrieben. Experimente mit diesen Kameras und deren Ergebnisse werden in Kapitel 5 erörtert. Diese und die gesamten Erkenntnisse dieser Diplomarbeit werden in Kapitel 6 zusammengefasst. Darüber hinaus wird in diesem Kapitel ein Aus-

---

<sup>4</sup>Maß für die Ebenheit der Spiegeloberfläche.

blick auf die Zukunft des relativ jungen Felds der *Computational Camera* im Allgemeinen und der katadioptrischen Stereokamera im Speziellen gegeben.



# Kapitel 2

## Grundlagen

Dieses Kapitel behandelt die Grundlagen von katadioptrischen Stereosystemen. Zuerst wird in Abschnitt 2.1 ein Überblick über wichtige Begriffe der Optik gegeben. In Abschnitt 2.2 folgt die Beschreibung von Korrelationsverfahren zur Erstellung von dichten Tiefenkarten. Abschließend wird in Abschnitt ?? auf mögliche Darstellungsformen eingegangen.

### 2.1 Optik

Im folgenden Abschnitt sollen die für diese Diplomarbeit relevanten Begriffe der Optik erklärt werden. Es kann hier nicht in aller Ausführlichkeit auf die verschiedenen Aspekte der Optik eingegangen werden, da dies über den Rahmen dieser Diplomarbeit hinausgehen würde. Für weiterführende Informationen über Optik sei deshalb auf [HZ05] und [Sch02] verwiesen.

#### 2.1.1 Brennweite

Die Brennweite  $f$  ist eine Abbildungseigenschaft eines optischen Systems. Sie stellt den Abstand zwischen dem Fokus  $F$  (vgl. Abschnitt 2.1.3) und dem Hauptpunkt  $H$  auf der

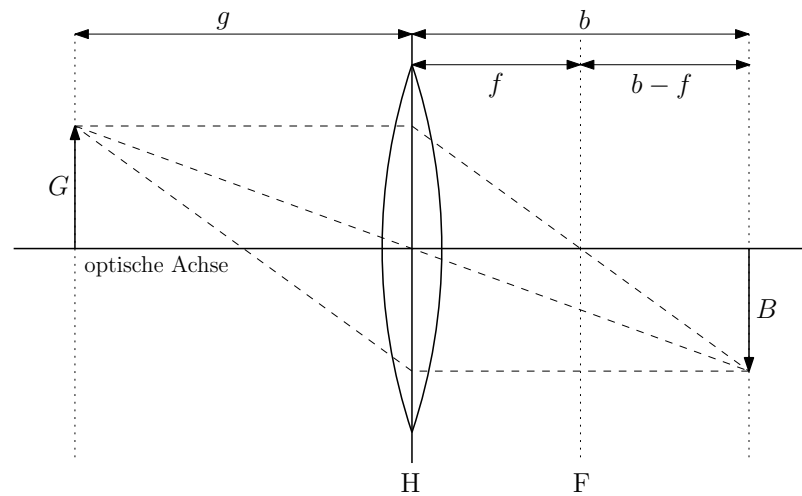


Bild 2.1: Linsengleichung

Linse (vgl. Abbildung 2.1) bei parallel einfallendem Licht dar. Der Zusammenhang zwischen der Gegenstandsgröße  $G$ , der Gegenstandsweite  $g$ , der Bildgröße  $B$ , der Bildweite  $b$  und der Brennweite  $f$  ergibt sich direkt aus der Geometrie (vgl. Abbildung 2.1):

$$\frac{G}{B} = \frac{g}{b} = \frac{f}{b-f}$$

Hieraus ergibt sich die Linsengleichung:

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g}$$

Nach Umformen der Linsengleichung ergibt sich die die Gleichung für die Brennweite:

$$f = \frac{g \cdot B}{G + B}$$

### 2.1.2 Blende

Die Blende ist eine Vorrichtung am Objektiv einer Kamera, die die einfallende Lichtmenge reguliert. Dies geschieht zumeist mechanisch mit Hilfe von Lamellen, die sich nach innen und außen verschieben lassen und somit die Öffnung kreisförmig verkleinern oder vergrößern können. Diese Art der Blende wird Lamellen- oder Irisblende genannt.

Die Blendenöffnung wird als Blendenzahl  $\kappa$  angegeben und stellt das Verhältnis zwischen Brennweite  $f$  und Durchmesser  $D$  der wirksamen Öffnung des Objektivs in Millimetern dar:

$$\kappa = \frac{f}{D}$$

Ist die Brennweite beispielsweise sechzehn Mal größer als der Durchmesser so ergibt sich eine Blendenzahl von  $\kappa = 16$ . Wird der Durchmesser der Objektivöffnung beziehungsweise der Blendenöffnung verdoppelt, so verringert sich dieses Verhältnis auf ein achtfaches und es ergibt sich eine Blendenzahl von  $\kappa = 8$ . Dies bedeutet, dass je kleiner die Blendenzahl ist, desto größer ist die Blendenöffnung. Die kleinste einstellbare Blendenzahl ist ein Maß für die Lichtstärke eines Objektivs, da sie die maximal einfallende Lichtmenge beschreibt.

Die Blendenöffnung ist auch für die Schärfentiefe verantwortlich. Hierbei gilt: Je kleiner die Blendenöffnung desto größer ist die Schärfentiefe. Näheres über diesen Zusammenhang kann in Abschnitt 2.1.5 gefunden werden.

### 2.1.3 Fokus

Das Fokussieren oder Scharfstellen eines Objektes hat zum Ziel, die Bildweite  $b$  der Brennweite  $f$  anzunähern. Beim Einstellen des Fokus  $F$  an einem Objektiv wird die Bildweite so verändert, dass dieses Ziel möglichst erreicht wird.

Das Fokussieren kann motorisiert oder wie bei dem in dieser Diplomarbeit verwendeten Sigma Fisheye-Objektivs ausschließlich manuell erfolgen. Für Systeme, die auf eine

Kalibrierung angewiesen sind, wäre eine Arretierung der Fokussierungseinstellung wünschenswert, damit diese nicht aus Versehen verstellt wird. Bei den hier vorgestellten Systemen spielt die Fokussierung aber keine so große Rolle für die Kalibrierung.

### 2.1.4 Bildwinkel

Die Auswahl des Objektivs wird normalerweise auf Basis der Gegenstandsgröße und Gegenstandsweite der aufzunehmenden Objekte getroffen. Dies ist bei katadioptrischen Stereosystemen nicht anders, allerdings muss darüber hinaus darauf geachtet werden, dass die Szene mehrfach auf dem Sensor abgebildet wird, der effektive Bildwinkel sich also verringert. Wie der effektive Bildwinkel berechnet wird, ist von System zu System unterschiedlich. Für eine nähere Untersuchung des Bildwinkels in Planspiegel-Stereosystemen siehe Abschnitt 3.4.2 und in radialen katadioptrischen Stereosystemen siehe Abschnitt 3.5.1.

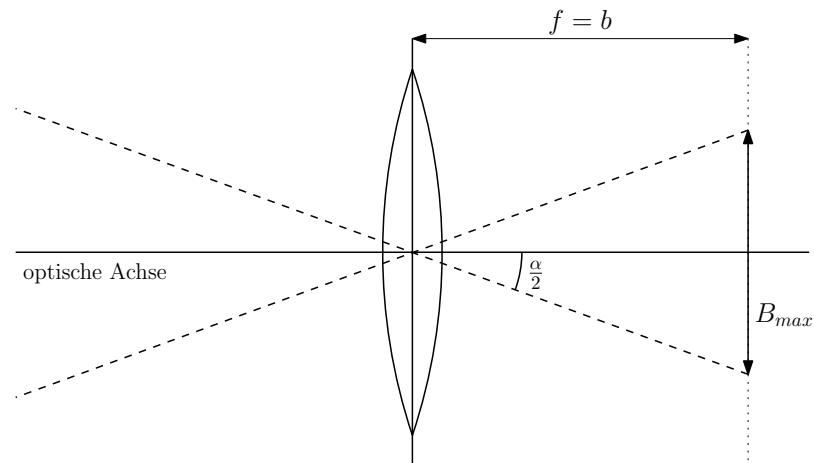


Bild 2.2: Bildwinkel

Der Winkel  $\frac{\alpha}{2}$  zwischen optischer Achse und dem äußersten Lichtstrahl, der noch auf den Sensor trifft, lässt sich leicht trigonometrisch berechnen:

$$\tan\left(\frac{\alpha}{2}\right) = \frac{B_{max}}{f}$$

Nach  $\alpha$  aufgelöst ergibt sich:

$$\alpha = 2 \cdot \arctan\left(\frac{B_{max}}{2f}\right)$$

Da das Format der Bildebene im Gegensatz zur Projektion nicht kreisförmig ist, muss der Bildwinkel für die Breite, Höhe oder Diagonale des Sensors getrennt bestimmt werden. Bei einer Kamera mit einem Sensor der Dimension  $22,2 \times 14,8$  mm sowie einem Objektiv mit einer Brennweite von 50 mm, ergibt sich ein horizontaler Bildwinkel von  $25,03^\circ$  und ein vertikaler Bildwinkel von  $16,84^\circ$ .

Ein Objektiv mit einem Bildwinkel von  $\alpha = 50^\circ$  gilt als Standardobjektiv. Objektive mit einem Bildwinkel größer als  $50^\circ$  weisen auf ein Weitwinkelobjektiv und eines mit einem kleineren Bildwinkel auf ein Teleobjektiv hin. Das in dieser Diplomarbeit verwendete Sigma Fisheye-Objektiv hat einen Bildwinkel von über  $180^\circ$ , was nur durch eine besondere Abbildungsfunktion erreicht werden kann (vgl. hierzu Abschnitt 2.1.6).

### 2.1.5 Schärfentiefe

Das Objektiv bildet lediglich eine zur Bildebene parallel stehende Ebene scharf ab. Objekte mit räumlicher Ausdehnung abzubilden ist daher nur möglich, wenn eine gewisse Unschärfe in Kauf genommen wird. Ein Weltpunkt wird bei dieser Form der Unschärfe nicht als Punkt, sondern als Kreisfläche auf dem Sensor abgebildet. Diese Kreisfläche wird Unschärfekreis genannt (vgl. Abbildung 2.3).

Um die Schärfentiefe zu bestimmen, muss zuerst ein Maximalwert für diesen Unschärfekreis festgelegt werden, der aussagt bis zu welcher Schwelle ein Punkt als scharf abgebildet gilt. Ist die Blende weit geöffnet, dann treffen die Lichtstrahlen in einem stumpfen Winkel aufeinander, und der Bereich der scharf abgebildet wird ist relativ klein (vgl. Abbildung 2.4 oben). Diese Konfiguration hat deshalb eine geringe Schärfentiefe. Wird die Blende weiter geschlossen, so treffen die Lichtstrahlen in einem spitzeren Winkel aufeinander und der Bereich des Scharfsehens verlängert sich, wodurch eine größere Schärfentiefe erzielt wird (vgl. Abbildung 2.4 unten). Bei Kameras (z. B. Kompaktkameras), die ein

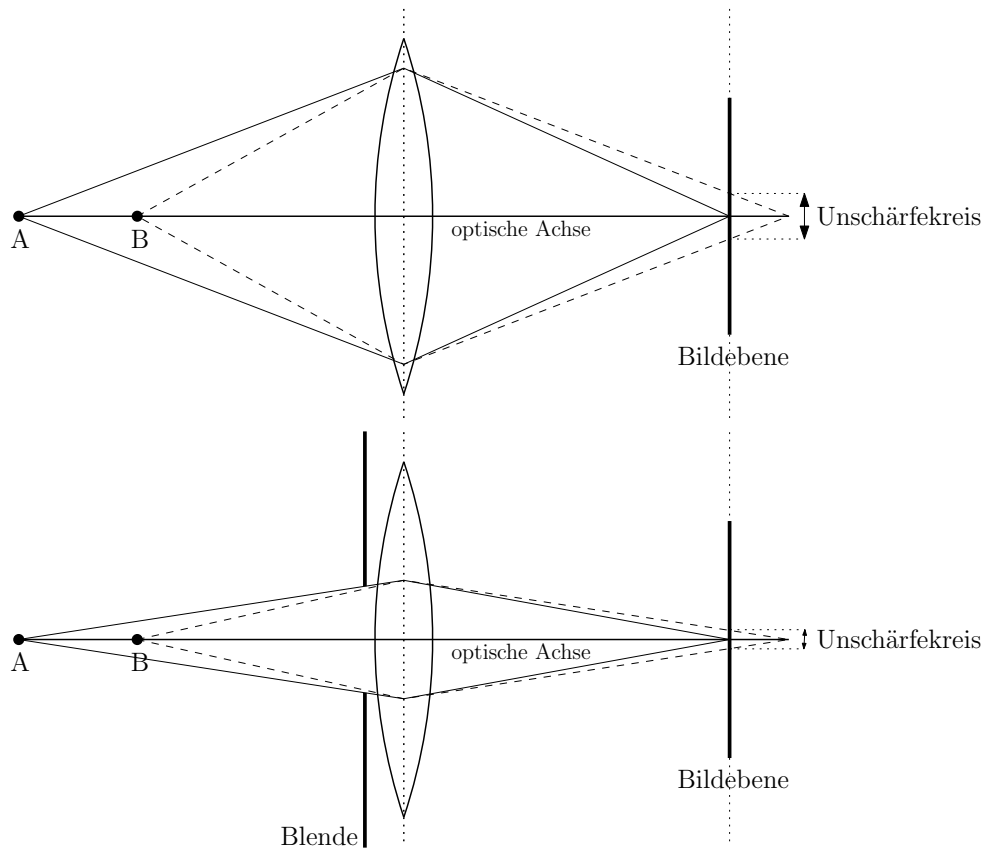


Bild 2.3: Schärfentiefe

kleine maximale Blendenöffnung aufweisen, ist es deshalb nur schwer möglich, die in der Fotografie häufig als Stilmittel eingesetzte geringe Schärfentiefe (auch Bokeh<sup>1</sup> genannt) zu erzielen.

### 2.1.6 Fisheye-Abbildungsfunktion

In diesem Abschnitt soll auf die unterschiedlichen Abbildungsfunktionen von Fisheye-Objektiven eingegangen werden. Bei dem in dieser Diplomarbeit realisierten radialen kadioptrischen Stereosystem, kommt ein Sigma Fisheye-Objektiv zum Einsatz. Die Abbil-

<sup>1</sup>vom japanischen boke, „unscharf“ „verschommen“

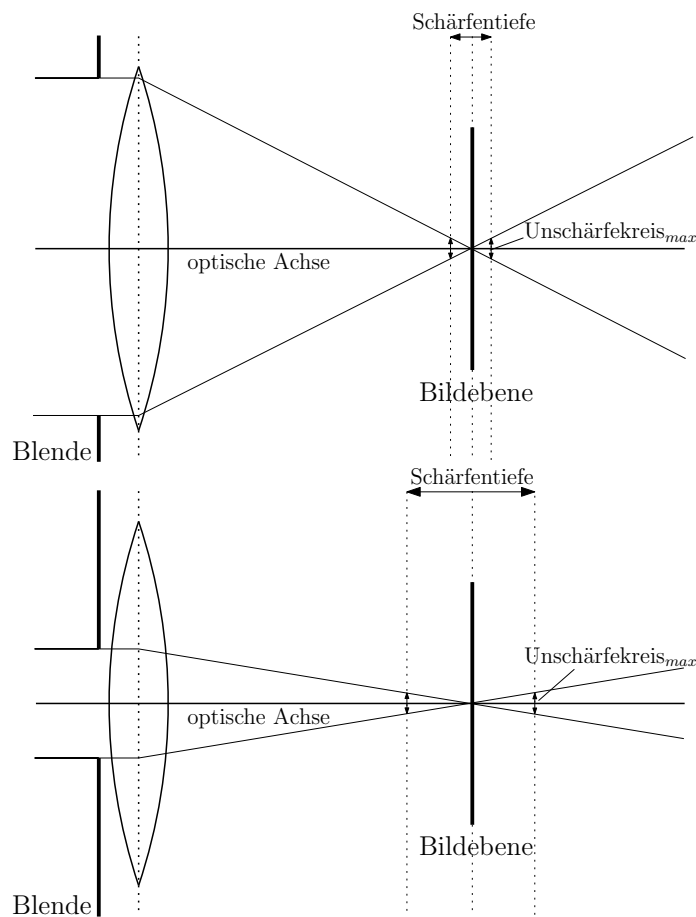


Bild 2.4: Schärfentiefe

dungsfunktion muss die Szene verzerren, damit der große Bildwinkel von etwa  $180^\circ$  auf dem Sensor abgebildet werden kann.

Zunächst lassen sich zwei Arten von Fisheye-Objektiven unterscheiden. Zum einen existieren so genannte *Full-Frame-Fisheyes*, deren Bildkreisdurchmesser mindestens so groß ist wie die Diagonale des Bildaufnahmeformates, es also keine ungenutzten Bereiche auf dem Sensor gibt. Zum anderen gibt es so genannte *zirkuläre Fisheyes*, deren Bildkreisdurchmesser höchstens so groß ist wie die kürzeste Kante des Bildaufnahmeformates.

Diese Unterscheidung gilt aber nur in Bezug auf das Kleinbildformat mit den Dimensionen  $36 \times 24$  mm. Die meisten digitalen Spiegelreflexkameras haben deutlich kleiner di-

mensionierte Sensoren und können deshalb nicht den vollen Kreis der Fisheye-Projektion erfassen. Nur Kameras der Profiklasse mit Vollformatsensoren ist es möglich die vollständige Projektion eines zirkulären Fisheye-Objektivs aufzufangen.

*Full-Frame-Fisheyes* weisen eine Brennweite von zirka 15 mm auf und erreichen ihren vollen Bildwinkel von  $180^\circ$  nur in der Diagonalen. *Zirkuläre Fisheyes* nutzen nicht den gesamten Sensor, sie projizieren ein kreisrundes Bild innerhalb des rechteckigen Aufnahmeformates und haben eine Brennweite von ungefähr 8 mm. Seit kurzem werden auch zirkuläre Objektive mit einer Brennweite von 4,5 mm angeboten (wie z. B. das Sigma 4,5mm F2,8 EX DC HSM), mit denen es auch unter Verwendung von kleineren Sensoren möglich ist den gesamten Bildkreisdurchmesser einzufangen.

Um in der Lage zu sein, einen Bildwinkel von über  $180^\circ$  auf dem Sensor abbilden zu können, wird die Bildlage von Objekten entsprechend ihres Winkels zur optischen Achse, dem Feldwinkel  $\omega$ , auf einen bestimmten Radius  $r$  abgebildet. Dies wird mit Hilfe spezieller Abbildungsfunktionen realisiert, aufgrund derer Fisheye-Objektive weiterhin unterschieden werden können. Die wichtigsten Vertreter sind *Equisolid-Angle-* und die *äquidistante Abbildungsfunktion* und sollen im Folgenden näher beschrieben werden.

### **Equisolid-Angle-Abbildungsfunktion**

Die am weitesten verbreitete Abbildungsfunktion ist die flächentreue Abbildung *Equisolid Angle*. Sie bildet die Fläche eines Objekts proportional zum Raumwinkel ab, das heißt die Fläche von Objekten wird relativ zum Raumwinkel korrekt abgebildet. So kann beispielsweise aus einer Aufnahme des Himmels der Bewölkungsgrad korrekt abgelesen werden. Dies bedeutet aber auch, dass die Entfernung von Objekten richtig eingeschätzt werden kann und es sich somit für den Einsatz in einer Stereokamera eignet.

Die *Equisolid-Angle-*Abbildungsfunktion lautet:

$$r = 2 \cdot f \cdot \sin\left(\frac{\omega}{2}\right) \quad (2.1)$$

Das in dieser Diplomarbeit verwendete zirkuläre Fisheye-Objektiv mit einer Brennweite



von 8 mm<sup>2</sup> weist eine *Equisolid-Angle*-Abbildungsfunktion auf. Der Bildkreisdurchmesser für 180° beträgt nach Formel 2.1 22,6 mm und passt somit bei der zu diesem Objektiv verwendeten Nikon D60 mit einem 23,6 × 15,8 mm großen Sensor horizontal vollständig auf die Projektionsfläche. In vertikaler Richtung reicht der Chip hingegen nur für einen Bildwinkel von 118.4° aus. Der Bildkreis wird in vertikaler Richtung abgeschnitten und kann nicht vollständig dargestellt werden. In Kombination mit dieser Kamera gehört es im engeren Sinne folglich nicht mehr in die Kategorie zirkulärer Fisheye-Objektive.

### Äquidistante Abbildungsfunktion

Einige ältere Objektive nutzen noch die äquidistante Abbildungsfunktion, um die Szene auf dem Sensor abzubilden. Es wurde hier mit mehr oder weniger Erfolg versucht eine lineare Abbildungsfunktion zu erschaffen. Sie eignet sich besonders für Winkelmessungen. Da Objekte am Bildrand nicht flächentreu abgebildet werden, eignet sich diese Funktion nicht zur Entfernungsmessung und kann nur bedingt in Stereokameras eingesetzt werden. Ist der Winkel  $\omega$  im Bogenmaß angegeben so lautet die äquidistante Abbildungsfunktion:

$$r = f \cdot \omega$$

Würde das Objektiv im obigen Beispiel eine äquidistante Abbildungsfunktion nutzen, so würde hier der Abbildungskreis sogar in horizontaler Richtung abgeschnitten werden, da 25,1 mm für einen Bildkreisdurchmesser von 180° benötigt werden.

## 2.2 Korrelationsverfahren

In diesem Abschnitt sollen verschiedene Verfahren zur Lösung des Korrespondenzproblems (*Stereo Matching*) im Hinblick auf die Erstellung von dichten Tiefenkarten vorge-

---

<sup>2</sup>Das Objektiv hat tatsächlich eine Brennweite von 7,8 mm und bietet einen Bildwinkel von mehr als 180°.

stellt werden. Hierfür müssen Korrespondenzen für alle Bildpunkte<sup>3</sup> rektifizierter Stereobildpaare gefunden werden.

Allen Verfahren ist die Annahme gemein, dass die Aufnahmen der Szene ähnlich sind. Punkte sollen einander zugeordnet (*gematcht*) werden, die nach einem bestimmten Ähnlichkeitsmaß als korrelierend angesehen werden. Diese Ähnlichkeitsmaße verwenden hierzu unterschiedliche Kriterien, wie beispielsweise die Helligkeit, Textur, Farbe oder Kantensorientierung. Aus Geschwindigkeitsgründen werden meist Ähnlichkeitsmaße genutzt, die die Ähnlichkeit anhand der Intensitätswerte einer Region schätzen, also auf Grauwertbildern arbeiten. Weiterhin können Korrelationsverfahren nach normalisierenden und nicht-normalisierenden Verfahren unterschieden werden. Normalisierende Korrelationsmaße sind im Gegensatz zu nicht-normalisierenden invariant gegenüber konstant additiven und/oder multiplikativen Helligkeitsunterschieden zwischen den Bildern  $I_1$  und  $I_2$ . Für unterschiedliche Intensitätswerte können verschiedene Kameraeinstellungen wie Brennweite, Linsenunschärfe oder die unterschiedlich reagierende *Automatic Gain Control* (AGC) verantwortlich sein.

Da katadioptrische Stereosysteme zumeist nur eine Kamera nutzen<sup>4</sup> treten solche Probleme nicht auf. Einzig Systeme, die die Szene auch direkt aufnehmen, leiden bei nicht idealen Spiegeln, vor allem wenn es sich um einen minderwertigen Spiegel handelt, unter ähnlichen Problemen. Da eine Ansicht direkt und eine andere durch die Reflexion im Spiegel entsteht, können die Ansichten unterschiedliche Helligkeiten aufweisen. Bei Systemen, die keine direkte Ansicht verwenden, werden folgerichtig keine rechenintensiven normalisierenden Verfahren benötigt, die diese Unterschiede auszugleichen versuchen.

Im Folgenden wird zuerst eine allgemeine Definition für Korrelationsmaße gegeben und anschließend die wichtigsten Korrelationsverfahren in ihren normalisierenden und nicht-normalisierenden Formen beschrieben. Eine ausführliche Untersuchung aktueller *Stereo-Matching*-Verfahren findet sich in [SS02].

---

<sup>3</sup>Irrelevante Bereiche können ausgelassen werden.

<sup>4</sup>Es wären zwar Systeme mit mehreren Kameras denkbar, diese würden sich allerdings des Vorteils der identischen Kameraparameter berauben.

### 2.2.1 Allgemeine Definition

Da ein Vergleich zwischen einzelnen Bildpunkten keine schlüssige Aussage über eine Korrelation zwischen ihnen liefern kann, werden Bildausschnitte, so genannte Fenster, auf ihre Korrelation untersucht. Bei diesen Fenstern handelt es sich um quadratische Graustufenbilder mit ungerader Kantenlänge ( $n = 2k + 1, k \geq 1$ ). Eine ungerade Kantenlänge hat den gleichen Vorteil den sich auch Filter zu nutze machen, denn der zu untersuchende Pixel  $(k, k)$  befindet sich somit im Zentrum des Bildausschnitts. Das Fenster  $F_1$  soll im Folgenden einen Bildausschnitt des Bildes  $I_1$  und  $F_2$  analog des Bildes  $I_2$  sein.

Je nach Verfahren wird entweder ein zu minimierendes Fehlermaß, oder ein zu maximierendes Ähnlichkeitsmaß berechnet. Die Grundformel für das Korrelationsmaß zweier Fenster  $F_1$  und  $F_2$  lautet:

$$c(F_1, F_2) = \sum_u \sum_v f(F_1(u, v), F_2(u, v)), \quad \text{mit } 0 \leq u < n \wedge 0 \leq v < n$$

In den beiden folgenden Abschnitten werden zuerst die drei bekanntesten nicht-normalisierenden Korrelationsverfahren, die Summe der quadrierten Differenzen, die Summe der absoluten Differenzen und die Kreuzkorrelation, und danach ihre normalisierenden Varianten vorgestellt.

### 2.2.2 Nicht-Normalisierende Korrelationsverfahren

Dieses häufig eingesetzte Fehlermaß nutzt die Summe der quadrierten Differenzen (SSD<sup>5</sup>) mit der Funktion  $f(x, y) = (x - y)^2$ :

$$\text{SSD}(F_1, F_2) = \sum_u \sum_v (F_1(u, v) - F_2(u, v))^2$$

Dieses Verfahren berechnet einen Fehler und muss somit minimiert werden. Bei identischen Bildern ergibt sich entsprechend der Wert Null. Es liefert bei gleichen oder sehr ähn-

---

<sup>5</sup>SSD steht für *Sum of Squared Differences*

lichen Helligkeitsbedingungen gute Ergebnisse, reagiert aufgrund der Quadrierung aber sehr stark auf Ausreißer. Das Maß der Summe der absoluten Differenzen (SAD) ist gegenüber solcher Ausreißer deutlich robuster, da hier die Beträge der Differenzen genutzt werden,  $f(x, y) = |x - y|$ :

$$\text{SAD}(F_1, F_2) = \sum_u \sum_v |F_1(u, v) - F_2(u, v)|$$

Die Summe absoluter Differenzen liefert ebenfalls ein Fehlermaß und muss minimiert werden, wohingegen bei der Kreuzkorrelation ein zu maximierendes Ähnlichkeitsmaß entsteht. Als Funktion nutzt sie  $f(x, y) = x \cdot y$ :

$$\text{CC}(F_1, F_2) = \sum_u \sum_v F_1(u, v) \cdot F_2(u, v)$$

Dieses Verfahren wird in der Praxis nur selten in ihrer nicht-normalisierenden Variante genutzt, da es sich relativ effizient in seiner normalisierenden Form implementieren lässt (vgl. Abschnitt 2.2.3).

### 2.2.3 Normalisierende Korrelationsverfahren

Die in diesem Abschnitt vorgestellten normalisierenden Varianten der drei Korrelationsverfahren sind invariant gegenüber konstanten Helligkeitsunterschieden der Form  $F_1(u, v) = \alpha I_2(u, v) + \beta$ .

Um die Verfahren invariant gegenüber additiven Helligkeitsunterschieden zu machen, wird das arithmetische Mittel  $\bar{F}$  eines Grauwertbildes benötigt:

$$\bar{F} := \frac{1}{n^2} \sum_u \sum_v F(u, v)$$

Die additiv normalisierten Fenster  $F'_1$  und  $F'_2$  lassen sich durch Subtraktion des jeweiligen Mittelwerts berechnen:

$$F'_1 = F_1(u, v) - \bar{F}_1 \quad \text{und}$$

$$F'_2 = F_2(u, v) - \bar{F}_2$$

Um die Fenster gegenüber multiplikativen Helligkeitsunterschieden invariant zu machen, bedarf es der Frobeniusnorm  $\|F\|_F$  eines Grauwertbildes:

$$\|F\|_F := \sqrt{\sum_u \sum_v F^2(u, v)}$$

Mittels Division durch die jeweilige Frobeniusnorm entstehen die hinsichtlich multiplikativer Helligkeitsunterschiede normierten Grauwertbilder  $F''_1$  und  $F''_2$ :

$$F''_1 = \frac{F_1(u, v)}{\|F_1\|_F} \quad \text{und}$$

$$F''_2 = \frac{F_2(u, v)}{\|F_2\|_F}$$

Um eine Unempfindlichkeit gegenüber additiven *und* multiplikativen Helligkeitsunterschieden zu erreichen, wird die Frobeniusnorm der additiv normalisierten Grauwertbilder benötigt:

$$\|F'\|_F = \sqrt{\sum_u \sum_v (F(u, v) - \bar{F})^2}$$

Die gegenüber additiven und multiplikativen Helligkeitsunterschieden invarianten Formen der genannten Korrelationsverfahren leiten sich direkt hieraus ab.

Die *Zero Mean Normalized Sum of Squared Differences* (ZNSSD) berechnet sich wie folgt:

$$\text{ZNSSD}(F_1, F_2) = \sum_u \sum_v \left( \frac{F_1(u, v) - \bar{F}_1}{\|F'_1\|_F} - \frac{F_2(u, v) - \bar{F}_2}{\|F'_2\|_F} \right)^2$$

Die *Zero Mean Normalized Sum of Absolute Differences* (ZNSAD) lautet:

$$\text{ZNSAD}(F_1, F_2) = \sum_u \sum_v \left| \frac{F_1(u, v) - \bar{F}_1}{\|F'_1\|_F} - \frac{F_2(u, v) - \bar{F}_2}{\|F'_2\|_F} \right|$$

Bei der normalisierenden Kreuzkorrelation (ZNCC) lassen sich die Frobeniusnormen aus der Summe ziehen, was sich äußerst positiv auf die Laufzeit auswirkt und sie sich deshalb vor allem in ihrer normalisierenden Variante anbietet:

$$\begin{aligned} \text{ZNCC}(F_1, F_2) &= \sum_u \sum_v \frac{F_1(u, v) - \bar{F}_1}{\|F'_1\|_F} \cdot \frac{F_2(u, v) - \bar{F}_2}{\|F'_2\|_F} \\ &= \frac{\sum_u \sum_v (F_1(u, v) - \bar{F}_1) \cdot (F_2(u, v) - \bar{F}_2)}{\|F'_1\|_F \cdot \|F'_2\|_F} \end{aligned}$$

## 2.3 Disparitätsschätzung

In diesem Abschnitt soll die Schätzung der Disparität und die Erstellung einer dichten Tiefenkarte beschrieben werden. Je heller ein Bildpunkt in einer solchen Tiefenkarte ist, desto näher befindet er sich an der Kamera. Hierbei wird die Disparität  $d$  genannte Differenz zwischen den Positionen korrespondierender Bildpunkte in einem rektifizierten Stereobildpaar unter Zuhilfenahme der Stereobasis dazu genutzt, die Tiefe dieses Punktes zu rekonstruieren. Die Punkte  $p_l$  und  $p_r$  in Abbildung 2.5 sind korrespondierende Bildpunkte, die beide den Weltpunkt  $p_w$  abbilden. Der Punkt  $p'_r$  befindet sich an der zu  $p_r$  korrespondierenden Position auf der linken Bildebene, wobei die linke Bildebene in der Abbildung als Referenz dient.

Die Disparität  $d$  zwischen  $p_x^l$  und  $p_x^r$  und ist wie folgt definiert:

$$d = |p_x^l - p_x^r|$$

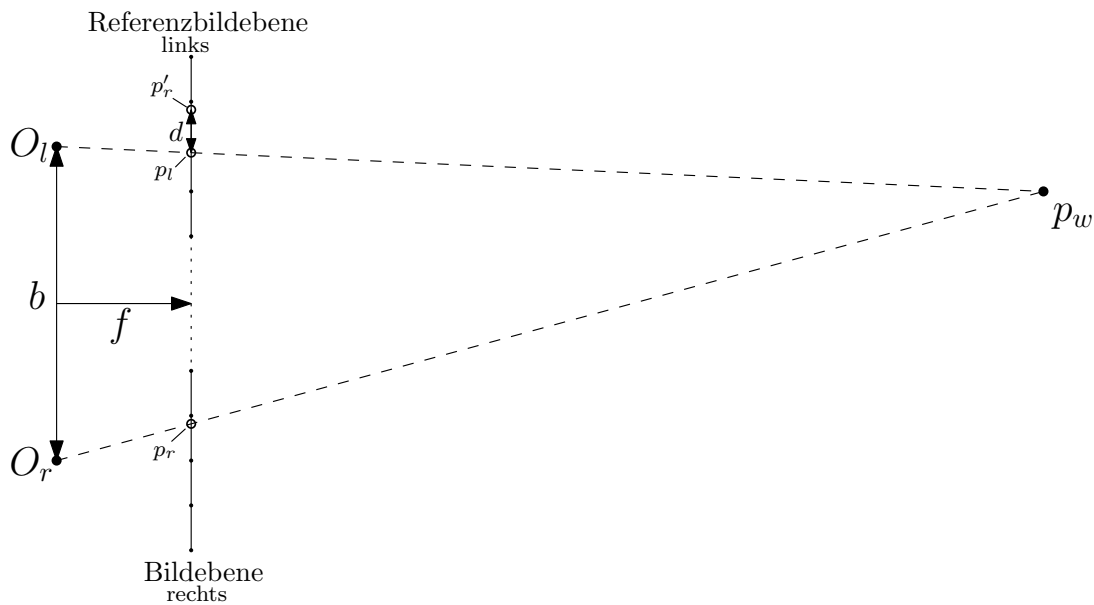


Bild 2.5: Bildwinkel

Wenn die Stereobasis  $b$  bekannt ist, kann mit Hilfe des Strahlensatzes ein Zusammenhang zwischen der Disparität und der Entfernung entlang der  $z$ -Achse des Weltpunktes  $p_w$  hergestellt werden:

$$p_z^w = f \cdot \frac{b}{d} \quad (2.2)$$

Für eine dichte Tiefenkarte wird im Gegenzug auch ein dichtes Stereo Matching benötigt. Das heißt, dass für jeden Punkt im Referenzbild oder zumindest für jeden Punkt des zu messenden Objektes eine Punktkorrespondenz benötigt wird. In einer solchen Tiefenkarte wird an jedem Punkt die dort gemessene Tiefe als Intensität eingetragen.





# Kapitel 3

## Katadioptrische Stereokamera

### 3.1 Überblick

Viele Anwendungen in der Bildverarbeitung und der Computergraphik, wie beispielsweise die Gewinnung von Tiefe, benötigen mehrere Aufnahmen der gleichen Szene. Die herkömmliche Herangehensweise ist, eine Kamera entweder innerhalb einer möglichst statischen Szene zu bewegen, oder die Szene gleichzeitig mit mehreren Kameras aus unterschiedlichen Positionen aufzunehmen. Ein Ein-Kamera-Stereosystem nach dem Prinzip der „Struktur aus Bewegung“ zu verwenden hat mehrere Vorteile gegenüber Mehr-Kamera-Systemen, da unter anderem das spektrale Verhalten, die Gain (Verstärkung) und das Rauschenverhalten in den Stereoaufnahmen identisch sind. Die Aufnahmen müssen einerseits nicht synchronisiert werden, andererseits müssen jedoch die Kamerabewegungen präzise gemessen werden, was nicht immer einfach ist. Der größte Nachteil dieses Verfahrens ist aber, dass die Szene statisch sein muss. Mehr-Kamera-Systeme haben diese Nachteile nicht, müssen aber synchronisiert, kalibriert und aufeinander abgestimmt werden. Weiterhin stellen mehrere Kameras, die für eine dichte Abtastung der Szene benötigt werden auch einen nicht zu geringen Kostenfaktor dar.

Das folgende Kapitel beschreibt die Grundlagen von katadioptrischen Stereokameras. Zuerst wird in Abschnitt 3.2 ein allgemeiner Überblick über *Computational Cameras* gegeben. Anschließend folgen Einführungen in katadioptrische Stereokameras (Abschnitt

3.3), katadioptrische Planspiegel-Stereosysteme (3.4), radiale katadioptrische Stereosysteme (3.5) sowie ein Überblick über andere Formen von katadioptrischen Systemen (3.6).

## 3.2 Computational Camera

In diesem Abschnitt wird ein allgemeiner Überblick über *Computational Cameras* gegeben, da es sich bei den in dieser Diplomarbeit vorgestellten Stereokameras um eine Unterkategorie dieser handelt. Die Aufnahmen jener Stereokameras liegen immer in einer noch zu bearbeitenden Rohform vor, in der die Informationen der Szene kodiert sind. Diese gilt es durch Methoden der Bildverarbeitung wieder zu dekodieren und weiter zu verarbeiten. Somit erfüllen diese Stereokameras die Bedingungen einer *Computational Camera*, die im Folgenden näher erläutert werden sollen.

Die Fototechnik hat in den letzten hundert Jahren eine rasante Entwicklung erlebt, die seit der Einführung von Digitalkameras erneut einen großen Innovationsschub erfahren hat. Das fast allen Kameras zugrundeliegende Modell, der Camera Obscura (lat. *camera* „Gewölbe“; *obscura* „dunkel“), ist aber bis heute unverändert geblieben. Dieses traditionelle Modell besteht aus einem Detektor (ein Sensor oder Film) und einer Linse beziehungsweise einem Linsensystem, dessen Aufgabe es ist, die durch die Kameraöffnung (Pinhole) fallenden Lichtstrahlen einzufangen und auf den Detektor zu fokussieren. Eine solche Kamera fängt alle Lichtstrahlen der Szene ein, die sich in ihrem Sichtfeld befinden. Das spätere Bild entsteht somit direkt auf dem Detektor und kann praktisch unverändert verwendet werden.

Das Modell der *Computational Camera* ermöglicht es, die Szene in einer radikal anderen Weise abzutasten und dadurch neue visuelle Informationen zu erhalten. Es entstehen unter Umständen Bilder, die in ihrer Rohform nicht direkt nutzbar sind, da sie die Informationen in kodierter Form enthalten, die entsprechend dekodiert neue und nützliche visuelle Informationen für Mensch oder Maschine liefern können.

Eine konventionelle Kamera<sup>1</sup> bildet eine Szene immer in einer bestimmten Weise ab und unterliegt dadurch modellbedingt einigen Restriktionen. Durch die Verschmelzung von

---

<sup>1</sup>Im Folgenden ist hiermit immer eine Kamera gemeint, die nur Linsen und keine Spiegel einsetzt.

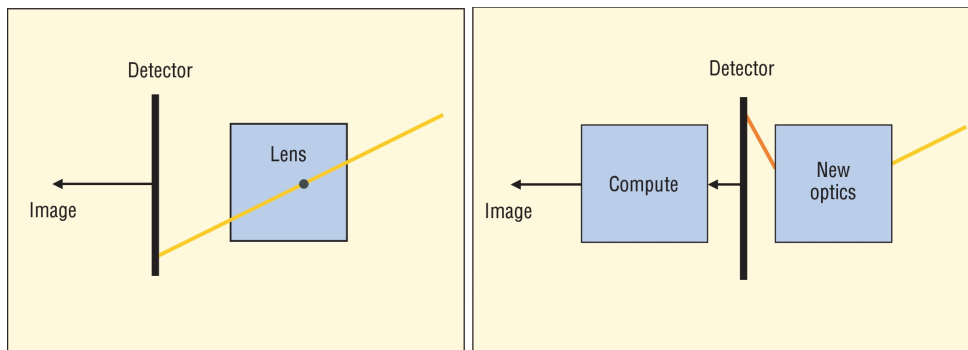


Bild 3.1: links: Camera Obscura, rechts: *Computational Camera* (aus [Nay06])

Kamera und Computer werden diese Restriktionen aufgehoben. Dies wird erreicht, indem die Optik einer *Computational Camera* die Szene in einer unkonventionellen Art und Weise visuell kodiert. Das bedeutet, dass die Lichtstrahlen nicht direkt auf dem Sensor abgebildet werden müssen und somit zusätzlich die Möglichkeit besteht die Intensität und Farbe des Lichtstrahls in der Optik zu verändern.

Erste Ansätze einer *Computational Camera* können in der Erstellung von Panoramabildern aus Aufnahmen gefunden werden, die mit einem Fisheye-Objektiv aufgenommen wurden. Hierbei wird die Szene durch eine spezielle Optik (einem Fisheye-Objektiv), die es ermöglicht mehr als  $180^\circ$  der Szene auf dem Chip abzubilden, in einer bestimmten Abbildungsfunktion abgetastet (vgl. Abschnitt 2.1.6). Diese Aufnahmen werden entzerrt und zu einem Panorama zusammengefügt, was der Dekodierung und Aufbereitung der Informationen entspricht. Da diese Schritte noch wenig automatisiert sind und die Aufnahmen per Hand gemacht werden müssen, kann hierbei jedoch nur bedingt von einer *Computational Camera* gesprochen werden.

Als Beispiel für eine biologische *Computational Camera* könnte der menschliche Sehapparat dienen, welcher eine Scharfsehzone von nur etwa  $1,5^\circ$  aufweist. Das Gehirn „berechnet“ daraus ein Bild, das in allen Teilen scharf zu sein scheint. Ebenso sind alle Ein-Chip-Digitalkameras im weiteren Sinne *Computational Cameras*, da sie eine *Bayer-Matrix* verwenden, um die Farbinformationen der Szene zu ermitteln. Hierbei wird ein Farbfilter (der so genannte *Bayer-Filter*) vor dem Sensor angebracht, der auf 50% der Pixel den Grünanteil und auf jeweils 25% den Rot- beziehungsweise Blauanteil des einfallenden Lichts

durchlässt. Die Szene liegt also kodiert vor und muss anschließend noch algorithmisch aufbereitet werden.

Während sich die Auflösung der Digitalkameras seit ihrer Einführung rasant erhöht hat, blieb die Anzahl der Helligkeitsstufen hingegen praktisch unverändert. Die meisten Kameras können nur 256 Helligkeitsstufen (8 Bit) in jedem Farbkanal unterscheiden. Bei Szenen mit großen Helligkeitsunterschieden können deshalb entweder die hellen oder die dunklen Bereiche detailreich abgebildet werden, aber nicht beide. Bei der High Dynamic Range (HDR) Fotografie werden deshalb mehrere Aufnahmen mit unterschiedlichen Belichtungszeiten zu einer Aufnahme mit höherem Dynamikumfang kombiniert. Dieses Vorgehen hat aber mehrere Nachteile, da es sich um mehrere Aufnahmen der Szene handelt, die nacheinander (also mit einem zeitlichen Versatz) aufgenommen werden. Sie müssen deshalb zum einen nachträglich aneinander ausgerichtet werden und zum anderen darf sich die Szene nicht verändern, da sonst Geisterbilder oder Bewegungsunschärfe in der Aufnahme entstehen würden.

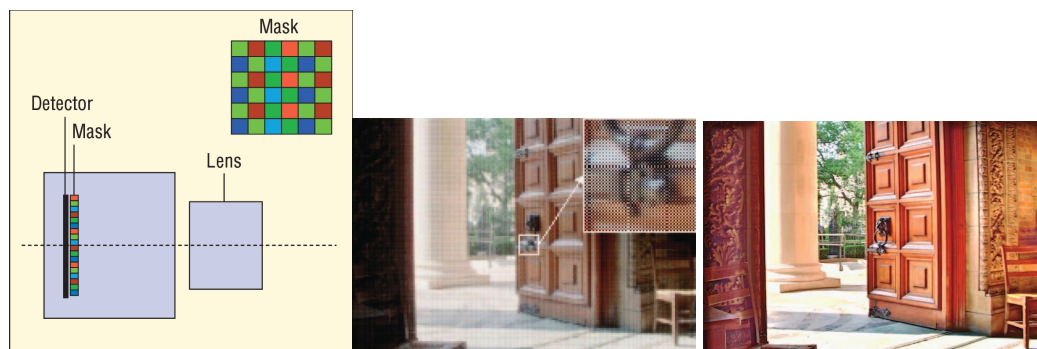


Bild 3.2: links: Schema einer Computational HDR Camera, rechts: Rohdaten, unten: daraus erstellte HDR-Aufnahme (aus [Nay06])

Eine *Computational Camera*, die ein ähnliches Prinzip wie der *Bayer-Filterung* zu Grunde liegt, erhöht das Kontrastverhältnis einer Aufnahme, indem der Filter nicht nur Farben sondern auch unterschiedliche Helligkeiten voneinander trennt (vgl. Abbildung 3.2). Den unterschiedlichen Helligkeiten entsprechend entstehen so mehrere unterschiedlich „belichtete“ Aufnahmen der Szene, die zu einer Aufnahme mit einem höheren dynamischen Umfang kombiniert werden. Mit diesem Verfahren können auch dynamische Szenen ab-

gelichtet werden, ohne dass es zu Bewegungsunschärfe kommen würde.

Eine weitere erstrebenswerte Eigenschaft einer Kamera wäre, wenn ihr Sichtfeld erweitert werden könnte. Ein Superweitwinkelobjektiv<sup>2</sup> hat bei der Verwendung einer digitalen Spiegelreflexkamera, wie der in dieser Diplomarbeit verwendeten Nikon D60 mit einem APS-C Sensor, in der Diagonale nur ein Sichtfeld von ungefähr 100°. Selbst ein Fisheye-Objektiv, welches die Lichtstrahlen vor allem im Randbereich stark beugt, kann nur etwas mehr als 180° des gesamten Umfelds erfassen. Aufgrund der starken Verzerrung in den Randbereichen ist es schwer ein Fisheye-Objektiv zu konstruieren, das bei einem Sichtfeld von deutlich mehr als 180° noch qualitativ hochwertige Ergebnisse in diesen Bereichen liefern kann.

Um den Moment als Ganzes einzufangen, wäre es wünschenswert das gesamte Umfeld mit einer einzigen Kamera zu erfassen und somit nichts mehr zu „verpassen“. Dabei würde es sich nicht nur um Panoramabilder sondern omnidirektionale Bilder handeln, also Bilder mit einem Sichtfeld von 360° nicht nur in horizontaler sondern auch vertikaler Richtung. So wäre es dem Betrachter später möglich, sich im aufgenommenen Moment umzusehen, als ob er ein Teil der Szene wäre. Würde dies schließlich mit bewegten oder gar stereoskopischen Bildern kombiniert werden, könnte ein solcher Ansatz zu ungeahnten Möglichkeiten in der Berichterstattung führen, die möglicherweise sogar ein immersives Erleben dieses Moments ermöglichen könnte.

Um die Einschränkungen von Fisheye-Objektiven zu umgehen, wird versucht die Szene mit Hilfe von gekrümmten Spiegeln so einzufangen, dass sie möglichst verzerrungsarm rekonstruiert werden kann. Diese Kombination aus Linsen und Spiegeln wird *katadioptrisches System* oder *Spiegellinsenobjektiv* genannt. Bisher fanden katadioptrische Systeme vornehmlich in Teleskopen Verwendung, welche ein relativ kleines Sichtfeld besitzen. Sie können aber ebenfalls dazu eingesetzt werden das Sichtfeld zu erhöhen. Hierbei können ellipsoide, hyperboloide und auch paraboloiden Spiegel zum Einsatz kommen.

Unter den beiden Bedingungen, dass sich alle Lichtstrahlen in einem einzelnen Punkt treffen sowie die Geometrie der Optik bekannt ist, kann jedes noch so verzerrte Bild in ein normales perspektivisches Bild umgewandelt werden [Nay97]. Für die Darstellung solcher

---

<sup>2</sup>Superweitwinkelobjektive haben Brennweiten von 14-24 mm.

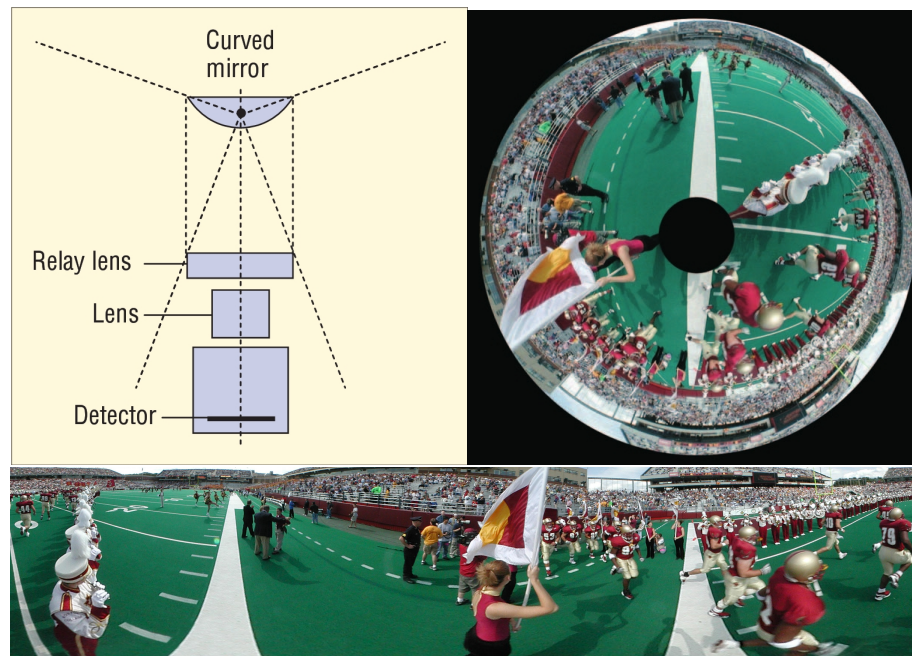


Bild 3.3: links: Schema einer *Computational Camera* mit großem Sichtfeld, rechts: Beispielaufnahme, Unten: daraus erstelltes Panorama (aus [Nay06])

Bilder wird eine emulierte Kamera benutzt, die sich frei in der aufgenommenen Szene drehen lässt und so den Eindruck vermittelt, der Betrachter sei Teil der Szene. Bild 3.3 zeigt das Schema, eine Beispielaufnahme und ein daraus erstelltes Panorama der Szene einer solchen Kamera. Die dargestellte Kamera hat ein Sichtfeld von  $360^\circ$  in horizontaler und  $220^\circ$  in vertikaler Richtung. Im Vergleich zum *Mosaicing* entfällt hierbei das *Stitching* der von einer rotierenden Kamera aufgenommenen Bilder. Ein weiterer wichtiger Aspekt ist, dass die Szene nicht statisch sein muss, sodass ein solches System unter Verwendung von Videokameras auch für Videokonferenzen oder zur Überwachung eingesetzt werden kann.

Das letzte Beispiel für die vielfältigen Möglichkeiten, die sich durch das Modell der *Computational Camera* ergeben, ist die Rekonstruktion von Tiefeninformationen aus einer einzelnen Aufnahme. Dieser Aspekt soll in der vorliegenden Diplomarbeit genauer untersucht werden (siehe Abschnitt 3.3). Auch hier kommen katadioptrische Systeme zum Einsatz, die diesmal jedoch nicht das Sichtfeld erweitern, sondern Teile davon so umlenken, dass die Szene mehrfach aus unterschiedlichen Perspektiven aufgenommen wird und es so-



Bild 3.4: links: Schema eines radialen Stereosystems, mitte: Rohaufnahme dieses Systems rechts: texturiertes 3D-Modell (aus [Nay06])

mit ermöglicht wird Tiefeninformationen der Szene aus nur einer Aufnahme zu gewinnen (vgl. Abbildung 3.4). Verschiedene Systemaufbauten kommen für diesen Einsatzzweck in Frage, die sich grob in Systeme mit ebenem Spiegel (siehe Abschnitt 3.4), und welche mit gekrümmtem Spiegel (siehe Abschnitt 3.5) einteilen lassen. Jeweils ein Vertreter dieser beiden unterschiedlichen Systeme soll untersucht werden. Der Fokus wird jedoch auf der Variante mit gekrümmtem Spiegel, oder genauer einem innen verspiegelten konischen Rohr, liegen. Dieses Rohr wird so vor einer Kamera mit einem Fisheye-Objektiv platziert, dass die optischen Achsen beider Komponenten aneinander ausgerichtet sind. Die Kamera lichtet die Szene zum einen direkt und zum anderen in der spiegelnden Innenseite des Rohrs ab. Ein Weltpunkt wird in diesem System einmal direkt und zweimal durch den Spiegel abgebildet. Durch diese Eigenschaft des Systems lassen sich drei Ansichten der Szene erstellen, die zur Rekonstruktion der Tiefeninformationen der Szene genutzt werden können (siehe Abschnitt 3.5.1).

### 3.3 Katadioptrische Stereokamera

Katadioptrische Kameras sind optische Systeme, welche Linsen und Spiegel in sich vereinen [HZ05]. Der Begriff wurde aus Katoptrik, dem Teilgebiet der Optik, welches sich mit Spiegeln und Reflexionen beschäftigt, und Dioptrik, einem veralteten Begriff für das Teilgebiet, das sich mit Linsen und der Brechung des Lichts beschäftigt, zusammengesetzt.

Mit solchen Systemen ist es möglich, Stereokameras herzustellen, die lediglich einen Sensor benötigen [GG93], [IHI93], [SBF96], [NN98], [GN01], [KN06]. In all diesen Systemen kommen ein oder mehrere Spiegel in einer Vielzahl unterschiedlicher Formen und Konstellationen zum Einsatz, mit dem Ziel die Szene aus mehreren Blickwinkeln (Viewpoints) aufzunehmen. Sie vereinen dadurch die Vorteile von Ein-Kamera-Systemen mit denen von Mehr-Kamera-Systemen: Da nur eine Kamera benötigt wird, besitzen die Ansichten identische radiometrische Eigenschaften und sie müssen dabei weder synchronisiert, noch müssen die Kamerabewegungen gemessen werden, da sich die Ansichten in einer Aufnahme befinden. Für eine dichte Abtastung der Szene werden in einem minimalen katadioptrischen Stereosystem lediglich ein Spiegel und eine Kamera benötigt. Ein Nachteil, der allen katadioptrischen Stereosystemen gemein ist, ist die Reduktion der Auflösung, da entweder ein größeres Sichtfeld auf dem Sensor abgebildet werden muss und sich dadurch die Pixeldichte pro Sichtwinkel verringert oder sich mehrere Aufnahmen der Szene die räumliche Auflösung des Sensors teilen müssen. Die hier vorgestellten katadioptrischen Kameras sind auch immer *Computational Cameras*, das heißt die Aufnahmen sind in ihrer Rohform nicht direkt verwendbar, was in gewissem Sinne auch einen Nachteil darstellt.

Übersicht 3.3 stellt die Vor- und Nachteile von katadioptrischen Stereosystemen noch einmal gegenüber.

### 3.4 Katadioptrische Planspiegel-Stereosysteme

In diesem Abschnitt soll die Klasse der katadioptrischen Stereokameras vorgestellt werden, die Planspiegel nutzen um aus der Szene Tiefeninformationen zu rekonstruieren. Bei allen Spiegeln deren spiegelnde Oberfläche eben ist handelt es sich um Planspiegel. Es wird mindestens ein solcher Spiegel eingesetzt, um neben der direkten Aufnahme der Szene eine weitere Ansicht zu erhalten. Am weitesten verbreitet sind Systeme mit ein oder zwei Spiegeln. Zwei-Spiegel-Lösungen nehmen die Szene oft nicht direkt, sondern ausschließlich durch die Spiegel auf. Dies hat den Vorteil, dass beide nicht nur die gleichen radiometrischen Eigenschaften aufweisen, sondern auch auf dem gleichen Weg als Reflexion entstanden sind und somit keine Qualitätsunterschiede zwischen den beiden Ansich-



**Vorteile:**

- **Identische Systemparameter:**

Die Linsen- und Sensor-Parameter (Linsenunschärfe und -verzerrung, Brennweite und Pixelgröße) sind für die Stereobildpaare identisch. Diese Tatsache erleichtert das Stereo Matching.

- **Identische radiometrische Eigenschaften:**

Das spektrale Verhalten, die Gain und das Rauschenverhalten sind identisch.

- **Einfache Datengewinnung:**

Die Stereodaten befinden sich in einem Bild und müssen nicht synchronisiert oder speziell gespeichert werden.

**Nachteile:**

- **Reduktion der Auflösung:**

Mehrere Aufnahmen der Szene werden auf dem Sensor abgebildet und teilen sich somit die verfügbare Auflösung.

- **Rohform der Bilder nicht aussagekräftig:**

Um an die gewünschten Informationen zu gelangen, müssen die Rohdaten erst verarbeitet werden.

ten entstehen.

Es wird auch eine Unterklasse von diesen Systemen vorgestellt, welche durch die spezielle Konfiguration von Kamera und Spiegeln in der Lage ist wie ein ideales Stereo-Rig direkt rektifizierte Ergebnisse zu liefern und darüber hinaus die Vorteile von katadioptrischen Systemen besitzt.

### 3.4.1 Konfigurationen

In diesem Abschnitt werden verschiedene Konfigurationen von Planspiegelsystemen vorgestellt und gezeigt was es hierbei für katadioptrische Stereosysteme zu beachten gilt. Wird ein Szenenpunkt wie in Abbildung 3.5 in einem Spiegel reflektiert, so ist es, als ob er von einer virtuellen Kamera hinter dem Spiegel aufgenommen würde. Um die virtuelle Kamera zu erhalten, muss die reale Kamera an der Spiegelebene reflektiert werden. Die exakte Berechnung dieses virtuellen Viewpoints wird in Abschnitt 3.4.2 beschrieben.

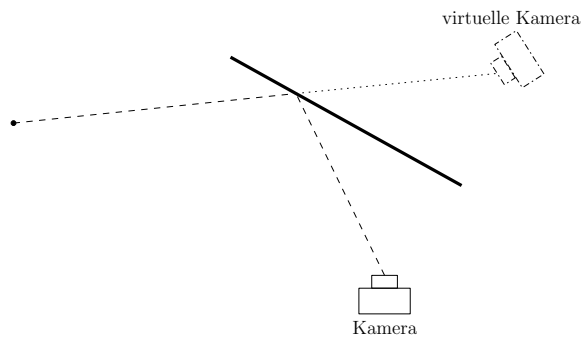


Bild 3.5: Ein gespiegelter Szenenpunkt wird von einer virtuellen Kamera eingefangen.

Wird die Szene an mehreren Spiegeln reflektiert, so kann die virtuelle Kamera ermittelt werden, indem die reale Kamera an den aufeinander folgenden Spiegelebenen gespiegelt wird (vgl. Abbildung 3.6). Die Kameras befinden sich hierbei alle auf einer Ebene, sodass die Transformationen zwischen ihnen nur fünf Freiheitsgrade aufweisen (vgl. Abschnitt 3.4.2).

Wird die Szene durch eine ungerade Spiegelanzahl reflektiert, so findet ein Wechsel von einem *Linke-Hand-* in ein *Rechte-Hand-System* statt und die virtuelle Kamera liefert spiegelverkehrte Bilder. Bei einer Reflexion an einer geraden Spiegelanzahl heben sich die Wechsel gegenseitig auf und die virtuelle Kamera liefert entsprechend nicht seitenvertauschte Bilder.

Die bisherigen Konfigurationen sind noch nicht in der Lage Stereoaufnahmen eines Szenenpunktes zu erstellen, da dieser maximal von einer Kamera aus einer Perspektive aufgenommen wird. Um stereoskopisches Sehen zu ermöglichen, werden mehrere Perspektiven des Szenenpunktes benötigt (vgl. Abbildung 3.7). Die Bildwinkel der involvierten Kame-

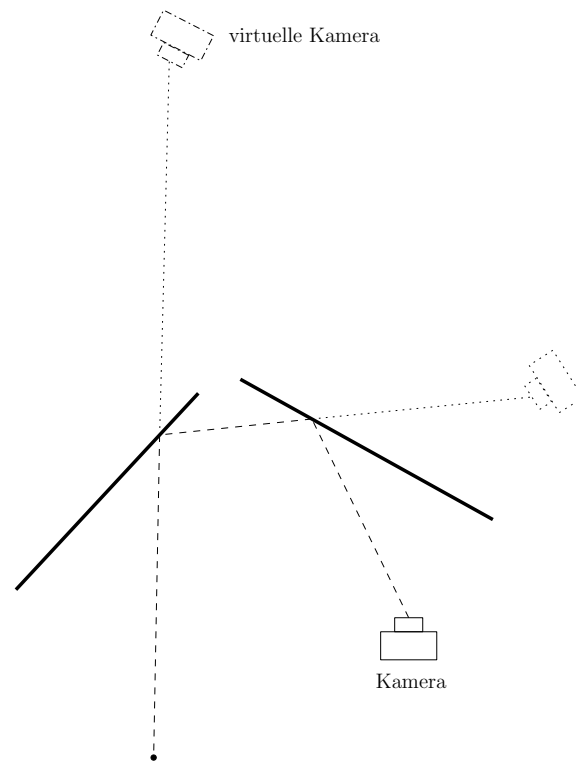


Bild 3.6: Ein Szenenpunkt wird über mehrere Spiegel hinweg aufgenommen.

ras müssen sich an der Stelle des aufzunehmenden Punktes überschneiden (vgl. Abbildung 3.11).

### 3.4.2 Eigenschaften

Dieser Abschnitt erklärt die allgemeinen Eigenschaften von katadioptrischen Planspiegel-Stereokameras. Ein System mit einer Kamera und zwei Spiegeln in beliebiger Anordnung dient als Beispielkonfiguration, an der sich die Eigenschaften eines solchen Systems gut zeigen lassen. Abbildung 3.8 zeigt ein solches System.

Der Szenenpunkt  $p$  von den beiden virtuellen Viewpoints  $V_1$  und  $V_2$  gesehen und wird auf den virtuellen Bildebenen abgebildet. Die Position der virtuellen Viewpoints kann ermittelt werden, indem der reale Viewpoint  $O$  am jeweiligen Spiegel  $m_1$  beziehungsweise

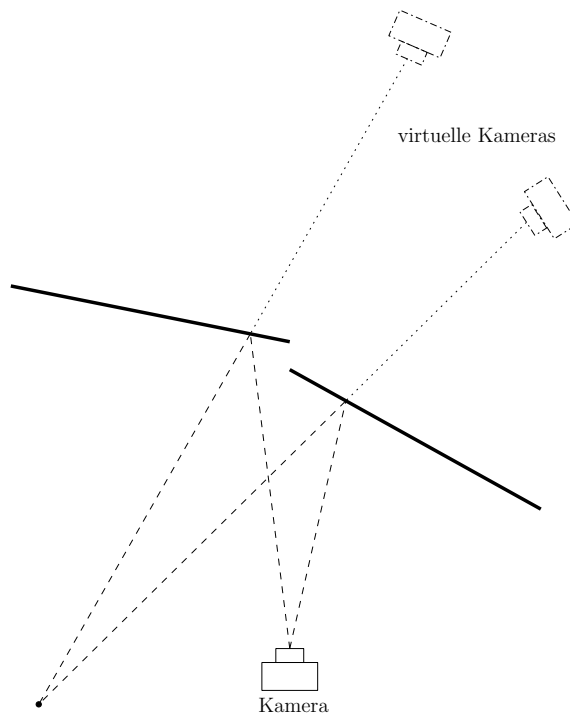


Bild 3.7: Ein Szenenpunkt wird über mehrere Spiegel hinweg aufgenommen.

$m_2$  reflektiert wird. Wird die optische Achse der realen Kamera ebenfalls an den Spiegeln reflektiert, so lassen sich die optischen Achsen und somit die Orientierung der virtuellen Kameras ermitteln. Die Brennweite  $f$  sowie die restlichen intrinsischen Kameraparameter sind wie bei allen katadioptrischen Systemen identisch. Mit Hilfe der Entfernung der realen Kamera von den Spiegeln und deren Orientierung in Relation zur Kamera, können die Positionen der virtuellen Viewpoints sowie deren Orientierung ermittelt werden (siehe Abschnitt 3.4.2).

Im Folgenden sollen die Eigenschaften eines katadioptrischen Systems mit zwei Planspiegeln näher beschrieben werden. Für eine tiefergehende Analyse dieser Stereokameraklasse sei auf [GN01] verwiesen.

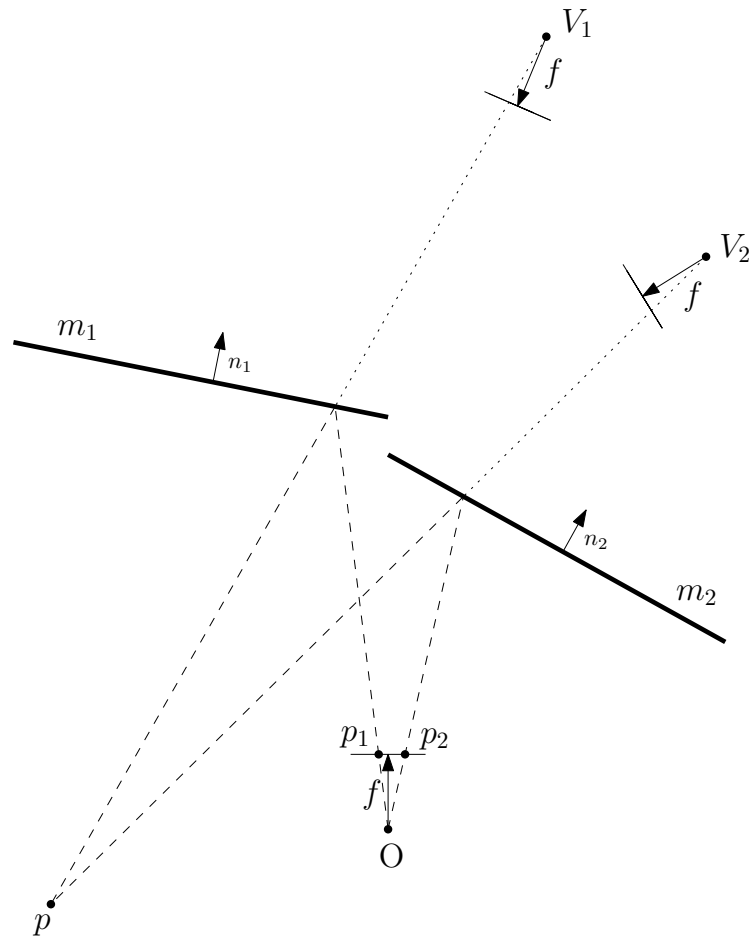


Bild 3.8: katadioptrisches System mit zwei Planspiegeln

### Position der Viewpoints

Die beiden virtuellen Viewpoints liegen auf einer Ebene. Um sie ineinander zu überführen bedarf es lediglich einer Rotation um die Achse  $S$  (vgl. Abbildung 3.10). Bei der Rotationsachse  $S$  handelt es sich um die Schnittlinie zwischen den beiden Spiegelebenen, also um das Kreuzprodukt der zwei Spiegelnormalen  $\mathbf{n}_1 \times \mathbf{n}_2$ . Die virtuellen Viewpoints liegen somit auf einer Ebene die orthogonal zu dieser Achse steht, wodurch sich die Freiheitsgrade der Transformation zwischen ihnen von sechs<sup>3</sup> auf fünf reduzieren, da die Translation

<sup>3</sup>Drei Freiheitsgrade für die Rotation und drei für die Translation

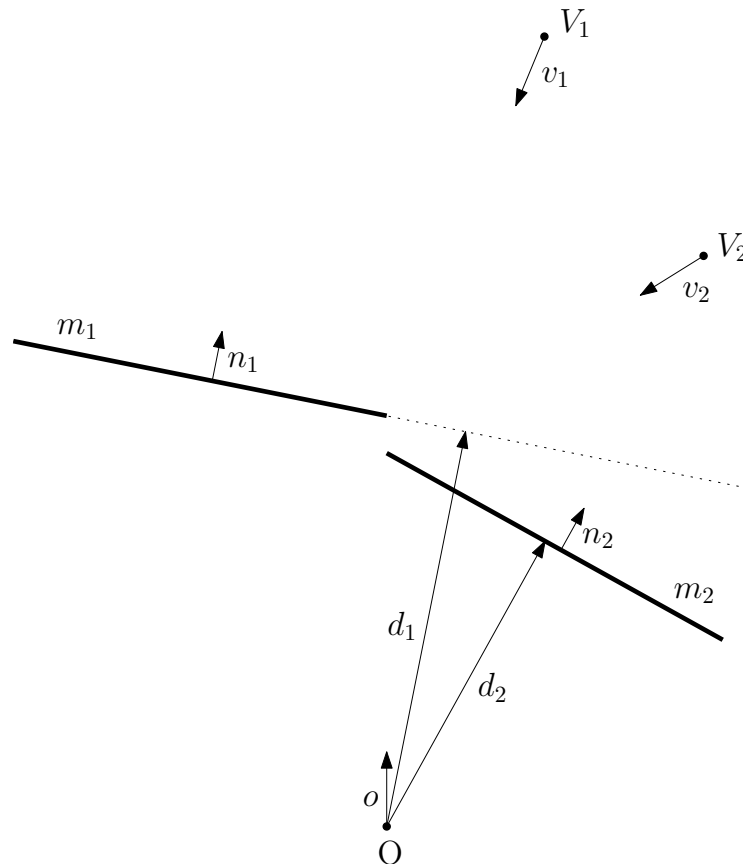


Bild 3.9: Position und Orientierung der Viewpoints

auf einer Ebene nur zwei Freiheitsgrade besitzt.

Um die virtuellen Viewpoints zu ermitteln, muss der reale Viewpoint  $O$  am jeweiligen Spiegel reflektiert werden (vgl. Abbildung 3.9). Die Transformation zwischen realer und virtueller Kamera kann wie folgt beschrieben werden:

$$\mathbf{V}_1 = 2d_1 \mathbf{n}_1 \cdot \mathbf{O}$$

und

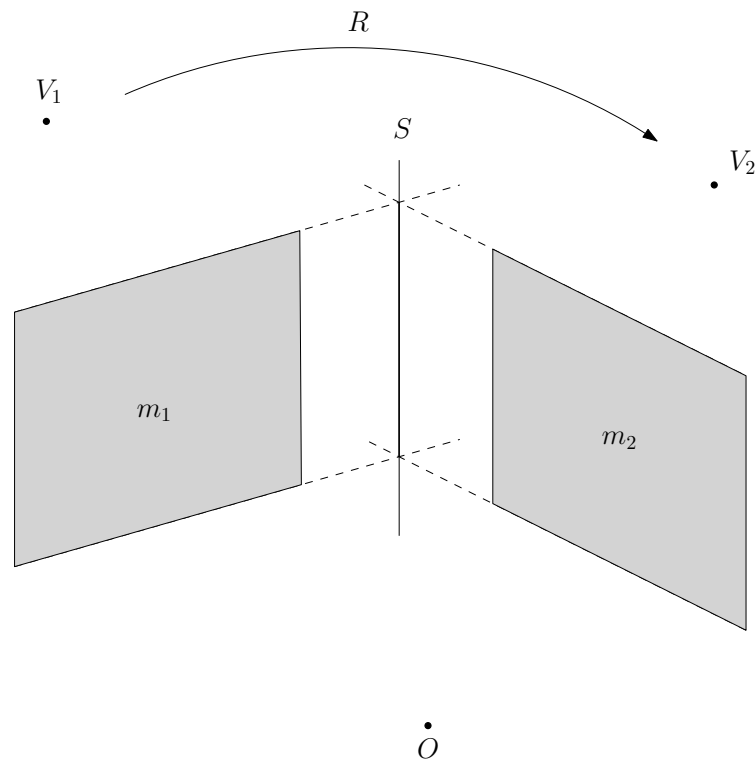


Bild 3.10: Die Rotationsachse der beiden virtuellen Viewpoints

$$\mathbf{V}_2 = 2d_2\mathbf{n}_2 \cdot \mathbf{O}$$

Hierbei sind  $\mathbf{n}_1$  und  $\mathbf{n}_2$  die Spiegelnormalen von  $m_1$  respektive  $m_2$  und  $d_1$  und  $d_2$  sind analog die Abstände zwischen realer Kamera und den jeweiligen Spiegeln (vgl. Abbildung 3.9). Da es sich hier um eine Spiegelung handelt, erfolgt die Translation in Richtung der Spiegelnormalen mit der doppelten Länge des Abstands zum Spiegel.

Die Rotation der Orientierung wird wie folgt berechnet:

$$\mathbf{v}_1 = \mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1^T \cdot \mathbf{o}$$

und

$$\mathbf{v}_2 = \mathbf{I} - 2\mathbf{n}_2\mathbf{n}_2^T \cdot \mathbf{o}$$

Bei den Vektoren  $\mathbf{o}$ ,  $\mathbf{v}_1$  und  $\mathbf{v}_2$  handelt es sich um die jeweiligen optischen Achsen der realen und virtuellen Kameras. Sie verlaufen in Blickrichtung ihrer zugehörigen Kamera, wobei  $\mathbf{I}$  die  $3 \times 3$  Einheitsmatrix ist.

Die relative Rotation  $\mathbf{R}$  zwischen den virtuellen Kameras ergibt sich aus:

$$\begin{aligned} \mathbf{R} &= \mathbf{I} - 2\mathbf{n}_2\mathbf{n}_2^T \cdot (\mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1^T)^{-1} \\ &= \mathbf{I} + 4(\mathbf{n}_1 \cdot \mathbf{n}_2)\mathbf{n}_1\mathbf{n}_2^T - 2\mathbf{n}_1\mathbf{n}_1^T - 2\mathbf{n}_2\mathbf{n}_2^T \end{aligned}$$

Weiterhin ergibt sich die Translation  $\mathbf{t}$  zwischen den virtuellen Kameras aus:

$$\begin{aligned} \mathbf{t} &= 2d_2\mathbf{n}_2 \cdot 2d_1\mathbf{n}_1 \\ &= 2d_1\mathbf{n}_1 - (4d_1(\mathbf{n}_1 \cdot \mathbf{n}_2) + 2d_2)\mathbf{n}_2 \end{aligned}$$

## Sichtfeld

Da sich die virtuellen Kameras das Sichtfeld der realen Kamera teilen, entspricht ihr maximal erreichbarer Bildwinkel der Hälfte des Bildwinkels der realen Kamera. Das gemeinsame Sichtfeld der beiden virtuellen Viewpoints, also der Bereich des bifokalen Sehens kann je nachdem in welchem Winkel die Spiegel zueinander stehen noch geringer ausfallen. Das gemeinsame Sichtfeld der beiden virtuellen Viewpoints entspricht, wie in Bild 3.11 abgebildet, dem Doppelten der relativen Orientierung  $\alpha$  der beiden Spiegel zueinander und ist somit  $2\alpha$ . Infolgedessen muss das Sichtfeld der realen Kamera mindestens einen Winkel von  $2\alpha$  aufweisen, um das Sichtfeld der virtuellen Viewpoints nicht weiter einzuschränken.



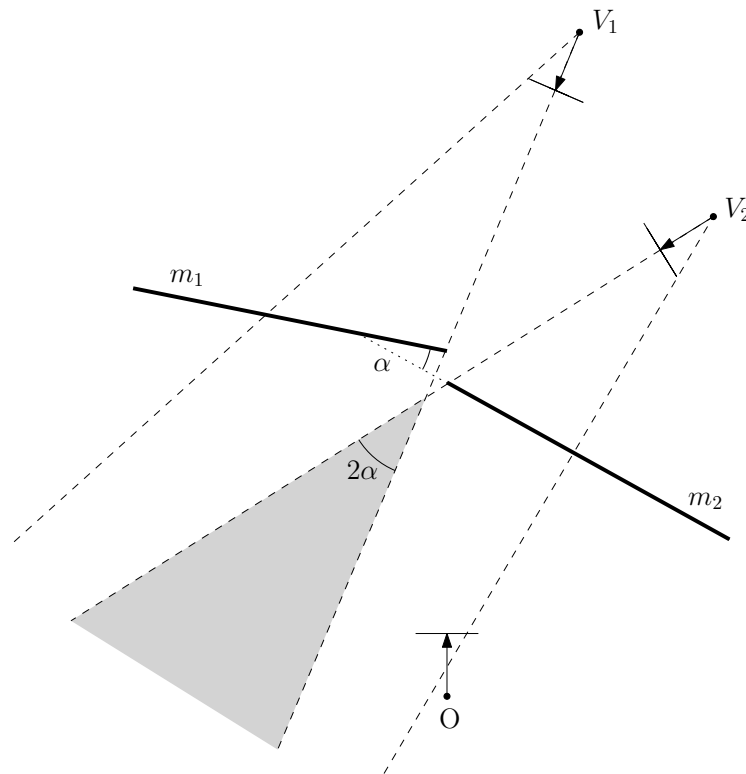


Bild 3.11: Das bifokale Sichtfeld entspricht dem Doppelten der relativen Orientierung  $\alpha$  der beiden Spiegel zueinander

### 3.4.3 Rektifizierte Systeme

In diesem Abschnitt werden Planspiegelsysteme vorgestellt, die aufgrund der besonderen Anordnung von Spiegeln und Kamera in der Lage sind, bereits rektifizierte Aufnahmen der Szene zu erstellen. Dies bietet zwei Vorteile. Zum einen müssen die Bilder nicht transformiert werden, was ein Qualitätsgewinn bedeutet, da die endgültigen Pixel nicht erst durch Interpolation entstehen, sondern direkt aus der Aufnahme verwendet werden können. Zum anderen kann so der Rechenaufwand einer nachträglichen Rektifizierung vermieden werden. Im Vergleich zu konventionellen Stereosystemen sind diese Vorteile durch eine Verringerung des Bildwinkels erkauft, da sich wie bei allen Planspiegelsystemen die virtuellen Kameras das Sichtfeld der realen Kamera teilen müssen.

Damit ein Stereosystem rektifizierte Ergebnisse liefert, muss es folgende Voraussetzungen erfüllen:

- Es darf keine relative Rotation zwischen den optischen Achsen der beteiligten Kameras existieren.
- Die Viewpoints müssen parallel zu den horizontalen oder vertikalen Scanlinien transliert sein.
- Die Kameraparameter müssen nahezu identisch sein.

Die letzte Bedingung ist bei katadioptrischen Kameras systemimmanent, da nur eine Kamera verwendet wird.

Sind all diese Voraussetzungen erfüllt, so verlaufen die Epipolarlinien horizontal und entsprechen den Scanlinien in den Stereoaufnahmen. Dies ermöglicht es, direkt eine dichte Tiefenkarte (vgl. Abschnitt 2.3) des Stereobildpaares zu erstellen.

### Konfigurationen

Eine Konfiguration die rektifizierte Bilder liefert muss oben genannte Bedingungen erfüllen. Identische Kameraparameter sind in einem Ein-Kamera-System gegeben. Um den weiteren Voraussetzungen gerecht zu werden, muss die Rotation  $\mathbf{R}$  zwischen den Kameras in folgender Form der Einheitsmatrix entsprechen:

$$\mathbf{R} = \mathbf{R}_1 \cdots \mathbf{R}_i \cdots \mathbf{R}_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.1)$$

Hier ist  $n$  die Anzahl der Spiegel an denen die reale Kamera gespiegelt wird und  $\mathbf{R}_i$  die durch den  $i$ -ten Spiegel verursachte Rotation. Des Weiteren muss die Translation zwischen den Kameras folgende Bedingung erfüllen:

$$t = t_1 \cdots t_i \cdots t_n = \begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix} \quad (3.2)$$

Auch hier ist  $t_i$  die durch den  $i$ -ten Spiegel verursachte Translation. Der Abstand  $b$  zwischen den Kameras ist folglich die Basislinie des Stereosystems.

Obwohl die Bedingungen 3.2 und 3.1 ausreichen, damit ein rektifiziertes System entsteht, so bedarf es noch einer weiteren Bedingung, um es auch zu einem Stereosystem zu machen. Für stereoskopische Aufnahmen werden mehrere Perspektiven der Szene benötigt, weshalb sichergestellt werden muss, dass sich die Sichtfelder der Kameras schneiden.

Beispiele für Ein- beziehungsweise Drei-Spiegel-Systeme finden sich in den Abbildungen 3.12 und 3.13. Das dargestellte Drei-Spiegel-System kann sich je virtueller Kamera das maximal mögliche halbe Sichtfeld der realen Kamera nutzbar machen. Anhand des Ein-Spiegel-Systems werden in den folgenden Abschnitten die Eigenschaften von rektifizierten Stereokameras erläutert.

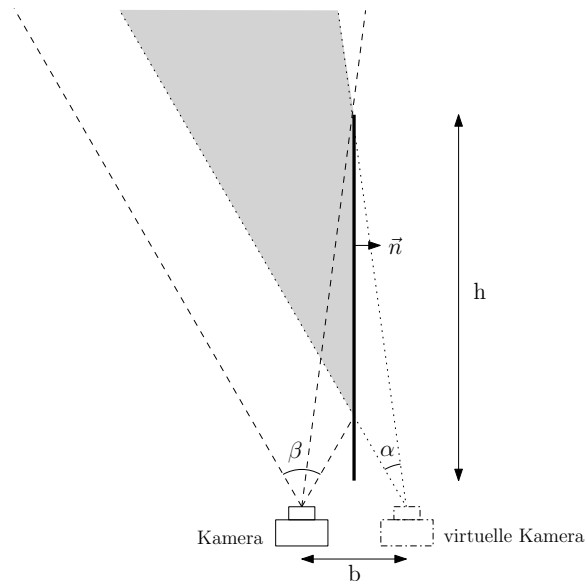


Bild 3.12: Rektifiziertes Ein-Spiegel-System

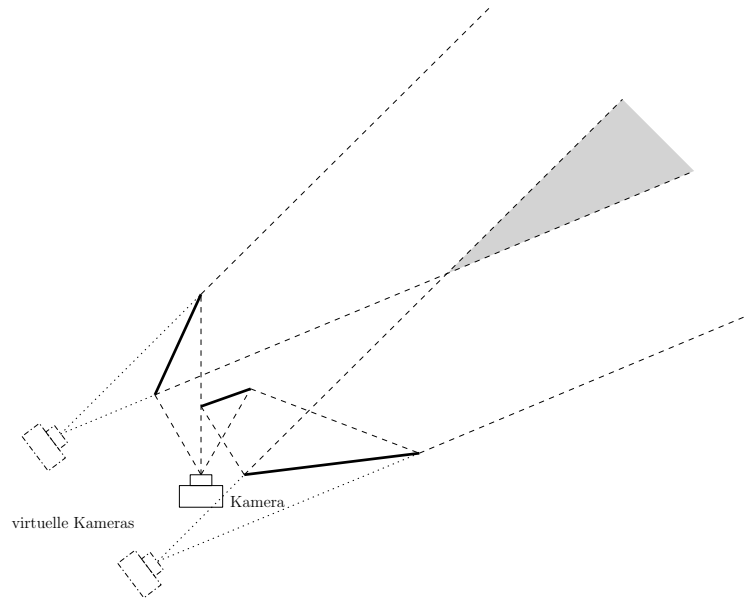


Bild 3.13: Rektifiziertes Drei-Spiegel-System

### Eigenschaften

In diesem Abschnitt werden an einem Ein-Spiegel-System (vgl. Abbildung 3.12) exemplarisch die Eigenschaften rektifizierter katadioptrischer Stereokameras gezeigt. Um die Voraussetzungen rektifizierter Systeme zu erfüllen, muss die Rotation  $R_1$ , die durch den Spiegel im System verursacht wird, Bedingung 3.1 einhalten:

$$\mathbf{R}_1 = (\mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1^T) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.3)$$

mit der Spiegelnormale  $\mathbf{n}_1$ . Die einzige Lösung für diese Normale ist  $\mathbf{n}_1 = [1, 0, 0]^T$ , das heißt die einzige Möglichkeit einer Ein-Spiegel-Konfiguration besteht darin, dass die Normale der Spiegelebene  $\mathbf{n}_1$  parallel zu den Scanlinien (x- oder y-Achse) der Kamera verläuft, wie es in Abbildung 3.12 dargestellt ist.

Um die Translationsbedingung zu erfüllen, muss  $\mathbf{t}_1$  der Bedingung 3.2 genügen:

$$\mathbf{t}_1 = (2d_1 \mathbf{n}_1) = \begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix} \quad (3.4)$$

$d_1$  ist hier die Entfernung von Kamera und Spiegel. Die Basislinie ist folglich  $b = 2d_1$ . Der Abstand zwischen realer und virtueller Kamera entspricht in dieser Konfiguration erwartungsgemäß dem doppelten Abstand von realer Kamera und Spiegel.

Der Bildwinkel der virtuellen Kamera wird durch die Größe des Spiegels beschränkt und hängt nicht nur vom Bildwinkel  $\beta$  der realen Kamera, sondern auch von der Basislinie und der Länge des Spiegels ab:

$$\alpha = \arctan\left(\frac{2h}{b}\right) - \frac{\pi}{2} + \frac{\beta}{2} \quad (3.5)$$

Der Nachteil dieser Konfiguration besteht darin, dass ein sehr großer Spiegel benötigt wird, um eine große Basislinie zu erzielen. Des Weiteren sind die Bildwinkel der Kameras asymmetrisch, was sich in dieser Konfiguration nur beheben ließe, indem der Sensor in Richtung des Spiegels verschoben würde, die Kamera also modifiziert würde.

### **Einsatzgebiete**

Die Einsatzmöglichkeiten von Planspiegelsystemen sind nahezu unbeschränkt, solange der effektive Bildwinkel dafür ausreicht. Rektifizierte sowie nicht-rektifizierte Verfahren eignen sich für Echtzeitanwendungen. Da bei den rektifizierten Systemen die Rektifizierungstransformationen wegfallen, eignen sie sich jedoch in noch höherem Maße für echtzeitfähige Anwendungen [FHM<sup>+</sup>93].

## **3.5 Radiale katadioptrische Stereosysteme**

In diesem Abschnitt werden radiale katadioptrische Systeme beschrieben, die zur Klasse katadioptrischer Systeme mit gekrümmten Spiegeln gehören. Vorgestellt wurden diese Sy-

steme in der Arbeit von Kuthirummal und Nayar [KN06]. Alle Spiegel dieser Klasse sind rotationssymmetrische Kegelstümpfe, deren Innenseite verspiegelt ist. Sie sind in der Lage die Szene gleichzeitig aus mehreren Viewpoints in einem einzelnen Bild einzufangen. Da in diesen Systemen nur eine einzige Kamera zum Einsatz kommt, sind alle Bildpunkte den genau gleichen radiometrischen Eigenschaften dieser Kamera ausgesetzt, was einen großen Vorteil darstellt. Darüber hinaus müssen die Aufnahmen nicht synchronisiert werden, da sie sich in einem Bild befinden und somit zum gleichen Zeitpunkt aufgenommen werden.

Um das zugrundeliegende Prinzip von radialen katadioptrischen Stereosystemen zu verstehen, ist es hilfreich sich eine Beispielkonfiguration eines solchen Systems anzuschauen (siehe Bild 3.27). Diese Systeme bestehen aus einer konventionellen Kamera, die durch einen hohlen Rotationskörper blicken<sup>4</sup>, dessen Innenseite verspiegelt ist. Die optischen Achsen von Kamera und Optik müssen übereinstimmen, was vorab in einem Konfigurationsschritt sichergestellt wird. Die Szene wird auf diese Weise direkt und noch einmal nach der Reflexion durch den Spiegel auf dem Sensor abgebildet. Das Sichtfeld der Kamera wird also durch den Spiegel nach innen gefaltet, sodass die Szene aus mehreren Perspektiven innerhalb eines einzelnen Bilds aufgenommen wird.

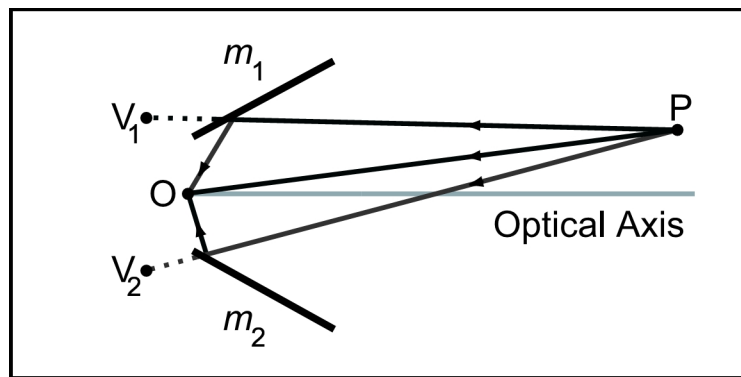


Bild 3.14: Die reale Kamera  $O$  und die virtuellen Kameras  $V_1$  und  $V_2$  sehen den Weltpunkt  $p$  aus unterschiedlichen Perspektiven (aus [KN06])

Das System in Abbildung 3.27 fängt die Szene zum einen direkt vom realen Viewpoint

<sup>4</sup>Es wird hier nur näher auf Spiegel in Form eines rotationssymmetrischen Kegelstumpfs eingegangen, auch wenn andere komplexere Formen ebenfalls denkbar wären.

der Kamera und zum anderen von den auf einer kreisförmigen Ortslinie (Locus) liegenden virtuellen Viewpoints des Spiegels aus ein. In einem Längsschnitt (vgl. Abbildung 3.14) dieses Systems, der durch die optische Achse des Gesamtsystems verläuft, ist dies gut zu erkennen. Der reale Viewpoint befindet sich an Punkt  $O$ . Der Spiegel erscheint im Längsschnitt als die zwei getrennten Spiegellinien  $m_1$  und  $m_2$ , welche wiederum für die virtuellen Viewpoints  $V_1$  und  $V_2$  verantwortlich sind. Sie entstehen aus der Spiegelung des realen Viewpoints  $O$  an den Linien  $m_1$  und  $m_2$ . Somit hat jeder durch die optische Achse verlaufende Längsschnitt zwei symmetrische virtuelle Viewpoints, da es sich um Spiegelungen des auf der optischen Achse liegenden realen Viewpoints handelt. Das System besteht aus unendlich vielen solcher Längsschnitte, sodass die virtuellen Viewpoints auf einer kreisförmigen Ortslinie um die optische Achse des Gesamtsystems liegen.

Im Querschnitt des Systems (vgl. Abbildung in 3.15) liegen die virtuellen und der reale Viewpoint auf einer Linie. Alle Rohaufnahmen radialer Systeme entsprechen diesem Muster. Eine Bildlinie verläuft durch den Mittelpunkt und kann in drei Abschnitte unterteilt werden: die Linie  $\overline{JK}$  wird vom virtuellen Viewpoint  $V_1$  auf das Bild projiziert, die Linie  $\overline{LM}$  analog vom virtuellen Viewpoint  $V_2$  und der mittlere Abschnitt  $\overline{KL}$  entspricht der Projektion der direkten Ansicht des realen Viewpoints  $O$ . Ausschließlich der auf dieser Linie liegende Teil der Szene, wird in dieser Weise von den drei Viewpoints abgebildet. Dabei wird ein Punkt  $p$  der Szene, falls er auf dieser Linie liegt und von allen Viewpoints aus gesehen werden kann, dreimal abgebildet. Der reale Viewpoint sieht den Punkt  $p$  dort wo er sich in der Szene befindet, die beiden virtuellen Viewpoints wiederum sehen ihn an den Positionen  $p_1$  respektive  $p_2$ . Da dies für alle Punkte auf der Bildlinie gilt, handelt es sich bei ihnen um die Epipolarlinien dieser Systeme.

Aufgrund dieser sternförmig verlaufenden Epipolarlinien kommt es zu weniger Mehrdeutigkeiten während des Stereo Matchings, was einen großen Vorteil radialer Stereosysteme darstellt. Mehrdeutigkeiten entstehen nur bei Kanten, die auf einer der radialen Epipolarlinien liegen, was nur selten in gehäufte Form auftritt.

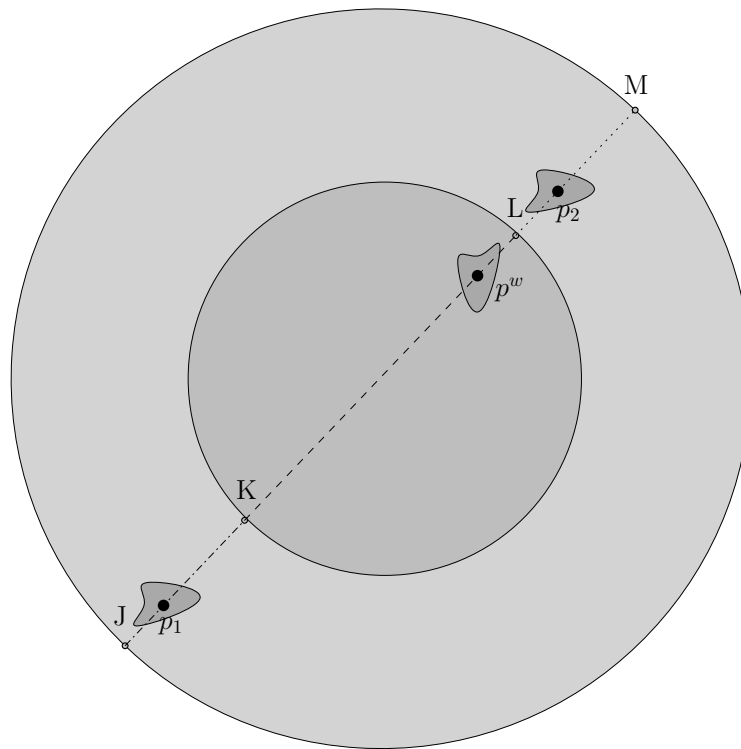


Bild 3.15: Aufbau eines mit einem radialen System aufgenommenen Bildes

### 3.5.1 Eigenschaften

Die Form des Spiegels, die einem rotationssymmetrischen Kegelstumpf entspricht und seine Lage im Gesamtsystem lässt sich durch drei Parameter beschreiben (siehe hierzu Bild 3.16):

- **Der Radius  $r$**   
Um die Orientierung des Spiegels im Gesamtsystem kenntlich zu machen, wird hier der Kamera zugewandte Öffnungsradius angegeben.
- **Die Länge  $l$**   
Es handelt sich hierbei geometrisch betrachtet um die Höhe des Kegelstumpfs.
- **Der halbe Öffnungswinkel  $\beta$**   
Der Winkel zwischen der (optischen) Achse und der Mantellinie der Innenseite des



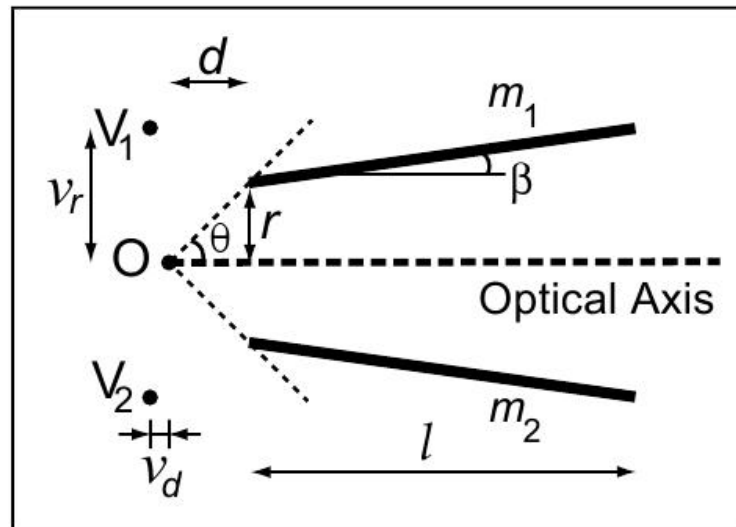


Bild 3.16: Längsschnitt (aus [KN06])

Kegels. Liegt die „Spitze“ des Kegels auf der Seite der Kamera, so handelt es sich um einen Öffnungswinkel, der größer Null ist. Liegt sie auf der der Kamera abgewandten Seite, so ist er kleiner Null.

Um die vollständige Konfiguration des Gesamtsystems anzugeben, bedarf es noch eines weiteren Parameters, dem Sichtfeld  $2\theta$ . Das Sichtfeld eines katadioptrischen Stereosystems ist das Minimum des horizontalen und vertikalen Sichtfelds der Kamera. Der Spiegel sollte in einer Distanz  $d = r \cot(\theta)$  vor der Kamera angebracht werden, sodass das gesamte Sichtfeld der Kamera genutzt wird.

Es lassen sich anhand des halben Öffnungswinkels  $\beta$  (vgl. Abschnitt 3.5.1) drei unterschiedliche Kategorien von Systemen des Spiegels unterscheiden, die sich mit verschiedenen Eigenschaften für jeweils andere Anwendungsgebiete eignen:

- ist  $\beta < 0$ , so handelt es sich bei dem Spiegel um einen konischen Kegelstumpf mit dem es möglich ist, die vollständige Textur eines konvexen Objektes in einer Aufnahme zu erfassen beziehungsweise unter Verwendung von zwei Aufnahmen mit einer Parallaxe entlang der optischen Achse die vollständige Geometrie eines konvexen Objektes zu ermitteln (vgl. Abschnitt 3.5.3).

- ist  $\beta > 0$ , so handelt es sich prinzipiell um den gleichen Spiegel wie bei  $\beta < 0$ , er wird nur andersherum vor der Kamera platziert. Diese Variante wird in dieser Diplomarbeit näher untersucht, da sie sich besonders als Nahbereichsscanner eignet (vgl. Abschnitt 3.5.3).
- ist  $\beta = 0$ , so handelt es sich um einen zylindrischen Spiegel (ein Spezialfall eines Kegelstumpfs) mit dem es beispielsweise möglich ist die bidirektionale Reflektanzverteilungsfunktion (BRDF) eines Materials zu ermitteln oder die 3D-Textur von Oberflächen zu vermessen (vgl. Abschnitt 3.5.2).

Ein weiterer Systemparameter der variiert werden kann ist die Länge des Spiegels. Bei einer bestimmten Spiegellänge wird die Szene genau einmal innerhalb des Spiegels reflektiert (vgl. Abschnitt 3.5.1). Wird diese Länge verringert, so kann nur noch ein Teil der Szene von den virtuellen Viewpoints erfasst werden. Wird sie hingegen verlängert, so werden die Lichtstrahlen mehrfach im Spiegel reflektiert, bevor sie auf den Sensor treffen. Es entstehen mehrere Ansichten der Szene, die von weiteren, wiederum auf einer kreisförmigen Ortslinie gelegenen, Viewpoints aufgenommen werden. Die aus einem solchen System resultierenden Bilder entsprechen immer noch dem in Abbildung 3.15 dargestellten Muster, allerdings unterteilt sich der Ring pro Ansicht in einen weiteren Ring. Den Jahresringen eines Baums ähnlich, befinden sich diese Ansichten jeweils in einem konzentrischen Ring um die direkte Ansicht in der Mitte.

Soll beispielsweise ein System mit zwei kreisförmigen Ortslinien erstellt werden, dürfen sich die Lichtstrahlen maximal zweimal im Spiegel reflektieren. In diesem Beispielsystem gäbe es demzufolge unverändert den Mittelkreis der direkten Ansicht, diesmal allerdings umlagert von zwei Ringen für die beiden virtuellen Viewpoint-Ortslinien. Ein möglicher Nutzen eines solchen Systems ist die Ermittlung der BRDF eines Materials, da es hierbei notwendig ist, das zu untersuchende Objekt aus möglichst vielen unterschiedlichen Perspektiven aufzunehmen.

Im Folgenden sollen einige Eigenschaften dieser Systeme näher beschrieben werden. Die maximale Länge die ein Spiegel haben darf, damit ein Lichtstrahl nur einmal darin reflektiert wird soll hergeleitet, die Position der virtuellen Viewpoints ermittelt und das Sichtfeld des Gesamtsystems beschrieben werden.

## Länge des Spiegels

Die Länge des Spiegels kann beliebig variiert werden. Systeme, die nur eine Ortslinie besitzen sollen, müssen die Länge  $l'$  aufweisen, bei der Lichtstrahlen nur einmal im Spiegel reflektiert werden. Diese Länge ermöglicht auch eine einfachere Analyse der weiteren Eigenschaften radialer Stereosysteme.

Um diese Länge zu berechnen wird ein Sehstrahl entlang des maximalen Bildwinkels  $\theta$  der Kamera verfolgt. An dem Punkt, an dem dieser Sehstrahl den Spiegel ein zweites Mal treffen würde, muss der Spiegel enden, damit die geforderte Bedingung erfüllt ist.

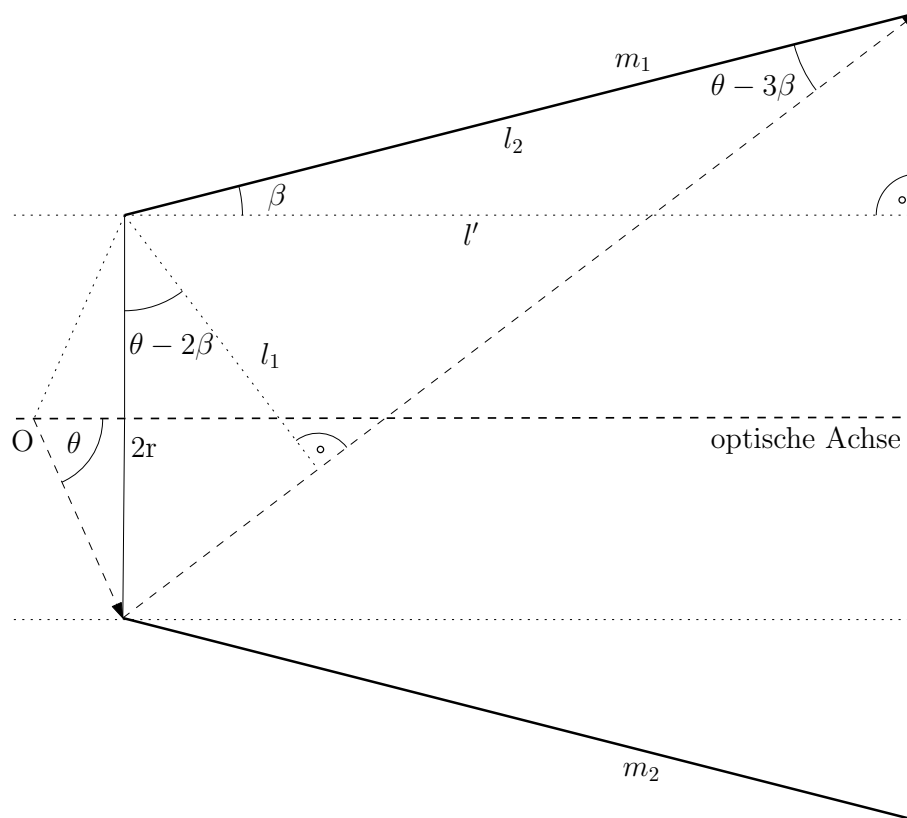


Bild 3.17: Länge  $l'$

Zunächst wird hierfür die Länge  $l_1$ , welche die Ankathete des rechtwinkligen Dreiecks mit der Hypotenuse  $2r$  und dem Winkel  $\theta - 2\beta$  ist, errechnet:

$$l_1 = 2r \cdot \cos(\theta - 2\beta)$$

Danach ist es möglich die Länge  $l_2$  zu berechnen:

$$l_2 = \frac{l_1}{\sin(\theta - 3\beta)}$$

Mit der Länge  $l_2$  ist es nun möglich die maximale Länge  $l'$  des Spiegels zu errechnen, ohne dass ein Lichtstrahl mehrfach reflektiert werden würde:

$$\begin{aligned} l' &= l_2 \cdot \cos(\beta) \\ &= \frac{2r \cdot \cos(\theta - 2\beta)}{\sin(\theta - 3\beta)} \end{aligned} \quad (3.6)$$

Um die weitere Analyse der Eigenschaften zu vereinfachen wird davon ausgegangen, dass Lichtstrahlen nur einmal innerhalb des Spiegels reflektiert werden, also  $l = l'$  gilt.

## Position der Viewpoints

Wie in Abschnitt 3.5 beschrieben, liegen die virtuellen Viewpoints auf einer kreisförmigen Ortslinie. Die Position dieser Ortslinie lässt sich mit Hilfe der Systemparameter berechnen. Da das System rotationssymmetrisch ist, kann die Analyse anhand eines Längsschnitts des Systems durchgeführt werden (vgl. Abbildung 3.16). Die virtuellen Viewpoints  $V_1$  und  $V_2$  entstehen durch die Reflexionen des realen Viewpoints an den Spiegeln  $m_1$  und  $m_2$ . Im Längsschnitt durch die optische Achse des Systems befinden sich zwei Spiegel, da der Kegelstumpf an zwei Stellen geschnitten wird, und zwar an den beiden Mantellinien  $m_1$  und  $m_2$ .

Der Abstand  $v_d$  zwischen den virtuellen und dem realen Viewpoint entlang der optischen Achse und der Abstand  $v_r$  zwischen den virtuellen Viewpoints und der optischen Achse

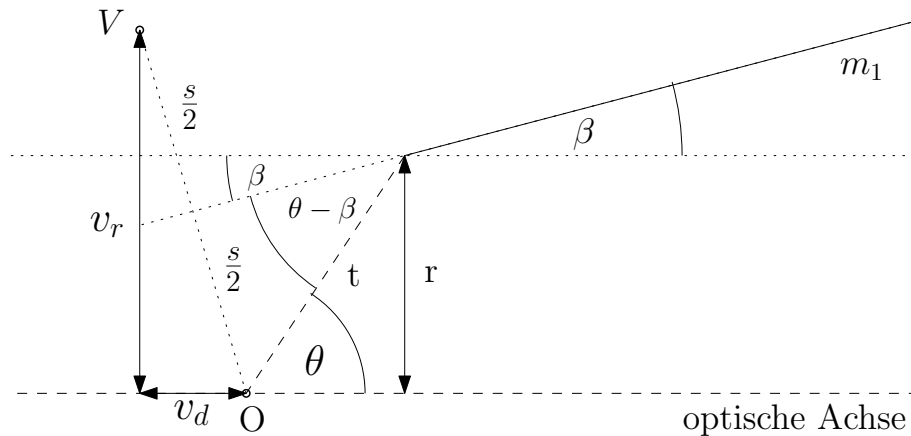


Bild 3.18: Position der Viewpoints

kann berechnet werden. Um diese Größen herzuleiten genügt die obere Hälfte des Längsschnittes, da das System achsensymmetrisch ist. Dazu werden die zwei Zwischengrößen  $s$  und  $t$  benötigt. Bei  $t$  handelt es sich um die Distanz des realen Viewpoints  $O$  zur Spiegelkante und bei  $s$  um die Distanz zwischen dem realen Viewpoint  $O$  und dem virtuellen Viewpoint  $V$  (vgl. Abbildung 3.18).

$$t = \frac{r}{\sin(\theta)}$$

$$\frac{s}{2} = t \cdot \sin(\theta - \beta)$$

$$s = 2 \cdot \frac{r}{\sin(\theta)} \cdot \sin(\theta - \beta)$$

Mit Hilfe dieser Zwischengrößen lässt sich der Abstand entlang der optischen Achse zwischen realem und virtuellem Viewpoint berechnen:

$$v_d = \sin(\beta) \cdot s \quad (3.7)$$

Analog lässt sich der Abstand der virtuellen Viewpoints  $V$  von der optischen Achse berechnen:

$$v_r = \cos(\beta) \cdot s \quad (3.8)$$

Bei einem Öffnungswinkel von  $\beta < 0$ , lässt sich beobachten, dass für den Abstand  $v_d > 0$  gilt, das heißt die virtuellen Viewpoints in einem solchen System befinden sich vor der Kamera. Dies begünstigt die Aufnahme von Objekten, die sich näher am Spiegel oder gar im Spiegel selbst befinden. Bei einem Öffnungswinkel von  $\beta > 0$  rücken die virtuellen Viewpoints hinter die Kamera und Aufnahmen von weiter entfernten Objekten werden möglich ( $v_d < 0$ ). In dem Spezialfall  $\beta = 0$ , bei dem der Kegelstumpf zu einem Zylinder wird, befinden sich die virtuellen Viewpoints auf gleicher Höhe mit der realen Kamera ( $v_d = 0$ ).

## Sichtfeld

Das Sichtfeld eines radialen katadioptrischen Systems kann innerhalb eines Längsschnittes in verschiedene Bereiche aufgeteilt werden: den bifokalen und den trifokalen Bereich. Je nachdem wo sich ein beobachteter Punkt befindet, kann er einfach, zweifach oder dreifach abgebildet werden. Befindet er sich innerhalb des bifokalen Bereichs so wird er vom realen und einem der virtuellen Viewpoints aus gesehen. Liegt er sogar im trifokalen Bereich, kann er vom realen und zwei virtuellen Viewpoints aus gesehen werden, wie in Abbildung 3.19 zu erkennen ist.

Eine weitere geometrische Analyse ermöglicht es das Sichtfeld der virtuellen und des realen Viewpoints zu errechnen und so die Bereiche auszumachen, in denen ein bifokales beziehungsweise trifokales Sehen möglich ist. Das Sichtfeld  $\phi$  der virtuellen Viewpoints wird durch den dazugehörigen Spiegel bedingt:

$$\phi = \cot\left(\frac{2\cos(\theta - 2\beta) \sin(\theta) \sin(\theta - \beta)}{\sin(\theta - 3\beta) + 2 \sin(\theta) \cos(\theta - 2\beta) \cos(\theta - \beta)}\right) \quad (3.9)$$

Mit Hilfe des Winkels  $\phi$  lässt sich das Sichtfeld  $\psi$  des realen Viewpoints errechnen:

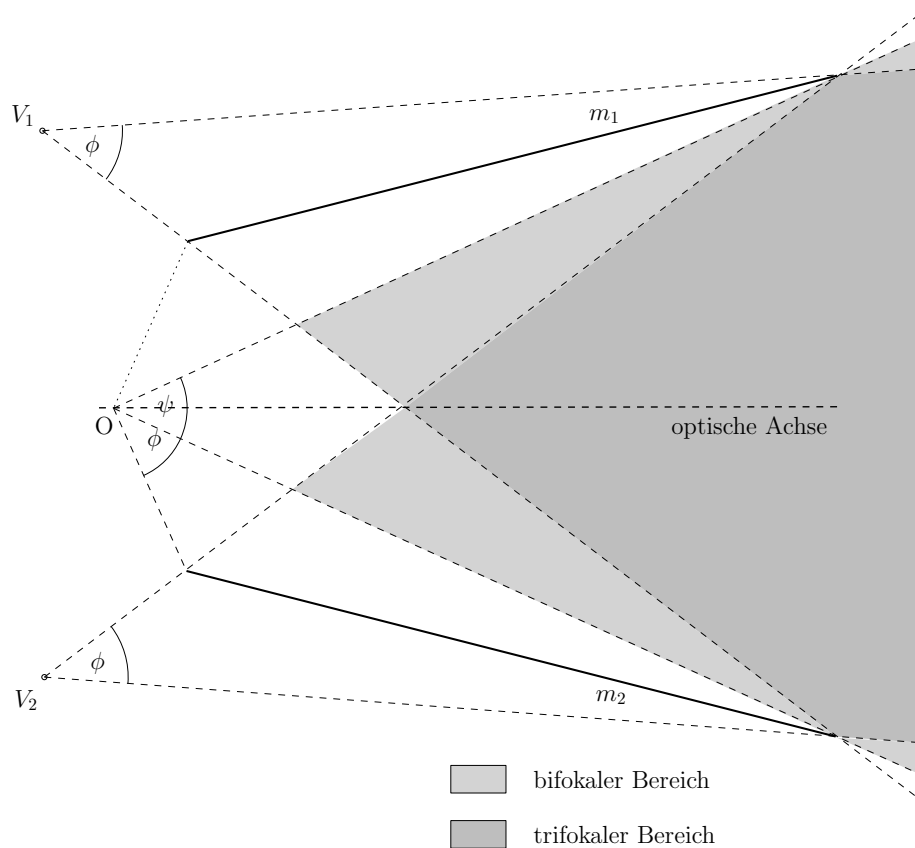


Bild 3.19: Sichtfeld teilt sich auf in bifokale und trifokale Bereiche

$$\psi = 2(\theta - \phi) \quad (3.10)$$

## Unschärfe

Neben den herkömmlichen Gründen für Unschärfe (*Defocus Blur*) in dioptrischen Systemen, kommen bei katadioptrischen Systemen mit gekrümmten Spiegeln zwei weitere Faktoren für Unschärfe hinzu. Zum einen führt die begrenzte Blendenöffnung [HZ05] und zum anderen die Krümmung des Spiegels [BN99] zu unscharfen Bereichen im Bild.

Bei der Unschärfe, die aus der begrenzten Blendenöffnung resultiert, handelt es sich um

Tiefenunschärfe (vgl. Abschnitt 2.1.5). Dies ist auf die stark variierende ideale Fokuseinstellung für unterschiedlich entfernte Punkte auf dem Spiegel zurückzuführen. Somit ist es schwierig, alle Bereiche der Szene gleichzeitig scharf zu stellen. Entweder ist der vordere oder der hintere Rand des Spiegels im Fokus, aber nicht beide gleichzeitig. Dieser Effekt lässt sich reduzieren, indem die Blende möglichst klein eingestellt wird, um die Schärfentiefe zu maximieren. Falls, wie in beiden hier realisierten Kameras, die Szene nicht nur indirekt durch den Spiegel, sondern auch direkt abgeleuchtet wird, muss die Schärfeneinstellung der direkten Aufnahme angepasst werden. Dies ist im besonderen Maße notwendig, wenn die direkte Aufnahme des Objekts im resultierenden 3D-Modell als Textur dienen soll.

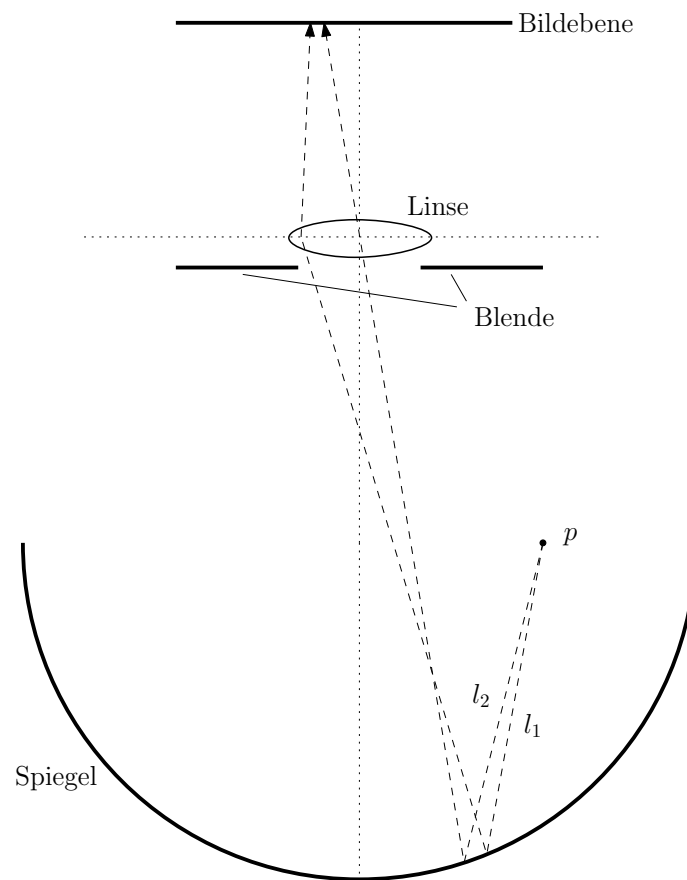


Bild 3.20: Ein Lichtstrahl wird auf unterschiedlichen Wegen auf den Sensor projiziert.



Durch die Krümmung des Spiegels und die Wirkung der Sammellinse des Objektivs der konventionellen Kamera werden von einem Weltpunkt ausgehende Lichtstrahlen mehrfach auf dem Sensor abgebildet (vgl. Abbildung 3.20). Dies führt dazu, dass ein Szenenpunkt nicht als Punkt, sondern als Kreisfläche abgebildet und somit als unscharf wahrgenommen wird. Die Lichtstrahlen  $l_1$  und  $l_2$  in Abbildung 3.20 gehen von demselben Weltpunkt aus, werden an verschiedenen Stellen des Spiegels reflektiert, und treffen, nachdem sie von der Linse eingefangen wurden, schließlich an leicht unterschiedlichen Stellen auf den Sensor. Der maximale Abstand zwischen zwei Punkten auf dem Sensor, die demselben Weltpunkt zuzuordnen sind, entspricht dem Durchmesser des Unschärfekreises an dieser Stelle. Eine Möglichkeit diesem *Field Curvature* genannten Effekt [HZ05] entgegen zu wirken, wäre es weitere Linsen im System zu platzieren. Allerdings gilt: Je kleiner der Spiegel, desto weniger tritt dieser Effekt auf. Diese Eigenschaft ließe sich nutzbar machen, wenn es darum ginge die Gesamtoptik zu miniaturisieren, um Linsen und Spiegel des Systems in einem Gesamtobjektiv zu vereinen.

### 3.5.2 Zylindrischer Spiegel $\beta = 0$

In diesem Abschnitt wird der Spezialfall  $\beta = 0$  behandelt, bei dem der Kegelstumpf zu einem Zylinder wird. Das entstehende System zeichnet sich durch besondere Eigenschaften und die sich daraus ergebenden Anwendungsgebiete aus.

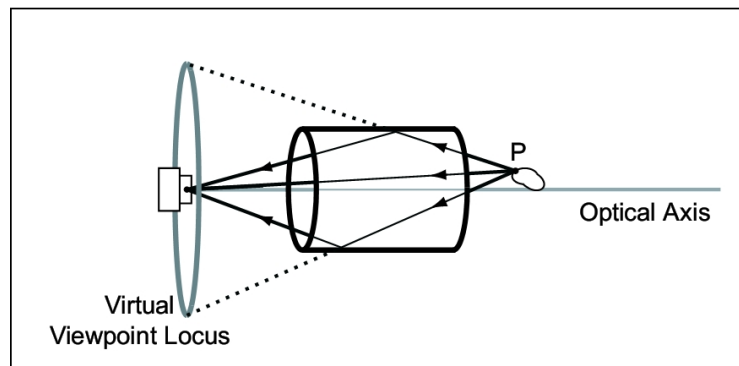


Bild 3.21: System mit  $\beta = 0$  (aus [KN06])

### Eigenschaften

Da bei diesem Spezialfall  $\beta = 0$  gilt, vereinfachen sich die Gleichungen 3.7 und 3.8 zu  $v_d = 2r$  und  $v_r = 0$ . Das bedeutet, dass die zirkuläre Ortslinie der virtuellen Viewpoints den realen Viewpoint als Zentrum besitzen und ihr Radius dem doppelten Radius des Zylinders entspricht.

### Einsatzmöglichkeiten zylindrischer Spiegel

In diesem Abschnitt sollen mögliche Anwendungsgebiete für Systeme mit einem Öffnungswinkel von  $\beta = 0$  beschrieben werden. Eine Möglichkeit liegt in der Approximation der bidirektionalen Reflektanzverteilungsfunktion (BRDF). Die BRDF eines Materials wird in vielen Anwendungen der Computergraphik und der Bildverarbeitung benötigt. Ihre Approximation ist aufgrund der Tatsache, dass das Material aus vielen verschiedenen Blickwinkeln aufgenommen werden muss, ein aufwendiger und komplizierter Prozess. Mit einem katadioptrischen System mit einer zylindrischen Optik ist es möglich, dies in nur einer Aufnahme zu realisieren. Hierfür wird ein Muster des Materials durch die zylindrische Optik aufgenommen, welches in der verspiegelten Innenseite vielfach reflektiert wird<sup>5</sup>. Diese Reflexionen entstehen jeweils aus einer anderen Blickrichtung auf das Material und damit wird es möglich die BRDF dieses Materials aus einer Aufnahme abzuschätzen (vgl. Abbildung 3.22).

Ein weiteres Anwendungsgebiet ist die Rekonstruktion der 3D-Textur von Oberflächen. Raue oder leicht unebene Oberflächen, die von einem „normalen“ 3D-Scanner nicht erfasst werden würden, können so erfasst werden. In gewissem Sinne handelt es sich um eine Stereokamera mit Makroobjektiv. Hiermit ist es beispielsweise möglich mit nur einer Aufnahme die Oberflächenstruktur von Brot oder der Rinde eines Baums zu erfassen (vgl. Abbildung 3.23).

---

<sup>5</sup>Ein Fisheye-Objektiv wird benötigt, damit durch den großen Bildwinkel  $\theta$  die Lichtstrahlen möglichst häufig im Spiegel zu reflektieren.

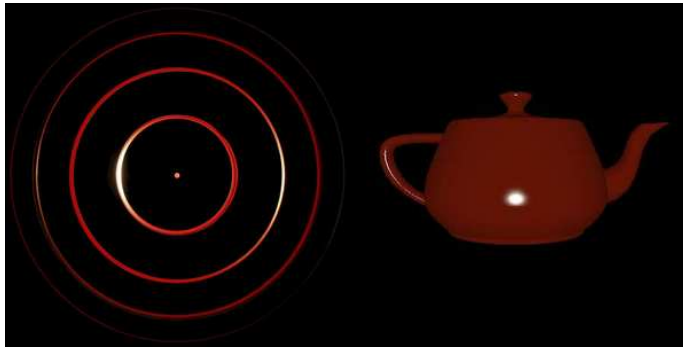


Bild 3.22: Approximation der BRDF eines Materials (aus [KN06]). links: Rohaufnahme der Materialprobe. rechts: *Utah Teapot* mit den gewonnenen Materialeigenschaften.

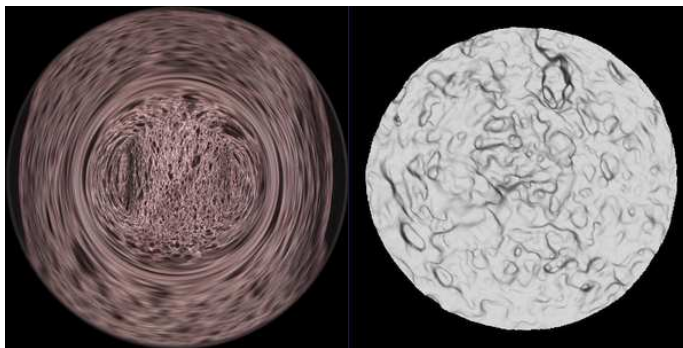


Bild 3.23: 3D-Textur der Oberfläche einer Brotscheibe (aus [KN06]). links: Rohaufnahme der Brotscheibe. rechts: Rekonstruierte Oberflächenstruktur.

### 3.5.3 Konischer Spiegel mit $\beta \neq 0$

In diesem Abschnitt werden die Eigenschaften von konischen Spiegeln beschrieben. Die Anwendungsgebiete und Eigenschaften unterscheiden sich weiter je nachdem, ob  $\beta$  kleiner oder größer Null ist. In der Konzeption dieser Kategorie von katadioptrischen Systemen bieten sich zwei Vorteile gegenüber zylindrischen Spiegeln. Zum einen sind diese Systeme flexibler in der Positionierung der zirkulären Ortslinien der virtuellen Viewpoints und zum anderen bietet sich die Möglichkeit den Spiegel umzudrehen, sodass daraus ein anderes katadioptrisches System mit völlig anderen Anwendungsmöglichkeiten entsteht.

### Eigenschaften

Die Position der virtuellen Ortslinie liegt bei Systemen mit einem Öffnungswinkel von  $\beta > 0$  zwischen der realen Kamera und  $v_d = -r \tan(\theta/2)$ , also hinter der realen Kamera. Der maximale Abstand des virtuellen Viewpoints zur realen Kamera wird erreicht, wenn der Öffnungswinkel  $\beta = \frac{\theta}{2}$  ist. Der Bereich des trifokalen Sehens rückt so weiter von der realen Kamera weg, was diese Variante dazu befähigt Objekte aufzunehmen, die sich vor dem Gesamtsystem befinden. Das Sichtfeld der virtuellen Viewpoints ist in dieser Konfiguration größer als das des realen Viewpoints.

Bei Systemen mit einem Öffnungswinkel von  $\beta < 0$  liegt die Ortslinie zwischen der realen Kamera und  $v_d = r \tan(\theta/2)$ , also vor der Kamera. Hier wird der maximale Abstand analog bei  $\beta = -\frac{\theta}{2}$  erreicht. Somit rückt der Bereich des trifokalen Sehens näher an die Kamera heran und es wird möglich Objekte aufzunehmen, die sich innerhalb des Spiegels befinden. Das Sichtfeld der virtuellen Viewpoints ist bei einem Öffnungswinkel von  $\beta < 0$  kleiner als das der realen Kamera.

### Einsatzmöglichkeiten konischer Spiegel mit $\beta < 0$

In diesem Abschnitt soll auf die Anwendungsmöglichkeiten eines Spiegels mit  $\beta < 0$  eingegangen werden. Eine Möglichkeit besteht darin, die vollständige Textur eines konvexen Objektes in einer Aufnahme zu erfassen. Die Ortslinie der virtuellen Viewpoints muss vor der Kamera liegen und größer als der kleinstmögliche Zylinder, der das Objekt umschließt, sein, damit es von allen Seiten von den virtuellen Viewpoints erfasst werden kann. Systeme mit einem Öffnungswinkel von  $\beta < 0$  können diese Bedingungen erfüllen. Das Objekt wird hierfür so innerhalb des Spiegels platziert, dass die Achse des minimalen umschließenden Zylinders des zu scannenden Objekts mit der optischen Achse des Gesamtsystems übereinstimmt. Nur so ist es möglich das Objekt von allen Seiten zu erfassen (vgl. Abbildung 3.24). Falls das Objekt auf dem Spiegelboden steht, kann die Unterseite selbstverständlich nicht erfasst werden. Das resultierende Bild hat die aus Abbildung 3.15 bekannte Gestalt. In der direkten Ansicht wird die Front des Objekts und im Spiegelring die Seiten und falls sichtbar die Unterseite abgebildet (vgl. Abbildung 3.25).

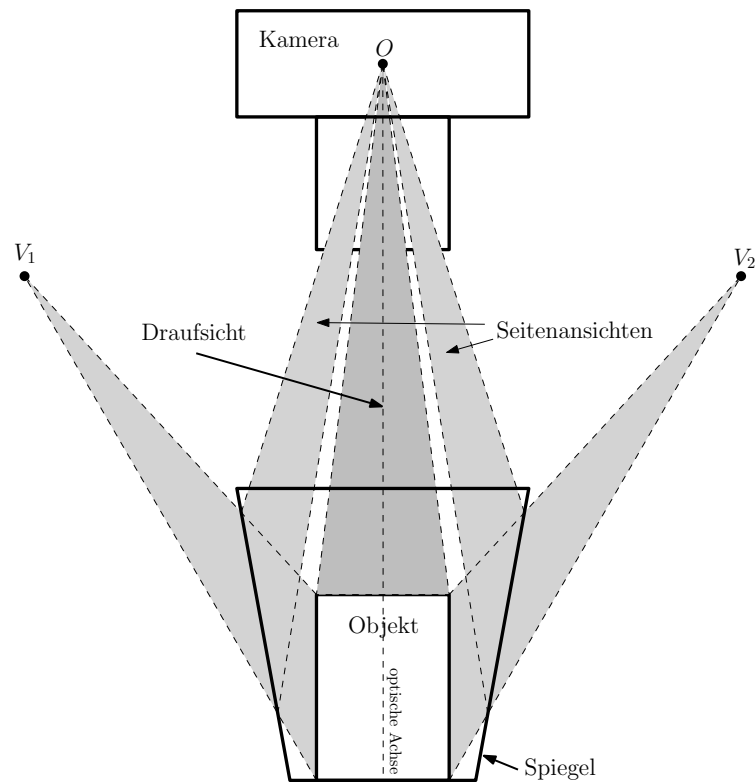
Bild 3.24: System mit  $\beta < 0$ 

Bild 3.25: Erfassen der vollständigen Textur eines konvexen Objektes (aus [KN06]). links: Rohaufnahme des Objektes im Spiegel. rechts: 3D-Modell mit der gewonnenen Textur.

Wird dieses Verfahren zweimal mit einer Parallaxe entlang der optischen Achse zwischen den beiden Aufnahmen von etwa einem halben Zentimeter ausgeführt, so ist es möglich die vollständige Geometrie eines Objekts zu erfassen (vgl. Abbildung 3.26). Dafür ist es notwendig Kamera und Spiegel auf einer Halterung zu befestigen, die es ermöglicht diese Translation entlang der z-Achse relativ präzise durchzuführen. Die für diese Diplomarbeit hergestellte Halterung käme für dieses Verfahren in Frage (siehe Abschnitt 4.4).

Die Epipolarlinien der resultierenden Bilder sind wie in Abschnitt 3.5 beschrieben radial. Um konventionelles Stereo Matching einsetzen zu können, müssen die sternförmig verlaufenden Linien in horizontale Linien umgewandelt werden. Dies kann erreicht werden, indem die Bilder vom kartesischen ins Polarkoordinatensystem transformiert werden. Punkte, die sich im kartesischen Koordinatensystem auf einer radialen Linie befinden, werden im Polarkoordinatensystem auf horizontale Linien abgebildet. Bevor die Bilder *gematcht* werden können, müssen sie noch rektifiziert werden, da die Spiegellinien nicht orthogonal zur Bildebene stehen (vgl. Abschnitt 3.4.3 und 4.4.4).

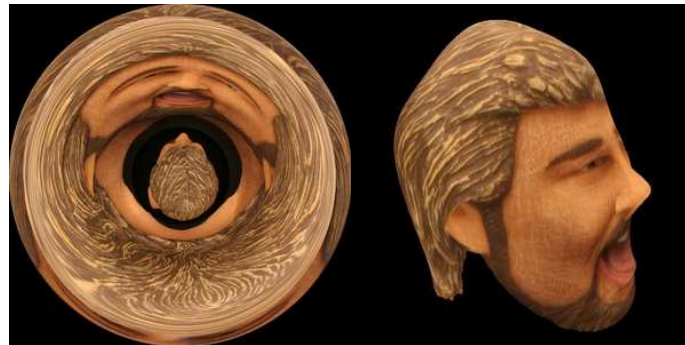
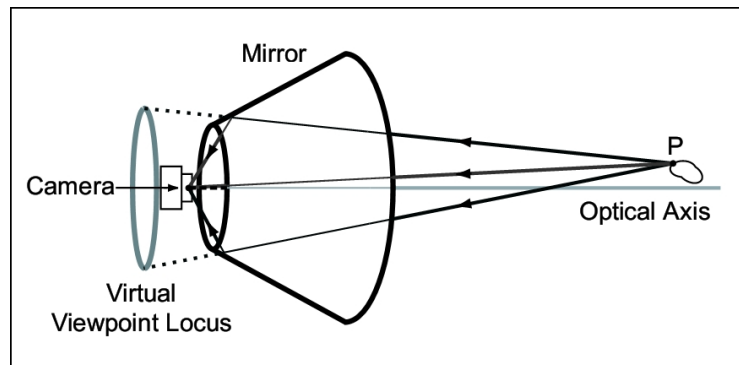


Bild 3.26: Erfassen der Geometrie eines konvexen Objektes (aus [KN06]). links: Eine der Rohaufnahmen des Objektes im Spiegel. rechts: Rekonstruiertes 3D-Modell.

Ein Nachteil dieser beiden Verfahren ist, dass die Größe des Objekts durch die Größe des Spiegels begrenzt ist, um größere Objekte aufnehmen zu können, würden zunehmend größere Spiegel benötigt. Solche Spiegel wären schwieriger herzustellen und damit auch kostenintensiver.

**Einsatzmöglichkeiten konischer Spiegel mit  $\beta > 0$** Bild 3.27: System mit  $\beta > 0$  (aus [KN06])

In diesem Abschnitt wird nur kurz auf das Haupteinsatzgebiet dieses Spiegels eingegangen, da es eins der realisierten Verfahren zur Rekonstruktion von 3D darstellt und in Abschnitt 4.4 im Detail vorgestellt wird. In dieser Variante liegt die Ortslinie der virtuellen Viewpoints hinter der Kamera, weshalb es sich hervorragend als Nahbereichsscanner eignet. Die Größe des Objekts wird in diesem System lediglich durch das Sichtfeld der jeweiligen Konfiguration begrenzt (vgl. Abbildung 3.19).

**3.6 Andere Formen**

Es existieren noch viele weitere Arten von katadioptrischen Stereosystemen, auf die in dieser Diplomarbeit nicht näher eingegangen werden kann. Eine allgemeine Übersicht über mögliche Systemkonfigurationen gibt [BN99]. Eine Einführung in Systeme mit mehreren gekrümmten Spiegeln, so genannte *gefaltete katadioptrische Systeme*, kann in [NP99] gefunden werden.





# Kapitel 4

## Bau und Implementierung

### 4.1 Übersicht

Zur Evaluierung der Leistungsfähigkeit von katadioptrischen Stereosystemen wurden zwei vielversprechende Ansätze näher untersucht: ein System mit einem Planspiegel und eines mit einem gekrümmten Spiegel.

Das Planspiegelsystem liefert aufgrund der Anordnung von Kamera und Spiegeln direkt rektifizierte Ergebnisse und weist dadurch einen nur geringen Nachbearbeitungsaufwand auf (vgl. Abschnitt 3.4.3). Es werden lediglich die beiden relevanten Rechtecke aus der Aufnahme extrahiert, die im Spiegel entstandene Aufnahme des Objekts entsprechend gespiegelt und anschließend mit konventionellen Stereo Matching Verfahren in eine Tiefenkarte überführt (siehe Abschnitt 4.3). Dieses Verfahren kann wohl als das einfachste katadioptrische Stereosystem bezeichnet werden. Die Schwierigkeit liegt hier vor allem im Systemaufbau.

Bei dem System mit gekrümmtem Spiegel wird dieser in Form eines innen verspiegelten konischen Rohrs vor eine Kamera mit einem Fisheye-Objektiv angebracht. Die Szene wird einmal direkt durch die Öffnung des Rohrs und ein weiteres Mal in der Spiegelung aufgenommen (vgl. Abschnitt 3.5). Da die genauen Dimensionen des Spiegels und die Abbildungsfunktion des Fisheye-Objektivs bekannt sind, können die Verzerrungen der Spie-

gelung und die Fisheye-Verzerrungen herausgerechnet werden und es entstehen neben der direkten Ansicht zwei weitere Ansichten. Diese werden nach einigen Bearbeitungsschritten wiederum mit konventionellen *Stereo-Matching*-Verfahren in zwei Tiefenkarten überführt. Beiden Tiefenkarten sind an der direkten Ansicht ausgerichtet und werden zu einer finalen Tiefenkarte zusammengeführt (siehe Abschnitt 3.5).

## 4.2 Realisierte System

### 4.2.1 System mit Planspiegel

In diesem Abschnitt soll die Komponentenwahl für das Planspiegel-Stereosystem erläutert werden. Ich habe mir zum Ziel gesetzt, dieses System möglichst einfach und kostengünstig zu gestalten. Es soll leicht nachzubauen sein und das Potential katadioptrischer Systeme demonstrieren, eine Stereokamera mit relativ geringem technischem Aufwand erstellen zu können.

Um dieses Ziel zu erreichen, habe ich mich für einen handelsüblichen Haushaltsspiegel und die digitale Kompaktkamera Olympus C-300 Zoom entschieden. Mit einer Sensorgröße von  $5,75 \times 4,31$  mm und einem Brennweitenbereich von 5,8 - 16,2 mm kann ein völlig ausreichender horizontaler Bildwinkel von  $52,7^\circ$  erreicht werden. Ob die Auflösung von  $1984 \times 1488$  für ein Ein-Planspiegel-Stereosystem ausreichend ist, muss sich in anschließenden Tests zeigen. Die für Kompaktkameras übliche kleine Blendenöffnung stellt kein Problem dar. Um Bewegungsunschärfe zu vermeiden, sollte das Objekt jedoch gut ausgeleuchtet sein, damit die Belichtungszeit nicht zu lang ist.

### 4.2.2 System mit gekrümmtem Spiegel

In diesem Abschnitt soll erläutert werden, welche Komponenten für das Stereosystem mit gekrümmtem Spiegel ausgesucht wurden.

Ich habe einen konischen Spiegel mit einem Öffnungswinkel  $\beta \neq 0$  gewählt, da dieser aufgrund der unterschiedlichen Anbringungsmöglichkeiten viele Einsatzmöglichkeiten bietet

(siehe die Abschnitte 3.5.3 und 3.5.3). Ein solcher Spiegel wurde im Labor für Mikrospannung der Universität Bremen aus der Aluminiumlegierung AlMg3 hergestellt. Der Innenkegel wurde durch Diamantdrehen mit einer optischen Oberflächengüte von  $R_a^1 < 10$  nm versehen. Das zum Einsatz kommende innen verspiegelte Rohr, bei dem es sich geometrisch um einen Kegelstumpf handelt, hat eine Länge von 12,7 cm sowie die Radien 3.4 cm und 7.4 cm und daraus resultierend einen halben Öffnungswinkel  $\beta$  von  $17.48^\circ$  (vgl. Konstruktionszeichnung C.1).

Da ein Nahbereichsscanner aufgebaut werden soll, muss der Spiegel so angebracht werden, dass der Öffnungswinkel größer Null ist, das Spiegelrohr öffnet sich also in Blickrichtung. Für dieses System werden eine digitale Spiegelreflexkamera und ein Fisheye-Objektiv benötigt. Hier habe ich mich für eine Nikon D60 mit einer Auflösung von 10 Megapixeln und das Sigma 8mm F3.5 EX DG Objektiv entschieden. Eine relativ hohe Auflösung der Kamera ist für katadioptrische Stereosysteme im Allgemeinen von Bedeutung, da sich alle Ansichten den verfügbaren Platz auf dem Sensor teilen müssen. Das Sigma Objektiv zeichnet sich durch eine gute Umsetzung der flächentreuen Fisheye-Abbildungsfunktion (*Equisolid Angle*) (vgl. Abschnitt 2.1.6) und eine geringe chromatische Aberration aus.

Mit dieser Kombination aus Kamera und Objektiv ist eine Autofokussierung der Szene nicht möglich, was jedoch kein Problem darstellt, da aufgrund des sich nur geringfügig ändernden Objektstands und der großen Schärfentiefe (siehe Abschnitt 3.5.1 und 4.4.5 warum die Schärfentiefe möglichst groß sein sollte) nur selten die Fokussierung nachjustiert werden muss. Die manuelle Fokussierung gewährleistet darüber hinaus besser vergleichbare Aufnahmen und sollte deshalb ohnehin genutzt werden, wenn es die Dynamik der Szene zulässt. Zumindest jedoch sollten die Fokussierungspunkte der Kamera fixiert werden, da Messungen an Spiegelpunkten zu falschen Ergebnissen führen können (siehe Abschnitt 3.5.1).

---

<sup>1</sup> $R_a$  ist der so genannte Mittenrauhwert, das arithmetische Mittel der Abweichungen von der Mittellinie.

### 4.3 Rektifiziertes Planspiegel-Stereosystem

In diesem Abschnitt soll der Aufbau und die Implementierung einer katadioptrischen Stereokamera beschrieben werden, die mit Hilfe eines Planspiegels einen weiteren Viewpoint der Szene erfasst und dadurch die Möglichkeit bietet Tiefe zu rekonstruieren. Spiegel und Kamera werden so platziert, dass die Aufnahmen bereits in ihrer Rohform rektifiziert sind. Dadurch entsteht eine einfache katadioptrische Stereokamera, deren Bilder nur geringfügig nachbearbeitet werden müssen und sich diese Klasse von Stereokameras somit besonders für Echtzeitanwendungen eignet. Es wird ein Ein-Spiegel-Stereosystem aufgebaut, das durch seine simple Konfiguration besonders leicht nachgebaut werden kann.

#### Ausrichten des Gesamtsystems

In diesem Abschnitt wird die Ausrichtung eines rektifizierten Planspiegel-Stereosystems mit einfachsten Mitteln erläutert. Hierfür wird eine Kompaktkamera an einem Stativ befestigt und vor dem auf einer waagrechten Fläche liegenden Spiegel positioniert. Das Gesamtsystem muss so ausgerichtet werden, dass die Bedingungen für direkt rektifizierte Aufnahmen in einem katadioptrischen System (vgl. Abschnitt 3.4.3) erfüllt sind. Da nur ein Sensor verwendet wird, genügt es, wenn die optische Achse und die Scanlinien der Kamera parallel zur Spiegelebene verlaufen. Der Spiegel muss lediglich so positioniert werden, dass das zu scannende Objekt ganz darin abgebildet wird.

Da es keine Rolle spielt, ob die x- oder y-Achse der Kamera am Spiegel ausgerichtet wird, kann die Kamera auch im Hochformat angebracht werden, um den größtmöglichen Bildwinkel zu nutzen. Zum Erreichen dieser Bedingung werden zunächst die Scanlinien mit Hilfe eines Anschlagwinkels justiert. Hierfür wird ein Bild des auf dem Spiegel stehenden Winkels gemacht. Ziel ist es die Kamera so auszurichten, dass der Winkel parallel zu den Scanlinien der Kamera verläuft (vgl. linke Abbildung 4.1). Am einfachsten gelingt dies, indem die Aufnahme auf dem Display der Kamera vergrößert und das Bild so ausgerichtet wird, dass der Winkel sich am Rand des Displays befindet. Auf diese Weise ist leicht zu erkennen, ob sich der Winkel parallel zum Bildrand beziehungsweise den Scanlinien der Kamera befindet. Tut er dies nicht, muss solange nachgebessert werden, bis diese Bedin-

gung erfüllt ist.



Bild 4.1: links: Ausrichten der Scanlinien an der Spiegelebene (senkrechte weiße Linie zur Orientierung), rechts: „Zielvorrichtung“ zur Ausrichtung der optischen Achse.

Um die Neigung der Kamera so auszurichten, dass die optische Achse parallel zur Spiegelebene verläuft, wird das Prinzip einer Zielvorrichtung genutzt. Sehr gut eignet sich hierfür eine auf den Spiegel gestellte Leiste<sup>2</sup> (vgl. rechte Abbildung 4.1). Die Kamera muss in Höhe und Neigungswinkel ausgerichtet werden, sodass die obere Kante der Leiste gerade nicht mehr zu sehen ist, sie also parallel zur optischen Achse der Kamera verläuft. Eine an der gegenüberliegenden Seite der Leiste angebrachte Stecknadel erfüllt hierbei zwei Aufgaben. Zum einen erleichtert sie es die richtige Position der Kamera zu erkennen und zum anderen kann die Position des farbigen Stecknadelkopfs mittels Farbsegmentie-

<sup>2</sup>Eine zweite identische Leiste oder ein Gegenstand mit einer größeren Oberseite wären noch besser für diese Aufgabe geeignet.

rung<sup>3</sup> ermittelt und somit die Position der Hauptachse des Systems bestimmt werden (vgl. Abbildung 4.2). Die Hauptachse ist zur Ermittlung der Disparität der korrespondierenden Punkte des entstehenden Stereobildpaares von Bedeutung (siehe Abschnitt 4.3.2).

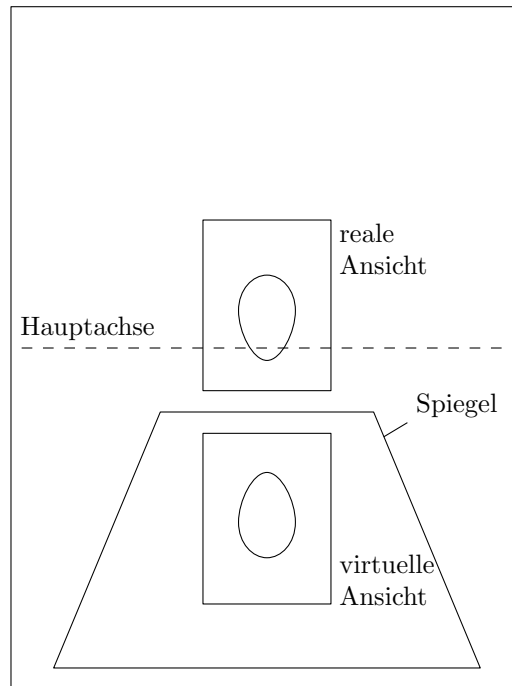


Bild 4.2: Schema der Rohaufnahme eines rektifizierten Planspiegel-Stereosystems

Ist die Kamera ausgerichtet, so entspricht die Höhe der Leiste  $l_h$  darüber hinaus der Entfernung zwischen Kamera und Spiegel. Die Stereobasis  $b_{mm}$  des Gesamtsystems folgt somit aus Formel 3.4 und entspricht der doppelten Höhe der Leiste (vgl. Abschnitt 3.4.3 und Abbildung 3.12):

$$b_{mm} = 2 \cdot l_h \quad (4.1)$$

<sup>3</sup>Es sollte hierbei ein einfarbiger Hintergrund mit einem starken Kontrast zur Farbe des Stecknadelkopfs gewählt werden.

### 4.3.1 Extraktion der Ansichten

In diesem Abschnitt wird die Extraktion der Ansichten erläutert. Im Folgenden wird davon ausgegangen, dass sich der Spiegel wie in Abschnitt 4.3 beschrieben, auf einer horizontalen Ebene vor der im Hochformat angebrachten Kamera befindet (vgl. Abbildung 4.2). In einer Rohaufnahme dieses Systems befinden sich die reale und die im Spiegel entstandene virtuelle Ansicht der Szene. Da die optische Achse der Kamera parallel zum Spiegel verläuft, liegt dieser immer vollständig in der unteren Hälfte der Rohaufnahme. In Folge dessen ist der Bildwinkel der virtuellen Kamera durch die Länge des Spiegels begrenzt (vgl. Abschnitt 3.4.3).

Zunächst muss die Lage der Bildausschnitte ermittelt werden. Diese können durch die Position ihrer oberen linken Ecke und ihrer Ausmaße beschrieben werden und unterliegen folgenden Bedingungen: Einerseits müssen die Höhe sowie die Breite der Ansichten für das anschließende *Stereo Matching* identisch sein und andererseits müssen die x-Positionen der beiden Rechtecke ebenfalls identisch sein, damit die Rektifizierung intakt bleibt. Außerdem muss die y-Position der realen Ansicht so gewählt werden, dass der Bildausschnitt mit der Hauptachse abschließt. Weiterhin entspricht die gespiegelte Ansicht in dieser Konfiguration der unteren Kamera des Stereosystems und sollte deshalb die Szene leicht nach oben versetzt zeigen. Dieser Versatz kann so gewählt werden, dass der Suchraum des *Stereo Matchings* minimal wird, da die eigentliche Disparität von der Hauptachse gemessen wird. Das heißt, befindet sich die Hauptachse an der y-Position  $h$  und die Lage der virtuellen Ansicht lautet  $v_x \times v_y$  mit  $v_x > h^4$ , so muss der Abstand zur Hauptachse  $v_x - h$  auf die gefundene Disparität  $d$  aufaddiert werden. Für die tatsächliche Disparität  $d'$  gilt also:

$$d' = d + (v_x - h) \quad (4.2)$$

Sind die Positionen der Ansichten bekannt, so handelt es sich bei der Extraktion lediglich um das Ausschneiden dieser rechteckigen Bildausschnitte aus der Rohaufnahme. Die

---

<sup>4</sup>In der beschriebenen Konfiguration mit einem Bild, dessen Ursprung sich links oben befindet, gilt diese Bedingung stets, da es eines unendlich großen Spiegels bedürfte, damit er bis zur Hauptachse heranreichen würde.

virtuelle Ansicht muss noch gespiegelt werden, da sie durch eine Reflexion an einer ungeraden Spiegelanzahl entstanden und deshalb spiegelverkehrt ist (vgl. Abschnitt 3.4.1).

### 4.3.2 Stereo Matching

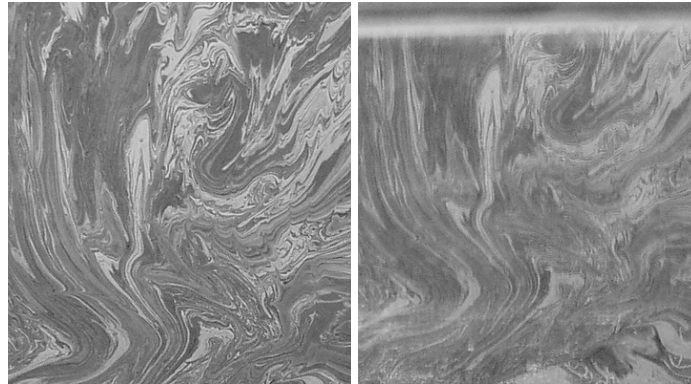


Bild 4.3: links: reale Ansicht eines parallel zur Bildebene stehenden Musters, rechts: gespiegelte virtuelle Ansicht

Nach Extraktion der Bildausschnitte und spiegeln der virtuellen Ansicht an der Hauptachse könnten sie direkt mit konventionellen Stereo Matching Verfahren zu einer dichten Tiefenkarte transformiert werden. Die Ansichten unterscheiden sich im vorliegenden System jedoch relativ deutlich voneinander (vgl. Abbildung 4.3). Für diese Unterschiede ist der in diesem Aufbau verwendete Haushaltsspiegel verantwortlich. Die spiegelnde Oberfläche befindet sich bei dieser Sorte von Spiegeln unter einer Glasscheibe, was vor allem bei einem flachen Betrachtungswinkel zu Unschärfen führt. Da es nur schwer möglich ist diese Unschärfe der virtuellen Ansicht zu reduzieren, wird die direkte Ansicht mit einer leichten Gauß-Unschärfe versehen. Dies gleicht die beiden Ansichten einander an und verbessert so das *Stereo Matching*. Ein weiteres Problem entsteht durch die mangelnden Reflexionseigenschaften des Spiegels, er absorbiert einen Teil des Lichts, weshalb die virtuelle Ansicht dunkler erscheint. Dieser Helligkeitsunterschied führt dazu, dass nur normalisierende Korrelationsverfahren für das *Stereo Matching* in Frage kommen. Abbildung 4.4 zeigt eine im ZNCC-Verfahren (vgl. Abschnitt 2.2.3) entstandene Tiefenkarte eines parallel zur Bildebene stehenden Musters.



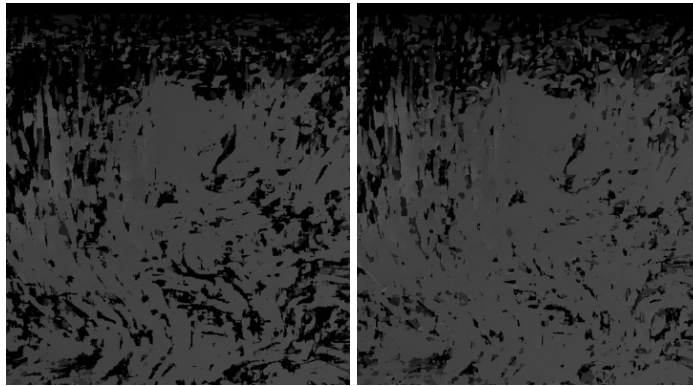


Bild 4.4: Tiefenkarte des Musters, links: mit unveränderter realer Ansicht, rechts: mit Gauß-Unschärfe versehene reale Ansicht

## 4.4 Radiales katadioptrisches Stereosystem

In diesem Abschnitt wird der Aufbau und die Implementierung einer katadioptrischer Stereokamera beschrieben, die mit Hilfe eines konischen Spiegels aus mehreren Viewpoints die Szene erfasst und es dadurch ermöglicht Tiefe zu rekonstruieren. Im Gegensatz zu dem zuvor vorgestellten System, welches direkt rektifizierte Bilder liefert und dadurch äußerst ressourcenschonend arbeitet, bedarf es bei diesem System deutlich mehr Nachbearbeitung. Das liegt zum einen an der Tatsache, dass die von den virtuellen Viewpoints aufgenommenen Ansichten der Szene mit einem komplexen Verfahren extrahiert werden müssen. Zum anderen entstehen hier zwei anstatt nur einer weiteren Ansicht. Es müssen pro Bild zwei Tiefenkarten erstellt werden, die dann in eine gemeinsame Tiefenkarte zusammengeführt werden müssen.

Die Szene wird hierbei direkt durch die Öffnung des Rohrs und indirekt durch die Spiegelung aufgenommen. Das System hat somit einen realen Viewpoint und unendlich viele virtuelle Viewpoints. Diese Viewpoints liegen auf einer Ortslinie (vgl. Abbildung 3.27) mit der optischen Achse als Mittelpunkt.

Die in den folgenden Abschnitten erläuterte Vorgehensweise soll zur leichteren Orientierung in einer kompakten Übersicht dargelegt werden:

Vorgehensweise:

1. Extraktion der relevanten Bereiche
2. Fisheye-Entzerrung
3. Aufteilen in verschiedene Ansichten
4. Rektifizierung
5. Dense Stereo Matching

## **Halterung**

Die Halterung ist ein selbstgebauter Prototyp aus Aluminium, Kunststoff und Holz, der eine Ausrichtung der Kamera in vier Freiheitsgraden ermöglicht (vgl. Abbildung 4.5). Die Verschiebung entlang der x-, y- und z-Achse, sowie eine Rotation um die y-Achse sind möglich. Wobei sich letztere als problematisch erwies, da das Stativgewinde, mit dem die Kamera an der Halterung befestigt wird, nicht deckungsgleich mit der y-Achse der Bildebene ist. Weitere Informationen hierzu können im Abschnitt über die Ausrichtung des Gesamtsystems 4.4 gefunden werden. Ebenfalls problematisch ist die fehlende Möglichkeit um die x-Achse rotieren zu können. Da die Kamera allerdings schon fast richtig rotiert in der Halterung sitzt, bedarf es nur einer sehr kleinen Rotation, die glücklicherweise durch Anziehen einer Feststellschraube zu erreichen war, die eigentlich für die Fixierung der Translation entlang der y-Achse gedacht war. Dieses Manko ließe sich beheben, indem ein Stativkopf als Befestigung für die Kamera verwendet werden würde. Eine Rotation um die z-Achse ist bei dem achsensymmetrischen Gesamtsystem nicht notwendig.

Des Weiteren ist es möglich den gesamten Spiegel um  $180^\circ$  zu drehen, um durch die große Öffnung des konischen Rohrs fotografieren zu können. Somit ist die Halterung auch für andere Aufgaben, wie beispielsweise die Extraktion der Textur eines im Spiegelbauteil befindlichen Objektes oder Ermittlung dessen vollständiger Geometrie, geeignet.

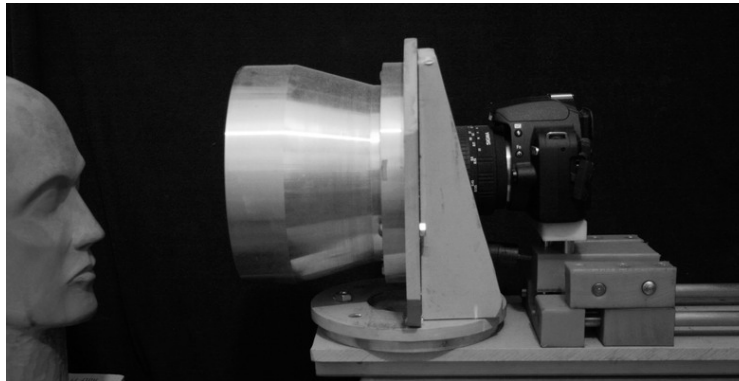


Bild 4.5: Halterung des radialen Systems

### Ausrichten des Gesamtsystems

Wie schon im Abschnitt 4.4 über die Halterung dargelegt, ist die Ausrichtung der Rotation um die  $y$ -Achse der Kamera besonders schwierig, da das Stativgewinde der Kamera und damit die Rotationsachse der Kamera nicht deckungsgleich mit der  $y$ -Achse der Bildebene ist. Es handelt sich bei einer Drehung um diese Achse zusätzlich auch immer um eine Translation, wodurch die Kalibrierung des Systems deutlich erschwert wird. Als hilfreich hat sich herausgestellt, die Kamera plan an das Spiegelbauteil heranzufahren und so die Orientierungen aneinander anzugleichen. Möglich ist dies nur wenn der Adapterring des Fisheye-Objektivs aufgesetzt ist, da sonst die hervorstehende äußere Linse des Fisheye-Objektivs den Spiegel berühren und dadurch zerkratzen würde. Der Adapterring erweist sich aber auch sonst als nützlich, da so weniger Licht von hinten ins System gelangen kann. Haben Kamera und Spiegel die gleiche Orientierung, so muss die Kamera zunächst entlang der  $z$ -Achse so verschoben werden, dass sich Kamera und Spiegel nicht mehr berühren, damit die Ausrichtung der anderen Achsen die Orientierung der beiden Bauelemente nicht beeinflusst.

Für die Ausrichtung der  $x$ - beziehungsweise der  $y$ -Achse ist es notwendig ein Kalibrierungsmuster (siehe Abbildung 4.6) an der Öffnung des Spiegels anzubringen. Um eine möglichst einfache und genaue Ausrichtung des Musters zu gewährleisten, wurde es an der Innenseite eines Rings angebracht, der wiederum genau auf die Öffnung des Spiegels passt. Die konzentrischen Kreise werden nur dann ebenfalls als konzentrische Kreise

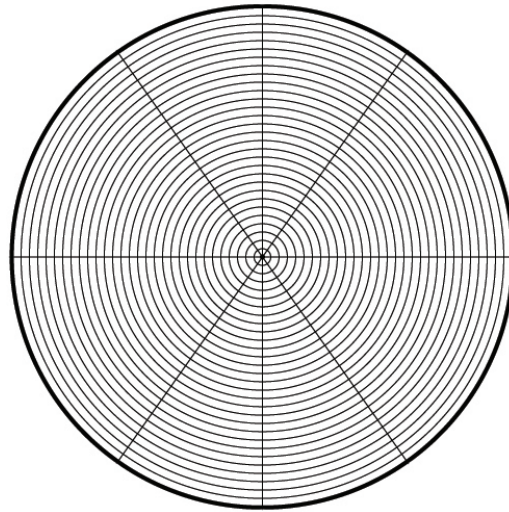


Bild 4.6: Kalibrierungsmuster (Quellcode: siehe Anhang B.1.1)

durch den achsensymmetrischen Spiegel reflektiert, wenn der Mittelpunkt des Musters die optische Achse des Spiegels schneidet.

Weist der äußerste Kreis des Musters einen konstanten Abstand zu seiner Reflexion auf, so sitzt das Muster richtig auf dem Spiegel. Alternativ kann der korrekte Sitz des Muster überprüft werden, indem eine Aufnahme in ihre Ansichten, wie in Abschnitt 4.4.3 beschrieben, getrennt wird. Bei diesem Vorgang werden die durch den Mittelpunkt verlaufenden Epipolarlinien extrahiert und zu den verschiedenen Ansichten aufgestapelt. Konzentrische Kreise werden somit als parallele Linien abgebildet. Ist dies bei der Probeaufnahme der Fall, so ist die Bedingung für den richtigen Sitz des Musters erfüllt. Damit liegt der Mittelpunkt des Musters auf der optischen Achse des Spiegels.

Um die optische Achse der Kamera an die des Spiegels anzugleichen, müssen noch die durch den Mittelpunkt des Musters verlaufenden Linien in ihrer Fortsetzung im Spiegel ebenfalls eine Linie bilden. Zur Veranschaulichung soll Abbildung 4.12 dienen, das aus der Kalibrierung des nicht-orthogonal stehenden Spiegels stammt, aber die gleiche Charakteristik aufweist wie das Kalibrierungsmuster. Es hat sich als hilfreich erwiesen mit der horizontalen Ausrichtungslinie zu beginnen und dann sukzessive die anderen Linien auszurichten.

Sind x- und y-Achsen so ausgerichtet, dass beide Bedingungen erfüllt sind, muss die Kamera so weit entlang der z-Achse transliert werden, dass das Verhältnis zwischen dem Sichtfeld der virtuellen und der realen Kamera 1:2 ist. Diese Reihenfolge der Kalibrierung kann nur eingesetzt werden, wenn die Halterung diese Translation ermöglicht, ohne dass die anderen Parameter dadurch verstellt werden. Eine Zeile der direkten Ansicht nimmt nun die doppelte Strecke einer Zeile der reflektierten Ansicht ein. Damit ist insgesamt das Verhältnis zwischen virtuellem und tatsächlichem Sichtfeld ausgeglichen. Ist dieser Abstand zusätzlich zu den oben erwähnten Bedingungen erreicht, so gilt das System als kalibriert.

#### 4.4.1 Extraktion der relevanten Bereiche

Das aufgenommene Bild enthält nicht nur die direkte Ansicht und die Spiegelung der Szene, sondern auch Teile des Spiegeläußeren und Bereiche, welche nicht von dem Fisheye-Objektiv belichtet wurden. Um nun die relevanten Bereiche aus dem Bild zu extrahieren, wird der Mittelpunkt des Abbildungskreises der Fisheye-Abbildung und der Durchmesser des äußeren Rings, also der Spiegelöffnung, benötigt.

Dieser äußere Durchmesser des Systems wird per Hand ermittelt. Hierfür wird eine Aufnahme des Kalibrierungsmusters benötigt. In diese werden ein Punkt und zwei konzentrische Kreise eingezeichnet. Diese gilt es dann per Hand so auszurichten, dass der Punkt dem Mittelpunkt des Kalibrierungsmusters, der kleinere der konzentrischen Kreise also dem inneren Durchmesser  $d_i$  und der größere dem äußeren Durchmesser  $d_o$  entspricht (vgl. Abbildung 4.7). Da dieser Vorgang nur einmalig für jede Konfiguration des Systems durchgeführt werden muss und zusätzlich sehr kritisch ist, wurde davon abgesehen ihn zu automatisieren<sup>5</sup>.

Mit diesen Informationen kann die *Bounding Box* des relevanten Bereichs ermittelt werden. Dieser quadratische Ausschnitt der Rohaufnahme enthält alle benötigten Bildinformationen und bildet die Basis aller weiteren Berechnungen.

---

<sup>5</sup>Durch Mängel in der Halterung war es leider allzu häufig notwendig das System neu zu kalibrieren.

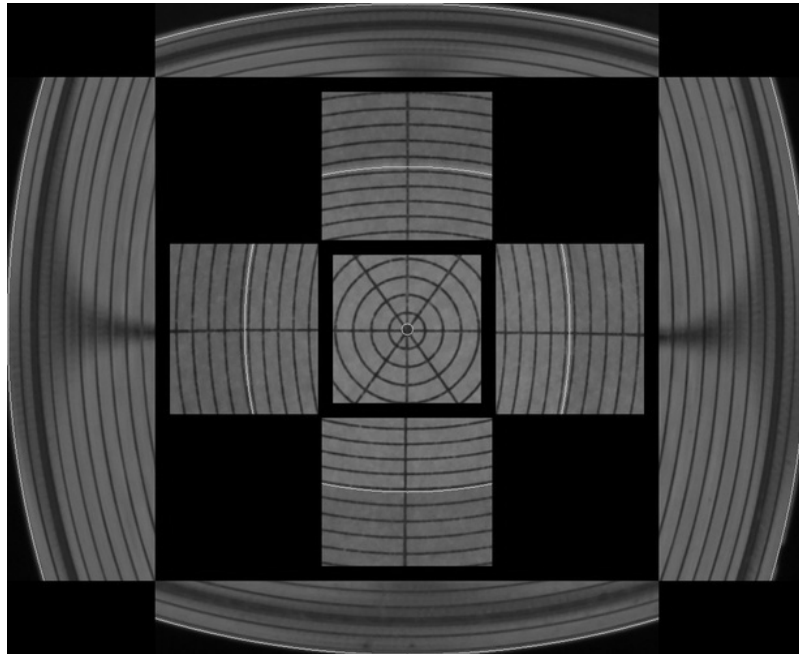


Bild 4.7: Visualisierung des Hilfsprogramms zur Bestimmung der relevanten Bereiche

#### 4.4.2 Fisheye-Entzerrung

In diesem Abschnitt wird ein Verfahren zur Entzerrung der Fisheye-Abbildungsfunktion vorgestellt. Das in diesem System zum Einsatz kommende Sigma Fisheye-Objektiv weist, wie die meisten modernen Fisheye-Objektive, eine *Equisolid-Angle*-Abbildungsfunktion auf (vgl. Abschnitt 2.1.6). Um die Verzerrung rückgängig zu machen, muss diese Abbildungsfunktion umgekehrt werden:

$$r = 2 \cdot f \cdot \sin\left(\frac{\omega}{2}\right) \quad (4.3)$$

In einem ersten Schritt wird die Bildgröße nach der Entzerrung berechnet und ein Bild mit diesen Dimensionen erstellt, im zweiten Schritt wird dieses Bild durchlaufen und für jeden „entzerrten“ Pixel der entsprechende verzerrte Punkt im Eingangsbild gesucht, mit Hilfe eines Interpolationsverfahrens dessen Wert ermittelt und an die aktuelle Position geschrieben.

Für die Berechnung der Bildgröße nach der Entzerrung wird die nach  $\omega$  aufgelöste *Equisolid-Angle*-Abbildungsfunktion benötigt:

$$\omega = 2 \arcsin \left( \frac{r}{2f} \right) \quad (4.4)$$

Die Brennweite  $f$  in Pixeln ergibt sich aus der Auflösung  $m \times n$  und den Dimensionen des Sensors  $w \times h$  in mm und der Brennweite  $f_{mm}$  in mm:

$$f = \frac{1}{2} \cdot f_{mm} \cdot \left( \frac{m}{w} + \frac{n}{h} \right) \quad (4.5)$$

Die halbe Breite des  $m \times n$  Eingangsbildes  $\frac{m}{2}$  entspricht dem Radius des maximalen Blickwinkels der Fisheye-Aufnahme in x-Richtung und analog entspricht die halbe Höhe des Eingangsbildes  $\frac{n}{2}$  dem Radius des maximalen Blickwinkels in y-Richtung. Werden  $r_x = \frac{m}{2}$  beziehungsweise  $r_y = \frac{n}{2}$  und die Brennweite  $f$  in die Formel 4.4 eingesetzt, so kann der maximale Blickwinkel der Fisheye-Aufnahme ermittelt werden:

$$\omega_{max_x} = 2 \arcsin \left( \frac{m}{4f} \right) \quad (4.6)$$

$$\omega_{max_y} = 2 \arcsin \left( \frac{n}{4f} \right) \quad (4.7)$$

Die Auflösung des entzerrten Bildes  $r \times s$  ergibt sich mit Hilfe der perspektivischen Abbildungsfunktion, da hierbei der Blickwinkel dem Abbildungswinkel  $\omega$  entspricht. Dieses Problem kann durch einfache Trigonometrie gelöst werden. Da die Brennweite  $f$  senkrecht auf der Bildebene steht, bildet sie die Ankathete des Winkels  $\omega_{max_x}$  und es lässt sich so der Radius  $r_p$  berechnen, an dem die perspektivische Abbildungsfunktion diesen Bildpunkt abbilden würde:

$$r := 2 \cdot r_{px} = 2 \cdot f \tan(\omega_{max_x}) \quad (4.8)$$

$$s := 2 \cdot r_{py} = 2 \cdot f \tan(\omega_{max_y}) \quad (4.9)$$

Im Folgenden wird davon ausgegangen, dass der gemeinsame Mittelpunkt der *perspektivischen* und der *Equisolid-Angle*-Abbildungsfunktion auch der Ursprung des entzerrten Bildes  $(k, k)$  darstellt. Dass dieser Mittelpunkt mit dem Brennpunkt der Aufnahme übereinstimmt, wird bei der Extraktion der relevanten Bildbereiche sichergestellt (vgl. Abschnitt 4.4.1).

Die Parameterdarstellung in Form von Radius und Winkel eines jeden Pixels des entzerrten Bildes muss berechnet werden. Der Winkel wird durch die Parametergleichung  $x = r \cos(\alpha)$  ermittelt:

$$\alpha = \arccos\left(\frac{x}{r_p}\right) \quad (4.10)$$

Der Radius  $r_p$  wird mit Hilfe des Satzes des Pythagoras berechnet:

$$r_p = \sqrt{x^2 + y^2} \quad (4.11)$$

Dieser wird benötigt, um den perspektivischen Abbildungswinkel  $\omega_p$  zu ermitteln:

$$\omega_p = \arctan\left(\frac{r_p}{f}\right) \quad (4.12)$$

Nun ist es möglich den Fisheye-Radius mit der *Equisolid-Angle*-Abbildungsfunktion zu berechnen:

$$r = 2 \cdot f \cdot \sin\left(\frac{\omega_p}{2}\right)$$

Mit Hilfe des zwischen perspektivischer und *Equisolid-Angle*-Abbildungsfunktion identischen Winkels  $\alpha$  und dem ermittelten Fisheye-Radius, kann die Position im verzerrten Eingangsbild, die der aktuellen Position im „entzerrten“ Bild entspricht, berechnet werden. Da es sich um ein Ergebnis im Subpixelbereich handelt, muss der Pixelwert der gefundenen Position noch mittels Interpolation (*Sampling*) ermittelt werden.



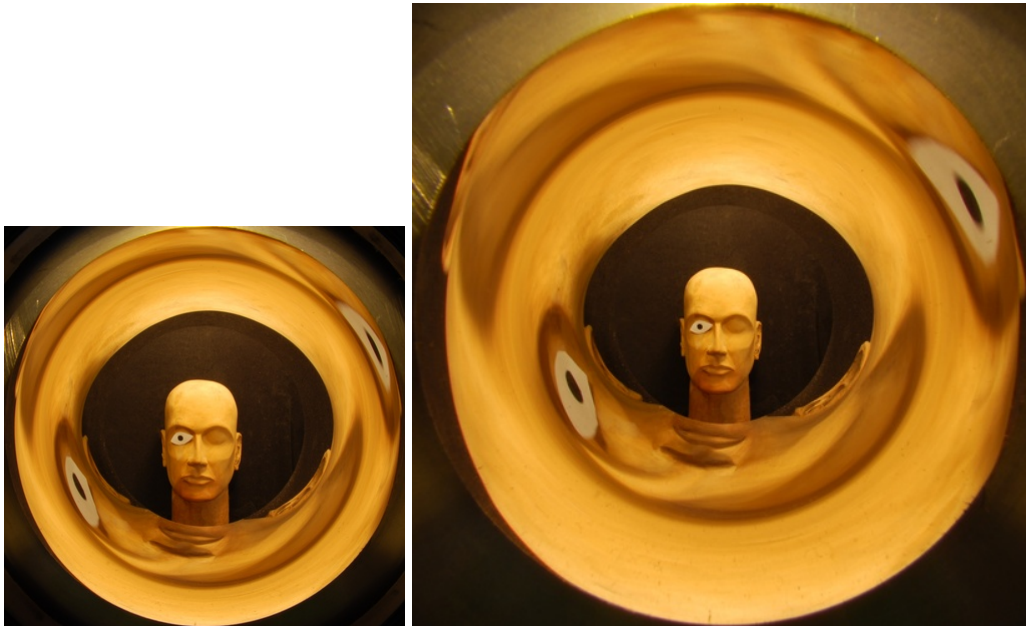


Bild 4.8: links: relevanter Bereich der Rohaufnahme, rechts: die *Equisolid-Angle*-Abbildungsfunktion wurde in eine perspektivische überführt (das Größenverhältnis wurde beibehalten).

Um das entzerrte Ergebnisbild zu erhalten, muss dieser Vorgang für jedes Pixel durchgeführt werden.

### 4.4.3 Aufteilen in verschiedene Ansichten

In diesem Abschnitt wird das Verfahren erläutert, mit dem das entzerrte Eingangsbild in drei Ansichten aufgeteilt wird (vgl. Abschnitt 3.5.1). Ähnlich dem Prozess der Entzerrung werden Bilder der angestrebten Auflösung durchlaufen und mit den entsprechenden Informationen aus dem Eingangsbild gefüllt. Es kann hierbei eine beliebige Auflösung der Ansichten gewählt werden, allerdings muss bedacht werden, dass sich die Informationsdichte durch eine Interpolation nicht erhöhen lässt. Umgekehrt gilt, dass die Auflösung auch nicht kleiner gewählt werden sollte als die tatsächliche Pixeldichte, da ansonsten Informationen verloren gehen.

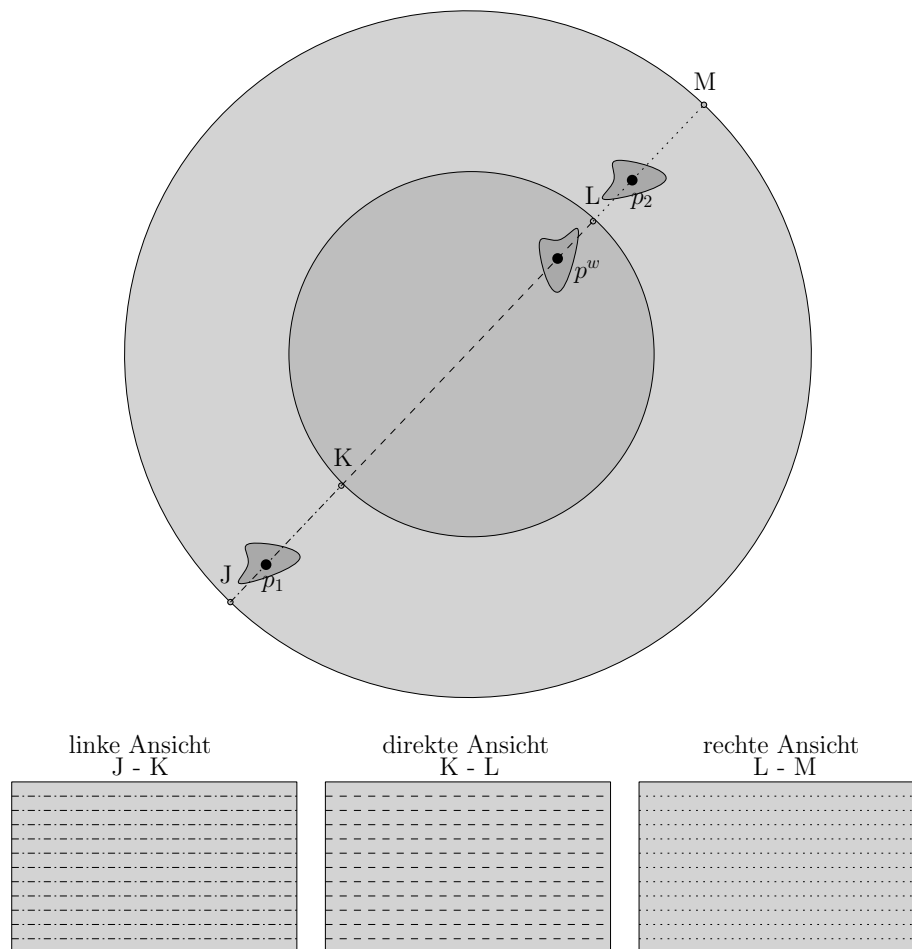


Bild 4.9: oben: Aufbau eines mit einem radialen System aufgenommenen Bildes, unten: Aufbau der extrahierten Ansichten

Wenn sich ein Weltpunkt  $p^w$  im trifokalen Bereich des Systems (vgl. Abschnitt 3.5.1) befindet, wird dieser dreimal auf dem Sensor abgebildet (siehe Abbildung 4.9). Einmal direkt ( $p^w$ ) und zweimal indirekt über den Spiegel ( $p_1$  und  $p_2$ ). Wichtig hierbei ist die Tatsache, dass sich die drei Punkte alle auf einer Linie befinden, die durch den Mittelpunkt ( $k, k$ ) des Gesamtsystems geht. Die Reflexion eines auf einer solchen Linie liegenden Weltpunktes befindet sich immer ebenfalls auf dieser Linie. Durch diese Eigenschaft werden Linien, die durch den Mittelpunkt verlaufen, zu den Epipolarlinien des Systems. Das heißt die Epipolarlinien eines solchen Systems verlaufen radial.

Eine solche Linie lässt sich in drei Abschnitte unterteilen (vgl. Abbildung 4.9): das äußere Liniensegment links des Mittelpunkts  $\overline{JK}$  (dessen  $x$ -Werte stets kleiner gleich dem Mittelpunkt  $k$  sind,  $x \leq k$ ), das mittlere Liniensegment  $\overline{KL}$  und das rechte Liniensegment  $\overline{LM}$  (mit  $x > k$ ). Um das Eingangsbild in drei Ansichten zu zerteilen, werden Linien mit dem Winkel  $0^\circ \leq \alpha < 180^\circ$  extrahiert und in die Abschnitte aufgeteilt. Diese Abschnitte werden in vorbereiteten Bildern für die drei Ansichten gestapelt. So bildet die erste extrahierte Linie mit dem Segment  $\overline{JK}$  die erste Zeile der linken Ansicht, mit Segment  $\overline{KL}$  die erste Zeile der direkten Ansicht und mit Segment  $\overline{LM}$  die erste Zeile der rechten Ansicht. Dieser Vorgang wird entsprechend der gewünschten  $y$ -Auflösung wiederholt.

Um die angestrebte Auflösung von  $m \times n$  zu erreichen, wird der zu durchlaufende Halbkreis für die  $y$ -Auflösung in entsprechend viele Abschnitte unterteilt. Die Schrittweite  $\text{step}_\alpha$  ergibt sich aus:

$$\text{step}_\alpha = \frac{\pi}{n}$$

Für die Auflösung in  $x$ -Richtung müssen die Segmente entsprechend unterteilt werden. Hierfür werden die jeweiligen Durchmesser des Gesamtkreises  $d_o$  und des inneren Kreises  $d_i$  benötigt. Die Schrittweiten für die direkte Ansicht  $\text{step}_i$  sowie die beiden äußeren Ansichten  $\text{step}_o$  unterscheiden sich und werden wie folgt berechnet:

$$\text{step}_o = \frac{\frac{d_o}{2} - \frac{d_i}{2}}{m}$$

$$\text{step}_i = \frac{d_i}{m}$$

Mit Hilfe dieser Winkel- und Raduisschrittweite können die gesuchten Positionen basierend auf der Parameterdarstellung des Kreises berechnet und der entsprechende Pixelwert durch Interpolation ermittelt werden. Sind die Ansichten so erstellt, muss die aus einer Reflexion entstandene linke und rechte Ansicht noch horizontal gespiegelt werden, um der mittleren Ansicht zu entsprechen.

Nach diesem Prozess sind die Epipolarlinien horizontal ausgerichtet und die Ansichten somit in  $y$ -Richtung rektifiziert (vgl. Abbildung 4.10). Da die linke und rechte Ansicht aus

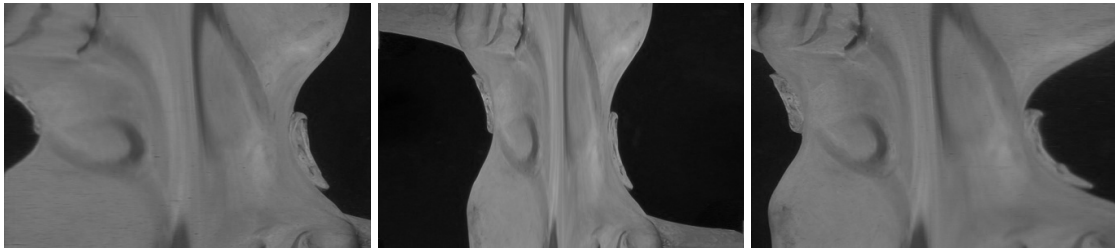


Bild 4.10: Die extrahierte linke, direkte und rechte Ansicht.

einer Reflexion an einem nicht orthogonal zur Bildebene stehenden Spiegel entstanden sind, müssen sie noch in x-Richtung rektifiziert werden. Die Rektifizierung dieser Ansichten wird im folgenden Abschnitt 4.4.4 erläutert.

#### 4.4.4 Rektifizierung des nicht-planaren Spiegels

Nur orthogonal zur Bildebene stehende Spiegel bilden die Szene ohne Verzerrungen ab. Wie aus dem Querschnitt des Systems, beziehungsweise aus der Tatsache, dass es sich um ein konisches Rohr handelt, deutlich ersichtlich ist, stehen die Spiegel in diesem System nicht orthogonal zur Bildebene und müssen deshalb rektifiziert werden. Wie stark die Verzerrungen ausfallen, hängt von der Entfernung zwischen Kamera und Spiegel ab und muss deshalb nur neu kalibriert werden, wenn sich dieser Abstand ändert. Aus diesem Grund ist es ausreichend, das für das Ausrichten des Gesamtsystems zum Einsatz kommende Kalibrierungsmuster zu analysieren.

Um dies zu verifizieren, wird das System mit Hilfe zweier Muster (siehe Bild 4.11) kalibriert, die vor dem Gesamtsystem angebracht werden. Die beiden Muster werden hierfür bündig übereinander gelegt und aneinander befestigt. Anschließend werden sie parallel zur Bildebene vor dem Gesamtsystem befestigt.

Die Linien im Ausrichtungsmuster helfen es so auszurichten, dass der Mittelpunkt auf der optischen Achse des Gesamtsystems liegt. Dies ist, wie schon bei der Ausrichtung der optischen Achsen der Kamera und des Spiegels (vgl. Abschnitt 4.4), dann der Fall, wenn sie auch in der Reflexion eine Linie bilden. Dies gilt insbesondere für den Bereich außerhalb der etwas dickeren Kreislinie, die durch die Spiegelung des Mittelpunkts entsteht

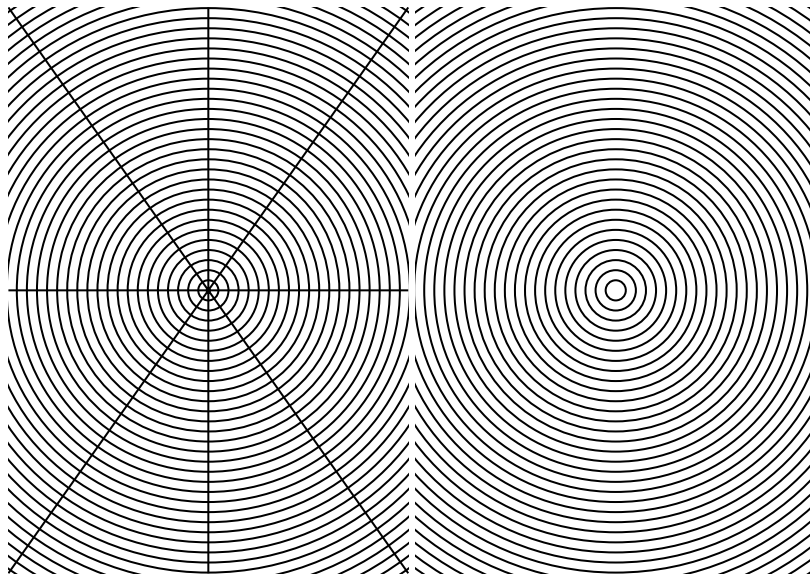


Bild 4.11: links: Ausrichtungsmuster, rechts: eigentliches Kalibrierungsmuster

(vgl. Abbildung 4.12). Dass sie sich zum Rand hin aufzulösen scheinen, ist auf den *Field Curvature* genannten Effekt zurückzuführen (siehe Abschnitt 4.4.5). Ist das Muster ausgerichtet, so befindet sich dessen Mittelpunkt auf der optischen Achse des Gesamtsystems und es kann durch das eigentliche Kalibrierungsmuster ersetzt werden.

In Tests hat sich gezeigt, dass das Ausrichtungsmuster unter Umständen auch zur Kalibrierung genutzt werden kann. Da im Rektifizierungsverfahren mit einem Sobelfilter nach vertikalen Kanten gesucht wird, stellen die nach der Aufteilung in die verschiedenen Ansichten (vgl. Abschnitt 4.4.3) horizontal verlaufenden Ausrichtungslinien kein Problem dar. Hierfür sollte jedoch nicht genau in der Bildmitte gesucht werden, da dort eine der Ausrichtungslinien liegt und somit das Auffinden der vertikalen Kanten an dieser Stelle erschwert wird.

Um die zur Berechnung nötigen Punkte im Kalibrierungsmuster finden zu können, wird noch eine weitere Markierung benötigt. Es muss der äußerste Punkt der direkten Ansicht gefunden werden, der noch von der Spiegelung erfasst wird. Um diese Stelle zu finden, werden rechts und links des Musters farbige Stecknadeln so platziert, dass sie nur halb von der Spiegelung erfasst werden. Die farbigen (in diesem Fall roten) Stecknadeln kön-

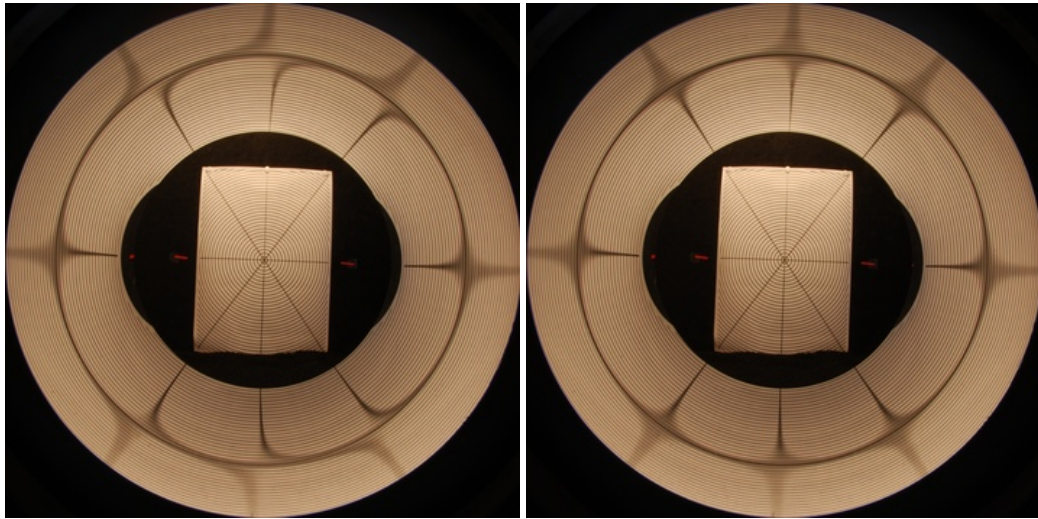


Bild 4.12: links: unausgerichtetes Muster, rechts: ausgerichtetes Muster

nen leicht durch Farbsegmentierung gefunden werden. Dazu wird das Bild in den HSV-Farbraum überführt und die Farbe der Markierung mit einem entsprechenden Farbintervall herausgefiltert.

Nach dem Trennen der Ansichten müssen Punktkorrespondenzen zwischen linker und direkter Ansicht und zwischen rechter und direkter Ansicht gefunden werden. Die Rektifizierungsfunktion lässt sich aus diesen Punkten berechnen, allerdings wird dafür der gemeinsame Ursprung benötigt. Zu diesem Zweck wurden die Markierungen angebracht, stellen sie doch genau diesen gemeinsamen Nullpunkt der Rektifizierungsfunktion dar (vgl. Abbildung 4.13). Die linke Markierung stellt hierbei den Nullpunkt der linken Ansicht und die rechte analog für die rechte Ansicht dar. Um die Analyse dieser beiden Fälle zu vereinheitlichen wurde die linke Ansicht vertikal gespiegelt und die Reihenfolge der Korrespondenzen in der direkten und als Referenz dienenden Ansicht für diesen Fall umgekehrt (vgl. Abbildung 4.14).

Die Rektifizierungsfunktion des nicht-orthogonal stehenden Spiegels hat die Form:

$$ax_{referenz}^2 + x_{undistorted} = x_{distorted} \quad (4.13)$$

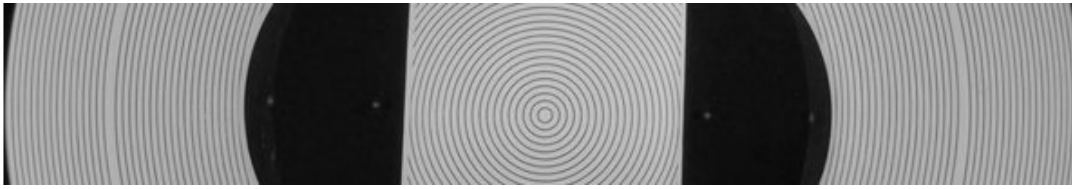


Bild 4.13: Ausschnitt des Kalibrierungsmusters mit roten Spiegelungsmarkierungen

Hierbei handelt es sich um die x-Werte der gefundenen Punktkorrespondenzen von direkter Ansicht  $x_{undistorted}$  und den verzerrten Spiegelansichten  $x_{distorted}$ . Der Rektifizierungsparameter  $a$  beschreibt den Grad der Verzerrung. Bei  $a = 0$  würde dies bedeuten, dass es keine Verzerrung zwischen direkter und Spiegelansicht gibt. Der Rektifizierungsparameter in dieser Konfiguration ist beispielsweise  $\approx 0.0028$ .

Berechnet wird dieser Parameter indem für jede Punktkorrespondenz die Gleichung analytisch gelöst wird, und dann alle anderen Korrespondenzen anhand eines Schwellwertes in zwei Gruppen unterteilt werden. Punkte die innerhalb dieses Schwellwertes liegen kommen in die Gruppe der *Inlier* und der Rest in die Gruppe der *Outlier*. Diejenige Korrespondenz, die die größte Anzahl von Inliern hatte wird als Parameter der jeweiligen Punktkorrespondenzen angenommen. Da der Anfangspunkt oft nicht ganz korrekt ermittelt wird und deshalb eine Verschiebung der Kurve zu unterschiedlich guten Ergebnissen führt, wird diejenige Lösung gewählt, die die meisten Inlier aufweisen kann, oder falls beide die gleiche Anzahl Inlier haben, werden die Ergebnisse der linken und rechten Punktkorrespondenzen anschließend gemittelt, um das finale Ergebnis zu erhalten. Somit kann das Ergebnis zwar nur unwesentlich besser als die beste Korrespondenz werden, dieses an RANSAC angelehnte Verfahren hat sich in der Praxis aber als völlig ausreichend erwiesen (vgl. Abbildung 4.15).

Für die Rektifizierung wird nun zunächst die Breite des rektifizierten Bildes ermittelt, um anschließend mit Hilfe der Formel 4.13 für jeden Punkt des rektifizierten Bildes den entsprechenden Punkt im nicht rektifizierten Bild zu approximieren. Hierbei sind  $a$  sowie  $x_{undistorted}$  bekannt und  $x_{distorted}$  muss berechnet werden. Es wird also ein Bild der Größe des rektifizierten Bildes Pixel für Pixel durchlaufen und mit Hilfe der Formel ermittelt, welcher Punkt im nicht-rektifizierten Bild diesen Punkt beeinflusst. Werden die beiden

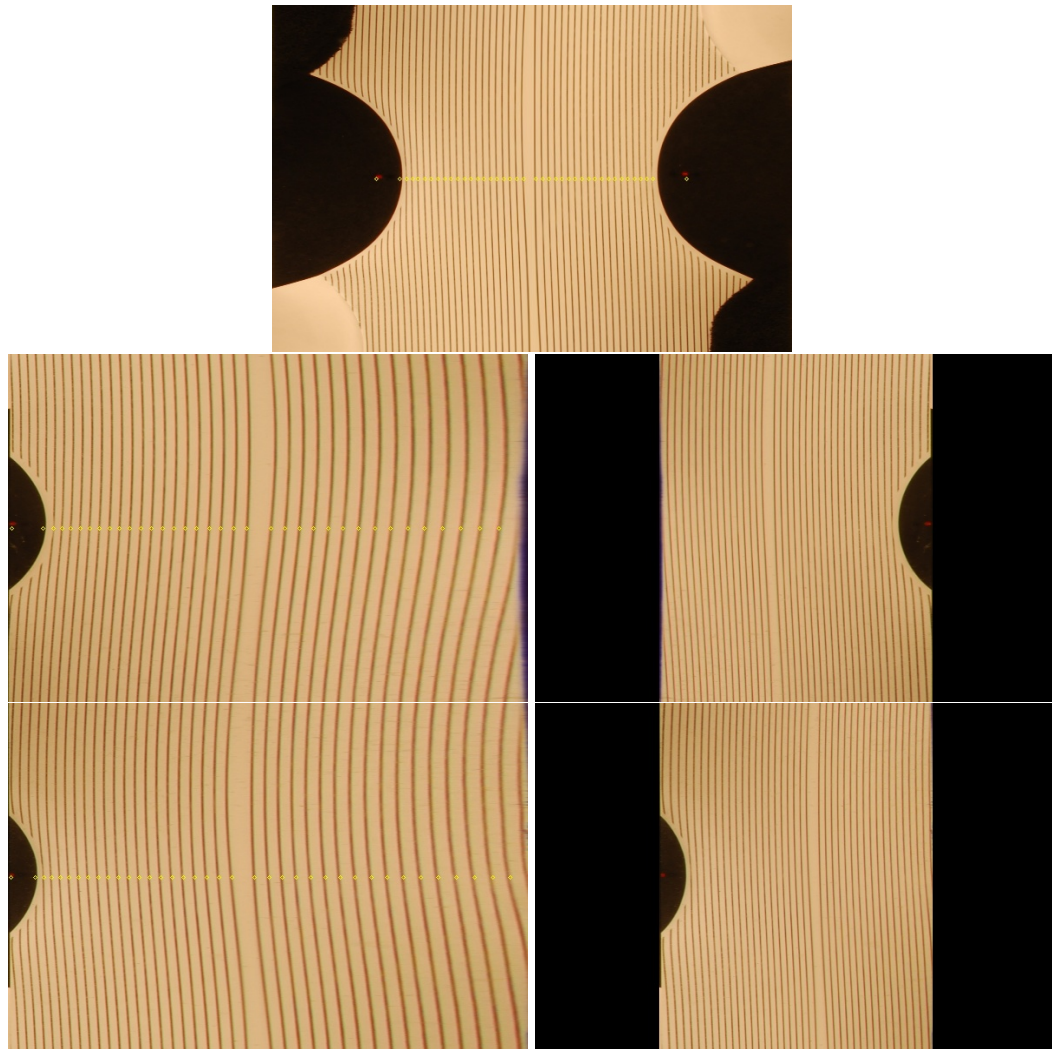


Bild 4.14: das getrennte und analysierte Kalibrierungsmuster

Spiegelansichten nun mit diesem Verfahren rektifiziert, so sind die Abstände zwischen den Linien identisch mit denen der direkten Ansicht.

Die beiden reflektierten Ansichten sind nun horizontal wie vertikal rektifiziert und können somit mit traditionellen Stereo Matching Verfahren in eine dichte Tiefenkarte überführt werden (vgl. Abbildung 4.16).



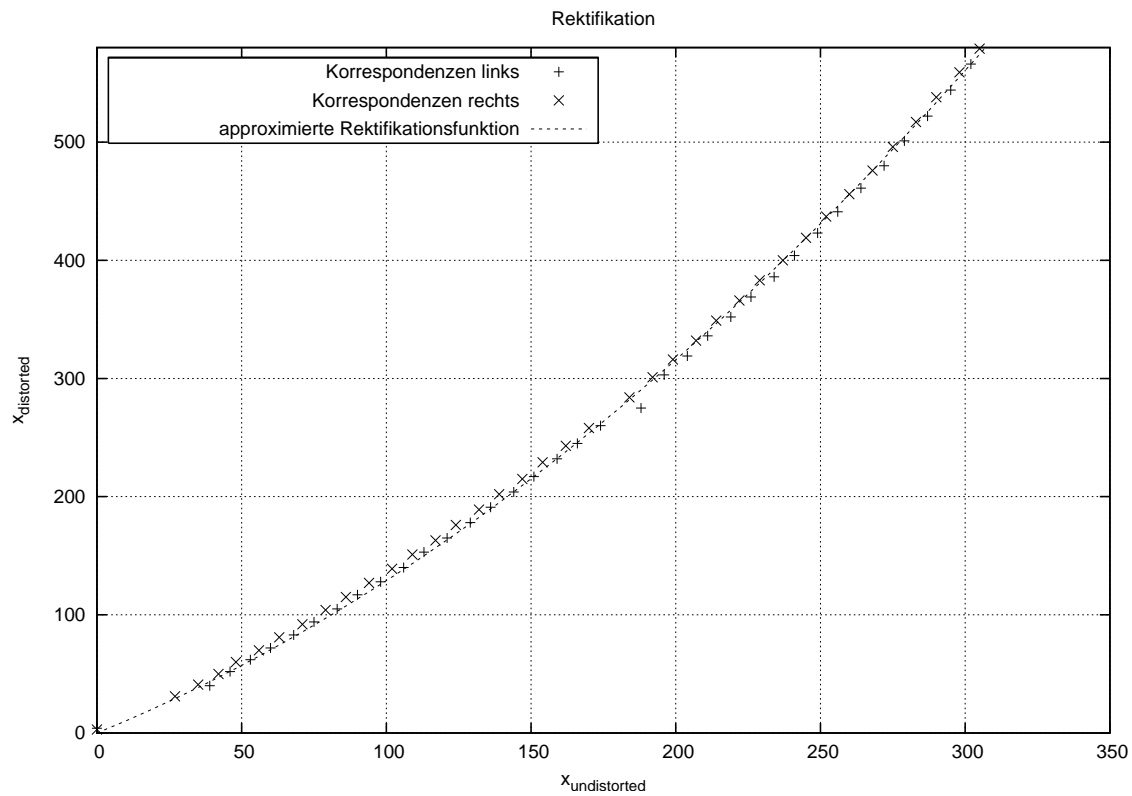


Bild 4.15: Rektifizierungsgenauigkeit

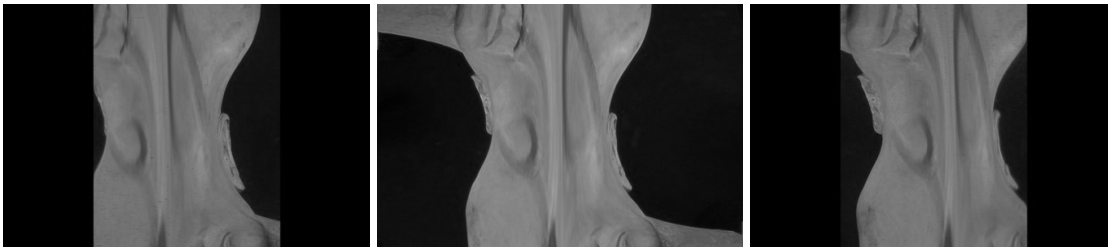


Bild 4.16: Die rektifizierten Ansichten.

#### 4.4.5 Dense Stereo Matching

Die nun rektifizierten Ansichten werden mit der mittleren Ansicht als Referenz *gematcht*, das heißt die linke Ansicht bildet mit der mittleren und die mittlere mit der rechten Ansicht

jeweils ein Stereopaar. Die beiden reflektierten Ansichten weisen zum linken respektive rechten Rand hin eine zunehmend stärker werdende Unschärfe auf (vgl. Abschnitt 3.5.1), die sich negativ auf die Ähnlichkeit auswirkt und somit das Stereo Matching weniger robust werden lässt. Um die Ähnlichkeit der zu matchenden Ansichten zu erhöhen, kann die mittlere Ansicht, die als Referenz dient, an die Unschärfe der mit ihr zu matchenden Ansicht angepasst werden (vgl. Abschnitt 4.4.5).

### Unschärfeausgleich



Bild 4.17: Die rechte Ansicht eines Kopfmodells, zum rechten Rand hin zeigt sich die zunehmend stärker werdende Unschärfe.

In den äußeren Regionen wird die in Abschnitt 3.5.1 erläuterte Unschärfe so stark, dass das Stereo Matching negativ beeinflusst wird. Konkret bedeutet dies, dass die Unschärfe in der linken Ansicht zum linken Rand und in der rechten Ansicht zum rechten Rand hin stark ansteigt, wie im Bild der rechten Ansicht 4.17 deutlich zu erkennen ist.

In Abschnitt 3.5.1 wurde beschrieben, dass eine kleine Blendenöffnung dieses Problem reduziert, da so eine große Schärfentiefe erzielt werden kann. Abbildung 4.17 ist bei maximaler Blendenzahl ( $f/22$ ) entstanden, der kleinstmöglichen Blende der verwendeten Kamera. Durch die kleine Blendenöffnung fällt allerdings auch weniger Licht auf den Sensor, was im Umkehrschluss bedeutet, dass die Szene entweder sehr hell sein muss, oder eine sehr lange Belichtungszeit benötigt wird, um das Motiv in ausreichender Helligkeit abzubilden. Bei Gesichtern ist beides nur bedingt möglich, da diese sich zum einen bewegen (sei es nur ein Blinzeln) und zum anderen die Beleuchtung nicht zu grell sein darf, um Augenschäden zu vermeiden.

Um dem Effekt auf anderer Art entgegenzuwirken, können die zusammengehörenden Stereopaare, jeweils linke und rechte mit direkter Ansicht, aneinander angeglichen werden. Da ein nachträgliches Schärfen dieser Bereiche nur schwer möglich ist, wird die umgekehrte Herangehensweise gewählt: um die unscharfen Bereiche der Spiegelansichten der direkten Ansicht ähnlicher zu machen, wird die von der Unschärfe nicht beeinflusste direkte Ansicht mit Unschärfe versehen. Dies geschieht mit einer zum Rand hin zunehmend stärker werdenden gaußschen Unschärfe. Für das Matching der linken mit der mittleren Ansicht wird die mittlere Ansicht zur linken Seite verstärkt mit Unschärfe versehen. Für das Matching der rechten mit der mittleren Ansicht, wird analog die mittlere Ansicht zum rechten Rand mit Unschärfe versehen.

#### 4.4.6 Aufnahme

Bei der Aufnahme gilt es zu beachten, dass das Objekt möglichst schattenarm ausgeleuchtet ist. Dies ist besonders wichtig, wenn die Aufnahme auch als Textur für die Darstellung fungieren soll. Ein sich auf der Textur befindender Schatten wäre nicht nur unschön, sondern würde auch den Eindruck von Dreidimensionalität verschlechtern, da der Schatten sich natürlich nicht entsprechend der Lichtquellen verändern würde. Um eine schattenarme Darstellung zu gewährleisten, sollte das Objekt bei Studiobedingungen aufgenommen werden. Die hier gezeigten Aufnahmen sind in einem improvisierten Studio entstanden. Es kamen zwei Baustrahler mit jeweils 500 Watt Leistung zum Einsatz, die ein etwas zu hartes Licht produzierten und deshalb mit einem Diffusor ausgestattet wurden. Hierzu

wurden weiße Regenschirme vor den Leuchten montiert, sodass eine möglichst schattenarme Ausleuchtung erreicht wurde.

Die Kamera soll das Objekt nicht nur direkt, sondern auch im deutlich näher an der Kamera gelegenen Spiegel scharf abbilden, dafür muss eine kleine Blende verwendet werden, um eine möglichst große Schärfentiefe zu erzielen. Daraus ergibt sich der negative Nebeneffekt, dass sich die Belichtungszeit dramatisch erhöht, da weniger Licht durch die kleine Blendenöffnung auf den Sensor fällt.

Bei nicht statischen Objekten, wie beispielsweise Gesichtern, ist eine starke Beleuchtung ebenfalls von besonderer Bedeutung, da dadurch die Belichtungszeit und damit auch die Bewegungsunschärfe reduziert werden kann.

### **Reinigung des Spiegels**

Da das Bauteil nicht zu allen Zeiten abgedeckt gehalten werden konnte, haben sich Staub und andere Schmutzpartikel an der Innenseite festgesetzt. Die Reinigung der sehr empfindlichen Oberfläche ist nur indirekt möglich. Dabei wird die Reinigungsflüssigkeit (Penthan oder Chloroform) lediglich auf der Fläche ablaufen gelassen, ohne mechanische Einwirkung, wie etwa Abwischen oder Polieren. Auch der Gebrauch von spezieller Pressluft ist möglich, die jedoch keinesfalls Öl- oder Schmiermittel beinhalten darf. Druckbehälter mit hochreiner Druckluft sind im Fachhandel für den Labor- und Messbereich erhältlich.

# Kapitel 5

## Experimente und Ergebnisse

### 5.1 Erstellung synthetischer Testbilder mit POV-Ray

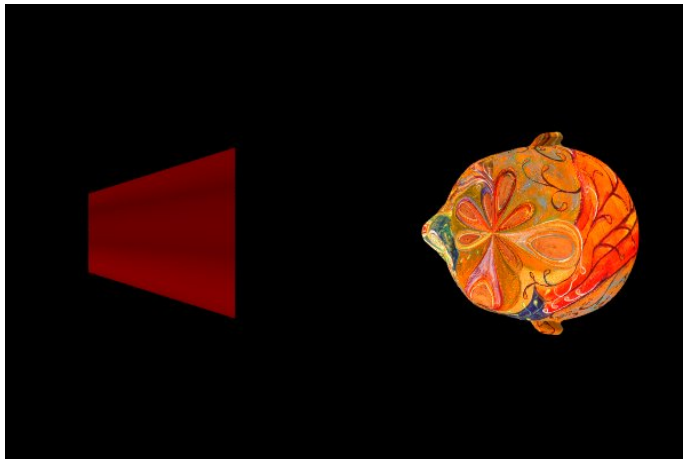


Bild 5.1: Draufsicht auf die synthetische Szene

In der Anfangsphase dieser Diplomarbeit wurden die Algorithmen aufgrund der fehlenden Ausrüstung anhand von synthetischen Bildern entwickelt. Diese wurden im Raytracing-Verfahren erstellt, da es sich besonders für die Simulation von Spiegeln eignet. Für diese

Aufgabe wurde der Raytracer POV-Ray<sup>1</sup> benutzt, da dieser plattformunabhängig und im Quellcode verfügbar ist. Der Spiegel wurde hierfür exakt in POV-Ray nachgebildet sowie das Modell eines Kopfes und die Kamera entsprechend platziert. Bei der in POV-Ray eingebauten Fisheye-Abbildungsfunktion handelt es sich allerdings nicht um die für eine Simulation des in dieser Diplomarbeit verwendeten Sigma 8mm Fisheye-Objektivs benötigte *Equisolid-Angle-Abbildungsfunktion* (siehe Abschnitt 2.1.6). Deshalb wurde POV-Ray um eine Kamera mit dieser Abbildungsfunktion erweitert (vgl. B.2). Auf diesem Weg ist Bild 5.2 (links) entstanden.

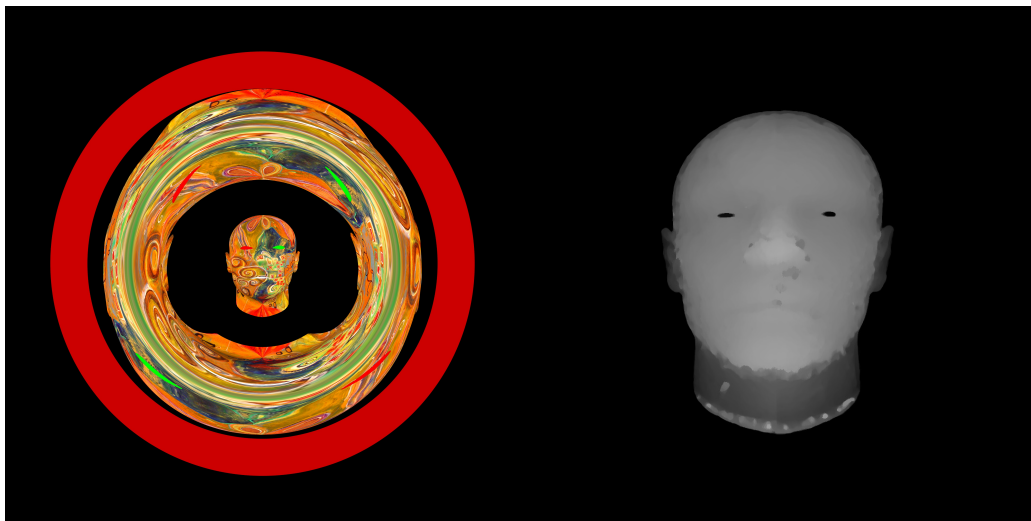


Bild 5.2: links: Mit POV-Ray entstandenes synthetische Rohaufnahme eines radialen kadioptrischen Stereosystems. rechts: Daraus erstellte Tiefenkarte des Kopfes (Median gefiltert)

Tests mit diesen Aufnahmen und denen, die mit dem realen Spiegel entstanden sind haben gezeigt, dass sich die Rektifizierungsfunktion in den beiden Fällen unterscheidet und ein Vergleich zwischen synthetischen und realen Aufnahmen somit schwierig wird.

Die Rektifizierungsfunktion der synthetischen Bilder weist im Vergleich zu der Rektifizierungsfunktion des realen Systems 4.13 einen weiteren Parameter  $b$  auf:

---

<sup>1</sup><http://www.povray.org>

$$ax_{referenz}^2 + bx_{undistorted} = x_{distorted} \quad (5.1)$$

Der Grund hierfür liegt wahrscheinlich in der simulierten Kamera. Mögliche Ursachen könnten sein, dass der Strahlengang in der realen Kamera nicht dem der simulierten Kamera entspricht oder dass die reale *Equisolid-Angle*-Abbildungsfunktion nicht mit der idealen Abbildungsfunktion identisch ist.

Vieles spricht für erstere Möglichkeit, da auch die Entfernung der simulierten Kamera vom Spiegel bei gleichem horizontalem Bildwinkel, wie er in den realen Aufnahmen auftritt, in keiner Weise der Entfernung der realen Kamera entspricht. Es wäre auch möglich, dass die Rektifizierungsfunktionen sich gerade deshalb unterscheiden, weil die Entfernung der Kamera zum Spiegel in beiden Systemen unterschiedlich ist.

Es können trotzdem wichtige Erkenntnisse aus diesen Bildern gewonnen werden. Die Ergebnisse des *Stereo Matchings* auf Basis synthetischer Bilder sind denen realer Aufnahmen überlegen (vgl. rechte Abbildung 5.2). Mehrere Ursachen kommen hierfür in Frage. Die Fisheye-Entzerrung ist im synthetischen Fall die exakte Umkehrfunktion der Fisheye-Abbildungsfunktion, was im realen Fall durch Ungenauigkeiten im Fisheye-Objektiv nicht der Fall ist. Der wahrscheinlich ausschlaggebende Punkt ist, dass das reale System nicht so präzise ausgerichtet werden kann wie das synthetische. Die Halterung hätte hierfür präzisere Einstellungsmöglichkeiten und eine bessere Fixierung der Kamera bieten müssen. Vor allem durch die Befestigung am Stativgewinde war eine ausreichende Fixierung der Rotation um die y-Achse nicht möglich, sodass schon so geringfügige Einwirkungen wie das An- und Ausschalten sowie das Wechseln des Akkus oftmals eine Neukalibrierung des Systems notwendig machten. Mit einer verbesserten Halterung könnten die Auswirkungen der Ausrichtung des Gesamtsystems auf die Ergebnisse weiter erforscht werden.

## 5.2 Planspiegel-Stereosystem

### 5.2.1 Unschärfeausgleich

In diesem Abschnitt sollen die Auswirkungen des Unschärfeausgleichs auf das *Stereo Matching* untersucht werden. Hierfür wurde eine parallel zur Bildebene stehende ebene Fläche mit dem in Abschnitt 4.3 beschriebenen Planspiegel-Stereosystem aufgenommen. Um die Auswirkung des Unschärfeausgleichs zu testen wurde die reale Ansicht mit zunehmend stärker werdender Gauß-Unschärfe versehen und jeweils mit der virtuellen Ansicht *gematcht*. Bei diesem Test kam das normalisierte Kreuzkorrelationsverfahren (siehe Abschnitt 2.2.3) mit einem Schwellwert von 0.5 zum Einsatz. Es wurde zum einen der prozentuale Anteil nicht zuordenbarer Pixel im Verhältnis zur Gesamtpixelzahl gemessen und zum anderen der optische Eindruck beurteilt.

Tabelle 5.1: Fehler bei unterschiedlichem Unschärfeausgleich

Unschärferadius	prozentualer Fehler
-	21,21 %
0,2	21,21 %
0,4	18,99 %
0,6	13,01 %
0,8	9,95 %
1,0	8,11 %
1,2	6,89 %
1,4	6,19 %
1,6	5,70 %
1,8	5,56 %
2,0	5,64 %
2,2	5,77 %
2,4	6,16 %

Wie in Tabelle 5.1 zu sehen ist, sinkt der prozentuale Fehler bis zu einem Unschärferadius von 1,8 und steigt bei größerem Unschärferadius wieder an. Der optische Eindruck



verbessert sich jedoch nur bis zu einem Unschärferadius von 1,0. Ein stärkerer Unschärfeausgleich bewirkt in diesem System keine weitere Verbesserung des optischen Eindrucks (vgl. Abbildung 5.3).

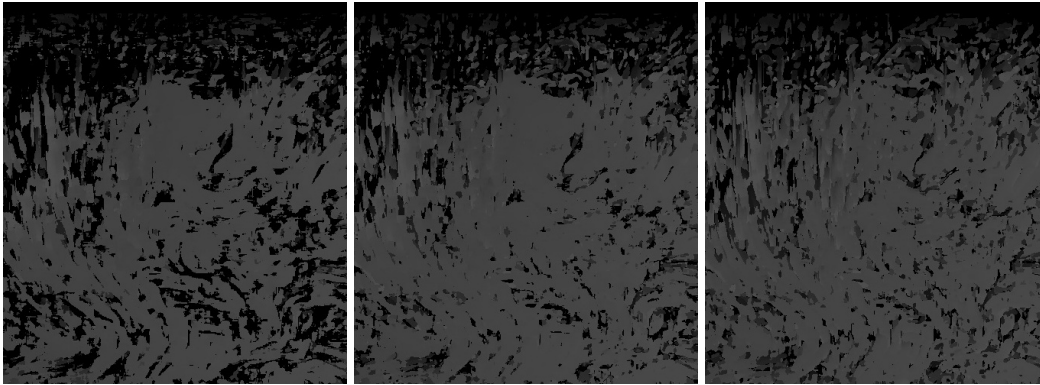


Bild 5.3: links: kein Unschärfeausgleich, mitte: optisch bester Unschärfeausgleich rechts: Unschärfeausgleich mit dem geringsten Fehler

Interessant wäre es die Unschärfe in der virtuellen Ansicht zu messen und den Unschärfeausgleich entsprechend anzupassen. Die im Haushaltsspiegel entstandene Unschärfe könnte sich hierbei jedoch als problematisch herausstellen, da sie eine ungewöhnliche Charakteristik aufweist. Es scheint sich hierbei um eine Art Dopplungseffekt zu handeln, der durch die Glasplatte vor der eigentlichen Spiegelfläche hervorgerufen wird (vgl. Abbildung 5.4). Der auf diese Weise ermittelte optimale Unschärfeausgleich hat unter anderen Bedingungen und unterschiedlichen Objektentfernungen gute Ergebnisse geliefert, was die Vermutung nahe legt, dass er lediglich systemabhängig ist und somit nur einmalig bestimmt werden muss.



Bild 5.4: Der durch die Glasplatte vor der Spiegelfläche hervorgerufene Dopplungseffekt.

### 5.2.2 Genauigkeit

In diesem Abschnitt soll die Messgenauigkeit des Planspiegel-Systems getestet werden. Hierfür wurde eine ebene Fläche parallel zur Bildebene in einem abgemessenen Abstand vor der Kamera platziert. Zwei Probleme traten hierbei auf: Zum einen ist die Messgenauigkeit mit einem Zollstock relativ gering und zum anderen ist von außen nicht zu erkennen, wo sich die Bildebene der Kamera genau befindet. Für ersteres Problem würde eine Präzisionsvorrichtung benötigt, bei der es möglich ist die Ebene exakt parallel zur Bildebene auszurichten und ihren Abstand zur Kamera genau zu messen. Um die genaue Lage der Bildebene, also des Sensors der Kamera, zu ermitteln, müssten die Baupläne der Kamera ausgewertet oder das Gehäuse geöffnet werden. Beides war leider nicht möglich.

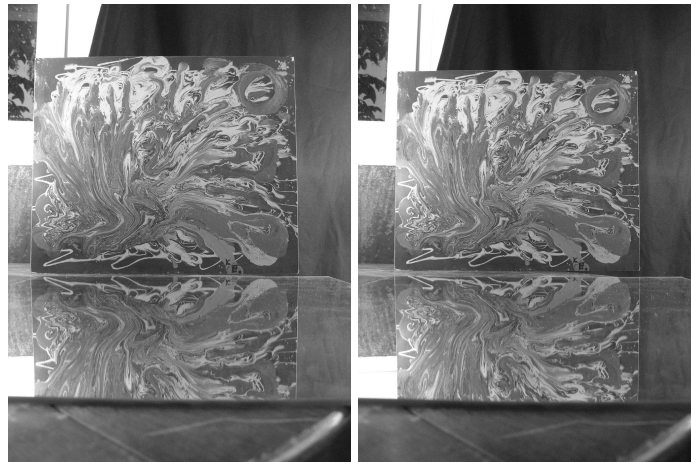


Bild 5.5: Ebenen mit unterschiedlichen Abständen. links: von zirka 75 cm, rechts: zirka 82 cm

Die Tests dienen aber auch der Überprüfung der in Abschnitt 4.3.1 vorgestellten Disparitätsformel 4.2 dieses Systems:

$$d' = d + (v_x - h)$$

Mit Hilfe der technischen Daten der Kamera (vgl. Abschnitt 4.2.1) können Millimeter in Pixel umgerechnet und so die für die Triangulierung wichtigen Größen ermittelt werden.

Der Umrechnungsfaktor  $u$  kann mit der leicht abgewandelten Formel 4.5 berechnet werden, hierbei gilt  $m = 1984$ ,  $n = 1488$ ,  $w = 5,75mm$  und  $h = 4,31mm$ :

$$u := \frac{1}{2} \cdot \left( \frac{m}{w} + \frac{n}{h} \right) = 345,14 \quad (5.2)$$

Die Brennweite der Kamera bei maximalem Bildwinkel entspricht  $f_{mm} = 5,8mm$  und die Stereobasis  $b_{mm}$  kann mit Formel 4.1 aus der Höhe der zur Ausrichtung der optischen Achse verwendeten Leiste  $l_h = 60mm$  errechnet werden:

$$\begin{aligned} f &= f_{mm} \cdot u = 2001,83 \\ b &= b_{mm} \cdot u = 2 \cdot l_h \cdot u = 41417,23 \end{aligned}$$

Mit Formel 2.2 kann nun die Tiefe trianguliert werden. Für den durchschnittlichen Disparitätswert wurde der Median der gefundenen Werte gewählt. Nicht zugeordnete Pixel wurden hierbei nicht berücksichtigt.

Bei der zirka 75 cm von der Kamera entfernten Ebene ergibt sich so ein durchschnittlicher Disparitätswert  $d$  von 43. Die tatsächliche Disparität  $d'$  ergibt sich aus Formel 4.2 mit der Position der virtuellen Ansicht in der Rohaufnahme  $v_x = 1280$  und der Position der Hauptachse  $h = 1000$  zu  $d' = 323$  (vgl. Abschnitt 4.3.1). Die Entfernung  $p_{z\ mm}^w$  kann nun mit Formel 2.2 berechnet werden:

$$\begin{aligned} p_z^w &= f \cdot \frac{b}{d'} \\ p_{z\ mm}^w &= \frac{p_z^w}{u} = 743,71mm \end{aligned}$$

Für die 82 cm entfernte Ebene mit  $d' = 296$  ergibt sich analog:

$$p_{z\ mm}^w = 811,55mm$$

Die Ergebnisse deuten auf eine relativ gute Messgenauigkeit dieses System hin. Weiter bestätigt wird dieser Eindruck durch Messungen der durchschnittlichen Disparität einzelner Spalten. In den Diagrammen 5.6 und 5.7 sind diese Ergebnisse mit dem Gesamtmedian der Disparität abgetragen.

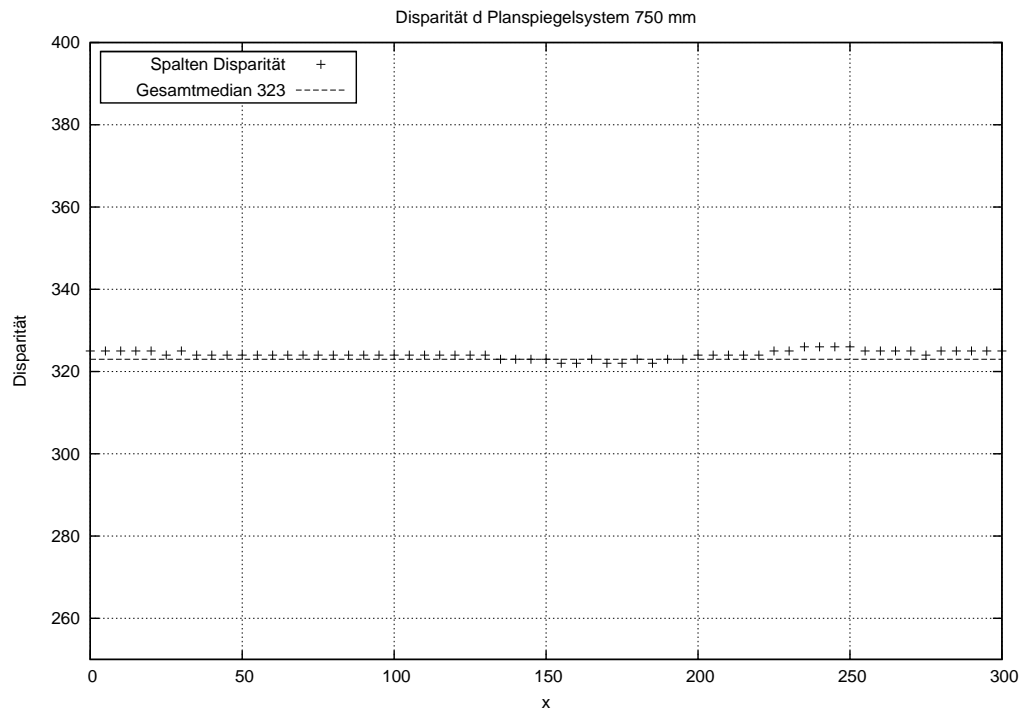


Bild 5.6: Disparität entlang der  $x$ -Achse (Ebenenabstand 75 cm)

### 5.2.3 Ergebnisse

In diesem Abschnitt werden einige der mit diesem System erstellten Tiefenkarten vorgestellt und auf die Schwierigkeiten dieses Systems eingegangen. Das Hauptproblem dieses Systems liegt in der niedrigen Erkennungsrate beim *Stereo Matching*, wie in den beiden Bildserien 5.8 und 5.9 deutlich zu erkennen ist.

In den weniger beleuchteten Bereichen der Bildserie 5.8 konnten deutlich weniger Stereokorrespondenzen gefunden werden. Hier wird deutlich wie wichtig eine gute Ausleuch-

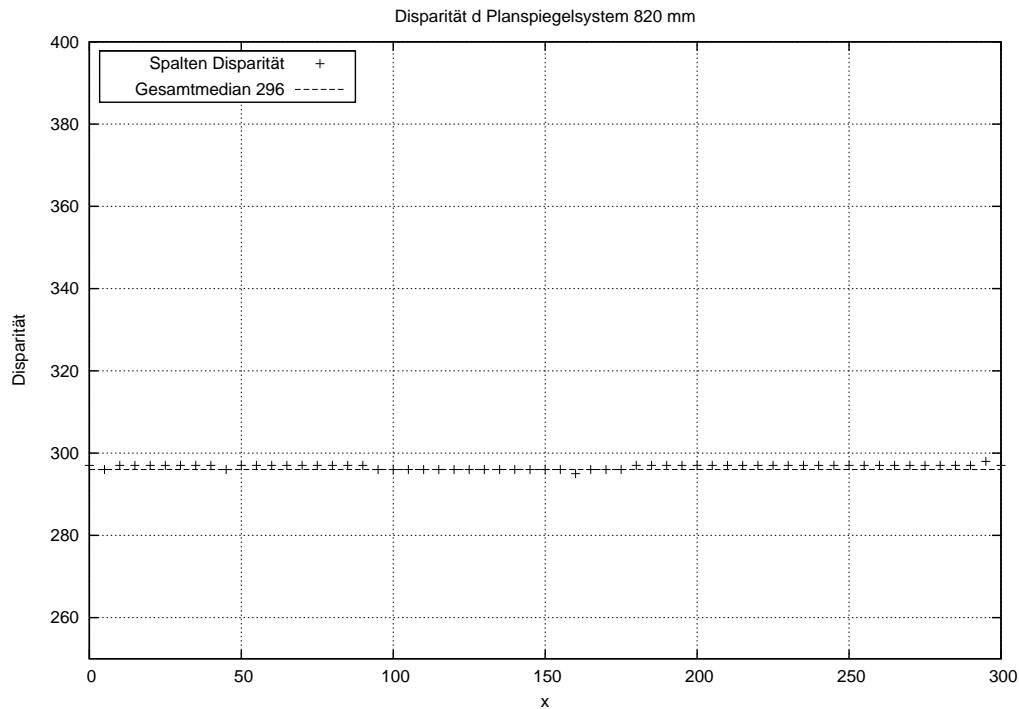


Bild 5.7: Disparität entlang der x-Achse (Ebenenabstand 82 cm)

tung des zu scannenden Objektes ist. In den Glanzpunkten auf der Oberfläche wurden erstaunlicherweise relativ viele Korrespondenzen gefunden. Das System findet bei dunkleren Oberflächen mehr Korrespondenzen als im relativ blassen realen Gesicht in Bildserie 5.9. Insgesamt kann die Rekonstruktion von Tiefe in diesem System als nicht ausreichend bezeichnet werden, dies liegt an der geringen Informationsdichte der Oberflächenstruktur. Für weitere Tests rektifizierender Planspiegelsysteme sollte deshalb zur Verbesserung der Ergebnisse eine höher auflösende Kamera und eventuell ein besserer Spiegel eingesetzt werden.



Bild 5.8: links: direkte Ansicht, mitte: reale Ansicht, rechts: Tiefenkarte (aufgehellt)



Bild 5.9: links: direkte Ansicht, mitte: reale Ansicht, rechts: Tiefenkarte (aufgehellt)

## 5.3 Radiales katadioptrisches Stereosystem

### 5.3.1 Genauigkeit

In diesem Abschnitt wird die Genauigkeit des radialen Stereosystems untersucht. Die Herangehensweise entspricht der des Planspiegelsystems und leidet deshalb unter den gleichen Ungenauigkeitsfaktoren. Es wurden ebenfalls parallel zur Bildebene stehende ebene Flächen mit dem Ziel aufgenommen, die Genauigkeit des aufgebauten radialen katadioptrischen Stereosystems zu testen.

Zunächst muss die Stereobasis des Systems berechnet werden. Wie bei der Extraktion des

relevanten Bereichs (vgl. Abschnitt 4.4.1) ermittelt, hat die Eingangsöffnung des Spiegels den Durchmesser  $d_o = 2390$ . Mit der durch Formel 4.5 ermittelten Brennweite  $f$  und der *Equisolid-Angle*-Abbildungsfunktion 2.1 kann der Bildwinkel  $\omega$  errechnet werden, den das Gesamtsystem in dieser Konfiguration besitzt.

$$\omega = 2 \cdot \arcsin\left(\frac{d_o}{4f}\right) = 55,67^\circ$$

Mit Hilfe des halben Öffnungswinkels des Spiegels  $\beta = 17,48^\circ$ , des Radius der Spiegelöffnung  $r$  und Formel 3.8 ist es nun möglich die Stereobasis zwischen einer virtuellen und der realen Kamera zu berechnen:

$$s = 2 \cdot \frac{r}{\sin(\theta)} \cdot \sin(\theta - \beta)$$

$$v_r = \cos(\beta) \cdot s = 7966,72$$

Die Ebene wurde in den Abständen 31 cm, 43 cm, 50 cm, 55 cm und 64 cm aufgenommen. Mit Hilfe des Umrechnungsfaktors  $u = 164,06$  (vgl. Formel 5.2) und der Formel 2.2 kann die Entfernung der Ebene  $p_z^w$  und die Ground Truth der Disparitäten  $d_{groundtruth}$  berechnet werden. In Tabelle 5.2 werden diese gegenübergestellt.

Tabelle 5.2: Gemessene Disparität im Vergleich zur Ground Truth

$p_z^w mm$	$p_z^w$	$d_{groundtruth}$	$d$
310	50861,7	200,46	47
430	70550,1	144,52	27
500	82035,0	124,29	18
550	90238,5	112,99	14
640	1005004,8	97,10	9

Die gemessenen Disparitätswerte weichen deutlich von der Ground Truth ab. Die Vermutung liegt nahe, dass die Rektifizierung die eigentliche Disparität zwischen den Ansichten verändert. Es konnte lediglich eine Funktion gefunden werden, die in der Lage ist den

Fehler in dem Bereich von 31 - 64 cm ungefähr auszugleichen und es ist somit nur schwer möglich, eine Aussage über die Genauigkeit des Systems zu treffen. Diese approximierte Funktion hat die Form  $d' = (d+x) \cdot y$  und wurde mit Hilfe der Daten der Abstände 430 mm und 640 mm berechnet ( $x = 27,86$  und  $y = 2,6344$ ). Tabelle 5.3 zeigt die approximierte Disparitäten  $d'$ .

Tabelle 5.3: approximierte Disparität

$d_{groundtruth}$	$d$	$d'$
200,46	47	197,21
144,52	27	144,52
124,29	18	120,81
112,99	14	110,28
97,10	9	97,10

### 5.3.2 Ergebnisse

In diesem Abschnitt werden einige der mit diesem System erstellten Tiefenkarten vorgestellt. In den beiden Bildserien 5.10 und 5.11 ist zu erkennen, dass das System von einer Segmentierung des zu scannenden Bereichs profitieren würde, da so keine Stereokorrespondenzen an unerwünschten Stellen im Bild gefunden würden. Des Weiteren stellen Haare ein Problem dar (vgl. Abbildung 5.11), wenn ein 3D-Modell des Gesichts erstellt werden soll. Deshalb wäre es besser, wenn die zu scannende Person eine Kopfbedeckung in der Farbe des Hintergrunds tragen würde, die ihre Haare verdeckt. Die wichtigste Bedingung ist jedoch, dass sich keine Haarsträhnen im Gesicht befinden. Ebenso stellen überbelichtete Bereiche, wie die Stirn in Abbildung 5.10 aufgrund der dort schlechter zu erkennenden Oberflächenstruktur ein Problem für das *Stereo Matching* dar.

Dort wo sich die sternförmig verlaufenden Epipolarlinien schneiden, kommt es zu relativ vielen Fehlkorrespondenzen. Dies liegt daran, dass die Abtastrate in der Nähe des Zentrums sehr hoch ist und Punkte deshalb mehrfach für viele der hier zusammenlaufenden Epipolarlinien *gesampelt* werden. Sie weisen somit in den getrennten Ansichten eine sehr



große Ausdehnung bei sehr geringer Informationsdichte auf (vgl. Abschnitt 4.4.3). In Folge dessen können dort nur wenige gute *Features* gefunden werden.



Bild 5.10: links: Rohaufnahme, rechts: Tiefenkarte (aufgehellt)

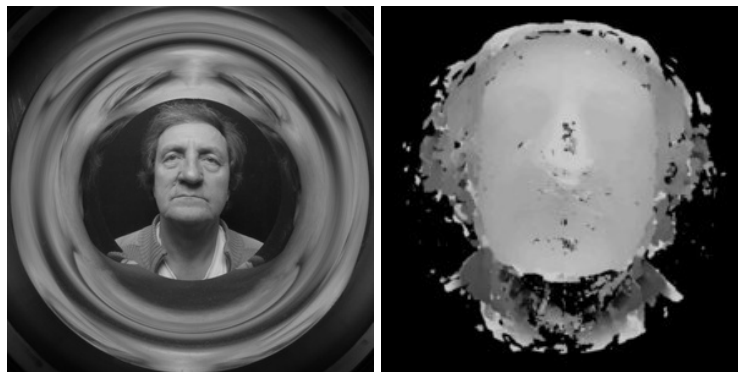


Bild 5.11: links: Rohaufnahme, rechts: Tiefenkarte (aufgehellt)



# Kapitel 6

## Zusammenfassung und Ausblick

In dieser Diplomarbeit sollte ein Überblick über katadioptrische Stereosysteme geben und zwei solcher Systeme gebaut werden. Implementiert wurde zum einen ein katadioptrisches Stereosystem, welches mit Hilfe einer digitalen Kompaktkamera und eines handelsüblichen Spiegels realisiert wurde. Aufgrund der besonderen Anordnung von Kamera und Spiegel liefert dieses System rektifizierte Ansichten der Szene (vgl. Abschnitt 4.2.1). Ziel bei diesem System war es mit möglichst geringen Kosten eine Stereokamera herzustellen. Zum anderen wurde ein Stereosystem implementiert, welches aus einem konischen Spiegel und einer Spiegelreflexkamera mit einem Fisheye-Objektiv besteht. Dieses System nimmt die Szene aus unendlich vielen und auf einer Ortslinie liegenden virtuellen Viewpoints auf (vgl. Abschnitt 4.2.2). Hier war das Ziel die Möglichkeiten gekrümmter Spiegel für den Einsatz in katadioptrischen Stereosystemen zu erforschen.

In Kapitel 5 wurden diese beiden Systeme getestet. Anhand des Planspiegelsystems wurde der Einfluss des Unschärfeausgleichs untersucht. Dieser verbessert das *Stereo Matching* zwar wie erhofft, jedoch stellt dieses Vorgehen einen Kompromiss dar. Durch das Unschärfezeichnen der direkten Ansicht werden sich die beiden Ansichten zwar ähnlicher, es gehen jedoch auch wertvolle Informationen verloren, die sonst das *Matching* verbessert hätten.

Außerdem wurde die Genauigkeit der Systeme untersucht. Das Planspiegelsystem lieferte insgesamt eine ausreichende Genauigkeit, die entstehenden Tiefenkarten sind jedoch sehr

lückenhaft, da das *Stereo Matching* bei der schlechten Bildqualität des Systems Schwierigkeiten hatte, ausreichend viele Korrespondenzen zu finden. Die Genauigkeit des radialen Stereosystems ließ sich nur bedingt beurteilen, da die Disparität zwischen den Ansichten durch die Rektifizierung verändert wird und es dadurch nicht mehr möglich ist die Tiefe direkt zu triangulieren. Aus diesem Grund mussten sie durch eine Ausgleichsfunktion korrigiert werden, deren Genauigkeit weiter untersucht werden sollte. Die Ergebnisse dieses Systems zeigen jedoch insgesamt eine gute Qualität, die sich sicher noch weiter steigern ließe, indem untenstehende Verbesserungen ins System integriert werden.

In der Anfangsphase dieser Diplomarbeit wurden die Komponenten für die zu bauenden Systeme ausgewählt und Hersteller gesucht, die in der Lage waren den Prototyp des konischen Spiegels herzustellen. Die Wahl fiel auf das Labor für Mikrozerspannung der Universität Bremen, welches den Spiegel im Diamantdrehverfahren aus einer Aluminiumlegierung fertigte (vgl. Abschnitt 4.2.2). Da sich dieser Prozess und vor allem die Lieferung des Spiegels bis drei Monate vor Ablauf der vorgesehenen Zeitspanne für die Diplomarbeit hinausgezögerte, war es erforderlich die Grundlagen und Algorithmen anhand von synthetischen Bildern zu erforschen. Hierfür kam das Programm POV-Ray<sup>1</sup> zum Einsatz, welches für diesen Zweck leicht modifiziert werden musste (vgl. Abschnitt B.2).

Die Entwicklung der Algorithmen mit den auf Basis von POV-Ray entstandenen Bildern hat insgesamt gut funktioniert. Lediglich die Rektifizierungsfunktion musste angepasst werden, da sie sich zwischen synthetischem und realem System unterscheidet (vgl. Abschnitt 5.1). Ein weiterer Unterschied zwischen synthetischen und realen Bildern ist, dass es in den durch einen Raytracer entstandenen Bildern weder Tiefenunschärfe noch die durch *Field Curvature* verursachte Unschärfe existiert. Aus diesem Grund war es leider nicht möglich den Unschärfeausgleich für das radiale Stereosystem rechtzeitig fertig zu stellen. Eine direkte Folge davon ist, dass das *Stereo Matching* in den unscharfen Bereichen keine ausreichenden Ergebnisse mehr lieferte. Wie sehr ein Unschärfeausgleich das *Matching* verbessert, kann in Test 5.2.1 beobachtet werden. Hinzu kommt, dass die linke Ansicht zum linken Rand hin nicht nur unschärfer wird, sondern auch Teile des zu scannenden Bereichs abgeschnitten werden. Das Gleiche gilt analog für den rechten Rand der rechten Ansicht (vgl. Abschnitt 4.4.5). In den Tiefenkarten entspricht dies jeweils den

---

<sup>1</sup><http://www.povray.org>

gegenüberliegenden Seiten. Die einfachste Lösung dieses Problems war deshalb, die qualitativ besseren Hälften der beiden entstandenen Tiefenkarten zu einer gemeinsamen zu vereinen (vgl. Abbildung 6.1).

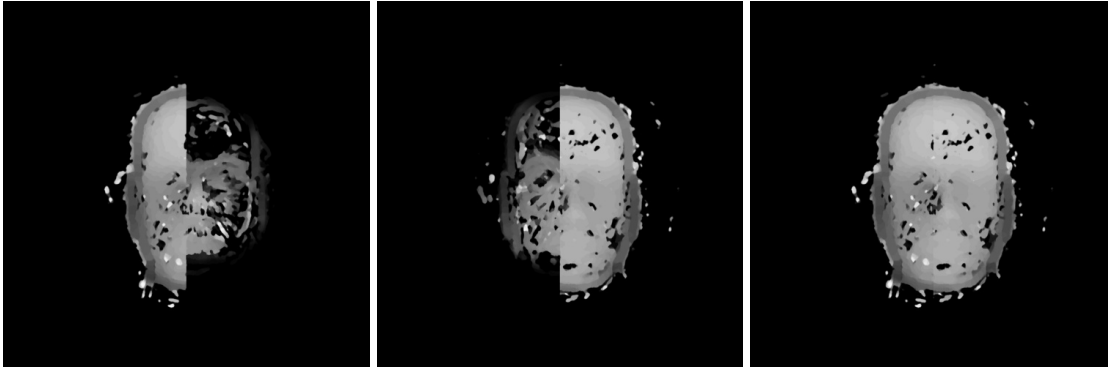


Bild 6.1: links: Tiefenkarte linke mit direkter Ansicht, mitte: direkte mit rechter Ansicht, rechts: kombinierte Tiefenkarte (aufgehellt)

Ein weiteres Problem entsteht bei der Rektifizierung des nicht-orthogonalen Spiegels dieses Systems, da hierbei die virtuellen Ansichten neu positioniert werden und sich dadurch die Disparität zwischen den virtuellen sowie der realen Ansicht verändert. Der relative Tiefenunterschied zwischen den Ansichten bleibt erhalten, eine direkte Triangulierung der Tiefe mit Hilfe der Stereobasis und der Brennweite ist aber nicht mehr möglich (vgl. Abschnitt 5.3.1). Weitere Untersuchungen müssen zeigen, ob die gefundene approximierende Ausgleichsfunktion ausreicht diesen Fehler zu kompensieren. Eventuell müssen andere Wege gefunden werden die virtuellen Ansichten zu rektifizieren. Beispielsweise wäre eine analytische Lösung des Problems denkbar.

Als äußerst problematisch hat sich erwiesen, lediglich unzureichende Halterungen für die jeweiligen Systeme einzusetzen. Um qualitativ bessere und vergleichbarere Aufnahmen herstellen zu können, wären bei beiden Halterungen präzisere Einstellungsmöglichkeiten und eine bessere Fixierung des Gesamtsystems notwendig gewesen. Geringfügige Einwirkungen, wie das An- und Ausschalten oder das Wechseln des Akkus, zog oftmals eine Neukalibrierung des Systems nach sich. Da diese Veränderungen nicht sofort sichtbar waren, kam es so häufig zu vermeidbaren Ungenauigkeiten.

Ebenso könnten die Ergebnisse verbessert werden, wenn die Umgebungsbedingungen weniger variieren würden und durch professionelle Beleuchtung Studiobedingungen hergestellt werden könnten. In dem hier vorgestellten System kamen zwei Baustrahler mit jeweils 500 Watt Leistung zum Einsatz. Es wurde versucht eine schattenarme Ausleuchtung zu erzeugen, indem weiße Regenschirme als Diffusor verwendet wurden. Mit professionellem Equipment wäre sicher eine weitere Steigerung der Aufnahmequalität zu erzielen.

Die Segmentierung von Gesichtern könnte durch die Verwendung der Greenscreen-Technik verbessert werden, da das menschliche Gesicht nur einen sehr geringen Anteil dieses Grüntons aufweist. Da Haar, besonders voluminösere Frisuren ein Problem für die Rekonstruktion von 3D darstellen, würde das Tragen einer Mütze in entsprechender Farbe in solchen Fällen die Ergebnisse verbessern. Ferner könnte die Kalibrierung einfacher gestaltet werden, indem einige der Kalibrierungseinstellungen, wie beispielsweise das Auffinden des inneren und äußeren Spiegelrands im radialen Stereosystem, automatisiert werden würden (vgl. Abschnitt 4.4.1).

Des Weiteren verschmutzt der Spiegel im momentanen Aufbau sehr leicht, da er zumindest für Aufnahmen, aber auch sonst nicht vollständig abgedeckt gehalten werden kann. Es sollte getestet werden, ob die Ergebnisse durch eine vor dem Spiegel platzierte dünne Glasplatte negativ beeinflusst wird. Falls dies nicht der Fall ist, sollte die Konstruktion entsprechend angepasst werden.

Um einen gleichbleibenden Abstand zwischen Objekt und Spiegel zu gewährleisten, wäre außerdem ein außerhalb des Sichtfelds des Gesamtsystems angebrachter Abstandshalter von Vorteil, da für einen kleineren Bereich von Objektabständen die Ausgleichsfunktion besser approximiert werden könnte.

Das Ausrichten des Planspiegelsystems (vgl. 4.3) funktioniert trotz der einfachen Hilfsmittel sehr gut und liefert außerordentlich gut rektifizierte Ansichten der Szene. Es können mit diesem System jedoch keine wirklich dichten Tiefenkarten erstellt werden, da das *Stereo Matching* keine ausreichend guten Ergebnisse liefert. Die Gründe hierfür liegen darin, dass der Haushaltsspiegel es nötig macht die qualitativ bessere direkte Ansicht mit relativ starker Unschärfe zu versehen, damit mehr korrespondierende Pixel gefunden werden können. Die Auflösung der Kamera, die durch die systembedingte Reduzierung (vgl. Abschnitt 3.3) noch weiter verkleinert wird, lässt es nicht zu die für gutes *Matching* benötigte

Oberflächenstruktur des zu scannenden Objektes sichtbar zu machen. Hinzu kommt, dass es mit der verwendeten Kompaktkamera kaum Möglichkeiten gibt in den Aufnahmeprozess einzugreifen. Beispielsweise bereitet die Belichtungszeit nicht einstellen zu können, Probleme dabei das Beste aus der gegebenen Beleuchtungssituation zu machen. Die Auswirkungen der Beleuchtung auf dieses System sind enorm, da die Oberfläche von heller Haut schon bei geringer Überbelichtung viel ihrer Struktur verliert. Für weitere Untersuchungen mit dieser Art von Stereosystem wäre es deshalb von Vorteil auf eine bessere oder zumindest höher auflösende Kamera zurückzugreifen.

Beim radialen System hat der Spiegel hervorragende Reflexionseigenschaften, lediglich die durch Tiefenunschärfe und *Field Curvature* bedingte Unschärfe am äußeren Rand verringert die Qualität der Aufnahmen etwas. Durch die hohe Auflösung und die vielfältigen Einstellungsmöglichkeiten der verwendeten Kamera beeinflusst die Beleuchtung der Szene die Ergebnisse in geringerem Maß als dies bei dem Planspiegelsystem der Fall ist.

Zusammenfassend ist zu sagen, dass es mit beiden Systemen möglich ist 3D zu rekonstruieren, wobei beide Systeme aus den genannten Gründen etwas hinter den Erwartungen zurückgeblieben sind. Werden die beschriebenen Probleme beseitigt und eventuell bessere *Stereo-Matching*-Verfahren eingesetzt (für mögliche Verfahren siehe [SS02]), so sollte es möglich sein die Qualität deutlich zu erhöhen. Würde der radiale Spiegel als ein auf ein Fisheye-Objektiv montiertes Objektiv realisiert und in Serie gefertigt werden, so würden sich die Kosten hierfür wahrscheinlich innerhalb üblicher Objektivpreise bewegen. Somit wäre es zum einen möglich einen Nahbereichs-Objektscanner zu den Kosten einer Fotoausrüstung zu erhalten und zum anderen werden durch die direkte Befestigung an der Kamera optimale Bedingungen für beste Aufnahmequalität geschaffen. Aufgrund der Mobilität eines solchen Scanners wären viele Einsatzgebiete vorstellbar, besonders falls sich durch die dritte Dimension eine Gesichtserkennung verbessern ließe. Denkbar wären Einsatzmöglichkeiten in der Portrait-Fotografie oder der Einsatz in Streifenwagen zur besseren Identifikation von Personen.

Noch höhere Auflösungen scheinen kein Verkaufsargument für Digitalkameras mehr darzustellen, weshalb die Kameras mit zusätzlichen *Features* ausgestattet werden müssen. Hier kommt das Konzept der *Computational Camera* ins Spiel. Das Interesse an Bildern mit großem Dynamikumfang oder die sich großer Beliebtheit erfreuende Panorama-

Fotografie hat gezeigt, dass viele Anwender sogar bereit sind einen relativ hohen Aufwand auf sich zu nehmen, um an solche Bilder zu gelangen. Eine „Ein-Knopf“-Variante dieser und anderer Verfahren (vgl. 3.2) würde sicher viele Anwender finden. Auf diesem Weg könnten auch katadioptrische Stereokameras ihren Markt finden.



# Anhang A

## Softwaredesign

### A.1 Radiale katadioptrische Stereokamera

Der Leitgedanke beim Softwaredesign des radialen Stereosystems war es, eine katadioptrische Kamera (Conical Catadioptric Camera - *CCCamera*) aus verschiedenen Modulen zusammenzusetzen. Diese Module werden von der *CCCamera*-Klasse wiederum zu einer Art Pipeline zusammengefügt. Zunächst wird der relevante Bereich des Bildes extrahiert (*Extractor*), danach werden die Fisheye-Verzerrungen beseitigt (*FisheyeRemapper*), die drei verschiedenen Ansichten getrennt (*CCCSplitter*), die in den nicht orthogonalen Spiegeln entstandenen Ansichten rektifiziert (*CCCRectifer*) und nach einem Unschärfeausgleich<sup>1</sup> (*CCCBlur*, Blur mit einem graduellen Verlauf) gematcht (*StereoMatcher*). Dabei sind Klassen, die ausschließlich in einer solchen katadioptrischen Kamera verwendet werden können, durch das Präfix *CCC* gekennzeichnet (vgl. Abbildung A.1). Ein analoges Design kam bei dem deutlich einfacher strukturierten Planspiegelsystem zum Einsatz.

---

<sup>1</sup>Der Unschärfeausgleich konnte leider nicht fertig gestellt werden.

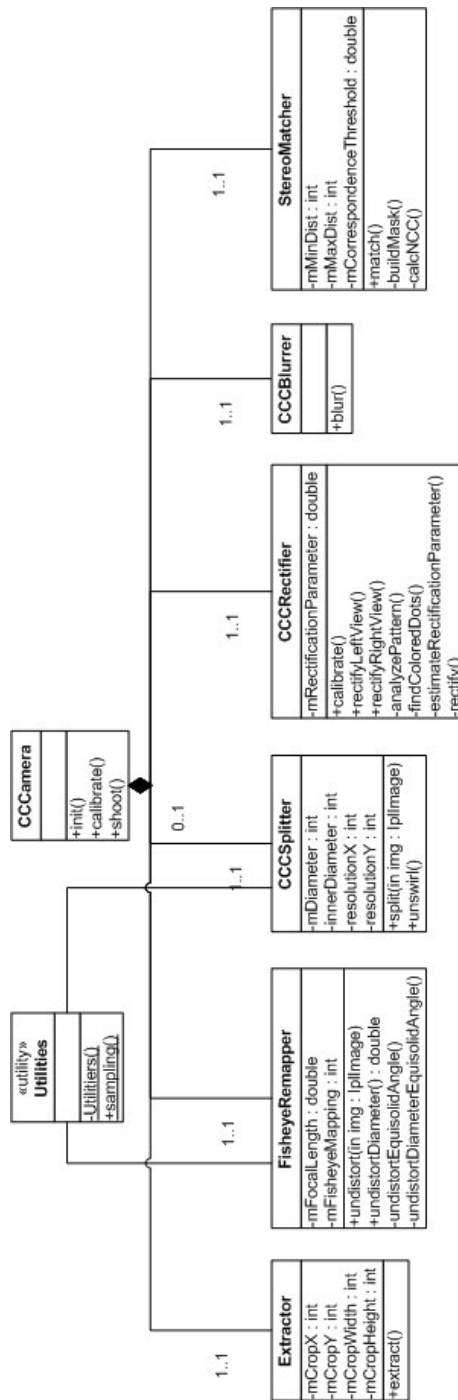


Bild A.1: UML Klassendiagramm der radialen katadioptrischen Stereokamera

# Anhang B

## Quelltexte

### B.1 PostScript

#### B.1.1 Kalibrierungsmuster für das Gesamtsystem

```
%!PS-Adobe-2.0
/cm {28.34645669 mul} def

/halfPageWidth 297.5 def
/halfPageHeight 421 def

0 setgray
4 setlinewidth

halfPageWidth halfPageHeight translate % move cursor to center

0 0 .5 0 360 arc stroke % center dot
0 0 8.5 cm 0 360 arc stroke % radius cone outside

1 setlinewidth
0 setgray
/radius 0 def
8 8 240
```

```

{
    /radius exch def
    0 0 radius 0 360 arc stroke
} for

% draw cross
0 0 moveto
0 halfPageHeight rlineto
0 0 moveto
0 -1 halfPageHeight mul rlineto
0 0 moveto
halfPageWidth 0 rlineto
0 0 moveto
-1 halfPageWidth mul 0 rlineto
% diagonal
0 0 moveto
halfPageWidth halfPageHeight rlineto
0 0 moveto
-1 halfPageWidth mul halfPageHeight rlineto
0 0 moveto
halfPageWidth -1 halfPageHeight mul rlineto
0 0 moveto
-1 halfPageWidth mul -1 halfPageHeight mul rlineto
stroke

showpage

```

## B.2 Equisolid-Angle-Abbildungsfunktion für POV-Ray

Auszug aus der Datei „*render.cpp*“ von POV-Ray, der eine neue Kamera mit *Equisolid-Angle-Abbildungsfunktion* hinzufügt<sup>1</sup>:

```

/*
* Fisheye camera (equisolid angle).

```

<sup>1</sup>Weitere Anpassungen in den Dateien: „*camera.h*“, „*parse.cpp*“, „*parse.h*“ und „*tokenize.h*“ sind notwendig um die Kamera in POV-Ray einzubinden.

```
* R = 2 * f * sin(theta/2)
*/

case FISHEYE_EQUISOLID_CAMERA:

// Convert the x coordinate to be a DBL from -1.0 to 1.0.
tmp = min((DBL)Frame.Screen_Width, (DBL)Frame.Screen_Height);
x0 = 2.0 * x / tmp - 1.0;
y0 = 2.0 * y / tmp - 1.0;

if (Precompute_Camera_Constants)
{
VLength(lx, FCR);
VLength(ly, FCU);
}

R = sqrt(x0*x0 + y0*y0);

// f = R / (2 * sin(theta/2))
// max R is 1.0
tmp = (Frame.Camera->Angle / 360) * M_PI; // conversion DEG to RAD of half the angle of view
focal_length = 1.0 / (2 * sin(tmp/2));

// If the pixel lies outside the unit circle no ray is traced.
if (R > 1.0)
{
return(false);
}
if (R == 0.0)
{
X = 0.0;
Y = 0.0;
Z = 1.0;
}

theta = 2 * (asin(R/(2*focal_length)));

Make_Vector(Xc, 1.0, 0.0, 0.0);
```

```
Make_Vector(P, x0, y0, 0.0);
VDot(tmp, Xc, P);
DBL l1;
VLength(l1, Xc);
DBL l2;
VLength(l2, P);

if (y0 > 0.0)
{
phi = 2*M_PI - acos(tmp / (l1 * l2));

X = R * sin(theta) * cos(phi);
Y = R * sin(theta) * sin(phi);
Z = R * cos(theta);

}
else
{
phi = acos(tmp / (l1 * l2));

X = R * sin(theta) * cos(phi);
Y = R * sin(theta) * sin(phi);
Z = R * cos(theta);

/* equidistant
tmp = sqrt(x0*x0 + y0*y0);
X = (x0/tmp)*(sin(tmp));
Y = (y0/tmp)*(sin(tmp));
Z = cos(tmp);*/

/* stereographic
X = (2*x0) / (1 + x0*x0 + y0*y0);
Y = (2*y0) / (1 + x0*x0 + y0*y0);
Z = (1 - (x0*x0 + y0*y0)) / (1 + x0*x0 + y0*y0);*/
}

if (Precompute_Camera_Constants)
{
```

```
VNormalize(FCR, FCR);
VNormalize(FCU, FCU);
VNormalize(FCD, FCD);
}

VLinComb3(Ray->Direction, X, FCR, Y, FCU, Z, FCD);

if (Focal_Blur_Is_Used)
{
jitter_camera_ray(Ray, ray_number);
}

initialize_ray_container_state(Ray, Precompute_Camera_Constants);

Precompute_Camera_Constants = false;

break;
```





# **Anhang C**

## **Konstruktionszeichnung**

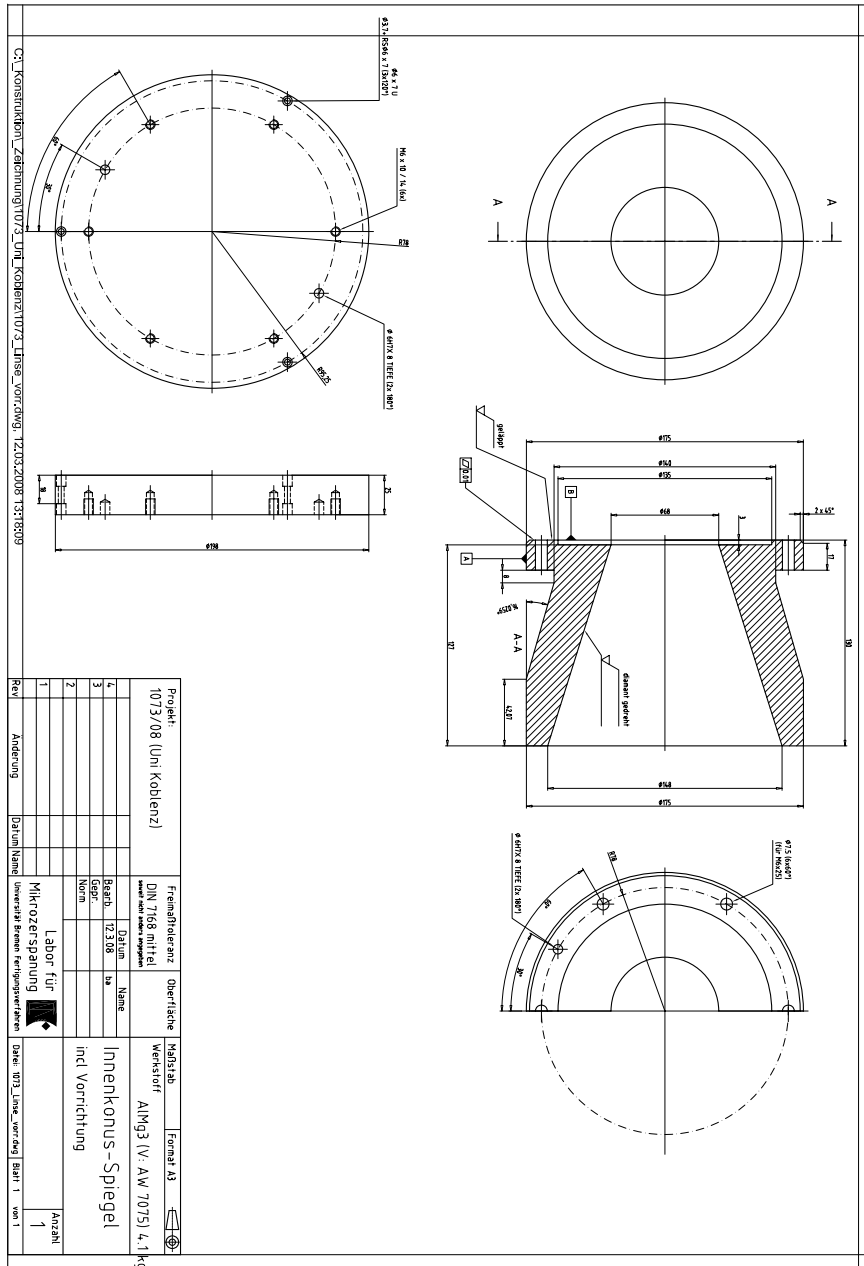


Bild C.1: Konstruktionszeichnung des innen verspiegelten konischen Rohrs

# Literaturverzeichnis

- [BN99] BAKER, S. ; NAYAR, S.K.: A Theory of Single-Viewpoint Catadioptric Image Formation. In: *International Journal on Computer Vision* 35 (1999), Nov, Nr. 2, S. 175–196
- [FHM<sup>+</sup>93] FAUGERAS, Olivier ; HOTZ, B. ; MATHIEU, Hervé ; VIÉVILLE, T. ; ZHANG, Zhengyou ; FUA, Pascal ; THÉRON, Eric ; MOLL, Laurent ; BERRY, Gérard ; VUILLEMIN, Jean ; BERTIN, Patrice ; PROY, Catherine: Real time correlation based stereo: algorithm implementations and applications / INRIA. Version: 1993. <http://perception.inrialpes.fr/Publications/1993/FHMOVZFTMBVBP93>. 1993 (RR-2013). – Forschungsbericht
- [GG93] GOSHTASBY, A. ; GRUVER, W. A.: Design of a Single-Lens Stereo Camera System, 1993
- [GN01] GLUCKMAN, J. ; NAYAR, S.K.: Catadioptric Stereo Using Planar Mirrors. In: *International Journal on Computer Vision* 44 (2001), Aug, Nr. 1, S. 65–79
- [HZ05] HECHT, E. ; ZAJAC, A.: *Optics*. 4th. Addison Wesley, 2005
- [IHI93] INABA, M. ; HARA, T. ; INOUE, H.: A Stereo Viewer Based on a Single Camera with View-Control Mechanism. In: *International Conference on Robotics and Systems*, 1993

- [KN06] KUTHIRUMMAL, S. ; NAYAR, S. K.: Multiview Radial Catadioptric Imaging for Scene Capture. In: *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)* (2006), Jul
- [Nay97] NAYAR, S.K.: Catadioptric Omnidirectional Camera. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997, S. 482–488
- [Nay06] NAYAR, S. K.: Computational Cameras: Redefining the Image. In: *IEEE Computer Magazine, Special Issue on Computational Photography* (2006), Aug, S. 30–38
- [NN98] NENE, S.A. ; NAYAR, S.K: Stereo with Mirrors. In: *6th International Conference on Computer Vision*, 1998
- [NP99] NAYAR, S.K. ; PERI, V.: Folded Catadioptric Cameras. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Bd. 2, 1999, S. 217–223
- [SBF96] SOUTHWELL, D. ; BASU, A. ; FIALA, M.: Panoramic Stereo. In: *International Conference on Pattern Recognition*, 1996
- [Sch02] SCHRÖDER, Gottfried: *Technische Optik*. 8. Vogel Verlag Und Druck, 2002
- [SS02] SCHARSTEIN, Daniel ; SZELISKI, Richard: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: *International Journal of Computer Vision* 47 (2002), S. 7–42