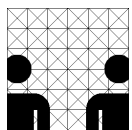


Diplomarbeit  
im Studiengang Informatik

# Grundlagen und Methodik der 3D-Rekonstruktion und ihre Anwendung für landmarkenbasierte Selbstlokalisierung humanoider Roboter

am Arbeitsbereich für  
Technische Aspekte Multimodaler Systeme,  
Universität Hamburg

vorgelegt von  
**Denis Klimentjew**  
**Andre Stroh**  
Januar 2008



betreut von  
Prof. Dr. Jianwei Zhang  
Dr. Werner Hansmann





*“Es ist nicht schwierig, Computer zu bauen, die Schach spielen oder Rechenaufgaben lösen können. Computern fällt es leicht, zu tun, was man in der Schule lernt. Schwierigkeiten haben sie jedoch, zu lernen, was Kinder lernen, bevor sie in die Schule kommen: eine auf dem Kopf stehende Tasse als Tasse zu identifizieren, sich in einem Garten zurechtzufinden, ein Gesicht wiederzuerkennen, zu sehen.”*

Tor Nørretranders



## **Zusammenfassung**

In der Robotik, besonderes im Bereich humanoider Roboter, bilden die Stereokamerasysteme nach wie vor eine der wichtigsten Schnittstellen des Roboters zur Außenwelt. Der Vorteil dieser Systeme liegt in der Möglichkeit, die Umwelt wahrzunehmen und auf Ereignisse gezielt zu reagieren. Dabei stellt der Verlust der Tiefeninformation bei der Bildaufnahme einen der größten Nachteile dieser Systeme dar. Kann die Tiefe rekonstruiert werden, bietet ein Stereokamerasystem eine ausreichende Menge an Daten, um die Navigation des Roboters zu realisieren.

Die vorliegende Arbeit beschäftigt sich mit den Grundlagen und Methoden der 3D-Rekonstruktion, speziell anhand der Daten eines Stereokamerasystems, mit dem Ziel, die Tiefeninformationen zu bestimmen. Dadurch wird es möglich, jedem Objekt der Szene eine eindeutige Position im Raum zuzuordnen. Darauf basierend werden die Landmarken detektiert und ihre relativen Positionen in Bezug auf den Roboter berechnet. Anschließend wird der Roboter anhand der vorhandenen Daten, also der absoluten Positionen der Landmarken im Raum, lokalisiert.

## **Abstract**

In the field of autonomous robots, particularly in the domain of humanoid robots, stereo vision systems are still the most important interface between the robot and the outside world. The advantage of these systems lies in the possibility to perceive the environment and to specifically react to events. One of the biggest disadvantages of this approach lies in the loss of the depth information during image acquisition. If the depth can be reconstructed, stereo camera systems deliver enough data for the navigation of robots.

This thesis deals with the theoretical foundations and methods of dimensional reconstruction, particularly based on the data of a stereo vision system, with the purpose of determining the depth information. This enables us to assign an unequivocal position in space to every object in the scene. Based on this, the guiding points will be detected and their relative position calculated in relation to the robot. Then we localise the robot on the basis of the available data, i.e., the absolute position of the guiding points in space.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation und Ziel dieser Arbeit . . . . .	3
1.2	Vergleichbare Arbeiten . . . . .	5
1.2.1	Statische Stereoanalyse . . . . .	5
1.2.2	Selbstlokalisierung anhand von Landmarken . . . . .	6
1.3	Gliederung der Arbeit . . . . .	8
1.4	Hinweise . . . . .	9
<b>2</b>	<b>Grundlagen der Bildaufnahme</b>	<b>11</b>
2.1	Kamera-Modell . . . . .	11
2.2	Bildaufnahme . . . . .	13
2.3	Zusammenfassung . . . . .	16
<b>3</b>	<b>Kamerakalibrierung</b>	<b>17</b>
3.1	Bestimmung der Kameraparameter . . . . .	18
3.2	Intrinsische Kameraparameter . . . . .	19
3.3	Extrinsische Parameter . . . . .	20
3.4	Verzerrungen . . . . .	21
3.5	Kamerakalibrierungsverfahren . . . . .	23
3.5.1	Nach Roger Tsai . . . . .	24
3.5.2	Nach Zhengyou Zhang . . . . .	26
3.5.3	Kamerakalibrierung eines Stereosystems . . . . .	32
3.6	Zusammenfassung . . . . .	33
<b>4</b>	<b>Statische Stereoanalyse</b>	<b>35</b>
4.1	Die Standardstereogeometrie . . . . .	36
4.2	Konvergente Stereogeometrie . . . . .	38
4.3	Epipolargeometrie . . . . .	40
4.4	Rektifikation der Stereobildpaare . . . . .	42
4.4.1	Rektifikation mit bekannten Kalibrierungsparameter . . . . .	42
4.4.2	Schätzung der Fundamentalmatrix . . . . .	46
4.4.3	Rektifikation mit Homographien . . . . .	50
4.5	Zusammenfassung . . . . .	53
<b>5</b>	<b>Korrespondenzanalyse</b>	<b>55</b>
5.1	Annahmen und Einschränkungen . . . . .	55
5.2	Korrespondenzanalyseverfahren . . . . .	57

5.2.1	Merkmalbasierte Verfahren . . . . .	57
5.2.2	Flächenbasierte Verfahren . . . . .	62
5.3	Zusammenfassung . . . . .	67
<b>6</b>	<b>Tiefenrekonstruktion</b>	<b>69</b>
6.1	Die Stereotriangulation . . . . .	69
6.1.1	Homogenes Lösungsverfahren . . . . .	71
6.1.2	Inhomogenes Lösungsverfahren . . . . .	71
6.1.3	Vergleich der beiden Lösungsverfahren . . . . .	72
6.2	Zusammenfassung . . . . .	72
<b>7</b>	<b>Erkennung und Selbstlokalisierung anhand von Landmarken</b>	<b>73</b>
7.1	Sensorik . . . . .	74
7.2	Landmarken . . . . .	78
7.2.1	Natürliche Landmarken . . . . .	79
7.2.2	Künstliche Landmarken . . . . .	80
7.2.3	Erkennung und Segmentierung passiver künstlicher Landmarken	81
7.3	Zusammenfassung . . . . .	95
<b>8</b>	<b>Ausgangssystem</b>	<b>97</b>
8.1	Hardware . . . . .	97
8.2	Software . . . . .	100
8.2.1	Eingesetzte Software . . . . .	101
8.2.2	Open Computer Vision Library . . . . .	101
<b>9</b>	<b>Experimentelle Ergebnisse</b>	<b>103</b>
9.1	Kamerakalibrierung . . . . .	103
9.1.1	Mit Camera Calibration Toolbox für Matlab . . . . .	104
9.1.2	Mit Algorithmen der OpenCV-Bibliothek . . . . .	111
9.2	Bildvorverarbeitung und Tiefenrekonstruktion . . . . .	117
9.3	Selbstlokalisierung anhand von Landmarken . . . . .	128
9.4	Zusammenfassung . . . . .	135
<b>10</b>	<b>Fazit und Ausblick</b>	<b>137</b>
10.1	Fazit . . . . .	137
10.2	Ausblick . . . . .	138



# Abbildungsverzeichnis

1.1	Modulare Architektur . . . . .	4
2.1	Das Lochkamera-Modell . . . . .	12
2.2	Schematischer Aufbau eines CCD-Sensor . . . . .	14
2.3	Ausleseverlauf der Daten eines CCD-Sensors . . . . .	15
3.1	Beziehungen zwischen den Koordinatensystemen . . . . .	18
3.2	Perspektivische Projektion . . . . .	21
3.3	Tangentiale und radiale Linsenverzerrungen . . . . .	22
3.4	Tonnenförmigen und kissenförmigen Verzerrungen . . . . .	23
3.5	Kalibrierungskörper . . . . .	27
4.1	Standardstereogeometrie . . . . .	37
4.2	Konvergente Stereogeometrie . . . . .	38
4.3	Epipolargeometrie . . . . .	41
4.4	Rektifikation . . . . .	42
5.1	2D Gauß-Operator . . . . .	60
5.2	2D Laplacian of Gaussian-Operator . . . . .	61
6.1	Stereotriangulation . . . . .	70
7.1	Navigationsstruktur eines mobilen Roboters . . . . .	73
7.2	Funktionsprinzip eines optischen Encoders . . . . .	74
7.3	Darstellung eines Laserkopfes . . . . .	77
7.4	Darstellung der Triangulation mit einem 3D Laserscanner . . . . .	78
7.5	Vorgang der Lokalisation . . . . .	81
7.6	Zweifarbiger Landmarken nach dem Verfahren von Yoon und Kweon . . . . .	83
7.7	Ein Beispiel einer Landmarke aus Permutationen mehrerer Farben . . . . .	86
7.8	Darstellung zweier $p$ -similarer Funktionen . . . . .	87
7.9	Die grafische Darstellung der Funktion $S(x)$ . . . . .	90
7.10	Generator- sowie Funktion der max. Unähnlichkeit bzw. Ähnlichkeit . . . . .	92
7.11	Landmarke von Scharnstein und Briggs . . . . .	93
7.12	Erweiterung der Landmarke mit einem Barcode . . . . .	94
8.1	HOAP-2 . . . . .	98
8.2	Gelenke und deren Freiheitsgrade des HOAP-2 Roboters . . . . .	99
8.3	Kommunikation zwischen dem externen Rechner und dem Roboter . . . . .	100

9.1	Oberflächen des Camera Calibration Toolbox für Matlab . . . . .	104
9.2	Auswahl der Eckpunkte . . . . .	106
9.3	Auflistung intrinsischer Parameter . . . . .	107
9.4	Ergebnisse der Stereokamerakalibrierung . . . . .	108
9.5	Extrinsische Parameter der Kamerakalibrierung . . . . .	109
9.6	Detektierte Eckpunkte des Kalibrierungskörpers . . . . .	112
9.7	Intrinsische Parameter für die linke und rechte Kamera . . . . .	112
9.8	Rektifikation nach Fusiello, Trucco und Verri [FTV00] . . . . .	114
9.9	Die nach dem Verfahren von Fusiello, Trucco und Verri berechneten Transformationsmatrizen . . . . .	114
9.10	Ergebnisse der Tiefenschätzung . . . . .	115
9.11	Grafische Darstellung der Anzahl der Gesamtpunkte sowie der Bild- punkte für die keine Korrespondenz gefunden werden konnte. . . . .	116
9.12	In Grauwerte transformiertes Stereobildpaar . . . . .	117
9.13	Stereobildpaar nach durchgeführter Entzerrung . . . . .	118
9.14	Rektifiziertes Stereobildpaar . . . . .	119
9.15	Original- und das resultierende Kantenbild nach Anwendung des LoG	120
9.16	Kantenbild und dazugehörige Disparitätskarte . . . . .	121
9.17	Stereobildpaar einer Tischszene . . . . .	122
9.18	Entzerrtes und rektifiziertes Stereobildpaar. . . . .	123
9.19	Mit dem Schirai-Algorithmus generierte Disparitätskarte . . . . .	124
9.20	Mit Block-Matching generierte Disparitätskarte . . . . .	125
9.21	Mit dynamischer Programmierung generierte Disparitätskarte . . . .	126
9.22	3D-Rekonstruktion der mit dem Schirai-Algorithmus generierten Dis- paritätskarte . . . . .	127
9.23	3D-Rekonstruktion der mit Block-Matching generierten Dispa- ritätskarte . . . . .	127
9.24	3D-Rekonstruktion der mit dem Algorithmus von S. Birchfield und C. Thomasi generierten Disparitätskarte . . . . .	128
9.25	Stereobildpaar mit Landmarken . . . . .	129
9.26	Stereobildpaar mit detektierten Landmarken. . . . .	130
9.27	Grafische Darstellung der Zusammenhänge des zweiten Strahlensatzes	131
9.28	Serie von Bildern mit detektierten Landmarken und dazugehöriger Ausgangsbilder . . . . .	134

# Einleitung

# 1

---

Roboter sind seit mehreren Jahrzehnten ein fester Bestandteil der Industrie, meistens werden diese in einer festdefinierten Umgebung wie zum Beispiel Produktionslinien eingesetzt. Schon 1954 patentierte George Devol einen Industrieroboter, der 1961 in den Werken von General Motors installiert wurde. In Deutschland wurde der erste Industrieroboter 1970 in der Automobilindustrie von Mercedes-Benz in Betrieb genommen.

Die typische Definition eines Roboters aus damaliger Zeit ist die von RIA (Robot Institute of America): *Ein Roboter ist ein programmierbares Mehrzweck-Handhabungsgerät für das Bewegen von Material, Werkstücken, Werkzeugen oder Spezialgeräten. Der frei programmierbare Bewegungsablauf macht ihn für unterschiedliche Aufgaben einsetzbar.*

Seit mehreren Jahren zeichnet sich aber eine Entfernung von stationären klassischen Robotern zu Lauf- oder Servicerobotern hin. Um den Ängsten der Menschen entgegen zu wirken und die bestmögliche Anpassung der Roboter an die menschliche Umgebung zu gewährleisten wird verstärkt die Entwicklung humanoider Roboter angestrebt. Aber nicht nur das Aussehen, sondern auch das Nachbilden menschlicher Fähigkeiten stehen im Mittelpunkt der Forschung.

Dabei stellt die Interaktion mit der Umwelt und den Menschen eine der wichtigsten Voraussetzungen für den erfolgreichen Einsatz der Roboter dar. Dies schließt vor allem präzise Kenntnisse der Umgebung, Wahrnehmung und Reaktion auf bestimmte Ereignisse mit ein, was ein nicht triviales Problem der Robotik darstellt.

Das Sehen und damit verbundene Wahrnehmung weckte schon sehr früh Interesse des Menschen. Schon ca. 300 v. Chr. untersuchte griechische Mathematiker Euklid *die Natur des Sehens mit zwei Augen*. Dabei erkannte er, dass die Welt dreidimensional ist und verfasste ein Buch *Optika*, in dem er als einer der ersten einige Gesetze der Optik formulierte.

Die menschliche Wahrnehmung ist immernoch ein aktuelles Thema der Wissenschaft, zum Beispiel in der Robotik. Dabei wird die Wahrnehmung der Roboter unter Zuhilfenahme einer oder mehreren Kameras, angelehnt an das menschliche

Sehen, untersucht. Dieser Vorgang wird in der Robotik als maschinelles Sehen bezeichnet. Erst die Aufnahme einer Szene aus unterschiedlichen Perspektiven erlaubt die Rekonstruktion der Tiefeninformation, die durch Einzelaufnahmen verloren geht.

Die Evolution des Menschen und die Vielfalt der Tierwelt entwickelten über Jahrtausende Organe, die verschiedenen Aufgaben des Körpers übernehmen und im Großen das Dasein des Wesens sichern. Die Augen samt visuellem Kortex des Menschen dienen dabei der visuellen Erfassung und Verarbeitung der räumlichen Information der Umgebung. Anhand der gewonnenen und zur Referenz stehenden Information kann der Mensch eine Schätzung der räumlichen Tiefe vornehmen und sich in einer Umgebung orientieren. Da die Vorgänge im visuellen Kortex immer noch nicht vollständig entschlüsselt sind, bleibt die dreidimensionale Rekonstruktion der Umgebung in der Robotik angelehnt an das menschliche Sehen ein schwieriges Unterfangen.

Auch beim Menschen ist die Entfernungsmessung ein komplexer Vorgang, der auf mehrere Ebenen verteilt ist. Dadurch dass die Augen des Menschen nah an einander platziert sind, funktioniert die Tiefenwahrnehmung nur auf kurzen Distanzen verlässlich. Die Eigenschaft der guten Tiefenschätzung der Menschen ist eine Symbiose aus beiden Augen, so genanntes Stereosystem, und Erfahrung. Erst das Gehirn wertet die ankommenden Informationen der Augen aus und schätzt mit hinzugezogenen Erfahrungswerten den Abstand zum signifikanten Objekt.

Der Vorgang auf bereits verhangenes Wissen zurückzugreifen bildet die Basis der Disziplinen "Kognitive Systeme und Wissens- und Sprachverarbeitung". Dabei wird versucht mit so genannten "Kognitiven Maps" die Reaktion des Roboters an gewisse Ereignisse anzupassen. Dennoch bleibt es ein diffiziler und langwieriger Prozess, der die Wissenschaftler womöglich noch Jahrzehnte beschäftigen wird.

Aber auch andere Wege zur Verbesserung der Orientierung in der Einsatzumgebung sind möglich, wie zum Beispiel Verwendung mehrerer unterschiedlicher Sensorenarten, Kenntnisse über Objektdimensionen im Raum oder Einsatz von natürlichen oder künstlichen Landmarken. Die Definition des Begriffes *Landmarke*, sowie deren Eigenschaften werden im Kapitel 7 ausführlich behandelt. Dabei stellt die Orientierung und Navigation anhand von Landmarken für den Menschen einen absolut natürlichen Prozess dar. In der Robotik ist dieser Vorgang jedoch mit vielen Problemen verbunden, rauschende Sensordaten, Ungenauigkeiten bei der Kalibrierung, Verdeckungen von Objekten verursachen Fehler und Unsicherheiten bei der Navigation und Orientierung in dreidimensionaler Umgebung.

Die praktische Lösung der Rekonstruktion und Selbstlokalisierung des Roboters in dreidimensionalen Räumen hängt meistens von der Struktur der Umgebung und eingesetzten Hardwarekomponenten ab. Anhand beider Faktoren wird schließlich die bestmögliche Strategie ausgearbeitet und realisiert.

## 1.1 Motivation und Ziel dieser Arbeit

Diese Arbeit entstand am Arbeitsbereich *TAMS*, Technische Aspekte Multimodaler Systeme, der Universität Hamburg in der Fakultät für Mathematik, Informatik und Naturwissenschaften. Unter dem Einfluss der Zusammenarbeit des Arbeitsbereiches mit dem *CINACS*-Projekt (Cross-Modal Interaktion in Natural and Artificial Cognitive Systems / Intermodale Interaktion in natürlichen und künstlichen kognitiven Systemen) und dem Arbeitsbereich *WSV* (Wissens- und Sprachverarbeitung) entstand ein weiterführendes Szenario dessen Grundlagen teilweise die vorliegende Arbeit realisieren sollte.

An dem *CINACS*-Projekt sind Universität Hamburg, mit dem Universitätsklinikum Eppendorf (UKE), Bereichen Biopsychologie und Neuropsychologie, sowie Department Informatik, und die Tsinghua Universität Beijing in China beteiligt.

Ziel dieses Projektes ist das Verstehen der natürlichen Mechanismen der intermodalen Verarbeitung sowie derer Einfluss auf die Wahrnehmung und Verhaltenskontrolle. Außerdem sollen Modelle, Architekturen und Algorithmen für künstliche multimodale Systeme entworfen werden, die an natürliche Systeme angelehnt sind [CIN07].

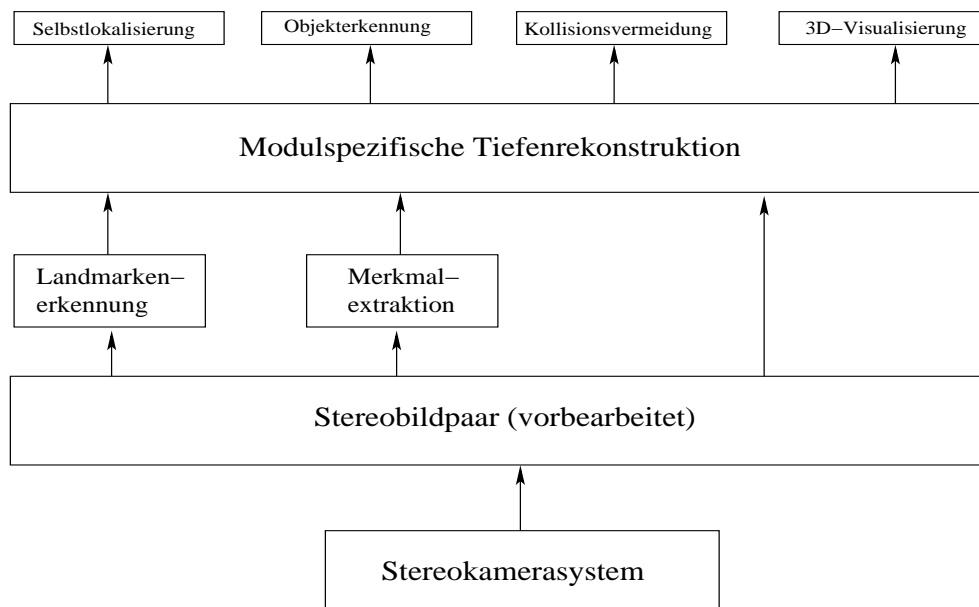
Angeregt durch den *CINACS*-Projekt entwickelten die oben erwähnte Arbeitsbereiche *TAMS* und *WSV* eine weiterführende Idee eines eventuellen Projektes. Dabei soll ein komplexes Umgebungsmodell, Miniaturnachbildung der Hamburger Stadtzentrums gebaut werden, auf der die beiden HOAP 2 Roboter der Fujitsu Automation Co. unter Zuhilfenahme von kognitiven Maps navigieren sollen.

Ziel dieser Arbeit ist eine Ausgangsbasis zu Orientierung der Roboter auf dem Gelände zu realisieren, sodass spätere weitere Ergänzungen der vorhandenen Struktur, durch die kognitiven Maps oder Einsatz anderer künstlicher multimodaler Systeme, einfacher vorgenommen und getestet werden konnte.

Durch den Einsatz beider Roboter wird die Auswahl der Raumerfassungssensoren auf das Stereokamerasystem reduziert, mit deren Hilfe die dreidimensionale Umgebung rekonstruiert werden kann. Da der Abstand der beiden Kameras zu einander gering und die Auflösung der Bilder niedrig ist, kann mit zunehmender Entfernung von immer größeren Verfälschungen der ausgewerteten Sensordaten ausgegangen werden. Außerdem stellt die Selbstlokalisierung in dreidimensionaler Umgebung ohne Hintergrundwissen, nur durch die Entfernungsmessung ein komplizierten und vielschichtigen Prozess dar.

Damit die resultierenden Funktionen und Ergebnisse aus der vorliegenden Arbeit für weitere Entwicklung verwendet werden können, wird zuerst eine universale modulare Architektur entworfen. Durch die universale modulare Struktur können spätere Modifikationen beziehungsweise Austausch einzelner Module leichter vollzogen werden, sodass der gesamte Aufbau erhalten und gegebenenfalls als Testumgebung für

daraufbasierende Algorithmen verwendet wird. Die Ausgangsbasis bilden dabei die von dem Stereokamerasystem gelieferten Stereobildpaare, die durch Linsenentzerrung und Rektifikation vorverarbeitet werden. Die gesamten für die Beschreibung der Architektur genutzten Fachbegriffe werden an den entsprechenden Stellen dieser Arbeit ausführlich erklärt. Dabei wird zuerst auf den Originalbildern nach bestimmten Objekten gesucht, zum Beispiel Landmarken oder anderen physikalischen Gegenständen. Sind die Merkmale extrahiert, werden auf den vorverarbeitenden Bildern die korrespondierende Punkte bestimmt und dem Objekt eine eindeutige Position im Raum durch die Tiefenrekonstruktion zugeordnet. Falls erwünscht kann auch der gesamte Raum rekonstruiert werden. Die resultierende Plattform wird in der Abbildung 1.1 grafisch dargestellt. Dieses auf der Tiefenrekonstruktion basierende Verfahren ermöglicht schnellere Realisierung mehrerer für die Roboternavigation benötigter Funktionen wie 3D-Visualisierung, Selbstlokalisierung, Objekterkennung und Kollisionsvermeidung.



**Abbildung 1.1:** Grafische Darstellung der universalen modularen Architektur, die in dieser Arbeit die Realisierung der Tiefenrekonstruktion und darauf basierende Selbstlokalisierung humanoider Roboter spezifiziert.

Im Mittelpunkt der vorliegenden Arbeit steht die 3D-Rekonstruktion und darauf basierende Selbstlokalisierung humanoider Roboter. Somit werden zuerst alle nötigen Grundlagen und Methoden theoretisch behandelt und anschließend, orientiert an der oben beschriebenen Architektur, realisiert. Um das Ziel dieser Arbeit zu erreichen und die Funktion der entworfenen Architektur zu verifizieren, wird die Selbstlokalisierung anhand von Landmarken basierend auf der Tiefenrekonstruktion implementiert.

Um die entstehende Fehler der Tiefenschätzung zu minimieren und der Prozess der Selbstlokalisierung zu optimieren wird beschlossen die Einsatzumgebung mit passiven, künstlichen und eindeutigen Landmarken zu präparieren. Da jeder im Bild gefundenen Landmarke eine eindeutige Position im Raum zugeordnet wird, können die Ergebnisse der Rekonstruktion verbessert und die Selbstlokalisierung unter Zuhilfenahme der Triangulation durchgeführt werden.

Damit sollen in der vorliegenden Arbeit theoretische Grundlagen und darauf basierende Algorithmen erarbeitet und implementiert werden, die die sichere Ausführung folgender Schritte ermöglichen. Zuerst wird ein dreidimensionaler Raum unter Verwendung der Daten des Stereokamerasystems rekonstruiert. In dem rekonstruierten Raum werden die künstlichen Landmarken detektiert und anschließend extrahiert. Durch das Stereokamerasystem und die bekannten Koordinaten der Landmarken kann die Position des Roboters unter Zuhilfenahme der Triangulation bestimmt werden.

## 1.2 Vergleichbare Arbeiten

Die vorliegende Arbeit befasst sich mit zwei großen Themengebieten. Damit ist es schwierig vergleichbare Arbeiten zu finden, die beide Themengebiete gleichzeitig behandeln. Um jedoch diese Arbeit mit Anderen vergleichen zu können, werden die beiden Themengebiete in diesem Abschnitt getrennt und die einzelnen Themengebiete mit verwandten Arbeiten referenziert.

Als erstes werden einige Arbeiten vorgestellt, die sich mit der statischen Stereoanalyse befassen. Im Anschluss darauf wird dann auf Arbeiten eingegangen, die sich mit visueller Erkennung von künstlichen Landmarken und der Selbstlokalisierung von Robotern auseinander setzen.

### 1.2.1 Statische Stereoanalyse

Die Gewinnung der räumlichen Tiefe am Beispiel des menschlichen Sehens beschäftigt seit Jahrhunderten interdisziplinäre Wissenschaftsbereiche (siehe Kapitel 1) und ist somit ausgiebig bearbeitet. Dementsprechend ist auch die Anzahl der Arbeiten, die sich mit diesem Thema auseinander setzen, recht groß. An dieser Stelle werden nur ein Paar dieser Arbeiten angegeben und auf deren behandelten Methoden eingegangen.

Die Arbeit [Som05] von Katharina Somieski befasst sich mit dem Aufbau eines Forschungsplatzes für die geometrische Stereoanalyse. Der Schwerpunkt dieser Arbeit

wird auf die Rekonstruktion der Tiefe aus Stereobildern gelegt. Dabei werden unterschiedliche Stereogeometrien sowie verschiedenen Korrespondenzanalyseverfahren untersucht. Bei der Korrespondenzanalyse werden vor allem die Verfahren *Block-Matching*, *Dynamische Programmierung* und *Graph-Cut* mit Energieminimierung gegenübergestellt. Die Korrespondenzanalyse wird auf rektifizierte Stereobildpaare angewendet, sodass die Rektifikation ein weiterer Bestand dieser Arbeit ist. Unter den Rektifikationsverfahren wird die Rektifikation mit Kalibrierung sowie ohne Kalibrierung des Stereosystems untersucht.

Um die Korrespondenzanalyseverfahren nach ihrer Güte ordnen zu können, werden die Disparitätskarten mit unterschiedlichen Parametern der Algorithmen berechnet. Die gelieferten Ergebnisse werden dann qualitativ und quantitativ verglichen. Zuletzt werden die Disparitätskarten dieser Verfahren dreidimensional visualisiert.

Eine weitere Arbeit [Bla04], die sich mit der Tiefeninformationsgewinnung aus Stereobildern befasst, ist von Roman Blaschek. Direkt nach der Einleitung der Arbeit, werden als erstes die verschiedenen Geometrien der Stereokamerasysteme behandelt. Dabei geht der Autor auf die Standardstereogeometrie und die konvergente Ausrichtung der Kameras ein. Über einen Exkurs zur Epipolargeometrie werden verschiedene Rektifikationsverfahren vorgestellt. Es wird ausgiebig die Rektifikation unter Kenntnis der internen und externen Kameraparameter, über Polarkoordinaten und mittels Homographien behandelt. Bei der Berechnung der Disparität wird auf die flächen- und merkmalsbasierte Korrespondenzanalyseverfahren eingegangen. Da die Implementierungen für das Bildverarbeitungsprogramm AdOculus bestimmt sind, werden auch die Experimente in diesem System durchgeführt.

### 1.2.2 Selbstlokalisierung anhand von Landmarken

Die Selbstlokalisierung von Robotern ist eins der wichtigen Kriterien für die Navigation in einer Umgebung. Zur Erleichterung dieser Aufgabe werden meistens Landmarken verwendet. Sind zu einem Zeitpunkt mehrere Landmarken erkannt worden, so kann anhand dessen die Position des Roboters in Weltkoordinaten berechnet werden. Dafür werden die Weltkoordinaten der Landmarken als bekannt vorausgesetzt.

Die Dissertation [Stü04] von Wolfgang Stürzl befasst sich mit der Verwendung von Landmarken für Navigationsaufgaben. Dabei wird speziell das Wiederfinden bekannter Orte anhand visueller Information untersucht, das als *visuelles Heimfinden* (engl. *visual homing*) oder *visuelle Zielanfahrt* bezeichnet wird. Der Begriff *Landmarke* wird in dieser Arbeit für extrahierte Merkmale der aktuell vorliegenden Sensorinformationen verwendet. Diese Landmarken werden mit einer gespeicherten Repräsentation des Zielorts verglichen und die Position des Roboters in Weltkoordinaten bestimmt.

Für die Durchführung der Experimente wird in dieser Dissertation ein kleiner Tischroboter eingesetzt, der mit einem Sensor für omnidirektionales Stereo ausgestattet



tet ist. Dieser Sensor ermöglicht Aufnahmen von Stereopanoramabildern, die im Anschluss zu Disparitätssignaturen umgerechnet werden. Die Disparitätssignaturen bilden eine unmittelbare Verknüpfung mit der Geometrie der Aufnahmeorte und können somit als Landmarken verwendet werden. Im Laufe der Dissertation wird ein Algorithmus zum Heimfinden anhand von Disparitätssignaturen vorgestellt. Dieser wird dann mit dem in [FSMB98] vorgestellten Algorithmus des bildbasierten Heimfindens erweitert. Durch die Kombination der Bild- und Disparitätssignaturen werden die Vorteile beider Repräsentationen verwendet und führen somit zu besseren Ergebnissen.

Da in dieser Dissertation Bild- und Disparitätssignaturen als Landmarken verwendet werden, müssen diese für die Navigation in einer großen Umgebung gespeichert und verglichen werden. Um das Problem der großen Datenmenge zu umgehen, wird ein auf die Fouriertransformation basierender Algorithmus zur Datenreduktion untersucht. Die Methode die Datenmenge mit PCA (*Principal Component Analysis*) zu reduzieren erwies sich in dieser Arbeit wegen omnidirektionaler Bilder eher ungeeignet, da die Koeffizienten des PCA von der Aufnahmerichtung abhängen und daher den Bildvergleich erschweren würden.

Der Vorteil des auf Fouriertransformation basierenden Algorithmus ist, dass bei einer Rotation eines Bildes lediglich die Phasen der Fourierkoeffizienten variieren. Des Weiteren zeigt dieser eine lineare Komplexität zur Anzahl der Fourierkoeffizienten auf. Damit besitzt der Algorithmus zum Vergleich zu den herkömmlichen Methoden mit quadratischer Abhängigkeit von der Anzahl der Bildpixel einen geringeren Speicherbedarf und Rechenaufwand.

Zum Abschluss der Dissertation wird vom Autor ein aktives Stereokamerasystem vorgestellt, das basierend auf disparitätsselektiven Energieneuronen Vergenzbewegungen ausführt (siehe [Stü04]) und lokale Disparitäten bestimmt. In Bezug auf landmarkenbasierte Navigation wird die Zielsetzung verfolgt, die Aktivität der Neuronenpopulation als Ortssignatur zu verwenden und damit nach Implementierung des Systems auf einem mobilen Roboter eine Rückkehr zu bekannten Orten zu ermöglichen.

Die Dissertation [Web02] von Joachim Weber befasst sich mit globaler Selbstlokalisierung mobiler Service Roboter. Die Zielsetzung dieser Dissertation ist die Bereitstellung globaler echtzeitfähiger Lokalisationstechniken. Diese sollen vor allem die durch Dynamik der Umgebung erzeugte Fehler korrigieren, initiale Positionshypothesen generieren und mit vorhandenen Mehrdeutigkeiten der Umgebungsteile zurecht kommen. Zur Gewinnung des Maximums an Effektivität und Robustheit basiert die entwickelte Lokalisationsarchitektur explizit auf der Auswertung einer Vielzahl verschiedener Mustererkenner und Sensoren.

In [Web02, Kapitel 4] wird die Vereinigung von Informationen mehrerer Lokalisationstechniken und Aufstellung von Hypothesen zur aktuellen Position behandelt. Zu den Techniken gehören in erster Linie die Positionsbestimmung mit 2D-Laserscans

und bildbasierte Lokalisation. Unter Verwendung dieser Informationen werden dann uni- und multimodale probabilistische Methoden der Selbstlokalisierung untersucht.

Da die implementierten Algorithmen für das CAROL-Forschungsprojekt (*Camera Based Adaptive Robot Navigation and Learning*) vorgesehen sind, wurden diese in Rahmen des CAROL-Systems getestet.

### 1.3 Gliederung der Arbeit

Im ersten Kapitel wird die grundlegende Architektur vorgestellt, die auf dem Stereokamerasystem basiert und unterschiedliche Verwendung der Tiefenrekonstruktion ermöglicht. Die 3D-Rekonstruktion und darauf aufbauende Selbstlokalisierung werden als Bestandteile dieser Architektur implementiert.

Das zweite Kapitel befasst sich mit den Grundlagen der Bildaufnahme. Des Weiteren wird das Prinzip des Lochkamera-Modells sowie Projektion eines 3D-Punktes auf eine 2D-Ebene dargestellt und erläutert.

Die mathematische Darstellung der oben genannten Projektion, die Notwendigkeit der Kamerakalibrierung sowie die Bedeutung der Kameraparameter wird im Kapitel drei beschrieben. Im Weiteren werden Kalibrierungsmethoden kategorisiert und in kurzer Form dargestellt. Einige Verfahren werden ausführlich erläutert und in mathematischer Form präsentiert. Außerdem wird das System um eine weitere Kamera zu einem Stereokamerasystem ergänzt. Diese Option erlaubt Durchführung der Tiefenrekonstruktion, die ein der wichtigsten Teile dieser Arbeit bildet.

Im vierten Kapitel wird ein Einblick in die statische Stereoanalyse gegeben und ihr Ziel erläutert. Dabei bilden die Bildaufnahme und Kamerakalibrierung, die in den Kapiteln davor beschrieben sind, ein Bestandteil der statischen Stereoanalyse. Außerdem werden geometrische Zusammenhänge eines Stereosystems und die unterschiedliche Anordnung der Kameras untersucht. Danach wird auf die weitere zwei Verarbeitungsschritte der statischen Stereoanalyse eingegangen die Rektifikation und Tiefenrekonstruktion. Es werden unterschiedliche Methoden der Rektifikation und Tiefenrekonstruktion vorgestellt. Der Schwerpunkt der Tiefenrekonstruktion bildet die Korrespondenzanalyse, die sehr rechenintensiv ist. Deswegen werden im Folgenden die Korrespondenzanalyseverfahren betrachtet mit dem Ziel die Komplexität zu reduzieren.

Die Korrespondenzanalyse bildet das Hauptmerkmal des fünften Kapitels. Hier werden zuerst die Annahmen und Einschränkungen diskutiert. Danach werden die Korrespondenzanalyseverfahren klassifiziert und deren Ergebnisse beschrieben, die den Ausgangspunkt der Tiefenrekonstruktion darstellen.

Die Erörterung der Tiefenrekonstruktion und dafür benötigten Stereotriangulation finden im Kapitel sechs statt. Damit wird in Abhängigkeit vom gewählten Korrespondenzanalyseverfahren die Rekonstruktion beziehungsweise Teilrekonstruktion von Objekten oder Szenen ermöglicht.

Im Kapitel sieben werden unterschiedliche Sensortechniken und daraufbasierende Lokalisierungsverfahren theoretisch vorgestellt und behandelt. Die Wahl passiver künstlicher Landmarken sowie die Auswahl in dieser Arbeit eingesetzten Landmarken werden begründet. Dafür werden mehrere Landmarken und falls vorhanden ihre mathematische Modelle vorgestellt und miteinander verglichen.

Die zur Verfügung stehende und in der vorliegenden Arbeit eingesetzte Hard- und Software werden im Kapitel acht beschrieben.

Im Kapitel neun werden eingesetzte und implementierte Algorithmen und damit erreichte Resultate vorgestellt. Abschließend wird die vorliegende Arbeit zusammengefasst und mögliche Erweiterungen diskutiert.

## 1.4 Hinweise

In dieser Arbeit wird für einige Begriffe, die einerseits sich in der Informatik etabliert haben oder für die keine präzise Übersetzung gefunden werden konnte, die englische Nomenklatur verwendet. Außerdem wurde darauf geachtet eine einheitliche Notation während der gesamten Arbeit zu verwenden. Für die bessere Lesbarkeit wird die genutzte Nomenklatur an den entsprechenden Stellen und im bestehenden Kontext erläutert. Die grundlegenden mathematischen Kenntnisse werden als bekannt vorausgesetzt.

Des Weiteren werden einige Grundlagen aus den Bereichen 3D Vision und Navigation nicht referenziert, da sie zu dem festen Bestandteil der Standardliteratur gehören. Dieses Wissen wird in dieser Arbeit als bekannt vorausgesetzt und kann in den meisten Standardwerken nachgeschlagen werden. Diese Maßnahme soll der besseren Lesbarkeit dieser Arbeit dienen und ist keinesfalls anderes als angegeben zu verstehen. Alle diese Grundlagen kommen mehrmals in hier angegebenen Literaturquellen vor, so das ein Verweis auf den konkreten Autor oder eine Gruppe von Autoren sich nur schwer gestaltet.



# Grundlagen der Bildaufnahme

# 2

---

## 2.1 Kamera-Modell

Einer der primitivsten Modelle einer Kamera ist die so genannte *Camera obscura* (lateinisch *Camera* für „Kammer“ und *obscura* für „dunkel“). Diese kann als eine Art Kasten vorgestellt werden, der mit einem Loch, dem optischen Zentrum, versehen ist. Die dem Loch gegenüberliegende Seite ist halbtransparent. Das Licht, das durch das Loch einfällt, erzeugt auf der diametralen, halb-transparenten Seite ein skaliertes, an horizontaler und vertikaler Achse gespiegeltes Abbild, der sich vor der Kamera befindlichen Szene.

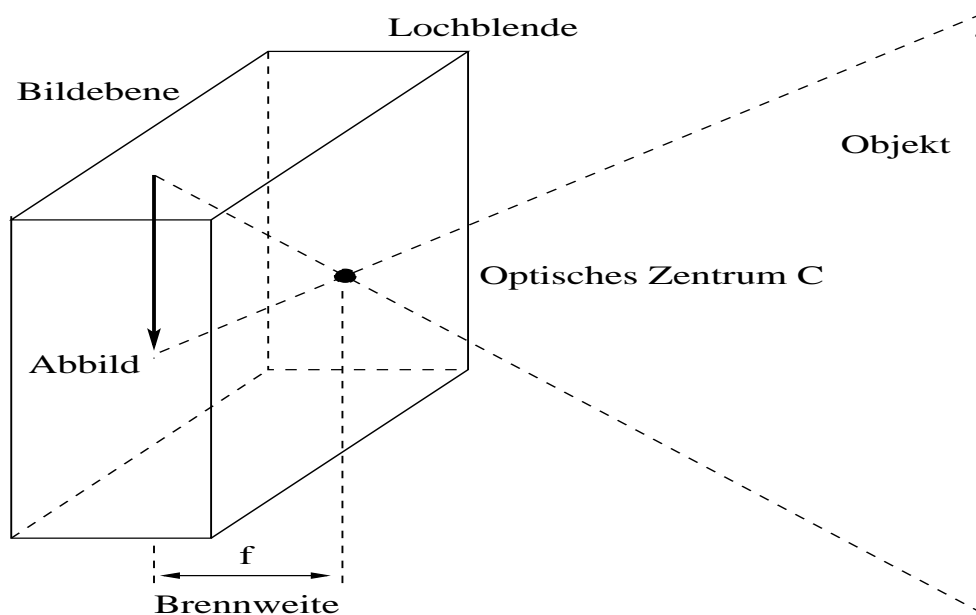
Das Prinzip der *Camera obscura* ist schon seit Jahrhunderten bekannt. Zum ersten Mal wird es vom Aristoteles (384 – 324 vor Chr.) in seinem Werk *Problemata physica* beschrieben. Um 980 wurden erste dokumentierte Experimente mit der *Camera obscura* von Abu Ali al-Hasan Ibn Al-Haitham (965 - 1040) durchgeführt. Viele Maler und Wissenschaftler benutzten bis in das 19. Jahrhundert die *Camera obscura* als Werkzeug für Ihre Arbeiten.

Da die *Camera obscura* nicht der Gegenstand dieser Arbeit ist, wird für weitere Informationen über Geschichte und Entwicklung der *Camera obscura* auf das Buch von Eric Renner [Ren04] verwiesen.

Die *Camera obscura* kann durch eine konvexe Linse erweitert werden. Fehlt diese, wird das Prinzip unter dem Begriff *Lochkamera-Modell* zusammengefasst. Das Lochkamera-Modell beschreibt die perspektivische Projektion eines dreidimensionalen Raumes, über das optische Zentrum, auf eine zweidimensionale Ebene [Fau95]. Auf dieses Modell wird in der vorliegenden Arbeit noch öfteren Bezug genommen.

Die Abbildung 2.1 stellt das Lochkamera-Modell anschaulich dar. Dabei wird durch die Zentralprojektion eine dreidimensionale Szene auf eine Zweidimensionale Ebene abgebildet.

Somit wird ein Punkt  $P_w$  im Raum mit den Koordinaten  $(x, y, z)$  auf ein Punkt  $P(fx/z, fy/z)$  in der Bildebene projiziert, wobei  $f$  die Brennweite der verwendeten Kamera darstellt. Wie schon oben erläutert, findet durch die Abbildung eine



**Abbildung 2.1:** Das Lochkamera-Modell. Das einfallende Licht erzeugt auf der gegenüberliegenden, halb-transparenten Seite ein skaliertes, spiegelverkehrtes und auf dem Kopf stehendes Abbild, der sich vor der Kamera befindlichen Szene. Mathematisch ausgedrückt wird eine dreidimensionale Szene auf eine zweidimensionale Ebene unter Zuhilfenahme der Zentralprojektion abgebildet.

Überführung von  $\mathbb{R}^3$  nach  $\mathbb{R}^2$  statt. Es ist ersichtlich, dass das beschriebene Modell nur in eine Richtung eindeutig ist, da bei der Projektion die Tiefeninformation verloren geht.

Das Lochkamera-Modell stellt nur ein abstraktes Modell einer realen Kamera dar und wird in der Wissenschaft zur Beschreibung der grundlegenden mathematischen Zusammenhänge einer realen Kamera verwendet.

Um Beugungsfehler und Schärfeverlust zu vermeiden, wird das Loch des Modells als unendlich klein angenommen, was nur in der theoretisch möglich ist. Um dieser Annahme nahe zu kommen, werden in der Praxis Objektive verwendet, die aus einer oder mehreren Linsen bestehen. Aber auch die Verwendung von Objektiven hat Nachteile, so werden zum Beispiel nur die Gegenstände scharf abgebildet, die sich in einem bestimmten Abstand zu der Kamera befinden. Außerdem verursacht jede Linse und die Geometrie des Sensor-Chips, durch die Ungleichmäßigkeiten und Fehler in der Struktur, eine Verzerrung. Die, unter anderem, in folgenden Abschnitten dieser Arbeit ausführlich behandelt wird.

## 2.2 Bildaufnahme

Das Lochkamera-Modell hat sich über Jahrhunderte als sehr nützlich erwiesen, auch moderne Digital-Kameras bauen auf demselben Prinzip auf. Das Modell wird durch ein Objektiv erweitert und anstatt halb-transparenter Rückseite wird eine lichtempfindliche Sensorfläche eingesetzt. Mit dem Einsatz der Sensorfläche wird ermöglicht, wie schon beim Einsatz eines Films in den analogen Kameras den größten Nachteil der *Camera obscura* auszugleichen, nämlich Bilder dauerhaft zu speichern.

Die Funktionsweise ist auch weitgehend ähnlich, das einfallende Licht wird durch ein Objektiv fokussiert und auf eine dahinterliegende Fläche projiziert. Dabei werden die Elektronen, die sich auf dem Sensor-Chip befinden, durch das Licht erregt, was zu einer Änderung der Ladung führt. In der Physik wird dieser Vorgang unter dem Begriff des lichtelektrischen Effektes [Tip94] zusammengefasst. Durch die Stärke der Ladung ist möglich auf die Helligkeit der Pixel zu schließen und damit die Bildinformation zu digitalisieren.

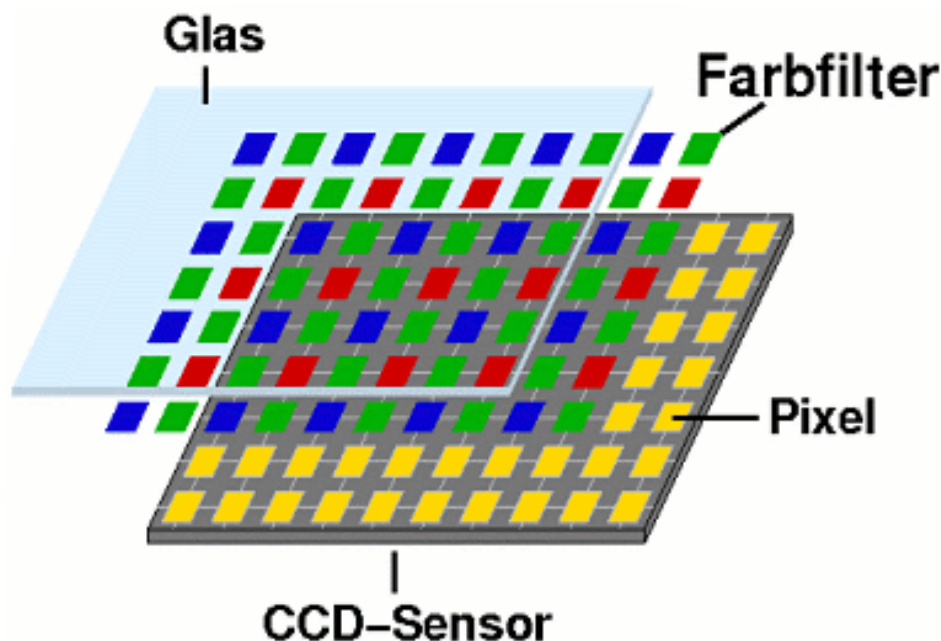
Der Bildsensor besteht größtenteils aus Fotodioden, die oft als Pixel bezeichnet werden [WW86]. Je nach Art der Verknüpfung der Fotodioden wird zwischen CCD- und CMOS-Chips unterschieden. Während ein CCD-Sensor (Charged Coupled Device) aus ladungsgekoppelten Bauelementen besteht, ist ein CMOS-Sensor (Complementary Metal Oxide Semiconductor) aus paarweise komplementär zueinander angeordneten Transistoren aufgebaut.

Da die meisten Kameras mit CCD-Sensoren ausgestattet sind, wird auf das Prinzip des CCD-Sensors näher eingegangen. Dennoch wird auch der CMOS-Sensor oberflächlich beschrieben, anschließend werden beide Techniken in Relation zueinander gesetzt.

Eine Fotodiode eines CCD-Sensors nutzt die materialimmanente Lichtempfindlichkeit des Siliziums. Eintreffende Photonen, aus denen Licht besteht, werden in elektrische Ladung umgewandelt und räumlich fixiert. Diese Ladung wird aus dem Sensor spaltenweise sequenziell in ein Schieberegister ausgelesen und anschließend an den Ladungsverstärker weitergeleitet. Dieser Vorgang wird in der Abbildung 2.3 grafisch verdeutlicht.

Beim Transport kann der so genannte Blooming-Effekt auftreten, einer der größten Nachteile des CCD-Sensors. Der Blooming-Effekt tritt immer dann auf, wenn der CCD-Sensor einer so großen Lichtintensität ausgesetzt wird, dass die Speicherkapazität des Ladungspools überschritten wird. Die Ladung kann sich auf benachbarte Fotodioden ausbreiten, was eine allgemeine Überbelichtung des Bildes zur Folge hat [Goe02]. Um eine Überbelichtung des Bildes zu vermeiden, wird in den meisten Kameras ein mechanischer Verschluss (engl. *Shutter*) verbaut.

Da die Eigenschaften der einzelnen Fotodioden weitgehend identisch sind, entsteht



**Abbildung 2.2:** Schematischer Aufbau eines CCD-Sensor [Bau07], deutlich zu sehen sind die Farbfilter, die das Extrahieren der Farbinformationen ermöglichen. Um die Farbinformationen aus dem CCD-Chip auszulesen, werden die Zellen abwechselnd mit verschiedenen Farbfiltern versehen. So resultiert die Farbe eines Punktes aus Zusammenfassung mehrerer für verschiedene Farben zuständige Zellen, was an das menschliche Sehen angelehnt ist.

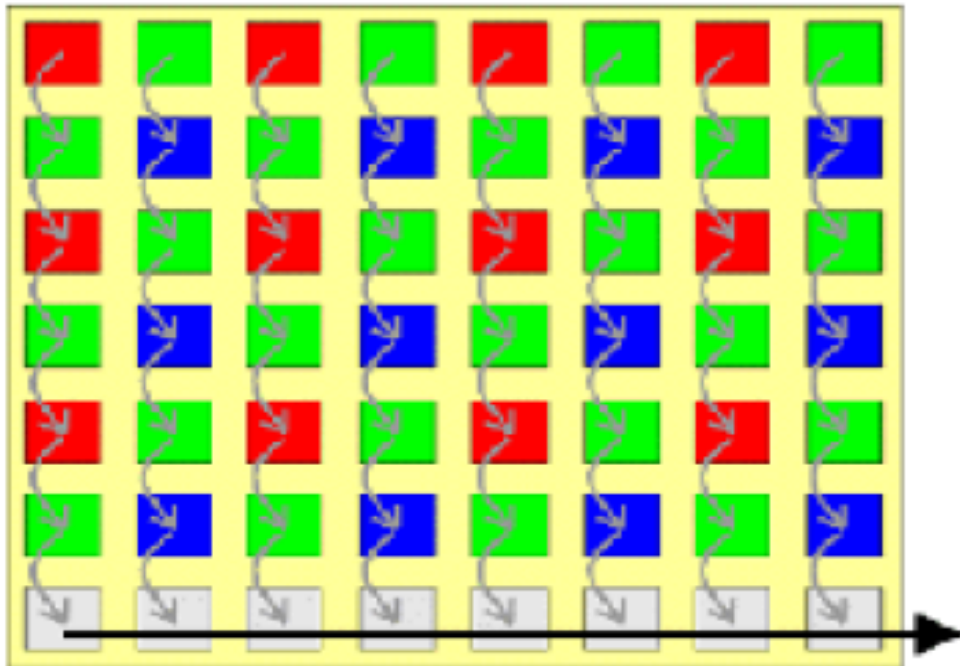
nur ein geringes ortsfestes Rauschen. In der nachfolgenden Abbildung 2.2 von Elmar Baumann [Bau07] wird der Aufbau eines CCD-Sensors schematisch dargestellt.

Um nicht nur die Helligkeitsinformation, sondern auch Farbinformationen aus dem CCD-Chips auslesen zu können, werden die Zellen abwechselnd mit verschiedenen Farbfiltern versehen. So resultiert ein Farbpunkt aus mehreren für verschiedene Farben zuständige Zellen. Diese Vorgehensweise ist mit der rezeptiven Aufnahme der Farbinformation im Auge des Menschen vergleichbar. In der Praxis kommt oft der Bayer-Sensor zum Einsatz, der auch in dem Bild 2.2 zu sehen ist. Durch den Gewinn der Farbinformation entstehen aber auch Nachteile, wie zum Beispiel die durch den Farbfilter verringerte Auflösung des Bildes.

Bei den CMOS-Sensoren, wie der Name schon sagt, sind die Transistoren paarweise komplementär zueinander ausgerichtet. Sodass eine bestimmte Spannung einen der beiden Transistoren leiten lässt und den anderen sperrt. Die Pixel sind einzeln verdrahtet, damit kann jeder Pixel adressiert und ausgelesen werden.

Die einzelne Verbindung der Zellen erlaubt es, weitere Transistoren in der Nähe der einzelnen Fotosensoren zu platzieren. Damit ähnelt der Aufbau eines CMOS-





**Abbildung 2.3:** Auslesung der Daten eines CCD-Sensors. Die eintreffende Lichtphotonen werden in elektrische Ladung umgewandelt und räumlich fixiert. Beim Auslesen werden einzelne Ladungen spaltenweise in ein Schieberegister und dann an einen Ladungsverstärker transportiert und anschließend ausgewertet.

Sensors einem RAM-Chip. Der Aufbau des Sensors in CMOS-Technik bietet die Möglichkeit zur Realisierung von mehreren Verarbeitungsschritten, wie die Analog-Digital-Wandlung, Weißabgleich und erste Bildverarbeitungsschritte direkt auf dem Chip. Durch die zusätzlichen Transistoren und die einzelne Anbindung der Sensorelemente wird jedoch eine Flächenvergrößerung des Sensors verursacht.

Der Gewinn der Farbinformationen, bei dem Einsatz des CMOS-Sensors, erfolgt durch Farbfilter, ähnlich dem CCD-Sensor.

Die wichtigsten Nachteile der CMOS-Sensoren gegenüber den CCD-Sensoren sind, die gesamte Größe des Bausteins, durch lokale Verstärkung entstehendes stärkeres Rauschen und die durch die eigene Kapazität verursachte Verzögerung des Signals. Jedoch sind auch die Vorteile der CMOS-Sensoren enorm, es kann kein Blooming-Effekt entstehen, da die Zellen einzeln ausgelesen werden. Zusätzlich benötigen Sensoren weniger Strom und können schneller ausgelesen werden [Goe02].

Es ist unmöglich zu sagen, welcher der beiden Sensoren besser ist. Die Auswahl für den Einsatz einer der beiden Sensoren hängt eher von der Aufgabe ab. Dennoch zeigt sich in den letzten Jahren die Tendenz, da die Herstellungskosten der CMOS-Chips auf das Niveau der Kosten der CCD-Chips reduziert werden konnten, dass die CMOS-Sensoren die CCD-Sensoren in der Zukunft vom Markt verdrängen werden.

Durch die nicht perfekte Lichtbrechung des Objektivs und der Lage des Sensors entstehen Aufnahmefehler, die die Verwendung des Lochkamera-Modells beeinträchtigen. Um das Prinzip jedoch einsetzen zu können, muss die Kamera in das Lochkamera-Modell überführt werden. Dafür sollen die Kameraparameter bestimmt werden. In dem nächsten Kapitel werden zuerst die Kameraparameter und deren Bedeutung beschrieben. Bei der Kamerakalibrierung wird die Bestimmung der Parameter dargestellt, sodass die eigentliche Überführung vollzogen werden kann.

### 2.3 Zusammenfassung

In diesem Kapitel wurden die Grundlagen der Bildaufnahme dargestellt. Dabei wurde das Prinzip des Lochkamera-Modells beschrieben und auf digitale Kameras übertragen. Durch die Kamera verursachte Abbildung einer 3D-Szenen auf eine 2D-Bildebene wurde erklärt, sodass in folgenden Kapiteln der Projektionsvorgang mathematisch dargestellt werden kann. Zusammen mit der Erläuterung der Funktion des Kamerasensors entstand eine Wissensbasis, sodass die gesamten späteren Transformationen leichter nachvollzogen werden können.

Der Prozess der Kamerakalibrierung ermöglicht die Beziehung zwischen den Weltkoordinaten und Bildkoordinaten der Kamera mathematisch zu beschreiben [KKS96]. In der Abbildung 3.1 werden die Beziehung zwischen den Koordinatensystemen der Welt, der Kamera und auf der Sensorfläche entstehendes Bildes grafisch verdeutlicht. Dabei wird ein Objekt im Weltkoordinatensystem über das optische Zentrum  $C$  in das Kamerakoordinatensystem transformiert und auf die Bildfläche projiziert.

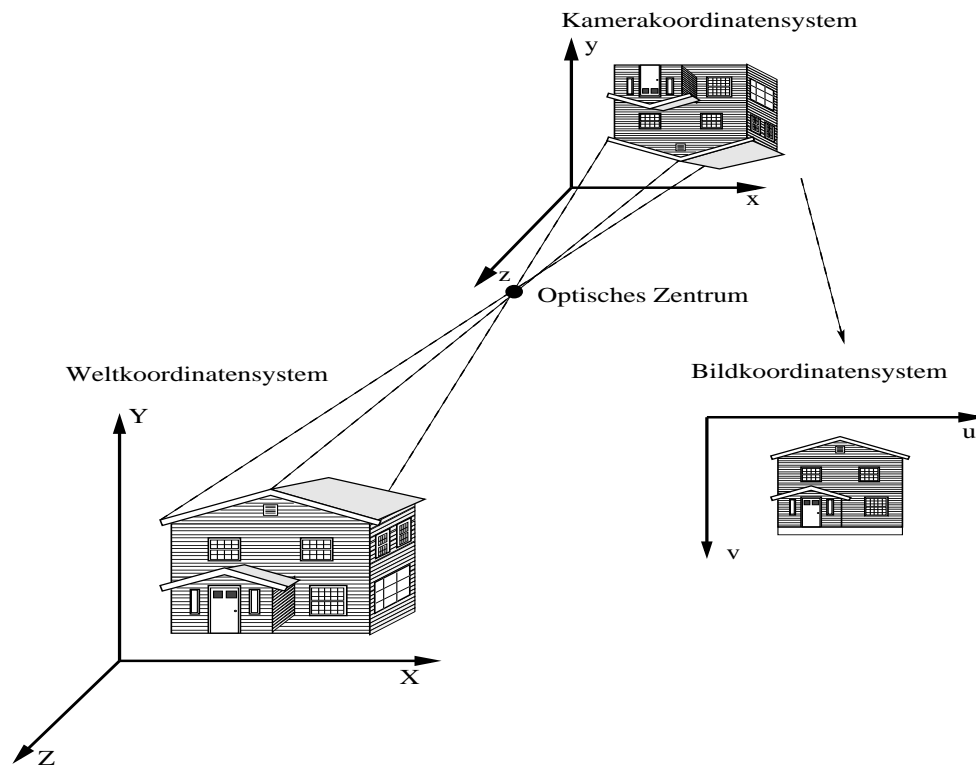
Durch den lichtelektrischen Effekt, der in dem Abschnitt 2.2 beschrieben ist, kann ein zweidimensionales Bild digitalisiert werden. Um von dem gewonnenen Bild auf die Weltkoordinaten zurückschließen zu können, werden kameraspezifische Parameter benötigt. Sind diese bekannt, so ist es möglich die tatsächliche Größe des abgebildeten Objektes sowie die durch den Abbildungsprozess verlorene Tiefeninformation zu rekonstruieren.

Durch die Kalibrierung werden Kameraparameter bestimmt, die aus der extrinsischen Parameter der Kamera, Rotation und Translation, sowie der intrinsischen, Brennweite, Bildhauptpunkt und anderen Parameter für die Linsenverzeichnungen bestehen.

Zuerst wird in diesem Kapitel auf die extrinsischen und intrinsischen Parameter einer Kamera eingegangen. Danach werden unterschiedliche Verfahren der Kamerakalibrierung vorgestellt.

Es existieren mehrere Kamerakalibrierungsverfahren. In diesem Abschnitt werden einige davon vorgestellt und ausführlich behandelt. Da wie in vorangegangenen Kapiteln erläutert bei der Abbildung einer dreidimensionalen Szene auf eine zweidimensionale Ebene die Tiefeninformation verloren geht, wird nachfolgend beschrieben, wie die Tiefeninformation zurückgewonnen werden kann. Um diese Information aus den Bildern zu extrahieren, werden mindestens zwei Bilder einer und derselben Szene aus unterschiedlichen Perspektiven aufgenommen. Einige Experimente dieser Arbeit basieren auf einem Stereokamerasystem, deswegen wird zusätzlich die Kalibrierung eines Stereosystems beschrieben. Sowie in dieser Arbeit eingesetzte Verfahren und Resultate.

Im nächsten Abschnitt wird auf die Bedeutung und Bestimmung der Kamerapara-



**Abbildung 3.1:** Beziehungen zwischen den Koordinatensystemen bei der Bildaufnahme. Über das optische Zentrum  $c$  wird ein Objekt in das Kamerakoordinatensystem transformiert und auf die Bildfläche projiziert. Das einfallende Licht verursacht auf der Sensorfläche einer Kamera Ladungsänderung und ermöglicht durch anschließende Quantisierung eine Digitalisierung der Bildinformation.

meter sowie in der vorliegenden Arbeit eingesetzte Methoden vorgestellt und theoretisch erläutert.

### 3.1 Bestimmung der Kameraparameter

In dieser Arbeit kam die Kamerakalibrierungsmethode nach Zhengyou Zhang zum Einsatz. Die Methode liefert die intrinsischen und extrinsischen Kameraparameter sowie die Koeffizienten der Linsenverzeichnung anhand mehrerer aufgenommenen Bilder eines Schachbrettmusters, mit den davor bekannten Dimensionen der eingesetzten Kalibrierungsvorlage.

Durch die Kamerakalibrierung bestimmen wir die Linsenverzerrung, so wie die restlichen intrinsischen Parameter, wie auch die extrinsischen Parameter einer Kamera

[Sch05b]. Dabei beschreiben die intrinsischen Parameter einer Kamera die Projektion der Punkte einer 3D-Welt in das lokale Kamerakoordinatensystem, das Abbild. Die extrinsischen Parameter geben Auskunft über die Lage der Kamera bezüglich eines globalen Koordinatensystems wieder.

Um die Prozedur der Kamerakalibrierung zu vereinfachen, wird die Kalibrierungsmethode in C++, unter Zuhilfenahme der Methoden aus OpenCV-Bibliothek implementiert, sodass die Kalibrierung nur wenig Zeit in Anspruch nimmt und jederzeit durchführbar ist. So kann sichergestellt werden, dass die Kameraparameter, die für spätere Berechnung benötigt werden, immer aktuell sind und damit die Berechnungen nur geringfügig verfälschten.

Um die Kalibrierungsergebnisse zu verifizieren, wird des Weiteren die Kamerakalibrierung mit der „Camera Calibration Toolbox“ für Matlab® durchgeführt. Auf die Ergebnisse der Verifikation wird in späteren Kapiteln noch ausführlich eingegangen. Auch die erst später mögliche indirekte Verifikation durch die Disparitätsbildung und daraus resultierende Entfernungsmessung werden in folgenden Kapiteln beschrieben.

## 3.2 Intrinsische Kameraparameter

Die intrinsischen Parameter beschreiben die interne Geometrie der Kamera, dazu gehören die Auflösung des verwendeten Sensor-Chips und Positionierung des Ursprungs des Koordinatensystems auf dem Chip, so wie die Koeffizienten der Lin senverzerrung [FFH<sup>+</sup>92], die in dieser Arbeit in einem gesonderten Unterkapitel behandelt werden. Da die intrinsischen Parameter die Zusammenhänge nur innerhalb der Kamera beschreiben, sind diese von der Lage und Ausrichtung der Kamera im Weltkoordinatensystem unabhängig.

Der Abbildungsprozess eines Weltpunktes auf die Bildebene kann vollständig über die perspektivische Projektion aus Gleichung (3.1) beschrieben werden.

$$P_c = A [R \ T] \cdot P_w \quad (3.1)$$

Die Rotation R und die Transformation T, die als extrinsische Kameraparameter zusammengefasst werden, sind in dem entsprechenden Kapitel ausführlich behandelt. Die Matrix A beinhaltet die intrinsischen Parameter der Kamera, die aus der Brennweite f und der  $u_0$ - und  $v_0$ -Koordinate des Hauptpunktes bestehen. Die  $fk_u$  und  $fk_v$  stellen dabei die horizontalen und vertikalen Skalierungsfaktoren dar. Der Parameter s, der so genannte *skew Parameter* charakterisiert die bei der Projektion entstehende Verzerrung. Der Aufbau der Matrix sieht dann wie folgt aus

$$A = \begin{bmatrix} fk_u & s & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

Bei der Betrachtung der Matrix  $A$  wird ersichtlich, dass bekannte intrinsische Parameter für die Wiederherstellung des Zusammenhanges zwischen dem Kamera- und Bildkoordinatensystem genutzt werden können. Diese Zusammenhänge bilden gemeinsam mit den extrinsischen Parameter und Koeffizienten der Verzerrung notwendige Basis für die später folgende Tiefenrekonstruktion, einer der wichtigsten Aufgaben dieser Arbeit.

Im nächsten Unterkapitel werden weitere Eigenschaften der Kamera theoretisch behandelt, die extrinsischen Parameter. Zuerst wird auf deren Bedeutung eingegangen. Anschließend wird die Bestimmung der Rotation und Transformation mathematisch beschrieben.

### 3.3 Extrinsische Parameter

Extrinsische Parameter beschreiben die Lage der Kamera bezüglich eines globalen Koordinatensystems [FFH<sup>+</sup>92]. Die Parameter bestehen aus der Rotation und Translation der Kamera. Somit lässt sich die Transformation vom Weltkoordinatensystem in das Kamerakoordinatensystem durch zwei Matrizen darstellen, einer 3x3 Rotationsmatrix und einem 3x1 Translationsvektor. Die mathematische Abbildung eines 3D-Punktes des Raumes auf einen 2D-Punkt des Kamerakoordinatensystems wird mit der Gleichung

$$P_c = P_w \cdot R + T \quad (3.3)$$

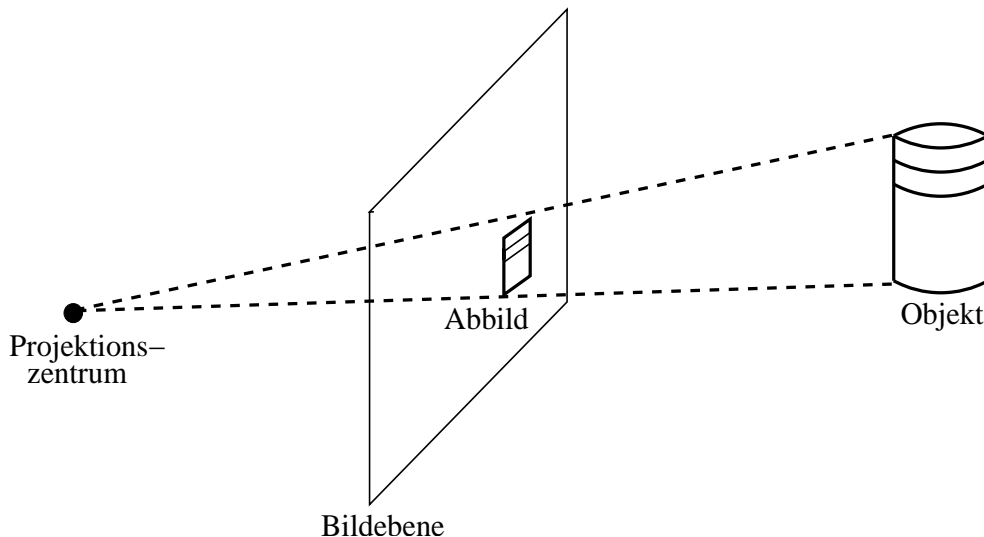
dargestellt, wobei  $P_c$  und  $P_w$  die Koordinate eines Punktes im Kamerakoordinatensystem, beziehungsweise Weltkoordinatensystem beschreiben und  $R$  und  $T$  jeweils Rotationsmatrix beziehungsweise Translationsvektor darstellen. Die Rotation und die Translation können in einer Matrix zusammengefasst werden, siehe Gleichung (3.4).

$$P_c = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} P_w \quad (3.4)$$

Dabei werden beide Matrizen um jeweils eine Dimension erweitert, sodass ein 3D-Punkt  $P(x, y, z)$ , der auch als Vektor zwischen dem Ursprung und Punkt  $P$  aufgefasst

werden kann, zu einem 4D-Punkt  $P'$  ( $kx, ky, kz, k$ ) wird. Die Komponente  $k$  stellt einen beliebigen von 0 unterschiedlichen Skalar dar.

In der Mathematik wird das oben beschriebene Verfahren als perspektivische Projektion bezeichnet. Dabei werden alle Punkte eines 3D-Modells entlang einer Linie über das optische Zentrum auf eine Fläche der Bildebene projiziert. Dadurch entsteht ein zweidimensionales Abbild eines dreidimensionalen Modells.



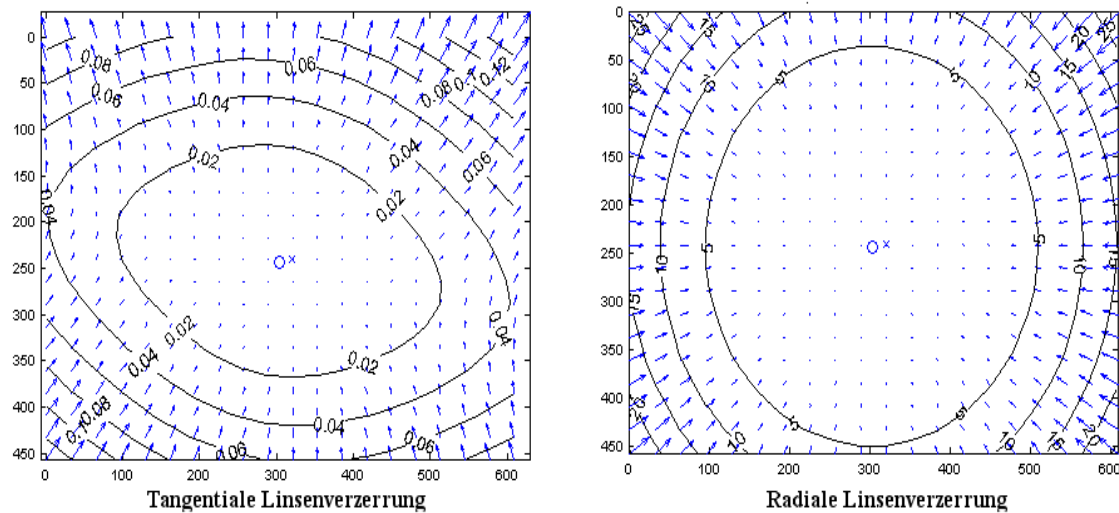
**Abbildung 3.2:** Die Darstellung der perspektivischen Projektion. Punkte eines 3D-Objekts werden auf eine zweidimensionale Bildebene projiziert. Einfachheit halber wird häufig, wie auch hier, die Projektionsebene vor dem Projektionszentrum platziert.

Aus der perspektivischen Projektion lassen sich zwei grundlegenden Eigenschaften ableiten. Zum einen, um zu den ursprünglichen euklidischen Koordinaten eines Punktes zurückzukehren, reicht es die ersten  $n$ -Elemente eines Vektors durch das  $n + 1^{\text{te}}$ -Element zu dividieren. Zum anderen werden zwei Punkte eines dreidimensionalen Modells genau dann auf einen zweidimensionalen Punkt projizieren, wenn sie durch die Skalierung ineinander überführt werden können.

### 3.4 Verzerrungen

Wie schon im vorangegangenen Abschnitt erwähnt, gibt es keine ideale Linse oder ideale Sensoren. Somit entstehen Verzerrungen in der Abbildung und Abweichungen zum Lochkamera-Modell. Diese so genannten nicht linearen Effekte sollen unter Zuhilfenahme inverser Transformation behoben werden. Dafür werden bei der Kamerakalibrierung die Eigenschaften der Verzerrung bestimmt. Die Faktoren der Ver-

zerrung gehören zu intrinsischen Kameraparametern, die später noch ausführlich beschrieben werden. Es wird zwischen zwei Arten der Linsen- beziehungsweise Sensor-Verzeichnung unterschieden, die tangentielle und die radiale Verzerrung.



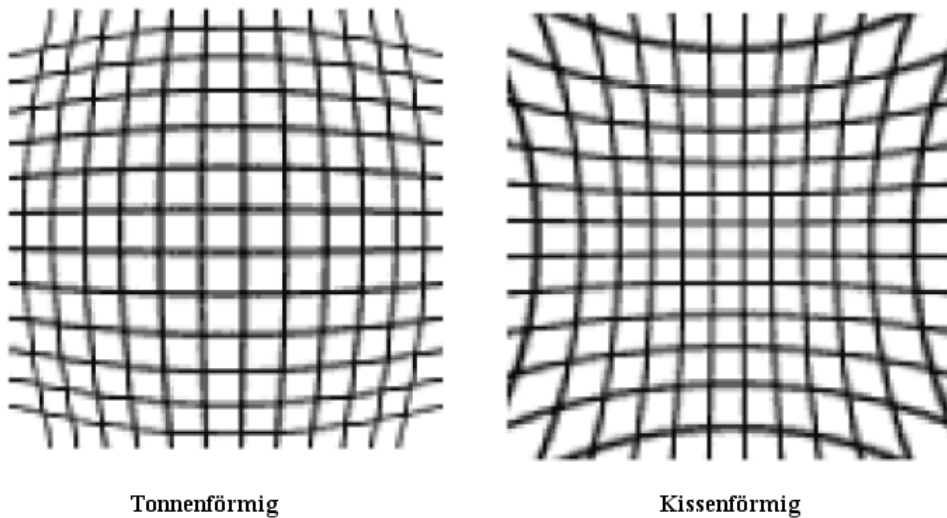
**Abbildung 3.3:** Beispiele für tangentielle und radiale Linsenverzerrung. Wie aus den Bildern ersichtlich, wird der Punkt bei der tangentialen Verzerrung entlang der Tangente durch den Mittelpunkt verschoben. Bei der radialen Linsenverzerrung wird der Abstand des Punktes zum Mittelpunkt verändert.

In der Abbildung 3.3 wird die tangentielle und radiale Linsenverzerrung grafisch dargestellt. Da die tangentielle Linsenverzerrung das Bild nur sehr geringfügig verfälscht, wird sie bei der hier verwendeten Kamerakalibrierungsmethode nach Zhengyou Zhang [Zha00], wie auch nach der Methode von Roger Tsai [Tsa86] nicht berücksichtigt.

Bei der radialen Linsenverzerrung wird wiederum zwischen zwei Arten unterschieden, der kissenförmigen und tonnenförmigen Verzerrung. Die Abbildung 3.4 stellt die beiden Arten grafisch dar.

Um das Modell der Lochkamera verwenden zu können, soll die radiale Verzerrung aus den Bildern herausgerechnet werden. Dafür werden mit bei der Kamerakalibrierung gewonnenen Daten die Bilder entzerrt und zur Gewinnung der intrinsischen und extrinsischen Kameraparameter weiterverwendet. Im Kapitel 9, bei der Beschreibung der in dieser Arbeit angewandten Kamerakalibrierungsmethode, wird auf die radiale Entzerrung noch näher eingegangen.





**Abbildung 3.4:** Darstellung der tonnenförmigen und kissenförmigen Verzerrungen, die aus der radialen und tangentialen Verzerrung resultieren.

### 3.5 Kamerakalibrierungsverfahren

Nach Luhmann existieren drei grundlegenden Verfahren zur Kamerakalibrierung, *Laborkalibrierung*, *Simultankalibrierung* und *Testfeldkalibrierung*. Diese werden anhand von Ort, Zeit oder eingesetzten Referenzkörper unterschieden. In der vorliegenden Arbeit werden nur die gängigsten Methoden behandelt, da die vollständige Schilderung einzelner Kalibrierungsverfahren [Luh00] den Rahmen dieser sprengen würde.

#### 1. Laborkalibrierung

Bei der Laborkalibrierung wird mithilfe eines Goniometers und eines hochpräzisen Gitters die innere Orientierung der Kamera bestimmt. Das Verfahren liefert sehr genaue Ergebnisse, ist aber nur für Messkameras sinnvoll und soll, wie die anderen Verfahren auch, in regelmäßigen Abständen wiederholt werden. Zusätzlich kann die Laborkalibrierung meistens nicht vom Anwender selbst durchgeführt werden.

#### 2. Simultankalibrierung

Bei der Simultankalibrierung wird das zu vermessende Objekt als Kalibrierungskörper verwendet. Die Kalibrierung findet meistens kurz vor dem geplanten Experiment statt. Dadurch liefert das Verfahren bessere Ergebnisse als die Testfeldkalibrierung, da die Parameter exakt zur Objektaufnahme bestimmt werden. Das Verfahren setzt aber voraus, dass die genaue Geometrie für jeden

Kalibrierungskörper neu bestimmt werden soll. Falls die Simultankalibrierung nur mit fotogrammetrischen Beobachtungen durchgeführt wird, wird dieses Verfahren als Selbstkalibrierung bezeichnet.

#### 3. Testfeldkalibrierung

Die Testfeldkalibrierung stellt, dass am weitesten verbreitete Verfahren zur Kamerakalibrierung dar. Dabei wird für die Kalibrierungszwecke immer derselbe Kalibrierungskörper verwendet, entsprechend werden dessen Abmessungen nur einmal bestimmt. Durch Erschütterungen oder Wechsel der Zoomeinstellungen kann sich die Geometrie der Kamera ändern. Um die Genauigkeit der Kamera konstant zu halten, sollte diese in regelmäßigen Abständen kalibriert werden. Der größte Vorteil der Testfeldkalibrierung resultiert aus den Tatsachen, dass das Verfahren einfach ist und permanent der gleiche Kalibrierungskörper eingesetzt wird. Dadurch nimmt die Kalibrierung nur wenig Zeit in Anspruch und kann auch von beliebigen Anwendern selbstständig durchgeführt werden.

Zhengyou Zhang unterscheidet seinerseits nur zwei Kalibrierungsmethoden, nämlich Kalibrierung mit und ohne Kalibrierungskörper. Diese Unterscheidung ähnelt der Klassifikation nach Luhmann, außer dass die Laborkalibrierung wegen den oben aufgeführten Nachteilen nicht berücksichtigt wird [Zha00].

In dem Buch von Hornberg wird die Laborkalibrierung zwar als Kalibrierungsmethode beschrieben, dennoch werden die Nachteile dieser Methode explizit aufgeführt, was die praxisbezogene Relevanz in Frage stellt. Die Methode benötigt hochwertiges Equipment und kann in den meisten Fällen nicht vom Anwender selbstständig durchgeführt werden. Damit verursacht die Laborkalibrierung einen enormen zeitlichen und finanziellen Aufwand [Hor06].

Im Folgenden werden zwei Verfahren der Testfeldkalibrierung vorgestellt, die am häufigsten in der Literatur referenziert und in der Praxis angewendet werden. Zuerst wird das Verfahren nach Roger Y. Tsai [Tsa86] in knapper Form erläutert und mathematisch beschrieben, anschließend wird etwas ausführlicher auf die Kamerakalibrierungsmethode nach Zhengyou Zhang [Zha00] eingegangen.

##### 3.5.1 Nach Roger Tsai

Mitte der achtziger Jahre stellte Roger Y. Tsai sein Verfahren der Kamerakalibrierung zum ersten Mal vor. Das leicht verbesserte und modifizierte Verfahren nach Tsai [Tsa86] ist auch nach über zwanzig Jahren sehr beliebt und findet häufig Anwendung in der Praxis.

Das Verfahren benötigt mindestens sieben korrespondierende Paare von Punkten, die nicht in einer Ebene liegen und deren Position vor der Kalibrierung als bekannt vorausgesetzt wird. Einer der größten Nachteile dieses Verfahren ist die Positionierung der korrespondierenden Punktepaare, die zwar beliebig aber dennoch vor der Kalibrierung bekannt sein muss.

Das Kamerakalibrierungsverfahren nach Tsai liefert, im Gegensatz zu der Kalibrierung durch die direkte lineare Transformation (DLT-Verfahren), zwei Koeffizienten der radialen Linsenverzeichnung [KKS96]. Die tangentielle Linsenverzerrung wird, wie auch im später beschriebenen Verfahren nach Zhang, nicht berücksichtigt.

Roger Tsai unterteilt sein Verfahren in 4 Schritte [Tsa87], die im Folgenden schematisch dargestellt werden. Im ersten Schritt wird auf die bereits bekannte Projektionsgleichung zurückgegriffen

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (3.5)$$

wobei  $R$  die  $3 \times 3$  Rotationsmatrix und  $T$   $1 \times 3$  Translationsvektor darstellen. Als Resultat liefert die Gleichung (3.5) ein Vektor  $(x, y, z)^T$ , Koordinaten eines 3D-Punktes im Kamerakoordinatensystem. Damit wird der Übergang von Weltkoordinatensystem in das Kamerakoordinatensystem mathematisch beschrieben.

Im zweiten Schritt stellt Roger Tsai die Gleichungen für das Lochkamera-Modell mithilfe der perspektivischen Projektion auf. Die Gleichungen beschreiben die Berechnung der idealen, nicht verzerrten Bildkoordinaten  $(X_u, Y_u)$ , also das Ergebnis einer Transformation von Kamerakoordinatensystem in Koordinatensystem des Bildes.

$$X_u = f \cdot \frac{x}{z} \qquad Y_u = f \cdot \frac{y}{z} \quad (3.6)$$

Dabei ist  $f$  die zu kalibrierende, effektive Brennweite der Kamera.

Der dritte Schritt behandelt die Linsenverzeichnung, dafür werden von Tsai folgende Gleichungen betrachtet

$$X_d + D_x = X_u \qquad Y_d + D_y = Y_u \quad (3.7)$$

Dabei sind  $X_d$  und  $Y_d$  verzerrte Koordinaten des Bildes in der Bildebene. Falls ein ideales System ohne jeglichen Verzerrungen betrachtet wird, stellen die beiden Koordinaten, die genauen Koordinaten des Bildes dar.  $D_x$  und  $D_y$  sind durch folgende mathematische Gleichungen

$$D_x = X_d \cdot (k_1 r^2 + k_2 r^2 + \dots) \quad D_y = Y_d \cdot (k_1 r^2 + k_2 r^2 + \dots) \quad (3.8)$$

definiert. Der Parameter  $r$  ist als  $r = \sqrt{X_d^2 + Y_d^2}$  gegeben.

Mit dem dritten Schritt werden die Koeffizienten der Linsenverzeichnung geschätzt, damit wird es möglich eine beliebige Kamera in das Lochkamera-Modell zu überführen. Nach Untersuchungen von Tsai reicht es für die meisten industriellen Anwendungen nur die radialen Verzerrungen und ein Koeffizient der Linsenverzeichnung zu betrachten. Weitere Modellbildungen bringen nicht nur keine besseren Erkenntnisse, sondern verursachen numerische Instabilität [Tsa87].

Im vierten Schritt der Kamerakalibrierung nach Roger Tsai wird die Transformation aus dem metrischen Bildkoordinatensystem in die Pixelkoordinaten des Bildes betrachtet.

$$X_f = s_x d_x'^{-1} X_d + C_x \quad Y_f = d_y^{-1} Y_d + C_y \quad (3.9)$$

wobei  $(X_f, Y_f)$  die Zeilen- und Spaltennummer der Pixelkoordinaten des gespeicherten Bildes und  $(C_x, C_y)$  die Koordinaten des Bildhauptpunktes darstellen. Des Weiteren ist  $d_x' = d_x \cdot \frac{N_{cx}}{N_{fx}}$  mit  $N_{cx}$  als Anzahl der Sensorpixel in X-Richtung.  $N_{fx}$  ist die Anzahl der Pixel, die durch den Computer abgetastet worden sind. Die Parameter  $d_x$  und  $d_y$  stehen für den Abstand zwischen den Zentren der benachbarten Pixel in X- beziehungsweise Y-Richtung.

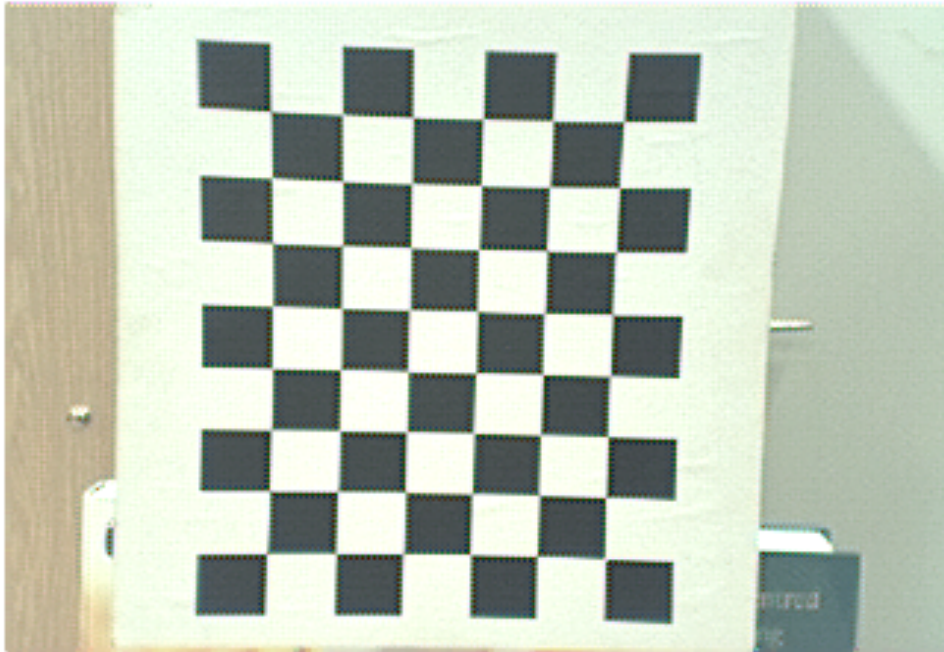
Somit wird die Kamerakalibrierung abgeschlossen, wobei zuerst die extrinsischen Parameter und in weiteren Schritten die intrinsischen Parameter geschätzt werden. Auch hier, wie auch im folgenden Kamerakalibrierungsverfahren nach Zhang, gilt, je größer die Anzahl der Punkte und genauer ihre Positionen, desto präziser funktioniert das Verfahren.

### 3.5.2 Nach Zhengyou Zhang

In dieser Arbeit wird das Kalibrierungsverfahren nach Zhengyou Zhang verwendet [Zha00], ein Testfeldkalibrierungsverfahren mit einem planaren Schachbrettmuster als Kalibrierungskörper, dass in diesem Abschnitt ausführlich behandelt wird.

Zhengyou Zhang stellte zum ersten mal 1998 ein Kamerakalibrierungsverfahren [Zha00] vor, das sich besonderes durch seine Einfachheit, Flexibilität und Stabilität auszeichnet.

Für die Kalibrierung kommt ein planarer Kalibrierungskörper mit einem Schachbrettmuster zum Einsatz. Dabei werden die Ecken einzelner Quadrate des Schachbrettmusters als Kalibrierungsmarken genutzt.



**Abbildung 3.5:** Der in dieser Arbeit eingesetzte Kalibrierungskörper. Ein planares Schachbrettmuster, wie es in dem Verfahren nach Zhengyou Zhang beschrieben ist.

Zur Kalibrierung werden mindestens zwei Bilder des Kalibrierungsobjekts benötigt, die aus unterschiedlichen Perspektiven aufgenommen werden. Die Orientierung des Körpers in den Bildern ist nicht festgelegt, den Kalibrierungsvorgang enorm erleichtert. Durch möglichst viele Aufnahmen des Kalibrierungskörpers aus unterschiedlichsten Perspektiven wird unmittelbar die Genauigkeit der extrinsischen und intrinsischen Parameter erhöht.

Die Ergebnisse der Kalibrierung, die innere und äußere Orientierung, werden in Pixel berechnet und zurückgegeben. Außerdem werden vier Koeffizienten der Linienverzerrung bestimmt, mit denen die entstehende Verzeichnung korrigiert werden kann.

Wegen den oben genannten Eigenschaften und der Tatsache, dass in der Fachliteratur das vollautomatische Verfahren als zuverlässig und die bestimmten Kamera- kalibrierungsparameter als präzise erachtet werden, kommt das Verfahren in der vorliegenden Arbeit zum Einsatz.

Das Verfahren nach Zhang [Zha00] kann in drei Schritte unterteilt werden, die nach-

folgend aufgezählt und erläutert werden.

### 1. Abschätzung der Homografie $H$

Es wird angenommen, dass das Kalibrierungsobjekt in der  $Z = 0$  Ebene liegt, was nicht der Allgemeingültigkeit des Verfahrens widerspricht. Wenn jetzt auf perspektivische Projektion, die in früheren Kapiteln vorgestellt wurde, zurückgegriffen wird, ergibt sich für einen Punkt der Bildebene  $P_c = [x, y, 1]^T$  und dem entsprechenden Punkt  $P_w = [X, Y, 0, 1]^T$  des Weltkoordinatensystems folgende Beziehung

$$\begin{aligned} s \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} &= A [R \ T] \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} \\ &= A[r_1 r_2 r_3 T] \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = A[r_1 r_2 T] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \end{aligned} \tag{3.10}$$

wobei die Parameter  $r_1$ ,  $r_2$  und  $r_3$  die entsprechenden Spaltenvektoren der Rotationsmatrix  $R$  darstellen. Da das Kalibrierungsobjekt in der Ebene  $Z = 0$  liegt, kann die Gleichung vereinfacht werden. Dafür werden bei der Berechnung die  $Z$ -Ebene und der Spaltenvektor  $r_3$  weggelassen. Das Produkt der Operanden  $A[r_1 \ r_2 \ T]$  kann als die Homografie  $H$  zusammengefasst werden.

$$H = A[r_1 r_2 T] \tag{3.11}$$

Damit ergibt sich für die Beziehung zwischen  $P_c$  und  $P_w$

$$sP_c = P_w H \tag{3.12}$$

Mit der Gleichung (3.12) wird für jedes Paar (mindestens vier) korrespondierender Punkte jeweils eine Gleichung aufgestellt. Diese werden, nach dem Algorithmus von Levenberg-Marquard [Mor77] vereinfacht und in geschlossener Form gelöst. Um die Homographie zu bestimmen, werden mindestens vier Punktepaare benötigt. Für die Berechnung der Matrizen wird auf ein Buch von Golub und van Loan [GL96] verwiesen.

Abschließend wird die Homographie  $H$  optimiert, da die Koordinatenpaare nicht exakt bestimmt werden können. Die Optimierung geschieht durch die

Minimierung des quadratischen Fehlers zwischen den gemessenen und berechneten Punkten.

## 2. Bestimmung der inneren und äußeren Orientierung aus H

Um die intrinsischen Parameter zu schätzen, greift Zhang [Zha00] auf die Gleichung (3.3) der Projektion eines Weltkoordinatenpunktes auf die Bildebene zurück. Matrix A beinhaltet die gesuchten intrinsischen Parameter.

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.13)$$

wobei  $u_0$  und  $v_0$  die Koordinaten des Hauptpunktes,  $\alpha$  und  $\beta$  die Skalierungsparameter in X- beziehungsweise in die Y-Richtung und  $\gamma$  den Versatz des CCD-Chips darstellen. Im nächsten Schritt konstruiert Zhang eine Matrix B, dabei stellt  $\lambda$  einen beliebigen Skalar da.

$$B = \lambda \cdot A^{-T} \cdot A^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (3.14)$$

Im Detail sieht die Matrix B dann folgendermaßen aus

$$B = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2\beta} & \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ -\frac{\gamma}{\alpha^2\beta} & \frac{\gamma^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} & -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix} \quad (3.15)$$

Damit können die intrinsischen Parameter einer Kamera aus der oberen Gleichung (3.15) folgend extrahiert werden.

$$v_0 = (B_{12} \cdot B_{13} - B_{11} \cdot B_{23}) / (B_{11} \cdot B_{22} - B_{12}^2) \quad (3.16)$$

$$\lambda = B_{33} - [B_{13}^2 + v_0(B_{12} \cdot B_{13} - B_{11} \cdot B_{23})] / B_{11} \quad (3.17)$$

$$\alpha = \sqrt{\lambda / B_{11}} \quad (3.18)$$

$$\beta = \sqrt{\lambda \cdot B_{11} / (B_{11} \cdot B_{22} - B_{12}^2)} \quad (3.19)$$

$$\gamma = -B_{12} \cdot \alpha^2 \beta / \lambda \quad (3.20)$$

$$u_0 = \gamma \cdot v_0 / \beta - B_{13} \cdot \alpha^2 / \lambda \quad (3.21)$$

Mit den oben aufgeführten Formeln kann die innere Orientierung der Kamera geschätzt werden. Falls die intrinsischen Parameter bekannt sind, können die extrinsischen Parameter folgend berechnet werden.

Aus der Gleichung der Homographie (3.11) kann anschließend folgende Gleichung (3.22) abgeleitet werden.

$$H = [h_1 h_2 h_3] = \lambda \cdot A \cdot [r_1 r_2 T] \quad (3.22)$$

Mit den Gleichungen (3.16 - 3.21) werden die intrinsischen Parameter einer Kamera geschätzt. Zusammen mit der Gleichung 3.22 ergibt sich folgende Möglichkeit die extrinsischen Parameter zu bestimmen.

$$r_1 = \lambda \cdot A^{-1} \cdot h_1 \quad (3.23)$$

$$r_2 = \lambda \cdot A^{-1} \cdot h_2 \quad (3.24)$$

$$r_3 = r_1 \times r_2 \quad (3.25)$$

$$T = \lambda \cdot A^{-1} \cdot h_3 \quad (3.26)$$

Somit werden intrinsische und extrinsische Parameter gewonnen, das Verfahren wird nun durch die Modellierung der Verzeichnung und anschließenden



Optimierung der Ergebnisse abgeschlossen.

### 3. Modellierung der Verzeichnung und Optimierung der Ergebnisse

Die Linsenverzerrung ist ein Teil der intrinsischen Parameter einer Kamera. In Folgenden wird die Bestimmung der Verzeichnung im Bild beschrieben.

Die in den Abschnitten 1 und 2 dieses Kapitels aufgeführte Formeln gelten nur für das Lochkamera-Modell. Um dem Lochkamera-Modell nahe zu kommen und die dazugehörigen Formeln einsetzen zu können, müssen die Bilder entzerrt werden. Erst die Entzerrung ermöglicht Durchführung weiterer Transformationen, wie die Rektifikation, auf die später noch eingegangen wird.

Wie schon in vorangegangenen Abschnitten dargestellt, beeinflusst die tangentielle Verzerrung das Bild nur unwesentlich. Daher konzentriert sich das Verfahren nach Z. Zhang [Zha00] nur auf die radiale Verzeichnung, die von den Eigenschaften der Kamera abhängen.

Nach dem Verfahren von Zhang werden nur die ersten beiden Faktoren der radialen Verzeichnung bestimmt, die Einflussgrößen höherer Ordnung werden vernachlässigt. Es wird angenommen, dass  $x_c$  und  $y_c$  die Kamerakoordinaten,  $x_w$  und  $y_w$  die korrespondierenden Weltkoordinaten nach dem Lochkamera-Modell darstellen. Genauso wie die  $u_c$  und  $v_c$  sowie  $u_w$  und  $v_w$ . Dann gilt

$$\begin{aligned}x_w &= x_c + x_c \cdot [k_1(x_c^2 + y_c^2) + k_2(x_c^2 + y_c^2)] \\y_w &= y_c + y_c \cdot [k_1(x_c^2 + y_c^2) + k_2(x_c^2 + y_c^2)]\end{aligned}\quad (3.27)$$

dabei sind  $k_1$  und  $k_2$  die Koeffizienten der radialen Verzerrung. Der Hauptpunkt stellt das Zentrum der radialen Verzeichnung da. Aus den folgenden Gleichungen

$$\begin{aligned}u_w &= u_0 + \lambda \cdot x_w + c \cdot y_w \\v_w &= v_0 + \beta \cdot y_w\end{aligned}\quad (3.28)$$

kann abgeleitet werden

$$\begin{aligned}u_w &= u_c + (u_c - u_0) \cdot [k_1(x_c^2 + y_c^2) + k_2(x_c^2 + y_c^2)^2] \\v_w &= v_c + (v_c - v_0) \cdot [k_1(x_c^2 + y_c^2) + k_2(x_c^2 + y_c^2)^2]\end{aligned}\quad (3.29)$$

Damit werden mindestens vier Paare korrespondierender Punkte benötigt um die radiale Verzerrung zu bestimmen.

Mit den berechneten Koeffizienten der radialen Verzerrung und allen anderen oben berechneten und geschätzten Koeffizienten sowie Projektionen ist es möglich, ein entzerrtes, normalisiertes Bild als Resultat zu erhalten.

Zum Schluss wird die resultierende Gesamtlösung durch die folgende Formel nach dem Verfahren von Levenberg-Marquard [Mor77] optimiert.

$$\sum_{j=1}^n \sum_{i=1}^m \|P_{w_{ij}} - P_c \cdot (A, k_1, k_2, R_j, T_j)\|^2 \quad (3.30)$$

wobei  $P_{w_{ij}}$  und  $P_c$  einen entsprechenden Weltkoordinatenpunkt und seine Projektion auf die Bildebene darstellen. Für die Faktoren  $A$ ,  $R$  und  $T$  werden die geschätzte Werte eingesetzt. Die Koeffizienten  $k_1$  und  $k_2$  werden zum Anfang der Optimierung auf null gesetzt.

In dieser Arbeit kam das Verfahren nach Zhang in zwei unterschiedlichen Implementierungen zum Einsatz. Zu einem, wie schon oben erwähnt, aus der Camera Calibration Toolbox für Matlab® und zum anderen aus einer OpenCV-Bibliothek.

Das Verfahren der Kamerakalibrierung nach Zhang wird oft in der Literatur als sehr stabil, flexibel und einfach in der Handhabung bezeichnet [Sta07], in späteren Kapiteln werden in dieser Arbeit gewonnene Ergebnisse präsentiert und ausgewertet.

#### 3.5.3 Kamerakalibrierung eines Stereosystems

Es existieren mehrere Möglichkeiten um aus einer zweidimensionalen Bildebene die Tiefeninformationen einer dreidimensionalen Szene wieder rekonstruieren zu können: die Originalgröße eines Objektes im Bild ist bekannt, eine und dieselbe Szene wird aus zwei oder mehr unterschiedlichen Blickwinkeln aufgenommen oder es werden zwei Kameras eingesetzt.

Da die beiden humanoiden Roboter, die in dieser Arbeit verwendet und in dem Kapitel 8 beschrieben werden, mit jeweils zwei Kameras ausgestattet sind, erschien die letztere der Möglichkeiten und damit der Einsatz eines Stereosystems als plausibel. Mit ihm wird eine und dieselbe Szene mit beiden Kameras zum gleichen Zeitpunkt aufgenommen, anschließend wird nach korrespondierenden Punkten in beiden Bildern gesucht. Ausgehend von dem Abstand der korrespondierenden Punkte, der als

Disparität bezeichnet wird, kann die Tiefe dieses Punktes im Raum mittels Triangulation rekonstruiert werden.

Dennoch bevor die Suche nach korrespondierenden Punkten durchgeführt werden kann, müssen die beiden Kameras kalibriert werden. Damit können die Epipolarlinien parallel ausgerichtet werden, die so genannte Rektifikation der Bilder, was die Suche auf eine zu durchsuchende Zeile beschränkt. Die Korrespondenzsuche, die Disparität, Epipolarlinien und die Rektifikation werden in dem Kapitel 4 ausführlich behandelt.

In der Literatur werden mehrere Wege ein Stereosystem zu kalibrieren beschrieben, hier werden die Grundlagen und Ziele der Kalibrierung eines Stereosystems skizziert. In später folgenden Kapiteln werden, in dieser Arbeit, eingesetzte Verfahren ausführlich beschrieben und erläutert.

Um das Stereosystem kalibrieren zu können, werden die beiden Kameras erstmal einzeln kalibriert, dabei sind nur die intrinsischen Parameter relevant. Mit den intrinsischen Parametern wird die Linsenverzeichnung aus den Bildern entfernt. Anschließend werden die neuen extrinsischen Parameter geschätzt, dabei wird nicht die Rotation und Translation bezogen auf den Koordinatenursprung bestimmt, sondern eine Kamera als Referenzpunkt betrachtet [GH01]. Bei der in dieser Arbeit verwendeten Kalibrierung wird die Rotation und Translation der rechten Kamera bezüglich der linken Kamera bestimmt.

Somit werden die Kameras bezüglich einander ausgerichtet, die Parameter der beiden Kameras können zu einem Stereokamera-Modell verbunden werden. Mit den geschätzten Parametern kann ein ideales Stereokamerasystem simuliert werden, in dem die Bildebenen koplanar und optischen Achsen parallel ausgerichtet sind. Damit liegen die beiden Epipole im Unendlichen, die Epipolarlinien sind in der gleichen Höhe, parallel der X-Achse und die zu suchende Korrespondenzpunkte in der gleichen Zeile der beiden Bilder.

## 3.6 Zusammenfassung

In diesem Kapitel wurde zuerst auf die Notwendigkeit und Bedeutung der Kamera- kalibrierung eingegangen. Das Modell der Abbildung einer dreidimensionalen Szene auf eine zweidimensionale Bildebene mathematisch dargestellt und erläutert. Danach wurden die intrinsischen und extrinsischen Parameter der Kameras behandelt. Nachdem der Zusammenhang zwischen den Welt-, Kamera- und Bildebenenkoordinaten relativiert war, wurden die Kamerakalibrierungsverfahren kategorisiert und in kurzer Form dargestellt. Einige der Kalibrierungsmethoden wurden ausführlich

behandelt und in mathematischer Form dargestellt. Anschließend wurde auf den Zweck der Kamerakalibrierung eines Stereosystems, nämlich die Parametergewinnung, die später zu Berechnung der Tiefeninformation aus 2D-Bildern eingesetzt werden können, eingegangen und die theoretische Bestimmung dieser wiedergegeben.

Damit wurden die Grundlagen geschaffen, um in dem nächsten Kapitel auf die Epipolargeometrie, Rektifikation und anschließende Bestimmung der Tiefeninformation ausführlich eingehen zu können.

# Statische Stereoanalyse

# 4

---

Die statische Stereoanalyse befasst sich mit der Untersuchung des geometrischen Zusammenhangs zwischen Stereobildpaaren. Es wird zunächst die geometrische Anordnung der Kameras untersucht und im nachhinein, je nach geometrischer Anordnung und der zugrunde liegender Kalibrierungsmethode die Tiefeninformation einer Szene bestimmt.

Bei der statischen Stereoanalyse wird davon ausgegangen, dass sich die Objekte einer Szene und die Kameras im Moment der Aufnahme des Stereobildpaares nicht bewegen oder bewegt werden. Falls mögliche Bewegungen im Moment der Aufnahme mitberechnet werden sollen, wird auf die dynamische Stereoanalyse zurückgegriffen, siehe [KKS96]. Ziel der statischen Stereoanalyse ist die Bestimmung der Tiefeninformation von Objekten einer Szene, die sich in mehrere Verarbeitungsschritte zerlegen lässt: **Bildaufnahme**. Der Bildaufnahmeprozess wird in erster Linie von den Eigenschaften der Lichtquellen, Abtastfrequenz und Auflösung des Kamerasensors beeinflusst. Eine detaillierte Beschreibung des Bildaufnahmeprozesses ist Kapitel 2 zu finden.

**Kameramodellierung / Kalibrierung**. Durch die Kalibrierung werden die kameraspezifischen Parameter und die Position des Stereosystems in Weltkoordinaten bestimmt. Die einzelnen Schritte zur Bestimmung der kameraspezifischen Parameter werden im Kapitel 3 ausführlich behandelt.

**Merkmalsextraktion**. In diesem Verarbeitungsschritt werden signifikante Merkmale eines Bildes extrahiert. Die Merkmalsextraktion wird dann angewandt, wenn zum Beispiel Kanten und ihre Orientierung in den Bildern bestimmt werden sollen. Die Beschreibung einiger Verfahren zur Extraktion von Merkmalen einer Szene sind im Unterkapitel 5.2.1 aufgeführt.

**Korrespondenzanalyse**. Mit der Korrespondenzanalyse werden in Relation stehende Elementenpaare der Stereobilder automatisch bestimmt. Es existieren eine Menge unterschiedlicher Verfahren zur Bestimmung korrespondierender Elemente. Der wesentliche Unterschied zwischen den Verfahren liegt in algorithmischer Vorgehensweise. Eine detaillierte Behandlung einiger Algorithmen der Korrespondenzanalyse erfolgt im Kapitel 5.

**Tiefenrekonstruktion**. In diesem Verarbeitungsschritt wird aus den Ergebnissen

des vorangegangenen Schrittes und der geometrischen Anordnung die Tiefeninformation einer Szene bestimmt. Die Tiefenrekonstruktion einer Szene ist in Kapitel 6 beschrieben.

**Interpolation.** Rekonstruktion der aufgenommenen Szene.

Die Gewinnung der Tiefeninformation einer Szene kann durch aktive oder passive Verfahren erfolgen. Aktive Verfahren werden im Kapitel 7.1 behandelt. Die Gewinnung der Tiefeninformation unter Zuhilfenahme eines Stereokamerasystems mit Anwendung der statischen Stereoanalyse gehört zu den passiven Verfahren. Die Tiefeninformation wird am Beispiel des natürlichen räumlichen Sehens von Menschen durch eine Stereoskopie bestimmt. Wie im Kapitel 2 beschrieben wird, geht bei der Abbildung einer Szene durch eine Kamera die Tiefeninformation verloren. Um die Tiefe einer Szene rekonstruieren zu können, wird redundante Information der Szene benötigt. Diese erhält man durch die Abbildung der gleichen Szene aus zwei oder mehr Perspektiven. Kann in jeder Abbildung die Position der Punkte bestimmt werden, so lässt sich die räumliche Tiefe aus bekannten Parameter der Kameras und deren Anordnung im Raum berechnen.

Anhand der geometrischen Ausrichtungen von Kameras können Stereokamerasysteme in unterschiedliche Klassen unterteilt werden. Bevor die verschiedenen Klassen von Stereokamerasystemen betrachtet werden, wird an dieser Stelle noch der Begriff *Disparität* eingeführt. Die Disparität ist ein wesentlicher Bestandteil zur Berechnung der räumlichen Tiefeninformation ausgehend von einem Stereokamerasystem und wird in den nächsten Kapiteln des öfteren vorkommen.

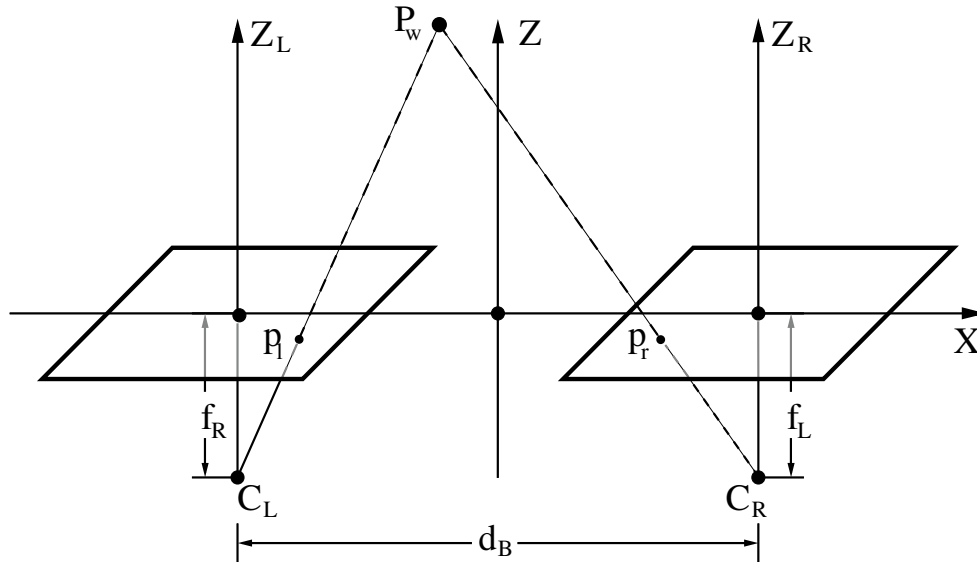
Das Wort *Disparität* setzt sich aus der lateinischen Vorsilbe *dis-* und dem Wort *par* zusammen. Die Vorsilbe *dis-* bedeutet “auseinander” und das Wort *par* bedeutet “gleich”. Somit bedeutet das Wort *Disparität* so viel wie “Ungleichheit” oder “Verschiedenheit”. Unter dem Begriff *Disparität* in der die Stereoskopie wird der Unterschied zwischen den Abbildungskordinaten ein und desselben Punktes der abgebildeten Szene verstanden.

In folgenden Unterkapiteln werden nun einige Geometrien von Stereokamerasystemen veranschaulicht.

### 4.1 Die Standardstereogeometrie

In der Standardstereogeometrie geht man davon aus, dass die Kameras in derselben Bildebene liegen und die gleichen Kameraparameter besitzen. Des Weiteren werden die optischen Achsen exakt parallel und senkrecht zur Basislinie (Linie über die beiden optischen Zentren  $C_L$  und  $C_R$  der Kameras) angenommen. Der Basisabstand zwischen den optischen Zentren  $C_L$  und  $C_R$  der Kameras wird als  $d_B$  angenommen. Der Abstand zwischen den optischen Zentren und den Bildebenen wird durch  $f_k$ , mit

$k = \{L, R\}$ , angegeben und als Fokallänge bezeichnet. Abbildung 4.1 repräsentiert den schematischen Aufbau der Standardstereogeometrie.



**Abbildung 4.1:** Schematischer Aufbau eines Stereokamerasystems in Standardgeometrie. Die Bildebenen beider Kameras liegen parallel zur Basislinie. Die optischen Achsen stehen senkrecht auf der Basislinie und verlaufen parallel zueinander.

Der Vorteil dieser geometrischen Anordnung der Kameras ist die Lage korrespondierender Punkte  $p_l(x, y)$  und  $p_r(x, y)$  eines Stereobildpaares in der gleichen Bildzeile (engl. *scanline*). Damit wird die Suche nach Korrespondenzen von zwei auf eine Dimension eingeschränkt. Dient das linke Bild als Referenz, so gilt ohne Einschränkung der Allgemeinheit  $x_l \geq x_r$  und die Disparität kann mit  $d = x_l - x_r$  bestimmen werden. Mithilfe der Disparität und den Parametern des Kamerasystems (Basisabstand  $d_B$  und die Kamerakonstante  $f_k$ ) kann auf die 3D-Weltkoordinaten eines Punktes  $P_w(X, Y, Z)$  mit den Gleichungen

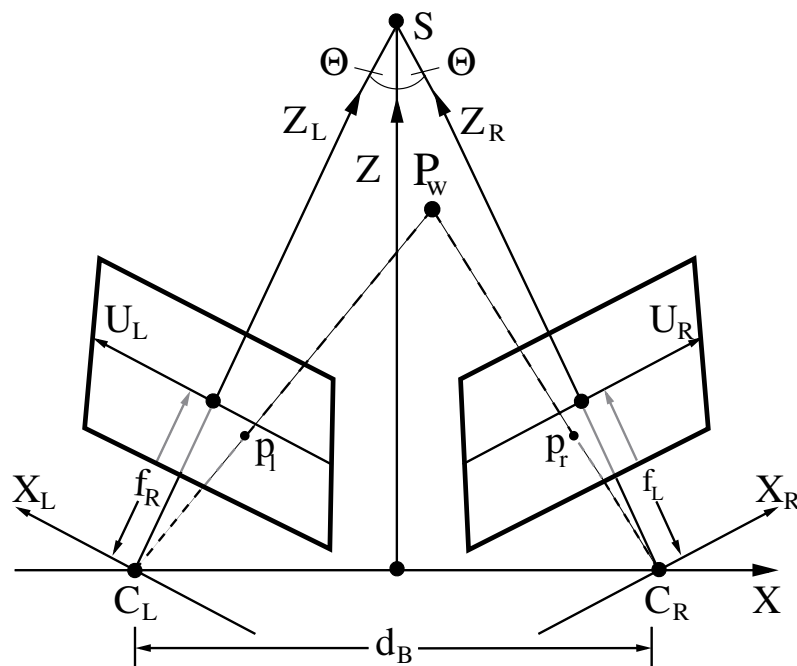
$$\begin{aligned} Z &= \frac{d_B \cdot f_k}{x_l - x_r} = \frac{d_B \cdot f_k}{d} \\ X &= \frac{d_B \cdot x_l}{x_l - x_r} = \frac{d_B \cdot x_l}{d} \\ Y &= \frac{d_B \cdot y_l}{x_l - x_r} = \frac{d_B \cdot y_l}{d} \end{aligned} \quad (4.1)$$

zurückgeschlossen werden. Der Nachteil der Standardstereogeometrie ist jedoch der geringausfallende Überschneidungsbereich der beiden Bilder. Des Weiteren lässt sich diese Anordnung, wegen den Abweichungen in der Anfertigung und ungenauen Positionierung der Kameras, physikalisch nicht präzise aufbauen, sodass eine Kalibrie-

zung des Stereosystems und anschließende Rektifikation der Bilder notwendig ist. Der Prozess der Rektifikation wird im Kapitel 4.4 behandelt.

## 4.2 Konvergente Stereogeometrie

Die konvergente Stereogeometrie ist die gängigste Methode und wird deswegen auch als allgemeine Stereogeometrie bezeichnet. Bei der konvergenten Anordnung werden die Kameras auf ein signifikantes Objekt beziehungsweise Szene ausgerichtet. Dadurch ergeben sich verschiedene Überschneidungsbereiche des Stereobildpaares. Mit der Ausrichtung auf ein signifikantes Objekt wird eine vollständige Erfassung des Objektes in beiden Kamerabildern erzielt. Der Überschneidungsbereich der Bilder außerhalb des Objektes wird dabei vernachlässigt. Bei der konvergenten Ausrichtung auf eine Szene wird ein größtmöglicher Überschneidungsbereich der Kamerabilder angestrebt, um den erfassten Weltausschnitt vollständig rekonstruieren zu können.



**Abbildung 4.2:** Schematischer Aufbau eines Kamerasystems in konvergenter Geometrie. Die optischen Achsen der beiden Kameras sind konvergent zueinander ausgerichtet und schneiden sich in einem Punkt  $S$ . Der Schnittwinkel zwischen den Achsen beträgt  $2\Theta$ .

In der konvergenten Stereoanordnung werden, wie bei der Standardstereogeometrie, Kameras mit gleicher Brennweite vorausgesetzt, die in dieselbe Ebene gelegt werden. Der Unterschied zu der Standardstereogeometrie liegt in der Ausrichtung der beiden Kameras. Die optischen Achsen der beiden Kameras sind konvergent und schneiden



sich in einem Punkt S. Der Winkel zwischen den optischen Achsen beträgt  $2\Theta$ . In Abbildung 4.2 wird die konvergente Ausrichtung der Bildebenen dargestellt. Das gemeinsame Koordinatensystem  $X, Y, Z$  liegt exakt zwischen den beiden Kameras und lässt sich durch Rotation um den Winkel  $\Theta$  an der  $Y$ -Achse und Transformation um  $d_B/2$  in die Koordinatensysteme  $X_l, Y_l, Z_l$  und  $X_r, Y_r, Z_r$  der einzelnen Kameras überführen. Mit den Gleichungen (4.2) gilt für die Koordinatensysteme der Kameras

$$\begin{pmatrix} X_l \\ Y_l \\ Z_l \end{pmatrix} = \begin{pmatrix} \cos(\Theta) & 0 & \sin(\Theta) \\ 0 & 1 & 0 \\ -\sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \begin{pmatrix} X - \frac{d_B}{2} \\ Y \\ Z \end{pmatrix} \quad (4.2)$$

$$\begin{pmatrix} X_r \\ Y_r \\ Z_r \end{pmatrix} = \begin{pmatrix} \cos(\Theta) & 0 & -\sin(\Theta) \\ 0 & 1 & 0 \\ \sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \begin{pmatrix} X + \frac{d_B}{2} \\ Y \\ Z \end{pmatrix}$$

Ein Punkt der Welt  $P_w(X, Y, Z)$  wird in den Bildebenen an die Punkte  $p_l(x, y)$  und  $p_r(x, y)$  projiziert. Der Zusammenhang zwischen den Punkten der Welt und Bildebenen wird durch die Abbildungsgleichungen 4.3 beschrieben.

$$x_l = \frac{f \cdot X_l}{Z_l} \quad y_l = \frac{f \cdot Y_l}{Z_l} \quad (4.3)$$

$$x_r = \frac{f \cdot X_r}{Z_r} \quad y_r = \frac{f \cdot Y_r}{Z_r}$$

Durch Einsetzen der Gleichungen aus (4.2) in (4.3) folgt

$$\begin{aligned} x_l &= f \cdot \frac{\cos(\Theta) \cdot (X - \frac{d_B}{2}) + \sin(\Theta) \cdot Z}{-\sin(\Theta) \cdot (X - \frac{d_B}{2}) + \cos(\Theta) \cdot Z} \\ y_l &= f \cdot \frac{Y}{-\sin(\Theta) \cdot (X - \frac{d_B}{2}) + \cos(\Theta) \cdot Z} \\ x_r &= f \cdot \frac{\cos(\Theta) \cdot (X + \frac{d_B}{2}) - \sin(\Theta) \cdot Z}{\sin(\Theta) \cdot (X + \frac{d_B}{2}) + \cos(\Theta) \cdot Z} \\ y_r &= f \cdot \frac{Y}{\sin(\Theta) \cdot (X + \frac{d_B}{2}) + \cos(\Theta) \cdot Z} \end{aligned} \quad (4.4)$$

Sind die korrespondierende Punkte  $p_l(x, y)$  und  $p_r(x, y)$  eines Stereobildpaares bekannt, so können die Koordinaten des zugehörigen Weltpunktes  $P_w(X, Y, Z)$ , aus den Gleichungen (4.4) durch Auflösen nach  $X$ ,  $Y$  und  $Z$  bestimmt werden. Zur Übersichtlichkeit des Gleichungssystems können zuvor folgende Koeffizienten ex-

trahiert werden

$$\begin{aligned}
 a_0 &= -\frac{d_B}{2}x_l \sin(\Theta) - \frac{d_B}{2}f \\
 a_1 &= -x_l \sin(\Theta) - f \cos(\Theta) \\
 a_3 &= x_l \cos(\Theta) - f \sin(\Theta)
 \end{aligned} \tag{4.5a}$$

$$\begin{aligned}
 b_0 &= \frac{d_B}{2}y_l \sin(\Theta) \\
 b_1 &= -y_l \sin(\Theta) \\
 b_2 &= -f \\
 b_3 &= y_l \cos(\Theta)
 \end{aligned} \tag{4.5b}$$

$$\begin{aligned}
 c_0 &= -\frac{d_B}{2}x_r \sin(\Theta) - \frac{d_B}{2}f \\
 c_1 &= x_r \sin(\Theta) - f \cos(\Theta) \\
 c_3 &= x_r \cos(\Theta) + f \sin(\Theta)
 \end{aligned} \tag{4.5c}$$

$$\begin{aligned}
 d_0 &= -\frac{d_B}{2}y_r \sin(\Theta) \\
 d_1 &= y_r \sin(\Theta) \\
 d_2 &= -f \\
 d_3 &= y_r \cos(\Theta)
 \end{aligned} \tag{4.5d}$$

Nach dem Einsetzen der Koeffizienten ergibt sich folgendes Gleichungssystem

$$\begin{aligned}
 a_1X + a_3Z &= a_0 \\
 b_1X + b_2Y + b_3Z &= b_0 \\
 c_1X + c_3Z &= c_0 \\
 d_1X + d_2Y + d_3Z &= d_0
 \end{aligned} \tag{4.6}$$

Bei der konvergenten Stereogeometrie gilt die Lagebeziehung korrespondierender Punkte der Standardstereogeometrie nicht mehr, sodass die Korrespondenzsuche anderweitig erfolgen muss. Dazu wird im nächsten Unterkapitel die Epipolargeometrie eines Stereosystems untersucht.

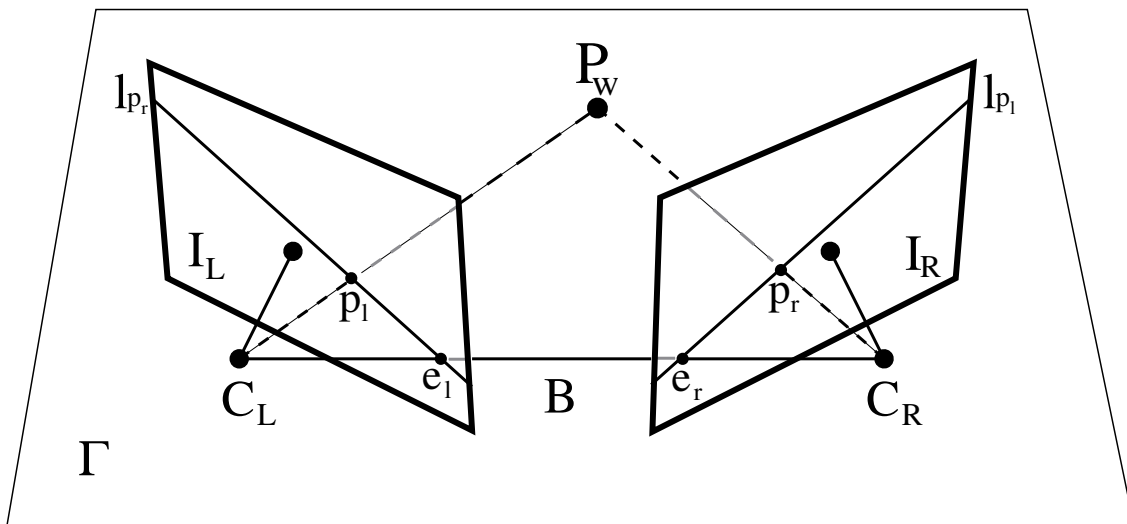
### 4.3 Epipolargeometrie

Durch die konvergente Ausrichtung der Kameras der allgemeinen Stereogeometrie liegen die korrespondierenden Punkte nicht in der gleichen y-Koordinate der beiden Bildebenen, sodass die Suche nach Korrespondenzen über die x- sowie der

y-Koordinate erfolgen muss. Die Epipolargeometrie eines Stereosystems bietet die Möglichkeit, den Suchbereich korrespondierender Punkte  $p_l(x, y)$  und  $p_r(x, y)$  auf eine Gerade einzuschränken [KKS96].

Der Sichtstrahl einer Kamera, durch das Linsenzentrum und einen Punkt in der Bildebene, sowie der Basislinie über die beiden Linsenzentren, spannen eine Epipolarebene auf. Alle Punkte, die auf der Epipolarebene  $\Gamma$  liegen, werden in der rechten beziehungsweise linken Bildebene auf die jeweilige Epipolarlinie abgebildet, die durch den Schnitt zwischen der Epipolar- und Bildebene entsteht. Das bedeutet, dass der zu einem im linken Bild liegenden Punkt  $p_l(x, y)$  korrespondierender Punkt  $p_r(x, y)$  nur entlang der Epipolarlinie in der rechten Bildebene gesucht werden muss. In der Abbildung 4.3 wird die epipolare Geometrie veranschaulicht.

Punkte, die ober- oder unterhalb der Epipolarebene liegen, erzeugen eine andere Epipolarebene und somit auch neue Epipolarlinien. Die Epipolarebenen bilden ein Ebenenbündel mit gemeinsamer Basislinie zwischen den Linsenzentren  $C_L$  und  $C_R$ . Bei der konvergenten Stereogeometrie schneiden sich die Paare der Epipolarlinien in den jeweiligen Epipolen.

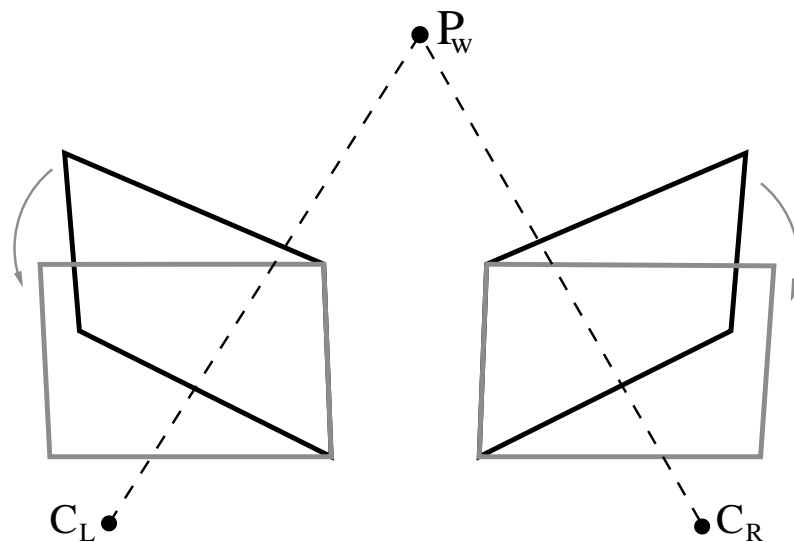


**Abbildung 4.3:** Epipolargeometrie eines Stereokamerasystems. Der Sichtstrahl von  $p_l$  über das optische Zentrum  $C_L$  und die Basislinie  $B$  spannen eine Ebene  $\Gamma$  auf. Der Schnitt dieser Ebene mit der Bildebene  $I_R$  ergibt die Epipolarlinie  $l_{p_l}$ .

Unter Verwendung der Epipolargeometrie wird zwar der Suchraum nach korrespondierenden Bildpunkten eingeschränkt, der Rechenaufwand für die Bestimmung der Epipolarlinien für jeden Bildpunkt ist jedoch nicht unerheblich. Bei einer konstanten Anordnung des Stereosystems ist es deswegen vorteilhaft die Epipolarrelationen zu berechnen und abzuspeichern. Eine andere Möglichkeit den Suchraum korrespondierender Punkte einzuschränken, bietet die Überführung in die Standardstereogeometrie. Dies lässt sich durch Rektifikation eines Stereobildpaares realisieren.

## 4.4 Rektifikation der Stereobildpaare

Die Rektifikation der Stereoanalyse befasst sich mit der Neuausrichtung von Stereobildpaaren, in der die Epipolarlinien beider Bilder auf die dieselbe Bildzeile projiziert werden und somit das Kamerasystem in die Standardstereogeometrie überführt wird. In Abbildung 4.4 wird Neuprojektion der beiden Bildebenen visuell veranschaulicht. Es existieren verschiedene Ansätze zur Rektifikation von Stereobildpaaren, die entweder die intrinsischen und extrinsischen Parameter oder äquivalente Lageinformationen (Fundamentalmatrix) des Kamerasystems als bekannt voraussetzen. Die intrinsischen und extrinsischen Parameter des Stereokamerasystems werden durch Kalibrierung bestimmt, siehe Kapitel 3. Die gewonnenen Parameter können dann unmittelbar für die Rektifikation nach [FTV97] verwendet werden.



**Abbildung 4.4:** Rektifikation eines Stereobildpaares. Durch die Rektifikation werden beide Bilder des Kamerasystems neu projiziert, sodass die Aufnahme einem Stereobildpaar in Standardstereogeometrie gleicht.

Im nächsten Unterkapitel werden die intrinsischen und extrinsischen Parameter als bekannt vorausgesetzt und ein Rektifikationsverfahren unter Verwendung dieser Parameter behandelt. In direktem Anschluss wird auf Rektifikationsverfahren unter Ausnutzung anderer äquivalenter Lageinformationen eingegangen, die auf Kenntnis der Kameraparameter verzichten.

### 4.4.1 Rektifikation mit bekannten Kalibrierungsparameter

Der Abbildungsprozess eines Punktes der Welt  $P_w$  in ein Stereokamerasystem kann durch zwei perspektivische Projektionsmatrizen  $M_l$  und  $M_r$  beschrieben werden.

Diese beinhalten die intrinsischen und extrinsischen Parameter der einzelnen Kameras des Stereosystems. In Kapitel 9.1.2 werden die Projektionsmatrizen genauer behandelt. Bei der Rektifikation werden die beiden Bildebenen auf eine gemeinsame Ebene projiziert, die parallel zu der Verbindungslinie zwischen den optischen Zentren  $C_L$  und  $C_R$  verläuft. Diese Projektionen lassen sich als zwei weitere Matrizen  $M_{nl}$  und  $M_{nr}$  beschreiben. In [FTV97] wird ein Algorithmus vorgestellt, der aus den bekannten und Neuberechneten Projektionsmatrizen für die Rektifikation notwendige Transformationen berechnet. Zur Berechnung der neuen Projektionsmatrizen  $M_{nl}$  und  $M_{nr}$  wird ein lineares und homogenes Gleichungssystem aus den Bedingungen für rektifizierte Stereosysteme aufgestellt, auf die in diesem Abschnitt noch eingegangen wird.

Die neuen Projektionsmatrizen beinhalten genau wie die Ausgangsmatrizen einen Rotations- und Transformationsanteil. Die  $3 \times 3$  Rotationsmatrizen der neuen Projektionsmatrizen sind jedoch gleich. Zu übersichtlicher Darstellung folgender Gleichungen wird der Rotationsanteil  $R$  der Neuberechneten Projektionsmatrizen  $M_{nl}$  und  $M_{nr}$  als Vektor von Zeilenvektoren aufgefasst und wie folgt notiert

$$R = \begin{bmatrix} a_1^T \\ a_2^T \\ a_3^T \end{bmatrix} \quad (4.7)$$

Verwendet man die Notation in den Projektionsmatrizen  $M_{nl}$  und  $M_{nr}$ , so können diese wie folgt dargestellt werden

$$\tilde{M}_{nl} = \left( \begin{array}{c|c} a_1^T & a_{14} \\ a_2^T & a_{24} \\ a_3^T & a_{34} \end{array} \right) = (M_{nl} | \tilde{m}_{nl}) \quad \tilde{M}_{nr} = \left( \begin{array}{c|c} b_1^T & b_{14} \\ b_2^T & b_{24} \\ b_3^T & b_{34} \end{array} \right) = (M_{nr} | \tilde{m}_{nr}) \quad (4.8)$$

Die Bedingungen für rektifizierte Stereosysteme liefern folgende Gleichungen:

1. **Skalierungsfaktor.** Die Projektionsmatrizen sind bis auf den Skalierungsfaktor definiert. Eine beliebige Wahl dieser Faktoren würde zu zweierlei Problemen führen:

Erstens werden die intrinsischen Parameter von der Wahl des Weltkoordinatensystems abhängig.

Zweitens kann aus den resultierenden Projektionsmatrizen nicht mehr sinnvoll auf die Kalibrierungsparameter geschlossen werden. Um diese Probleme zu umgehen, werden folgende Bedingungen eingeführt

$$\|a_3\| = 1 \quad \|b_3\| = 1 \quad (4.9)$$

2. **Position der optischen Zentren.** Die Position der optischen Zentren darf sich durch die Rektifikation nicht verändern

$$M_{nl} \begin{pmatrix} C_L \\ 1 \end{pmatrix} = 0 \quad M_{nr} \begin{pmatrix} C_R \\ 1 \end{pmatrix} = 0 \quad (4.10)$$

Diese Bedingungen ergeben sechs unabhängige Gleichungen

$$\begin{aligned} a_1^T C_l + a_{14} &= 0, & b_1^T C_r + b_{14} &= 0 \\ a_2^T C_l + a_{24} &= 0, & b_2^T C_r + b_{24} &= 0 \\ a_3^T C_l + a_{34} &= 0, & b_3^T C_r + b_{34} &= 0 \end{aligned} \quad (4.11)$$

3. **Gemeinsame Fokalebene.** Die rektifizierten Projektionen müssen eine gemeinsame Brennpunktebene besitzen. Daraus folgt für die Projektionsmatrizen

$$a_3 = b_3 \qquad a_{34} = b_{34} \quad (4.12)$$

4. **Ausrichtung der Epipolarlinien.** Die auf die rektifizierte Bildebenen abgebildete Weltpunkte  $P_w$  müssen die gleiche vertikale Koordinate besitzen.

$$\frac{a_2^T P_w + a_{24}}{a_3^T P_w + a_{34}} = \frac{b_2^T P_w + b_{24}}{b_3^T P_w + b_{34}} \quad (4.13)$$

Mit den Bedingungen aus (4.12) gilt

$$a_2 = b_2 \qquad a_{24} = b_{24} \quad (4.14)$$

5. **Orientierung der rektifizierten Bildebenen.** Die rektifizierten Bildebenen werden parallel zur Verbindungslinie zwischen den optischen Zentren  $C_L$  und  $C_R$  platziert.

$$a_3^T (f_l \times f_r) = 0 \quad (4.15)$$

Die Brennweiten  $f_l$  und  $f_r$  beziehen sich jeweils auf die dritte Zeile der Projektionsmatrizen  $M_l$  und  $M_r$ . Da nach Gleichung (4.14)  $a_2 = b_2$  gilt, liefert die Gleichung  $b_2^T (f_l \wedge f_r) = 0$  redundante Information.

6. **Orthogonalität der rektifizierten Ansichten.** Die Schnittgeraden zwischen den Bildebenen des Ausgangssystems und den Ebenen  $a_1^T P_w + a_{14} = 0$  und  $a_2^T P_w + a_{24} = 0$  korrespondieren mit den x- und y-Achsen. Da das Ausgangskoordinatensystem orthogonal ist, muss das neue Koordinatensystem ebenfalls orthogonal sein. Hieraus folgt

$$a_1^T a_2 = 0 \qquad b_1^T a_2 = 0 \quad (4.16)$$

7. **Kamerahauptpunkt.** Ist eine 3x4-Matrix mit vollem Rang gegeben, so ist der Kamerahauptpunkt  $p_0(u, v)$  wie folgt definiert

$$u_0 = a_1^T a_3 \qquad v_0 = a_2^T a_3 \quad (4.17)$$

Die Kamerahauptpunkte werden wieder an den Koordinaten (0,0) platziert. Durch Verwendung der Gleichungen (4.12) und (4.14) ergeben sich folgende Bedingungen

$$\begin{aligned} a_1^T a_3 &= 0 \\ a_2^T a_3 &= 0 \\ b_1^T a_3 &= 0 \end{aligned} \quad (4.18)$$

8. **Brennweite in Pixel.** Die horizontalen und vertikalen Brennweiten sind in Pixel definiert und durch folgende Gleichungen beschrieben

$$a_u = \|a_1 \times a_3\| \quad a_v = \|a_2 \times a_3\| \quad (4.19)$$

Werden die Werte  $a_u$  und  $a_v$  auf die Werte der Projektionsmatrix  $\tilde{M}_l$  gesetzt, so ergeben sich die Bedingungen

$$\begin{aligned} \|a_1 \times a_3\|^2 &= a_u^2 \\ \|a_2 \times a_3\|^2 &= a_v^2 \\ \|b_1 \times a_3\|^2 &= a_u^2 \end{aligned} \quad (4.20)$$

Mit der Beziehung für Vektorprodukte  $\|x \times y\|^2 = \|x\|^2 \|y\|^2 - (x^T y)^2$  und den Bedingungen aus (4.18) können die Gleichungen zu

$$\begin{aligned} \|a_1\|^2 \|a_3\|^2 &= a_u^2 \\ \|a_2\|^2 \|a_3\|^2 &= a_v^2 \\ \|b_1\|^2 \|a_3\|^2 &= a_u^2 \end{aligned} \quad (4.21)$$

umgeschrieben werden.

Alle zuvor aufgezählten Bedingungen werden folgendermaßen in vier Gleichungssysteme zusammengefasst

$$\begin{array}{cccc} a_3^T C_L + a_{34} = 0 & a_2^T C_L + a_{24} = 0 & a_1^T C_L + a_{14} = 0 & b_1^T C_R = -b_{14} \\ a_3^T C_R + a_{34} = 0 & a_2^T C_R + a_{24} = 0 & a_1^T a_2 = 0 & b_1^T a_2 = 0 \\ a_3^T (f_l \times f_r) = 0 & a_2^T a_3 = 0 & a_1^T a_3 = 0 & b_1^T a_3 = 0 \\ \|a_3\| = 1 & \|a_2\| = a_v & \|a_1\| = a_u & \|b_1\| = a_u \end{array} \quad (4.22)$$

Dies ergibt jeweils ein lineares homogenes Gleichungssystem mit einer quadratischen Nebenbedingung der Form

$$Ax = 0 \text{ mit } \|x'\| = k \quad (4.23)$$

Die Gleichungssysteme sind unabhängig voneinander und können durch Zerlegung gelöst werden. Damit werden die beiden Projektionsmatrizen  $M_{nl}$  und  $M_{nr}$  eindeutig

bestimmt. Nach [FTV97] lassen sich die Transformationsmatrizen aus der Multiplikation der Rektifikations- und Projektionsmatrizen berechnen.

$$T_l = M_{nl}M_l^{-1} \qquad T_r = M_{nr}M_r^{-1} \qquad (4.24)$$

Die Rektifikation der Bilder kann schließlich durch die Matrizenmultiplikation durchgeführt werden.

Das hier behandelte Rektifikationsverfahren aus [FTV97] basiert auf bekannten Kalibrierungsparametern, diese sind jedoch nicht immer gegeben. Im Unterkapitel 4.4.3 wird eine weitere Rektifikationsmethode behandelt, die auf berechneten Homographien basiert.

Im folgenden Abschnitt wird der Begriff *Fundamentalmatrix* erläutert und deren Herleitung beschrieben.

#### 4.4.2 Schätzung der Fundamentalmatrix

Die Fundamentalmatrix  $F$  ist eine  $3 \times 3$  Matrix vom Rang zwei. Diese enthält sowohl die intrinsischen Parameter beider Kameras als auch die extrinsischen Parameter der euklidischen Transformation und beschreibt somit vollständig die Epipolargeometrie in Bildkoordinaten. Die Fundamentalmatrix besitzt zwar neun Elemente, aber nur sieben Freiheitsgrade und ist somit überparametrisiert. Sechs der sieben Freiheitsgrade resultieren aus den linear unabhängigen Spaltenvektoren  $f_1$  und  $f_2$  von  $F$ . Der dritte Spaltenvektor  $f_3$  ist eine Linearkombination aus  $f_1$  und  $f_2$ :  $f_3 = \alpha f_1 + \beta f_2$ .

Für alle korrespondierende Punkte  $p_l(x_l, y_l)$  und  $p_r(x_r, y_r)$  in einem Stereobildpaar gilt

$$p_l^T F p_r = 0 \qquad (4.25)$$

Die Gleichung (4.25) besagt nichts anderes, als dass die Multiplikation eines Punktes  $p_l(x_l, y_l)$  im linken Bild mit der Fundamentalmatrix  $F$  die zu diesem Punkt entsprechende Epipolarlinie  $l_{p_l}^r$  im rechten Bild ergibt und umgekehrt. Formal aufgeschrieben ergeben sich folgende Gleichungen

$$l_{p_l}^r = F p_l \text{ und } l_{p_r}^l = F p_r \qquad (4.26)$$

Für die Epipole in Bildkoordinaten gelten die Beziehungen

$$F \tilde{e}_l = 0 \text{ und } F \tilde{e}_r = 0 \qquad (4.27)$$

Die Fundamentalmatrix kann mittels Punktkorrespondenzen aus zwei stereoskopischen Bildern geschätzt werden. Die Vielfalt der verfügbaren Algorithmen zur Schätzung der Fundamentalmatrix breitet sich von den linearen über iterativ-linearen bis hin zu nichtlinearen Verfahren aus. Die Güte der Algorithmen ist von der Anzahl, Positionsgenauigkeit korrespondierender Punkte und dem Optimierungsprozess abhängig. In folgenden Abschnitten werden einige Ansätze zur Schätzung der Fundamentalmatrix vorgestellt.



### Linearer Ansatz

Der lineare Ansatz zur Schätzung der Fundamentalmatrix basiert auf Kenntnis mehrerer Punktkorrespondenzen, unter deren Zuhilfenahme ein lineares Gleichungssystem aufgestellt wird. Die Normierung der Messwerte hat einen wesentlichen Einfluss in allen linearen Schätzverfahren. Diese wirkt auf die Konditionierung der Gleichungssysteme und damit auf die Robustheit der Lösung. Der Einfluss der Normierung auf die Parameterschätzung wurde in [Har97] explizit nachgewiesen und wird in allen Fällen der Schätzung der Fundamentalmatrix empfohlen.

Zur Aufstellung eines homogenen linearen Gleichungssystems wird die Gleichung (4.25) ausgeschrieben

$$(x_l, y_l, 1) = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \cdot \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} \quad (4.28)$$

Für jedes korrespondierende Punktepaar kann mit (4.28) eine Gleichung folgender Form

$$\begin{aligned} x_l x_r f_{11} + x_l y_r f_{12} + x_l f_{13} + y_l x_r f_{21} + \\ y_l y_r f_{22} + y_l f_{23} + x_r f_{31} + y_r f_{32} + f_{33} = 0 \end{aligned} \quad (4.29)$$

oder in abgekürzter Schreibweise

$$\begin{aligned} (x_l x_r, x_l y_r, x_l, y_l x_r, y_l y_r, y_l, x_r, y_r, 1) \cdot f = 0 \\ \text{mit } (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^T = f \end{aligned} \quad (4.30)$$

aufgestellt werden.

Mit  $n$  korrespondierenden Punkten entsteht ein lineares homogenes Gleichungssystem

$$A \cdot f = 0 \quad (4.31)$$

wobei

$$A = \begin{pmatrix} x_{l_1} x_{r_1} & x_{l_1} y_{r_1} & x_{l_1} & y_{l_1} x_{r_1} & y_{l_1} y_{r_1} & y_{l_1} & x_{r_1} & y_{r_1} & 1 \\ x_{l_2} x_{r_2} & x_{l_2} y_{r_2} & x_{l_2} & y_{l_2} x_{r_2} & y_{l_2} y_{r_2} & y_{l_2} & x_{r_2} & y_{r_2} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{l_n} x_{r_n} & x_{l_n} y_{r_n} & x_{l_n} & y_{l_n} x_{r_n} & y_{l_n} y_{r_n} & y_{l_n} & x_{r_n} & y_{r_n} & 1 \end{pmatrix}$$

und

$$f = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^T$$

Es existieren verschiedene Ansätze um das Gleichungssystem (4.31) zu lösen, zum Beispiel Normalisierung, Singulärwertzerlegung (engl. *singular value decomposition*)

SVD) und einige Andere. Das Gleichungssystem ist bis auf den Skalierungsfaktor eindeutig definiert. Bei der Lösung des Gleichungssystems müssen jedoch numerische Randbedingungen beachtet werden: Rang-2-Bedingung und schlecht konditioniertes System. Im Folgenden werden einige Algorithmen zur Lösung des Gleichungssystems (4.31) vorgestellt.

**Der 8-Punkt-Algorithmus** Für den 8-Punkt-Algorithmus werden, wie der Name des Algorithmus preis gibt, mindestens acht Paare korrespondierender Punkte benötigt. Dieser Algorithmus wurde erstmals von Longuet-Higgins in [LH81] vorgestellt.

Um numerisch stabilere Ergebnisse zu erhalten, werden als erstes die Bildpunkte normalisiert. Dabei wird auf alle Punkte  $x_{l_i}$  des linken Bildes eine Transformation  $T_l$  und auf alle Punkte  $x_{r_i}$  des rechten Bildes eine Transformation  $T_r$  angewandt, die den Ursprung des jeweiligen Koordinatensystems auf das Mittel der jeweiligen Punktmenge verschiebt und die Koordinaten so skaliert, dass der durchschnittliche Abstand der Punkte zum Ursprung  $\sqrt{2}$  beträgt.

Anschließend wird das Gleichungssystem (4.31) der normalisierten Koordinaten gelöst. Dazu wird die Matrix A mittels der Singulärwertzerlegung in die Matrizen U, S und V zerlegt. Im Idealfall bildet die letzte Spalte von V den Kern von A. Dies ist aber aufgrund der Messfehler in den Bildpunkten nur bei genau 8 Punkten der Fall. Die Singularität wird dadurch erzwungen, dass als erstes Fundamentalmatrix F aus der letzten Spalte von V gebildet wird.

Um die bestimmte Fundamentalmatrix F auf Rang 2 zu bringen wird F per SVD zerlegt und der kleinste Singulärwert gleich Null gesetzt. Mit der neuen Matrix S und den Matrizen U und V, die bei der zweiten SVD entstanden sind, wird  $\hat{F}=USV^T$  berechnet. So wird die Singularität erzwungen.

Die berechnete Fundamentalmatrix  $\hat{F}$  bezieht sich durch die Normalisierung auf Punktkorrespondenzen  $(T_l p_l, T_r p_r)$ . Es gilt also

$$(p_r^T T_r^T) \hat{F} (T_l p_l) = 0$$

und damit

$$p_r^T (T_r^T) \hat{F} (T_l) p_l = 0$$

Um sich auf die ursprünglichen Punktkorrespondenzen  $(p_l, p_r)$  zu beziehen, müssen die Daten denormalisiert werden. Dies wird durch die Bestimmung der ursprüngliche Fundamentalmatrix F mit der Gleichung

$$F = T_r^T \hat{F} T_l \tag{4.32}$$

realisiert.

Ein weiteres lineares Verfahren zur Schätzung der Fundamentalmatrix ist der *7-Punkt-Algorithmus*, der dem *8-Punkt-Algorithmus* ziemlich nahe kommt. Dieser Algorithmus ist nun das Thema des nächsten Abschnittes.

**7-Punkt-Algorithmus** Wie in der Einführung des Kapitels (4.4.2) schon erwähnt wurde, besitzt die Fundamentalmatrix nur sieben unabhängige Parameter. Demnach sind sieben Punktkorrespondenzen für die Bestimmung der Fundamentalmatrix ausreichend. Die Gleichung (4.31) mit sieben Punktkorrespondenzen führt zu einer  $7 \times 9$  Matrix  $A$ , die den Rang sieben aufweist.

Die Lösung des Gleichungssystems (4.31), mit der Singulärwertzerlegung, ist der zweidimensionale Nullraum und hat die Form

$$F = \alpha F_1 + (1 - \alpha) F_2 \quad (4.33)$$

Da für die Fundamentalmatrix  $F$  die Rang-Bedingung und somit  $\det(F) = 0$  gelten muss, kann die Fundamentalmatrix durch lösen der kubischen Gleichung,

$$\det(F) = \det(\alpha F_1 + (1 - \alpha) F_2) = 0 \quad (4.34)$$

aufgelöst nach  $\alpha$ , bestimmt werden. Die kubische Gleichung (4.34) liefert genau eine oder drei Lösungen für  $\alpha$ . Aus diesen Lösungen wird diejenige ausgewählt, welche den geringsten geometrischen Fehler aufweist.

Um numerisch stabilere Ergebnisse zu erhalten, sollten die Daten analog zum *8-Punkt-Algorithmus* normalisiert werden.

Der Vorteil des *7-Punkt-Algorithmus* gegenüber dem *8-Punkt-Algorithmus* ist, dass die Fundamentalmatrix nicht nachträglich auf Rang 2 gebracht und somit die Singulärwertzerlegung nur einmal durchgeführt werden muss.

Nach [Sch05b] besteht das Interesse den Prozess zur Bestimmung korrespondierender Punkte und somit auch der Fundamentalmatrix automatisch zu gestalten. Die automatische Zuordnung korrespondierender Punkte liefert jedoch nicht immer eindeutige Ergebnisse, sodass die iterativen nichtlinearen Verfahren durch Elimination stark fehlerbehafteter Punktkorrespondenzen einen enormen Vorteil besitzen. Der *RANSAC*-Algorithmus gehört zu der Klasse iterativen nichtlinearen Verfahren und kann in Kombination mit dem *7-* und *8-Punkt-Algorithmus* in automatischen Methoden zur Bestimmung der Fundamentalmatrix eingesetzt werden.

**RANSAC-Algorithmus** *RANSAC (RANDOM SAMPLE CONSENSUS)* ist ein mathematischer Algorithmus zur Elimination von Ausreißern beziehungsweise groben Fehlern innerhalb einer Menge von Datenpunkten. Dieser Algorithmus wurde in [FB81] von Martin A. Fischler und Robert C. Bolles in *Communications of the ACM* vorgestellt. Die Voraussetzung für diesen Algorithmus ist, dass die Anzahl der Datenpunkte größer als zur Bestimmung der Modellparameter tatsächlich benötigten werden. Die Funktionsweise des Algorithmus kann dann durch folgende Verarbeitungsschritte beschrieben werden:

1. Zufälliges Auswählen von  $n$  Datenpunkten, die für die Bestimmung der Modellparameter benötigt werden. Dies geschieht in Erwartung, dass diese Menge frei von Ausreißern ist.

2. Ermittlung der Modellparameter aus den gewählten Messpunkten.
3. Bewertung der Modellparameter mittels eines Gütekriterium. Die Messpunkte werden entsprechend ihrer Güte entweder weiterverwendet (diese Teilmenge wird „Consensus set“ genannt) oder verworfen. Alle Punkte, die eine größere Abweichung haben, werden für dieses Modell als grobe Fehler angesehen und sind nicht in der Teilmenge enthalten. Sind ausreichend Punkte in der Teilmenge vorhanden, werden diese gespeichert.
4. Wiederhole die Schritte 1–3  $N$  mal.

Nach  $N$  Durchläufen entsteht als Resultat ein reduzierter Umfang von Messwerten, der keine Ausreißer mehr enthält.

Die Anzahl der nötigen Durchläufe  $N$  kann durch Angabe einer festdefinierten Wahrscheinlichkeit  $p$  zum Beispiel  $p=0.99$  berechnet werden. Dabei ist  $p$  die Wahrscheinlichkeit, dass eine Zufallswahl aus dem reduzierten Umfang von Messwerten keine (tatsächlichen) Ausreißer beinhaltet. Die Gleichung zur Berechnung von  $N$  ist folgend

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \quad (4.35)$$

definiert, wobei  $s$  die Anzahl der Modellparameter und  $\epsilon$  der Anteil von Ausreißern in der Gesamtdatenmenge ist.

Der Anteil an Ausreißern in der Gesamtmenge der Datenpunkte ist oft unbekannt. Somit ist es nicht möglich, die benötigte Zahl der Iterationen  $N$  und die Größe des Consensus set  $t$  zu bestimmen. In diesem Fall wird der Algorithmus mit der Worst-Case-Annahme eines Ausreißeranteils von 50% initialisiert. Nach jeder Iteration wird dann dieser Wert angepasst, falls eine größere konsistente Menge gefunden wurde. Wird zum Beispiel mit dem Initialwert  $\epsilon = 50\%$  begonnen und enthält aber damit den berechneten Consensus set von 80% aller Datenpunkte, ergibt sich für die nächste Iteration ein verbessertes  $\epsilon$  von 20%. Analog werden  $N$  und  $t$  festgelegt. Beide werden unter der Worst-Case-Annahme von  $\epsilon = 50\%$  berechnet und nach jeder Änderung von  $\epsilon$  neu bestimmt.

Nachdem der Begriff Fundamentalmatrix und deren Herleitung behandelt sind, wird im folgenden Abschnitt auf die Rektifikation der Stereobildpaare mittels Homographien unter möglicher Verwendung der Fundamentalmatrix behandelt.

#### 4.4.3 Rektifikation mit Homographien

Bevor auf den Ansatz der Rektifikation mittels Homographien eingegangen wird, soll an dieser Stelle der Begriff *Homographie* der statischen Stereoanalyse eingeführt werden. Der Begriff *Homographie* setzt sich aus den griechischen Wörter *homos* und *graphiein* zusammen und bedeuten so viel wie gleich und zeichnen, schreiben. Die Homographie ist nach [Sch05b] eine Abbildung zwischen zwei Ebenen, bei der

allen geometrischen Elementen der ersten Ebene entsprechende Elemente der zweiten Ebene zugeordnet werden können.

Die Homographie in der statischen Stereoanalyse ist eine  $3 \times 3$  Matrix und hat folgende Form

$$H = \begin{bmatrix} u_a & u_b & u_c \\ v_a & v_b & v_c \\ w_a & w_b & 1 \end{bmatrix} \quad (4.36)$$

Es existieren eine Menge Verfahren zum Beispiel [Har98], [Sun03] oder [Zha98], die aus bekannter Lage des Epipols die Möglichkeit bieten Homographien  $H_l$  und  $H_r$  zu bestimmen. Diese Homographien überführen die Ausgangsbilder eines Stereokamerasystems in rektifizierte Ansichten. Das Gemeinsame der meisten Verfahren ist das Bestreben den Epipol eines Bildes um eine der Achsen zu rotieren und anschließend nach unendlich zu projizieren. Schließlich wird das neu erzeugte Bild auf die erforderliche Bildgröße skaliert.

Der Ausgangspunkt für die Berechnung der Homographien ist die Epipolargeometrie der Standardstereoanordnung. Das in [Zha98] publizierte Verfahren von Z. Zhang benötigt zusätzlich noch die Kenntnis der Fundamentalmatrix.

Die Fundamentalmatrix für rektifizierte oder aus standardstereogeometrischer Kameraanordnung stammende Bilder hat die Form

$$\hat{F} = [i]_{\times} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad (4.37)$$

Die Fundamentalmatrix  $\hat{F}$  hat die Eigenschaft, beliebige Bildpunkte  $p_{(i,j)} = [u_i, v_j, 1]^T$  auf Geraden entlang derselben v-Koordinate abzubilden

$$\hat{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_i \\ v_j \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ v_j \end{bmatrix} \quad (4.38)$$

Durch die Homographien  $H_l$  und  $H_r$  werden die Punkte  $\hat{p}_l$  und  $\hat{p}_r$  der Ausgangsbilder mit

$$p_l = H_l \hat{p}_l \quad p_r = H_r \hat{p}_r \quad (4.39)$$

neu projiziert. Mit den Gleichungen aus (4.39) kann die Gleichung (4.25) wie folgt umgeschrieben werden

$$p_l^T \hat{F} p_r = 0 \quad (4.40)$$

$$\hat{p}_l H_l^T \hat{F} H_r \hat{p}_r = 0 \quad (4.41)$$

Die Fundamentalmatrix hat also die Faktorisierung

$$F = H_l^T [i]_{\times} H_r \quad (4.42)$$

Da die Fundamentalmatrix wie schon im Abschnitt 4.4.2 erwähnt skalierungsinvariant und somit nicht eindeutig bestimmt ist, sind auch die Homographien aus der Gleichung (4.39) keinesfalls eindeutig definiert.

Die Homographien werden von der Standardstereogeometrie ausgehend berechnet. Nach Definition der Standardstereogeometrie liegen die Epipolarlinien beider Kameras parallel zur u-Achse und die Epipole demzufolge im Unendlichen.

$$\tilde{e}_l = \tilde{e}_r = (1, 0, 0)^T \quad (4.43)$$

Unter Verwendung der Homographien  $H_l$  und  $H_r$  zur projektiven Rektifikation kann die Gleichung (4.43) für nichtrektifizierte Bilder erweitert werden.

$$H_l \cdot \tilde{e}_l = (1, 0, 0)^T \quad H_l \cdot \tilde{e}_l = (1, 0, 0)^T \quad (4.44)$$

Im Folgenden wird eine mögliche Zerlegung der Homographien vorgestellt, die das Bestimmen dieser erleichtert. Auf die Berechnung der einzelnen Teilmatrizen wird jedoch nicht weiter eingegangen, da dies den Rahmen dieser Arbeit übersteigt. Der interessierte Leser wird somit mit dem Abschluss des folgenden Absatzes auf weiterführende Literatur verwiesen.

**Faktorisierung der Homographie** Eine Homographiematrix  $H$  kann in vier Teilmatrizen zerlegt werden

$$H = H_u \cdot H_s \cdot H_r \cdot H_p \quad (4.45)$$

wobei  $H_u$  eine uniforme Skalierungsmatrix,  $H_s$  eine Scherungsmatrix,  $H_r$  eine Rotationsmatrix und  $H_p$  eine Projektionsmatrix ist. Da die Homographien skalierungsinvariant sind, ist es zweckmäßig, sie über ihr skalierungsvariantes Gegenstück zu berechnen

1. Faktorisierung der Homographie  $H$  in affine  $H_a$  und projektive  $H_p$  Matrizen. Dadurch wird das Minimieren der projektiven Verzerrungen vereinfacht, da es nur einen Teil der Matrix betrifft.

$$H = H_a \cdot H_p \quad (4.46)$$

wobei

$$H_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ w_a & w_b & 1 \end{pmatrix} \quad (4.47)$$

und  $H_a$  unter Zuhilfenahme der Gleichungen (4.36) und (4.47) wie folgt berechnet wird

$$H_a = H H_p^{-1} = \begin{pmatrix} u_a - u_c w_a & u_b - u_c w_b & u_c \\ v_a - v_c w_a & v_b - v_c w_b & v_c \\ 0 & 0 & 1 \end{pmatrix} \quad (4.48)$$

2. Faktorisierung der affinen Matrix  $H_a$  in Scherungs- und Rotationsmatrizen. Die Scherungsmatrix  $H_s$  wird für den Ausgleich der durch  $H_p$  eingebrachten projektiven Verzerrungen benötigt. Die Rotationsmatrix  $H_r$  dreht das Bild um einen Winkel  $\alpha$ , sodass die Epipolarlinien achsenparallel zur u-Koordinate liegen. Des Weiteren werden durch die Matrix  $H_r$  die korrespondierende Epipolarlinien aufeinander auszurichten, sodass diese dieselbe v-Koordinate haben. Dies wird durch eine Verschiebung eines der beiden Bilder in v-Richtung ermöglicht.

$$H_a = H_s H_r \quad (4.49)$$

wobei

$$H_r = \begin{pmatrix} v_b - v_c w_b & v_c w_a - v_a & 0 \\ v_a - v_c w_a & v_b - v_c w_b & v_c \\ 0 & 0 & 1 \end{pmatrix} \quad (4.50)$$

und

$$H_s = \begin{pmatrix} s_a & s_b & s_c \\ 0 & 1 & 0 \\ w_a & w_b & 1 \end{pmatrix} \quad (4.51)$$

3. Uniforme Skalierung der Bilder mit der Matrix  $H_u$ , falls die Bilder nicht mehr der ursprünglichen Größe entsprechen.

Auf die Bestimmung der einzelnen Matrizen wird in dieser Arbeit nicht weiter eingegangen und auf weiterführende Literatur verwiesen.

## 4.5 Zusammenfassung

Am Anfang dieses Kapitels wurde ein Einblick in die statische Stereoanalyse gegeben und deren behandelnde Gebiete sowie verfolgtes Ziel beschrieben. Der Gesamtprozess zur Verfolgung des gesetzten Ziels der statischen Stereoanalyse behandelt ein umfangreiches Gebiet, sodass es in mehrere Verarbeitungsschritte unterteilt wurde. Einige Verarbeitungsschritte, die in die statische Stereoanalyse miteinfließen, wurden schon in Kapiteln 2 und 3 behandelt, dies sind der Bildaufnahme- und Kamerakalibrationsprozess. In diesem Kapitel wurde der Schwerpunkt auf die geometrischen Zusammenhänge eines Stereosystems gelegt und die unterschiedlichen Anordnungen der Kameras untersucht, dabei wurde in den Unterkapiteln 4.1 und 4.2 die Standardstereogeometrie und die konvergente Stereogeometrie behandelt. Weitere Verarbeitungsschritte der statischen Stereoanalyse sind Korrespondenzanalyse und Tiefenrekonstruktion, die in den Kapiteln 5 und 6 behandelt werden. Der Prozess der Korrespondenzanalyse weist ein hohes Maß an Komplexität der Datenverarbeitung auf. Um die Komplexität der Korrespondenzanalyse zu reduzieren, kann eine Rektifikation der Stereobildpaare durchgeführt werden. Dies ist ein weiterer Bestandteil dieses Kapitels gewesen und behandelte die verschiedenen Methoden der Rektifikation.





Die Korrespondenzanalyse in der statischen Stereoanalyse beschreibt Such- und Zuordnungsverfahren, die jedem Punkt eines Bildes, nach festgelegten Kriterien, ein Punkt eines anderen Bildes zuordnen. In diesem Kapitel werden einige Algorithmen der Korrespondenzsuche vorgestellt und nach ihren Merkmalen kategorisiert. Eine mögliche Kategorisierung der Algorithmen kann zum Beispiel anhand der verwendeten Suchkriterien oder des zu durchsuchenden Korrespondenzenraums vorgenommen werden. In dieser Arbeit wird eine Kategorisierung in merkmal- und flächenbasierte Verfahren nach [Müh02] angewendet.

Da die Korrespondenzsuche immens rechenintensiv sein kann und die Zuordnung von korrespondierenden Punkten nicht zwangsläufig eindeutig ist, werden unter Kenntnis der geometrischen Zusammenhänge des Stereokamerasystems und der Eigenschaften von Objekten unserer natürlichen Umwelt gewisse Annahmen und Einschränkungen getroffen. Diese finden in der Korrespondenzsuche Anwendung und werden vor der Beschreibung der Korrespondenzanalyseverfahren behandelt.

## 5.1 Annahmen und Einschränkungen

In [KKS96] werden mögliche Annahmen und Einschränkungen, bei bekannter Stereogeometrie und Verwendung der Eigenschaften von Objekten, vorgestellt. Diese werden im Folgenden eingeführt und in einzeln behandelt.

Die Annahmen und Einschränkungen nach [KKS96] werden in zehn Punkte unterteilt:

1. Epipolare Geometrie,
2. Eindeutigkeit der Zuordnung,
3. Verträglichkeit von Intensitätswerten,
4. Geometrische Ähnlichkeit der Merkmale (Winkelkriterium, Längenkriterium),
5. Kontinuität der Disparitäten,
6. Kontinuität entlang der Kanten,

7. Verträglichkeit von Merkmalen,
8. Das Disparitätslimit,
9. Das Disparitätsgradientenlimit und
10. Reinform der Punkte in den Bildern.

Die hier genannten Annahmen und Einschränkungen sind nicht alle gleichzeitig anwendbar, sondern kommen abhängig vom gewählten Korrespondenzanalyseverfahren zum Einsatz. Im Folgenden werden nur ausgewählte, für diese Arbeit relevante, Annahmen und Einschränkungen erörtert.

**Epipolare Geometrie.** Der zu einem Punkt im linken Bild korrespondierende Punkt im rechten Bild, liegt auf zugehöriger Epipolarlinie des linken Punktes. Mit dieser Einschränkung reduziert sich die Suche des korrespondierenden Punktes von einem zweidimensionalen auf ein eindimensionales Problem. Bei der Standardstereogeometrie stimmen die Epipolarlinien mit der Scannlinie überein. Allgemein können bei fester Anordnung der Kameras die Epipolarlinien einmalig bestimmt und abgespeichert werden.

**Eindeutigkeit der Zuordnung.** Für jeden Punkt im linken Bild existiert nur bis auf wenige Ausnahmen ein korrespondierender Punkt im rechten Bild. Eine Ausnahme tritt nur dann auf, wenn zwei Punkte der Welt auf einen Sichtstrahl der Kamera fallen, die von der Anderen getrennt sichtbar sind.

**Verträglichkeit von Intensitätswerten.** Zwei Punkte  $p_l(x_l, y_l)$  und  $p_r(x_r, y_r)$  eines Stereopaars sind nur dann Korrespondent, wenn die Differenz  $|I_{p_l(x_l, y_l)} - I_{p_r(x_r, y_r)}|$  der Intensitäten klein ist.

In der folgenden Annahme geht man davon aus, dass die Oberflächen eines Objektes kontinuierlich sind und nur Objektgrenzen eine Ausnahme bilden, die jedoch den unwesentlichen Anteil eines Bildes belegen.

**Kontinuität der Disparitäten.** Sind zwei Punkte  $p_{l1}(x_{l1}, y_{l1})$  und  $p_{r1}(x_{r1}, y_{r1})$  eines Stereobildpaares korrespondent, so können die Nachbarpunkte  $p_{l2}(x_{l2}, y_{l2})$  und  $p_{r2}(x_{r2}, y_{r2})$  nur dann korrespondieren, wenn die absolute Differenz der Disparitätswerte

$$| \sqrt{(x_{l1} - x_{r1})^2 + (y_{l1} - y_{r1})^2} - \sqrt{(x_{l2} - x_{r2})^2 + (y_{l2} - y_{r2})^2} | \quad (5.1)$$

bzw. bei der Standardstereogeometrie  $||x_{l1} - x_{r1}| - |x_{l2} - x_{r2}||$  gering ist.

**Das Disparitätslimit.** Für sämtliche Punkte  $P_l(x_l, y_l)$  im linken Bild, für die Korrespondenzen  $P_r(x_r, y_r)$  im rechten Bild existieren, gibt es eine maximale Disparität  $d_{max}$  und es gilt folgende Ungleichung

$$\sqrt{(x_l - x_r)^2 + (y_l - y_r)^2} < d_{max}. \quad (5.2)$$

Kommt die Standardstereogeometrie zum Einsatz, so gilt  $|x_l - x_r| < d_{max}$ . Das Disparitätenlimit definiert zugleich den Mindestabstand zwischen den Objekten der Szene und dem Stereosystem.

**Das Disparitätsgradientenlimit.** Ein Paar benachbarter Punkte  $p_{l1}(x_{l1}, y_{l1})$  und  $p_{l2}(x_{l2}, y_{l2})$  im linken Bild korrespondieren nur dann die mit einem benachbarten Punktepaar  $p_{r1}(x_{r1}, y_{r1})$  und  $p_{r2}(x_{r2}, y_{r2})$  im rechten Bild, wenn der Disparitätsgradient

$$\Gamma_d = \frac{|\Delta x_l - \Delta x_r|}{\sqrt{\frac{1}{4}(\Delta x_l - \Delta x_r)^2 + \Delta y^2}} \quad (5.3)$$

einen festen Wert  $\Gamma_{max}$  nicht überschreitet  $\Gamma_d < \Gamma_{max}$ . Wobei  $\Delta x_l$  und  $\Delta x_r$  die Differenzen der jeweiligen x-Koordinaten sind und  $\Delta y$  die Differenz zwischen  $y_{l1}$  und  $y_{l2}$ .

**Reinfolge der Punkte in den Bildern.** Liegen die Punkte eines Stereobildpaares in einem Bild auf einer Epipolarlinie, so werden diese in gleicher Reinfolge auf einer korrespondierenden Epipolarlinie des anderen Bildes abgebildet. Bei dieser Annahme wird davon ausgegangen, dass sich die Objekte der Szene ungefähr im gleichen Abstand zum Kamerasystem befinden.

## 5.2 Korrespondenzanalyseverfahren

Die Korrespondenzanalyseverfahren können, wie oben erwähnt, in merkmals- und flächenbasierte Klassen unterteilt werden, siehe [Müh02]. In der flächenbasierten Korrespondenzanalyse wird die Suche der Korrespondenzen für alle Punkte der Bildfläche durchgeführt. Das Ergebnis dieser Verfahren ist üblicherweise eine dichte Disparitätskarte mit einem Disparitätswert pro Bildpunkt. In der merkmalsbasierten Korrespondenzanalyse werden vor der Korrespondenzsuche Bildmerkmale extrahiert zum Beispiel Kanten, Punkte oder andere typische Merkmale. Das Ergebnis der merkmalsbasierten Korrespondenzanalyse ist eine auf den Bildmerkmalen berechnete Disparitätskarte.

Für das folgend beschriebene Korrespondenzanalyseverfahren wird zunächst der besseren Übersicht die Standardstereogeometrie vorausgesetzt. Im Folgenden werden Algorithmen der flächen- beziehungsweise merkmalsbasierten Korrespondenzanalyse vorgestellt.

### 5.2.1 Merkmalsbasierte Verfahren

Bei der merkmalsbasierten Korrespondenzanalyse wird vor der eigentlichen Korrespondenzsuche eine Merkmalsextraktion durchgeführt. Dies kann eine Extraktion von Eckpunkten, Kanten, Bildausschnitte mit hohem Kontrast oder besonderen

Farbwerten, bis hin zum Erkennen einzelner Objekte sein. Die Extraktion erfolgt mithilfe eines entsprechenden Operators, zum Beispiel kann der Algorithmus aus [HS88] zur Extraktion von Eckpunkten und der Laplacian of Gaussian zur Extraktion von Kanten verwendet werden.

In folgenden Abschnitten werden zwei Algorithmen vorgestellt, die eine Extraktion typischer Bildmerkmale durchführen. Als erstes wird der Algorithmus aus [HS88] vorgestellt, der aus einem Bild Eckenpunkte extrahiert. Im direkten Anschluss wird dann der Laplacian of Gaussian behandelt, der aus einem Bild Kanten extrahiert. Zur Extraktion anderer typischer Bildmerkmale wird auf weiterführende Literatur verwiesen.

### Extraktion von Punktmerkmalen

Für die Bestimmung der Eckpunkte nach [HS88] werden als erstes die Gradienten in horizontaler und vertikaler Richtung berechnet. Die diskrete Approximation der Gradienten in beide Richtungen wird, wie folgt definiert

$$\frac{df}{du} = f(u - 1, v) - f(u + 1, v), \quad \frac{df}{dv} = f(u, v - 1) - f(u, v + 1) \quad (5.4)$$

Um die Extraktion der Kantenpunkte unempfindlich gegenüber Bildrauschen zu gestalten, werden die Quadrate der örtlichen Ableitungen einer Tiefpassfilterung unterzogen. Die Filterung des Bildes wird durch Faltung  $\otimes$  mit einer Gewichtsfunktion  $g$  realisiert und ergibt folgende Ergebnisse

$$\begin{aligned} \left(\frac{\delta f}{\delta u}\right)^2 &= g \otimes \left(\frac{df}{du} \cdot \frac{df}{du}\right) \\ \left(\frac{\delta f}{\delta v}\right)^2 &= g \otimes \left(\frac{df}{dv} \cdot \frac{df}{dv}\right) \\ \left(\frac{\delta f}{\delta uv}\right)^2 &= g \otimes \left(\frac{df}{du} \cdot \frac{df}{dv}\right) \end{aligned} \quad (5.5)$$

Damit kann dann folgende Matrix aufgestellt werden

$$M = \begin{bmatrix} \left(\frac{\delta f}{\delta u}\right)^2 & \left(\frac{\delta f}{\delta uv}\right)^2 \\ \left(\frac{\delta f}{\delta uv}\right)^2 & \left(\frac{\delta f}{\delta v}\right)^2 \end{bmatrix} \quad (5.6)$$

Die in Gleichung (5.6) angegebene Matrix  $M$  liefert das Kriterium für die Erkennung von Ecken. Die Determinante der Matrix kann folgend

$$\det(M) = \left(\frac{\delta f}{\delta u}\right)^2 \left(\frac{\delta f}{\delta v}\right)^2 - \left(\frac{\delta f}{\delta uv}\right)^2 \quad (5.7)$$

berechnet werden. Diese nimmt nur dann einen von Null verschiedenen Wert an, wenn eine große Änderung in horizontaler und vertikaler Richtung vorliegt. Zur Unterscheidung zwischen Kanten und Punkten wird zusätzlich die Spur der Matrix, das heißt die Summe der Hauptdiagonale herangezogen

$$\text{trace}(M) = \left(\frac{\delta f}{\delta u}\right)^2 + \left(\frac{\delta f}{\delta v}\right)^2 \quad (5.8)$$

Liegt eine Intensitätsänderung in beide Richtungen vor, so nimmt die Spur einen großen Wert an und es handelt sich um einen Eckpunkt. Bei Kanten hingegen verläuft die Intensitätsänderung nur in eine Richtung so, dass die Spur der Matrix relativ klein ausfällt. Das Auswahlkriterium wird dann wie folgt

$$K = \det(M) - k(\text{trace}(M))^2 \quad (5.9)$$

definiert, wobei  $k$  ein Gewichtungsfaktor ist und nach [HS88] zu 0.04 gewählt wird.

Im nächsten Abschnitt wird die Extraktion von Kantenmerkmalen unter Verwendung des Laplacian of Gaussian-Operators behandelt.

### Extraktion von Kantenmerkmalen

Kantenfilter werden zum Auffinden von starken Gradienten der Helligkeit innerhalb eines Bildes eingesetzt. Aufgrund des differenzierenden Charakters reagieren Kantenfilter sehr empfindlich auf Bildrauschen, sodass vor dem Anwenden eines Kantenoperators meistens eine Glättung des Bildes durchgeführt werden muss. Die Glättung eines Bildes kann zum Beispiel mithilfe eines Gauß-Filters stattfinden. Im folgendem präsentierte Laplacian of Gaussian-Filter ist eine Kombination des Gauß- und Laplace-Filters, sodass die Glättung des Bildes vor der Gradientenberechnung entfällt. Die Vorteile dieses Filters liegen in der Hand, der Prozess der Kantenextraktion erfolgt in einem Verarbeitungsschritt und somit die Verarbeitungsdauer eines Bildes kurz gehalten wird.

Der Laplacian of Gaussian resultiert aus dem Assoziativgesetz der Faltung

$$(\Delta^2 G) * I = \Delta^2(G * I) \quad (5.10)$$

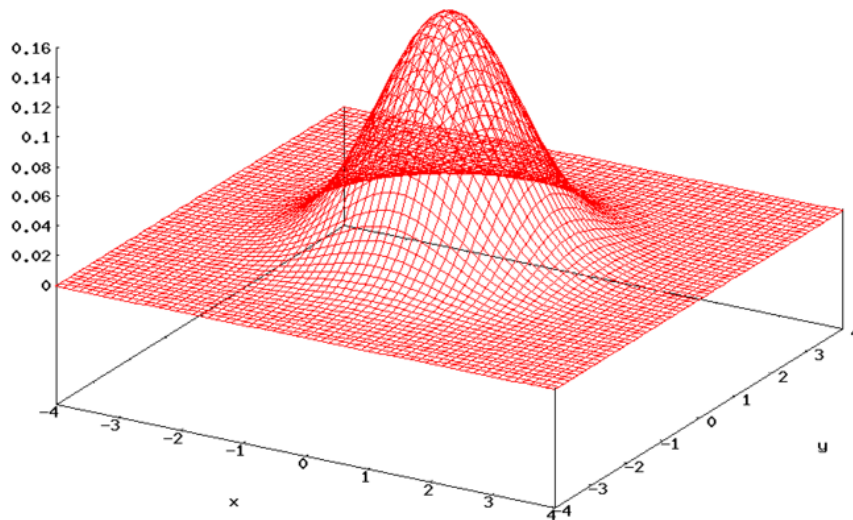
wobei  $\Delta^2$  der Laplace-Filter,  $G$  der Gauß-Filter und  $I$  das zu verarbeitende Bild sind. Durch Anwenden des Laplace-Operators auf den Gauß, ist der Laplacian of Gaussian folgend gegeben

$$\Delta^2 G(x, y) = -\frac{1}{\pi\sigma^4} \cdot \left[1 - \frac{x^2+y^2}{2\sigma^2}\right] \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5.11)$$

Zur Verwendung des Laplacian of Gaussian für die digitale Bildverarbeitung, wird der kontinuierliche Kantenoperator diskret approximiert. Ein diskreter Kernel der Größe  $7 \times 7$  mit  $\sigma = 1.6$  könnte folgend aussehen

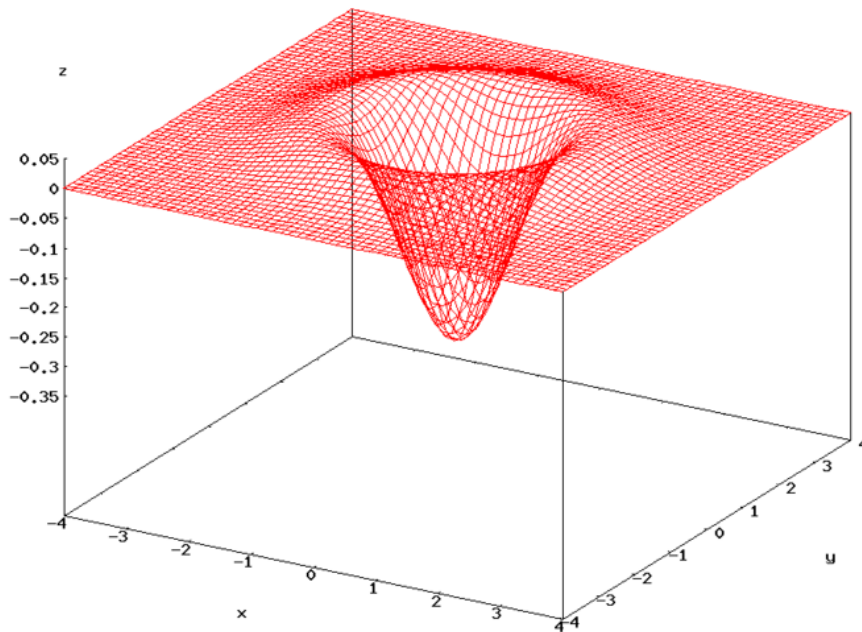
$$\Delta^2 G(x, y) = \begin{bmatrix} 1 & 3 & 4 & 4 & 4 & 3 & 1 \\ 3 & 4 & 3 & 0 & 3 & 4 & 3 \\ 4 & 3 & -9 & -17 & -9 & 3 & 4 \\ 4 & 0 & -17 & -30 & -17 & 0 & 4 \\ 4 & 3 & -9 & -17 & -9 & 3 & 4 \\ 3 & 4 & 3 & 0 & 3 & 4 & 3 \\ 1 & 3 & 4 & 4 & 4 & 3 & 1 \end{bmatrix} \quad (5.12)$$

Die grafische Repräsentation des Laplacian of Gaussian ähnelt stark einem mexikanischen Sombrero und wird daher auch gerne als Mexican-Hat bezeichnet. In den Abbildungen 5.1 und 5.2 werden die zweidimensionalen Gauß- und Laplacian of Gaussian-Funktionen grafisch dargestellt.



**Abbildung 5.1:** Grafische Darstellung des zweidimensionalen Gauß-Operators. Die Anwendung dieses Operators auf ein Bild führt eine Glättung in diesem durch. Somit kann im Bild vorhandenes Rauschen eliminiert werden.

In diesem Unterkapitel wird Laplacian of Gaussian behandelt, der Kanten aus Bildern extrahiert. Im Folgendem wird ein Korrespondenzanalyseverfahren vorgestellt,



**Abbildung 5.2:** Grafische Darstellung des zweidimensionalen Laplacian of Gaussian-Operators. Die Anwendung dieses Operators auf ein Bild hebt die Kanten in diesem hervor.

das in [KKS96] zur Berechnung von Disparitätskarten auf Kantenbilder verwendet wird.

### Shirai-Algorithmus

Der hier behandelte Algorithmus wird in [KKS96] vorgestellt. Dieser berechnet aus Grauwert-Kanten eines Stereobildpaares die Disparitätskarte. Für die Berechnung des Ähnlichkeitsmaßes zwischen zwei Kantenpunkte wird die von Y. Shirai definierte Funktion  $\text{SIMILARITY}(p,q)$  verwendet. Diese basiert auf dem Intensitätsvergleich zweier um die Punkte gebildete Fenster. Dafür wird im linken Bild ein Kantenpunkt  $p_l(x, y)$  fixiert, der quadratischer Fehler über die Intensität der Fenster um die Punkte  $p_l(x, y)$  und  $p_r(x, y)$  berechnet und durch die Varianz des fixierten Punktes  $p_l(x, y)$  mit gleicher Fenstergröße dividiert. Diese Prozedur wird für alle auf der Scanlinie liegende Kantenpunkte im rechten Bild durchgeführt. Der Kantenpunkt  $p_r(x, y)$ , der den kleinsten  $\text{SIMILARITY}$ -Wert aufweist, gilt als korrespondierend zu  $p_l(x, y)$ .

Die Berechnung des quadratischen Fehlers ist mit einer Fenstergröße  $(2k+1)$  und der Intensitätsfunktion der Pixel  $E_{\{l,r\}}$  folgend definiert

$$SE(p_l, p_r) = \sum_{i=-k}^k \sum_{j=-k}^k (E_l(x+i, y+j) - E_r(x+i, y+j))^2 \quad (5.13)$$

Die Varianz des fixierten Punktes  $p_l(x, y)$  im linken Bild, ist mit derselben Fenstergröße durch

$$\begin{aligned} VAR(p_l) &= \frac{1}{(2k+1)^2} \sum_{i=-k}^k \sum_{j=-k}^k (E_l(x+i, y+j) - AVER(p_l))^2 \\ &= \frac{1}{(2k+1)^2} \sum_{i=-k}^k \sum_{j=-k}^k E_l(x+i, y+j)^2 - AVER(p_l)^2 \end{aligned} \quad (5.14)$$

gegeben, wobei AVER den arithmetischen Mittelwert des Fensters beschreibt. Mit den Gleichungen (5.13) und (5.14) wird letztendlich das Ähnlichkeitsmaß mit

$$SIMILARITY(p_l, p_r) = \frac{SE(p_l, p_r)}{VAR(p)} \quad (5.15)$$

berechnet.

In [KKS96] Abschnitt 1.4.2 ist zu diesem Algorithmus der Korrespondenzanalyse zusätzlich der Pseudocode unter Verwendung der SIMILARITY-Funktion vorzufinden. Durch die passende Einstellung der im Pseudocode vorkommenden freien Variablen lässt sich die Anzahl korrespondierender Punkte zusätzlich steigern. Für weitere Details, siehe [KKS96].

### 5.2.2 Flächenbasierte Verfahren

In der flächenbasierten Korrespondenzanalyse erstreckt sich die Korrespondenzsuche über alle Punkte eines Referenzbildes. Dient das linke Bild einer Stereoaufnahme als Referenzbild, so wird versucht jedem Punkt dieses Bildes einen korrespondierenden Punkt im rechten Bild zuzuordnen. Als mögliches Vergleichskriterium bietet sich die Intensität der Bildpunkte an. Da jedoch der Vergleich zwischen zwei Bildpunkten stark fehleranfällig ist, nutzen die meisten flächenbasierte Korrespondenzanalyseverfahren, die Intensität eines Blockes um einen Punkt herum.

Der Vorteil der flächenbasierten Korrespondenzanalyse ist die Möglichkeit der vollständigen Rekonstruktion der Tiefe von Objekten in Raum. Der Nachteil, die Notwendigkeit der immens-hohen Rechenleistung. Je nach Anwendung und Hardwaresystem muss deswegen eventuell ein Kompromiss, zwischen Geschwindigkeit der Verarbeitung und Qualität der Korrespondenzanalyse, eingegangen werden.

In folgenden Abschnitten werden Korrespondenzanalyseverfahren zur Berechnung dichter Disparitätskarten vorgestellt.

### Block-Matching

Block-Matching ist ein weitverbreitetes Standardverfahren zur Encodierung von Videosequenzen. Ziel des Block-Matching-Verfahrens ist die effektive Schätzung und



Speicherung von Bewegungsverläufen zweier nacheinander folgender Bilder. Dafür wird für jeden Punkt des aktuellen Bildes ein Bewegungsvektor aus dem zeitlich-vorhergehenden Bild berechnet und gespeichert. Verändert sich die Position eines Punktes in zwei aufeinanderfolgenden Bildern nicht, so wird der dazugehörige Bewegungsvektor gleich null gesetzt. Durch Weglassen dieser Vektoren ergibt sich eine Codierung einer Videosequenz.

Die Grundidee des Verfahrens für die Berechnung von Disparitätskarten zu verwenden, basiert auf der Annahme gleicher Disparität innerhalb eines Blockes und Vergleich der Grauwertverteilung zweier Blöcke mit der gleichen Größe. Die Berechnung von Disparitätskarten mit Block-Matching setzt sich aus mehreren Verarbeitungsschritten zusammen, die im Folgendem beschrieben werden:

1. Unterteilung eines Bildes zum Beispiel des Linken in eine konstante Anzahl von gleichgroßen Blöcken  $(2n + 1) \times (2m + 1)$
2. Auffinden korrespondierender Blöcke im rechten Bild und Bestimmung der Blockdisparitäten. Zum Vergleich der Blöcke können unterschiedliche Metriken verwendet werden, die nach [BBH03] in drei Klassen unterteilt werden können. Dies sind Korrelation (engl. *correlation*), jegliche Art von Intensitätsdifferenzen (engl. *intensity differences*) und Rangmetriken (engl. *rank metrics*).

Seien  $E_{\{l,r\}}(x, y)$ ,  $E'_{\{l,r\}}(x, y)$  die Intensitäten der Punkte  $p_{\{l,r\}}$  an Position  $(x, y)$  und  $\bar{E}_{\{l,r\}}(x, y)$  die mittlere Intensität über das Fenster  $(2n+1) \times (2m+1)$  mit dem Mittelpunkt an  $(x, y)$ , so können die Metriken der drei Klassen folgend beschrieben werden.

Ein Pendant der Korrelationsklasse ist zum Beispiel die normierte Kreuzkorrelation (engl. *normalized crosscorrelation*, *NCC*). Diese ist mit

$$\bar{E}_{\{l,r\}}(x, y) = \frac{1}{(2n + 1) \cdot (2m + 1)} \sum_{i=-n}^n \sum_{j=-m}^m E_{l,r}(x + i, y + j) \quad (5.16)$$

als

$$NCC(x, y, \Delta) = \sum_{i=-n}^n \sum_{j=-m}^m \frac{(E_l(x + i, y + j) - \bar{E}_l(x, y)) \cdot (E_r(x + i + \Delta, y + j) - \bar{E}_r)}{\sqrt{(E_l(x + i, y + j) - \bar{E}_l(x, y))^2 \cdot (E_r(x + i + \Delta, y + j) - \bar{E}_r)^2}} \quad (5.17)$$

definiert. Mit der NCC wird die statistische Ähnlichkeit zwischen den Blöcken gemessen, wobei die Kreuzkorrelation umso größer wird, je ähnlicher sich die Intensitätsverläufe der beiden Bildausschnitte sind.

Zu der Klasse der Intensitätsdifferenzen gehören zum Beispiel die Summe der quadratischen Abstandsdifferenzen (engl. *sum of squared differences*, *SSD*),

Summe der absoluten Abstandsdifferenzen (engl. *sum of the absolute differences, SAD*), mittlere quadratische Fehler (engl. *mean square error, MSE*) oder mittlere absolute Fehler (engl. *mean absolute error, MAE*). Im Folgendem werden die Gleichungen der hier erwähnten Intensitätsdifferenzen nacheinander aufgelistet

$$SSD(x, y, \Delta) = \sum_{i=-n}^n \sum_{j=-m}^m (E_l(x_l + i, y_l + j) - E_r(x_r + i + \Delta, y_r + j))^2 \quad (5.18)$$

$$SAD(x, y, \Delta) = \sum_{i=-n}^n \sum_{j=-m}^m |E_l(x_l + i, y_l + j) - E_r(x_r + i + \Delta, y_r + j)| \quad (5.19)$$

$$MSE(x, y, \Delta) = \frac{1}{(2n + 1) \cdot (2m + 1)} \sum_{i=-n}^n \sum_{j=-m}^m |E_l(x_l + i, y_l + j) - E_r(x_r + i + \Delta, y_r + j)|^2 \quad (5.20)$$

$$MAE(x, y, \Delta) = \frac{1}{(2n + 1) \cdot (2m + 1)} \sum_{i=-n}^n \sum_{j=-m}^m |E_l(x_l + i, y_l + j) - E_r(x_r + i + \Delta, y_r + j)| \quad (5.21)$$

Bei optimaler Übereinstimmung zweier Blöcke wird das Ergebnis der Intensitätsdifferenzen zum Vergleich mit der Korrelation jedoch minimal.

In der Klasse der Rangmetriken wird vor dem Matching eine nicht-parametrische Transformation beider Bilder durchgeführt. In [SS02] werden *rank transform* und die *census transform* als nicht-parametrische Transformationen vorgeschlagen. Diese sind in Gleichungen (5.22) und (5.23) aufgeführt.

$$E'_{\{l,r\}}(x, y) = \sum_{i=-n}^n \sum_{j=-m}^m E_{\{l,r\}}(x_l + i, y_l + j) < E_{\{l,r\}}(x, y) \quad (5.22)$$

$$E'_{\{l,r\}}(x, y) = BITSTR_{i,j} E_{\{l,r\}}(x_l + i, y_l + j) < E_{\{l,r\}}(x, y) \quad (5.23)$$

mit  $i = -n, \dots, n$  und  $j = -m, \dots, m$

Nach der durchgeführten Vorverarbeitung mit der Gleichung (5.22), können die korrespondierende Blöcke mit den vorher vorgestellten Metriken bestimmt werden.

Auf den berechneten Bitstrings mit der Gleichung (5.23) werden korrespondierende Blöcke anhand des kleinsten Hammingabstandes zwischen allen Pixeln

innerhalb der Blöcke bestimmt. Dafür wird folgende Gleichung

$$R(x, y, \Delta) = \sum_{i=-n}^n \sum_{j=-m}^m \text{HAM}(E_{\{l,r\}}(x_l + i, y_l + j) < E_{\{l,r\}}(x, y)) \quad (5.24)$$

verwendet.

Der Parameter  $\Delta$  beschreibt ein Versatz, der die Verschiebung der Matrix in Spalten im Rechten im Bezug zum linken Bild wiedergibt. Der höchste Versatz wird meistens zur Effizienz der Verarbeitung durch ein Disparitätslimit  $d_{max}$  beschränkt. Die Blockdisparität gibt den Abstand in Spalten zwischen den Blöcken mit kleinstem quadratischen Fehler wieder. Diese ist jedoch nur dann definiert, wenn ein eindeutiges Minimum existiert, ansonsten wird ein zusätzliches Entscheidungskriterium benötigt. Das Ergebnis der Block-Matching-Verfahrens ist eine Disparitätskarte von Blöcken.

3. Verfeinerung der Ergebnisse auf Pixelebene, sodass jedem Punkt des linken Bildes ein Disparitätswert  $DISP(x'_l, y'_l)$  zugeordnet wird. Die Disparität ist durch den Wert  $D(k)$  definiert, für den die Funktion

$$DIFF(k) = | E_l(x'_l, y'_l) - E_r(x'_r + D(k), y'_r) | \quad \text{mit } k = 1, \dots, n \cdot m \quad (5.25)$$

ihr Minimum annimmt. Die Funktion  $DIFF(k)$  ist die Differenz zwischen den Intensitätswerten der Punkte an den Positionen  $(x'_l, y'_l)$  und  $(x'_r + D(k), y'_r)$  in den jeweiligen Bildern. Diese wird für alle Disparitäten  $D(k)$ , mit  $(1 < k < n \cdot m)$  aus der  $n \times m$  Blockumgebung berechnet.

## Dynamische Programmierung

Ein weiteres Pendant der flächenbasierten Korrespondenzanalyse ist der in [BT98] publizierte Algorithmus von S. Birchfield und C. Thomasi. Dieser Algorithmus setzt wie die meisten anderen Korrespondenzanalyse-Algorithmen auch ein rektifiziertes Stereobildpaar voraus und macht die Annahme, dass die Epipolarlinien mit den Bildzeilen übereinstimmen. Die Berechnung der Disparitätskarte eines Stereobildpaares dieses Algorithmus basiert auf dem Prinzip der Dynamischen Programmierung und setzt sich aus zwei Verarbeitungsschritten zusammen.

Das Prinzip der Dynamischen Programmierung wurde von [Bel57] vorgestellt und gehört zu der Klasse der Optimierungs-Algorithmen. Die Idee dieses Algorithmus ist, ein komplexes Problem (in Korrespondenzanalyse: Matchen zweier Bildzeilen) in kleinere Teilprobleme (in Korrespondenzanalyse: Matchen zweier Bildpunkte) zu zerlegen und die Lösungen für diese zu berechnen. Die optimale Lösung des ursprünglichen Problems resultiert aus der Zusammensetzung der Teillösungen mit geringsten Kosten. Der Vorteil dieser Vorgehensweise gegenüber rekursiven Ansätzen

ist, dass Teilergebnisse nicht unabhängig voneinander mehrmals berechnet werden, sondern in einer Kostenmatrix abgespeichert werden.

Die Anwendung des Prinzips der Dynamischen Programmierung in der Bildverarbeitung wurde erstmals in [Bak81] vorgestellt. Im Folgendem wird der auf Dynamische Programmierung basierender Korrespondenzanalyse-Algorithmus von [BT98] behandelt und die beiden Verarbeitungsschritte des Algorithmus veranschaulicht:

1. Die Einschränkung der Suche nach korrespondierenden Punkten auf eine Bildzeile ergibt sich aus dem Verlauf der Epipolarlinien der Standardstereogeometrie, die, wie oben erwähnt, für diesen Algorithmus als Voraussetzung gilt. In diesem Verarbeitungsschritt wird jedem Bildpunkt  $x_i$  der linken Bildzeile ein Bildpunkt  $y_i$  der rechten Bildzeile zugeordnet (engl. *matching*) und jede Übereinstimmung als Tupel  $(x_i, y_i)$  in einer Match-Sequenz (engl. *match sequence*) festgehalten. Für jede übereinstimmende Match-Sequenz  $M$  wird anschließend die Kostenfunktion

$$\gamma(M) = N_{occ}k_{occ} - N_mk_r + \sum_{i=1}^{N_m} d(x_i, y_i) \quad (5.26)$$

berechnet, die eine Wahrscheinlichkeit beschreibt, dass  $M$  die wahre Korrespondenz ist. Die Kosten der Match-Sequenzen sind durch eine konstante Strafe verdeckter Sequenzen  $k_{occ}$ , eine konstante Belohnung für korrekte Sequenzen  $k_r$  und der Summe über das Unähnlichkeitsmaß  $d(x_i, y_i)$  der gepaarten Punkte bestimmt.  $N_{occ}$  und  $N_m$  beschreiben die Anzahl von verdeckten und korrekten Sequenzen.

2. Nachdem die Kosten der Match-Sequenzen berechnet sind, wird mithilfe der Methode der dynamischen Programmierung die optimale Sequenz entlang einer Bildzeile aus den Match-Sequenzen mit minimalen Kosten bestimmt.

Zur Berechnung der Kosten der Match-Sequenzen benutzen S. Birchfield und C. Thomasi eine eigendefinierte Unähnlichkeitsfunktion  $d(x_i, y_i)$ , die im Folgenden beschrieben wird. Die Unähnlichkeitsfunktion hat die Form

$$d(x_i, y_i) = \min\{\bar{d}(x_i, y_i, I(x_i), I(y_i)), \bar{d}(y_i, x_i, I(y_i), I(x_i))\} \quad (5.27)$$

und erweist sich als unempfindlich gegenüber Bildabtafelfehlern, da diese nicht nur die Differenz der Intensitäten  $I(x_i)$  und  $I(y_i)$  einzelner Pixel betrachtet, sondern auch die Nachbarpixel in die Berechnung mit einbezieht. Die aktuell betrachteten Pixel auf einer Zeile haben die Intensität  $I(x_i)$  im Linken beziehungsweise  $I(y_i)$  im rechten Bild. Die Nachbarn von  $I(x_i)$  auf der Zeile seien  $I(x_{i-1})$  und  $I(x_{i+1})$ . Zwischen  $I(x_i)$  und den Nachbarn werden die Intensitäten  $I(x_{i-\frac{1}{2}})$  und  $I(x_{i+\frac{1}{2}})$  interpoliert. Aus den Intensitätswerten  $I(x_i)$ ,  $I(x_{i-\frac{1}{2}})$  und  $I(x_{i+\frac{1}{2}})$  wird ein Intervall mit den Grenzen  $[I_{min}, I_{max}]$  durch

$$\begin{aligned} I_{min} &= \min(I(x_i), I(x_{i-\frac{1}{2}}), I(x_{i+\frac{1}{2}})) \\ I_{max} &= \max(I(x_i), I(x_{i-\frac{1}{2}}), I(x_{i+\frac{1}{2}})) \end{aligned} \quad (5.28)$$

definiert. Aus den Ergebnissen der Gleichungen (5.28) wird der erste Parameter für (5.27) durch

$$\bar{d}(x_i, y_i, I(x_i), I(y_i)) = \{0, I(x_i) - I_{max}, I_{min} - I(x_i)\} \quad (5.29)$$

berechnet.

Der zweite Parameter der Gleichung (5.27) lässt sich analog zu (5.28) und (5.29) bestimmen.

### 5.3 Zusammenfassung

Das vorliegende Kapitel befasste sich mit der Korrespondenzanalyse, dabei wurde in der Einleitung auf den Begriff selbst eingegangen und im weiteren Verlauf des Kapitels die einzelnen Bestandteile behandelt. Da die Korrespondenzsuche immens rechenintensiv sein kann und die Zuordnung korrespondierender Punkte nicht zwangsläufig eindeutig ist, wurden vor der Beschreibung der Korrespondenzanalyseverfahren einige Annahmen und Einschränkungen getroffen. Diese dienen der einfacheren Gestaltung der Such- und Zuordnungsalgorithmen und setzen die Voraussetzungen zur Erlangung bessere Ergebnisse. Die vorgestellten Korrespondenzanalyseverfahren wurden der besseren Strukturierung des Kapitels in merkmals- und flächenbasierte Verfahren klassifiziert. Das Ergebnis dieser Algorithmen sind Punktzuordnungen, die den Ausgangspunkt für die, im nächsten Kapitel behandelten, Tiefenrekonstruktion der 3D-Weltpunkte setzen.

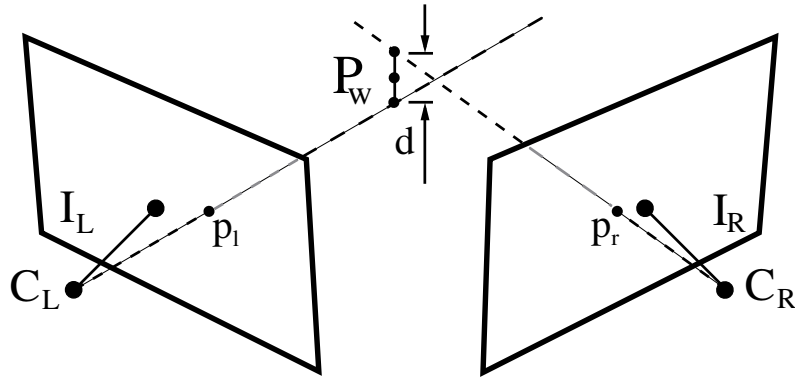


Wie im Kapitel 2 schon erwähnt ist, geht beim Abbildungsprozess eines Objektes beziehungsweise einer Szene in eine Kamera die Tiefeninformation verloren. Im Laufe vorangegangenen Kapiteln ist nun gezeigt worden, dass durch zusätzliche Information, in dieser Arbeit durch Aufnahme des gleichen Objektes beziehungsweise der gleichen Szene aus einer differenten Perspektive, die Tiefeninformation wieder gewonnen werden kann.

Die Rekonstruktion der Tiefeninformation kann auf zwei unterschiedlichen Wegen erfolgen. Erstens, durch Einsetzen korrespondierender Punkte in die Gleichungen (4.1) oder (4.6) aus den Unterkapiteln 4.1 beziehungsweise 4.2 je nach Geometrie des Stereokamerasystems. Die Berechnung auf diesem Wege setzt jedoch absolut identische Kameras voraus, sodass die intrinsischen Parameter des Kamerasystems gemittelt werden müssen. Zweitens, durch Stereotriangulation unter Verwendung der perspektivischen Projektionsmatrizen beider Kameras. Die Berechnung der Tiefeninformation auf diesem Wege setzt die Voraussetzung, dass der Abbildungsprozess keiner nichtlinearen Verzerrung unterliegt. In folgenden Unterkapiteln wird das Gleichungssystem der Stereotriangulation aufgestellt und die daraufbezogenen Lösungsverfahren behandelt.

## 6.1 Die Stereotriangulation

In diesem Unterkapitel wird die sogenannte Stereotriangulation behandelt, die aus den im Kapitel 5 bestimmten Korrespondenzen die Tiefeninformation der Punkte bestimmt. Um die 3D-Weltkoordinaten eines Punktes rekonstruieren zu können, ist für die Stereotriangulation in erster Linie die Kenntnis der perspektivischen Projektionsmatrix beider Kameras vorausgesetzt. Die Beschreibung dieser Matrizen ist in Kapitel 9 zu finden. Da von der Annahme ausgegangen werden muss, dass die Punktkorrespondenzen nicht absolut exakt sind, werden sich ihre optischen Strahlen nicht genau in einem Punkt schneiden. Nach [Sch05b, Kap. 9] ist eine gute Approximation der tatsächlichen Lage eines 3D-Punktes die Mittelposition der kürzesten Strecke zwischen den beiden optischen Strahlen. Zur visuellen Veranschaulichung siehe Abbildung 6.1.



**Abbildung 6.1:** Stereotriangulation. Ausgehend von der Annahme, dass die korrespondierenden Punkte nicht absolut exakt sind, schneiden sich ihre optischen Strahlen nicht genau in einem Punkt. Eine gute Approximation der tatsächlichen Lage des Weltpunktes  $P_w$  ist nach [Sch05b, Kap. 9] durch  $d/2$  gegeben, wobei  $d$  den kürzesten Abstand zwischen den Geraden darstellt.

In der folgenden Beschreibung der Rekonstruktion wird davon ausgegangen, dass die Projektion einer Szene in die Bildebenen der Kameras keiner nichtlinearen Verzerrung unterliegt, das heißt, der Abbildungsprozess wird nicht durch die Linsen der Kameras radial verzerrt. Unter Zuhilfenahme der allgemeinen Projektionsgleichung aus Kapitel 3 werden die Bestimmungsgleichungen nach [Sch05b] Kapitel 9 hergeleitet.

Ein 3D-Punkt  $P_w$  im Weltkoordinatensystem wird durch die Projektionsmatrizen  $M^i$  der Kameras in die Bildpunkte  $P_c^i$  transformiert

$$sP_c^i = M^i P_w \quad (6.1)$$

wobei das  $i = \{l, r\}$  für die rechte beziehungsweise linke Kamera steht. Im folgendem bezeichnet  $m_{jk}^i$  das  $(j, k)$ -Element der Projektionsmatrix der entsprechenden Kamera. Werden die ersten drei Elemente einer Zeile der Matrix  $M^i$  zu einem Vektor  $m_j^i = (m_{j1}^i, m_{j2}^i, m_{j3}^i)$  zusammengefasst, so kann durch Anwendung des Vektorprodukts der homogene Skalierungsfaktor nach [Sch05b] Kapitel 9 eliminiert werden

$$\begin{aligned} y_i(m_3^i P_w) - (m_2^i P_w) &= 0 \\ x_i(m_3^i P_w) - (m_1^i P_w) &= 0 \\ x_i(m_2^i P_w) - (m_1^i P_w) &= 0 \end{aligned} \quad (6.2)$$

In den folgenden Unterkapiteln werden zwei Verfahren vorgestellt, die unter Zuhilfenahme des Gleichungssystems (6.2) 3D-Rekonstruktionen aus Punktkorrespondenzen vornehmen.



### 6.1.1 Homogenes Lösungsverfahren

Durch Einsetzen der in Unterkapitel 5.2 berechneten Punktkorrespondenzen für die beiden Komponenten  $x$  und  $y$  in die zwei linear unabhängigen Gleichungen des Gleichungssystems (6.2) entsteht eine Gleichung folgender Form

$$BP_w = 0$$

$$\text{mit } B = \begin{bmatrix} x_l m_3^{lT} - m_1^{lT} \\ y_l m_3^{lT} - m_2^{lT} \\ x_r m_3^{rT} - m_1^{rT} \\ y_r m_3^{rT} - m_2^{rT} \end{bmatrix} \quad (6.3)$$

Die Gleichung 6.2 kann mittels Direkten Linearen Transformation unter Verwendung der Nebenbedingung  $\|P_w\| = 1$  durch Lösung des Eigenwertproblems gelöst werden.

### 6.1.2 Inhomogenes Lösungsverfahren

Durch Erweiterung des Weltpunktes  $P_w$  um eine Komponente, die gleich eins gesetzt wird, führt zu einem Satz von vier inhomogenen Gleichungen mit drei unbekanntem  $X_w, Y_w$  und  $Z_w$ . Für die korrespondierenden Punkte  $p_i$  ergeben sich dann die vier Gleichungen

$$\begin{aligned} (m_1^l - x_l \cdot m_3^l)^T \cdot P_w + m_{14}^l - x_l \cdot m_{34}^l &= 0 \\ (m_2^l - y_l \cdot m_3^l)^T \cdot P_w + m_{24}^l - y_l \cdot m_{34}^l &= 0 \\ (m_1^r - x_r \cdot m_3^r)^T \cdot P_w + m_{14}^r - x_r \cdot m_{34}^r &= 0 \\ (m_2^r - y_r \cdot m_3^r)^T \cdot P_w + m_{24}^r - y_r \cdot m_{34}^r &= 0 \end{aligned} \quad (6.4)$$

Diese Gleichungen können mit

$$B_i = \begin{pmatrix} (m_1^i - x_i \cdot m_3^i) \\ (m_2^i - y_i \cdot m_3^i) \end{pmatrix} \text{ und } c_i = \begin{pmatrix} x_i \cdot m_{34}^i - m_{14}^i \\ y_i \cdot m_{34}^i - m_{24}^i \end{pmatrix} \text{ für } i = l, r \quad (6.5)$$

in Matrixschreibweise als

$$B \cdot P_w = c, \text{ mit } B = \begin{pmatrix} B_l \\ B_r \end{pmatrix}, P_w(X_w, Y_w, Z_w, 1) \text{ und } c = \begin{pmatrix} c_l \\ c_r \end{pmatrix} \quad (6.6)$$

aufgefasst werden.

Die Lösung für das Problem der Minimierung nach dem kleinsten quadratischen Fehler ist schließlich nach [Sch05b] Kapitel 9 durch

$$P_w = (B^T \cdot B)^{-1} \cdot B^T \cdot c \quad (6.7)$$

definiert. Dabei wird  $B^T B$  als invertierbar vorausgesetzt.

### 6.1.3 Vergleich der beiden Lösungsverfahren

Beide Verfahren setzen eine Bedingung als Voraussetzung: das homogene Verfahren setzt  $\|P_w\| = 1$  voraus und das inhomogene Verfahren  $P_w = (X_w, Y_w, Z_w, 1)$ . Weder die Bedingung für das Homogene, noch für das inhomogene Verfahren sind invariant gegenüber projektiven Transformationen. Das heißt, dass eine Lösung für das ursprüngliche Problem nicht mit einer Lösung nach einer Transformation korrespondiert.

## 6.2 Zusammenfassung

Im vorliegenden Kapitel wurde der Begriff der Stereotriangulation eingeführt und einige Verfahren zur Berechnung der Tiefeninformation eines 3D-Weltpunktes unter Zuhilfenahme der perspektivischen Projektionsmatrizen beider Kameras und der in Kapitel 5.2 berechneten Punktkorrespondenzen vorgestellt. Je nach gewähltem Korrespondenzanalyseverfahren (siehe Kapitel 5.2) ist es nun möglich eine Voll- oder Teilrekonstruktion eines Objektes beziehungsweise Szene durchzuführen.

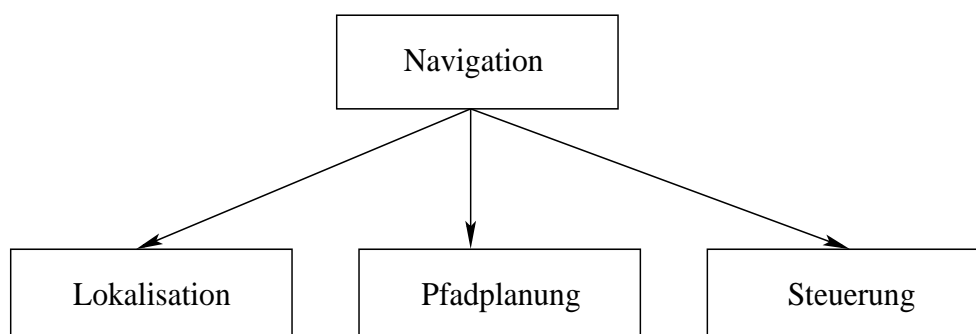
Es ist denkbar, die durch statische Stereoanalyse rekonstruierten Objekte beziehungsweise Szenen als Landmarken zu identifizieren. Die Fähigkeit Objekte beziehungsweise Szenen als Landmarken zu benutzen, spielt bei der Navigation des Menschen eine wichtige Rolle. Dies ist jedoch kein Bestandteil dieser Arbeit so, dass auf weiterführende Literatur verwiesen wird.

Im nächsten Kapitel wird die Selbstlokalisierung mobiler Roboter basierend auf visuelle Erkennung von künstlichen Landmarken behandelt.

# Erkennung und Selbstlokalisierung anhand von Landmarken

# 7

Damit ein mobiler Roboter in einem 3D Umfeld selbstständig operieren kann, benötigt er die Kenntnisse über seine Umgebung. Das Wissen über die jetzige Position, die Zielkoordinaten, die Berechnung des Pfades und Bewegung entlang der bestimmten Trajektorie wird in der Informatik unter dem Begriff der Navigation zusammengefasst [LDW91]. Die Struktur der Navigation wird in der Abbildung 7.1 grafisch dargestellt.



**Abbildung 7.1:** Grafische Darstellung der Navigationsstruktur eines mobilen Roboters, die aus drei Teilgebieten besteht. Bei der Lokalisierung wird die jetzige Position des Roboters bestimmt. Die Pfadplanung berechnet einen der möglichen Pfade ausgehend von Start- und Zielposition des Roboters. Die Steuerung überwacht die Bewegung entlang der kalkulierten Trajektorie.

In dieser Arbeit wird nur ein Teil der Navigation eines mobilen Roboters behandelt, die Selbstlokalisierung, auf die in diesem Kapitel theoretisch und in späteren Kapiteln auf deren Implementation und resultierende Ergebnisse eingegangen wird.

Das erste Teilgebiet der Navigation ist die Bestimmung der eigenen Position im Raum. Es existieren mehrere Methoden der Orientierung, wobei die meisten Vertreter dieses Teilgebietes auf der Sensortechnik basieren. Dabei wird die Lokalisierung in relative und absolute unterteilt. Bei der Absoluten wird die genaue Position an-

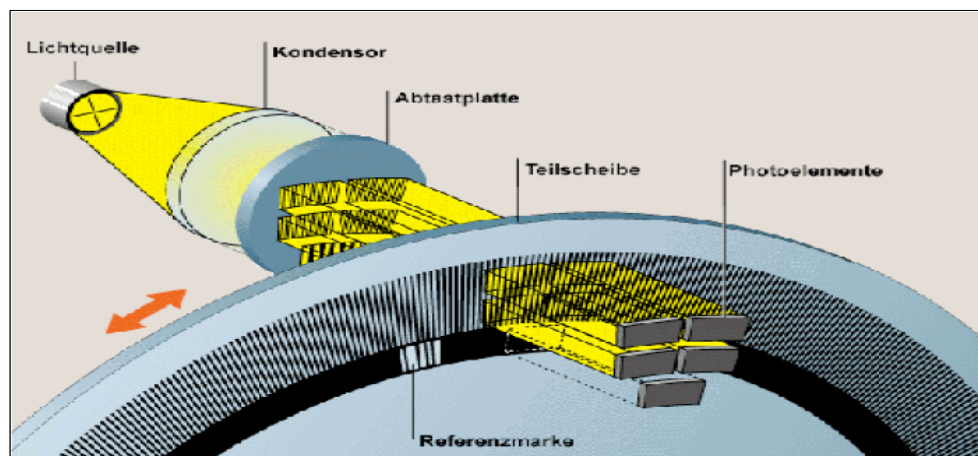
hand von Merkmalen berechnet, zum Beispiel die Koordinaten im Raum anhand von Landmarken. Bei der relativen Positionsbestimmung wird auf die letzte berechnete Position zurückgegriffen, um die neue Position zu bestimmen.

Im nächsten Abschnitt werden unterschiedliche Sensorarten vorgestellt. Basierend auf den Eigenschaften dieser Sensoren wird erläutert, wie die Position des Roboters bestimmt werden kann und welche Art der Lokalisierung möglich ist. Ein ausführlicher Überblick über unterschiedliche Sensoren und deren Einsatzgebiete wird in der [Zha05] gegeben.

### 7.1 Sensorik

In der entsprechenden Literatur werden häufig die vier bekanntesten Sensortechniken, die zum Zwecke der Orientierung eingesetzt werden, nämlich Odometrie, Ultraschallsensoren, Laserscanner und Videokameras vorgestellt [BEFW96]. Der hier folgender Überblick der meist eingesetzten Sensoren für die Navigation basiert auf der oben genannten Vorlesung *angewandte Sensorik*.

Die **Odometrie** ist ein sehr verbreitetes Verfahren der relativen Positionsbestimmung. Es wird als relative Positionsbestimmung bezeichnet, da die neue Position von der davor bestimmten Position abhängig ist.



**Abbildung 7.2:** Funktionsprinzip eines optischen Encoders [Sch05a]. Eine Maske bestehend aus transparenten und nicht-transparenten gleich großen Streifen. Dabei fällt ein Strahl auf die Maske und wird gegebenenfalls von dem dahinter liegenden Sensor empfangen, so ist es möglich die Zeit für eine Umdrehung oder eine Winkeländerung zu messen.

Dafür wird ein optischer Encoder eingesetzt, der auf einer Maske, die auf einer Plastik- oder Metall-Scheibe aufgedruckt ist, basiert. Die Maske besteht abwech-

sind aus transparenten und nicht-transparenten gleich großen Streifen. Dabei fällt ein Lichtstrahl auf die Maske und wird gegebenenfalls von dem dahinter liegenden Sensor empfangen. So ist es möglich eine Winkeländerung zu messen. In Kombination mit dem Wissen über angeschlossene Getriebe und Räder, sowie der benötigten Zeit kann die zurückgelegte Strecke bestimmt werden. Die Abbildung 7.2 stellt die Funktion eines optischen Encoders grafisch dar.

Das Prinzip des Verfahrens ist sehr einfach und ermöglicht die relative Positionsbestimmung des Roboters. Die neue Position wird durch die Addition der zurückgelegten Strecke und der alten Position errechnet. Auf der einfachen ebenen Fläche kann die neue Position des Roboters auf die folgende Art und Weise bestimmt werden

$$P_t = \begin{pmatrix} X_t \\ Y_t \\ \alpha \end{pmatrix} \implies P_{t+1} = \begin{pmatrix} X_t + s \cdot \cos(\alpha) \\ Y_t + s \cdot \sin(\alpha) \\ \alpha + \beta \end{pmatrix} \quad (7.1)$$

wobei  $P_t$  den Ausgangspunkt und  $P_{t+1}$  die Endposition eines Intervalls darstellen. Der Winkel  $\beta$  steht für die Richtungsänderung und  $s$  für die zurückgelegte Strecke.

Mit der Odometrie können vor allem auf kurzen Distanzen genaue Ergebnisse erreicht werden, jedoch summieren sich entstehende Fehler über die Zeit auf. Die häufigsten Fehler entstehen durch Unebenheiten im Boden, Unwucht oder unterschiedlicher Luftdruck der Räder. Deswegen wird die Odometrie des Öfteren in Verbindung mit einem oder mehreren anderen Verfahren der Positionsbestimmung eingesetzt.

Auch die **Ultraschallsensoren** werden häufig zur Entfernungsmessung und Lokalisierung eingesetzt, besonders wegen ihrer geringen Größe werden sie des Öfteren verwendet. Unter dem Begriff Ultraschall werden die Schallwellen mit einer Mindestfrequenz von 20 kHz verstanden. Die Ausbreitungsgeschwindigkeit hängt vom Medium und seiner Temperatur ab und beträgt in der normalen Luft circa  $331 \frac{m}{s}$ .

Es existieren mehrere Ansätze der Orientierung mit dem Ultraschall, die nahe liegende ist die Messung der Zeit zwischen einem und demselben Signal, das abgesendet, durch die Reflexion zurückgeschickt und schließlich empfangen wird. Da die Ausbreitungsgeschwindigkeit bekannt ist, kann auf die Entfernung zurückgeschlossen werden.

Aber auch umgekehrter Ansatz ist denkbar, mehrere sich im Raum befindliche Sender senden jeweils zwei Signale aus, ein Funk- und ein Ultraschallsignal. Da die Ausbreitungsgeschwindigkeit des Ultraschalls ca.  $10^6$  langsamer ist als beim Funk entsteht eine Zeitdifferenz. Dadurch, dass die Geschwindigkeit der beiden Signale bekannt ist und der zeitliche Unterschied zwischen zwei empfangenen Signalen gemessen werden kann, ist es möglich die Entfernung zum jeweiligen Sender zu bestimmen. Falls mehrere Sender oder Empfänger vorhanden sind, kann die Position im Raum bestimmt werden [Rot02].

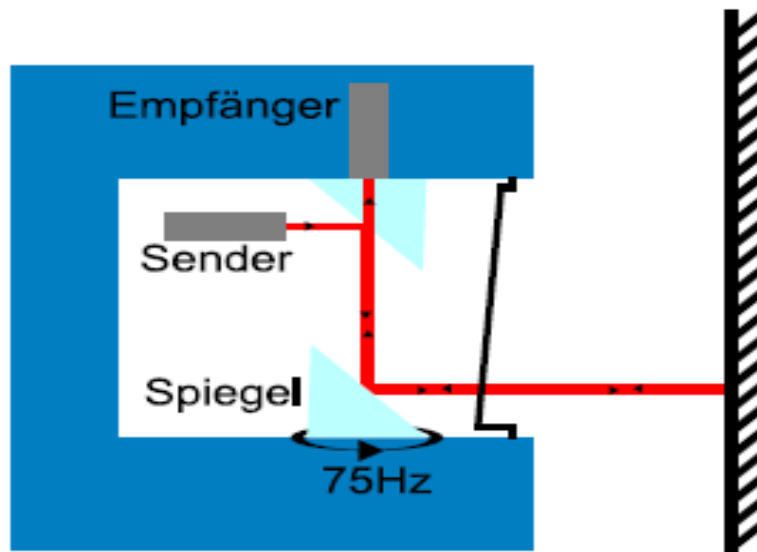
Ein anderer Ansatz ist die Präparation der Räume mit Ultraschallempfänger, so genannten *Ultrasonic Receiver Units*. Die meistens werden in der Decke einer Büroumgebung platziert und per Netzwerk miteinander verbunden. An dem Objekt, dessen Position bestimmt werden soll, wird einer oder mehrere *Bats* befestigt. Dabei handelt es sich um ein Gerät mit einer eindeutigen ID, das ein Funkempfänger und Ultraschallsender beinhaltet. Zur Steuerung des gesamten Systems werden Basisstationen eingesetzt. Diese senden ein Funksignal aus, der aus einer, der im Raum vorhandenen *Bat-ID* besteht. Der entsprechende *Bat* empfängt dieses Signal und sendet darauf ein Ultraschallsignal aus, das von den Sensoren in der Decke empfangen wird. Aus dem Unterschied der Ankunftszeiten kann die Position des Objektes bestimmt werden [War98]. Falls ein Objekt mit mehreren *Bats* ausgestattet wird, kann zusätzlich die Orientierung dessen eindeutig bestimmt werden. Das Verfahren liefert präzise Ergebnisse, die Installation- und Wartungskosten sind aber enorm.

Der Einsatz der Ultraschallsensoren, ausgenommen des *Bat*-Systems, ist kostengünstig und liefert auf kurze Distanzen präzise Ergebnisse. Beide Arten der oben beschriebenen Positionsbestimmung sind möglich. Bei längeren Entfernungen entstehen Schwierigkeiten, da nicht mehr angenommen werden kann, dass der Strahl nur geradeaus verläuft. Dennoch bleibt ein großer Vorteil gegenüber der Odometrie, dass die entstehende Fehler sich im Laufe der Zeit nicht summieren, sondern durch den Einsatz von entsprechenden Algorithmen sogar reduziert werden.

Das am häufigsten anzutreffendes Verfahren zur Entfernungsmessung und Lokalisierung greift auf die **Laserscansensoren** zurück, ein optischer Vorgang, bei dem die Umgebung berührungslos abgetastet wird, [BP06]. Der maximale Messbereich eines Laserscanners ist generell nicht beschränkt, bei dem in dem Arbeitsbereich TAMS verwendete LMS 200 vom SICK liegt dieser bei ca. 80 Metern. Zur Entfernungsmessung werden keine Reflektoren benötigt. Die Abbildung 7.3 stellt der Aufbau eines Laserkopfes grafisch dar.

Einige Geräte, wie zum Beispiel der Minolta 3D Laserscanner [Min07] zur Laserabtastung basieren auf dem Verfahren der Triangulation, dabei wird ein Laserstrahl erzeugt und durch ein, im Laserkopf eingebauten Spiegel reflektiert. Um die Triangulation zu ermöglichen, wird zwischen dem Laser und dem Sensor ein Abstand geschaffen. Der Prozess der Triangulation mit einem Laserscanner wird in der Abbildung 7.4 grafisch verdeutlicht. Der von den Objekten im Raum zurückreflektierter und von den Sensoren empfangener Laserstrahl schließt ein Dreieck, anhand dessen die Entfernung zu dem Objekt berechnet wird.

Aber auch so genannte *Time of Flight* Verfahren, die auf der Zeitdifferenz der Laserstrahlen basieren, werden des Öfteren verwendet. Der Aufbau des Laserkopfes ist mit dem in der Abbildung 7.3 identisch. Dabei wird ein Laserimpulse gesendet und zurückempfungen, nachdem er von den Objekten des Raumes reflektiert worden ist. Die Zeitdifferenz wird gemessen und anhand der bekannten Ausbreitungsgeschwindigkeit, kann die Entfernung zu dem Objekt berechnet werden, ähnlich dem Ultra-



**Abbildung 7.3:** Schematische Darstellung eines Laserkopfes. Dabei wird ein Laserstrahl erzeugt und durch ein, im Laserkopf eingebautes, Spiegel mit Zuhilfenahme der Reflexion gelenkt. Der von den Objekten im Raum zurückreflektierter Strahl wird von den Sensoren empfangen. Anschließend ist es möglich die Entfernung zu dem Objekt zu berechnen [BP06].

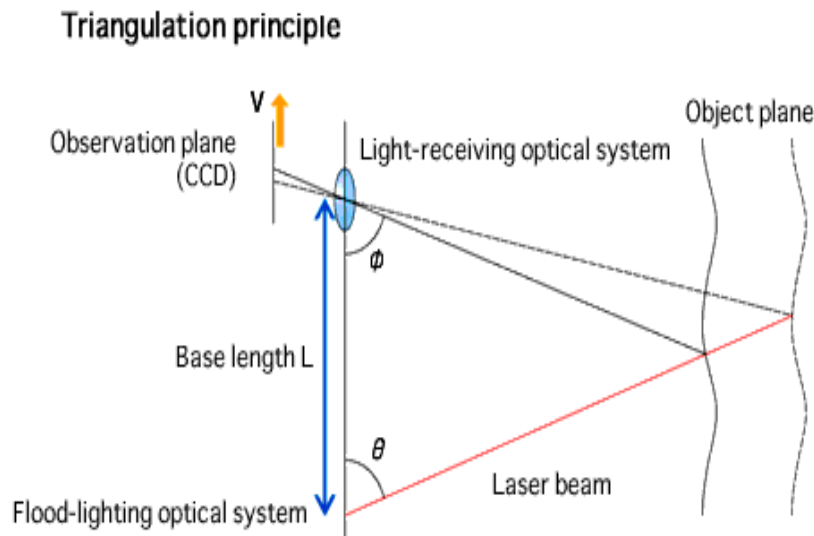
schall. In beiden Fällen kann der Winkel des Spiegels dynamisch verändert werden, so dass eine große Fläche schnell erfasst werden kann.

Auch die mehrdimensionale Raumerfassung mit dem Laserscanner ist möglich, dabei wird eine spezielle verstellbare Optik benötigt.

Die Vorteile der Laserscanner sind enorm, die Abtastung erfolgt schnell und stabil, die gelieferten Ergebnisse sind präzise. Das Verfahren ist beleuchtungsunabhängig und erzeugt eine 3D Struktur, wobei der Aufwand sehr gering ist. Es existieren aber auch einige Nachteile, so können einige Oberflächenstrukturen nicht erkannt werden, die die Strahlen absorbieren oder ablenken können, wie zum Beispiel Glas. Genau so wie beim Einsatz des Ultraschalls sind beide Arten der Positionsbestimmung möglich.

Da die beiden Hoap-Roboter, die im Kapitel 8 beschrieben sind, mit jeweils einem **Stereokamerasystem** ausgestattet und heutige Rechner in der Lage sind die von Kameras ankommende Menge an Informationen zu verarbeiten, werden Kameras in dieser Arbeit zur Lokalisierung eingesetzt. Über Funktion der Kameras, die Anordnung und anschließende Tiefenmessung kann in den entsprechenden Kapiteln der vorliegenden Arbeit nachgelesen werden.

Wie bei dem Laserscanner wird auch bei den Kameras zur Entfernungsmessung



**Abbildung 7.4:** Grafische Darstellung der Triangulation mit einem 3D Laserscanner von Minolta [Min07].

keine Landmarken benötigt. Dennoch ist die Bestimmung eigener Position mit einem Stereosystem in einem 3D-Raum ohne die Landmarken unabdingbar.

Im nächsten Abschnitt dieser Arbeit werden die Landmarken kategorisiert und beschrieben. Außerdem wird in diesem Kapitel auf die existenziellen Probleme beim Einsatz von Landmarken eingegangen, das Auffinden und Erkennen der Marken im Bild.

## 7.2 Landmarken

Es existieren mehrere Definitionen des Begriffes *Landmarke*, in der vorliegenden Arbeit wird die Definition nach [WKBM<sup>+</sup>97] verwendet, die wie folgt lautet:

*Landmarken sind distinkte, stationäre und saliente Objekte oder Reize, die als Referenzpunkte dienen können.*

Also ist eine Landmarke ein eindeutiges, stationäres und hervorstechendes Objekt oder Signal, das als Orientierungspunkt genutzt werden kann.

Nach [SW75] sind Landmarken notwendige und ausreichende Elemente, um das *Wegfinden* zu ermöglichen.



Um diese Voraussetzungen zu erfüllen, sollen die eingesetzten Landmarken bestimmte Kriterien erfüllen. Wie zum Beispiel der Kontrast zwischen der Landmarke und dem Hintergrund, eine bestimmte geometrische Form und das Beinhalten gewisser Informationen, wie zum Beispiel die Position im Raum. Bevor die Landmarke von den Robotern zur Navigation genutzt werden kann, sollte das Aussehen und die Form der eingesetzten Marken den Robotern bekannt sein [BEFW96].

In der Informatik wird zwischen natürlichen und künstlichen Landmarken unterschieden, die im Folgenden beschrieben werden.

### 7.2.1 Natürliche Landmarken

Die ältesten bekannten Orientierungsmerkmale sind die natürlichen Landmarken, wie der Name schon sagt, kommen diese im menschlichen Umfeld dauerhaft vor. Dabei ist die Bezeichnung *natürliche* eventuell irreführend, da natürliche Landmarken im engeren Sinne nichts mit der Natur zu tun haben. Die Bezeichnung soll eher darauf hinweisen, dass diese Landmarken schon vorhanden sind und es keine Veränderung der Umgebung vorgenommen werden soll. Meistens sind es herausragende geografische Objekte, wie zum Beispiel Berge, aber auch Dächer oder Türen und Kanten [Rie04], falls die Navigation in geschlossenen Räumen stattfindet.

Natürliche Landmarken werden in globale und lokale Landmarken unterteilt [EPK06]. Globale Landmarken sind aus großer Entfernung zu erkennen, somit kann die Orientierung an diesen sehr früh beginnen und lange Zeit andauern. Als Beispiel wären da die Sterne, Sonnenstand, Magnetfeldrichtung oder der Spitzenturm einer Kirche.

Lokale Landmarken sind nur aus kurzer Distanz zu sehen und dienen meistens dazu eigene Position zu bestimmen beziehungsweise zu verifizieren. Nach solchen Marken wird oft vor Verzweigungen gesucht, um Richtungsänderung eindeutig beschreiben zu können. Es ist leicht nachvollziehbar, dass die häufigsten Probleme bei der Erkennung von lokalen natürlichen Landmarken die Verdeckungen sind.

Zusätzlich wird hier darauf hingewiesen, dass die natürlichen, im Gegensatz zu den künstlichen, Landmarken immer passiv sind. Dabei bedeutet passiv, dass die Landmarken keine Handlungen ausüben, sondern von denen betroffen sind. Im Falle visueller Landmarken werden diese in Bildern wiedergefunden.

Die Orientierung an natürlichen Landmarken ist intuitiv und kommt sehr nah an das menschliche Verhalten. Dennoch wird eine Datenbank oder eine interne Karte benötigt. Das Verfahren kann nicht in dynamischer Umgebung eingesetzt werden, außer der Roboter soll die ihm unbekanntes Umgebung modellieren. Die Algorithmen sind oft sehr rechenintensiv, da das Umfeld meistens keine Strukturierungen

aufweist. Andererseits findet kein Eingriff in die Umgebung statt und die aufwendige Erstellung und Installation von Landmarken entfällt.

Für den Roboter ist es jedoch enorm schwer die natürlichen Landmarken aus den Bildern zu selektieren, mit seinen internen Daten zu vergleichen und anhand derer die eigene Position exakt zu bestimmen.

### 7.2.2 Künstliche Landmarken

Nach Borenstein sind künstliche Landmarken wie folgt definiert: *Künstliche Landmarken sind speziell von Menschen zum Zwecke der Lokalisierung in einer bestimmten Umgebung geschaffene Objekte [BEFW96].*

Da die Landmarken mit dem entsprechenden Kontrast zu dem Umfeld oder mit einer bestimmten geometrischen Form kreiert werden, gestaltet sich die Detektion der Landmarken in den Bildern einfacher und verlässlicher als bei natürlichen Landmarken [AH93]. Außerdem ist meistens nur wenig Rechenaufwand nötig.

Andererseits setzt die Lokalisierung anhand von künstlichen Landmarken voraus, dass die Umgebung bekannt ist und mit den Landmarken vor dem Einsatz des Roboters präpariert wird.

Dabei werden die wichtigsten Nachteile von künstlichen Landmarken sichtbar. Der Einsatz in unbekannter Umgebung ist nicht möglich und es entstehen die Kosten für die Erstellung, Aufstellung, Justierung und Wartung der Landmarken.

Wie schon im vorangegangenen Abschnitt angedeutet, unterteilen sich die künstlichen Landmarken in aktive und passive. Die aktiven übernehmen das Ausstrahlen vom Signal, das von dem Roboter empfangen und verarbeitet wird. Als Beispiel wären da die Leuchttürme, Ampeln, GPS oder Lichtquellen.

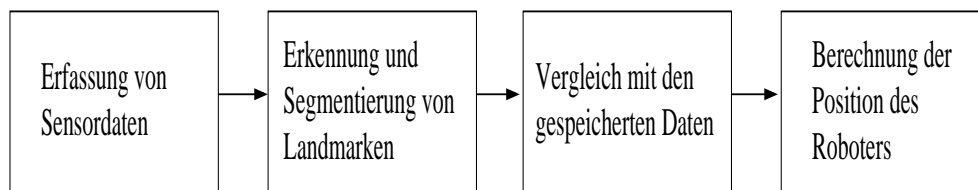
Die passiven Landmarken sind diejenigen, die das Signal reflektieren, zum Beispiel die Reflexionsstreifen oder in Bildern wiedergefunden werden, wie die Barcodestreifen.

In der vorliegenden Arbeit kamen passive künstliche Landmarken zum Einsatz, die mit einem Stereokamerasystem aufgenommen und später in den Bildern erkannt werden. Anhand der gefundenen Landmarken wird die aktuelle Position des Roboters unter Zuhilfenahme der Triangulation ermittelt, die in einem späteren Kapitel ausführlich behandelt wird.

Doch zuerst soll entschieden werden, welche Landmarken am besten erkannt und lokalisiert werden können. Dafür werden im Folgenden die gängigsten Vertreter von Landmarken und entsprechende Suchalgorithmen beschrieben und analysiert.

### 7.2.3 Erkennung und Segmentierung passiver künstlicher Landmarken

In den meisten bekannten Verfahren zur Selbstlokalisierung mit einem Stereokamerasystem werden einfarbige oder zweifarbige Landmarken eingesetzt. Besonderes schwarz-weiße Landmarken sind sehr verbreitet. Die Vorgehensweise bei der Lokalisierung bleibt aber weit gehend identisch und wird in der Abbildung 7.5 verdeutlicht [BEFW96].



**Abbildung 7.5:** Vorgang der Lokalisierung, der meistens in vier Schritte unterteilt wird. Zuerst werden die Sensordaten erfasst und ausgewertet. Anhand von gewonnenen Daten und weiteren Bildverarbeitungsschritten werden Landmarken erkannt und segmentiert. Danach folgt ein Abgleich mit den gespeicherten Informationen. Anschließend wird eine Hypothese über die Position des Roboters aufgestellt oder die Selbst-Lokalisierung unter Zuhilfenahme der Triangulation oder anderen Verfahren durchgeführt.

Zuerst werden die Sensordaten erfasst und ausgewertet, anhand derer und weiterer Bildverarbeitungsschritte werden Landmarken erkannt und segmentiert. Danach findet ein Abgleich mit den gespeicherten Daten statt, zum Beispiel Position der Marke im Weltkoordinatensystem oder Bezug der Originalgröße zur Größe im Bild. Anschließend wird eine Hypothese über die Position des Roboters aufgestellt oder, falls mehrere Landmarken im Bild zu sehen sind, die Selbst-Lokalisierung unter Zuhilfenahme der Triangulation oder anderen Verfahren durchgeführt.

Bei den meisten Verfahren, die farbige Landmarken zur Lokalisation nutzen, werden gezielt Farben ausgewählt, die in der Einsatzumgebung nicht vorkommen. Wie zum Beispiel bei Robocup, wo der Ball lila, Feldabgrenzung weiß und Tore gelb oder blau sind. Es existieren mehrere Möglichkeiten der Selbstlokalisierung, die im Folgenden vorgestellt werden.

In [RS88] und [RS90] wird das Template-Matching Verfahren beschrieben. Dabei werden die gesuchten Objekte aus mehreren möglichen Perspektiven aufgenommen und gespeichert. Das aus dem aufgenommenen Bild extrahierte Muster wird als eine Art Schablone genutzt und mit neuen aufgenommenen Bildern gefaltet. Dabei wird versucht das Objekt in dem neuen Bild zu lokalisieren. Dem Algorithmus werden zwei Matrizen übergeben, eine  $N \times N$  Matrix  $I[0, \dots, N-1, 0, \dots, N-1]$  des Bildes und eine  $M \times M$  Matrix  $T[0, \dots, M-1, 0, \dots, M-1]$  des gesuchten Musters.

$$C_{2D}[i, j] = \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} I[(i+u) \bmod N, (i+v) \bmod N] \cdot T[u, v] \quad (7.2)$$

Ausgegeben wird eine  $N \times N$  große Matrix  $C_{2D}$ , die aus der oberen Gleichung (7.2) resultiert. Mathematisch ist die Matrix  $C_{2D}$  eine zweidimensionale Faltung von  $I$  und  $T$ , wobei Werte der Matrix die Übereinstimmung des Bildes mit dem Muster darstellen.

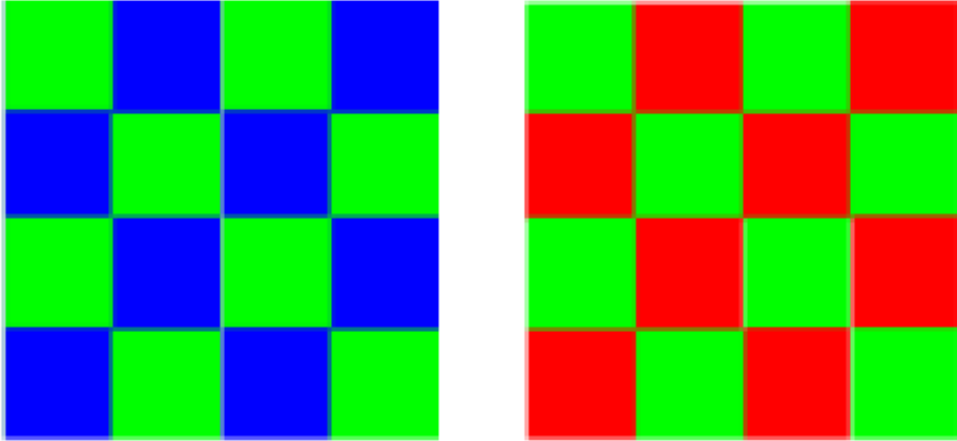
Als Nachteil des Template-Matching Verfahrens kann die immer anwachsende Menge an gespeicherten Daten aufgeführt werden, auch der Aufbau dieser Wissensbasis ist nicht einfach und muss strukturiert und verwaltet werden. Deswegen wird meistens ein Lernverfahren verwendet, um die Wissensbasis zu erstellen und zu vervollständigen.

Das Auffinden von Landmarken gestaltet sich am einfachsten, falls die Farbe oder Form in der Umgebung sonst nicht vorkommt und damit der Kontrast zu der Umgebung groß ist. In diesem Fall werden meistens die Schwellenwerte (engl. Thresholds) eingesetzt. Dabei wird mit dem Schwellenwert das Bild durchlaufen, wobei alle Werte unter der vorgegebenen Grenze auf null und alle anderen auf maximale Intensität der entsprechenden Farbe gesetzt werden. Später können nicht schwarze Bereiche noch einmal analysiert werden, sodass ein Objekt segmentiert werden kann. Für die Erkennung reicht in den meisten Fällen das Vorhandensein der gesuchten Farbe oder geometrischen Form im Bild.

Seit mehreren Jahren werden ähnliche Verfahren bei Robocup [VWL<sup>+</sup>00] eingesetzt, um zum Beispiel den Ball oder die Tore zu lokalisieren. Solche Algorithmen haben geringe Komplexität und können oft direkt auf dem Roboter ausgeführt werden. Falls die oben genannten Vorbedingungen erfüllt werden, liefern die Algorithmen schnelle und stabile Ergebnisse.

Der größte Nachteil solcher Algorithmen oder grundsätzlich aller Verfahren, bei denen farbliche Landmarken verwendet werden, sind unstabile Lichtverhältnisse [Web02]. Ändern sich diese, muss der Algorithmus angepasst werden. Des Öfteren wird das Problem damit umgangen, dass die Erkennung in geschlossenen Räumen unter konstanten Lichtverhältnissen stattfindet. Anderenfalls soll die Anpassung an die gegebenen Lichtverhältnisse vor jedem Einsatz des Algorithmus durchgeführt werden. Es existieren einige Veröffentlichungen, wo die Autoren durch die Benutzung unterschiedlicher Farbsystemen oder Trennung der einzelnen Farbkanäle das Problem zu lösen versuchen. Dennoch, wie die Experimente zeigen, bleibt eine große Abhängigkeit von Lichtverhältnissen bestehen, die meistens direkten Einfluss auf die Zuverlässigkeit und Schnelligkeit der Algorithmen ausübt.

Ein Verfahren in dem zweifarbiger Landmarken verwendet werden, ist in [YK01] präsentiert. Die Varianten der eingesetzten Farbmarken werden in der Abbildung 7.6 dargestellt.



**Abbildung 7.6:** Zwei unterschiedliche Varianten zweifarbiger Landmarken nach dem Verfahren von Yoon und Kweon [YK01]. Die Landmarken werden unter Zuhilfenahme von Histogrammen erkannt und segmentiert.

Um die Landmarke zu finden, werden mehrere Histogramme, über die Zeilen und Spalten des Bildes, gebildet und ausgewertet. Dabei wird in den Bildern nach der charakteristischen Farbverteilung gesucht, ohne die geometrische Form zu berücksichtigen. Die Resultate werden mit dem Landmarken-Modell verglichen, das dem Roboter bekannt sein muss. Anhand von gebildeten Ähnlichkeitswerten wird entschieden, ob die Landmarke im Bild vorhanden ist.

Zuerst wird der Ähnlichkeitswert zwischen dem abgetasteten Block  $i$  und dem Landmarken-Modell  $j$  wie folgt gebildet [YK01]:

$$M(i, j) = GHI(i, j) + LHI(i, j) \quad (7.3)$$

wobei GHI das globale Histogramm, die globale Farbverteilung und LHI das lokale Histogramm, die Farbverteilungseigenschaften der Landmarke darstellen und wie folgt bestimmt werden können

$$GHI(i, j) = \sum_r \sum_g [gh_i(r, g) \cap gh_{model\ j}(r, g)] \quad (7.4)$$

$$LHI(i, j) = \sum_r \sum_g \sum_w lh_i(w, r, g) + \sum_r \sum_g \sum_h lh_i(h, r, g) \cap lh_{model\ j}(h, r, g)$$

dabei stellen  $gh$  und  $lh$  globales beziehungsweise lokales Histogramm dar. Die Parameter  $r$  und  $g$  repräsentieren farbliche Histogramme im chromatischeren Farbraum, die wie folgt bestimmt werden

$$[r \ g] = \left[ \frac{R}{R + G + B} \quad \frac{G}{R + G + B} \right] \quad (7.5)$$

Um den Algorithmus effizienter zu gestalten und die Landmarken deduktiv, also ohne a priori Informationen über die Szene zu finden, wird das Bild in Blöcke unterteilt, die zufällig generiert und mit allen Landmarken-Modellen verglichen werden. Die Wahrscheinlichkeit einer Übereinstimmung zwischen dem Block  $i$  und der Landmarke  $j$  ist definiert als

$$prob(i | j) = \frac{M(i, j)}{\sum_{k=1}^N M(k, j)} \quad (7.6)$$

wobei  $N$  die Anzahl der Blöcke und  $M(i, j)$  der Ähnlichkeitswert aus der Gleichung (7.3) sind.

Die Gleichung (7.6) wird schrittweise ausgeführt und erlaubt anschließend die Landmarken exakt zu lokalisieren [YK01]. Um die schrittweise Ausführung zu ermöglichen sollen permanent neue Blöcke kalkuliert werden. Dafür wird zuerst die kumulative Wahrscheinlichkeitsverteilung der Ähnlichkeitswerte berechnet

$$c(i, j) = \sum_{k=1}^i prob(k | j) \quad (7.7)$$

Nachdem die Wahrscheinlichkeitsverteilung bestimmt ist, wird ein Zufallswert zwischen  $[0, 1]$  generiert mit dem ein abgetasteter Block  $n$ , mit der folgenden Formel, ausgewählt wird.

$$c(n, j) \leq \text{Zufallszahl} < c(n + 1, j) \quad (7.8)$$

Anhand vom Block  $n$ , der in der Gleichung (7.8) bestimmt wird, kann ein neuer Block mit den Koordinaten  $(x, y)$  erzeugt werden durch

$$\begin{bmatrix} x \\ y \end{bmatrix} = A \times \begin{bmatrix} x_n \\ y_n \end{bmatrix} + s \times b \quad (7.9)$$

Dabei sind  $(x_n, y_n)$  die Koordinaten des Blockes  $n$ , Matrix  $A$  präsentiert die dynamischen Kenndaten,  $s$  ist ein Skalierfaktor und  $b$  stellt den Vektor aus den vereinigten Zufallszahlen dar.

So wird die Effizienz des Algorithmus gesteigert, da Blöcke mit größerer Ähnlichkeit zu der Landmarke weiterhin eingesetzt werden. Blöcke mit kleinerer Ähnlichkeit werden verworfen.

Nachdem der vorherige Schritt abgeschlossen ist, werden Blöcke mit der größten Ähnlichkeit ausgewählt und die Landmarken endgültig lokalisiert, anschließend wird die Lage der Landmarke bestimmt, dabei stellen  $(x_k, y_k)$  die Koordinaten des Blocks mit dem Index  $k$  dar.

$$\begin{bmatrix} L_x^j \\ L_y^j \end{bmatrix} = \frac{1}{N} \sum_{k=1}^N \text{prob}(k | j) \times \begin{bmatrix} x_k \\ y_k \end{bmatrix} \quad (7.10)$$

Damit findet die exakte Lokalisierung der Landmarke in jedem Bild statt. Dabei wird aus aktueller Position der Landmarke im Bild die Lage der Marke im nächsten Bild vorhergesagt. Dadurch terminiert das Verfahren mit nur geringerer Anzahl an Wiederholungen.

Die Autoren stellen fest, dass symmetrische und sich wiederholende farbliche Muster gegenüber der Helligkeit und geometrischen Verzerrungen unempfindlich sind.

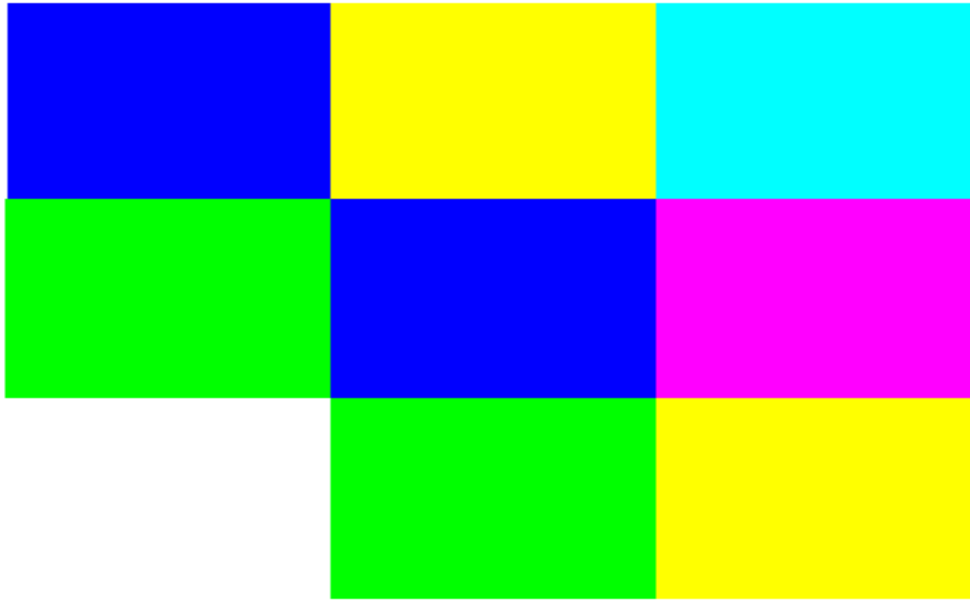
Außerdem zeigen sie experimentell, dass der beschriebene Algorithmus und das vorgestellte Landmarken-Modell sehr effektiv sind. Die Landmarken werden sogar in Bildern sicher erkannt, die viele Objekte unterschiedlicher Farben beinhalten.

An einem Projekt des Arbeitsbereiches TAMS der Universität Hamburg zusammen mit dem Prof. Y. F. Li von der Universität Hong Kong entstand eine Idee der Gewinnung der 3D-Daten eines Objektes unter Zuhilfenahme vom strukturierten Licht [CZL07]. Der Aufbau ähnelt dem Stereokamerasystem, wobei statt einer der beiden Kameras ein Projektor angeschlossen wird.

Der Projektor wirft strukturiertes farbkodiertes Licht auf ein Objekt, das mit der Kamera aufgenommen und von dem Rechner analysiert wird. Durch die farbliche Kodierung, ein eindeutiges Gitter aus farbunterschiedlichen gleich großen Rechtecken, wird die Tiefe und Struktur des Objektes rekonstruiert. Dieses Verfahren sollte vor allem in der Produktion eingesetzt werden, zum Beispiel für die Qualitätskontrolle.

Außerdem wird von den Autoren vorgeschlagen, Teile des oben genannten Gitters als Landmarken einzusetzen. Da das Gitter unterschiedlich groß gestaltet und unterteilt werden kann, ist auch das Aussehen der Landmarken flexibel.

Um die Eindeutigkeit und Anzahl der möglichen Landmarken zu bewahren, wird eine Struktur bestehend aus 9 Rechtecken als sinnvoll erachtet. Die Rechtecke werden in einer  $3 \times 3$  Matrix angeordnet und mit unterschiedlichen Farben versehen, wobei zwei gleiche Farben nicht nebeneinander positioniert werden dürfen und nicht alle eingesetzten Farben in der Landmarke vorhanden sein müssen. Die Erstellung der Landmarken basiert auf allen möglichen Permutationen von mehreren Farben, wobei keine der Marken mehr als einmal vorkommen darf und die oben genannten Bedingungen eingehalten werden sollen. Ein Beispiel einer der möglichen Landmarken wird in Abbildung 7.7 vorgestellt.



**Abbildung 7.7:** Ein Beispiel einer möglichen Landmarke, die aus Permutationen (hier 7 Farben) mehrerer Farben entsteht, wobei zwei gleiche Farben nicht neben einander positioniert werden können und nicht alle eingesetzten Farben in der Landmarke vorhanden sein müssen.

Der Vorteil der oben beschriebenen Landmarken liegt in ihrer Einzigartigkeit, so kann jeder Landmarke eine eindeutige Position im Raum zugeordnet werden. Beim Einsatz von 7 Farben können über 5 000 000 unterschiedliche Landmarken kreiert werden, was für die meisten Einsatzgebiete ausreichend ist.

Dennoch erfordert der Einsatz von mehrfarbigen Landmarken konstante Lichtverhältnisse oder die Helligkeitsanpassung der Kamera. Außerdem beinhalten die Landmarken kein wiederkehrendes Muster, so dass ein ausgefallener Suchalgorithmus entwickelt werden muss. Auch die Auffindung der Landmarke unter unterschiedlichen Blickwinkeln stellt kein triviales Problem dar.

Zum Zeitpunkt der Entstehung dieser Arbeit lag kein Algorithmus zur Auffindung dieser Landmarken vor. Deswegen wird in dieser Arbeit auf die Nutzung der oben genannten Landmarken verzichtet. Dennoch für bestimmte Einsatzgebiete und mit einem durchgedachten Suchalgorithmus könnten die Landmarken erfolgreich verwendet und ihre Vorteile vollständig ausgenutzt werden.

Die meisten Landmarken werden anhand von gesammelter Erfahrung, Beschaffenheit der Umgebung oder vorhandenen Hilfsprogrammen und Algorithmen entworfen. Dennoch existieren einige wenige Verfahren, die die Landmarke nach einem mathematischen Modell erstellen, das speziell für Zwecke der Detektion angepasst ist.

Im Folgenden werden das mathematische Modell und die resultierende Landmarke



nach Algorithmus von Daniel Scharnstein und Amy J. Briggs [SB99] und [SB01] vorgestellt und behandelt. Auch der oben beschriebene Algorithmus nach Yoon und Kweon greift auf dieses Verfahren zurück.

Das Ziel der Arbeit von Scharnstein und Briggs ist es die Landmarken zu konzipieren, die schnell und zuverlässig gefunden werden können. Außerdem soll der Algorithmus möglichst unempfindlich gegenüber Lichtverhältnissen, Orientierung und Entfernung der Landmarke sein.

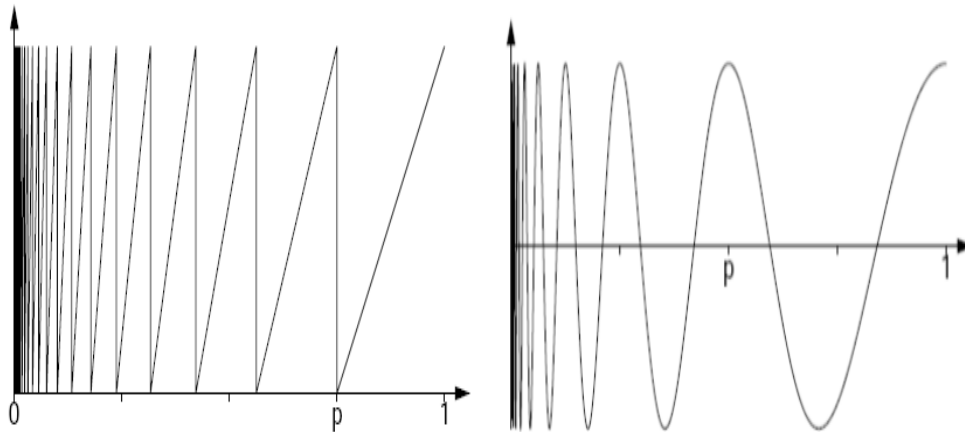
Dafür betrachten sie zuerst das größte Problem der Landmarkendetektion, die entstehende Verzerrungen bei der Abbildung eines Objektes der realen Welt in das Koordinatensystem der Kamera. Diese Verzerrungen gliedern sich in affine und perspektivische. Dabei werden die perspektivischen Verformungen von den Autoren nicht weiter betrachtet, um das Problem zu vereinfachen.

Die Hauptidee von Scharnstein und Briggs ist die Nutzung selbst-ähnlicher Muster als visuelle Landmarken. Dafür wird zuerst die selbst-ähnliche Funktion  $f: \mathbb{R}^+ \rightarrow \mathbb{R}$  definiert:

*Eine Funktion ist genau dann  $p$ -similar für alle festen, aber beliebigen Skalierungsfaktoren  $p$ ,  $0 < p < 1$ , falls*

$$\forall_{(x>0)} f(x) = f(px) \quad (7.11)$$

In der Abbildung 7.8 werden zwei Beispiele selbst-ähnlicher Funktionen grafisch dargestellt.



**Abbildung 7.8:** Grafische Darstellung zweier  $p$ -similarer Funktionen mit  $p = \frac{1}{2}$ .

Außerdem sind die Grafen der  $p$ -similaren Funktionen selbst-ähnlich, da der Vorgang mit dem Skalieren der Funktion mit dem Faktor  $p$  in die  $x$ -Richtung identisch ist. Damit ist die  $p$ -Similarität invariant bezüglich der Skalierung.

$$f(\alpha x) = f(p\alpha x) = f(\alpha px) \quad (7.12)$$

Die aus der Gleichung (7.12) resultierende Eigenschaft ist enorm wichtig für die Detektion von Mustern im Bild, da der Abstand zwischen dem Stereokamerasystem und der Landmarke nicht mehr von Bedeutung ist.

Damit die Detektion der Landmarke im Bild stattfinden kann, wird ein Ähnlichkeitsmaß benötigt, das angibt wie  $p$ -similar eine Funktion  $f$  zu sich selbst, skaliert um Faktor  $p$  ist. Der zu vergleichende Bereich kann durch die Breite  $w$  des Fensters eingeschränkt werden. Dann sieht die Formel für das Ähnlichkeitsmaß  $d$  folgendermaßen aus 7.13

$$d_{p,w}(f) = \frac{1}{w} \int_0^w |f(x) - f(px)| dx \quad (7.13)$$

Das Ähnlichkeitsmaß  $d_{p,w}(f)$  ist also ein Integral über die Differenz zwischen einer skalierten und unskalierten Funktion  $f$  auf dem Intervall  $w$ .

Nach der Definition 7.11 muss für die  $p$ -similare Funktion auf dem Intervall  $[0, w]$  das Ähnlichkeitsmaß  $d_{p,w}(f) = 0$  gelten. Damit wird deutlich, dass je näher das Ähnlichkeitsmaß  $d_{p,w}(f)$  an den Wert 1 kommt, desto weniger selbstähnlich sind die beiden zu untersuchenden Funktionen.

Dennoch reicht der Vergleich zweier Funktionen anhand des Ähnlichkeitsmaßes  $d_{p,w}(f)$  nicht aus, um ein charakteristisches Muster für die Landmarken zu erstellen. Problematisch sind dabei Bildbereiche mit konstanter Intensität, da sie mathematisch als Funktion mit einem konstanten Wert betrachtet werden und damit  $p$ -similar zu einander sind.

Zur Lösung dieses Problem setzen Scharnstein und Briggs eine weitere Bedingung, die erfüllt werden muss und für die konstanten Funktionen nicht gilt. Es wird gefordert, dass bei der Skalierung einer Funktion mit dem Faktor  $\sqrt{p}$  das Ähnlichkeitsmaß  $d_{\sqrt{p},w}(f)$  den maximalen Wert 1 erreicht.

Durch die Matching-Funktion  $m$  werden beide Bedingungen verbunden und ergeben folgendes Suchkriterium (Gleichung (7.14))

$$m_{p,w}(f) = d_{\sqrt{p},w}(f) - d_{p,w}(f) \quad (7.14)$$

Das neue Ähnlichkeitsmaß  $m_{p,w}(f)$  ist auf dem Intervall  $[-1, 1]$  definiert. Je mehr sich  $m_{p,w}(f)$  dem Wert 1 nähert, desto größer ist die Wahrscheinlichkeit, dass beide Funktionen  $p$ -similar sind und umgekehrt.

Als Nächstes suchen die Autoren eine Funktion, die das optimale Ergebnis für das Ähnlichkeitsmaß  $m_{p,w}(f) = 1$  liefert. Diese Funktion  $s_p(x)$  soll die Bedingungen der  $p$ -Similarität und  $\sqrt{p}$ -Antisimilarität erfüllen.

Dafür werden die Bedingungen für das Ähnlichkeitsmaß, die in der Gleichung (7.13) zusammengefasst sind, durch die oben eingeführte Bedingung erweitert und in beiden nachfolgenden Gleichungen (7.15) und (7.16) präsentiert.

$$\frac{1}{w} \int_0^w |f(x) - f(px)| dx = 0 \quad (7.15)$$

$$\frac{1}{w} \int_0^w |f(x) - f(\sqrt{p}x)| dx = 1 \quad (7.16)$$

Um die Funktion für die Generierung von Landmarken zu entwickeln, werden einfache Funktionen  $F$  mit  $\forall z F(z) = F(z + 1)$  betrachtet. Damit diese Funktionen genutzt werden können, wird Funktion  $f(x)$  wie folgt substituiert.

$$f(x) = F(\log_p x) = F(\ln(x)/\ln(p)) \quad (7.17)$$

Für die  $f(px)$  gilt dementsprechend die Gleichung (7.18)

$$\begin{aligned} f(px) &= F\left(\frac{\ln(px)}{\ln(p)}\right) = F\left(\frac{\ln(x) + \ln(p)}{\ln(p)}\right) \\ &= F\left(1 + \frac{\ln(x)}{\ln(p)}\right) = F\left(\frac{\ln(x)}{\ln(p)}\right) = f(x) \end{aligned} \quad (7.18)$$

Danach werden die Erkenntnisse der Gleichung (7.18) auf die Gleichungen (7.15) und (7.16) angewendet, die Grenzen des Intervalls  $w$  auf unendlich und die Gültigkeit auf alle möglichen  $x$ -Werte ausgeweitet. Dann kann die Normierung  $\frac{1}{w}$  und das Integral weggelassen werden, zusätzlich wird durch die oben beschriebene Funktion  $F$  substituiert und es entsteht für erste Bedingung folgende Gleichung

$$\begin{aligned} &|f(x) - f(\sqrt{p}x)| = 1 \\ \Rightarrow &|F(\log_p(x)) - F(\log_p(\sqrt{p} \cdot x))| = 1 \\ \Rightarrow &|F(\log_p x) - F(\log_p(\sqrt{p} \cdot x) + \log_p(x))| = 1 \\ \Rightarrow &|F(\log_p x) - F\left(\frac{1}{2} + \log_p(x)\right)| = 1 \end{aligned} \quad (7.19)$$

Da die Formel  $f(x)$  für alle  $x$  gilt und unter Zuhilfenahme der Substitution durch die Funktion  $S: \mathbb{R} \rightarrow [0, 1]$  dieses Verhalten sich nicht ändert, kann Gleichung (7.19) vereinfacht werden (Gleichung (7.20))

$$\forall x \quad |S(x) - S\left(\frac{1}{2} + x\right)| = 1 \quad (7.20)$$

Genauso wird mit der zweiten Bedingung verfahren und es entsteht folgende Gleichung (7.21)

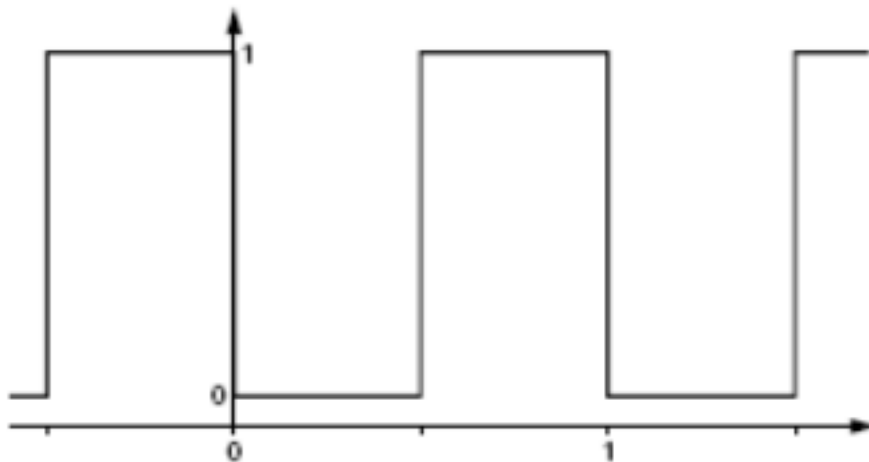
$$\begin{aligned} & | F(\log_p(x)) - F(\log_p(px)) | = 0 \\ \Rightarrow & | F(\log_p x) - F(\log_p(p) + \log_p(x)) | = 0 \\ \Rightarrow & | F(\log_p x) - F(1 + \log_p(x)) | = 0 \end{aligned} \quad (7.21)$$

Die resultierende vereinfachte Darstellung für die zweite Bedingung wird in der Gleichung (7.22) präsentiert.

$$\forall x \ S(x) = S(1 + x) \quad (7.22)$$

Aus den beiden Gleichungen (7.20 und 7.22) ergibt sich eine Funktion 7.23, die wie folgt aussieht

$$S(x) = \begin{cases} 0, & x - \lfloor x \rfloor < \frac{1}{2} \\ 1, & x - \lfloor x \rfloor \geq \frac{1}{2} \end{cases} = \lfloor 2(x - \lfloor x \rfloor) \rfloor \quad (7.23)$$



**Abbildung 7.9:** Die grafische Darstellung der Funktion  $S(x)$ .

In der Abbildung 7.9 wird die Funktion  $S(x)$  grafisch dargestellt, eine rechteckige periodische Funktion. Dennoch existiert eine Menge von Funktionen, die die Bedingungen der Gleichungen (7.20) und (7.22) erfüllen.

Die erste Bedingung, vergleiche Gleichung (7.20), erlaubt eine freie Wahl der Funktion  $S(x)$  auf einem Intervall  $[0, 1)$ , danach hat die Funktion einen festen Verlauf.

Die zweite Bedingung, die in der Gleichung (7.22) mathematisch dargestellt wird, verkürzt sogar das oben erwähnte Intervall auf  $[0, \frac{1}{2})$  und fordert zusätzlich, dass die Funktionswerte im Bereich zwischen 0 und 1 liegen.

Beide Bedingungen zusammengefasst ergeben eine Menge von Funktionen des Typs  $g : [0, \frac{1}{2}) \rightarrow \{0, 1\}$ , die alle genutzt werden können, um die Landmarken zu kreieren. Dabei entscheiden sich Scharnstein und Briggs und wählen die einfachste aller möglichen Funktionen, nämlich  $g(x) = 0$  aus.

Zurücktransformiert ergibt sich folgende Genirierungsfunktion  $s_p(x)$  (Gleichung (7.24))

$$s_p(x) = S(\log_p(x)) = \lfloor 2(\log_p(x) - \lfloor \log_p(x) \rfloor) \rfloor \quad (7.24)$$

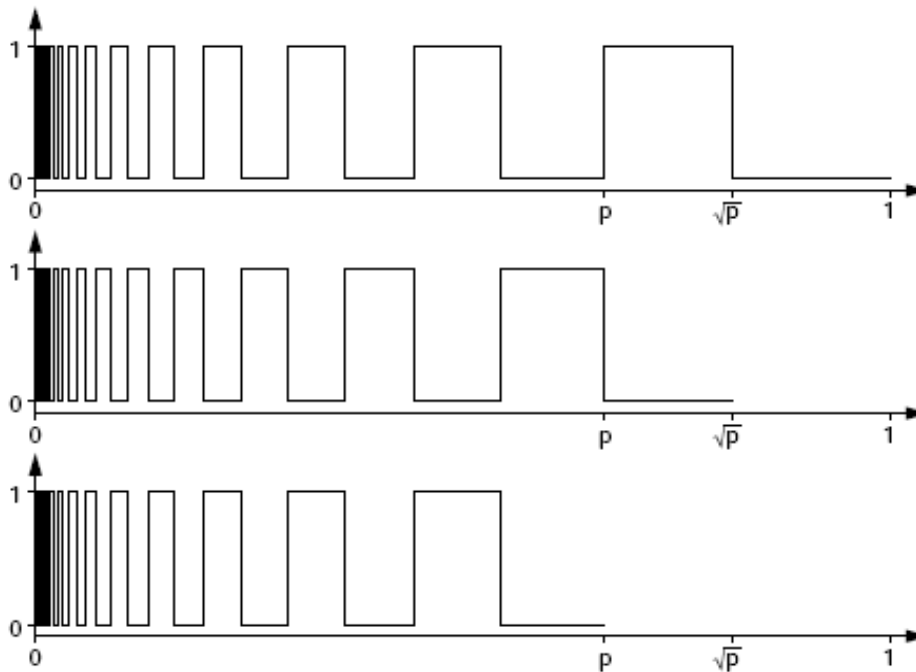
Wenn die Funktion  $s_p(x)$ , die in der Gleichung (7.24) dargestellt ist, betrachtet wird, so gilt für alle möglichen Intervalle  $w$   $d_{p,w}(s_p) = 0$  und  $d_{\sqrt{p},w}(s_p) = 1$ . Damit ist die Herleitung einer geeigneten Generatorfunktion für die Erstellung von Landmarken abgeschlossen. Wie von den Autoren gewünscht ist die resultierende Funktion  $s_p(x)$  p-ähnlich und unähnlich im Bezug auf  $\sqrt{p}$ , dieses wird zusätzlich grafisch in der Abbildung 7.10 verdeutlicht. Die erste Funktion von oben ist die resultierende Generatorfunktion  $s_p(x)$  mit  $p = \frac{2}{3}$ , in der Mitte wird die maximale Unähnlichkeit und unten die maximale Ähnlichkeit dargestellt.

Experimentell wird von den Autoren gezeigt, dass beim Parameter  $p = \frac{2}{3}$  die besten Resultate erzielt werden. Sodass die größte Unabhängigkeit gegenüber dem Kontrast und durch die affinen Transformationen entstehenden Verformungen erreicht wird. Somit kann schnelle Auffindung der Landmarken im Bild von dem Algorithmus gewährleistet werden. Die resultierende Landmarke, die durch die Generatorfunktion  $s_p(x)$  mit  $p = \frac{2}{3}$  erzeugt wird, ist in der Abbildung 7.11 dargestellt.

Abschließend wird der Algorithmus zur Detektion der Landmarken an die realen Bedingungen angepasst. Dabei werden die Kontraständerung, aliasing Artefakte und Sensorrauschen mitberücksichtigt, außerdem wird der Wert des Ähnlichkeitsmaßes  $m_{p,w}(f)$  angepasst. Die Anpassung des Ähnlichkeitsmaßes ist notwendig, da der Wert ein absolutes Maximum darstellt und unter oben aufgeführten Bedingungen der Realität nicht erreicht werden kann.

Der Algorithmus funktioniert dementsprechend folgendermaßen aus. Es werden entlang der Bildzeilen die Bildspalten nacheinander durchlaufen und das Ähnlichkeitsmaß  $m_{p,w}(f)$ , gemäß Gleichung (7.14), gebildet. Statt der Integration wird schrittweise aufsummiert, da dabei nicht ganzzahlige Koordinaten auftreten können, wird mit den benachbarten Pixeln bilinear interpoliert. Für die Detektion der Landmarke erfolgt die Überprüfung des Ähnlichkeitsmaßes  $m_{p,w}(f)$  in zwei Schritten

$$m_p(x) > c_{min} \quad (7.25)$$



**Abbildung 7.10:** Oben ist die Generatorfunktion  $s_p(x)$  mit  $p = \frac{2}{3}$ . In der Mitte und unten wird jeweils maximale Unähnlichkeit beziehungsweise Ähnlichkeit dargestellt.

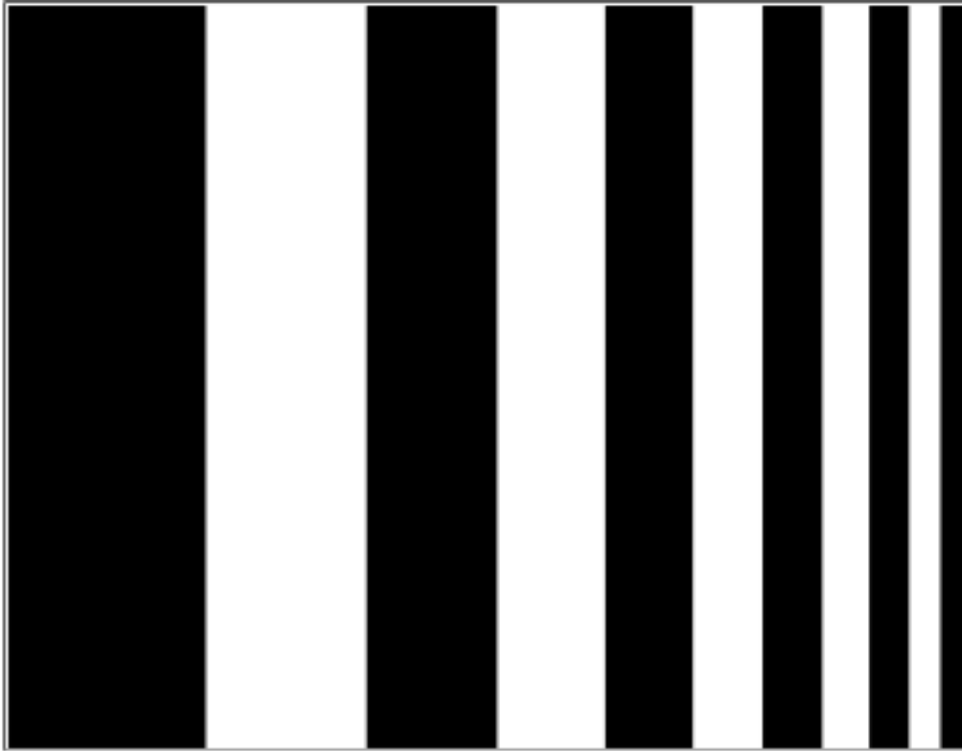
$$m_p(x + \delta) < \frac{c_{min}}{2} \quad (7.26)$$

Dabei stellt  $c_{min}$  einen absoluten Schwellenwert dar, über dem das Ähnlichkeitsmaß für die erfolgreiche Detektion liegen soll. Dieses wird durch die Gleichung (7.25) dargestellt und ermöglicht die Auffindung der Landmarken auch unter den oben aufgezählten und in der Realität herrschenden Konditionen.

Mit der zweiten Bedingung, Gleichung (7.26), wird versucht Detektion von p-similar ähnlichen oder zufälligen Mustern im Bild, die keine Landmarken sind, auszuschließen. Dabei werden die Ränder, um die Stelle, an der erste Bedingung wahr ist, betrachtet, falls es nicht wie bei dem Original der Landmarke einen starken Abfall der Funktionswerte existiert, wird das Muster verworfen und das Bild weiterdurchsucht.

Sind beide Bedingungen erfüllt, gilt der entsprechende Pixel als Anfang einer Landmarke, sonst wird zurückgegeben, dass die Landmarke im Bild nicht gefunden werden konnte.

Die Komplexität des Algorithmus soll bei  $O(\frac{nw}{k})$  liegen, wobei  $n$  die Anzahl der Pixel des Bildes,  $w$  das Suchintervall und  $k$  Anzahl der durchsuchten Spalten darstellt. Damit läuft der Algorithmus nur einmal beziehungsweise konstant oft über das Bild durch um die Landmarken zu detektieren.



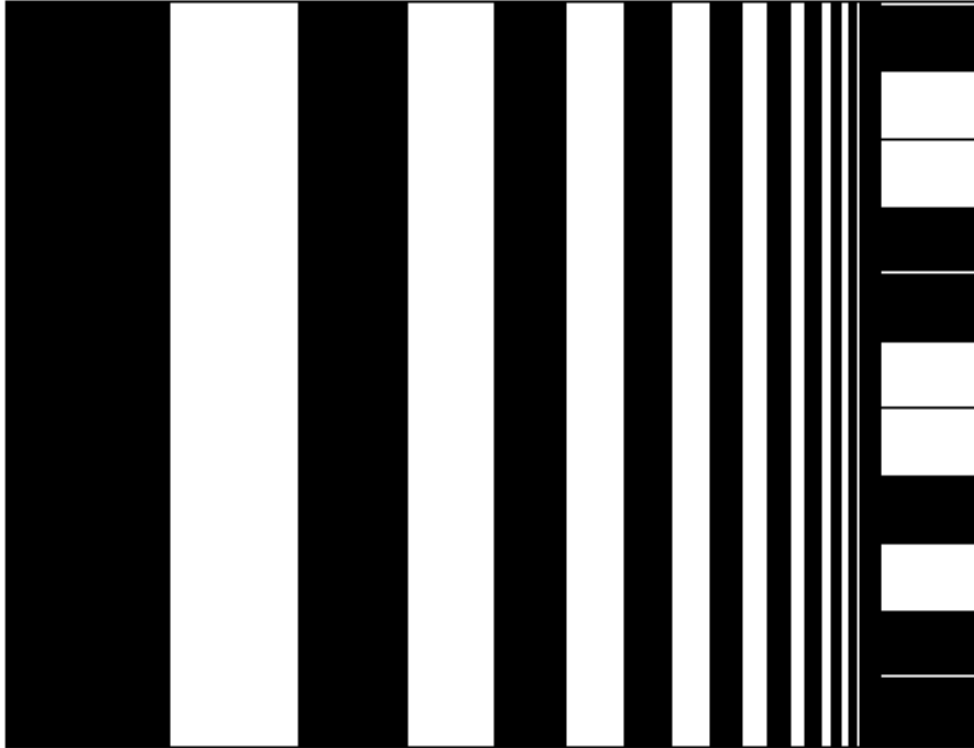
**Abbildung 7.11:** Die durch Generatorfunktion  $s_p(x)$  mit  $p = \frac{2}{3}$  erzeugte Landmarke.

Ein Vorteil des Algorithmus von Scharnstein und Briggs ist die Suche nach einem eindeutigen Muster im Bild, dennoch können dadurch die gefundenen Landmarken von einander nicht unterschieden werden. Um diesen Nachteil entgegen zu wirken, wird die Landmarke in der Veröffentlichung [SB01] erweitert. Die Landmarke wird auf der rechten Seite durch einen einfachen binären Barcode, der in der Abbildung 7.12 zu sehen ist, ergänzt.

Der binäre Barcode besteht aus 11 Blöcken, die jeweils ein Bit repräsentieren. Werden alle Blöcke zur Kodierung der Identität der Landmarke verwendet, so können 2048 verschiedene Landmarken kreiert werden. In der Realität werden aber einige davon als Paritätsbits genutzt, um mögliche entstehende Fehler zu korrigieren.

Um den Barcode auszulesen, wird zuerst die Landmarke im Bild detektiert und ihre Ausrichtung bestimmt. Ist die Ausrichtung und Position bekannt, kann die Lage jeden einzelnen Streifen bestimmt werden. Damit ist die Position des Barcodes bekannt und er kann problemlos ausgelesen werden, was die von den Autoren durchgeführte Experimente bestätigen.

Damit erschaffen Scharnstein und Briggs eindeutige Landmarken, die auf einem mathematischen Modell basieren und speziell für die Detektion der Landmarken im Bild optimiert sind. Der Algorithmus ist robust, schnell und zuverlässig. Die Komplexität



**Abbildung 7.12:** Die durch Generatorfunktion  $s_p(x)$  mit  $p = \frac{2}{3}$  erzeugte und durch der Barcode ergänzte Landmarke. Die Erweiterung der Landmarke erlaubt eine eindeutige Zuordnung dieser zu dem Hintergrundwissen, wie zum Beispiel den Koordinaten im Raum.

des Algorithmus ist konstant, zusätzlich durch den Einsatz zweier komplementärer Farben wird die maximale Kontrastunabhängigkeit erreicht und der Einfluss unterschiedlicher Beleuchtung stark minimiert. Durch die zuletzt vorgestellten Anpassungen, Gleichungen (7.25) und (7.26), ist der Detektionsalgorithmus weit gehend umgebungstolerant. Auch die Anzahl der eindeutigen Landmarken, die kreiert werden können, ist für die meisten Einsatzgebiete ausreichend.

Da der Algorithmus und die Landmarken für in dieser Arbeit gestellte Aufgaben, der Orientierung und Selbst-Lokalisierung eines Roboters, am geeignetsten erschienen und die Komplexität gegenüber anderen existierenden Detektionsalgorithmen gering ist, wird die Landmarke und dazugehöriger Detektionsalgorithmus in dieser Arbeit verwendet. Zusätzlich erweisen sich die Landmarken gegenüber dem Kontrast und durch die affinen Transformationen entstehenden Verformungen als robust, was die Detektion verbessert, dennoch die Komplexität konstant hält. Außerdem ist keine Veränderung oder Anpassung des Umfeldes nötig, da das Verfahren nach Experimenten, die von Scharnstein und Briggs durchgeführt wurden, zuverlässige Ergebnisse in beliebiger Umgebung und bei unterschiedlichen Lichtverhältnissen liefert.



Im Kapitel *Experimentelle Ergebnisse* werden, die in der vorliegenden Arbeit gewonnenen Erkenntnisse vorgestellt und erörtert. Zusätzlich werden die Effizienz und Verlässlichkeit des Algorithmus von Scharnstein und Briggs getestet und kritisch beurteilt.

### **7.3 Zusammenfassung**

In diesem Kapitel wurden zuerst unterschiedliche Sensortechniken und daraufstützende Lokalisierungsverfahren theoretisch vorgestellt und erläutert sowie die jeweiligen Vor- und Nachteile.

Basierend auf vorhandener Hardware, im Falle dieser Arbeit eines Stereokamerasystems, wurde für die Lokalisierung unter Zuhilfenahme von passiven künstlichen Landmarken entschieden. Daraufhin wurden unterschiedliche Arten von Landmarken und entsprechende Detektionsverfahren dargestellt und beschrieben.

Durch gewonnene Erkenntnisse wurden die, auf einem mathematischen Modell basierende und durch eine Generatorfunktion kreierten, Landmarken und dazugehöriger Detektionsalgorithmus von Scharnstein und Briggs ausgewählt, der in vorliegender Arbeit zum Zwecke der Orientierung und Lokalisation des Roboters im Weltkoordinatensystem verwendet wird.



# Ausgangssystem

# 8

---

In diesem Kapitel werden, die zur Verfügung stehende und in dieser Arbeit eingesetzte Hardware- und Softwarekomponente beschrieben.

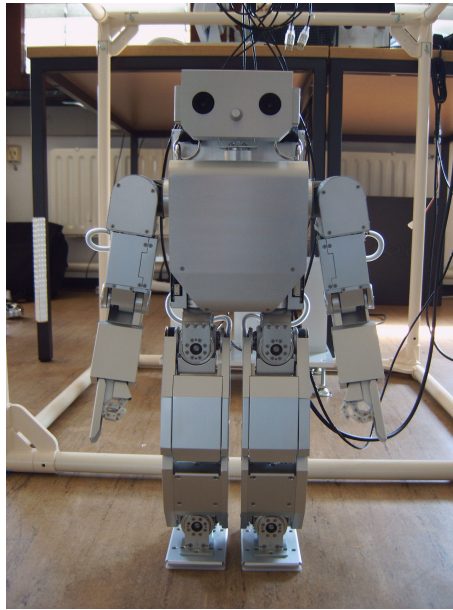
Diese Arbeit entstand am Arbeitsbereich TAMS, Technische Aspekte Multimodaler Systeme, der Universität Hamburg in der Fakultät für Mathematik, Informatik und Naturwissenschaften. Die gesamte hier beschriebene Hardwarekomponente ist Eigentum des oben genannten Arbeitsbereiches und wird für die Dauer der Diplomarbeit den Schreibenden zur Verfügung gestellt.

Das erste Teil dieses Kapitels beschäftigt sich mit den Hardwarekomponenten, die eingesetzt werden. Im zweiten Teil wird auf die genutzte Software eingegangen, dabei wird der Abschnitt in zwei Unterabschnitte geteilt. Der erste Unterabschnitt beschreibt die Programmierumgebung und Werkzeuge, die die Darstellung von Ergebnissen erleichtern. Im zweiten Unterabschnitt wird die Open Source Computer Vision Library beschrieben, da einige Algorithmen dieser Bibliothek in der vorliegenden Arbeit verwendet werden.

## 8.1 Hardware

Die gesamten Experimente fanden auf zwei humanoiden Robotern statt, den HOAP-2 der Fujitsu Automation Co. Ltd. Die Roboter sind jeweils 50 cm groß und wiegen 7 kg. Die Abbildung 8.1 zeigt einen der beiden HOAP-2 Robotern des Arbeitsbereiches TAMS.

Der Roboter besitzt fünfundzwanzig Freiheitsgrade, vier in jedem Arm und 6 in jedem Fuß. Außerdem 2 Gelenke für die Bewegung des Kopfes, ein im Körper und ein in jeder Hand. Die gesamten Gelenke des HOAP-2 werden in der Abbildung 8.2 grafisch dargestellt. Der Roboter ist mit einem Beschleunigungssensor und einem Gyrosensor ausgestattet, die jeweils drei Richtungen erfassen. Außerdem befinden sich in jedem Bein vier Drucksensoren, mithilfe derer auf die Stabilität der Lage des Roboters zurückgeschlossen werden kann.



**Abbildung 8.1:** Darstellung einer der beiden HOAP 2 Robotern des Arbeitsbereiches TAMS.

Die für diese Arbeit wichtigsten Sensoren sind die beiden Kameras, die im Kopf des Roboters in einem Abstand von 6 cm eingebaut sind. Bei den beiden Kameras handelt es sich um die Logitech Quickcams<sup>®</sup>, die jeweils ein  $\frac{1}{4}$  inch CMOS Sensor besitzen und nicht synchronisiert sind. Die maximale Auflösung beträgt  $324 \times 248$  Pixel und es können 16.77 Millionen Farben (24 Bit) dargestellt werden. Die Änderung des Fokus kann nur manuell erfolgen, über ein Drehrad um das Objektiv der Kameras. Der Datenaustausch findet über USB statt, dafür wird jede Kamera über jeweils ein USB Kabel direkt an den gleichen oder einen externen Rechner angeschlossen. Dabei soll ein anderer Controller, als der für die Befehlsübertragung, genutzt werden, da sonst der Roboter nicht bewegt werden kann.

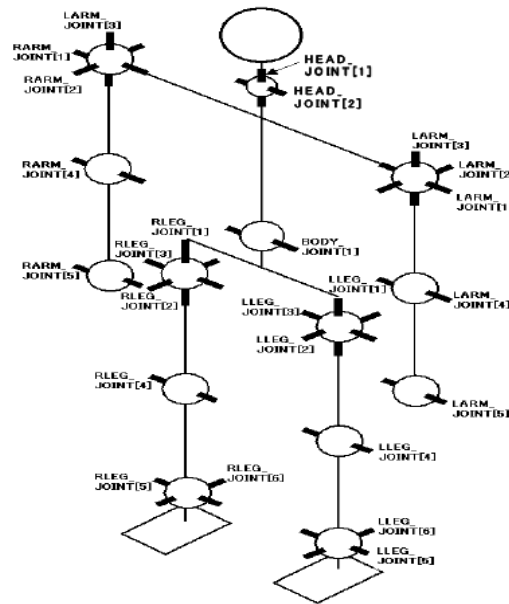
Der Roboter ist mit einem Geode GX1 Prozessor ausgestattet [FAC03], vergleichbar mit einem Pentium<sup>®1</sup> III CPU mit 700 MHz.

Damit stellt die Fujitsu den Entwicklern mehrere Möglichkeiten bereit. In Abhängigkeit von der Rechenintensität der Algorithmen kann die Ausführung direkt auf dem eingebetteten System des Roboters oder auf einem externen Rechner geschehen. Aber auch eine Verteilung der Algorithmen auf mehrere Systeme ist denkbar.

Die Architektur des Hoap 2 erlaubt ebenfalls verschiedene Herangehensweisen. So stellt jedes einzelne Gelenk des Roboters ein USB Endgerät dar, die einzelne USB

---

<sup>1</sup>Pentium ist ein eingetragener Name der Intel Corporation



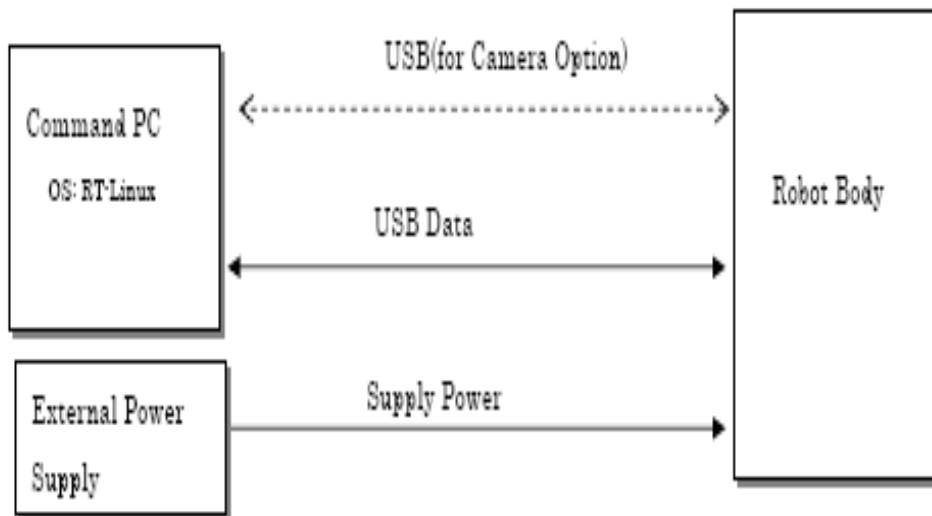
**Abbildung 8.2:** Grafische Darstellung der gesamten Gelenke und deren Freiheitsgrade des HOAP 2 Roboters.

Geräte werden durch mehrere USB Hubs zusammengeführt. Die miteinander verbundenen Hubs bieten die Schnittstelle nach außen, über die die Befehle an den Roboter übergeben und Daten abgefragt werden können. Dabei können die einzelnen USB-Geräte, sowie mehrere Geräte gleichzeitig angesteuert werden. Es kommt der USB 1.1 Standard zum Einsatz, der die Geschwindigkeit bis zu 12 Mbps erlaubt.

Um die Echtzeitausführung zu garantieren, läuft die Kommunikation über RT-Linux [FSM07] 3.2 pre 1, eine Erweiterung für Linux zur Steuerung der Echtzeitanwendungen [FAC03]. Das System übernimmt unter Linux die Prioritätenverteilung und garantiert damit, dass die ankommenden Befehle höher bewertet werden als eigene Programme.

Die gesamte Entwicklung findet auf einem oder mehreren externen Rechnern, auf denen RT-Linux läuft, statt. Die Kommunikation zwischen den externen Rechnern und dem Roboter läuft über die oben beschriebene USB Schnittstelle, siehe Abbildung 8.3. Dabei wird jedes USB-Gerät einzeln adressiert. Damit kann die Leistung des Systems optimiert werden, da die eigentliche Verarbeitung von eingehenden und ausgehenden Daten nicht auf dem Roboter stattfindet.

Für die Generierung und Übergabe der Befehle an den Roboter wurden vom Fujitsu zwei FMV Esprimo<sup>TM</sup> PCs mit den beiden HOAPs mitgeliefert. Zwei Pentium<sup>TM</sup> 4 @ 2.4 GHz Systeme mit mehreren USB Controller und jeweils 256 MB Ram, auf denen Red Hat Linux 7.3 mit 2.4 Kernel und RT-Linux 3.2 pre 1 laufen, um die Echtzeitausführung zu gewährleisten.



**Abbildung 8.3:** Veranschaulichung der Kommunikation zwischen dem externen Rechner und dem Roboter. Auf dem Rechner läuft das RT-Linux, um die Echtzeitausführung der Befehle zu garantieren. Der Datenaustausch, kabelgebunden oder per Funk, findet über die USB-Schnittstelle statt.

Die, für die Bildbearbeitung, eingesetzten Rechner sind zwei Optiplex 745 der Firma Dell<sup>®</sup> mit jeweils 1 GB Ram und einem Intel<sup>®</sup>Core<sup>™</sup>2 CPU @ 2.13 GHz und 2048 kB Cache. Für die grafischen Anwendungen sind die beiden Rechner mit jeweils einer ATI Radeon X1300 Grafikkarte mit 256 MB Speicher ausgestattet.

Als Betriebssystem kam OpenSuse Linux 10.2 mit dem Kernel 2.6 zum Einsatz. Somit stand genügend Leistung zur Verfügung, nicht nur um die gesamte Bildverarbeitung auf den externen Rechnern auszuführen und den Prozessor des Roboters damit zu entlasten, sondern sogar um auf den Bildsequenzen zu operieren.

## 8.2 Software

Wie schon im vorangegangenen Abschnitt erwähnt, wird Linux als Betriebssystem genutzt, also liefen alle im Nachfolgenden aufgezählten Programme und Algorithmen unter Linux.

### 8.2.1 Eingesetzte Software

Die gesamte Implementation fand in der Programmiersprache C/C++, unter Nutzung der frei verfügbaren Programmierumgebung Eclipse in der Version 3.3.0 und in das Framework eingebundenen Bibliothek CDT, statt.

Die grafische Darstellung der Ergebnisdaten wird mit den Programmen Gnuplot und OpenGL realisiert. Für die Kamerakalibrierung werden Camera Calibration Toolbox für Matlab<sup>®</sup> und OpenCV<sup>®</sup>-Bibliothek eingesetzt. Außerdem wird die OpenCV<sup>®</sup>-Bibliothek in der Version 1.0, auf die im nächsten Abschnitt näher eingegangen wird, zur Bildbearbeitung verwendet.

### 8.2.2 Open Computer Vision Library

Vollständig heißt die Bibliothek Intel<sup>®</sup> Open Source Computer Vision Library [Ope07]. Wie der Name schon sagt, wird der größte Anteil an der Entwicklung und Pflege der OpenCV-Bibliothek vom Intel Research Lab übernommen. Die erste Version erschien im Jahre 2000, seit dem 6. November 2006 ist die Version 1.0 verfügbar, die auch in der vorliegenden Arbeit zum Einsatz kam.

Die Bibliothek basiert auf dem so genannten IPP, Intel<sup>®</sup> Integrated Performance Primitives [IPP07], und ist speziell für den Befehlssatz der Intel<sup>®</sup> Prozessoren optimiert. Dadurch erreichen die Algorithmen und Funktionen der OpenCV-Bibliothek beim Einsatz auf einem Computersystem mit einem Intel<sup>®</sup>-Prozessor eine deutlich bessere Performance.

Die Bibliothek ist frei verfügbar, unterliegt der BSD-Lizenz<sup>2</sup> und besteht aus einer Ansammlung von Algorithmen und Beispielprogrammen. OpenCV-Bibliothek nutzt IPL, Image Processing Library, die von Intel<sup>®</sup> um einige komplexe Algorithmen und Funktionen erweitert wurde.

Der Verbreitungsgrad der OpenCV-Bibliothek ist sehr hoch. Sie läuft unter Linux und Windows und besteht aus vier Teilen, die als einzelne Bibliotheken eingebunden werden können:

- Grundlegende CV-Funktionen [cv.h]
- Experimentelle CV-Funktionen [cvaux.h]
- GUI-Werkzeuge [highgui.h]
- Basisstrukturen und Operationen [cxcore.h]

---

<sup>2</sup><http://www.intel.com/technology/computing/opencv/license.htm>, letzter Aufruf 22. 10. 2007

Die Bibliothek ist für die Programmiersprachen ANSI C/C++ geschrieben und beinhaltet über 300 Funktionen, sowie über 350 Bildverarbeitungsalgorithmen und stellt eine offene Infrastruktur bereit. Einsatzgebiete sind vor allem die Echtzeitanwendungen, Mensch-Maschine Schnittstelle [HCI] und Robotik.

Die Hauptbestandteile der OpenCV sind die Bewegungsanalyse, Objektverfolgung, so genanntes Tracking und statische Gestenerkennung. Des Weiteren stehen die Algorithmen für Bild- und Strukturanalyse, Erkennung von Konturen, Ecken und Kanten sowie mehrere sehr schnelle Filter zur Verfügung. Außerdem sind einige Verfahren zur Gesichts- und Objekterkennung, 3D-Rekonstruktion, Kamerakalibrierung und einige GUI-Werkzeuge implementiert.

Die OpenCV Bibliothek besitzt leider keine klare Struktur, die von Intel® bereitgestellte Dokumentation ist knapp gehalten und erlaubt keinesfalls einen schnellen problemlosen Einstieg [Her04]. Sodass die Einarbeitung einige Zeit in Anspruch nimmt und eine intensive Internetrecherche erfordert. Dennoch finden sich zunehmend immer mehr kompetente Diskussionsforen im Internet in denen sich viele Fragen beantworten und Probleme lösen lassen.

In dieser Arbeit werden nur einige wenige Funktionen der OpenCV-Bibliothek genutzt, an den entsprechenden Stellen wird darauf verwiesen und die Funktionen entsprechender Algorithmen ausführlich behandelt.

Im nächsten Kapitel 9 werden die eingesetzten Algorithmen und Funktionen sowie in der vorliegenden Arbeit erzielte Resultate dargestellt und beschrieben.



# Experimentelle Ergebnisse

# 9

---

Ziel dieser Arbeit ist die Selbstlokalisierung des humanoiden Roboters in der komplexen dreidimensionalen Umgebung. Die Realisierung dieses Vorhabens kann in mehrere Schritte unterteilt werden. Zuerst wird die Tiefeninformation des Raumes unter Verwendung der Daten des Stereokamerasystems rekonstruiert. In dem rekonstruierten Raum werden die künstlichen Landmarken detektiert und anschließend extrahiert. Durch das Stereokamerasystem und die bekannten Koordinaten der Landmarken wird die Position des Roboters unter Zuhilfenahme der Triangulation bestimmt.

In diesem Kapitel werden die eingesetzten Anwendungen, sowie die resultierende Ergebnisse vorgestellt und erörtert. Dabei werden, wie in der Formulierung des Ziels, einzelne Schritte nacheinander behandelt. In diesem Zusammenhang werden benutzte frei verfügbare sowie selbst implementierte Programme beschrieben und entsprechend gekennzeichnet.

Das Resümee der vorliegenden Arbeit, Bewertung und Diskussion der Ergebnisse, eventuelle Verbesserungsvorschläge, sowie Erweiterungs- und Ergänzungsmöglichkeiten werden im folgenden Kapitel *Fazit* ausführlich behandelt.

Es wird versucht weit gehend auf die Darstellung des Quelltextes zu verzichten, dennoch um einige Sachverhalte und Algorithmen zu verdeutlichen werden einige Programmausschnitte abgebildet. Der gesamte Quellcode wird dieser Arbeit in Form einer CD beigelegt.

## 9.1 Kamerakalibrierung

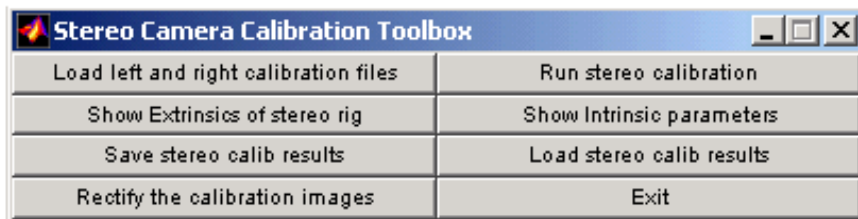
Wie schon im Kapitel 8 beschrieben, ist der humanoide Roboter HOAP 2 mit einem Stereokamerasystem ausgestattet. Die beiden Logitech Quickcams<sup>®</sup> Kameras besitzen jeweils ein  $\frac{1}{4}$ -inch-CMOS Sensor und sind nicht synchronisiert.

In dieser Arbeit werden zwei unterschiedliche Anwendungen zur Kamerakalibrierung der Kameras eingesetzt, Camera Calibration Toolbox für Matlab® und Algorithmen der OpenCV®-Bibliothek, die auch im Kapitel *Ausgangssystem* erwähnt werden.

Im Folgenden werden beide Verfahren vorgestellt und erläutert, danach werden die Ergebnisse präsentiert und miteinander verglichen. Der Abschnitt wird mit der Beurteilung der beiden vorgestellten Verfahren abgeschlossen.

### 9.1.1 Mit Camera Calibration Toolbox für Matlab

Die Camera Calibration Toolbox von Jean-Yves Bouguet und weiteren Autoren [BSS<sup>+</sup>07] ist eine Erweiterung für Matlab® in den Versionen 5.x, 6.x oder 7.x. Dabei erlaubt die Toolbox Kalibrierung einer Kamera, so wie auch eines Stereokamerasystems.



**Abbildung 9.1:** Die beiden grafischen Oberflächen der Kalibrierung des Camera Calibration Toolbox für Matlab. Oben wird die GUI für die Kalibrierung einzelner Kamera, unten eines Stereokamerasystems dargestellt.

Die Kalibrierung eines Stereokamerasystem setzt voraus, dass die Parameter der einzelnen Kameras schon bekannt sind. Aus den bekannten Parametern in Form jeweils einer Kalibrierungsdatei wird die Kalibrierung eines Stereosystems durchgeführt.

Die Toolbox bietet eine intuitive Nutzung des Programms durch eine grafische Benutzeroberfläche. In der Abbildung 9.1 werden die GUIs für Kalibrierung einer Kamera und eines Stereokamerasystems dargestellt.

Als Kalibrierungsmuster wird ein planares Schachbrettmuster, bestehend aus Aneinanderreihung weißer und schwarzer Quadrate, eingesetzt. Das, in dieser Arbeit verwendete, Kalibrierungskörper ist in der Abbildung 3.5 dargestellt. Das Schachbrettmuster ist zweidimensional mit einer Feldgröße von  $30 \times 30$  mm und erfüllt die Voraussetzungen der nicht Symmetrie, die für die korrekte Detektion der Eckpunkte notwendig ist.

Die Anwendung bestimmt die Linsenverzerrung, wie auch intrinsische und extrinsische Parameter, die schon im Kapitel 3 theoretisch behandelt werden. Im Folgenden wird die Bestimmung einzelner Parameter beschrieben, anschließend werden aus dem Verfahren resultierende Ergebnisse vorgestellt und erläutert.

### **Bestimmung intrinsischer Parameter**

Für die Bestimmung der intrinsischen Parameter wird ein halbautomatisches Verfahren genutzt, das auf den Arbeiten von Zhengyou Zhang [Zha00] basiert.

Zur Kalibrierung werden mehrere Bilder des Kalibrierungskörpers benötigt, die aus unterschiedlichen Perspektiven aufgenommen sind [BSS<sup>+</sup>07]. Dabei ist nicht nur die Entfernung der Kamera zu dem Kalibrierungskörper, sondern Bilder, die möglichst viele verschiedene affine und perspektivische Verzerrungen beinhalten vom Interesse.

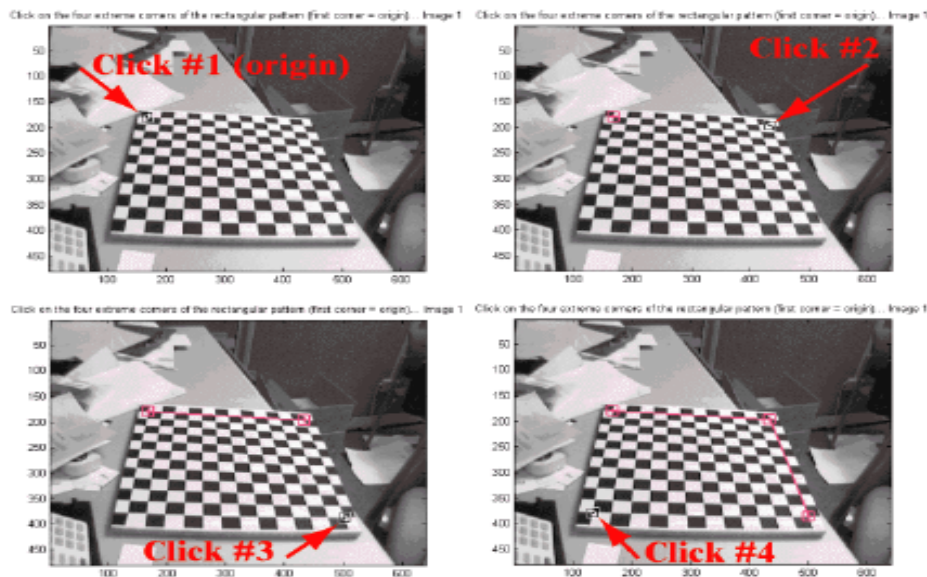
Die Bilder werden unter Zuhilfenahme der grafischen Oberfläche geladen, anschließend wird der Bereich mit dem Schachbrettmuster abhängig von dem Referenzpunkt bestimmt. Das Muster umrahmendes Quadrat wird von dem Benutzer in jedem einzelnen Bild per Hand gesetzt, diesem Umstand verdankt die Kamerakalibrierung der Toolbox die Bezeichnung halbautomatisch. Die Auswahl der Eckpunkte vor allem die richtige Reihenfolge wird in der Abbildung 9.2 veranschaulicht. Zusätzlich soll die Größe eines Schachbrettfeldes angegeben werden.

Nachdem das Muster in jedem Bild markiert ist, wird ein Kantendetektor zur Ermittlung innerer Knoten des Kalibrierungskörpers eingesetzt. Dieser Schritt erfolgt automatisch und orientiert sich an Berührungspunkten einzelner Felder basierend auf starker Kontraständerung des Übergangs zwischen zwei Feldern.

Falls alle Eckpunkte bestimmt werden können, werden die resultierende intrinsische Parameter durch das, in der Gleichung (3.30) dargestellte, Levenberg-Marquard Verfahren [Mor77] optimiert.

Dabei werden die detektierten Muster mit der bekannten Geometrie des Kalibrierungskörpers verglichen und die Abweichungen bestimmt. Dabei werden für die linke und rechte Kamera folgende intrinsische Parameter bestimmt

Die Brennweite  $f_c$  der beiden Kameras definiert jeweils den Abstand zwischen der Projektionsfläche, die Fläche des Kamerasensors und des Bildhauptpunktes, der



**Abbildung 9.2:** Auswahl der Eckpunkte und Veranschaulichung der richtigen Reihenfolge in Camera Calibration Toolbox für Matlab.

im Folgenden beschrieben wird. Die Brennweite  $f_c$  wird in Form eines 2D-Vektors  $[f_{c_x}, f_{c_y}]^T$  in Pixel angegeben.

In dem Bildhauptpunkt  $cc$  schneidet sich die optische Achse und die Bildebene. Bei dem, in der Abbildung 2.1 dargestellten Lochkamera-Modell liegt der Hauptpunkt an der optischen Achse, die senkrecht zu der Aufnahmeebene verläuft. Die Linsenverzeichnung, die in dem Abschnitt 3.4 beschrieben ist, verursacht die Verschiebung aller Bildpunkte, so dass der Hauptpunkt im Bezug zu dem optischen Zentrum versetzt wird. Ähnlich der Brennweite werden auch die Koordinaten des Hauptpunktes  $cc$  in Form eines 2D-Vektors  $[cc_x, cc_y]^T$  angegeben.

In dem  $5 \times 1$  großen Vektor  $k_c$  werden die Parameter der tangentialen und radialen Verzerrung des Bildes, die aus dem Kapitel 3.4 bekannt und in der Abbildung 3.3 grafisch dargestellt sind, zusammengefasst.

Nachdem die erste Kalibrierung abgeschlossen ist, kann die automatische Korrektur der detektierten Eckpunkte durchgeführt werden. Anschließend wird die Kamerakalibrierung noch einmal durchgeführt, im Normalfall werden danach die Kalibrierungsergebnisse deutlich verbessert. Zusätzlich stellt die Camera Calibration Toolbox für Matlab eine große Menge an Werkzeugen, um die Ergebnisse zu verbessern, wie zum Beispiel die neue Bearbeitung der Bilder mit der stärksten Abweichung. Weitere Möglichkeiten können direkt auf der Seite nachgelesen oder bei den Autoren erfragt werden [BSS<sup>+</sup>07].

In der Abbildung 9.3 ist der Versatzkoeffizient  $alpha_c$  nicht aufgeführt, da in der

<b>LINKE KAMERA</b>	<b>RECHTE KAMERA</b>
<code>fc = [473.02547; 477.66246]</code>	<code>fc = [471.47928; 475.94103]</code>
<code>fc_error = [1.98586; 2.20707]</code>	<code>fc_error = [2.20955; 2.42586]</code>
<code>cc = [175.64435; 138.86013]</code>	<code>cc = [148.95859; 136.92801]</code>
<code>cc_error = [0.9427927; 1.091550]</code>	<code>cc_error = [1.03019; 1.35150]</code>
<code>kc = [0.8498; -3.4915; 0.0007; -0.0029; 0.0000]</code>	<code>kc = [0.80397; -3.14618; 0.0058; -0.0042; 0.0000]</code>
<code>kc_error = [0.0206; 0.1489; 0.0019; 0.0015; 0.0000]</code>	<code>kc_error = [0.0234; 0.1613; 0.0023; 0.0017; 0.0000]</code>

**Abbildung 9.3:** Tabellarische Auflistung intrinsischer Parameter, bestimmt unter Zuhilfenahme der Kalibrierung nach Zhang mit Camera Calibration Toolbox für Matlab.

Camera Calibration Toolbox für Matlab davon ausgegangen wird, dass Sensorpixel idealerweise rechteckig sind. Damit wird  $alpha_c = 0$  angenommen.

Die restlichen Koeffizienten mit der Erweiterung `_error` im Namen stellen die Standardabweichung des jeweiligen Parameters dar. Somit sind intrinsische Parameter bestimmt und können für weitere Bildverarbeitungsschritte, wie zum Beispiel die Rektifikation, verwendet werden.

### Kalibrierung eines Stereokamerasystems

Die Camera Calibration Toolbox bietet auch die Möglichkeit ein Stereokamerasystem zu kalibrieren. Dafür werden zuerst die beiden Kameras einzeln kalibriert, da die Kalibrierungsdateien für die Kalibrierung des Stereosystems verwendet werden. Außerdem werden mehrere Bilder einer und derselben Szene aus unterschiedlichen Perspektiven der jeweiligen Kamera benötigt.

Nachdem die Kalibrierungsdaten der einzelnen Kameras und mehrere Bilder einer und derselben Szene, von beiden Kameras aufgenommen, geladen sind, kann die Stereokalibrierung durchgeführt werden. Aus dem Verfahren resultieren intrinsische Parameter für die linke und rechte Kamera, die auf den Ergebnissen der Einzelkamerakalibrierung basieren und durch das Programm, unter Zuhilfenahme neuer Bilder, optimiert werden. Eine weitere Optimierung der Kalibrierungsergebnisse ist durch die Zunahme von den Bildern der Einzelkamerakalibrierungen möglich und wird in dieser Arbeit angewendet. Die Ergebnisse der Stereokamerakalibrierung werden in der Abbildung 9.4 veranschaulicht.

Wie anhand der Standardabweichung ersichtlich wird, werden die Ergebnisse durch

<b>LINKE KAMERA</b>	<b>RECHTE KAMERA</b>
$fc = [473.30880; 477.98227]$	$fc = [471.78; 251.475.89065]$
$fc\_error = [1.62919; 1.276281]$	$fc\_error = [1.80269; 1.93757]$
$cc = [175.16522; 138.49399]$	$cc = [149.497465; 136.47167]$
$cc\_error = [0.676122; 0.89074]$	$cc\_error = [0.83237; 1.11019]$
$kc = [0.8401; -3.4899; 0.0003; -0.0038; 0.0000]$	$kc = [0.7869; -3.0803; 0.0049; -0.0035; 0.0000]$
$kc\_error = [0.0166; 0.1168; 0.0015; 0.0011; 0.0000]$	$kc\_error = [0.0185; 0.1302; 0.0019; 0.0013; 0.0000]$

**Abbildung 9.4:** Ergebnisse der Stereokamerakalibrierung. Es entstehen zwei separate Auflistungen der intrinsischen Parameter der beiden Kameras. Die Ergebnisse basieren auf den Resultaten der Einzelkamerakalibrierung, die durch neue Stereobilder verbessert werden. Zusätzlich wird in der vorliegenden Arbeit durch die Zusammenfügung der neuen Ergebnisse mit den Bildern der Einzelkamerakalibrierung eine weitere Optimierung durchgeführt.

die Stereokalibrierung und Zusammenfassung mit den Ergebnissen der Einzelkalibrierung verbessert.

Matlab bietet die Möglichkeit die Ergebnisse für die Weiterverarbeitung zu speichern. Zusätzlich liefert die Stereokalibrierung die extrinsischen Parameter des Stereosystems, auf die im nächsten Abschnitt näher eingegangen wird.

### Bestimmung extrinsischer Parameter

Wie im Kapitel *Extrinsische Parameter* 3.3 dargestellt, beschreiben die extrinsischen Parameter die Lage der Kamera bezüglich eines globalen Koordinatensystems [FFH<sup>+</sup>92]. Damit fallen die extrinsischen Parameter in Abhängigkeit von der Lage der Kamera zu den Kalibrierungsbildern unterschiedlich aus. Die Parameter bestehen aus Rotation und Translation der Kamera.

Dagegen sind die extrinsischen Parameter eines Stereokamerasystems fest und charakterisieren die Position einer Kamera bezüglich der anderen. In dieser Arbeit beschreiben die extrinsischen Parameter die Position der rechten Kamera in Bezug auf die linke und bestehen aus zwei  $3 \times 1$  großen Vektoren, dem Rotationsvektor  $om$  und der Transformation  $T$ .

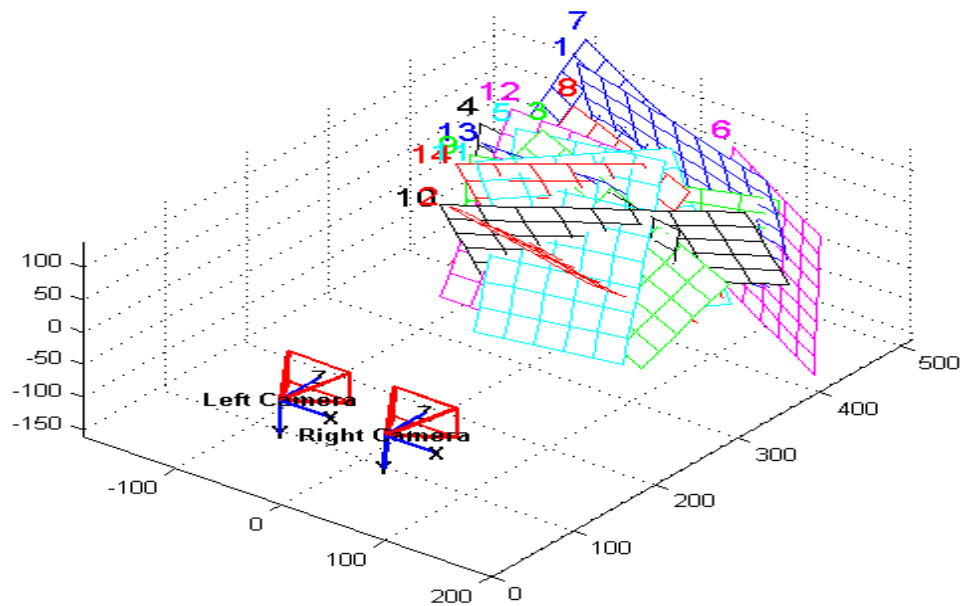
Die Camera Calibration Toolbox bietet im Anschluss der Stereokamerakalibrierung die Möglichkeit die extrinsischen Parameter des Stereosystems zu bestimmen, die in

der Gleichung (9.1) dargestellt werden.

$$om = [0.00356; 0.01012; 0.00716] \quad (9.1)$$

$$T = [-61.28621; 2.55919; -3.54608]$$

Beide Oberflächen der Kamerakalibrierung des Camera Calibration Toolbox für Matlab, die in der Abbildung 9.1 dargestellt sind, bieten jeweils eine Funktion *Show Extrinsic* beziehungsweise *Show Extrinsic of stereo rig*, die die Positionen einzelnen Kalibrierungsobjekte, sowie die Lage der Kameras anzeigt. Dabei wird bei der Stereokalibrierung nicht nur die Position der Kamera zu den Kalibrierungsobjekten, sondern auch die Lage der Kameras zueinander, vergleiche Abbildung 9.5, dargestellt.



**Abbildung 9.5:** Extrinsische Parameter der Kamerakalibrierung. Dargestellt wird die Position der Kamera relativ zu Kalibrierungsobjekten sowie die Lage der Kameras zueinander.

Die Kamerakalibrierung des Stereosystems ist damit abgeschlossen. Dennoch ist einer der Ziele dieser Arbeit eine zusammenhängende Anwendung, die die Position des Roboters bestimmt. Um die Effizienz des Programms zu steigern und die Ressourcen zu schonen werden im Folgenden die Bestimmung der Parameter dargestellt, sodass die Camera Calibration Toolbox für Matlab nur für Kamerakalibrierung eingesetzt und damit, als Extraanwendung benutzt wird. Damit ist die eigentliche Anwendung,

die die Landmarken in Bildern detektiert und die anschließende Position des Roboters bestimmt, vom Matlab unabhängig, sodass das Programm wenig Speicherplatz beansprucht und direkte Installation auf dem Roboter ermöglicht.

Mit den Kalibrierungsergebnissen, die von dem Programm in der Datei *Calib\_Result\_stereo.mat* gespeichert werden, ist es möglich die Bilder direkt im Matlab zu rektifizieren. Der Prozess der Rektifikation wird in dem Kapitel *Statische Stereoanalyse* der vorliegenden Arbeit ausführlich behandelt.

Die Camera Calibration Toolbox rektifiziert, die zur Kalibrierung, geladene Bilder und speichert diese automatisch. Dennoch falls die bestimmten Parameter für die weitere Bildverarbeitung nötig sind, wie zum Beispiel Rektifizierung neuer Bilder ohne Zugriff auf die Toolbox für Matlab, sollen die Rektifikationsmatrizen berechnet werden.

Unter Zuhilfenahme der Rotationsformel für Vektoren von Olinde Rodrigues kann aus dem  $3 \times 1$  Rotationsvektor  $om$  die  $3 \times 3$  Rotationsmatrix  $R$  bestimmt werden [Fau95]. Dabei stellt Matrix  $R$ , die aus der Gleichung (3.3) bekannte Rotationsmatrix dar und beschreibt die Rotation bei der Abbildung eines 3D-Punktes des Raumes auf einen 2D-Punkt im Kamerakoordinatensystem.

Die Camera Calibration Toolbox für Matlab stellt die Funktion, die die Formel von Rodrigues implementiert, standardmäßig zur Verfügung. Angewendet auf den Rotationsvektor  $om$  ergibt sich, für in vorliegender Arbeit eingesetzte Kameras, folgende Matrix 9.2

$$R = \begin{pmatrix} 0.9999 & -0.0071 & 0.0101 \\ 0.0072 & 1.0000 & -0.0035 \\ -0.0101 & 0.0036 & 0.9999 \end{pmatrix} \quad (9.2)$$

Durch die Kenntnisse der Abbildung eines Punktes des 3D-Raumes in das Kamerakoordinatensystem und in die Bildebene, kann die Abbildung zwischen den Weltpunkten und Punkten der Bildebene durch die Gleichung (3.4) mathematisch dargestellt werden. Danach werden zwei korrespondierende Punkte,  $P_{left}$  und  $P_{right}$  des linken beziehungsweise rechten Bildes, die einen und denselben 3D-Punkt des Raumes referenzieren, betrachtet. Für die Rektifikation der beiden Bilder werden die Matrizen  $T1$  und  $T2$  benötigt, die die Projektion zwischen den beiden Punkten beschreiben.

Die Camera Calibration Toolbox für Matlab stellt dafür die Funktion  $rectify(P_{left}, P_{right})$  bereit, die in der vorliegenden Arbeit folgende zwei Rektifikationsmatrizen liefert 9.3 und 9.4.

$$T1 = \begin{pmatrix} 0.9704 & -0.0468 & 17.6818 \\ 0.0294 & 0.9954 & -6.8439 \\ -0.0001 & 0.0000 & 1.0219 \end{pmatrix} \quad (9.3)$$



$$T_2 = \begin{pmatrix} 0.9791 & -0.0414 & 35.7155 \\ 0.0252 & 0.9989 & -2.3804 \\ -0.0001 & 0.0000 & 1.0169 \end{pmatrix} \quad (9.4)$$

Mit den Kenntnissen den Verzerrungskoeffizienten, der intrinsischen sowie extrinsischen Parameter und der Rektifikationsmatrizen sind alle notwendigen Parameter für die Selbstlokalisierung des Roboters anhand von Landmarken bestimmt. In weiteren Abschnitten wird darauf basierend erläutert, wie die Bildpaare rektifiziert und die Tiefeninformationen wiedergewonnen werden. Im folgenden Unterkapitel wird eine Alternative zur Kamerakalibrierung mit der Camera Calibration Toolbox für Matlab vorgestellt und erläutert.

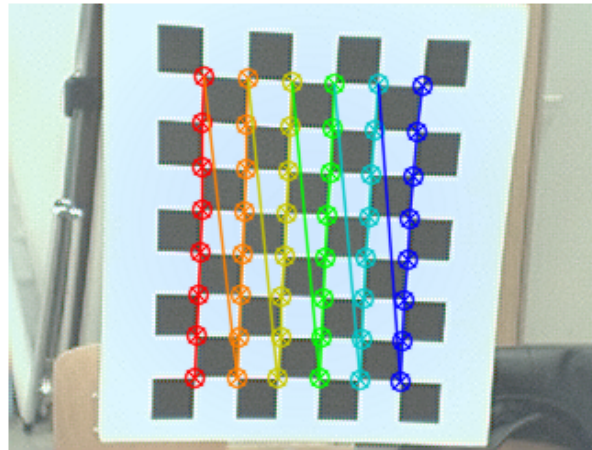
### 9.1.2 Mit Algorithmen der OpenCV-Bibliothek

Die Kamerakalibrierung unter Verwendung der OpenCV-Bibliothek gestaltet sich weit gehend ähnlich zu der Kalibrierung mit der Camera Calibration Toolbox für Matlab und ist genauso nach dem Verfahren von Zhengyou Zhang [Zha00] realisiert. Zuerst werden die beiden Kameras einzeln kalibriert, auch der eingesetzte Kalibrierungskörper bleibt der gleiche. In folgenden Abschnitten wird zuerst auf die Kalibrierung einzelner Kameras und anschließend auf die Kalibrierung eines Stereosystems eingegangen und die gewonnenen Ergebnisse präsentiert.

#### Kalibrierung einer Kamera

Um eine Kamera zu kalibrieren, werden mehrere Bilder aufgenommen und mit der Funktion `cvFindChessBoardCornerGuesses` die inneren Eckpunkte des Kalibrierungsmusters detektiert. Mit der Funktion `cvFindCornerSubPix` kann die Genauigkeit der Eckpunktepositionen verbessert werden. Die gefundenen Eckpunkte können mit der Funktion `cvDrawChessboardCorners` angezeigt werden, eine der möglichen grafischen Darstellungen wird in der Abbildung 9.6 veranschaulicht.

Durch unterschiedliche Positionierung und Neigung des Kalibrierungskörpers werden verschiedene affine und perspektivische Transformationen aufgefasst und bei der Berechnung der intrinsischen und extrinsischen Parametern mitberücksichtigt. Nachdem die Daten aus den möglichst vielen Bildern zur Verfügung stehen, werden mit der Funktion `cvCalibrateCamera2` die intrinsischen Parameter und die Linsenverzeichnung der jeweiligen Kamera berechnet. Sind diese Parameter bekannt, kann



**Abbildung 9.6:** Grafische Darstellung detektierter innerer Eckpunkte des Kalibrierungskörpers, die in das entsprechende Bild projiziert werden.

unter Zuhilfenahme der Funktion *cvFindExtrinsicCameraParams2* die Rotation und Translation der Kamera bezüglich des Kalibrierungskörpers bestimmt werden. Dadurch ist es möglich jedes Bild so neu zu projizieren, das die Kamera sich im rechten Winkel zu dem Kalibrierungskörper befindet und seine Mitte im optischen Zentrum liegt. In der Tabelle 9.7 werden die Ergebnisse der Kalibrierung der linken und rechten Kamera präsentiert.

	Linke Kamera	Rechte Kamera
Destortion:	0.807171 -2.910991 0.005936 0.011895	0.666211 -1.647859 0.039611 -0.019206
Intrinsische Parameter:	472.912231 0.000000 185.057663	473.821472 0.000000 143.811569
	0.000000 479.248444 141.710815	0.000000 480.072357 156.916199
	0.000000 0.000000 1.000000	0.000000 0.000000 1.000000

**Abbildung 9.7:** Tabellarische Darstellung der Parameter für die linke und rechte Kamera, die unter Zuhilfenahme der OpenCV-Bibliothek bestimmt sind. Dabei werden die extrinsischen Parameter nicht aufgelistet, da sie für jedes Bild unterschiedlich sind und daher kaum aussagekräftig sind.

Es ist ersichtlich, dass die Parameter sich nur geringfügig von den Ergebnissen der Camera Calibration Toolbox für Matlab unterscheiden. Damit ist die Kalibrierung einzelner Kamera abgeschlossen, die Ergebnisse sind stabil und das resultierende Programm läuft vollautomatisch ab.

## Kalibrierung eines Stereokamerasystems

Die OpenCv-Bibliothek stellt keine dokumentierte Funktionen zur Kalibrierung eines Stereokamerasystems bereit. Es existieren zwar einige Klassen wie die *cvCalibFilter*, die die Kalibrierung eines Stereosystems ermöglichen sollen und alle auf der Funktion *icvStereoCalibration* basieren. Dennoch ist die Kalibrierung ungenau und die Standardabweichung groß [BEK<sup>+</sup>04].

Auch der Versuch die Rotation und Translation des Stereosystems aus den extrinsischen Parametern beider Kameras zu bestimmen, die auf den Aufnahmen einer und derselben Szene aus unterschiedlichen Kameraperspektiven basiert, brachte keine stabilen und genauen Ergebnisse. Die dafür benötigten Gleichungen (9.5) und (9.6) werden nachfolgend präsentiert, eine ausführliche Herleitung, der beiden Formeln ist, in der Dissertation von Karsten Mühlmann zu finden [Müh02].

$$R_{RL} = R_R R_L^{-1} \quad (9.5)$$

$$t_{RL} = t_R - R_R R_L^{-1} t_L \quad (9.6)$$

Die Berechnung und Einbindung des Skalierungsfaktors verbesserte und stabilisierte die Ergebnisse nur geringfügig, sodass die Genauigkeit trotzdem nicht ausreichte um die Rektifikationsmatrizen zu berechnen.

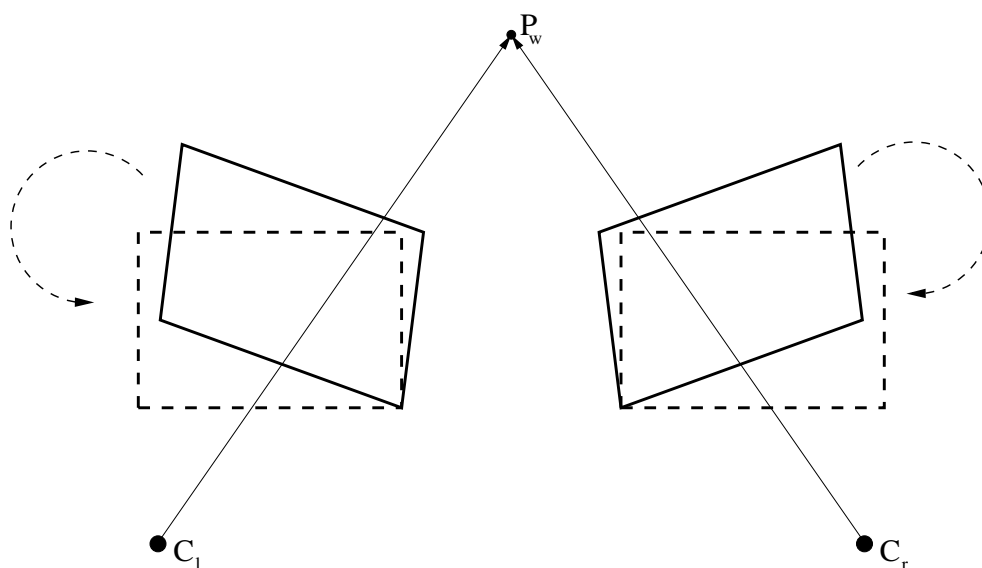
Deswegen wird in Rahmen dieser Diplomarbeit für die Bestimmung der Stereoparameter und im Besonderen der Rektifikationsmatrizen das Verfahren nach Fusiello, Trucco und Verri eingesetzt [FTV00].

Als Eingabe dienen dabei die Projektionsmatrizen der linken und rechten Kamera. Die beiden Projektionsmatrizen bestehen aus den intrinsischen und extrinsischen Parametern der jeweiligen Kamera und werden nach der Kalibrierung einzelner Kameras wie in der Gleichung (9.7) dargestellt berechnet.

$$P = A \cdot [R \ T] \quad (9.7)$$

Dabei ist A die Matrix mit den intrinsischen Parametern, R eine unter Zuhilfenahme der Rodrigues-Funktion berechnete  $3 \times 3$  Rotationsmatrix und T ein  $1 \times 3$  Translationsvektor.

Die optischen Zentren der beiden Kameras bleiben beim diesem Verfahren erhalten, lediglich die Rotation der Kameras wird verändert. Damit wird von den Autoren ein ideales Stereokamerasystem konstruiert, indem die Projektionsflächen der beiden um die Kamerazentren in die erwünschte Position gedreht werden. Die Vorgehensweise wird in der Abbildung 9.8 verdeutlicht.



**Abbildung 9.8:** Künstlich konstruierte Position der Bildebenen, die durch die Rektifikation erreicht wird.

Durch diesen Vorgang wird es möglich das neu Paar Projektionsmatrizen zu bestimmen. Anschließend werden die Transformationsmatrizen kalkuliert, die die als Eingabe dem Algorithmus mitgelieferte Matrizen auf die neue Bildebene projizieren. Die berechneten Transformationsmatrizen werden nach der Linsenverzerrung auf die Bilder angewandt und damit rektifiziert.

Der Quellcode für den Algorithmus steht nur für den Matlab zur Verfügung, um das Verfahren in der vorliegenden Arbeit einsetzen zu können, wird dieser in der Programmiersprache C implementiert. Die bestimmten Transformationsmatrizen werden in der Tabelle 9.9 vorgestellt. Die originalen und rektifizierten Bilder werden in folgenden Abschnitten vorgestellt.

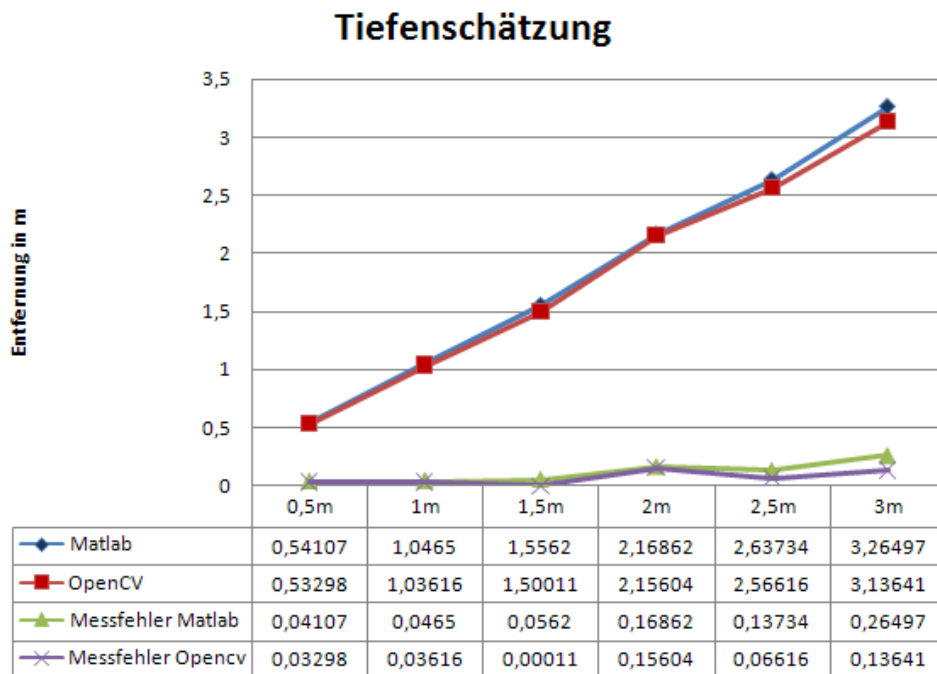
$\mathbf{T}_l$	$\mathbf{T}_r$
0.984750, 0.007055, 27.346222,	0.985330, 0.004180, 13.168516,
-0.016733, 0.996122, 5.801623,	-0.030008, 0.995588, 14.587992,
-0.000065, -0.000000, 1.009505	-0.000106, -0.000054, 1.025135

**Abbildung 9.9:** Tabellarische Darstellung der Transformationsmatrizen, die anhand des in OpenCV implementierten Algorithmus nach Fusiello bestimmt werden.

Damit entstand ein vollautomatisches Verfahren, dass die Kalibrierung einzelner Kamera sowie die anschließende Kalibrierung eines Stereosystems ermöglicht. Als

Ergebnis werden entzerrte rektifizierte Bilder geliefert, die direkt zur Wiedergewinnung der Tiefeninformation genutzt werden können. Der Vorgang der Rektifikation ermöglicht eine effiziente Suche nach korrespondierenden Pixel, da er die Suche auf eine Bildzeile reduziert.

Die berechneten Resultate sind durchaus mit den Ergebnissen der halbautomatischen Kalibrierung mit der Camera Calibration Toolbox für Matlab vergleichbar. Um diese These zu verifizieren, wird folgend vorgegangen, anhand der berechneten Parameter wird die Entfernung zu einem Objekt geschätzt und die Ergebnisse verglichen. Dabei wird die Strecke zwischen dem Stereokamerasystem und dem Objekt vom 0,5 m bis 3 m in 0,5 m Schritten vergrößert. Das Diagramm 9.10 präsentiert die geschätzten Entfernungen.

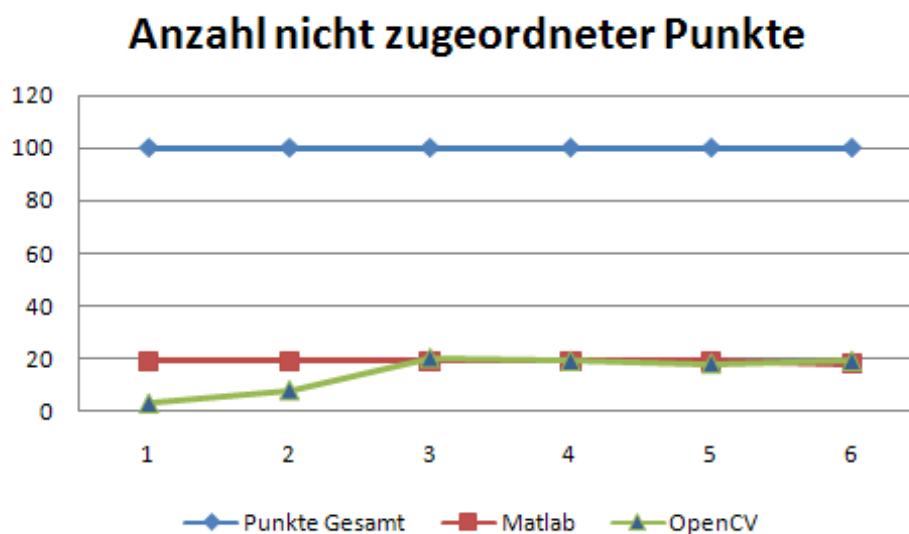


**Abbildung 9.10:** Vergleich der bestimmten Kamerakalibrierungsparameter aus den Verfahren von Camera Calibration Toolbox für Matlab und OpenCV, mit dem implementierten Verfahren nach Fusiello, Trucco und Verri, anhand von Tiefenschätzung.

Für die Entfernungsschätzung werden Blöcke von  $10 \times 10$  Bildpunkten gewählt und für jedes Pixel die Tiefenschätzung durchgeführt. Die Punkte für die keine Disparität gefunden werden könnte, werden verworfen. Um das präsentierte Ergebnis zu erreichen, wird über die restlichen Punkte gemittelt.

Das Diagramm 9.11 zeigt Anzahl der Punkte für unterschiedliche Tiefen, für die keine Korrespondierendepunkte gefunden und damit keine Entfernung geschätzt wer-

den konnten. Die blaue Linie stellt die Gesamtanzahl der Punkte, für die die Korrespondenzpunkte gesucht werden. Die roten und grünen Linien stellen die Anzahl der Punkte, für die keine Korrespondenz gefunden werden konnte, jeweils für die Kamerakalibrierung mit Camera Calibration Toolbox für Matlab beziehungsweise OpenCV dar. Dabei fällt auf, dass bei der Kalibrierung mit Camera Calibration Toolbox für Matlab die Anzahl der Punkte für die keine Disparität vorlag relativ konstant bleibt. Bei OpenCV fällt die Anzahl dieser Punkte bei geringerer Entfernung kleiner aus, steigt aber schnell an und stabilisiert sich bei ungefähren vergleichbaren Werten von Matlab und liegt zwischen 18 % - 20 %.



**Abbildung 9.11:** Grafische Darstellung der Anzahl der Gesamtpunkte sowie der Bildpunkte für die keine Korrespondenz mit den beiden oben beschriebenen Verfahren gefunden werden konnte.

Das auf den Algorithmen von OpenCV basierendes Verfahren mit der Bestimmung der Rektifikationsmatrizen nach Fusiello, Trucco und Verri zeichnet sich durch Robustheit, Schnelligkeit und Stabilität aus. Außerdem läuft das Verfahren automatisch ab, was die Kamerakalibrierung nicht nur erleichtert, sondern auch enorm beschleunigt. Die Genauigkeit der Ergebnisse ermöglicht eine Tiefenrekonstruktion und daraufbauende Schätzung der Position des Roboters.

Damit wird oben beschriebenes automatisches Verfahren zur Kalibrierung einer Kamera beziehungsweise eines Stereokamerasystems in der vorliegenden Arbeit eingesetzt. Die mit dem Camera Calibration Toolbox für Matlab bestimmten Kameraparameter werden in dieser Arbeit nur zum qualitativen Vergleich mit den Ergebnissen der Kamerakalibrierung mit Algorithmen von OpenCV und dem Verfahren nach Fusiello, Trucco und Verri verwendet. Im folgenden Abschnitt wird die Bildvorverarbeitung und Tiefenrekonstruktion basierend auf den Daten der Kalibrierung des

Stereokamerasystems behandelt.

## 9.2 Bildvorverarbeitung und Tiefenrekonstruktion

Die Tiefenrekonstruktion ausgehend von einem Stereokamerasystem kann für verschiedene Zwecke in der Robotik verwendet werden zum Beispiel Vermessung des Abstandes zu einer Landmarke, Kollisionsvermeidung, Erfassung von Objekten ect. Daher werden in dieser Arbeit die Bildvorverarbeitung und die anschließende Tiefenrekonstruktion von einem Stereobildpaar so ausgelegt, dass diese ein breites Spektrum der Weiterverwendung abdecken. Die einzelnen Schritte der Tiefenrekonstruktion sind im Kapitel 4 aufgelistet.

Nachdem die Parameter des Stereokamerasystems durch Kalibrierung (siehe Unterkapitel 3.5.3) bestimmt wurden, wird in diesem Abschnitt die Bildvorverarbeitung und Tiefenrekonstruktion signifikanter Elemente beziehungsweise Szene unter Verwendung dieser Parameter behandelt.



**Abbildung 9.12:** Aufnahme einer Büroumgebung mit dem Stereokamerasystem des HOAP 2. Auf diese Bilder ist zur Weiterverarbeitung eine Grauwerttransformation angewendet worden.

Da die Implementation der Algorithmen in dieser Arbeit unter Verwendung der OpenCV-Bibliothek erfolgt, werden einige Algorithmen mit dem Einsatz dieser Bibliothek realisiert. Um ein Überblick über die in diesem Unterkapitel realisierten Methoden zu schaffen, werden zunächst die einzelnen Verarbeitungsschritte aufgelistet. Als erstes wird die radiale Entzerrung der Kamerabilder durchgeführt, um die durch die Linsen der Kameras verursachte Verzeichnung zu eliminieren. Danach wird durch Neuprojektion das Stereobildpaar rektifiziert, sodass dieses einer Aufnahme eines Kamerasystems in Standardstereogeometrie gleicht. Im nachhinein werden die umgesetzten Methoden zur Extraktion signifikanter Elemente vorgestellt.

Schließlich wird durch Korrespondenzanalyse die Disparitätskarte signifikanter Elemente beziehungsweise der gesamten Szene berechnet. Zur qualitativen Bewertung der Algorithmen werden die Disparitätskarten der Korrespondenzanalyseverfahren verglichen und dreidimensional visualisiert.

Die Operationen auf Bildern sind immer mit der Verarbeitung von großen Datenmengen verbunden, so ist es sinnvoll von vornherein eine möglichst weitgehende Datenreduktion durchzuführen. Die in dieser Arbeit verwendeten Stereokamerasysteme bauen zwar auf Farbkameras auf, die algorithmische Verarbeitung wird jedoch auf Grauwertbildern durchgeführt. Dadurch wird die Bildverarbeitung beschleunigt und Speicherbedarf reduziert. In der Abbildung 9.12 wird eine Stereobildaufnahme der Ausgangsgrauwertbilder aufgenommen mit dem Stereokamerasystem des HOAP 2 präsentiert.



**Abbildung 9.13:** Stereobildpaar nach durchgeführter Entzerrung. Das Ergebnis der radialen Entzerrung ist gut erkennbar an langen Geraden in den Bildern (siehe Türrahmen und vergleiche mit dem Stereobildpaar aus Abbildung 9.12)

Verschiedenen Typen von Objektiven zum Beispiel Weitwinkel- oder Zoomobjektive besitzen unterschiedliche Brechungskoeffizienten des Lichts in Linsen. Die Vielfalt von Objektiven ermöglicht die Anpassung der Kameras an ein breites Spektrum von Aufgaben. Ein großer Brechungskoeffizient verursacht Verzerrungen in der Projektion auf die Bildebene. Der Vorteil eines Weitwinkelobjektives ist eine große Erfassung der Umgebung. Der Nachteil ist jedoch der große Brechungskoeffizient. Des Weiteren weisen Linsen wegen der physikalisch begrenzten Genauigkeit in der Fertigung einen ungleichmäßigen Schliff auf, sodass zusätzliche Fehler in der Abbildung verursacht werden. Um die Projektionsfehler in den Bildern zu eliminieren, müssen die Verzeichnungskoeffizienten bestimmt werden. Dies kann durch Kalibrierung der Kameras erfolgen. Die Ergebnisse der Kalibrierung werden in Tabelle 9.3 präsentiert. Für die Entzerrung der Bilder der Kameras werden zusätzlich die jeweiligen intrinsischen Matrizen benötigt. Diese sind für die linke und rechte Kamera in der Tabelle 9.7 gegeben. Die hier beschriebenen Elemente können in die OpenCV-Funktion



*cvUndistort2* eingesetzt und die radiale Entzerrung an den Bildern durchgeführt werden. Die Abbildung 9.13 präsentiert das Ergebnis der radialen Entzerrung eines Stereobildpaares.

Da die in dieser Arbeit implementierten und verwendeten Algorithmen für die Korrespondenzanalyse rektifizierte Stereobildpaare voraussetzen, werden die Bilder für die Weiterverwendung rektifiziert. Durch die Rektifikation werden die beiden Bilder neu projiziert, sodass diese einer Aufnahme der Standardstereogeometrie gleichen. Die Vorteile einer Standardstereogeometrie sind Kapitel 4.4 beschrieben. Die Rektifikation der Stereobildpaare lässt sich unter Verwendung der in dem Unterkapitel 9.1.2 bestimmten Transformationsmatrizen (siehe Tabelle 9.9) und der OpenCV-Funktion *cvWarpPerspective* durchführen. Eine weitere Möglichkeit die Bilder zu rektifizieren, besteht durch Bestimmung der Homographien (siehe Unterkapitel 4.4.3). In Abbildung 9.14 wird ein Stereobildpaar nach angewandter Rektifikation dargestellt.



**Abbildung 9.14:** Rektifiziertes Stereobildpaar. Die Rektifikation überführt die Stereoaufnahme der allgemeinen Stereogeometrie in die Bilder der Standardstereogeometrie. Dadurch wird die Suche korrespondierender Punkte der Korrespondenzanalyse von zwei auf eine Dimension eingeschränkt.

Der nächste Verarbeitungsschritt ist von der gewählten Weiterverwendung abhängig. Einer der implementierten Verarbeitungsschritte ist die Kantendetektion. Die Kantendetektion ist Teil einer Segmentierung in der Bildbearbeitung bei der versucht wird, flächige Bereiche in einem digitalen Bild von einander zu trennen. Die Kantendetektion kann zum Beispiel für die Wiedererkennung von Objekten verwendet werden. Es ist natürlich denkbar, ein Objektbild pixelweise mit einem unbekanntem Bild, das das Objekt enthält, zu vergleichen. Diese Methode wäre jedoch sehr fehleranfällig gegenüber Größenänderungen, Drehungen, Änderungen des Blickwinkels und den Belichtungsänderungen des Objektes im Bild. Des Weiteren wäre sie sehr ineffizient, da zum Beispiel auch Objekte verglichen werden, die eine eckige Form aufweisen, obwohl nach runden Objekten gesucht wird. Um diese Nachteile zu vermeiden wird versucht unter Verwendung eines Kantendoperators die Komplexität

des Informationsgehaltes eines Objektes zu erhöhen. Dies bringt auch den positiven Effekt der Datenreduktion und erlaubt schnellere Bildverarbeitung.



**Abbildung 9.15:** Originalbild (links) und Kantenbild (rechts). Das hier präsentierte Kantenbild einer Tischszene wurde durch Anwendung des Laplacian of Gaussian-Filters mit den Parameter  $\sigma = 0.7$  und Fenstergröße  $3 \times 3$  generiert.

In dieser Arbeit wird der Laplacian of Gaussian-Filter zur Kantenextraktion implementiert. Die Vorteile dieses Filters sind im Unterkapitel 5.2.1 vorgestellt. In der Abbildung 9.15 wird das Ergebnis des Laplacian of Gaussian-Filters mit den Parameter  $\sigma = 0.7$  und Fenstergröße  $3 \times 3$  vorgestellt. Im rechten Bild sind die Objekte durch ihre Kanten im Binärbild wiederzuerkennen. Wird in einem späteren Verarbeitungsschritt die Tiefe der einzelnen Kantenpunkte rekonstruiert, ergibt sich ein dreidimensionales Kantenbild. Dieses kann dann zur Wiedererkennung von Objekten verwendet werden.

Für die Korrespondenzanalyse ist der Shirai-Algorithmus, Block-Matching und der Algorithmus vom S. Birchfield und C. Thomasi implementiert worden. Auf die Berechnung flächendeckender Disparitätskarten mit diesen Algorithmen wird noch im Laufe dieses Kapitels näher eingegangen. An dieser Stelle soll nach erfolgter Kantenextraktion die dazugehörige Disparitätskarte berechnet werden. Dafür wird auf das Kantenbild der Shirai-Algorithmus aus [KKS96] angewendet. Dieser Algorithmus berechnet das Ähnlichkeitsmaß zwischen einzelnen Bildpunkten anhand der Intensität der Bildpunkte innerhalb einer festdefinierten Fenstergröße. Da der Shirai-Algorithmus ein intensitätsbasiertes Verfahren ist, muss das Kantenbild nicht binär, sondern in Graustufen vorliegen. In Abbildung 9.16 wird das Kantenbild in Graustufen und die mit dem Shirai-Algorithmus aus [KKS96] berechnete Disparitätskarte präsentiert.

Um einen Roboter innerhalb einer dynamischen Umgebung zum Beispiel eines Bürokomplexes navigieren zu können, ist das Vorhandensein einer Kollisionsvermeidung und der Selbstlokalisierung unumgänglich. Zur Realisierung einer Kollisionsvermeidung beziehungsweise Selbstlokalisierung ist eine permanente Erfassung



**Abbildung 9.16:** Kantenbild (links) und dazugehörige Disparitätskarte (rechts). Für die Berechnung der Disparitätskarte mit dem Shirai-Algorithmus aus [KKS96] wird das Kantenbild (links) in Graustufen generiert. Die Graustufen der Disparitätskarte repräsentieren die Tiefeninformationen der Bildpunkte im Raum. Je heller die Graustufe des Punktes, desto näher befindet sich dieser zur Kamera.

der Umgebung nötig. Das Ausgangssystem (humanoide Roboter HOAP 2) besitzt ein Stereokamerasystem als einzigen Sensor zur Erfassung der Umgebung, sodass sämtliche Methoden zur Wechselwirkung mit der Umgebung über diese Sensoren zu erfolgen haben.

Als mögliches Szenario die Tiefenrekonstruktion weiterzuverwenden ist wie schon oben erwähnt die Realisierung der Kollisionsvermeidung. Die Selbstlokalisierung des Roboters anhand von Landmarken wird im gesonderten Unterkapitel 9.3 behandelt. Die Kollisionsvermeidung kann unter Verwendung einer flächendeckenden Tiefenkarte der Umgebung erfolgen. Die Tiefeninformation einer Szene der Umgebung kann durch die Berechnung der Disparitätskarte mit einem Korrespondenzanalyseverfahren und anschließender Anwendung der Tiefenbestimmung wiedergewonnen werden. In den vorherigen Abschnitten sind die Bildvorverarbeitungsschritte zur Korrespondenzanalyse durchgeführt worden. Nachdem nun die Verzerrungen in den Bildern eliminiert sind und das Stereobildpaar rektifiziert ist, wird an dieser Stelle auf die Gewinnung der Disparitätskarte eingegangen. Aus der schließlich unter Zuhilfenahme der Parameter des Stereokamerasystems die Berechnung der Tiefenkarte erfolgen kann.

**Auswahl des geeigneten Korrespondenzanalyseverfahrens.** Im Kapitel 5.2 werden unterschiedliche Korrespondenzanalyseverfahren vorgestellt. Das Ergebnis der meisten Korrespondenzanalyseverfahren ist eine Disparitätskarte. Diese ergibt sich aus der Differenz korrespondierender Bildpunkte. Da die Zuordnung korrespondierender Bildpunkte jedoch nicht eindeutig ist, liefern die Algorithmen als Ergebnis qualitativ unterschiedliche Disparitätskarten.



**Abbildung 9.17:** Stereobildpaar einer Tischszene. Die beiden Bilder des Stereokamerasystems werden als Vorlage zur Untersuchung der Güte der Korrespondenzanalyseverfahren verwendet.

In diesem Abschnitt sollen die Algorithmen gegenübergestellt und somit ihre Güte verglichen werden. Zur Entscheidung der Güte muss jedoch zuvor ein Vergleichskriterium definiert werden. Ein mögliches Vergleichskriterium wird zum Beispiel in [SBK00] definiert. Dafür wird aus einem Bild unter Verwendung der Grauwerte der Disparitätskarte das zweite Bild des Stereopaars generiert und die Differenz zum Ausgangsbild mit

$$E_{diff} = \frac{1}{N} \sum_{(x,y)} |E_l(x, y) - E_r(x + d, y)| \quad (9.8)$$

berechnet, wobei  $N$  die Anzahl der Pixel und  $E_{\{l,r\}}(x, y)$  der Intensitätswert eines Pixels im linken beziehungsweise im rechten Bild ist. Das Korrespondenzanalyseverfahren mit dem kleinsten Ergebnis der Gleichung (9.8) liefert somit nach [SBK00] die qualitativ bessere Disparitätskarte.

Ein weiteres Vergleichskriterium ist in [SS02] gegeben. Für die Berechnung der Güte wird per Hand ein ideales Disparitätsbild erstellt und vom Disparitätsbild des Algorithmus subtrahiert. Die Güte wird mit der Gleichung

$$R = \sqrt{\frac{1}{N} \sum_{(x,y)} |d_C(x, y) - d_T(x, y)|^2} \quad (9.9)$$

bestimmt, wobei  $N$  wieder die Anzahl der Pixel im Bild,  $d_C$  per Hand erstellte und  $d_T$  vom Algorithmus gelieferte Disparitätskarte ist. Die Interpretation der Güte geschieht analog zum Vergleichskriterium aus [SBK00].

Die beiden vorgestellten Vergleichskriterien liefern als Ergebnis eine numerische Größe, anhand deren die Güte der Algorithmen bewertet wird. Diese Größe liefert jedoch keine Informationen über die Schwachstellen und Stärken der Algorithmen, sodass von dieser Art der Bewertung abgewichen wird. Um die Schwachstellen und



**Abbildung 9.18:** Entzerrtes und rektifiziertes Stereobildpaar einer Tischszene. Dieses Stereobildpaar diente als Vorlage zum Vergleich der implementierten Korrespondenzanalyseverfahren.

Stärken der Algorithmen herauszufinden, wird das Verhalten der Algorithmen im Bezug zu einer aufgenommenen Tischszene untersucht. Dabei werden vor allem Objektkanten, texturarme und verdeckte Bereiche der Szene betrachtet. In Abbildung 9.17 wird das als Vorlage verwendete Stereobildpaar der Tischszene präsentiert. Beim Aufbau der Szene ist darauf geachtet worden, die Objekte gleichmäßig in verschiedenen Tiefen zu platzieren. Um einen fairen Vergleich zwischen den Algorithmen zu ermöglichen, werden diese auf das gleiche rektifizierte Stereobildpaar aus Abbildung 9.18 angewendet.

Als erstes wird der Shirai-Algorithmus aus [KKS96, Kap. 1.4.2] implementiert und untersucht. Die theoretische Beschreibung dieses Algorithmus ist im Unterkapitel 5.2.1 zu finden. Da der Algorithmus die Zuordnung korrespondierender Punkte eines Stereobildpaars anhand der Intensitäten der Punkte innerhalb eines Fensters vornimmt, wird dieser in [KKS96, Kap. 1.4.2] auch für die Berechnung der Disparitätskarten von Kantenbildern verwendet. Bei der Anwendung dieses Algorithmus auf flächendeckende Bilder lassen sich seine Schwächen und Stärken erkennen.

Der Algorithmus liefert wegen seines lokalen Suchcharakters stabile Disparitäten an Kanten und stark texturierten Bereichen der Objekte. Für schwachtexturierte Bereiche weist der Algorithmus jedoch eine Schwäche auf. Des Weiteren ist festzustellen, dass der Algorithmus bei gleichmäßig sich wiederholenden Mustern keine eindeutige Zuordnung vornehmen kann.

Die einzelnen Objekte der Tischszene aus dem vorgelegten Stereobildpaar sind gegenüber dem Kamerasystem in eine Tiefenebene gelegt worden, sodass die Disparität über deren gesamten Bereich relativ konstant sein sollte. Durch die schwach und stark texturierten Bereiche der Objekte wird ein Großteil der Disparitäten innerhalb der Objekte vom Algorithmus jedoch falsch berechnet und über ein breites Spektrum zerstreut. Bei gleichmäßig sich wiederholende Muster in der Szene weist

der Algorithmus ein ähnliches Verhalten auf. In Abbildung 9.19 wird mit dem Shirai-Algorithmus generierte Disparitätskarte dargestellt.



**Abbildung 9.19:** Mit dem Shirai-Algorithmus generierte Disparitätskarte. Die Grauwerte des Bildpunktes bilden die Basis zur Rekonstruktion der Tiefeninformation der Punkte im Raum. Die Tiefe der Weltpunkte ist auch im Disparitätsbild erkennbar und zwar, je heller der Grauwert, desto näher befindet sich dieser zur Referenzkamera.

Als nächstes wird der Block-Matching-Algorithmus untersucht. Die Berechnung der Disparitätskarte mit Block-Matching lässt sich wie im Unterkapitel 5.2.2 schon beschrieben in mehrere Verarbeitungsschritte zerlegen. Als erstes wird das Referenzbild des Stereopaars in gleichgroße Blöcke unterteilt und die Blockdisparität berechnet. Anschließend wird die Blockdisparität auf die Pixelebene verfeinert.

Die Suche nach korrespondierenden Blöcken geschieht unter Verwendung einer Metrik. Verschiedene Metriken sind im oben genannten Unterkapitel beschrieben worden. Für die Durchführung der Experimente ist der Block-Matching-Algorithmus unter Verwendung der Metrik SAD (engl. *sum of the absolute differences*) implementiert worden. Da dieser Algorithmus durch den Vergleich der Intensitätswerte innerhalb eines Fensters und des lokalen Suchcharakters eine Ähnlichkeit mit dem Shirai-Algorithmus aufweist, ist auch das Verhalten gegenüber Objektkanten, texturreichen und verdeckten Bereichen mit dem Shirai-Algorithmus vergleichbar. Im Unterschied zum Shirai-Algorithmus sind die Kanten der Objekte durch die Verfeinerung der Blockdisparität auf Pixelebene eher unscharf. In Bereichen mit gleichmäßig

sich wiederholender Muster zum Beispiel die Maserung des Schrankes beziehungsweise Tisches (siehe Abbildung 9.20) versagt der Algorithmus schlechthin.



**Abbildung 9.20:** Mit Block-Matching generierte Disparitätskarte. Die Helligkeit der Bildpunkte gibt repräsentativ die Tiefe im Raum wieder. Je heller der Grauwert, desto näher befindet sich Weltpunkt zur Referenzkamera.

Abschließend wird der Algorithmus von S. Birchfield und C. Thomasi auf seine Güte untersucht. Der auf dynamische Programmierung basierende Algorithmus aus [BT98] ist Bestand der OpenCV-Bibliothek, sodass dieser unmittelbar qualitativ bewertet werden kann. Zum Vergleich mit den beiden vorher vorgestellten Korrespondenzanalyseverfahren besitzt der Algorithmus von S. Birchfield und C. Thomasi durch das Prinzip der dynamischen Programmierung einen globalen Suchcharakter, das heißt der Algorithmus bestrebt durch die lokal gefundenen Optima die Scanlinien der beiden Bilder optimal zu überlagern. Der Algorithmus liefert fast in allen untersuchten Bereichen gute Ergebnisse. Zu bemerken sind die wenigen Ausreißer an den Kanten der Objekte sowie der nicht festgestellte Tiefenunterschied zwischen dem Schrank und der Tür auf der rechten Bildseite (siehe Abbildung 9.21). Da der Algorithmus bereits in der OpenCV-Bibliothek zu Verfügung stand, lässt sich nur vermuten, dass in diesem Bereich des Bildes keine korrespondierende Punkte gefunden werden und die zuletzt gültige Disparität vom Schrank für diesen Bereich übernommen wird.

Unter den getesteten Korrespondenzanalyseverfahren lieferte der Algorithmus von S. Birchfield und C. Thomasi die besseren Ergebnisse, sodass dieser den anderen bei-



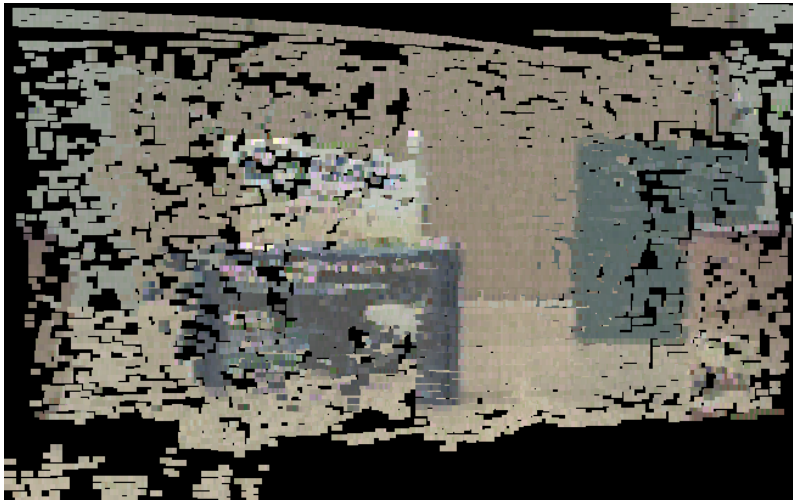
**Abbildung 9.21:** Mit dynamische Programmierung generierte Disparitätskarte. Die Helligkeit eines Bildpunktes repräsentiert den Abstand zum Weltpunkt. Je heller der Grauwert des Punktes im Bild, desto näher befindet sich der dazugehörige Punkt in Weltkoordinaten zur Referenzkamera.

den Verfahren zu bevorzugen ist. Es ist jedoch möglich, dass der Schirai-Algorithmus und Block-Matching in einer anderen Testumgebung zum Beispiel in Naturszenen bessere Ergebnisse liefern, da eine Naturszene nur wenige gleichmäßige Flächen beziehungsweise wiederholende Muster beinhaltet.

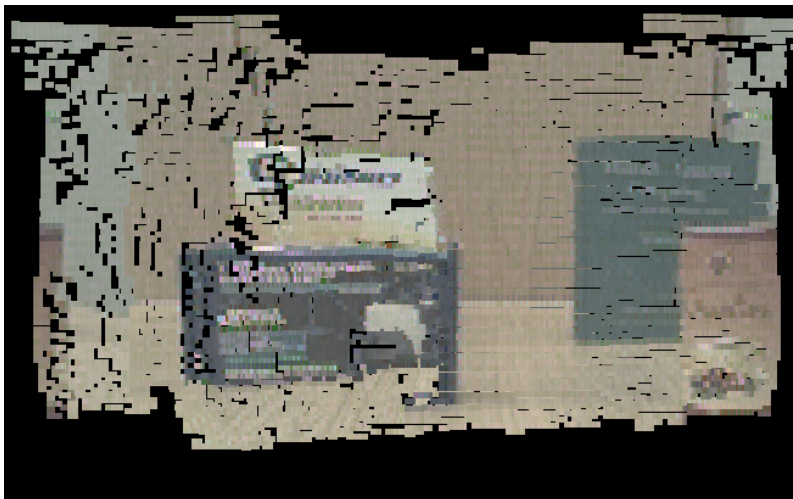
Nachdem die die einzelnen Algorithmen bewertet wurden, werden die berechneten Disparitätskarten mit OpenGL dreidimensional visualisiert. Die Ergebnisse der dreidimensionalen Visualisierung sind in den Abbildungen 9.22-9.24 dargestellt.

Im nächsten Abschnitt wird die Selbstlokalisierung unter Zuhilfenahme passiver künstlicher Landmarken behandelt. Dabei wird die visuelle Erkennung von Landmarken behandelt und anhand bekannter Positionen der Landmarken die Selbstlokalisierung des Roboters durchgeführt.





**Abbildung 9.22:** 3D-Rekonstruktion der mit dem Schirai-Algorithmus generierten Disparitätskarte



**Abbildung 9.23:** 3D-Rekonstruktion der mit Block-Matching generierten Disparitätskarte.



**Abbildung 9.24:** 3D-Rekonstruktion der mit dem Algorithmus von S. Birchfield und C. Thomasi generierte Disparitätskarte.

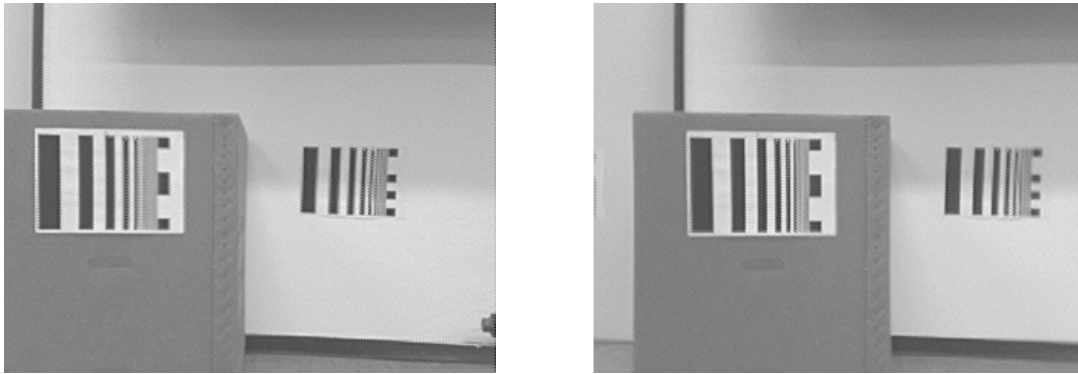
### 9.3 Selbstlokalisierung anhand von Landmarken

Die Selbstlokalisierung eines Roboters kann unter Verwendung von Landmarken realisiert werden. Verschiedene Landmarken zum visuellen Erkennen werden im Kapitel 7 beschrieben. Die visuelle Aufnahme von Landmarken ist permanent wechselnden Lichtverhältnissen, visuellen Verzerrungen und unterschiedlichen Skalierungen ausgesetzt. In dieser Arbeit wurde nach Lösungen gesucht, die sich größtenteils unempfindlich gegenüber diesen Einflüssen verhalten. Die Autoren Daniel Scharnstein und Amy J. Brigs erstellen in Ihrer Arbeit [SB01] ein mathematisches Modell einer Landmarke, die sich robust gegenüber hier genannten Einflüssen erweist. Ein weiterer Vorteil dieser Landmarke ist das gleichbleibende und einfach gehaltene Muster, sodass die Landmarke mit konstanter Komplexität von  $O(\frac{nw}{k})$  in einem Bild wiedergefunden wird. Die Berechnung der Komplexität setzt sich aus der Anzahl der Pixel  $n$  eines Bildes, dem Suchintervall  $w$  und der Anzahl der durchsuchten Spalten  $k$  zusammen.

Um die Landmarken eindeutig identifizieren zu können, werden diese mit einem Barcode erweitert. Dieser setzt sich aus  $i$  Informations- und  $j$  Paritätsbits zusammen. Durch die  $i$  Informationsbits wird die Identität der Landmarke kodiert.

Somit ist die maximale Anzahl unterschiedlicher Landmarken auf  $2^i$  begrenzt. Unter Verwendung der  $j$  Paritätsbits werden im ausgelesenen Barcode mögliche Fehler erkannt und korrigiert.

Die Größe der Barcodestruktur ist durch die Höhe der Landmarke begrenzt, sodass die steigende Anzahl von Informationsbits zu deren Verkleinerung führt. Die verklei-



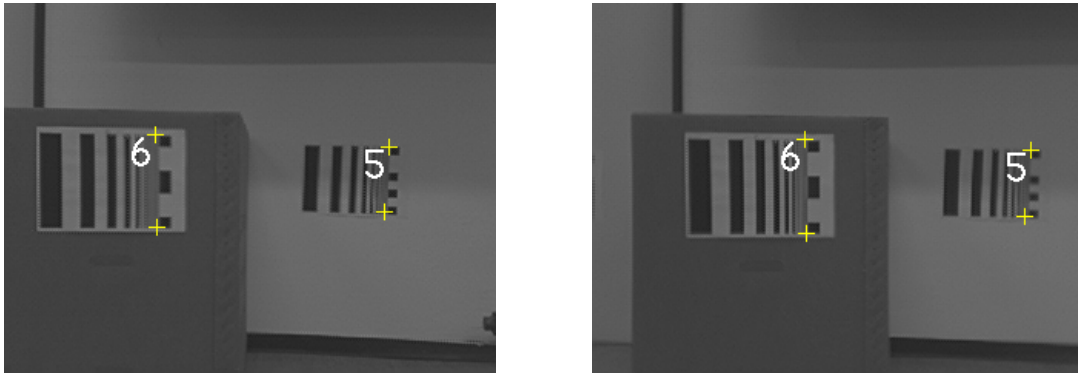
**Abbildung 9.25:** Ausgangsbildpaar des Stereokamerasystems zur Detektion von Landmarken. Das vorliegende Stereopaar wird auf in Bildern vorhandene Landmarken untersucht. Die Suche der Landmarken im Bild wird mit dem von Daniel Scharnstein und Amy J. Brigs vorgeschlagenen Algorithmus aus [SB01] durchgeführt.

nerte Struktur verschlechtert die Auslesequalität des Barcodes und schränkt somit die Landmarkenerkennung auf geringe Distanzen ein. Da das Kamerasystem der humanoiden Roboter (HOAP 2) nur QVGA-Auflösung bereitstellen und die Qualität der Bilder wegen starkem Rauschen eher minderwertig ist, werden die Landmarken für stabilere Erkennung nur mit fünf Informationsbits kodiert. Mit fünf Informationsbits ergeben sich nach oben genannter Berechnung  $2^5 = 32$  unterschiedliche Landmarken. Die zusätzlichen drei Bits des Barcodes stellen die Paritätsbit dar.

Daniel Scharnstein und Amy J. Brigs stellen zu ihrer Veröffentlichung den Quellcode für die Detektion der Landmarken in der Programmiersprache C bereit. Da die in dieser Arbeit implementierten Algorithmen unter Verwendung der OpenCV-Bibliothek realisiert werden, wird auch der Algorithmus zur Landmarkendetektion mit den Methoden der OpenCV-Bibliothek umgesetzt. Dieser Schritt erlaubt eine Interaktion zwischen den bestehenden Algorithmen der Bildvorverarbeitung und Tiefenrekonstruktion.

Da die Landmarken keine Farbinformationen beinhalten, wird der Algorithmus auf Grauwertbildern angewendet. Dies bringt den positiven Effekt der schnellen Verarbeitung und reduziert den Ressourcenverbrauch. In Abbildung 9.25 wird ein Bildpaar des Stereokamerasystems präsentiert, das für die Landmarkendetektion verwendet wird.

Werden die Landmarken in einer aufgenommenen Szene erkannt, so liefert der Algorithmus zu jeder Landmarke zwei Punkte  $p_t(u, v)$  und  $p_b(u, v)$  und die Nummer des ausgelesenen Barcodes. Die Indizes  $t$  und  $b$  stehen für *top* und *bottom* der Landmarke. In Abbildung 9.26 wird ein Stereobildpaar mit erkannten Landmarken präsentiert. Der Abstand zwischen den Punkten der Landmarke in Weltkoordinaten ist durch Vermessen gegeben und beträgt  $\overline{P_t P_b}$ . Auf der Bildebene kann dieser durch die vom



**Abbildung 9.26:** Stereobildpaar aus Abbildung 9.25 mit detektierten Landmarken. Eine Landmarke im Bild gilt als gefunden, wenn der Algorithmus zu dessen Detektion ein Punktpaar und die Identitätsnummer der Landmarke ausgibt. In dem vorliegenden Stereobildpaar sind die beiden Parameter der Landmarken eingeblendet, sodass eine visuelle Begutachtung der algorithmischen Richtigkeit stattfinden kann.

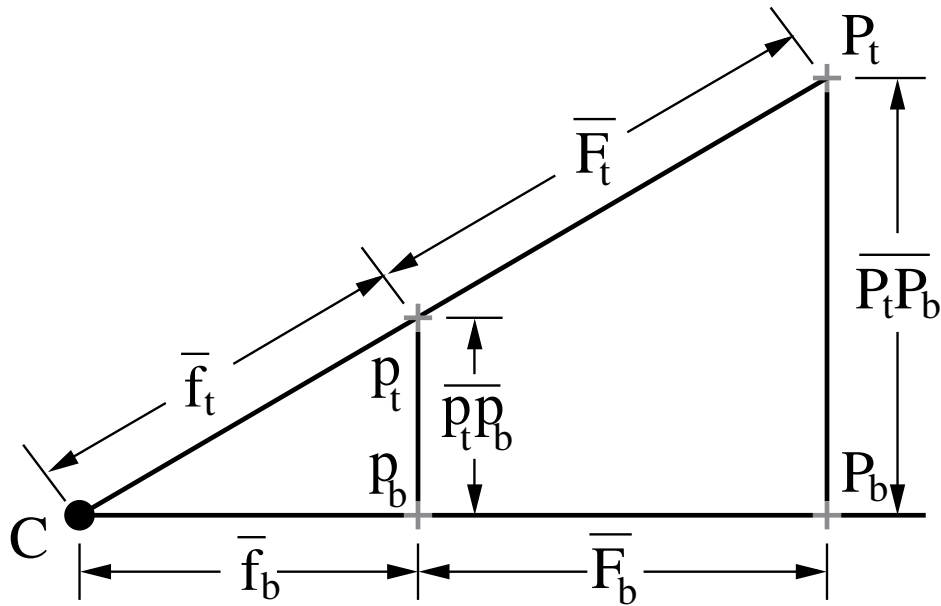
Algorithmus ausgegebenen Koordinaten der Punkte  $p_t(u, v)$  und  $p_b(u, v)$  unter Zuhilfenahme des Pythagorassatzes bestimmt werden und beträgt  $\overline{p_t p_b}$ . In Abbildung 9.27 werden, die hier genannten Größen grafisch dargestellt. Die beiden bekannten Abstände zwischen den Punkten und durch Kalibrierung bestimmte intrinsische Parameter liefern genügend Informationen um die Strecke zwischen Roboter und Landmarke mit dem zweiten Strahlensatz des antiken griechischen Philosophen und Mathematiker Thales von Milet zu berechnen. Die Berechnung des Abstandes zwischen Roboter und Landmarke mit dem zweiten Strahlensatz kann jedoch nur unter der Einschränkung erfolgen, dass die Landmarke in der Welt und der Bildsensor der Kamera in vertikaler Richtung parallel zueinander verlaufen.

*Werden zwei Strahlen gemeinsamen Anfangspunktes von Parallelen geschnitten, so verhalten sich die Längen auf den Parallelen wie die von Anfangspunkt gemessenen Längen der entsprechenden Abschnitte auf den Strahlen.*

Durch diesen Satz ergeben sich die Verhältnisgleichungen

$$\frac{\overline{p_t p_b}}{\overline{P_t P_b}} = \frac{\overline{f_t}}{\overline{F_t}} \quad \frac{\overline{p_t p_b}}{\overline{P_t P_b}} = \frac{\overline{f_b}}{\overline{F_b}} \quad (9.10)$$

Eine weitere Möglichkeit den Abstand zwischen dem Roboter und einer Landmarke zu bestimmen, besteht unter Verwendung der Methoden der statischen Stereoanalyse. Für diesen Ansatz wird aber vorausgesetzt, dass dieselbe Landmarke in beiden Bildern des Stereosystems gefunden wird. In den Abbildungen 9.25 und 9.26 wird ein Stereopaar der Ausgangsbilder und das gleiche Bildpaar mit gefundenen Landmarken präsentiert.



**Abbildung 9.27:** Grafische Darstellung der Zusammenhänge des zweiten Strahlensatzes. Laut Definition des zweiten Strahlensatzes verhält sich die Länge  $\overline{p_t p_b}$  zu  $\overline{P_t P_b}$  wie  $\overline{f_{\{t,b\}}}$  zu  $\overline{F_{\{t,b\}}}$ .

Im folgenden Abschnitt dieses Unterkapitels wird der Prozess der Selbstlokalisierung unter Zuhilfenahme hier genannten Methoden beschrieben. Dafür wird jedoch davon ausgegangen, dass die Landmarken in den Bildern bereits gefunden und dadurch die Koordinaten der Punkte samt Identitätsnummer bekannt sind.

Der Prozess der Selbstlokalisierung eines Roboters kann in mehrere Verarbeitungsschritte zerlegt werden:

1. Für die Berechnung des Abstandes zwischen der Landmarke und dem Roboter wird ein Referenzpunkt am Roboter festgelegt, bezüglich dessen die Vermessung erfolgt. Es ist sinnvoll, das optische Zentrum der Kamera als Ursprung des Referenzkoordinatensystems für die folgenden Berechnungsschritte festzulegen. Da in dieser Arbeit von einem Stereokamerasystem ausgegangen wird, wird der Referenzpunkt jedoch genau zwischen die optischen Zentren der beiden Kameras festgelegt. Da das Kamerasystem kalibriert und in die Standardstereogeometrie überführt wird, kann der Ursprung des Referenzkoordinatensystems durch einfache Translation um  $d_B/2$  beziehungsweise  $-d_B/2$  in die Koordinatensysteme der einzelnen Kameras verschoben werden.
2. Anhand bekannter Abstände zwischen den beiden Punkten einer Landmarke in Welt- und Bildkoordinaten, wird mit dem Strahlensatz der relative Abstand der Landmarke bezüglich des Roboters geschätzt. Wird eine Landmarke in beiden Bildern des Stereosystems erkannt, so kann der relative Abstand auch

mit den Methoden der statischen Stereoanalyse berechnet werden.

3. Mit dem Hauptpunkt der Kamera als Ursprung des Bildkoordinatensystems, wird die Position der Landmarke bezüglich dessen im Bild bestimmt. Diese Information bringt in Erfahrung, ob sich die Landmarke rechts beziehungsweise links vom Roboter befindet.
4. Die berechneten Ergebnisse aus den Schritten zwei und drei werden auf das Weltkoordinatensystem transformiert und die Verschiebung um die halbe Basislänge wieder entfernt. Damit wird der Roboter bezüglich der Landmarken positioniert und anhand deren bekannter Positionen in der Welt lokalisiert.

An dieser Stelle sollen die einzelnen Verarbeitungsschritte etwas ausgiebiger beschrieben werden. Als erstes wird der Ansatz mit einer Kamera untersucht. Wie schon oben erwähnt, wird der Ursprung des Referenzkoordinatensystems genau zwischen die beiden optischen Zentren  $C_L$  und  $C_R$  gelegt. Ausgehend von diesem Punkt, kann das Referenzkoordinatensystem, je nachdem, welche Kamera gerade betrachtet wird, einfach an die Position des jeweiligen optischen Zentrums  $C_L$  oder  $C_R$  um die halbe Basislänge transformiert werden.

Der Abstand  $\overline{P_t P_b}$  zwischen den zwei Punkten der Welt  $P_t$  und  $P_b$  ist durch Vermessen an der Landmarke bekannt. Der Abstand in Bildkoordinaten  $\overline{p_t p_b}$  kann mit dem Pythagorassatz berechnet werden.

$$\overline{p_t p_b} = \sqrt{(u_t - u_b)^2 + (v_t - v_b)^2} \quad (9.11)$$

Die Fokallängen des Stereokamerasystems sind durch Kalibrierung der einzelnen Kameras bestimmt worden. Die Fokallänge einer Kamera gibt den Abstand zwischen dem optischen Zentrum und dem Hauptpunkt  $p_0(u, v)$  der Bildebene an. Da die Sensorfläche der Kamera eine Ebene und kein Halbkreis ist, wird die Fokallänge zu den Rändern der Bildebene größer. Um den Abstand zwischen dem Roboter und einem Punkt der Landmarke mit dem zweiten Strahlensatz in Weltkoordinaten berechnen zu können, muss die Fokallänge zu dem gleichen Punkt in der Bildebene berechnet werden. Da das optische Zentrum der Kamera senkrecht zur Bildebene steht, kann die Berechnung der punktspezifischen Fokallänge mit dem Pythagorassatz erfolgen. Die punktspezifische Fokallänge wird in diesem Zusammenhang in Abhängigkeit vom ausgewählten Bildpunkt  $p_{\{t,b\}}$  mit  $\overline{f}_t$  beziehungsweise  $\overline{f}_b$  bezeichnet. Die Berechnung dieser erfolgt mit

$$\overline{p_{\{t,b\}} p_0} = \sqrt{(u_0 - u_{\{t,b\}})^2 + (v_0 - v_{\{t,b\}})^2} \quad (9.12)$$

wie folgt

$$\overline{f_{\{t,b\}}} = \sqrt{\overline{p_{\{t,b\}} p_0}^2 + \overline{f}^2} \quad (9.13)$$

Mit bekannter Fokallänge  $\overline{f}_{\{t,b\}}$  aus Gleichung (9.13) kann der Abstand  $\overline{F}_{\{t,b\}}$  zu dem Punkt  $P_{\{t,b\}}$  der Landmarke mit den Gleichungen (9.10) berechnet werden.

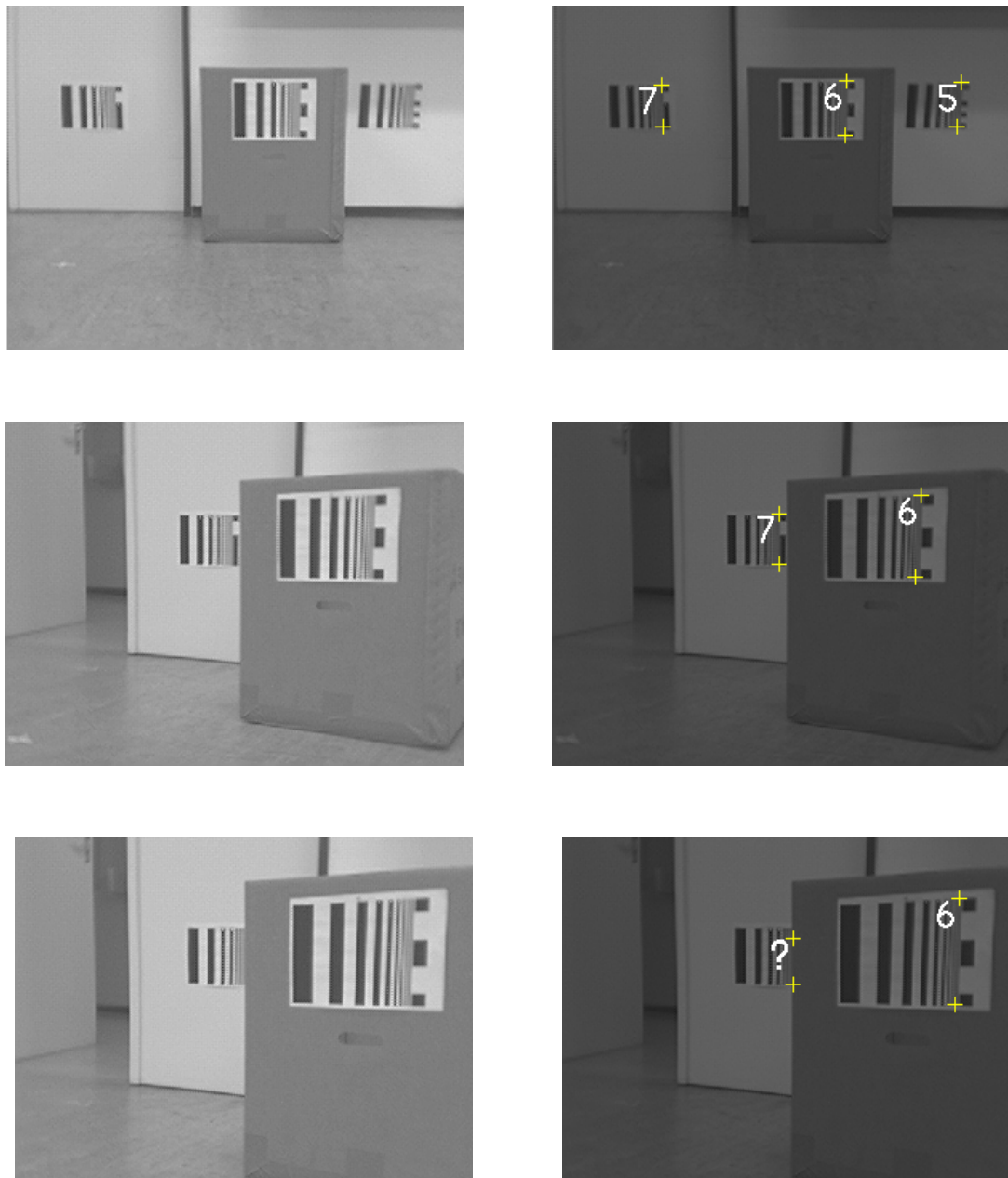
Eine weitere Möglichkeit den Abstand des Roboters zu einer Landmarke zu berechnen, kann unter Zuhilfenahme der Methoden der statischen Stereoanalyse erfolgen. Da das Stereokamerasystem in dieser Arbeit kalibriert und durch Rektifikation in die Standardstereogeometrie überführt wird, kann die Position der Landmarke im Bezug zum Roboter mit den Gleichungen der Standardstereogeometrie berechnet werden. Um diesen Ansatz der Berechnung nutzen zu können, muss jedoch wie schon oben erwähnt dieselbe Landmarke in beiden Bildern des Stereokamerasystems erkannt worden sein. Ist dies der Fall, wird der Abstand zu der Landmarke aus dem jeweiligen korrespondierenden Punktpaar mit der Gleichung 4.1 bestimmt.

Je nachdem welcher der beiden Punkte als Referenzpunkt der Landmarke dient, gibt der Abstand zu diesem den Abstand zur Landmarke wider.

Mit dem berechneten Abstand der Landmarke gegenüber dem Roboter kann dessen Position auf den Radius des Abstandes eingeschränkt werden. Eine weitere Einschränkung ist durch den Sichtbarkeitsbereich der Landmarke von  $180^\circ$  gegeben. Durch die beiden Einschränkungen ist jedoch noch keine eindeutige Positionierung möglich, außer der Roboter befindet sich direkt gegenüber der Landmarke. Die eindeutige Position kann durch die zusätzliche Information der seitlichen Verschiebung der Landmarke im Bezug zum Hauptpunkt des Bildes bestimmt werden. Dies wird das Thema des folgenden Absatzes sein.

Nachdem nun der Abstand zur Landmarke berechnet ist, wird an dieser Stelle die seitliche Verschiebung der Landmarke im Bezug zum Hauptpunkt  $p_0(u, v)$  in Bildkoordinaten bestimmt. Diese lässt sich durch einfache Differenzbildung der U-Koordinaten des Referenzpunktes der Landmarke bezüglich des Hauptpunktes berechnen. Durch die bekannte seitliche Verschiebung im Bild kann unter Zuhilfenahme der intrinsischen Parameter der Kamera die Position des Roboters gegenüber der Landmarke in Weltkoordinaten berechnet werden. Die Berechnung erfolgt mit den oben beschriebenen Zusammenhängen des zweiten Strahlensatzes.

Die bestimmte Position der Landmarke gegenüber dem Roboter kann auf die Position des Roboters bezüglich der Landmarke umgerechnet werden. Mit den bekannten Weltkoordinaten der Landmarke wird dann die Position des Roboters im demselben Koordinatensystem bestimmt. Da der Referenzpunkt des Roboters auf halber Länge der Basislinie festgelegt wurde, die durchgeführten Berechnungen jedoch bezüglich eines Kamerakoordinatesystems erfolgten, ist schließlich eine Korrektur der Position um eine halbe Basislänge notwendig.



**Abbildung 9.28:** Serie von Bildern mit detektierten Landmarken (rechts) und dazugehöriger Ausgangsbilder (links). In dieser Bildserie ist eine unterschiedliche Anzahl von Landmarken aus verschiedenen Perspektiven aufgenommen worden. Im oberen und mittleren Ausgangsbild ist darauf geachtet worden, dass die Landmarken von keinen Objekten der Szene verdeckt werden. Im unteren Ausgangsbild wird ein Teil der Landmarke von der Kartonkiste verdeckt. Die verdeckte Landmarke wird zwar im Bild gefunden, der Barcode kann jedoch nicht mehr ausgelesen werden.



## 9.4 Zusammenfassung

In diesem Kapitel sind die realisierten Methoden und Techniken zur Kalibrierung der einzelnen Kameras beziehungsweise des Stereokamerasystems vorgestellt worden. Zur Kalibrierung der einzelnen Kameras ist das Verfahren nach Z. Zhang [Zha00] verwendet worden, das im Kapitel 3 theoretisch behandelt wird. Dieses Verfahren wurde unter Zuhilfenahme der OpenCv-Bibliothek verwirklicht. Mit bekannten intrinsischen und extrinsischen Parametern einzelner Kameras und implementiertem Verfahren nach Fusiello, Trucco und Verri [FTV00] wurden anschließend die Transformationsmatrizen zur Rektifikation des Stereosystems bestimmt.

Ein weiterer Bestand dieses Kapitels ist die Bildvorverarbeitung und Tiefenrekonstruktion. Unter Verwendung intrinsischer und extrinsischer Parameter wurde in diesem Abschnitt die Linsenverzerrung aus den Bildern herausgerechnet und das Stereobildpaar rektifiziert. Die vorverarbeiteten Bilder wurden dann zur Berechnung der Disparitätskarte und darauf aufbauende 3D-Visualisierung verwendet.

Im letzten Abschnitt dieses Kapitels wurden dann schließlich die realisierten Methoden der Selbstlokalisierung unter Verwendung der gewonnenen Informationen aus einer Kamera beziehungsweise des Stereokamerasystems behandelt.

Im nächsten Kapitel werden eine Kurzfassung der gesamten Arbeit und ein Ausblick über mögliche Verbesserungen und Erweiterungen gegeben.



In folgenden Abschnitten werden die Ergebnisse der vorliegenden Arbeit zusammengefasst sowie deren mögliche Verbesserungen und Erweiterungen vorgestellt und erläutert.

## 10.1 Fazit

Die vorliegende Arbeit beschäftigt sich mit der 3D-Rekonstruktion und deren Anwendung für landmarkenbasierte Selbstlokalisierung mobiler Roboter. Als Ausgangssystem standen zwei humanoide HOAP-2 Roboter der Fujitsu Automation Co. Ltd. zur Verfügung, die mit jeweils zwei Kameras ausgerüstet sind. Die Logitech Quickcams sind nicht synchronisiert und haben eine maximale Auflösung von  $324 \times 248$  Bildpunkte. In dieser Arbeit wurden die beiden Kameras des jeweiligen Roboters als ein Stereokamerasystem aufgefasst und als solches verwendet.

Die Entwicklung der eingesetzten Software richtete sich an der entworfenen und in der Einleitung präsentierten Architektur, die in der Abbildung 1.1 grafisch dargestellt wird.

Dabei wurden zuerst die beiden Kameras einzeln nach der Methode von Z. Zhang [Zha00] und unter Verwendung der OpenCV-Bibliothek kalibriert. Mit den bestimmten intrinsischen Parametern einzelner Kameras wurde unter Zuhilfenahme des implementierten Verfahrens nach Fusiello, Trucco und Verri [FTV00] die Kalibrierung des Stereokamerasystems durchgeführt.

Danach wurden die Stereobildpaare einer Szene mit den ermittelten Parametern einzelner Kameras sowie des Kamerasystems vorverarbeitet. Dafür wurden die Bilder entzerrt und mit den bei der Stereokalibrierung berechneten Transformationsmatrizen rektifiziert. Dieser Schritt erlaubt die Suche nach korrespondierenden Punkten auf eine Zeile einzuschränken, was die Effizienz des Verfahrens enorm steigert. Für die Korrespondenzanalyse wurde der Algorithmus von Birchfield und Tomasi [BT98] verwendet. Der Algorithmus wird wegen seiner Schnelligkeit des Öfteren

verwendet, außerdem reduziert die Suche nach dem globalen Minimum einer Zeile die Anzahl der nicht zugeordneten Punkte. Als Ergebnis liefert der Algorithmus eine Disparitätskarte aller gefundenen Stereobildpaare. Aus den bekannten Disparitäten korrespondierender Punkte wurde unter Zuhilfenahme der Triangulation die Entfernung zu dem jeweiligen Punkt im Raum geschätzt. Anschließend wurden die Szenen mit dem bekannten Abstand zu möglichst vielen einzelnen Punkten rekonstruiert.

Aus den bekannten Tiefen einer Szene und mindestens zwei Punkten, deren Positionen im Weltkoordinatensystem bekannt sind, wurde unter Verwendung des Strahlensatzes die Selbstlokalisierung des Roboters durchgeführt. Die Koordinaten der dafür benötigten Punkte im Bild wurden anhand von Landmarken ermittelt.

Dafür wurden nach der intensiven Recherche die Landmarken von D. Scharnstein und A. J. Briggs [SB99] und [SB01] ausgewählt. Der Vorteil dieser Landmarken liegt in dem vorhandenen mathematischen Modell und geringerer Unempfindlichkeit gegenüber den wechselnden Lichtverhältnissen. Mit dem Algorithmus von Scharnstein und Briggs wurden die Landmarken im Bild detektiert und deren Identitätsnummer anhand des Barcodes ausgelesen. Für die Feststellung der Identitätsnummer der Landmarke sucht der Algorithmus die Koordinaten zweier Punkte im Bild. Diese Punkte zusammen mit deren gespeicherten Weltkoordinaten wurden eingesetzt, um die Selbstlokalisierung des Roboters durchzuführen.

Damit entstand ein Verfahren der unter Verwendung der OpenCV-Bibliothek die Selbstlokalisierung mobiler Roboter ermöglicht. Des Weiteren können die entworfene Struktur und die bestimmte Tiefeninformation einer Szene für die Realisierung der gesamten Navigation eines Roboters genutzt werden. Außerdem stellt eine dreidimensional rekonstruierte Szene eine Grundlage für die Verwirklichung der Interaktion zwischen Mensch und Roboter dar. Im nächsten Abschnitt werden die möglichen Verbesserungen und Erweiterungen der vorliegenden Arbeit diskutiert.

### 10.2 Ausblick

In diesem Abschnitt werden einige Vorschläge über die möglichen Verbesserungen und Erweiterungen zu den in dieser Arbeit realisierten Methoden und Techniken gegeben. Dafür wird die gesamte Arbeit in drei große Themengebiete unterteilt

1. Kalibrierung einer Kamera beziehungsweise eines Stereokamerasystems
2. Tiefenrekonstruktion ausgehend von einem Stereokamerasystem
3. Selbstlokalisierung anhand passiver künstlicher Landmarken

Eine mögliche Verbesserung für die “Kalibrierung einer Kamera beziehungsweise eines Stereokamerasystems” wäre zum Beispiel die Transformationsmatrizen des

Stereokamerasystems aus der Mittlung mehrerer Ergebnisse zu berechnen. Diese Erweiterung würde eventuell bessere Kalibrierungsparameter liefern und somit die Zuordnung korrespondierender Punkte verbessern.

Eine Erweiterung der Kalibrierung um eine grafische Oberfläche würde eventuell den Prozess der Kalibrierung noch weiter vereinfachen und intuitiver gestalten, so dass auch Benutzer ohne Wissen der Hintergründe die Kalibrierung durchführen könnten. Des Weiteren könnte die Kalibrierung des Stereokamerasystems um ein vollautomatisches Verfahren ergänzt werden, so dass dieser Kalibrierungsschritt für den Benutzer vollständig entfallen würde. Eine vollautomatische Kalibrierung des Stereokamerasystem lässt sich mit den im theoretischen Teil behandelten Methoden realisieren.

In dem Themengebiet “Tiefenrekonstruktion ausgehend von einem Stereokamerasystem” könnten die implementierten Korrespondenzanalyseverfahren um weitere Entscheidungskriterien erweitert werden, so dass die Zuordnung korrespondierender Punkte eventuell qualitativ bessere Ergebnisse liefern würde. Als zusätzliches Entscheidungskriterium könnte zum Beispiel die Farbinformation der Bilder verwendet werden. Eine weitere Verbesserung wäre die Beschleunigung der Algorithmen durch Optimierung.

Zur Durchführung der Experimente ist der Block-Matching-Algorithmus nur mit SAD (engl. sum of absolute differences) implementiert worden. Dieser könnte um die weiteren Metriken aus Unterkapitel 5.2.2 ergänzt werden, so dass eventuell noch bessere Ergebnisse erzielt werden können. Entsprechend dem Themengebiet “Kalibrierung einer Kamera beziehungsweise eines Stereokamerasystems” könnte auch hier eine Erweiterung um eine grafische Oberfläche stattfinden.

Die Berechnung der Position des Roboters gegenüber der Landmarke ausgehend von einer Kamera wird unter der Voraussetzung durchgeführt, dass die Bildfläche der Kamera zu der Landmarke in vertikaler Richtung parallel verläuft. Diese Voraussetzung könnte eliminiert werden, wenn auf der Landmarke zwei weitere Punkte definiert werden, so dass diese ein Dreieck auf der Landmarke bilden. Werden diese vier Punkte im Bild der Kamera gefunden, so kann anhand dieser, den bekannten Koordinaten der entsprechenden Punkte in Weltkoordinaten und durch Kalibrierung bestimmten intrinsischen Parameter der Kamera die extrinsischen Parameter berechnet werden. Anhand dieser Parameter kann die Position des Roboters gegenüber der Landmarke bestimmt und schließlich in der Welt lokalisiert werden.



## Danksagung

An dieser Stelle möchten wir uns bei allen den Jenigen bedanken, die uns die Erstellung der vorliegenden Diplomarbeit ermöglichten.

Wir bedanken uns bei allen Mitarbeitern des AB TAMS, die uns während der Erstellung dieser Diplomarbeit unterstützten und uns den Start in das wissenschaftliche Arbeiten so leicht wie möglich gestalteten. Besonderer Dank gilt Daniel Westhoff, Markus Hüser und Sascha Jockel für die unzähligen Tipps und unermüdliche Hilfsbereitschaft sowie das Korrekturlesen der vorliegenden Arbeit.

Herrn Prof. Dr. J. Zhang danken wir für das, durch die Vergabe dieser Arbeit, entgegengebrachtes Vertrauen und zur Verfügung gestellte Arbeitsmittel.

Beim Herrn Dr. W. Hansmann bedanken wir uns ganz herzlich für die Übernahme des Zweitgutachtens.

Des Weiteren bedanken wir uns bei Dr. Andreas Mäder für das Einrichten und die Bereitstellung des nötigen Hard- und Software-Equipments. Auch Tatjana Tetsis gehört unserer Dank für die Korrekturen, Abwicklung unzähliger organisatorischer Sachen sowie ständige gute Laune und Aufmunterung während der Anfertigung dieser Arbeit.

Im Folgenden wird jeder von uns ein Paar Worte an die Menschen aus seiner privaten Umgebung richten.

Ich möchte mich bei meiner Frau Svetlana bedanken, die mich nicht nur durch die Höhen und Tiefen dieser Arbeit begleitete, sondern seit mehreren Jahren mit mir das Leben meistert. Nicht zuletzt waren ihre Geduld, Verständnis und unermüdlicher Glaube an mich die Faktoren, die mir das Erstellen dieser Arbeit möglich machten. Ferner bedanke ich mich bei meinen Eltern für ihren Zuspruch und seelischen Beistand während der Diplomarbeit sowie ihrer finanziellen Unterstützung meines gesamten Studiums.

An dieser Stelle möchte ich meiner Frau Wiktoria Stroh und unserer Tochter Viviane dafür herzlich danken, dass sie im Laufe des Studiums und insbesondere während der Zeit, in der die vorliegende Diplomarbeit entstand, mich in jeglicher Hinsicht unterstützten. Des Weiteren möchte ich meinen Eltern Sina und Waldemar Stroh für die finanzielle Unterstützung danken, durch diese mir das Studium überhaupt ermöglicht wurde. Ein besten Dank an Denis Klimentjew für die sehr gute Zusammenarbeit während der gesamten Diplomarbeit und alle meine Freunde für die ich wenig Zeit aufbringen konnte.





## Literaturverzeichnis

- [AH93] ATIYA, S. ; HAGER, G.: Real-time Vision-based Robot Localization. In: *IEEE Transactions on Robotics and Automation*, Vol. 9, No. 6, pp. 785-800 (1993)
- [Bak81] BAKER, H. H.: Depth from Edge and Intensity Based Stereo. (1981)
- [Bau07] BAUMANN, E.: Fotografie-Informationen. (2007). <http://www.elmar-baumann.de>
- [BBH03] BROWN, M. Z. ; BURSCHKA, D. ; HAGER, G. D.: *Advances in Computational Stereo*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Bd. 25, S. 993 – 1008, 2003
- [BEFW96] BORENSTEIN, J. ; EVERETT, H. ; FENG, L. ; WEHE, D.: Navigation Mobile Robots-Systems and Techniques. In: *Journal of Robotic Systems, Special Issue on Mobile Robots*, pp. 231 – 249 14 (1996)
- [BEK<sup>+</sup>04] BRUNS, B. ; EICHNER, C. ; KÖWING, A. ; MAHN, M. ; MÜLLER, F. ; OLTMANN, C. ; PEI, J. ; PFEIFER, L. ; PÖPKEN, J. ; SCHWARZE, K.: Endbericht, Entwicklung eines Indoor Micro Air Vehicles, Carl von Ossietzky Universität Oldenburg. (2004). [http://www-is.informatik.uni-oldenburg.de/~dibo/pg\\_fb10/endberichte/pg-mav.pdf](http://www-is.informatik.uni-oldenburg.de/~dibo/pg_fb10/endberichte/pg-mav.pdf)
- [Bel57] BELLMAN, S. E.: Dynamic programming. (1957)
- [Bla04] BLASCHEK, R.: Tiefeninformationsgewinnung aus Stereobildern. (2004)
- [BP06] BISTRY, H. ; PÖHLSSEN, S.: *Entwicklung eines Eingebetteten Systems zur ressourcenschonenden und plattformunabhängigen Anbindung von SICK-Lasermesssystemen*. Diplomarbeit am Arbeitsbereich TAMS der Universität Hamburg, 2006
- [BSS<sup>+</sup>07] BOUGUET, J-Y. ; STROBL, K. ; SEPP, W. ; PAREDES, C. ; ARBTER, K.: Camera Calibration Toolbox for Matlab. (2007). [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- [BT98] BIRCHFIELD, S. ; THOMASI, C.: Depth Discontinuities by Pixel-to-Pixel Stereo. (1998)
- [CIN07] CINACS: Cross-Modal Interaktion in Natural and Artificial Cognitive Systems,. (2007). <http://cinacs.informatik.uni-hamburg.de/content/view/80/87/>

- [CZL07] CHEN, S. Y. ; ZHANG, J. ; LI, Y. F.: Vision Processing for Realtime 3D Data Acquisition Based on Coded Structured Light. In: *IEEE International Conference on Robotics and Automation, Roma, Italy* 10 (2007)
- [EPK06] ELIAS, B. ; PAELKE, V. ; KUHN, S.: Kartographische Visualisierung von Landmarken. In: *Aktuelle Entwicklungen in Geoinformation und Visualisierung, GEOVIS 2006, 5./6. April 2006, Potsdam, Kartographische Schriften Band 10, Kirschbaum Verlag* 10 (2006)
- [FAC03] FUJITSU AUTOMATION Co., Ltd.: HOAP-2 Instruction Manual. (2003). <http://jp.fujitsu.com/group/automation/downloads/en/services/humanoid-robot/hoap2/instructions.pdf>
- [Fau95] FAUGERAS, O.: Stratification of 3-Dimensional Vision: Projective, Affine and Metric Representations. In: *Journal of the Optical Society of America* 19 (1995)
- [FB81] FISCHLER, M. A. ; BOLLES, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. (1981)
- [FFH<sup>+</sup>92] FAUGERAS, O. ; FUA, P. ; HOTZ, B. ; MA, R. ; ROBERT, L. ; THONNAT, M. ; ZHANG, Z.: *Quantitative and Qualitative Comparison of some Area and Feature-Based Stereo Algorithms*. Wichmann, Karlsruhe, 1992
- [FSM07] FSMLABS: FSMLabs Real-Time Super Cluster. (2007). <http://www.fsmlabs.com/>
- [FSMB98] FRANZ, M. O. ; SCHÖLKOPF, B. ; MALLOT, H. A. ; BÜLTHOFF, H. H.: Where did I take that snapshot? Scene-based homing by image matching. (1998)
- [FTV97] FUSIELLO, A. ; TRUCCO, E. ; VERRI, A.: Rectification with unconstrained stereo geometry. (1997)
- [FTV00] FUSIELLO, A. ; TRUCCO, E. ; VERRI, A.: A compact algorithm for rectification of stereo pairs. In: *Machine Vision and Applications(2000) 12, Springer-Verlag, pp. 16-22* (2000)
- [GH01] GRUEN, A. ; HUANG, T. S.: *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag Berlin Heidelberg, 2001
- [GL96] GOLUB, G. ; LOAN, C. van: *Matrix Computations*. Hopkins University Press, Baltimore, Maryland, 3 edition, 1996
- [Goe02] GOEHRING, D.: Digitalkameratechnologien. (2002). <http://www2.informatik.hu-berlin.de/~meffert/Seminararbeiten/Weitere/Cmos/Ccd-cmos.pdf>

- [Har97] HARTLEY, R.: Self-Calibration of Stationary Cameras. (1997)
- [Har98] HARTLEY, R.: Theory and Practice of Projective Rectification. (1998). <http://users.rsise.anu.edu.au/~hartley/Papers/joint-epipolar/journal/joint3.pdf>
- [Her04] HERRMANN, H.: Ein System zur schnellen Entwicklung von Bildverarbeitungsalgorithmen, Doktorarbeit an der Universität Mannheim. (2004). [http://deposit.ddb.de/cgi-bin/dokserv?idn=972137300&dok\\_var=d1&dok\\_ext=pdf&filename=972137300.pdf](http://deposit.ddb.de/cgi-bin/dokserv?idn=972137300&dok_var=d1&dok_ext=pdf&filename=972137300.pdf)
- [Hor06] HORNBERG, A.: *Handbook of Machine Vision*. WILEY-VCH Verlag GmbH & Co. KGaA, 2006
- [HS88] HARRIS, C. ; STEPHENS, M.: A combined corner and edge detector. (1988)
- [IPP07] IPP: Intel® Integrated Performance Primitives. (2007). <http://www.intel.com/software/products/ipp/ippvm20/index.htm>
- [KKS96] KLETTE, R. ; KOSCHAN, A. ; SCHLÜNS, K.: *Computer Vision*. Friedrich Vieweg & Sohn Verlagsgesellschaft mbH, 1996
- [LDW91] LEONARD, J. J. ; DURRANT-WHYTE, H. F.: *Mobile robot localization by tracking geometric beacons*. 1991
- [LH81] LONGUET-HIGGINS, H. C.: A Computer Algorithm for Reconstructing a Scene from Two Projections. (1981)
- [Luh00] LUHMANN, T.: *Nachbereichsphotogrammetrie*. Wichmann, Heidelberg, 2000
- [Min07] MINOLTA, Konica: Minolta. (2007). <http://www.konicaminolta3d.com/index.php?id=9&L=1>
- [Mor77] MORE, J.: *The Levenberg-Marquard algorithm, Implementation and Theory*. Springer-Verlag, 1977
- [Müh02] MÜHLMANN, K.: Design und Implementierung eines Systems zur schnellen Rekonstruktion dreidimensionaler Modelle aus Stereobildern, Dissertation. (2002). [http://deposit.ddb.de/cgi-bin/dokserv?idn=966293746&dok\\_var=d1&dok\\_ext=pdf&filename=966293746.pdf](http://deposit.ddb.de/cgi-bin/dokserv?idn=966293746&dok_var=d1&dok_ext=pdf&filename=966293746.pdf)
- [Ope07] OPENCV: Intel® Open Source Computer Vision Library. (2007). <http://sourceforge.net/projects/opencvlibrary>
- [Ren04] RENNER, E.: *Pinhole Photography - Rediscovering a Historic Technique*. Third Edition, Focal Press, Boston, Oxford, 2004
- [Rie04] RIESNER, S.: Autonome Mobile Roboter, Dr. Stefan Riesner, TU München, Fakultät für Informatik,. (2004). <http://atknoll11.informatik.tu-muenchen.de:8080/tum6/lectures/>

courses/ss04/robotik/documents/Documents/1089189231.72/  
robotik2004\_06.pdf

- [Rot02] ROTH, J.: *Mobile Computing Grundlagen, Technik, Konzepte*. Dpunkt Verlag, 2002
- [RS88] RANKA, S. ; SAHNI, S.: Image Template Matching on SIMD Hypercube Multicomputers. In: *Proceedings of International Conference on Parallel Processing, volume 3, pp. 84-91. Penn State University Press* 10 (1988)
- [RS90] RANKA, S. ; SAHNI, S.: Image template matching on MIMD hypercube multicomputers. In: *J. Parallel Distrib. Comput., 10(1), pp. 79-84* 10 (1990)
- [SB99] SCHARSTEIN, D. ; BRIGGS, A.: Fast Recognition of Self-Similar Landmarks. In: *In Workshop on Perception for Mobile Agents, June 1999, pp. 74-81* (1999)
- [SB01] SCHARSTEIN, D. ; BRIGGS, A.: Real-time recognition of self-similar landmarks. In: *Image Vision Comput., 2001, pp. 763-772* (2001). [http://dx.doi.org/10.1016/S0262-8856\(00\)00105-0](http://dx.doi.org/10.1016/S0262-8856(00)00105-0)
- [SBK00] SCHREER, O. ; BRANDENBURG, N. ; KAUFF, P.: Disparitätsanalyse konvergenter und rektifizierter Stereoansichten: Eine vergleichende Untersuchung. (2000)
- [Sch05a] SCHNELLE, D.: Sensoren, ein Kompendium der Klasse PF3A. (2005). <http://www.linuspithan.de/schule/MST/Messtechnikkompendiumsensoren.pdf>
- [Sch05b] SCHREER, O.: *Stereoanalyse und Bildsynthese*. Springer-Verlag Berlin Heidelberg, 2005
- [Som05] SOMIESKI, K.: Aufbau eines Forschungsplatzes für die geometrische Stereoanalyse. (2005)
- [SS02] SCHARSTEIN, D. ; SZELISKI, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. (2002)
- [Sta07] STANKE, G.: *Bildanalytisch gestützte Messeinrichtung für die Montage von Fertigteilen (BiMes), Sachbericht, Projektlaufzeit 01.08.2004 - 31.07.2006*. Gesellschaft zur Förderung angewandter Informatik e. V. (GFaI) <http://www.gfai.de/forschungsbereiche/bv/bimes.pdf>
- [Stü04] STÜRZL, W.: *Sensorik und Bildverarbeitung für Landmarken-basierte Navigation*. Eberhard Karls Universität Tübingen, 2004
- [Sun03] SUN, C.: Uncalibrated Three-View Image Rectification. (2003)
- [SW75] SIEGEL, A. ; WHITE, S.: The Development of Spatial Representations of Large-Scala Enviromments. In: *Child Development and Behavior* 10, pp. 9-55 10 (1975)

- [Tip94] TIPLER, P.: *Physik*. Spektrum Akademischer Verlag, Heidelberg, 1994
- [Tsa86] TSAI, R.: An efficient and accurate camera calibration technique for 3D machine vision. In: *Proc International Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA 10* (1986)
- [Tsa87] TSAI, R.: A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses. In: *IEEE JOURNAL OF ROBOTICS AND AUTOMATION* 21 (1987)
- [VWL+00] VELOSO, M. ; WIENER, E. ; LENSER, S. ; BRUCE, J. ; BALCH, T.: Vision-Servoed Localization and Behavior-Based Planning for an Autonomous Quadruiped Legged Robot, School of Computer Science Carnegie Mellon University Pittsburgh. (2000). <http://www.cs.cmu.edu/~coral/publinks/mmv/aips00-cmtrio.pdf>
- [War98] WARD, A.: Sensor-driven Computing. PhD thesis, University of Cambridge. (1998). <http://citeseer.ist.psu.edu/ward99sensordriven.html>
- [Web02] WEBER, J.: Globale Selbstlokalisierung für mobile Service Roboter, Doktorarbeit an der Universität Kaiserslautern. (2002). <http://ag-vp-www.informatik.uni-kl.de/Projekte/CAROL/papers/Diss.pdf>
- [WKBM+97] WERNER, S. ; KRIEG-BRÜCKNER, B. ; MALLOT, H. ; SCHWEIZER, K. ; FREKSA, C.: *Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation*. In W. Brauer , Informatik '97. Informatik als Innovationsmotor, TG 27 in Informatik aktuell Berlin. GI - Gesellschaft für Informatik, Springer Verlag, 1997
- [WW86] WINSTEL, G. ; WEYRICH, C.: *Optoelektronik*. Springer Verlag, Berlin, 1986
- [YK01] YOON, K-J. ; KWEON, I-S.: Landmark design and real-time landmark tracking for mobile robot localization. (2001). [http://rcv.kaist.ac.kr/publication/file/foreign\\_conference/59\\_KukJinYoon\\_SPIE2001.pdf](http://rcv.kaist.ac.kr/publication/file/foreign_conference/59_KukJinYoon_SPIE2001.pdf)
- [Zha98] ZHANG, Z.: A Flexible New Technique for Calibration. (1998)
- [Zha00] ZHANG, Z.: A flexible new technique for camera calibration. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000)
- [Zha05] ZHANG, J.: *Angewandte Sensorik*. Vorlesung am Arbeitsbereich TAMS der Universität Hamburg, 2005



## **Aufteilung der Gruppenarbeit**

Da diese Arbeit von Denis Klimentjew und Andre Stroh gemeinsam erstellt wurde, werden im Folgenden einzelne Kapitel dem jeweiligen Autor zugeordnet.

Vom Denis Klimentjew, Matr.- Nr.: 5422349, erstellte Kapitel: 1.1, 1.3, 1.4, 2, 3, 7, 8, 9.1, 10.1

Vom Andre Stroh, Matr.- Nr.: 5408478, erstellte Kapitel: 1.2, 4, 5, 6, 9.2 - 9.3, 10.2

Die hier nicht aufgeführten Textpassagen, genauso wie der gesamte Quellcode wurde in Zusammenarbeit erstellt beziehungsweise programmiert.

## Erklärung

Ich, Denis Klimentjew, Matr.- Nr.: 5422349, versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Fachbereichs einverstanden.

(Ort, Datum)

(Unterschrift)

Ich, Andre Stroh, Matr.- Nr.: 5408478, versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Fachbereichs einverstanden.

(Ort, Datum)

(Unterschrift)