

Ansichtenauswahl für die 3-D-Rekonstruktion statischer Szenen

Der Technischen Fakultät der
Universität Erlangen-Nürnberg

zur Erlangung des Grades

DOKTOR-INGENIEUR

vorgelegt von

Stefan Wenhardt

Erlangen – 2013

Als Dissertation genehmigt von der
Technischen Fakultät der
Universität Erlangen-Nürnberg

Tag der Einreichung:	02.11.2012
Tag der Promotion:	17.06.2013
Dekan:	Prof. Dr.-Ing. Marion Merklein
Berichterstatter:	Prof. em. Dr.-Ing. Heinrich Niemann Prof. Dr. Günther Greiner

Danksagung

Die vorliegende Arbeit fasst die Ergebnisse meiner wissenschaftlichen Tätigkeit am Lehrstuhl für Mustererkennung der Technischen Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg zusammen. An dieser Stelle möchte ich mich bei allen bedanken, die mich während der vielen Jahre am Lehrstuhl und danach bei der Erstellung dieser Arbeit unterstützt haben.

Mein spezieller Dank gilt Herrn Prof. em. Dr.-Ing. Dr.-Ing. h.c. Heinrich Niemann, der mir die Möglichkeit gab und das Vertrauen entgegen brachte unter seiner Leitung den Themenkreis der Dissertation wissenschaftlich zu bearbeiten. Dabei möchte ich mich bei Prof. Niemann insbesondere dafür bedanken, dass er nachdem er 2005 emeritierte, die Möglichkeit gab die Arbeit unter seiner Leitung und Betreuung zu beenden. Dabei war die Zeit nach 2007, als ich den Lehrstuhl verlassen hatte, zunehmend schwierig. Trotzdem hatte Prof. Niemann die notwendige Geduld und gab mir die Zeit die Arbeit fertigzustellen.

Weiterhin möchte ich mich bei Herrn Prof. Dr. Günther Greiner für die Übernahme des Zweitgutachtens bedanken.

Im Besonderen möchte ich mich bei Prof. Dr.-Ing. Joachim Denzler bedanken. Er hatte die initiale Idee dieses Thema im Rahmen einer Dissertation zu bearbeiten. Auch als er seinem Ruf nach Passau und später nach Jena folgte, stand er für mich weiterhin als Ansprechpartner mit fachlichen Ratschlägen zur Seite.

Ganz herzlich möchte ich mich bei Prof. Dr.-Ing. Jochen Schmidt für die gemeinsame Zeit zunächst bei der Entwicklung des Kurses „Rechnersehen mit Anwendungen in der Augmented Reality sowie beim bildbasierten Rendering“ für die virtuelle Hochschule Bayern (vhb) bedanken. Später stand er mir als wissenschaftlicher Leiter der Bildverarbeiter-Gruppe am Lehrstuhl für Mustererkennung und auch danach mit vielen wertvollen Tipps zur Verfügung.

Bedanken möchte ich mich für die Unterstützung zahlreicher Kollegen am Lehrstuhl, insbesondere Dr.-Ing. Ingo Scholz und Dr.-Ing. Florian Vogt für die Hilfestellungen mit IGF3, bei Friedrich Popp für die Unterstützung mit dem Scrobot ER VII und dem Drehteller sowie meinen ehemaligen Bürokollegen Dr.-Ing. Rainer Deventer, Dr.-Ing. Andreas Maier, Benjamin Deutsch und Christian Derichs. Insbesondere haben sich durch die fachliche Nähe der Arbeitsgebiete mit Benjamin Deutsch und Christian Derichs zahlreiche fruchtbare Diskussionen ergeben, deren Resultate teilweise als Artikel veröffentlicht wurden. Ferner gilt der Dank meinen Korrekturlesern: Jochen Schmidt, Rainer Deventer, Joachim Denzler, Sabine Wenhardt, Ute Schramm und meinen Eltern.

Zu guter Letzt möchte ich mich bei meiner Familie bedanken. Bei meinen Eltern die mich immer unterstützt und das notwendige Vertrauen entgegengebracht haben. Sowie bei meiner Frau Sabine, die mich in meinem Vorhaben immer wieder ermuntert und die mir die Fertigstellung der Arbeit an unzähligen Wochenenden ermöglicht hat.

Stefan Wenhardt

Kurzfassung

Die 3-D-Rekonstruktion eines Objekts aus zwei oder mehr Intensitätsbildern gehört zu den klassischen Problemstellungen aus dem Bereich Rechnersehen. Kann man dabei aktiv in das System eingreifen, also gezielt auf die Parameter der Aufnahmen einwirken, so spricht man vom „aktiven Sehen“. Die aktiven Parameter sind hier die Brennweite, die durch ein Zoomobjektiv gesteuert werden kann, sowie Position, d. h. Translation und Rotation, der Kamera. Die Kamera wird dabei durch einen Roboterarm in die gewünschte Position gebracht.

Methoden des aktiven Sehens im Bereich der 3-D-Rekonstruktion befassen sich üblicherweise entweder mit dem Problem der Vollständigkeit der Rekonstruktion, im Sinne einer Szenenexploration, oder mit einer möglichst präzisen Bestimmung der 3-D-Geometrie. Ferner gibt es Ansätze die einen Kompromiss zwischen diesen beiden Zielsetzungen herstellen. In der vorliegenden Arbeit werden Methoden des aktiven Sehens benutzt, um eine möglichst präzise 3-D-Rekonstruktion zu berechnen. Hierzu werden aus den Aufnahmen Merkmalspunkte extrahiert und deren 3-D-Koordinaten anschließend geschätzt. In dieser Arbeit werden dazu zwei unterschiedliche Ansätze entwickelt und evaluiert: ein geometrischer Ansatz für Stereosysteme und ein probabilistischer Ansatz.

Der geometrische Ansatz betrachtet Stereosysteme, also Systeme, die aus genau zwei Kameras bestehen. Der Einfluss der aktiven Parameter (Translation, Rotation und Brennweite) auf die Genauigkeit der 3-D-Rekonstruktion wird systematisch untersucht und, wo es möglich ist, analytisch bewiesen.

Im probabilistischen Ansatz wird, basierend auf der aktuellen Schätzung der 3-D-Koordinaten, die nächste beste Ansicht ermittelt, d. h. die, die die Präzision der Schätzung bestmöglich erhöht. Hierzu wird die 3-D-Rekonstruktion als Zustandsschätzproblem formuliert, welches mit Hilfe des erweiterten Kalman-Filters gelöst wird. Dadurch ist es möglich, basierend auf der aktuellen Schätzung, weitere Aufnahmen zu machen und in die Schätzung zu integrieren. In dieser Arbeit wird bewiesen, dass sich der Rechenaufwand der Kalman-Filter-Gleichungen durch geeignete Annahmen drastisch reduzieren lässt. Es wird dabei diskutiert, warum diese Annahmen für die 3-D-Rekonstruktion sinnvoll sind.

Ein zu rekonstruierender 3-D-Punkt eines Objekts ist im Allgemeinen nicht von jeder beliebigen Ansicht aus zu sehen. Deshalb muss die Sichtbarkeit bei der Planung berücksichtigt werden. Zum einen wird in dieser Arbeit gezeigt, wie man die Wahrscheinlichkeit, ob ein Punkt in einer bestimmten Ansicht sichtbar ist, berechnen kann. Zum anderen wird bewiesen, wie sich die Sichtbarkeitsbedingung geschlossen in das Optimierungskriterium für eine neue Ansicht integrieren lässt.

Für die Ansichtenplanung ist schließlich entscheidend, die Kamera mit Hilfe eines Roboters an der gewünschten Stelle zu positionieren. Deshalb stellt sich die Frage, welche Positionen von dem verwendeten Roboterarm erreicht werden können. In der einschlägigen Literatur zur Ansichtenplanung für die 3-D-Rekonstruktion wird diese Fragestellung entweder gar nicht behandelt oder stets angenommen, dass sich die möglichen Kamerapositionen auf einer (Teil-) Kugel um das Objekt herum befinden. Im Gegensatz dazu wird hier mit Hilfe der sog. Denavit-Hartenberg-Matrix der Arbeitsbereich eines Roboterarms vollständig beschrieben. So können alle Positionen, die von dem jeweils verwendeten Roboterarm erreichbar sind, mit in die Ansichtenplanung einbezogen werden. Eine Beschränkung auf eine (Teil-)Kugel ist in diesem Fall dann unnötig.

Abstract

The problem of 3-D reconstruction is one of the main topics in computer vision. If certain imaging parameters can be modified to improve the 3-D reconstruction result, the question how to select these parameters belongs to a domain called active vision. The active parameters in our case are the focal length of the camera, which can be controlled by a zoom lens, and the pose, i. e. translation and rotation of the camera. The camera is mounted on a robot, so the position of the camera can be controlled.

Usually, active vision for 3-D reconstruction means either scene exploration or most accurate estimation of the 3-D structure of an object. Of course, there are approaches trying to find a trade-off of the two aspects. This thesis focuses only on the aspect of high accurate estimates. For this purpose feature points are extracted from the images to estimate their 3-D coordinates. Here two different approaches are developed and evaluated: a geometric approach for stereo camera systems and a probabilistic approach.

The geometric approach considers only stereo camera systems, i. e. systems which consist of exactly two cameras. The influence of the active parameters (translation, rotation and focal length) are evaluated and, if possible, analytically proven.

The probabilistic approach determines the next best view, to increase the accuracy of the current estimate. Therefore it is necessary to describe the problem of 3-D reconstruction as a state estimation problem. The state estimation is solved by the extended Kalman filter. So, it is possible to improve the current state estimate of the 3-D coordinates by additional observations. This thesis derives a modification of the Kalman filter, which allows to reduce the calculation complexity drastically. For this modification only some simple assumptions are necessary, but it is discussed, why these assumptions are meaningful in the application of 3-D reconstruction. This modification is exact, i. e. there is no approximation required.

A 3-D point to be reconstructed may be invisible, e. g. it is occluded by the object itself or its projection is outside of the field of view of the camera. Therefore, the next best view planning has to consider whether the point is visible or not from a certain view. We will show how the probability of visibility of a 3-D point can be calculated and further, how we can integrate the visibility issue into the closed-form optimization criteria for the next best view planning.

Another aspect of next best view planning is to move the camera to the desired position. Therefore, the question is, which position is reachable by the used robot device the camera is mounted on. In former publications, this aspect is either ignored or it is assumed the camera can move only on a (part-)sphere around the object. But this thesis describes the reachable workspace by the Denavit-Hartenberg matrix. This allows to consider the complete workspace of the used robot in the next best view planning, without any unnecessary limitation to a (part-)sphere.

Inhaltsverzeichnis

1	Einführung	1
1.1	Problemstellung und Motivation	3
1.1.1	Aufnahmemodalitäten	3
1.1.2	Wozu Ansichtenplanung	3
1.2	Stand der Kunst	6
1.2.1	3-D-Rekonstruktion mit Tiefenscannern	7
1.2.2	3-D-Rekonstruktion aus Intensitätsbildern	9
1.3	Beitrag der Arbeit	11
1.3.1	Geometrischer Ansatz	11
1.3.2	Probabilistischer Ansatz	11
1.4	Aufbau der Arbeit	12
2	3-D-Rekonstruktion	15
2.1	Kameraparameter und Projektionsmodelle	15
2.1.1	Koordinatensysteme	16
2.1.2	Intrinsische Kameraparameter	16
2.1.3	Extrinsische Kameraparameter	17
2.1.4	Projektionsmodelle	18
2.2	Kamerakalibrierung	22
2.2.1	Schätzung der Homographie	23
2.2.2	Bestimmung der Kameraparameter aus der Homographie . .	24
2.2.3	Nichtlineare Optimierung	26
2.3	Hand-Auge-Transformation	27
2.4	Direkte und inverse Kinematik	31
2.4.1	Direkte Kinematik	31
2.4.2	Inverse Kinematik	34
2.5	Triangulation und Bündelausgleich	36
2.5.1	Punktdetektion und Punktverfolgung	36
2.5.2	Triangulation	38
2.5.3	Bündelausgleich	40
3	Grundlagen für die Auswahl von Ansichten	43
3.1	Allgemeine geometrische Ansätze	44
3.1.1	Structure-from-controlled-motion	44
3.1.2	Shape-from-occluding-contour	46
3.2	Ein geometrischer Ansatz für Stereosysteme	48
3.2.1	Zweidimensionales Modell	49
3.2.2	Dreidimensionales Modell	58
3.3	Probabilistische Ansätze	64

4	Probabilistische 3-D-Rekonstruktion aus einer Folge von Einzelbildern	71
4.1	Zustandsschätzung statischer und dynamischer Systeme	72
4.1.1	Zustandsschätzer für statische Systeme	73
4.1.2	Zustandsschätzer für dynamische Systeme	74
4.2	Integration neuer Ansichten mit dem erweiterten Kalman-Filter . . .	78
4.2.1	3-D-Rekonstruktion mit Hilfe des erweiterten Kalman-Filters	79
4.2.2	Das sequentielle Kalman-Filter und eine weitere Modifikation	80
4.3	Auswahlkriterien für neue Ansichten	84
4.3.1	Die Entropie als Gütemaß (D-Optimalität)	85
4.3.2	Modifizierte E-Optimalität	88
4.3.3	T-Optimalität	90
4.4	Berücksichtigung von Nebenbedingungen für die Sichtbarkeit	91
4.4.1	Integration der Nebenbedingung in das Optimierungskriterium	91
4.4.2	Berechnung der Wahrscheinlichkeit für die Sichtbarkeit	93
4.5	Zusammenfassung	97
5	Ein System zur Realisierung des probabilistischen Ansatzes	99
5.1	Beschreibung der Roboter	99
5.2	Softwarekomponenten	102
5.2.1	Initiale 3-D-Rekonstruktion	102
5.2.2	Planung der nächsten besten Ansicht	103
5.2.3	3-D-Rekonstruktion	104
5.3	Das Gesamtsystem	105
6	Experimente	107
6.1	Gütekriterien der Rekonstruktionen	107
6.1.1	Kalibriermuster	108
6.1.2	Objekte mit einer planaren Fläche	108
6.1.3	Objekte mit mehreren planaren Flächen	109
6.1.4	Objekte mit nichtplanaren Flächen	109
6.2	Evaluierung des geometrischen Ansatzes mit Stereosystemen	110
6.2.1	Ergebnisse Kalibriermuster	110
6.2.2	L-förmiges konkaves Objekt	118
6.2.3	L-förmiges konvexes Objekt	130
6.3	Experimente zum probabilistischen Ansatz an einfachen Objekten . .	130
6.3.1	Ergebnisse Kalibriermuster	132
6.3.2	Ergebnisse Mousepad	133
6.3.3	L-förmiges konkaves Objekt	137
6.3.4	Weihnachtsmann	142
6.4	Experimente zum probabilistischen Ansatz an komplexen Objekten .	144
6.4.1	L-förmiges konvexes Objekt	144
6.4.2	Tetraeder	147
6.4.3	Prüfkörper	150
6.5	Zusammenfassung und Diskussion der Ergebnisse	152
7	Zusammenfassung und Ausblick	155
7.1	Ausblick	155
7.2	Zusammenfassung	158

A	Mathematische Symbole	165
B	Herleitung der Fehlerfunktion für das 2-D-Modell aus Abschnitt 3.2	171
C	Probabilistischer Ansatz für die Optimierung der Rotation im 2-D	177
D	L-förmiges konvexes Objekt	179
	Literaturverzeichnis	187
	Index	201

Verzeichnis der Bilder

1.1	Beispiel für Erweiterte Realität	2
1.2	Lichtfelder bei endoskopischen Eingriffen	2
1.3	PMD-Kamera und Streifenprojektor	4
1.4	Lochkameramodell und Triangulation	4
1.5	Triangulation aus reiner Rotationsbewegung	5
1.6	Triangulation bei genau gegenüberliegenden Kameras	6
1.7	Triangulation bei unterschiedlichen Konfigurationen	6
2.1	perspektivische Projektion	19
2.2	paraperspektivische Projektion	20
2.3	Kalibriermuster	23
2.4	Roboterarm mit Kamera	27
2.5	Hand-Auge-Transformation	28
2.6	Gelenke eines Roboters	33
2.7	Gelenke eines Roboters	35
2.8	Triangulation bei windschiefen Geraden	38
3.1	Shape from Occluding Contour	46
3.2	Typen von Oberflächen	47
3.3	Normiertes Stereosystem	48
3.4	Stereosystem maßstabsgetreu	51
3.5	Optimierung der Stereobasis	52
3.6	Optimierung der Translation	53
3.7	Fehler in Abhängigkeit der Translation	54
3.8	Fehlerverlauf bei Verschiebung der rechten Kamera in x -Richtung	55
3.9	Stereosystem mit rotierter Kamera	55
3.10	Fehler in Abhängigkeit des Rotationswinkels	56
3.11	Anfangskonfiguration im 3-D-Modell	59
3.12	Endkonfiguration bei Optimierung der Rotation	62
3.13	Endkonfiguration bei Optimierung aller Parameter	63
3.14	Doppelverhältnis bei einem Geradenbüschel	67
3.15	Doppelverhältnis - invariant gegenüber projektiven Transformationen	68
4.1	Höhenlinien	88
4.2	Kovarianzmatrix als Ellipse	88
4.3	Varianten des E-Kriteriums	89
4.4	Berechnung der Sichtbarkeitswahrscheinlichkeit durch eingeschränkten Sichtbereich	94
4.5	Integration in baryzentrischen Koordinaten	96
5.1	Diverse Roboterarme	101
5.2	Arbeitsbereich des Scorbot	101

5.3	Gesamtsystem	106
6.1	Stereorekonstruktion eines Kalibriermusters in Abhängigkeit der Brennweite	112
6.2	Kalibriermuster mit unterschiedlichen Brennweiten	113
6.3	Stereorekonstruktion eines Kalibriermusters in Abhängigkeit der Translation	114
6.4	Stereorekonstruktion eines Kalibriermusters in Abhängigkeit der Translation	115
6.5	Stereorekonstruktion eines Kalibriermusters in Abhängigkeit der Rotation	115
6.6	Optimale Position für die Rekonstruktion des Kalibriermusters	118
6.7	Rekonstruktion des Kalibriermusters	119
6.8	Kalibriermuster aus den optimalen Ansichten	119
6.9	L-förmiges konkaves Objekt	120
6.10	3-D-Fehler in Abhängigkeit der Brennweite	122
6.11	Fehlermaß in Abhängigkeit der Brennweite	122
6.12	3-D-Fehler in Abhängigkeit der Translation	123
6.13	Fehlermaß in Abhängigkeit der Translation	124
6.14	3-D-Fehler in Abhängigkeit der Translation	124
6.15	Fehlermaß Abhängigkeit der Translation	125
6.16	3-D-Fehler in Abhängigkeit der Rotation	125
6.17	Fehlermaß in Abhängigkeit der Rotation	126
6.18	Fehler der Stereorekonstruktion des L-förmigen konkaven Objekts in Abhängigkeit des Radius bei Kameraanordnung auf einer Halbkugel	129
6.19	L-förmiges konvexes Objekt	130
6.20	Simulation einer Kalibriermuster-Rekonstruktion	132
6.21	Kamerapositionen nach jedem Optimierungsschritt	134
6.22	Simulation einer Kalibriermuster-Rekonstruktion	135
6.23	Simulation einer Kalibriermuster-Rekonstruktion	135
6.24	Mousepad auf dem Drehteller	137
6.25	Rekonstruktion eines Mousepads	138
6.26	Kamerapositionen nach jedem Optimierungsschritt	139
6.27	Iterationen beim E-Kriteriums	140
6.28	Rekonstruktion eines Mousepads	141
6.29	Rekonstruktion eines Mousepads	141
6.30	Rekonstruktion des L-förmigen konkaven Objekts	142
6.31	Weihnachtsmann	142
6.32	Rekonstruktion des Weihnachtsmanns	143
6.33	Rekonstruktion des L-förmigen konvexen Objekts	146
6.34	Tetraeder	148
6.35	Rekonstruktion des Tetraeders	149
6.36	Prüfkörper	150
6.37	Rekonstruktion eines Prüfkörpers	151
C.1	Eigenschaften des E-Kriteriums	178
D.1	3-D-Fehler in Abhängigkeit der Brennweite	180
D.2	Fehlermaß in Abhängigkeit der Brennweite	180
D.3	3-D-Fehler in Abhängigkeit der Translation	181

D.4 Fehlermaß in Abhängigkeit der Translation	182
D.5 3-D-Fehler in Abhängigkeit der Translation	182
D.6 Fehlermaß Abhängigkeit der Translation	183
D.7 3-D-Fehler in Abhängigkeit der Rotation	183
D.8 Fehlermaß in Abhängigkeit der Rotation	184
D.9 Fehler der Stereorekonstruktion des L-förmigen konvexen Objekts in Abhängigkeit des Radius bei Kameraanordnung auf einer Halbkugel . .	184

Verzeichnis der Tabellen

3.1	Übersicht veränderbare Parameter und resultierende Fehler	64
5.1	Denavit-Hartenberg-Parameter für den Scrobot	100
5.2	Denavit-Hartenberg-Parameter für den Drehteller	100
6.1	3-D-Fehler in Abhängigkeit der Brennweite	111
6.2	Zuordnung der x - z -Position zur Bildnummer	114
6.3	Der 3-D-Fehler in Abhängigkeit der Rotation der linken und rechten Kamera.	116
6.4	3-D-Fehler in Abhängigkeit der Brennweite	121
6.5	Fehlermaß in Abhängigkeit der Brennweite	121
6.6	Zuordnung der x - z -Position zur Bildnummer	123
6.7	3-D-Fehler in Abhängigkeit der Rotation	127
6.8	Fehlermaß in Abhängigkeit der Rotation	128
6.9	Optimierung am L-förmigen konvexen Objekt	130
6.10	Anzahl der 3-D-Punkte für jede Ansicht des L-förmigen konvexen Objekts	145
6.11	Anzahl der 3-D-Punkte für jede Ansicht des Tetraeders	149
6.12	Anzahl der 3-D-Punkte für jede Ansicht des Prüfkörpers	151
D.1	3-D-Fehler in Abhängigkeit der Brennweite	179
D.2	Fehlermaß in Abhängigkeit der Brennweite	179
D.3	Zuordnung der x - z -Position zur Bildnummer	181
D.4	3-D-Fehler in Abhängigkeit der Rotation	185
D.5	Fehlermaß in Abhängigkeit der Roation	186

Kapitel 1

Einführung

Bereits Aristoteles (384 v. Chr. - 322 v. Chr.) teilte die verschiedenen Sinneseindrücke des Menschen in die noch heute gebräuchlichen 5 Kategorien ein [Goh47]: Sehsinn, Hörsinn, Tastsinn, Geschmackssinn, Geruchssinn. „Auch die Bedeutung der einzelnen Sinne, [...] wie andererseits die des Gehörs und des Sehvermögens, betrachtete er unter biologisch-teleologischem Gesichtspunkt, d. h. in ihrer entscheidenden Bedeutung für die Erhaltung und Förderung des Lebens.“ ([Cap71] S. 67). Selbst wenn in der modernen Physiologie noch vier weitere Sinne (Temperatursinn, Schmerzempfinden, Gleichgewichtssinn und Tiefensensitivität) bekannt sind, so haben doch die Fernsinne Hören und Sehen den wichtigsten Einfluss in unserem alltäglichen Leben. Insbesondere das Sehen vermittelt uns weitreichende Fähigkeiten, die uns dabei helfen, uns in unserer Umwelt zu orientieren und mit ihr zu interagieren.

Ein wichtiger Aspekt des Sehens ist dabei die Gewinnung von Tiefen- oder 3-D-Information. Erst dadurch kann man z. B. ein Fahrzeug einparken oder Gegenstände zielgenau greifen, wie ein Selbstversuch zeigt: Ein Auge wird geschlossen, anschließend nimmt man in linke und rechte Hand jeweils einen Stift und versucht nun die Spitze eines Stifts mit der Spitze des anderen zu berühren. Die Gewinnung der 3-D-Information geschieht beim Menschen überwiegend unterbewusst unter anderem durch das Augenpaar als Stereosystem, durch Bewegungsinformationen und/oder a priori Informationen (z. B. der Größe eines Objekts).

Im Forschungsbereich Mustererkennung versucht man die perzeptiven Fähigkeiten des Menschen mit Hilfe von mathematischen Methoden am Rechner nachzubilden [Nie03]. Dabei beschäftigt sich das Teilgebiet Rechnersehen mit der Verarbeitung und automatischen Auswertung von visuellen Informationen. Einer der Schwerpunkte ist dabei die Rekonstruktion von 3-D-Information für unterschiedliche Anwendungsbereiche, von denen einige im Folgenden vorgestellt werden.

Bestimmung von Entfernungen In der Astronomie ist die Entfernungsmessung an sich das Ziel. Bei relativ nahe gelegenen Objekten, wie Mond, Planeten oder nahen Fixsternen, werden die gleichen Triangulationsverfahren eingesetzt wie beim Rechnersehen (vgl. Abschnitt 2.5 und [Her96]).

Bildbasiertes Rendering Man möchte beispielsweise eine Szene aus einer Position betrachten, von der kein Bild aufgenommen wurde. Hat man aber genügend Aufnahmen der Szene aus anderen, geeigneten Positionen, sowie die 3-D-Geometrie der Szene, dann lässt sich ein Lichtfeld [Gor96, Lev96, Hei04, Sch08] erzeugen, mit dessen Hilfe man die gewünschte Ansicht fotorealistisch rendern kann.



Bild 1.1 Beispiel für Erweiterte Realität: In das linke Bild sollen drei künstliche Objekte eingefügt werden. In der Mitte ist eine sog. Tiefenkarte dargestellt, die für jeden Bildpunkt die Tiefe als Grauwert darstellt. Dies ist notwendig um Verdeckungen richtig zu berechnen. Das Resultat ist rechts zu sehen. Bildquelle: [Vog01, Sch02].

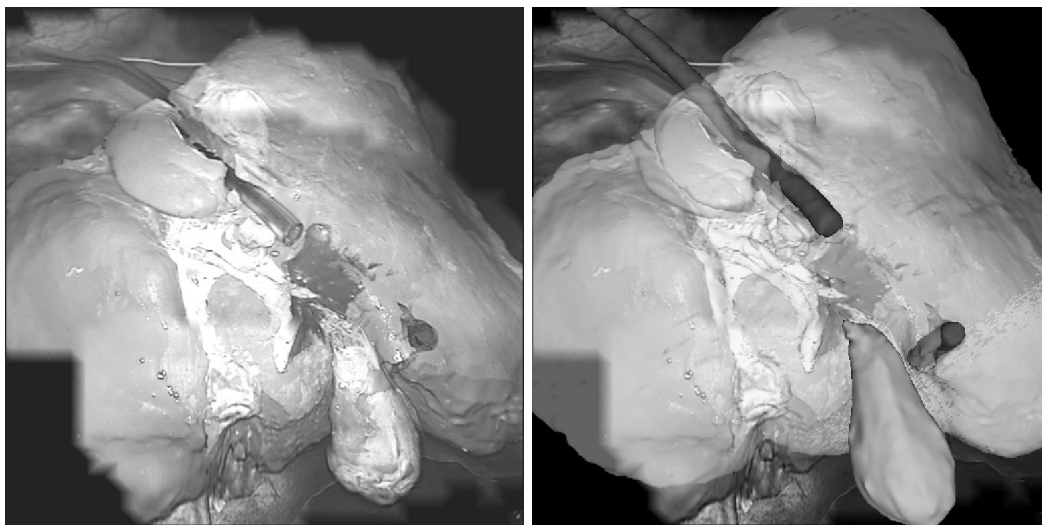


Bild 1.2 Links: Modell von Leber und Gallenblase. Rechts: Überlagerung mit einer CT-Aufnahme. Bildquelle: [Vog06].

Erweiterte Realität Die Erzeugung einer erweiterten Realität (engl.: Augmented Reality) ist ein anderer Aspekt der Visualisierung, der 3-D-Informationen benötigt [Oht99, Sch06a]. Virtuelle Objekte können in Bilder der Szene eingefügt werden. Durch die 3-D-Informationen können dann z. B. gegenseitige Verdeckungen der Objekte oder auch Schatten modelliert werden. Ein Beispiel sieht man in Bild 1.1: in das Originalbild wurden 3 Objekte (Schwein, Teekanne und Vase) eingefügt.

Computerunterstützte Endoskopie Eine Kombination von bildbasiertem Rendering und Augmented Reality wird in [Vog06] vorgestellt. Aus Endoskopieaufnahmen des Bauchraums wird ein 3-D-Modell erstellt, durch das der Arzt navigieren kann, ohne tatsächlich das Endoskop im Patienten zu bewegen. Zusätzlich können in die gerenderte Szene weitere Strukturen eingeblendet werden, wie beispielsweise der Verlauf der Aorta in Bild 1.2.

1.1 Problemstellung und Motivation

Wozu 3-D-Rekonstruktionen verwendet werden können, ist aus obigen Beispielen ersichtlich geworden. Aus Gründen der Vollständigkeit und um eine Abgrenzung dieser Arbeit gleich zu Beginn deutlich zu machen, wird im folgenden Abschnitt auf unterschiedliche Sensoren für die 3-D-Rekonstruktion eingegangen. In dieser Arbeit werden als Sensoren nur Kameras verwendet werden, die Intensitätsbilder liefern. Deshalb wird im darauf folgenden Abschnitt zunächst konzeptuell eine Grundtechnik für die 3-D-Rekonstruktion aus Intensitätsbildern erläutert: die Triangulation. Anhand der Triangulation lässt sich elegant zeigen, unter welchen Bedingungen zunächst eine 3-D-Rekonstruktion berechnet werden kann. Anschließend wird herausgearbeitet, dass die Wahl der Aufnahmepositionen einen Einfluss auf das Ergebnis einer Rekonstruktion hat. Dies motiviert das Kernthema der Arbeit: Welche Ansichten sind für eine 3-D-Rekonstruktion günstig?

1.1.1 Aufnahmemodalitäten

Um eine 3-D-Rekonstruktion zu berechnen, gibt es unterschiedliche Algorithmen und auch unterschiedliche Sensoren, die hier in zwei Gruppen eingeteilt werden: Sensoren mit aktiver und ohne aktive Beleuchtung.

Systeme mit aktiver Beleuchtung ermöglichen eine direkte Messung der 3-D-Struktur, z. B. durch Laufzeitmessungen von spektralreinem Infrarotlicht oder Laserimpulsen. Bild 1.3a zeigt eine PMD-Kamera¹, die Infrarotlicht einer bestimmten Wellenlänge aussendet. Anhand der Phasenverschiebung, die durch die entsprechende Laufzeit des Lichts von der Diode zum Objekt und zurück zum Sensor entsteht, lassen sich Tiefeninformationen gewinnen.

Ähnlich arbeiten Laserscanner, die jedoch nicht den kompletten Sichtbereich ständig mit Licht im Infrarotspektrum ausleuchten, sondern nur einzelne Impulse nacheinander in verschiedene Richtungen senden und die entsprechende Laufzeit messen.

Ein weiterer Ansatz, der aktive Beleuchtung verwendet, ist strukturiertes Licht. Durch Streifenprojektionen auf Oberflächen können dort 3-D-Strukturen vermessen werden. Bild 1.3b zeigt eine solche Streifenaufnahme. Der Streifenprojektor projiziert ein bekanntes Streifenmuster auf das Objekt. Dieses wird von der Kamera aufgenommen und aufgrund der Abstände zwischen den Linien und dem Verlauf der Kanten im Bild, die durch den Übergang zwischen hell und dunkel entstehen, kann die 3-D-Struktur rekonstruiert werden [Sal98, Fof01, Li03].

Im Gegensatz zu den Systemen mit aktiver Beleuchtung wird bei den passiven Systemen keine (gezielte) Änderung der Beleuchtung vorgenommen. In der Regel ist die Szene durch konstantes, größtenteils diffuses Licht ausgeleuchtet, während mittels Kameras Intensitätsbilder aufgenommen werden. Die in dieser Arbeit entwickelten Methoden betrachten ausschließlich Systeme ohne aktive Beleuchtung.

1.1.2 Wozu Ansichtenplanung

Ohne aktive Beleuchtung genügt im Allgemeinen eine einzelne Aufnahme nicht, um Tiefeninformation zu schätzen. Dies lässt sich am Lochkameramodell (eine

¹PMD = Photonic Mixer Device

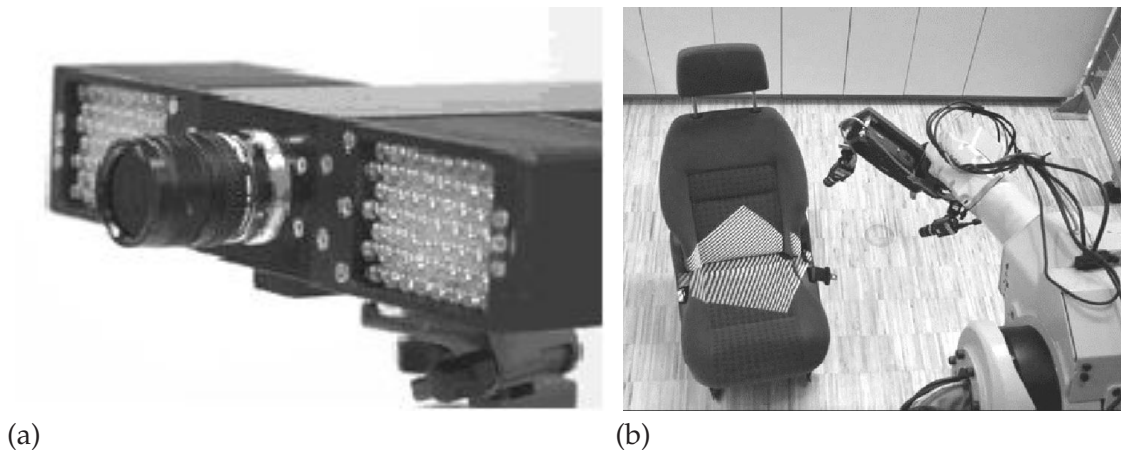


Bild 1.3 (a): Die Infrarot-Dioden der PMD-Kamera senden spektralreines Infrarotlicht. Durch die Phasenverschiebung des gemessenen Lichts auf dem Sensor lässt sich die Tiefe errechnen. (b): Beim Messsystem „kolibri ROBOT“ sind Streifenprojektor und Kameras an einem beweglichen mechanischen Arm befestigt.
Quellen: (b): mit freundlicher Genehmigung des Fraunhofer IOF Jena, 2006

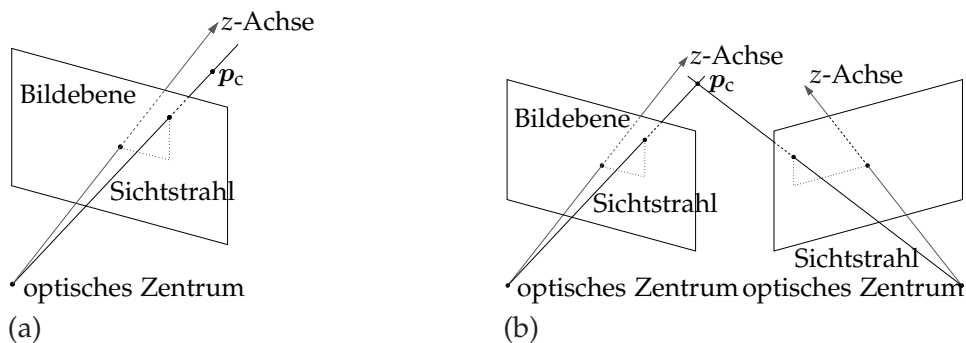


Bild 1.4 (a) Bei nur einem Bild kann auch nur ein Sichtstrahl bestimmt werden, nicht jedoch die Lage des Punkts p_c auf dem Sichtstrahl. (b): Von einem zweiten Bild kann man einen zweiten Sichtstrahl berechnen und mit dem ersten schneiden und erhält somit die Koordinaten des 3-D-Punkts

ausführliche Beschreibung des Lochkameramodells folgt in Abschnitt 2.1.4) plausibilisieren: Die 3-D-Koordinaten eines Punkts werden auf 2-D-Koordinaten im Bild abgebildet (vgl. Bild 1.4a). Bei der Rekonstruktion lässt sich der Sichtstrahl berechnen, der durch das optische Zentrum der Kamera und den Bildpunkt auf der Bildebene festgelegt wurde. Die Position des Punkts entlang des Sichtstrahls ist jedoch nicht bestimmbar. Erst durch eine weitere Aufnahme von anderer Position kann ein zweiter Sichtstrahl und damit der Schnittpunkt der Geraden berechnet werden (vgl. Bild 1.4b). Dieses Verfahren nennt man Triangulation.

Damit ist offensichtlich, dass mindestens zwei Aufnahmen für eine 3-D-Rekonstruktion benötigt werden. Somit stellt sich die nächste Frage: Genügen im Allgemeinen genau zwei Aufnahmen um die Koordinaten eines 3-D-Punkts bestimmen zu können? Hierzu zwei Beispiele, bei denen ein 3-D-Punkt durch Triangulation der Sichtstrahlen rekonstruiert wird:

1. Wurde die Kamera zwischen den Aufnahmen nur um das optische Zentrum rotiert, lässt sich keine 3-D-Information gewinnen. Dies lässt sich mathema-

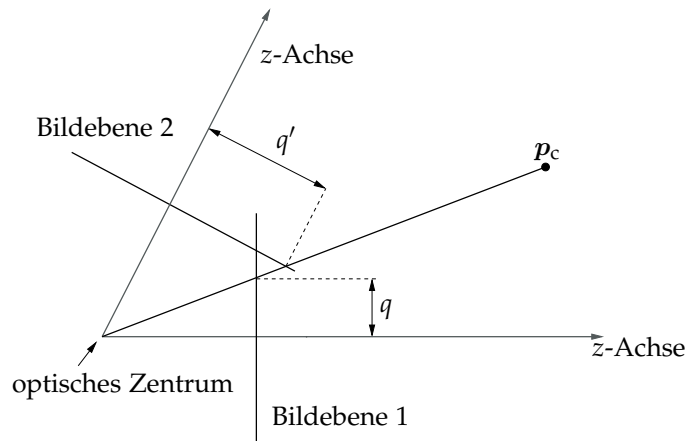


Bild 1.5 Der Sichtstrahl zum Punkt p_c ändert sich nicht, auch wenn die Bildkoordinaten im 1. Bild q und im 2. Bild q' unterschiedlich sind. Eine Triangulation ist hier nicht möglich.

tisch beweisen [Hee92, Oli94] oder anhand einer Skizze (vgl. Bild 1.5) plausibilisieren: Rotiert die Kamera um das optische Zentrum ändert sich der Sichtstrahl nicht, auch wenn sich die Koordinaten des Punkts auf der Bildebene ändern. Die Sichtstrahlen für die erste und zweite Aufnahme sind identisch, damit ist keine Triangulation möglich.

2. Ist die zweite Kameraposition genau so gewählt, dass sich die Sichtstrahlen nicht schneiden, sondern identisch sind, kann die Lage des Punkts auf dem Sichtstrahl nicht bestimmt werden (vgl. Bild 1.6). Ein solcher Fall tritt auf, wenn das optische Zentrum der einen Kamera durch zentrische Streckung, mit dem zu rekonstruierenden Punkt p_c als Zentrum, in das optische Zentrum der anderen Kamera überführt werden kann. Dies gilt insbesondere auch dann, wenn die Kamera nicht bewegt wird und zwischen zwei Aufnahmen nur die Brennweite geändert wird.²

Die obige Frage, ob im Allgemeinen genau zwei Aufnahmen genügen, um die 3-D-Position eines Punkts zu bestimmen, muss somit mit „nein“ beantwortet werden: Es gibt Fälle, in denen zwei Aufnahmen nicht genügen, da die Aufnahmepositionen degeneriert sind. In diesem Fall können die Koordinaten eines 3-D-Punkts nicht bestimmt werden.

Dies führt sofort zur nächsten Frage: Von welchen Kamerapositionen lässt sich eine 3-D-Rekonstruktion durchführen? Aus den obigen Erläuterungen ist klar, dass trianguliert werden kann, falls sich die Sichtstrahlen schneiden. Die bisherigen Überlegungen gehen davon aus, dass alle Werte exakt bestimmt sind. Bei realen Experimenten dürfen aber Aspekte, wie z. B. Rauschen, nicht vernachlässigt werden. So ist beispielsweise die Koordinate auf der Bildebene nicht immer exakt gegeben (z. B. Diskretisierungsfehler, Fehler bei der Zuordnung korrespondierender Punktpaare). Durch die verrauschten Bildpunkte wird die Triangulation fehlerhaft.

Die nächste Frage lautet damit: Kann, trotz verrauschten Eingangsgrößen, durch eine Ansichtenauswahl der Rekonstruktionsfehler minimiert werden, oder anders

²Da hier das Lochkameramodell verwendet wird, spielt Tiefenschärfe und die 3-D-Rekonstruktion daraus (Depth from Focus/Defocus [Kuh06]) keine Rolle.

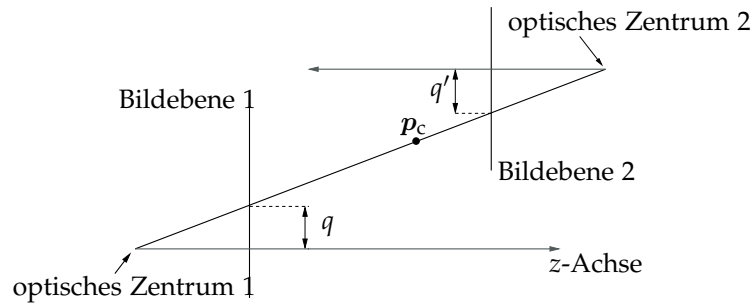


Bild 1.6 Der Sichtstrahl zum Punkt p_c ist von beiden Kamerapositionen aus der Gleiche. Seine ursprüngliche Lage auf dem Sichtstrahl ist nicht ermittelbar.

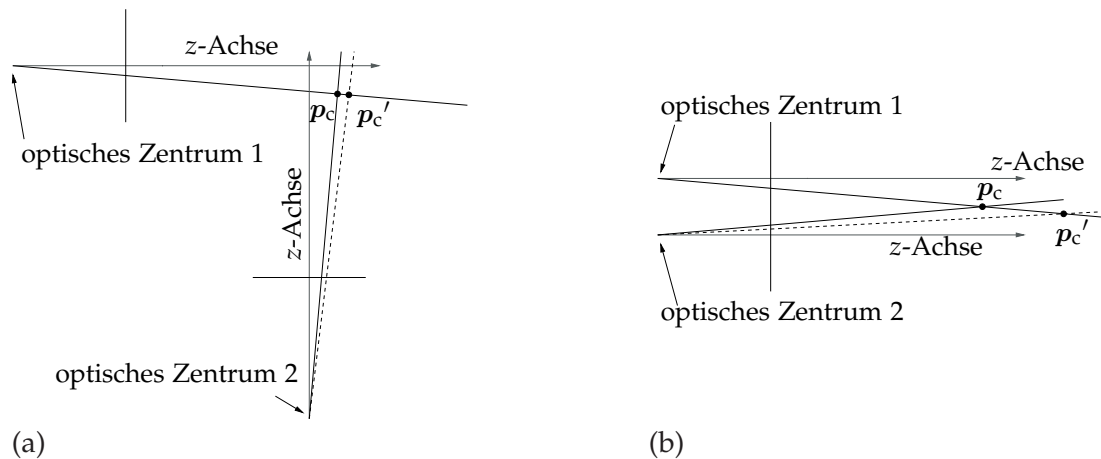


Bild 1.7 Der Sichtstrahl der unteren Kamera weicht aufgrund von Rauschen jeweils um 2° ab. Bei der Konfiguration (a) ist jedoch der resultierende Fehler wesentlich geringer als in Konfiguration (b).

ausgedrückt: gibt es einen Zusammenhang zwischen Aufnahmeposition und Rekonstruktionsgüte? Ein Eindruck lässt sich anhand Bild 1.7 gewinnen: Der Sichtstrahl der zweiten Kamera weicht um jeweils 2° ab (gestrichelte Linie). Der resultierende Rekonstruktionsfehler ist bei der nahezu rechtwinkligen Anordnung der Sichtstrahlen wesentlich kleiner als bei der spitzwinkligen Anordnung.

Es existiert also ein Zusammenhang zwischen Aufnahmeposition und Rekonstruktionsgüte. Dieser lässt sich unter gewissen Annahmen direkt herleiten. Ein entsprechendes Modell wird in Kapitel 3.2 für Stereosysteme hergeleitet.

Im Folgenden wird nun zunächst ein Überblick über bereits bestehende Arbeiten zum Thema Ansichtenplanung zur 3-D-Rekonstruktion gegeben. Anschließend wird der Beitrag dieser Arbeit erläutert.

1.2 Stand der Kunst

In der Literatur wird die Ansichtenplanung für unterschiedliche Applikationen eingesetzt: Neben der 3-D-Rekonstruktion wird sie z. B. in der Objektverfolgung [Den97, Zob04, Deu05, Deu06] oder Objekterkennung [Arb99, Arb01, Dei03, Dei06b, Der06] verwendet. Einen Überblick über Sensorplanung für verschiedene Anwendungen findet sich in [Tar95a, Den03a, Der08]. Bei der 3-D-Rekonstruktion wird die Ansichtenplanung überwiegend dann eingesetzt, wenn Tiefenscanner zur Ver-

fügung stehen. Ferner verfolgen diese Ansätze meist ausschließlich das Ziel, ein zu rekonstruierendes Objekt möglichst vollständig zu erfassen. In dieser Arbeit geht es jedoch um eine möglichst akkurate Rekonstruktion aus Intensitätsbildern. Wegen der unterschiedlichen Zielsetzung können die Algorithmen nicht einfach übertragen werden. Vergleichsweise gering ist dagegen die Literatur für Ansätze mit Kameras. Diese setzen häufig eine bestimmte geometrische Struktur des zu rekonstruierenden Objekts voraus. Die neuartigen, hier vorgestellten Algorithmen unterliegen keinen solchen Einschränkungen. Für sie konnten hingegen Ansätze aus dem Bereich Objektverfolgung [Den03b, Zob04, Deu06] weiterentwickelt werden.

Zunächst folgt nun ein Überblick über die Ansichtenplanung für die 3-D-Rekonstruktion mit Tiefenscannern, also mit aktiver Beleuchtung. Anschließend werden Ansätze unter Verwendung von Kameras ohne aktive Beleuchtung vorgestellt.

1.2.1 3-D-Rekonstruktion mit Tiefenscannern

Eine der frühesten Arbeiten zur Ansichtenplanung für die 3-D-Rekonstruktion mit Tiefenscannern stammt von Connolly [Con85]. Connolly startet mit einem unvollständigen Octree (eine Baumstruktur in der jeder Knoten 8 Nachfolger besitzt), dessen Wurzel zunächst den Status „noch nicht gesehen“ bekommt, was impliziert, dass dieser Bereich von noch keiner Ansicht gesehen wurde. Die anderen Zustände, die ein Knoten einnehmen kann, sind „leer“ (sie sind also „vor“ der Oberfläche des Objekts) oder „verdeckt“ (auf der Oberfläche oder dahinter). In der Arbeit werden der „planetarium“ und der „normale“ Algorithmus präsentiert. Der „planetarium“-Algorithmus sucht auf einem regulären Gitter auf einer Kugel um das Objekt herum diejenige Ansicht, die die meisten noch nicht gesehenen Knoten des Octrees enthält. Beim „normalen“ Algorithmus werden Nachbarschaftsbeziehungen zwischen den Knoten des Octrees mit einbezogen: Für jeden noch nicht gesehenen Knoten wird die Zahl der „leeren“ Nachbarn gezählt. Der Knoten, dessen drei Nachbarn die meisten „leeren“ Nachbarn enthalten, wird als nächstes beobachtet. Bei diesen Algorithmen geht es ausschließlich um die vollständige Rekonstruktion eines Objekts. Der „planetarium“-Algorithmus erreicht dies bei topologisch einfachen Objekten gewöhnlich mit 4–5 Aufnahmen [Won98].

In [Won98] wird ein Algorithmus für die vollständige Rekonstruktion eines Objekts mit der kleinstmöglichen Anzahl von Ansichten vorgestellt. Als Scanner wird eine perspektivische Kamera verwendet, die direkt die Tiefenwerte messen kann, ähnlich dem Prinzip einer PMD-Kamera. Für die Rekonstruktion wird ein Voxelgitter verwendet. Von einer Ansicht erfolgt eine Messung, durch die den einzelnen Voxeln Zustände zugewiesen werden können. Neue Ansichten werden so gewählt, dass man möglichst viele Voxel mit noch unbekanntem Zustand sieht. Für die Optimierung der Zielfunktion wird unter anderem eine vollständige Suche auf einem diskreten regulären Gitter auf einer Kugel um das Objekt herum durchgeführt. Die experimentelle Auswertung basiert auf simulierten Daten. Der Algorithmus ermöglicht eine schnelle vollständige Rekonstruktion, jedoch ohne eine hohe Rekonstruktionsgenauigkeit zu erzielen.

Massios und Fisher [Mas98] wählen ebenfalls ein Voxelgitter als Volumenrepräsentation. Dieses besteht aus würfelförmigen Voxeln, die den Status „empty“, „seen“, „unseen“ oder „occlusion plane“ besitzen. Wird eine Messung durchge-

führt, so werden alle Voxel, die vom Scanner aus gesehen hinter der gemessenen Oberfläche sind, als „unseen“ markiert, die davor als „empty“ und alle zwischen einem „unseen“ und einem „empty“ als „occlusion plane“. Das Qualitätskriterium für neue Ansichten besteht dann aus der gewichteten Summe zweier Gütemaße: Das erste Kriterium spiegelt die Zahl neuer Voxel wider, die noch nicht gesehen wurden; das zweite stellt einen Zusammenhang zwischen der Normalen der Oberfläche und dem Blickwinkel des Sensors her. Dieser Winkel sollte bei dem verwendeten Sensor möglichst klein sein, um eine hohe Messgenauigkeit zu erzielen.

In der Arbeit von Pito [Pit99] wird das Problem der nächsten besten Ansicht nicht nur auf die Frage, welcher Teil der Szene als nächstes betrachtet werden sollte, reduziert, sondern zusätzlich weitere Probleme für eine vollständig automatisierte 3-D-Rekonstruktion erörtert. Zum einen das Registrierungs- und das Integrationsproblem: Unter dem Registrierungsproblem werden mechanische Ungenauigkeiten und Kalibrierfehler verstanden, die keine exakte Bestimmung der Sensorposition erlauben. Die Integration wird benötigt, um die aus einer einzelnen Ansicht rekonstruierten Oberflächen in das bisherige globale Modell einfügen zu können. Dazu muss die neu rekonstruierte Oberfläche einen bestimmten Überlappbereich mit dem bisherigen Modell bilden. Zum anderen werden weitere Anforderungen an den Algorithmus formuliert: eine möglichst exakte Rekonstruktion, keine a priori Annahmen über die geometrische Struktur, möglichst effiziente Auswahl aller möglichen Aufnahmekonstellationen und eine automatische Terminierung, sobald das Objekt vollständig rekonstruiert wurde.

Klein und Sequeira [Kle00a, Kle00b] präsentieren Ansätze, die gut skalierbar sind. Im Gegensatz zu Ansätzen anderen Autoren muss die Position des Sensors nicht auf einer Kugel- oder Zylinderoberfläche liegen, was die Akquisition kompletter Räume ermöglicht. Es wird ein 8-dimensionaler Suchraum (3-D-Position und Orientierung sowie zwei Auflösungsparameter des Scanners) verwendet. Der Algorithmus zur Ansichtenplanung basiert auf der Analyse der Randbereiche der bisher rekonstruierten Oberfläche. Die durch diesen Ansatz notwendigen Berechnungen von Verdeckungen werden auf Prozessoren von Grafikkarten durchgeführt. Dadurch wird die Berechnungszeit deutlich reduziert.

Banta et. al. [Ban00] akquirieren in ihrem Ansatz zusätzlich Intensitätsbilder, die jedoch nicht für die 3-D-Rekonstruktion verwendet werden, sondern nur um später die Oberflächen rendern zu können. Für die Rekonstruktion wird ein 3-stufiger Algorithmus beschrieben. Die erste Stufe untersucht große, noch nicht gesehene Bereiche, ähnlich wie der Algorithmus aus [Mav93], anhand einer Analyse des Randbereichs der bisher rekonstruierten Oberfläche. Als zweites wird ein Voxelgitter aufgebaut und um noch fehlende Bereiche ergänzt. Im letzten Schritt erfolgt eine genaue lokale Rekonstruktion eines jeden Oberflächensegments, um eine möglichst akkurate 3-D-Rekonstruktion zu erhalten.

In [Sco01, Sco03] wird eine Messbarkeitsmatrix (erstmalig von Tarbox und Gottschlich [Tar95b] eingeführt) als zugrundeliegendes Datenmodell verwendet. Diese enthält als Zeilenindizes die Punktindizes der Punkte auf der Oberfläche des Objekts und als Spaltenindizes die Indizes möglicher Ansichten. Jedes Element enthält einen binären Wert, entsprechend der Sichtbarkeit des 3-D-Punkts von der jeweiligen Ansicht. Aufgrund der Größe der Messbarkeitsmatrix sind die Algorithmen von [Tar95b] jedoch sehr rechenintensiv. Deshalb wird in [Sco01, Sco03] ein schrittweises Vorgehen beschrieben: zunächst eine grobe Abtastung und an-

schließlich eine grobe Segmentierung; danach die Bestimmung der möglichen Ansichten, die Berechnung der Messbarkeitsmatrix und zuletzt die Optimierung der groben Rekonstruktion. Durch die Segmentierung von einzelnen Bereichen kann die Dimensionalität der Messbarkeitsmatrix reduziert werden. Dafür wird jedoch zum Schluss eine Registrierung notwendig, um ein komplettes und konsistentes Modell zu erhalten.

Die Arbeit von Munkelt et. al. [Mun06, Mun11] verwendet strukturiertes Licht für die Rekonstruktion. Dabei basiert die Ansichtenauswahl wiederum auf einem Voxelgitter; die möglichen Voxel haben die Werte, wie sie auch schon in [Mas98] verwendet wurden („empty“, „seen“, „unseen“ oder „occlusion plane“). Neben der vollständigen Rekonstruktion steht hier eine möglichst exakte Rekonstruktion im Vordergrund. Dazu werden für bereits rekonstruierte Flächen die günstigsten Scannerparameter ermittelt, um beim erneuten Scannen der Flächen eine möglichst hohe Genauigkeit zu erhalten.

1.2.2 3-D-Rekonstruktion aus Intensitätsbildern

Neben den oben aufgeführten Arbeiten zur Ansichtenplanung für die 3-D-Rekonstruktion mit Tiefenscannern existieren vergleichsweise wenig publizierte Ansätze, die ausschließlich aus Intensitätsbildern, d. h. ohne aktive Beleuchtung, 3-D-Informationen gewinnen. Die wichtigsten werden nun im Folgenden dargestellt.

Marchand und Chaumette [Mar94, Mar96, Mar99a, Mar99b] bilden bei der Rekonstruktion die Geometrie einer Szene approximativ durch Zylinder und Polyeder nach. In den Arbeiten [Mar94, Mar96] wird insbesondere das Problem der präzisen Rekonstruktion von Zylindern thematisiert. Die 3-D-Rekonstruktion eines Körpers läuft in vier Schritten ab: Fokussierung des Objekts, Erkennung des Objekttyps (Zylinder oder Polyeder), optimale Bestimmung der Parameter für die Vermessung der geometrischen Struktur des Objekts und Bestimmung der Lage des Objekts in der Szene. Der Typ des Objekts wird aufgrund detektierter Kanten mit Hilfe von statistischen Tests bestimmt. Nach der Bestimmung der Klasse werden die Kanten der Zylinder bzw. die 3-D-Segmente der Polyeder durch einen Structure-from-controlled-motion Ansatz rekonstruiert. Dieser Ansatz erlaubt die Rekonstruktion und ermöglicht Vorhersagen, bei welcher Kamerabewegung der Fehler möglichst klein wird [Cha94, Cha96]. Die Rekonstruktion der gesamten Szene erfolgt dann durch die Zusammensetzung der gefundenen 3-D-Segmente.

In den Arbeiten [Mar99a, Mar99b] wird der Ansatz zu einem 3-schichtigen System weiterentwickelt. Die erste Schicht gewährleistet die präzise Rekonstruktion von Zylindern und Polyedern, wie in den früheren Arbeiten. Die Zwischenschicht bestimmt inkrementell die Lage der einzelnen geometrischen Formen zueinander. Dabei wird eine 3-D-Karte erstellt, in die die bisher rekonstruierten Segmente eingefügt werden. Mit Hilfe eines Vorhersage- und Verifikationsschemas, welches auf Bayes-Netzen basiert [Pea88, Dev04], werden bereits rekonstruierte Segmente von neuen unterschieden. Der probabilistische Ansatz ist hier notwendig, da die Positionen der 3-D-Objekte in der Karte fehlerbehaftet sind. Schließlich wird in der dritten Schicht sichergestellt, dass die Szene vollständig rekonstruiert wurde. Dabei wird ein zu [Con85] ähnlicher Ansatz gewählt: Das Verhältnis zwischen dem von einer Ansicht insgesamt sichtbaren Volumen zum bisher nicht beobachteten muss minimal werden.

Eine Reihe von Arbeiten wurde von Kutulakos und Dyer veröffentlicht: [Kut94a, Kut94b, Kut94c, Kut94d, Kut95]. Dort wird für die 3-D-Rekonstruktion ein Shape-from-occluding-contour Algorithmus verwendet. Dieses Verfahren ist ähnlich zu Shape-from-silhouettes Ansätzen [Che05]: Die Kante im Bild zwischen Objekt und Hintergrund wird als Konturlinie bezeichnet. Aufgrund eines einfarbigen, homogenen Hintergrunds lässt sich die Konturlinie relativ einfach ermitteln. Bewegt sich die Kamera, verändert sich die Konturlinie im Bild entsprechend der 3-D-Struktur des Objekts. Dadurch kann die Krümmung der Oberfläche im Bereich der Konturlinie berechnet werden. Elliptisch gekrümmte Oberflächen sind dabei unkritisch. Schwieriger sind parabolische oder hyperbolische Krümmungen, da sie nur aus bestimmten Blickwinkelbereichen rekonstruierbar sind. Kritisch sind plane Oberflächenstücke, da hier bei der Verfolgung der Kontur Unstetigkeiten auftreten. Konkave Bereiche lassen sich überhaupt nicht rekonstruieren, da vom Inneren eines solchen Bereichs nie eine Konturlinie zwischen Objekt und Hintergrund entstehen kann. Für die Ansichtenauswahl sind dann je nach Krümmungsmodalität unterschiedliche Fallunterscheidungen notwendig um die jeweils nächste beste Ansicht zu ermitteln.

Niem [Nie96, Nie99] verwendet, analog zu Kutulakos und Dyer, für die Rekonstruktion ebenfalls einen Shape-from-occluding-contours Algorithmus. Erweitert wird der Algorithmus zum einen durch ein Verfahren, das Unstetigkeitsstellen bei der Verfolgung analysiert. Dadurch lassen sich ebene Flächen detektieren und rekonstruieren. Zum anderen wird bei der Ansichtenplanung auf die komplizierten Fallunterscheidungen von Kutulakos und Dyer verzichtet. Stattdessen wird angenommen, dass die bisher rekonstruierte Oberfläche aufgrund von Rauschen der Kameraparameter fehlerhaft ist. Der Fehler ist der Abstand zwischen der real beobachteten Kontur und der, aufgrund der bisher rekonstruierten Oberfläche, erwarteten. Das Fehlermaß ist dabei eine Funktion der Volumenparameter, die das Volumen des Objekts beschreiben, und der Kameraparameter. Die Kameraparameter werden optimiert, so dass das zu erwartende Fehlermaß minimal wird.

Ein Verfahren für die automatische Oberflächenmodellierung mittels Stereosystemen wurde von Koch [Koc97] entwickelt. Das Stereosystem ist kalibriert und besteht aus zwei starr verkoppelten Videokameras. Aus einer einzelnen Ansicht des Stereosystems wird ein Teilmodell der Oberfläche rekonstruiert. Die Ansichtenauswahl basiert dabei auf heuristischen Regeln, die sich aus den „Erfahrungen des Autors“ ([Koc97] S. 84) ergeben, und gewährleistet nur eine vollständige Rekonstruktion aber keine Verbesserung der Genauigkeit.

Olaque und Mohr schlagen ein Verfahren zur optimalen Triangulation von Targets vor [Ola97, Ola02]. Die Targets sind Marker auf dem Objekt, die in den Bildern leicht zu identifizieren sind. Zunächst wird für die Ansichtenplanung ein Gütemaß definiert, das den Triangulationsfehler im Dreidimensionalen misst. Mit Hilfe eines genetischen Ansatzes wird eine festgelegte Anzahl von optimalen Kamerapositionen für die Rekonstruktion ermittelt, die das Gütemaß optimiert.

Relativ neue Arbeiten zum Thema Ansichtenplanung für die 3-D-Rekonstruktion stammen von Trummer [Tru10, Tru11]. Trummer verwendet dazu ein System bei dem eine Kamera auf einen Roboterarm montiert ist. Für die 3-D-Rekonstruktion werden verschiedene Positionen mit Hilfe des Roboters angefahren und zwischen den Aufnahmepositionen Punktmerkmale verfolgt. Durch die Berücksichtigung der Kameraposition in jedem Schritt vereinfacht sich das Problem der Punkt-

verfolgung und zusätzlich kann dadurch die Unsicherheit der 3-D-Punkte robust geschätzt werden [Tru09a, Tru09b]. Basierend auf dieser Schätzung der Unsicherheit wird dann die nächste beste Ansicht berechnet, das ist die Ansicht, die die Unsicherheit am stärksten reduziert. Dazu wird, ähnlich wie in dieser Arbeit, ein Kriterium entwickelt, das auf der Eigenwertanalyse der Kovarianzmatrix der 3-D-Schätzung beruht. Möchte man mehr als nur einen Punkt rekonstruieren, so ist, im Gegensatz zu dem in dieser Arbeit vorgestellten Ansatz, bei Trummer [Tru11] nur der 3-D-Punkt mit größter Unsicherheit für die Optimierung ausschlaggebend.

1.3 Beitrag der Arbeit

Die Zielsetzung der Arbeit besteht in der Entwicklung von zwei Algorithmen zur Ansichtenplanung für eine möglichst akkurate 3-D-Rekonstruktion. Der erste Algorithmus optimiert direkt ein geometrisches Gütemaß für Stereokamerasysteme, während der zweite auf einer probabilistischen Schätzung beruht. In diese probabilistische Schätzung werden sukzessiv neue Ansichten integriert, um so die Unsicherheit in der Schätzung zu minimieren. Die Ansätze der beiden Algorithmen sind deshalb disjunkt. Es wird kein spezielles Anwendungsgebiet vorausgesetzt.

1.3.1 Geometrischer Ansatz

In diesem Teil der Arbeit wird ein neuartiger Ansatz für die Optimierung von Stereo- und Kameraparametern für eine möglichst akkurate 3-D-Rekonstruktion entwickelt.

Der Algorithmus zur 3-D-Rekonstruktion unterliegt dabei keinen Beschränkungen auf geometrische Grundkörper (Zylinder, Kugeln, etc.), wie sie in [Mar96, Cha96] notwendig sind. Stattdessen kann die Struktur beliebig aufgebaut sein. Ferner sind keine Fallunterscheidungen zwischen konkaven, konvexen und planaren Bereichen, wie in [Kut94a, Kut94b, Kut94c, Kut94d, Kut95], notwendig.

Im Gegensatz zur klassischen Fehlerrechnung für Stereosysteme [Kan96], wird hier ein eigenes Fehlermodell aufgestellt. Dieses erlaubt Zusammenhänge zwischen Stereo- und Kameraparameter und der Rekonstruktionsgüte herzustellen. Die Zusammenhänge werden, soweit wie möglich analytisch, andernfalls durch numerische Simulationen, bewiesen.

Neuartig an dem vorgestellten Ansatz ist darüber hinaus, dass an dieser Stelle der hergeleitete Zusammenhang zur Optimierung der entsprechenden Parameter verwendet wird. In der Literatur finden sich von anderen Autoren bislang keine so detaillierten Untersuchungen zu Stereosystemen, die den Einfluss der Parametersystematisch untersuchen und Aussagen formulieren, wie die Parameter optimal einzustellen sind. Durch die optimale Parameterkonfiguration kann das zu erwartende Rekonstruktionsergebnis erheblich verbessert werden.

1.3.2 Probabilistischer Ansatz

Dieser Ansatz basiert auf einem informationstheoretischen Konzept, welches in [Den02, Dei05] für die Ansichtenplanung zur Objekterkennung entwickelt und in [Zob04, Deu06] auf den dynamischen Fall der Objektverfolgung transferiert

wurde. Die allgemeinen Ideen und Ansätze aus den vier Publikationen werden hier erstmalig auf das Problem der Ansichtenauswahl für die 3-D-Rekonstruktion übertragen. Dies erlaubt eine probabilistische Beschreibung der 3-D-Rekonstruktion mittels Kalman-Filter und eine darauf basierende Ansichtenplanung.

Für das, in der Technik allgemein sehr weit verbreitete Kalman-Filter, wird in dieser Arbeit eine elegante Möglichkeit vorgestellt, wie für spezielle Arten von Zustandsübergangs- und Beobachtungsfunktionen die Berechnungskomplexität drastisch vereinfacht werden kann. Dieses Ergebnis kann auch in anderen Anwendungsfällen des Kalman-Filters eingesetzt werden.

Der probabilistische Ansatz für die Ansichtenplanung wird erstmalig in ein Gesamtsystem zur 3-D-Rekonstruktion und Ansichtenplanung integriert. Das entwickelte System kommt dabei ohne jegliche a priori Informationen aus: Es genügen zwei manuell vorgegebene Ansichten, um alle notwendigen Größen für die Ansichtenplanung zu berechnen. Ferner werden keinerlei Voraussetzungen an die 3-D-Geometrie der Szene gestellt: Der Algorithmus kann prinzipiell beliebig komplexe Szenen rekonstruieren, was die meisten Verfahren zur Ansichtenplanung für die 3-D-Rekonstruktion aus Intensitätsbildern nicht beherrschen.

Im Gegensatz zu [Zob04] wird nicht nur die Entropie als Gütemaß für die Bestimmung einer nächsten Ansicht verwendet, sondern noch zwei weitere Gütekriterien aus der Theorie abgeleitet. Diese Gütemaße erlauben mehr Freiheiten bei der Ausgestaltung des Algorithmus hinsichtlich der Entscheidung zwischen höherer Komplexität bei der Berechnung des Gütemaßes oder höherer Präzision der Rekonstruktion.

Entgegen der bisherigen Publikationen in diesem Bereich wird in dieser Arbeit explizit das Problem der Sichtbarkeit durch das eingeschränkte Sichtfeld der Kamera und Selbstverdeckungen der zu rekonstruierenden Szene, modelliert. Die Sichtbarkeit eines Szenenbereichs ist für die Ansichtenplanung von fundamentaler Bedeutung, denn ohne kann unter Umständen eine nutzlose Ansicht gewählt werden. Die Wahrscheinlichkeit der Sichtbarkeit wird hier automatisch aus den berechneten Informationen ermittelt. Hervorzuheben ist, dass die Sichtbarkeitsbedingungen in die Gütefunktionen integriert werden, so dass eine hinreichende und notwendige Bedingung für die nächste beste Ansicht in geschlossener Form formuliert werden kann. Die Ermittlung dieser Wahrscheinlichkeit ist auch in anderen Anwendungsgebieten, wie z. B. der Objektverfolgung, nützlich oder gar notwendig [Der08].

Ein weiterer Beitrag der Arbeit besteht in der allgemeinen Formulierung des Positionierungsproblems. Durch Methoden der Robotik kann die Beschränkung der Kamerapositionen auf eine Halb- oder Teilkugel, die bei fast allen Autoren aus dem Literaturüberblick zum Tragen kommt, aufgehoben werden und es kann die volle Bewegungsfreiheit eines Roboterarms ausgenutzt werden. Diese Methodik ist darüber hinaus auch noch sehr einfach auf andere Roboterarme übertragbar.

1.4 Aufbau der Arbeit

Kapitel 2 erläutert Grundlagen für die 3-D-Rekonstruktion, wie sie in den darauf folgenden Kapiteln verwendet werden. Zu den Grundlagen gehören dabei die Definition der einzelnen Begriffe, die Festlegung einer Notation, übliche Techniken

für die Bestimmung der notwendigen Parameter sowie eine Einführung in gängige Techniken zur 3-D-Rekonstruktion.

Ein Überblick über verschiedene Ansätze zur Ansichtenplanung für die 3-D-Rekonstruktion wird in Kapitel 3 gegeben. Dabei bietet sich eine Unterscheidung zwischen geometrischen und probabilistischen Ansätzen an. Bei den geometrischen Ansätzen werden zunächst die bereits unter Abschnitt 1.2.2 erwähnten nochmals ausführlicher dargestellt, damit eine Abgrenzung zu dem dort entwickelten Algorithmus deutlich herausgestellt werden kann. Anschließend folgen die Prinzipien für die probabilistische 3-D-Rekonstruktion und Ansichtenauswahl.

In Kapitel 4 wird ein Algorithmus für die probabilistische Ansichtenauswahl beschrieben. Dazu wird zunächst das Problem der 3-D-Rekonstruktion als Zustandsschätzproblem formuliert. Der Zustand lässt sich mit Hilfe eines erweiterten Kalman-Filters schätzen. Anschließend werden drei verschiedene Optimalitätskriterien für die Ansichtenauswahl diskutiert. Zusätzlich werden Nebenbedingungen, deren Beachtung essentiell für die Ansichtenplanung bei komplexen Objekten sind, diskutiert.

Die Beschreibung der experimentellen Umgebung folgt in Kapitel 5. Dabei wird insbesondere auf die verwendete Hardware sowie die wichtigsten Softwarekomponenten eingegangen. Am Ende des Kapitels folgt ein Überblick über das Gesamtsystem.

Die Ergebnisse ausführlicher Experimente werden in Kapitel 6 dargestellt. Für den Vergleich wird zunächst ein Gütemaß für die Rekonstruktion definiert. Anschließend erfolgen die Experimente für die in dieser Arbeit vorgestellten neuen Algorithmen für die geometrischen und probabilistischen Ansätze.

Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick über mögliche zukünftige Arbeiten in Kapitel 7.

Kapitel 2

3-D-Rekonstruktion

Dieses Kapitel beschreibt die theoretischen Grundlagen für die 3-D-Rekonstruktion mit einem System ohne aktive Beleuchtung, das ausschließlich Bilddaten benutzt. Für die Rekonstruktion ist es zunächst wichtig, wie das Bild entsteht. Deshalb befasst sich der Abschnitt 2.1 mit der Projektion der 3-D-Szene auf die 2-D-Bildkoordinaten. Hierzu werden die entsprechenden Projektionsmodelle und die zugehörigen Kameraparameter eingeführt. Die Bestimmung der Kameraparameter mittels Kalibriermuster wird in Abschnitt 2.2 erläutert. Eine wichtige Rolle für die Rekonstruktion spielen die extrinsischen Parameter, die Position und Orientierung der Kamera beschreiben. Ist die Kamera an einem mechanischen Arm befestigt, wie z. B. in Bild 2.4, und die Lage des Arms kann ausgelesen werden, so können die extrinsischen Parameter über die so genannte Hand-Auge-Transformation (vgl. Abschnitt 2.3) bestimmt werden. Der Abschnitt 2.4 erläutert die Kinematik von Roboterarmen, mit deren Hilfe man die Transformation vom Basis- zum Endeffektorkoordinatensystem aufgrund von Gelenkeinstellungen des Roboters und umgekehrt von der Transformation ausgehend die Gelenkeinstellungen berechnen kann. Zum Abschluss dieses Kapitels wird das Prinzip der Triangulation und des Bündelausgleichs vorgestellt. Die Triangulation ist ein elementares Verfahren für die 3-D-Rekonstruktion. Es kann in vielen Fällen zur Plausibilisierung von Sachverhalten herangezogen werden. Der Bündelausgleich ist ein Verfahren für die nichtlineare Optimierung des Rekonstruktionsergebnisses.

2.1 Kameraparameter und Projektionsmodelle

Ein Kamerabild enthält die Projektion einer 3-D-Szene auf die 2-D-Bildebene. Die Abbildung wird üblicherweise durch eine Linse oder durch ein System von Linsen (z. B. Zoomobjektiv) realisiert. Um die Projektion mathematisch zu beschreiben, müssen die physikalischen Eigenschaften der Linsen modelliert werden. Aufgrund der Komplexität der Linsenmodelle ist dies jedoch im Bereich Rechnersehen unüblich. Stattdessen wird entweder das Lochkameramodell (vgl. Bild 1.4) verwendet, dessen Projektionsgleichungen sehr einfach sind, oder eine affine Approximation des Lochkameramodells. In beiden Fällen ergeben sich die Koordinaten des homogenen¹ 2-D-Bildpunkts $\underline{q}$ aus dem 3-D-Punkt $\underline{p}_w$ durch die Multiplikation mit der 3×4 Projektionsmatrix $\mathbf{P}$. Diese kann wie folgt aufgespalten werden:

$$\underline{q} \sim \mathbf{P} \underline{p}_w \sim \mathbf{K} \mathbf{P}_M \mathbf{F} \underline{p}_w, \quad (2.1)$$

¹Punkte in homogenen Koordinaten sind unterstrichen, um sie von kartesischen zu unterscheiden. Eine detaillierte Beschreibung homogener Koordinaten findet sich in einschlägiger Literatur zu projektiver Geometrie, z. B. [Beu04]

dabei steht $\sim$ für „gleich bis auf einen Skalierungsfaktor“, da homogene Koordinaten immer nur bis auf einen Skalierungsfaktor eindeutig sind, $\mathbf{K}$ ist die Kameramatrix, die die intrinsischen Kameraparameter enthält, $\mathbf{P}_M$ die Matrix, die das Projektionsmodell beschreibt und $\mathbf{F}$ die Matrix der extrinsischen Parameter.

Im Folgenden werden zunächst verschiedene Koordinatensysteme eingeführt, die als Bezugssysteme dienen. Danach wird die Bedeutung der intrinsischen und extrinsischen Kameraparameter sowie der Aufbau der entsprechenden Matrizen $\mathbf{K}$ bzw. $\mathbf{F}$ erläutert. Zum Schluss dieses Unterkapitels werden die Projektionsmodelle für die perspektivische und die paraperspektivische Projektion eingeführt. Somit sind alle Komponenten aus Gleichung (2.1) definiert.

2.1.1 Koordinatensysteme

Im Rechnersehen sind drei verschiedene Koordinatensysteme von Relevanz:

1. Weltkoordinatensystem

Das Weltkoordinatensystem legt den Bezugspunkt für alle Elemente einer betrachteten Szene fest. Die Lage ist prinzipiell beliebig, aber fest. In manchen Fällen dienen spezielle Konventionen der Vereinfachung. Die Längen werden gewöhnlich in Meter oder Millimeter gemessen.

2. Kamerakoordinatensystem

Das Kamerakoordinatensystem ist ebenfalls dreidimensional. Die Lage wird durch die Kamera festgelegt: Die z -Achse entspricht der optischen Achse und zeigt von der Kamera in Richtung Objekt/Szene. Der Ursprung liegt im optischen Zentrum O_Z . Die x - bzw. y -Achse verläuft parallel zu den Kanten des optischen Sensors. Längen werden in der gleichen Einheit wie im Weltkoordinatensystem angegeben.

3. Bildkoordinatensystem

Das Bildkoordinatensystem ist zweidimensional. Der Ursprung liegt üblicherweise in der linken oberen Ecke des Bildes. Die x - bzw. y -Achse verläuft parallel zu den Kanten des Bildes. Längen werden in Pixel gemessen.

2.1.2 Intrinsische Kameraparameter

Als intrinsische Kameraparameter werden all diejenigen bezeichnet, die sich durch eine Kamerabewegung nicht ändern. Anders ausgedrückt, die intrinsischen Parameter beschreiben die Projektion der 3-D-Kamerakoordinaten eines Punkts in die 2-D-Bildkoordinaten [Tru98]. Sie enthalten keine Information über die Transformation von Welt- in Kamerakoordinaten und bleiben somit bei Bewegungen der Kamera konstant. Die intrinsischen Kameraparameter bestehen aus der Brennweite f , z. B. in Millimeter, der Größe eines Pixels auf dem Sensor in horizontaler bzw. vertikaler Richtung ζ_x bzw. ζ_y , z. B. in Millimeter pro Pixel, den Koordinaten des Hauptpunkts (das sind die Koordinaten des Schnittpunkts zwischen optischer Achse und Bildebene im Bildkoordinatensystem) (u_p, v_p) , dem Winkel zwischen den beiden Achsen des Bildkoordinatensystems $\beta_\angle$ und Verzerrungsparametern der Linse.

Die Brennweite und die Pixelgröße in die jeweilige Richtung werden zur effektiven Brennweite f_x bzw. f_y zusammengefasst: $f_x := f/\zeta_x$ bzw. $f_y := f/\zeta_y$. Aus dem Winkel $\beta_\perp$ wird der sog. Skew Parameter β gebildet: $\beta := -f/\zeta_x \tan(\pi/2 - \beta_\perp)$. Dieser wird Null, falls der Winkel zwischen den Koordinatenachsen des Sensors 90° beträgt, was bei den meisten Kameras der Fall ist [Har03, Hei04]; deswegen wird im weiteren Verlauf dieser Arbeit $\beta = 0$ angenommen.

Aus den abgeleiteten Kameraparametern wird die Kameramatrix aufgebaut:

Def: 2.1 (Kameramatrix)

Die Kameramatrix K ist eine 3×3 Matrix mit folgenden Einträgen:

$$K = \begin{pmatrix} f_x & \beta & u_p \\ 0 & f_y & v_p \\ 0 & 0 & 1 \end{pmatrix}.$$

Da Verzerrungen durch die Linse nicht durch lineare Abbildungen modelliert werden können, werden sie gesondert behandelt. Gängige Modelle unterscheiden zwischen radialen und tangentialen Verzerrungen [Sla80, Zha96]. Die Abbildung eines verzerrten Bildpunkts q_d auf den unverzerrten q wird durch eine Verschiebung realisiert:

$$\begin{pmatrix} q_x \\ q_y \end{pmatrix} = \begin{pmatrix} q_{dx} \\ q_{dy} \end{pmatrix} + \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix}, \quad (2.2)$$

mit der Verschiebung

$$\begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = \underbrace{\begin{pmatrix} q_{dx} - u_p \\ q_{dy} - v_p \end{pmatrix} \cdot (1 + \kappa_1 \cdot a^2 + \kappa_2 \cdot a^4 + \dots + \kappa_n \cdot a^{2n})}_{\text{radialer Anteil}} + \underbrace{\begin{pmatrix} (2\vartheta_1(q_{dx} - u_p)(q_{dy} - v_p) + \vartheta_2(a^2 + 2(q_{dx} - u_p)^2)) \cdot (1 + \vartheta_3 a^2 + \dots + \vartheta_m a^{2(m-2)}) \\ (2\vartheta_2(q_{dx} - u_p)(q_{dy} - v_p) + \vartheta_1(a^2 + 2(q_{dy} - v_p)^2)) \cdot (1 + \vartheta_3 a^2 + \dots + \vartheta_m a^{2(m-2)}) \end{pmatrix}}_{\text{tangentialer Anteil}}.$$

Dabei ist $a = \sqrt{(q_{dx} - u_p)^2 + (q_{dy} - v_p)^2}$ der Abstand des jeweiligen Punkts zum Hauptpunkt. Die Parameter $\kappa_1, \kappa_2, \dots$ modellieren die radialen Verzerrungen und $\vartheta_1, \vartheta_2, \dots$ die tangentialen.

In der Praxis ist es üblich, ein bis zwei Parameter für die radialen Verzerrungen zu bestimmen und die restlichen auf Null zu setzen [Tsa87, Tru98, Zha00, Hei04]. Tangentiale Verzerrungen werden häufig ganz vernachlässigt, da der Einfluss der radialen stark überwiegt. Werden tangentiale Verzerrungen berücksichtigt, arbeitet man gewöhnlich ebenfalls mit zwei tangentialen Verzerrungsparametern [Vog06].

In dieser Arbeit werden alle von einer Kamera aufgenommenen Bilder vor einer weiteren Verarbeitung mittels Gleichung (2.2) entzerrt. Dazu werden die radialen Parameter κ_1, κ_2 sowie die tangentialen Parameter ϑ_1, ϑ_2 verwendet, die vorher durch eine Kamerakalibrierung (siehe Abschnitt 2.2) bestimmt werden müssen.

2.1.3 Extrinsische Kameraparameter

Während die intrinsischen Parameter die Abbildung von Kamerakoordinaten auf Bildkoordinaten parametrisieren, definieren die extrinsischen die Transformation

von Weltkoordinaten in Kamerakoordinaten. Dazu wird Lage und Orientierung der Kamera im Weltkoordinatensystem mit Hilfe eines Translationsvektors $\mathbf{t} \in \mathbb{R}^3$ und einer Rotationsmatrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ angegeben. Dabei wird in dieser Arbeit die folgende Definition verwendet:

Def: 2.2

Für die Transformation von Kamera- in Weltkoordinaten bzw. für die umgekehrte Transformation von Welt- in Kamerakoordinaten gilt in kartesischen Koordinaten:

$$\mathbf{p}_w = \mathbf{R}\mathbf{p}_c + \mathbf{t} \quad \text{bzw.} \quad \mathbf{p}_c = \mathbf{R}^T(\mathbf{p}_w - \mathbf{t}).$$

In homogenen Koordinaten gilt entsprechend:

$$\underline{\mathbf{p}}_c = \underbrace{\begin{pmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \\ \mathbf{0}_3^T & 1 \end{pmatrix}}_{=: \mathbf{F}} \underline{\mathbf{p}}_{w'}$$

dabei ist $\mathbf{0}_3^T := (0 \ 0 \ 0)$ und $\mathbf{F}$ die Matrix der extrinsischen Kameraparameter, gem. Gleichung (2.1).

2.1.4 Projektionsmodelle

Als letzte Größe aus Gleichung (2.1) muss noch die Projektionsmodellmatrix $\mathbf{P}_M$ definiert werden. Das perspektivische Projektionsmodell entspricht dem Lochkameramodell und dient als Approximation der Projektion mit Linsen.

Ein Nachteil ist jedoch, dass die perspektivische Projektion in kartesischen Koordinaten eine nichtlineare Abbildung ist, eine Vielzahl von Algorithmen im Rechnersehen jedoch lineare oder affine Abbildungen voraussetzen. Dies betrifft in dieser Arbeit z. B. die Verwendung des Kalman-Filters (vgl. Abschnitt 4.1). Deswegen wird das perspektivische Projektionsmodell durch affine Projektionen approximiert. Drei affine Projektionsmodelle sind im Rechnersehen weit verbreitet: orthogonales, schwach perspektivisches und paraperspektivisches Projektionsmodell. Das orthogonale und das schwach perspektivische Modell sind nur aus Gründen der Vollständigkeit erwähnt; sie werden in dieser Arbeit nicht verwendet.

2.1.4.1 Perspektivisches Projektionsmodell

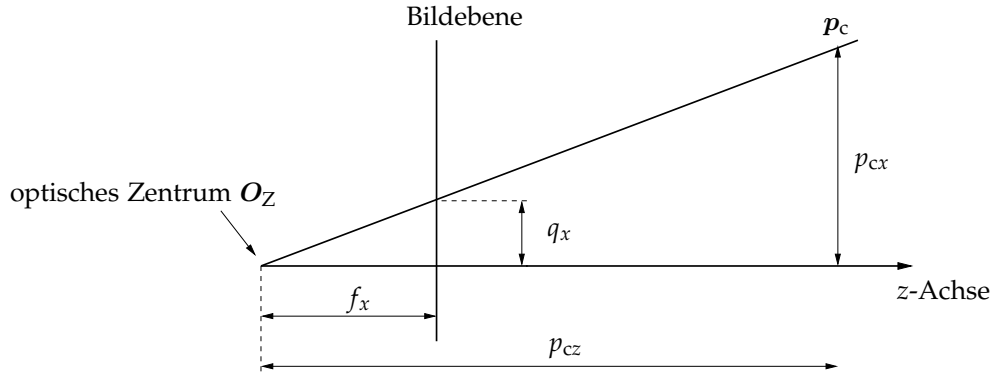
Die perspektivische Projektion eines Punkts in Kamerakoordinaten auf die Bildebene ist schematisch in Bild 2.1 dargestellt. Daraus lässt sich leicht die Beziehung für die einzelnen Koordinaten aufstellen:

$$q_x = f \frac{p_{cx}}{p_{cz}}, \quad q_y = f \frac{p_{cy}}{p_{cz}}. \quad (2.3)$$

Liegt der Hauptpunkt nicht auf der optischen Achse so müssen die Bildkoordinaten noch verschoben werden:

$$q_x = f \frac{p_{cx}}{p_{cz}} + u_p, \quad q_y = f \frac{p_{cy}}{p_{cz}} + v_p. \quad (2.4)$$

Die perspektivische Projektion ist eine nichtlineare Abbildung. Mittels homogenen Koordinaten lässt sich die Projektion jedoch als lineare Abbildung schreiben.

Bild 2.1 Perspektivische Projektion des Punkts p_c **Satz 2.3**

Die Projektionsmodellmatrix für die perspektivische Projektion in homogenen Koordinaten, gem. Gleichung (2.1) lautet:

$$P_M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (2.5)$$

Damit ist die Projektion von homogenen Weltkoordinaten in homogene Bildkoordinaten als lineare Abbildung darstellbar, wie in Gleichung (2.1).

Beweis: durch einsetzen erhält man²:

$$K P_M \underline{p}_c = \begin{pmatrix} f_x & 0 & u_p & 0 \\ 0 & f_y & v_p & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_{cx} \\ p_{cy} \\ p_{cz} \\ 1 \end{pmatrix} = \begin{pmatrix} f_x p_{cx} + u_p p_{cz} \\ f_y p_{cy} + v_p p_{cz} \\ p_{cz} \\ 1 \end{pmatrix}.$$

Durch Umrechnung der homogenen in kartesische Koordinaten ergeben sich die Gleichungen 2.4.

Gleichung (2.1) kann verkürzt dargestellt werden:

$$\underline{q} \sim P \underline{p}_w \sim K [R^T, -R^T t] \underline{p}_w. \quad (2.6)$$

Dabei steht $[A, b]$ für eine Matrix, die aus der Matrix A besteht, die um eine Spalte erweitert worden ist, in der die Komponenten des Vektors b stehen.

2.1.4.2 Paraperspektivisches Projektionsmodell

Das paraperspektivische Projektionsmodell, wurde in [Oht81] eingeführt [Poe97]. Seinen Namen bekam es erst später durch die Veröffentlichung [Alo90]. Es stellt eine Annäherung an das perspektivische Projektionsmodell dar, ist in kartesischen Koordinaten eine affine Abbildung.

Für die Berechnung der paraperspektivischen Projektion muss der Schwerpunkt der zu projizierenden 3-D-Punkte bekannt sein. Projiziert man mehrere 3-D-Punkte mit Hilfe des paraperspektivischen Projektionsmodells, dann kann der Schwerpunkt aus dem arithmetischen Mittel dieser 3-D-Punkte berechnet werden. Beim

²man beachte der Skew Parameter $\beta = 0$, wie oben erläutert

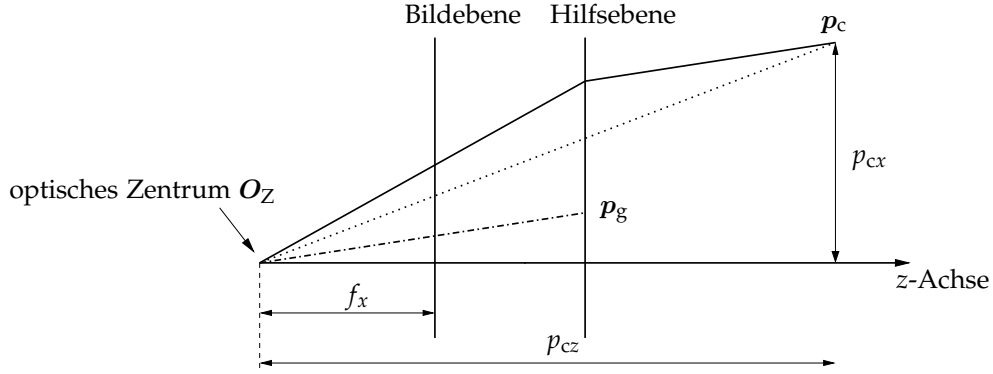


Bild 2.2 Paraperspektivische Projektion des Punkts p_c . Zunächst wird parallel zur Punkt-Strich-Linie auf die Hilfsebene projiziert, von dort aus perspektivisch auf die Bildebene. Zum Vergleich, wie gut die Approximation sich der perspektivischen Projektion annähert, ist diese als gepunktete Linie eingezeichnet.

umgekehrten Weg, der Rekonstruktion der 3-D-Punkte aus mehreren Kamerabilddern, muss der Schwerpunkt mitberechnet werden, wie z. B. in [Poe97]. In dieser Arbeit wird der Schwerpunkt durch einen a priori Schätzwert festgelegt. Details hierzu werden in den Abschnitten 4.1.2.4 und 4.2 genannt.

Im paraperspektivischen Projektionsmodell verläuft die Abbildung wie folgt: Zunächst wird eine Hilfsebene parallel zur Bildebene aufgespannt, die durch den Schwerpunkt der abzubildenden Punkte $p_g = (p_{gx} \ p_{gy} \ p_{gz})$ verläuft. Der Punkt p_c wird parallel zur Verbindungsgeraden des optischen Zentrums O_Z mit dem Schwerpunkt p_g auf die Hilfsebene und anschließend perspektivisch auf die Bildebene projiziert. Die Vorgehensweise ist in Bild 2.2 skizziert. Hier sieht man auch die Abweichung zwischen paraperspektivischer und perspektivischer Projektion.

Satz 2.4

Die Projektionsmodellmatrix für die paraperspektivische Projektion lautet:

$$P_M = \begin{pmatrix} 1 & 0 & -\frac{p_{gx}}{p_{gz}} & p_{gx} \\ 0 & 1 & -\frac{p_{gy}}{p_{gz}} & p_{gy} \\ 0 & 0 & 0 & p_{gz} \end{pmatrix}. \quad (2.7)$$

Beweis: Die Gerade parallel zu der Strecke $[O_Z, p_g]$ durch den Punkt p_c in Bild 2.2 besitzt die Gleichung

$$x = \frac{p_{gx}}{p_{gz}}(z - p_{cz}) + p_{cx}.$$

Schneidet man diese mit der Hilfsebene $z = p_{gz}$, so erhält man die x-Koordinate des Schnittpunkts:

$$\frac{p_{gx}}{p_{gz}}(p_{gz} - p_{cz}) + p_{cx}.$$

Der Schnittpunkt wird nun perspektivisch projiziert, wie anhand von Gleichung (2.3) und man erhält für die x-Koordinate:

$$q_x = f_x \frac{\frac{p_{gx}}{p_{gz}}(p_{gz} - p_{cz}) + p_{cx}}{p_{gz}} + u_p.$$

Dies ist identisch zu $\mathbf{K} \mathbf{P} \underline{p}_c$ und Umrechnung in kartesische Koordinaten. Der Beweis für die y-Koordinate ist völlig analog.

Das paraperspektivische Projektionsmodell stellt eine affine Abbildung in kartesischen Koordinaten dar, dies lässt sich wie folgt zeigen: Die Umrechnung von Welt- in Kamerakoordinaten gem. Definition 2.2 ist offensichtlich affin. Es bleibt also noch zu zeigen, dass die eigentliche Projektion der Kamerakoordinaten auf die Bildebene ebenfalls affin ist. Es gilt gem. Gleichung (2.1)

$$\underline{q} = \mathbf{K} \mathbf{P}_M \underline{p}_c = \begin{pmatrix} f_x(p_{cx} + p_{gx}) - \frac{f_x p_{gx}}{p_{gz}} p_{cz} + u_p p_{gz} \\ f_y(p_{cy} + p_{gy}) - \frac{f_y p_{gy}}{p_{gz}} p_{cz} + v_p p_{gz} \\ p_{gz} \end{pmatrix}.$$

Diese lässt sich in kartesische Koordinaten umrechnen:

$$\mathbf{q} = \begin{pmatrix} f_x \frac{p_{cx} + p_{gx}}{p_{gz}} - f_x \frac{p_{gx}}{p_{gz}^2} p_{cz} + u_p \\ f_y \frac{p_{cy} + p_{gy}}{p_{gz}} - f_y \frac{p_{gy}}{p_{gz}^2} p_{cz} + v_p \end{pmatrix}. \quad (2.8)$$

Die einzelnen Komponenten von $\mathbf{q}$ in Gleichung (2.8) bestehen aus einer Linearkombination aus den Komponenten des 3-D-Punkts $\underline{p}_c$ und damit ist die Abbildung, wie oben bereits gezeigt linear in den Weltkoordinaten. Dies war zu zeigen.

2.1.4.3 Zusammenhang zwischen perspektivischer und paraperspektivischer Projektion

Es existiert ein analytischer Zusammenhang zwischen perspektivischer und paraperspektivischer Projektion. Auf diesen wird hier kurz eingegangen, da er in Kapitel 4 von Relevanz ist.

Die perspektivische Projektion ist in kartesischen Koordinaten eine nichtlineare Abbildung $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$. Es gilt gemäß Gleichung (2.4):

$$\mathbf{q} = \phi(\underline{p}_c) = \begin{pmatrix} \phi_x(\underline{p}_c) \\ \phi_y(\underline{p}_c) \end{pmatrix} = \begin{pmatrix} f_x \frac{p_{cx}}{p_{cz}} + u_p \\ f_y \frac{p_{cy}}{p_{cz}} + v_p \end{pmatrix}. \quad (2.9)$$

Die Funktion ϕ ist für $p_{cz} > 0$ (d. h. der zu projizierende Punkt liegt vor der Kamera) beliebig oft partiell differenzierbar. Deshalb ist für eine affine Approximation eine Taylor-Reihen-Entwicklung im konvexen Gebiet $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ naheliegend. Mit dem Schwerpunkt der 3-D-Punkte $\underline{p}_g$ als Entwicklungspunkt ergibt sich [Mey01, Seite 395]:

$$\phi(\underline{p}_c) = \phi(\underline{p}_g) + \mathbf{J}(\phi(\underline{p}_g)) \cdot (\underline{p}_c - \underline{p}_g) + \rho_2(\underline{p}_c, \underline{p}_g), \quad (2.10)$$

mit der Jacobi-Matrix

$$\mathbf{J}(\phi(\underline{p}_g)) = \begin{pmatrix} \frac{\partial}{\partial p_{cx}} \phi_x & \frac{\partial}{\partial p_{cy}} \phi_x & \frac{\partial}{\partial p_{cz}} \phi_x \\ \frac{\partial}{\partial p_{cx}} \phi_y & \frac{\partial}{\partial p_{cy}} \phi_y & \frac{\partial}{\partial p_{cz}} \phi_y \end{pmatrix} \quad (2.11)$$

und dem Restglied $\rho_2(p_c, p_g)$ der Ordnung³ $o(p_c - p_g)$. Da der Schwerpunkt der 3-D-Punkte p_g zu allen anderen im Mittel den kleinsten Abstand hat, eignet er sich als Entwicklungspunkt gut. Dadurch wird der Fehler bei Vernachlässigung des Rests $\rho_2(p_c, p_g)$ klein. Durch Nachrechnen ergibt sich:

Satz 2.5

Die Taylor-Reihen-Approximation 1. Ordnung der perspektivischen Projektion, mit dem Schwerpunkt der 3-D-Punkte als Entwicklungspunkt, entspricht der paraperspektivischen Projektion.

Beweisskizze: Einsetzen von Gleichung (2.11) in Gleichung (2.10) und Vergleich mit dem Ergebnis aus Gleichung (2.8).

2.2 Kamerakalibrierung

Unter Kamerakalibrierung fasst man die Bestimmung der in- und extrinsischen Kameraparameter zusammen. Die Kalibrierverfahren werden in zwei verschiedene Kategorien eingeteilt [Zha00]: photogrammetrische Kalibrierung und Selbstkalibrierung. Unter Selbstkalibrierung versteht man die Ermittlung der Parameter ohne Kalibrierobjekt. Algorithmen zur Selbstkalibrierung, z. B. [Pol99], sind nur aus Gründen der Vollständigkeit erwähnt; sie sind in dieser Arbeit ohne Relevanz.

Bei der photogrammetrischen Kamerakalibrierung wird ein sogenanntes Kalibrierobjekt verwendet. Das ist ein Objekt, dessen 3-D-Geometrie bekannt ist, z. B. das Kalibriermuster aus Bild 2.3. Die Genauigkeit der Kalibrierung hängt im Wesentlichen von zwei Faktoren ab:

- a) von der Genauigkeit mit der das Kalibriermuster gefertigt wurde. Dabei gilt die Faustregel [Tru98]: Die Kalibriermustergenauigkeit sollte 1-2 Größenordnungen größer sein als die gewünschte Kalibriergenauigkeit.
- b) wie gut die vom Kalibrierobjekt bekannte 3-D-Struktur im Bild detektiert werden kann. Hier sind die Kalibrierpunkte durch die Kreise des Kalibriermusters aus Bild 2.3 gegeben, für die in [Vog06] eine automatische Erkennung vorgeschlagen wird.

Für die Experimente (Kapitel 6) wurde das Kalibriermuster aus Bild 2.3 verwendet. Für die Bestimmung der Kameraparameter werden die 7×7 Kreise, die 20 mm voneinander entfernt sind, verwendet. Die kleineren Kreise an den Ecken des Kalibriermusters dienen dazu, die Orientierung der Kamera relativ zum Kalibriermuster zu bestimmen. Ohne die Asymmetrie sind die extrinsischen Kameraparameter nicht eindeutig zu bestimmen.

Im Folgenden wird nun ein Kalibrierverfahren vorgestellt, das auf den Ausführungen von [Tsa87, Zha00, Vog06] basiert.

³für das Landau-Symbol o wird die Definition aus [Wal02, Seite 64] verwendet: „Die reellwertigen Funktionen f und g seien auf einer Teilmenge D eines metrischen Raumes mit dem Häufungspunkt ξ erklärt, f habe Werte in einem normierten Raum, g sei reellwertig und positiv. Dann schreiben wir [...] $f(x) = o(g(x))$ für $x \rightarrow \xi$, wenn $\lim_{x \rightarrow \xi} \frac{|f(x)|}{g(x)} = 0$ ist.“

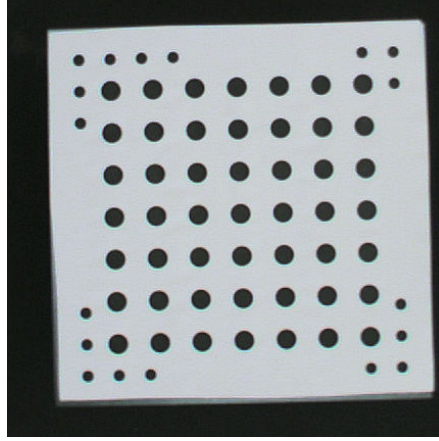


Bild 2.3 Kalibriermuster: Die 49 großen Kreise dienen zur Kalibrierung, die kleinen an den Ecken ermöglichen die Bestimmung der Orientierung

2.2.1 Schätzung der Homographie

Die Projektionsmatrix $\mathbf{P}$ beschreibt die Abbildung eines 3-D-Punkts in homogenen Weltkoordinaten $\underline{p}_w$ in 2-D-Bildkoordinaten $\underline{q}$ gem. Gleichung (2.1). Für die Berechnung müssen die Koordinaten des Bildpunkts $\underline{q}$ und die des zugehörigen 3-D-Punkts $\underline{p}_w$ bekannt sein. Die Bestimmung der Koordinaten der Bildpunkte wird über die automatische Detektion der Kalibrierpunkte des Kalibrierusters realisiert [Vog06]. Die Koordinaten der zugehörigen 3-D-Punkte lassen sich durch eine geeignete Festlegung des Weltkoordinatensystems bestimmen: die x - y -Ebene ist identisch mit der Ebene, in der die Kreise liegen, die x -Achse verläuft parallel zu einer Reihe von Kreisen und der Ursprung liegt in der Mitte des Kalibrierusters. Damit können die Weltkoordinaten der Kreismittelpunkte bestimmt werden.

Die Projektionsmatrix $\mathbf{P}$ ist eine 3×4 Matrix mit 11 Freiheitsgraden. Aufgrund der Wahl des Weltkoordinatensystems sind alle z -Komponenten der 3-D-Punkte gleich Null. Dadurch lässt sich Gleichung (2.1) vereinfachen; es genügt zunächst eine 3×3 Homographie $\mathbf{L}$ zu schätzen:

$$\nu \begin{pmatrix} q_x \\ q_y \\ 1 \end{pmatrix} = \mathbf{K} \begin{pmatrix} {}^1\tilde{\mathbf{r}} & {}^2\tilde{\mathbf{r}} & {}^3\tilde{\mathbf{r}} & \tilde{\mathbf{t}} \end{pmatrix} \begin{pmatrix} p_{wx} \\ p_{wy} \\ 0 \\ 1 \end{pmatrix} = \underbrace{\mathbf{K} \begin{pmatrix} {}^1\tilde{\mathbf{r}} & {}^2\tilde{\mathbf{r}} & \tilde{\mathbf{t}} \end{pmatrix}}_{:=\mathbf{L}} \begin{pmatrix} p_{wx} \\ p_{wy} \\ 1 \end{pmatrix}, \quad (2.12)$$

mit ${}^1\tilde{\mathbf{r}}$ der 1., ${}^2\tilde{\mathbf{r}}$ der 2. bzw. ${}^3\tilde{\mathbf{r}}$ der 3. Spalte der Rotationsmatrix $\tilde{\mathbf{R}} = \mathbf{R}^T$, dem Translationsvektor $\tilde{\mathbf{t}} = -\mathbf{R}^T \mathbf{t}$ und der Skalierung $\nu \neq 0$. Da die Homographie $\mathbf{L}$ nur bis auf einen Skalierungsfaktor eindeutig ist, besitzt sie zunächst 8 Freiheitsgrade. Es gelten jedoch noch zwei weitere Bedingungen: da die Vektoren ${}^1\tilde{\mathbf{r}}$ und ${}^2\tilde{\mathbf{r}}$ Spaltenvektoren einer Rotationsmatrix sind, gilt:

$${}^1\tilde{\mathbf{r}}^T {}^2\tilde{\mathbf{r}} = {}^1\mathbf{l}^T (\mathbf{K}^T)^{-1} \mathbf{K}^{-1} {}^2\mathbf{l} = 0, \quad (2.13)$$

$${}^1\tilde{\mathbf{r}}^T {}^1\tilde{\mathbf{r}} = {}^1\mathbf{l}^T (\mathbf{K}^T)^{-1} \mathbf{K}^{-1} {}^1\mathbf{l} = {}^2\mathbf{l}^T (\mathbf{K}^T)^{-1} \mathbf{K}^{-1} {}^2\mathbf{l} = {}^2\tilde{\mathbf{r}}^T {}^2\tilde{\mathbf{r}}. \quad (2.14)$$

Dabei sind ${}^1\mathbf{l}$, ${}^2\mathbf{l}$ und ${}^3\mathbf{l}$ die Spalten von $\mathbf{L}$. Diese Bedingungen müssen die intrinsischen Parameter erfüllen [Zha00]. Da die Homographie $\mathbf{L}$ 8 Freiheitsgrade be-

sitzt und davon 3 für die Rotation und 3 für die Translation sind, sind diese beiden Bedingungen die einzigen für die intrinsischen Parameter.

Es existieren verschiedene Wege, um die Homographie aus Gleichung (2.12) zu schätzen. Zhang [Zha00] schlägt eine Maximum-Likelihood-Schätzung vor, da die Koordinaten der Bildpunkte verrauscht sind. Unter der Annahme, das Rauschen ist ein mittelwertfreies Gauß-Rauschen mit Kovarianzmatrix Σ , folgt das zugehörige Minimierungsproblem für n Punkte:

$$\mathbf{L}^* = \underset{\mathbf{L}}{\operatorname{argmin}} \sum_{i=1}^n (\mathbf{q}_i - \hat{\mathbf{q}}_i)^T \Sigma_i^{-1} (\mathbf{q}_i - \hat{\mathbf{q}}_i). \quad (2.15)$$

Dabei ist $\mathbf{L}^*$ die optimale Lösung des Minimierungsproblems, $\mathbf{q}_i = \mathbf{q}_i(\mathbf{L})$ der i -te Punkt, der durch Gleichung (2.12) ins Bild projiziert wurde und $\hat{\mathbf{q}}_i$ der detektierte Punkt. Unter der Annahme, die Kovarianz sei von der Form $\Sigma_i = \sigma^2 \mathbf{I}_{2 \times 2}$, vereinfacht sich das Minimierungsproblem (2.15):

$$\mathbf{L}^* = \underset{\mathbf{L}}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{q}_i - \hat{\mathbf{q}}_i\|_2^2, \quad (2.16)$$

dabei bezeichnet $\|\mathbf{x}\|_2^2$ das Quadrat der euklidische Norm des Vektors $\mathbf{x}$.

Gleichung (2.12) wird verwendet, um eine geeignete lineare Lösung zu berechnen, die als Startlösung für eine anschließende nichtlineare Optimierung verwendet werden kann. Allgemein gilt im nichtverrauschten Fall:

$$\mathbf{q}_i - \hat{\mathbf{q}}_i = \mathbf{0}_2. \quad (2.17)$$

Durch Umrechnung der Gleichung (2.12) in kartesische Koordinaten, Einsetzen in Gleichung (2.17) und Umformungen erhält man

$$\begin{pmatrix} \mathbf{p}_{w_i}^T & \mathbf{0}_3^T & -\hat{\mathbf{q}}_{ix} \mathbf{p}_{w_i}^T \\ \mathbf{0}_3^T & \mathbf{p}_{w_i}^T & -\hat{\mathbf{q}}_{iy} \mathbf{p}_{w_i}^T \end{pmatrix} \begin{pmatrix} {}_1\mathbf{l}^T \\ {}_2\mathbf{l}^T \\ {}_3\mathbf{l}^T \end{pmatrix} = \mathbf{0}_2, \quad (2.18)$$

mit den zwei- bzw. dreidimensionalen Nullvektoren $\mathbf{0}_2$ bzw. $\mathbf{0}_3$ sowie den Zeilen ${}_1\mathbf{l}$, ${}_2\mathbf{l}$ bzw. ${}_3\mathbf{l}$ von $\mathbf{L}$. Da die Homographie $\mathbf{L}$ über 8 Freiheitsgrade verfügt, lässt sich mit 4 Punktkorrespondenzen eine Lösung errechnen. Diese kann mit Hilfe einer nichtlinearen Optimierung weiter verbessert werden [Zha00]. In dieser Arbeit wird hierzu das Levenberg-Marquardt-Verfahren [Pre02, Sch06b] eingesetzt.

2.2.2 Bestimmung der Kameraparameter aus der Homographie

Aus der Homographie $\mathbf{L}$ können die einzelnen Kameraparameter bestimmt werden. Sei

$$\mathbf{C}_a = (\mathbf{K}^T)^{-1} \mathbf{K}^{-1} = \begin{pmatrix} \frac{1}{f_x^2} & 0 & -\frac{u_p}{f_x^2} \\ 0 & \frac{1}{f_y^2} & -\frac{v_p}{f_y^2} \\ -\frac{u_p}{f_x^2} & -\frac{v_p}{f_y^2} & \left(\frac{u_p}{f_x}\right)^2 + \left(\frac{v_p}{f_y}\right)^2 + 1 \end{pmatrix}. \quad (2.19)$$

Das entspricht dem Bild der absoluten Konik [Har03, Luo97]. Nun lassen sich zunächst die intrinsischen Parameter mit Hilfe der beiden Nebenbedingungen (2.13)

und (2.14) bestimmen: Die Nebenbedingungen sind von der Form ${}^i\mathbf{l}^T \mathbf{C}_a {}^j\mathbf{l} = 0$, mit $i, j = 1, 2$. Dies lässt sich auch umformen:

$${}^i\mathbf{l}^T \mathbf{C}_a {}^j\mathbf{l} = \mathbf{k}_{ij}^{*T} \mathbf{c} \quad (2.20)$$

mit

$$\mathbf{k}_{ij}^{*T} = (l_{i1}l_{j1} \quad l_{i1}l_{j2} + l_{i2}l_{j1} \quad l_{i2}l_{j2} \quad l_{i3}l_{j1} + l_{i1}l_{j3} \quad l_{i3}l_{j2} + l_{i2}l_{j3} \quad l_{i3}l_{j3})$$

und dem Vektor $\mathbf{c} = (c_{11} \quad c_{12} \quad c_{22} \quad c_{13} \quad c_{23} \quad c_{33})^T$, der die Komponenten der absoluten Konik $\mathbf{C}_a$ enthält. Die Nebenbedingungen (2.13) und (2.14) können nun als homogenes Gleichungssystem formuliert werden:

$$\begin{pmatrix} \mathbf{k}_{12}^{*T} \\ (\mathbf{k}_{11}^* - \mathbf{k}_{22}^*)^T \end{pmatrix} \mathbf{c} = \mathbf{0}_2. \quad (2.21)$$

Verwendet man n Bilder, so lassen sich die Gleichungen aus (2.21) übereinander anordnen und man erhält ein Gleichungssystem der Form

$$\mathbf{K}^* \mathbf{c} = \mathbf{0}_{2n}, \quad \mathbf{K}^* \in \mathbb{R}^{2n \times 6}. \quad (2.22)$$

Die Matrix $\mathbf{K}^*$ enthält dabei die Parameter der Homographie $\mathbf{L}$.

Für einen eindimensionalen Lösungsraum des homogenen Gleichungssystems benötigt man also mindestens 3 Bilder. Damit lässt sich der Vektor $\mathbf{c}$ bis auf einen Skalierungsfaktor eindeutig bestimmen. Eine Lösung lässt sich mit einem mathematische Standardverfahren berechnen. In dieser Arbeit wird hierzu die Singulärwertzerlegung [Gol96, Tre97, Sch06b] verwendet. Die Matrix $\mathbf{K}^*$ wird zerlegt in

$$\mathbf{K}^* = \mathbf{U}_l \mathbf{D}_s \mathbf{V}_r^T. \quad (2.23)$$

Die Spaltenvektoren von $\mathbf{V}_r$, die mit den Singulärwerten gleich Null korrespondieren, bilden eine Basis des Nullraums von $\mathbf{K}^*$. Da der Nullraum hier eindimensional ist, darf nur ein Singulärwert gleich Null sein und die gesuchte Lösung für $\mathbf{c}$ ist gleich dem letzten Spaltenvektor von $\mathbf{V}_r$.

Mit Hilfe des Vektors $\mathbf{c}$ lassen sich über die Gleichung (2.19) die einzelnen intrinsischen Parameter schätzen. Da $\mathbf{c}$ nur bis auf einen Skalierungsfaktor bestimmt werden kann, setzt man nun

$$\mathbf{C}_a = \nu (\mathbf{K}^T)^{-1} \mathbf{K}^{-1}.$$

Dabei modelliert ν die Skalierung. Das nichtlineare Gleichungssystem lässt sich aber eindeutig lösen und man erhält:

$$\nu = c_{33} - \left(\frac{c_{11}^3}{c_{13}^2} + \frac{c_{22}^3}{c_{23}^2} \right), \quad (2.24)$$

$$u_p = -\frac{c_{11}}{c_{13}}, \quad v_p = -\frac{c_{22}}{c_{23}}, \quad (2.25)$$

$$f_x = \sqrt{\frac{\nu}{c_{11}}}, \quad f_y = \sqrt{\frac{\nu}{c_{22}}}. \quad (2.26)$$

Damit sind die effektiven Brennweiten und der Hauptpunkt bestimmt.

Die bisherigen Schritte werden noch einmal kurz zusammengefasst: Man muss die 3-D-Struktur des Kalibrieramusters kennen, um die Weltkoordinaten der Punkte auf dem Kalibrieramuster bestimmen zu können. Weiterhin werden die Bildkoordinaten der Kalibrierpunkte bestimmt. Aus Gleichung (2.18) kann mit diesen Daten die Homographie $\mathbf{L}$ geschätzt werden. Mit Gleichung (2.21) können dann die Komponenten der absoluten Konik bestimmt werden und daraus wiederum die einzelnen Parameter der Kameramatrix $\mathbf{K}$ über die Gleichungen (2.24)-(2.26). Nun sind noch die extrinsischen Parameter und die Verzerrungskoeffizienten zu bestimmen.

Die extrinsischen Parameter lassen sich bei bekannter Matrix $\mathbf{K}$ direkt aus der Definition der Homographie $\mathbf{L}$ (vgl. Gleichung (2.12)) bestimmen. Da die Homographie nur bis auf einen Skalierungsfaktor ζ zu bestimmen ist, wird angesetzt:

$$\begin{pmatrix} {}^1\mathbf{l} & {}^2\mathbf{l} & {}^3\mathbf{l} \end{pmatrix} = \zeta \mathbf{K} \begin{pmatrix} {}^1\tilde{\mathbf{r}} & {}^2\tilde{\mathbf{r}} & \tilde{\mathbf{t}} \end{pmatrix}. \quad (2.27)$$

Die Matrix $\mathbf{K}$ ist eindeutig und die Vektoren ${}^1\tilde{\mathbf{r}}$ und ${}^2\tilde{\mathbf{r}}$ sind die Spaltenvektoren der Rotationsmatrix $\tilde{\mathbf{R}}$. Deshalb wird der Skalierungsfaktor so gesetzt, dass er die Vektoren auf Länge 1 normiert:

$$\zeta = \frac{1}{\|\mathbf{K}^{-1} {}^1\mathbf{l}\|} = \frac{1}{\|\mathbf{K}^{-1} {}^2\mathbf{l}\|}. \quad (2.28)$$

Es gilt dann für die extrinsischen Parameter:

$${}^1\tilde{\mathbf{r}} = \zeta \mathbf{K}^{-1} {}^1\mathbf{l}, \quad {}^2\tilde{\mathbf{r}} = \zeta \mathbf{K}^{-1} {}^2\mathbf{l}, \quad {}^3\tilde{\mathbf{r}} = {}^1\tilde{\mathbf{r}} \times {}^2\tilde{\mathbf{r}}, \quad \tilde{\mathbf{t}} = \zeta \mathbf{K}^{-1} {}^3\mathbf{l}.$$

Die so errechnete Rotationsmatrix $\tilde{\mathbf{R}}$ ist, aufgrund von verrauschten Eingangsdaten, im Allgemeinen nicht orthogonal. Mit Hilfe der Singulärwertzerlegung [Gol96, Tre97, Sch06b] lässt sich $\tilde{\mathbf{R}}$ jedoch orthogonalisieren. Die Rotationsmatrix $\tilde{\mathbf{R}}$ wird zerlegt:

$$\tilde{\mathbf{R}} = \mathbf{U}_l \mathbf{D}_S \mathbf{V}_r^T. \quad (2.29)$$

Die Diagonalmatrix $\mathbf{D}_S$ enthält die Singulärwerte von $\tilde{\mathbf{R}}$. Weichen die Singulärwerte von $\tilde{\mathbf{R}}$ von 1 ab, so setzt man $\mathbf{D}_S = \mathbf{I}$ und berechnet die Matrix $\tilde{\mathbf{R}}$ gemäß Gleichung (2.29) erneut. Die so berechnete neue Rotationsmatrix $\tilde{\mathbf{R}}$ besitzt die Eigenschaft einer Rotationsmatrix (Orthogonalität) und hat im Sinne der Frobeniusnorm den geringsten Abstand zur ursprünglich berechneten Matrix $\tilde{\mathbf{R}}$.

2.2.3 Nichtlineare Optimierung

Die gefundene Lösung kann durch eine nichtlineare Optimierung verbessert werden. Hat man n Punkte und m Bilder, so muss folgender Ausdruck minimiert werden:

$$\min_{\mathbf{K}, \mathbf{R}, \mathbf{t}, \kappa_1, \kappa_2, \vartheta_1, \vartheta_2} \sum_{i=1}^n \sum_{j=1}^m \|q_{ij} - \phi(p_{wi}, \mathbf{K}, \mathbf{R}, \mathbf{t}, \kappa_1, \kappa_2, \vartheta_1, \vartheta_2)\|. \quad (2.30)$$

Dabei werden die Verzerrungskoeffizienten $\kappa_1, \kappa_2, \vartheta_1, \vartheta_2$ zunächst mit 0 initialisiert. Die Lösung kann mit einem nichtlinearen Minimierungsverfahren berechnet

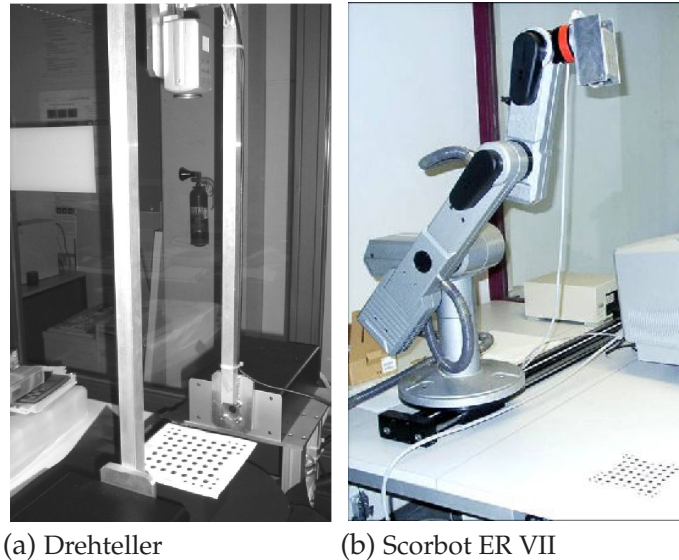


Bild 2.4 Links: Kamera an einem Schwenkarm über einem Drehteller. Rechts: Der Scorbot ER VII Roboterarm mit Kamera.

werden. In dieser Arbeit wird hierzu das Verfahren nach Levenberg-Marquardt [Pre02, Sch06b] verwendet.

Damit können die Kameraparameter, effektive Brennweiten, Hauptpunkt, Rotationsmatrix und Translationsvektor und die vier Verzerrungsparameter berechnet werden. Wie oben bereits erwähnt, wird der Skewparameter β auf Null gesetzt.

2.3 Hand-Auge-Transformation

Wie bereits im Abschnitt 1.1 festgestellt wurde, werden für eine 3-D-Rekonstruktion Aufnahmen der Szene aus unterschiedlichen Positionen benötigt. Bei den Aufnahmen müssen sich folglich die extrinsischen Kameraparameter ändern.

Die Bestimmung der extrinsischen Kameraparameter mittels Kalibriermuster führt aber in der Praxis zu folgenden Problemen: Das Kalibriermuster darf während der gesamten Aufnahme-prozedur nicht bewegt werden, da die Kameraposition und Orientierung relativ zum Weltkoordinatensystem angegeben werden, welches durch das Kalibriermuster festgelegt wird. Es muss aber in jeder Aufnahme sichtbar sein. Dies ist häufig aber nicht möglich, da z. B. Szenen zu groß sind und deshalb nur Teile aufgenommen werden können, oder weil das Kalibriermuster durch Verdeckungen nicht sichtbar ist. Es ist daher von Vorteil, wenn man die extrinsischen Parameter ohne Kalibriermuster ermitteln kann.

Ist die Kamera an einem mechanischen Arm befestigt, wie in Bild 2.4, lassen sich die extrinsischen Parameter ohne Kalibriermuster bestimmen: Position und Orientierung der Roboterhand ist durch die an den Gelenken des Roboters eingestellten Winkel definiert (vgl. Abschnitt 2.4). Die extrinsischen Parameter der Kamera ergeben sich aus der Lage der Roboterhand und der Hand-Auge-Transformation [Den94, Beß96]. Die Hand-Auge-Transformation beschreibt dabei die starre Transformation zwischen der Lage und Orientierung der Roboterhand und der Kamera, die fest auf der Hand montiert ist.

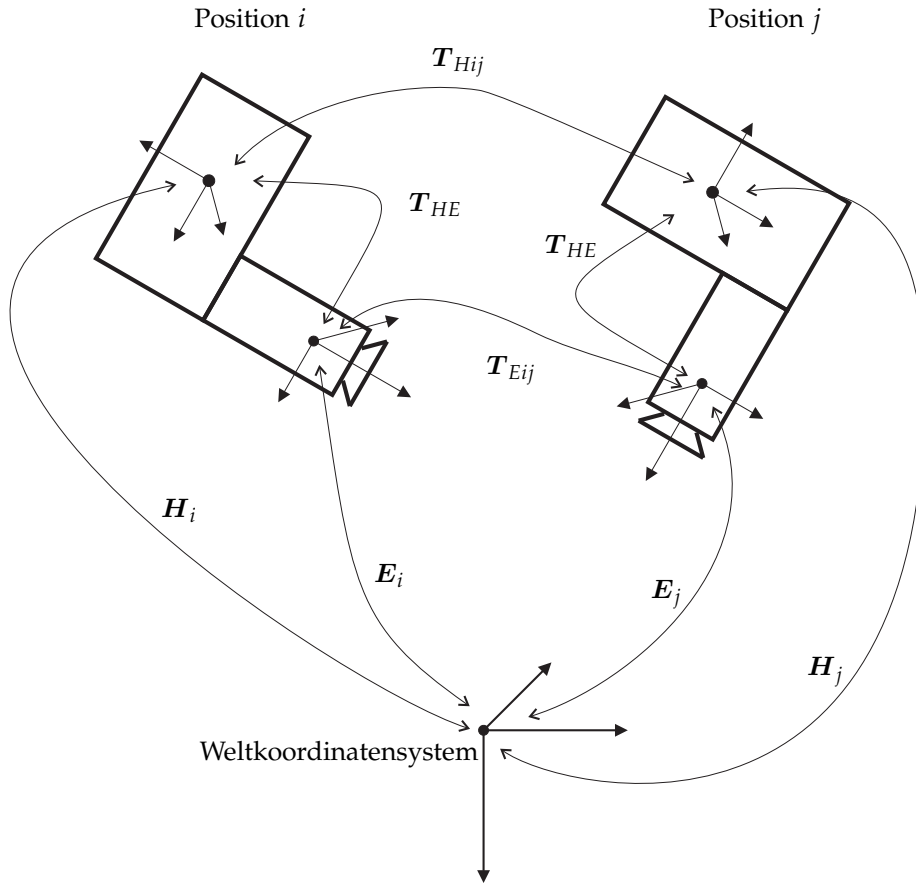


Bild 2.5 Für Position i (links) und Position j (rechts) sind alle Transformationen eingezeichnet. Wichtig für die Hand-Auge-Kalibrierung ist, dass die Hand-Auge-Transformation stets gleich bleibt. Die absoluten Positionsangaben $T_{Hi}, T_{Hj}, T_{Ei}, T_{Ej}$ beziehen sich auf das Weltkoordinatensystem. Zur Bestimmung der Hand-Auge-Transformation genügt es aber die relativen Bewegungen T_{Hij} und T_{Eij} zu kennen.

Die ersten Arbeiten zur Hand-Auge-Kalibrierung wurden von Shiu und Ahmad [Shi89], und von Tsai und Lenz [Tsa89] publiziert. Idee der Arbeiten ist, dass zwischen zwei Aufnahmepositionen i, j die Hand-Auge-Transformation gleich bleibt.

Für die folgenden Überlegungen werden die Transformationen, bestehend aus Rotationen und Translationen des Auges E bzw. der Hand H , in homogenen Koordinaten durchgeführt und dabei die folgenden 4×4 Matrizen verwendet:

$$E = \begin{pmatrix} R_E & t_E \\ 0_3 & 1 \end{pmatrix} \quad \text{bzw.} \quad H = \begin{pmatrix} R_H & t_H \\ 0_3 & 1 \end{pmatrix}. \quad (2.31)$$

Die Hand-Auge-Transformation T_{HE} ist für Positionen i bzw. j gleich:

$$E_i = T_{HE} H_i \quad \wedge \quad E_j = T_{HE} H_j. \quad (2.32)$$

Die Hand-Auge-Transformation in Position i bzw. j ist in Bild 2.5 zusehen. Dort sind zusätzlich auch die Transformationen für die Hand und das Auge eingezeichnet, die im Folgenden verwendet werden.

Wird eine Bewegung von Position i zur Position j durchgeführt, so lässt sich die Transformation T_{Hij} der Hand aufgrund der vom Roboter gelieferten Rotation und

Translation berechnen. Analog muss die Transformation für das Auge T_{Eij} , z. B. durch eine Kamerakalibrierung, bestimmt werden. Damit gelten die Gleichungen:

$$\mathbf{H}_j = \mathbf{T}_{Hij}\mathbf{H}_i \quad \wedge \quad \mathbf{E}_j = \mathbf{T}_{Eij}\mathbf{E}_i. \quad (2.33)$$

Aus Gleichung (2.32) und (2.33) lässt sich jeweils eine Gleichung für $\mathbf{E}_j$ aufstellen:

$$\mathbf{E}_j = \mathbf{T}_{HE}\mathbf{H}_j = \mathbf{T}_{HE}\mathbf{T}_{Hij}\mathbf{H}_i \quad \wedge \quad \mathbf{E}_j = \mathbf{T}_{Eij}\mathbf{E}_i = \mathbf{T}_{Eij}\mathbf{T}_{HE}\mathbf{H}_i. \quad (2.34)$$

Die Existenz der inversen Transformation $\mathbf{H}_i^{-1}$ ist gesichert, denn die Transformation $\mathbf{H}_i$ hat die Form (2.31) und die Determinante ergibt sich nach dem Laplace-Entwicklungssatz: Nach der letzten Zeile entwickelt ergibt sich:

$$\det(\mathbf{H}_i) = \det(\mathbf{R}_{Hi}) = 1, \quad (2.35)$$

da Rotationsmatrizen stets Determinante gleich Eins besitzen. Da die Determinante von $\mathbf{H}_i$ ungleich Null ist, ist $\mathbf{H}_i$ invertierbar. In den Gleichungen (2.34) sind die linken Seiten gleich, daher müssen auch die rechten gleich sein („Was demselben gleich ist, ist auch einander gleich“ [Euk97]). Durch die Multiplikation mit $\mathbf{H}_i^{-1}$ von rechts ergibt sich die Basisgleichung für die Hand-Auge-Kalibrierung:

$$\mathbf{T}_{Eij}\mathbf{T}_{HE} = \mathbf{T}_{HE}\mathbf{T}_{Hij}. \quad (2.36)$$

Gleichung (2.36) ausführlich geschrieben lautet:

$$\begin{pmatrix} \mathbf{R}_{Eij} & \mathbf{t}_{Eij} \\ \mathbf{0}_3 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_{HE} & \mathbf{t}_{HE} \\ \mathbf{0}_3 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{HE} & \mathbf{t}_{HE} \\ \mathbf{0}_3 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_{Hij} & \mathbf{t}_{Hij} \\ \mathbf{0}_3 & 1 \end{pmatrix}. \quad (2.37)$$

Der klassische Weg ist eine Aufspaltung von (2.37) in zwei Gleichungssysteme,

$$\mathbf{R}_{Eij}\mathbf{R}_{HE} = \mathbf{R}_{HE}\mathbf{R}_{Hij}, \quad (2.38)$$

$$(\mathbf{R}_{Eij} - \mathbf{I}_{3 \times 3})\mathbf{t}_{HE} = \mathbf{R}_{HE}\mathbf{t}_{Hij} - \mathbf{t}_{Eij}, \quad (2.39)$$

wobei zunächst aus Gleichung (2.38) die Rotation bestimmt wird. Anschließend kann mit bekannter Rotation aus (2.39) die Translation bestimmt werden.

Es existieren zahlreiche Publikationen mit diversen Varianten zur Berechnung der Hand-Auge-Transformation: In [Shi89, Tsa89] wird die Rotation über die Achse-Winkel-Darstellung parametrisiert, in [Cho91, Wan92] mit Hilfe von Quaternionen [Ham44, Ham47, Ham48]. In [Che91] wird vorgeschlagen, Rotation und Translation simultan zu berechnen. Das hat den Vorteil, dass der Fehler in der Rotation und Translation möglichst klein gehalten wird.

Daniilidis [Dan99, Dan01] formuliert das Problem mit Hilfe von dualen Quaternionen. Dadurch lassen sich Rotation und Translation gleichzeitig bestimmen. Darüber hinaus lassen sich die Gleichungen (2.37) linearisieren, was eine effiziente Bestimmung der Transformation ermöglicht.

Unabhängig welcher Algorithmus für die Bestimmung der Hand-Auge-Transformation eingesetzt wird, eine elementare Bedingung muss berücksichtigt werden: Es sind mindestens zwei Bewegungen nötig und die zugehörigen Rotationsachsen dürfen nicht parallel sein. Andernfalls sind die Gleichungen nicht eindeutig lösbar, wie in [Tsa89] algebraisch und in [Che91] geometrisch bewiesen wird.

Für die Translation lässt sich diese Bedingung in wenigen Schritten zeigen: Alle Punkte, die auf der Rotationsachse liegen, werden durch die Rotation auf sich selbst abgebildet, insbesondere der Richtungsvektor der Rotationsachse:

$$\mathbf{R}e_v = e_v. \quad (2.40)$$

Das bedeutet, der Eigenvektor von $\mathbf{R}$ zum Eigenwert $\lambda = 1$ repräsentiert die Rotationsachse. Damit lässt sich folgender Satz zeigen:

Satz 2.6

Sei $\mathbf{R}$ eine Rotationsmatrix. Dann hat das Gleichungssystem

$$(\mathbf{R} - \mathbf{I})\mathbf{x} = \mathbf{0}. \quad (2.41)$$

die nichttriviale Lösung $\mathbf{x} = \nu e_v$, mit dem Skalierungsfaktor $\nu \in \mathbb{R}$ und dem Eigenvektor e_v zum Eigenwert $\lambda = 1$.

Beweis:

$$(\mathbf{R} - \mathbf{I})\nu e_v = \nu(\mathbf{R}e_v - e_v) = \nu(e_v - e_v) = \mathbf{0}. \quad (2.42)$$

Das Gleichungssystem für die Translation (2.39) ist damit also nicht eindeutig lösbar. Hat man zwei Bewegungen, von Position i nach j und von j nach k , dann lautet das zugehörige homogene Gleichungssystem

$$(\mathbf{R}_{Eij} - \mathbf{I})\mathbf{t}_{HE} = \mathbf{0}, \quad (2.43)$$

$$(\mathbf{R}_{Ejk} - \mathbf{I})\mathbf{t}_{HE} = \mathbf{0}. \quad (2.44)$$

Dabei hat Gleichung (2.43) laut Satz 2.6 den Lösungsraum $\mathbf{t}_{HE} = \nu_{ij}e_{vAij}$ und Gleichung (2.44) den Lösungsraum $\mathbf{t}_{HE} = \nu_{jk}e_{vAjk}$. Da beide Gleichungen gleichzeitig erfüllt werden müssen, ist der zugehörige Lösungsraum die Schnittmenge. Diese Erkenntnisse werden in folgendem Satz zusammengefasst:

Satz 2.7

Sei das homogene Gleichungssystem

$$(\mathbf{R}_{Eij} - \mathbf{I})\mathbf{t}_{HE} = \mathbf{0},$$

$$(\mathbf{R}_{Ejk} - \mathbf{I})\mathbf{t}_{HE} = \mathbf{0},$$

gegeben. Genau dann, wenn die Eigenvektoren von $\mathbf{R}_{Eij}$ und $\mathbf{R}_{Ejk}$ zum Eigenwert $\lambda = 1$ kollinear sind, existiert eine nichttriviale Lösung des Gleichungssystems.

Beweis: Aus der Tatsache, dass eine 3×3 Rotationsmatrix stets einen Eigenwert $\lambda = 1$ besitzt und aus der Definition von Eigenwert und Eigenvektor folgt, dass $\mathbf{t}_{HE}$ Eigenvektor von $\mathbf{R}_{Eij}$ sein muss. Damit existieren nichttriviale Lösungen für

$$(\mathbf{R}_{Eij} - \mathbf{I})\mathbf{t}_{HE} = \mathbf{0}.$$

Analog ist offensichtlich, dass $\mathbf{t}_{HE}$ Eigenvektor von $\mathbf{R}_{Ejk}$ zum Eigenwert $\lambda = 1$ sein muss. Damit ist gezeigt, dass die Eigenvektoren von $\mathbf{R}_{Eij}$ und $\mathbf{R}_{Ejk}$ zum Eigenwert $\lambda = 1$ kollinear sein müssen.

Gegenrichtung: Besitzen $\mathbf{R}_{Eij}$ und $\mathbf{R}_{Ejk}$ kollineare Eigenvektoren e_v zum Eigenwert $\lambda = 1$, so ist $\mathbf{t}_{HE} = \nu e_v$, eine nichttriviale Lösung des Gleichungssystems.

Aus Satz 2.7 folgt: Werden für die Bestimmung der Hand-Auge-Transformation mittels Gleichung (2.37) zwei Bewegungen verwendet, bei denen die jeweiligen Rotationsachsen nicht kollinear sind, dann existiert höchstens eine Lösung des inhomogenen Gleichungssystems. Die Existenz wenigstens einer Lösung ergibt sich aus der geometrischen Herleitung.

Es lässt sich zusammenfassen: Um eine eindeutige Translation bestimmen zu können, müssen mindestens 2 Bewegungen gemacht werden, bei denen die Rotationsachsen nicht parallel sind. Zusätzlich wird in [Tsa89] der mittlere Fehler der Schätzung der Rotation in Abhängigkeit des Winkels, den die beiden Rotationsachsen einschließen, berechnet. Als Ergebnis erhält man: Der Fehler ist indirekt proportional zum Sinus des Winkels. Lässt man den Winkel zwischen den Rotationsachsen gegen Null gehen, so strebt der zu erwartende Fehler gegen unendlich. Für eine gute Hand-Auge-Kalibrierung sollten also die jeweiligen Rotationsachsen einen Winkel nahe 90° einschließen.

Für die Güte der Hand-Auge-Kalibrierung ist es entscheidend, welche Bewegungen vollzogen werden. Steht eine Vielzahl von Aufnahmen zur Verfügung, dann lässt sich durch eine geeignete Datenauswahl das Kalibrierergebnis verbessern. Untersuchungen zur Datenauswahl und dem damit verbundenen Fehler der Hand-Auge-Kalibrierung wurden von Schmidt veröffentlicht [Sch03, Sch06a].

Ist die Hand-Auge-Transformation bestimmt, so lassen sich die für eine 3-D-Rekonstruktion notwendigen extrinsischen Kameraparameter aus Position und Orientierung der Hand berechnen. Der nächste Abschnitt befasst sich nun damit, wie die Handparameter bestimmt werden können.

2.4 Direkte und inverse Kinematik

Wie im obigen Abschnitt beschrieben, lassen sich über die (von einem Roboter bereitgestellten) Positions- und Orientierungsdaten der Hand die extrinsischen Kameraparameter berechnen. Es gibt jedoch Roboter, die keine Rotationsmatrix oder keinen Translationsvektor bereitstellen. Diese müssen dann aus den Gelenken und Gliedern des Roboters (vgl. Bild 2.4b) berechnet werden, was man in der Literatur als „direkte Kinematik“ bezeichnet. Die umgekehrte Fragestellung, die „inverse Kinematik“, befasst sich mit dem Problem für eine gegebene Position und Orientierung der Hand die Gelenkeinstellungen zu finden. Die direkte Kinematik ist im Vergleich zur inversen ein relativ einfaches Problem, wie sich im Folgenden zeigen wird.

2.4.1 Direkte Kinematik

Als direkte Kinematik bezeichnet man die Bestimmung der Transformationsmatrix D , die die Transformation des Koordinatensystems am Fuß des Roboters in das Koordinatensystem der Hand angibt. Diese Transformation wird dabei aus den Einstellungen der Gelenke berechnet. Ein Gelenk ist hier laut Definition die Verbindung von zwei starren Gliedern des Roboters, welches die Lagebeziehung dieser Elemente verändern kann. Es wird davon ausgegangen, dass die hier verwendeten Gelenke ideal sind, d. h. jedes Gelenk ermöglicht eine räumliche Bewegung mit genau einem Freiheitsgrad [Gre97]. Unabhängig davon, wie viele Ge-

lenke ein Roboter besitzt, die Zahl der Freiheitsgrade des Endeffektors (der Hand) ist maximal sechs: je drei Parameter für die Translation und die Rotation.

Eine Folge von Gelenken heißt kinematische Kette. Um eine kinematische Kette zu beschreiben, führt man Koordinatensysteme ein, die den Gelenken zugeordnet sind. Eine Gelenkbewegung bewirkt eine Translation oder auch eine Rotation aller Koordinatensysteme der folgenden Gelenke und des Endeffektors. Außer den Gelenkkoordinatensystemen besitzt eine kinematische Kette noch das Basis- und das Endeffektorkoordinatensystem. Das Basiskoordinatensystem ist ortsfest (außer bei mobilen Robotern) und bildet das Bezugssystem. Eine offene, unverzweigte, kinematische Kette mit N_g Gelenken hat demnach bis zu $N_g + 2$ Koordinatensysteme. Bei stationären Roboterarmen ist es üblich das Basiskoordinatensystem in das erste Gelenk zu legen und nicht mit einer eigenen Nummer zu versehen. In diesem Fall sind nur $N_g + 1$ Koordinatensysteme zu betrachten. Die Gelenke und Koordinatensysteme werden von der Basis zum Endeffektor durchnummeriert. Das Basiskoordinatensystem hat dann die Nummer 0 und liegt im ersten Gelenk, im Gelenk Nummer 2 liegt das Koordinatensystem mit der Nummer 1, ..., im Gelenk Nummer N_g liegt das Koordinatensystem $N_g - 1$, das Endeffektorkoordinatensystem hat die Nummer N_g .

Denavit und Hartenberg schlagen eine Konvention für die Festlegung der Koordinatensysteme vor [Den55, Sie96]:

- Die Koordinatensysteme liegen fest in den Bewegungsachsen.
- Die z_i -Achse wird bei Rotationsgelenken längs der Rotationsachse des Gelenks $i + 1$ gelegt, bei Translationsgelenken längs der Translationsrichtung des Gelenks.
- Die x_i -Achse ist senkrecht zur z_{i-1} -Achse (und zeigt von ihr weg).
- Die y_i -Achse steht senkrecht auf der x_i und der z_i Achse und bildet ein rechtshändiges Koordinatensystem.

Sind die z_i - und die z_{i-1} -Achse windschief, so ist die x_i -Achse eindeutig festgelegt (und damit auch y_i). Andernfalls sind Sonderfälle zu beachten [Sie96]:

- Schneiden sich z_i - und z_{i-1} -Achse, gibt es zwei Möglichkeiten die x_i -Achse festzulegen. Beide Möglichkeiten sind gleichwertig.
- Sind die z_i - und die z_{i-1} -Achse parallel, so ist die x_i -Achse eindeutig, der Ursprung jedoch frei wählbar.
- Sind die z_i - und die z_{i-1} -Achse identisch, so ist die x_i -Achse und der Ursprung frei wählbar. Dies tritt z. B. bei Schubgelenken auf.

In Bild 2.6 sind zwei aufeinander folgende Gelenke aus der kinematischen Kette sowie die x - und z -Achsen der zugehörigen Koordinatensysteme dargestellt. Die y -Achsen wurden aus Gründen der Übersichtlichkeit nicht eingezeichnet. Anhand der Skizze lassen sich die folgenden Denavit-Hartenberg-Regeln [Den55, McC89, Sie96] für die Überführung des Koordinatensystems $i - 1$ in das Koordinatensystem i plausibilisieren:

1. Drehung des $(i - 1)$ -ten Koordinatensystems um θ_i um die z_{i-1} -Achse zur Ausrichtung der x -Achsen

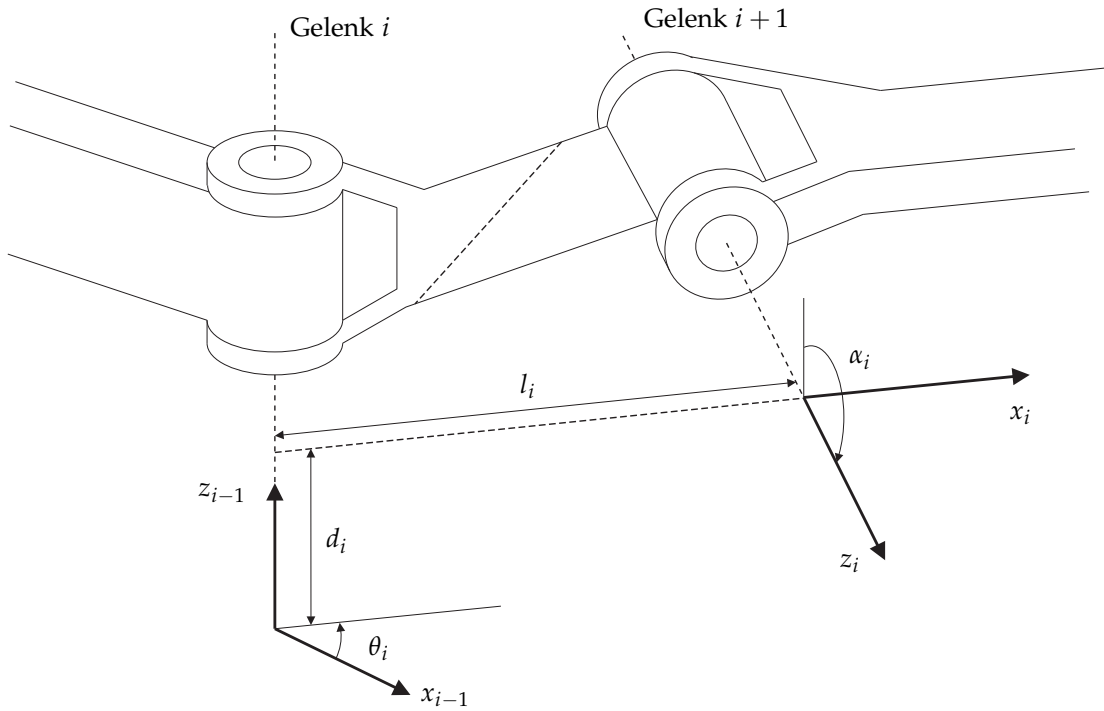


Bild 2.6 Zwei Gelenke eines Roboters mit den dazugehörigen Koordinatensystemen nach den Denavit-Hartenberg-Konventionen.

2. Positive Verschiebung um den Betrag d_i entlang der z_{i-1} -Achse
3. Positive Verschiebung um den Betrag l_i entlang der (bereits gedrehten) x_{i-1} -Achse
4. Drehung um den Winkel α_i um die x_i -Achse zur Ausrichtung der z -Achsen

Durch die Verwendung homogener Koordinaten können die Transformationen der Punkte 1-4 durch Matrixmultiplikationen ausgedrückt werden. Die Rotation um die z - bzw. x -Achse werden durch die Rotationsmatrizen

$$R_z = \begin{pmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ \sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{bzw.} \quad R_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha & 0 \\ 0 & \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.45)$$

definiert. Entsprechend werden die Verschiebungen entlang der z - bzw. x -Achse durch die Matrizen

$$T_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{bzw.} \quad T_x = \begin{pmatrix} 1 & 0 & 0 & l \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.46)$$

beschrieben. Mit Hilfe dieser Matrizen lassen sich die Transformationen der Punkte 1-4 zu einer Gesamttransformation D zusammensetzen. Da die Transformationen der Punkte 1-4 das Koordinatensystem betreffen, ist für die Abbildung der Punkte jeweils die inverse Transformation zu verwenden. Für die Koordinaten eines Punkts p im Koordinatensystem i und $i-1$ gilt somit:

$$p_{i-1} = D_i p_i. \quad (2.47)$$

mit

$$D := R_z T_z T_x R_x. \quad (2.48)$$

Def: 2.8

Die Matrix D im Sinne von Gleichung (2.48) heißt Denavit-Hartenberg-Matrix. Ausführlich geschrieben lautet sie:

$$D = \begin{pmatrix} \cos(\theta) & -\cos(\alpha) \sin(\theta) & \sin(\alpha) \sin(\theta) & \cos(\theta) l \\ \sin(\theta) & \cos(\alpha) \cos(\theta) & -\sin(\alpha) \cos(\theta) & \sin(\theta) l \\ 0 & \sin(\alpha) & \cos(\alpha) & d \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.49)$$

Die Parameter θ, α, l, d heißen Denavit-Hartenberg-Parameter. Die Längen l und d werden in Längeneinheiten, die Winkel θ und α in einem Winkelmaß (Bogenmaß, Grad) gemessen.

Aufgrund der Festlegung der Koordinatensysteme sind die Denavit-Hartenberg-Parameter α, l sowohl bei Rotations- als auch bei Translationsgelenken invariant. Sie lassen sich z. B. aus Konstruktionsplänen des Roboters entnehmen. Bei Translationsgelenken ändert sich nur d , bei Rotationsgelenken nur θ .

Um die Denavit-Hartenberg-Matrix für die komplette kinematische Kette zu bestimmen, müssen nur noch die einzelnen Denavit-Hartenberg-Matrizen der Gelenke zusammen multipliziert werden. Die Gesamttransformation D_{ges} ist

$$D_{ges} = \prod_{i=1}^{N_g} D_i. \quad (2.50)$$

Für den Scrobot ER VII aus Bild 2.4b werden die einzelnen Parameter und die Gesamttransformation in Abschnitt 5.1 ermittelt.

Über die Denavit-Hartenberg-Matrix lassen sich aus den Winkel- und Translationseinstellungen der Gelenke Position und Ausrichtung des Endeffektors ermitteln. Über die Hand-Auge-Kalibrierung lassen sich damit die extrinsischen Kameraparameter bestimmen.

2.4.2 Inverse Kinematik

Die inverse Kinematik ist das Gegenstück zur direkten Kinematik. Die inverse Kinematik ist dann nötig, wenn der Endeffektor eine bestimmte Position und Orientierung einnehmen soll. Für die Lösung des Problems gibt es, aus mathematischer Sicht, zunächst zwei Fragen [Cra04]: existiert eine Lösung, und ist sie eindeutig.

Existenz In [Cra04] wird der Arbeitsbereich des Roboters in zwei Teilbereiche gegliedert: den „dextrous workspace“, das ist der Bereich, in dem der Endeffektor die Positionen mit allen möglichen Orientierungen ansteuern kann, und den „reachable workspace“, den der Endeffektor mit mindestens einer Orientierung ansteuern kann. Hat der Roboter nicht genügend Freiheitsgrade, so kann der „dextrous workspace“ eine leere Menge sein.

Eindeutigkeit Bild 2.7 zeigt zwei unterschiedliche Stellungen der Arme des Roboters, die jedoch zu der gleichen Transformationsmatrix D führen, da der

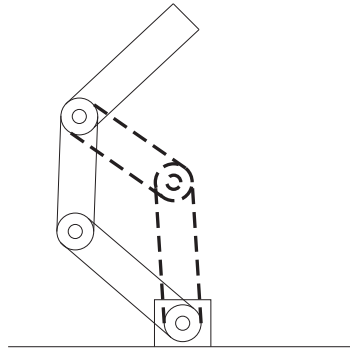


Bild 2.7 Für die Stellung des Endeffektors sind zwei unterschiedliche Konfigurationen der Gelenke und Arme möglich: die mit durchgezogenen Linien und die mit gestrichelten Linien.

Endeffektor in beiden Fällen die gleiche Position und die gleiche Ausrichtung hat. Es kann also im Allgemeinen nicht von einer eindeutigen Gelenkeinstellung für eine bestimmte Endeffektorstellung ausgegangen werden.

Existiert eine Lösung, so müssen für ihre Berechnung noch Randbedingungen beachtet werden [Sie96]:

- Die Gelenke haben aufgrund ihrer Konstruktion nur eingeschränkte Winkelbereiche. Diese dürfen nicht überschritten werden.
- Die Roboterarme, der Endeffektor oder die auf den Endeffektor montierte Kamera dürfen nicht mit anderen Roboterteilen oder mit Hindernissen im Arbeitsbereich kollidieren.

Für die Lösung der inversen Kinematik sind in der Literatur verschiedene Prinzipien beschrieben: direkte Berechnung der Winkel und Translationen aus der Denavit-Hartenberg-Matrix [Iso92, Sie96, Cra04], geometrische Berechnung über Kosinus- und Sinussatz [Cra04] sowie numerische, iterative Methoden [Tsa85]. Aus jüngerer Zeit existieren auch Arbeiten, die durch Methoden der Algebra Fragestellungen der inversen Kinematik lösen [Den04]. Bei allen Verfahren ist die Lösung der inversen Kinematik aber im Allgemeinen schwierig [Sie96], da rein mathematische Modelle zunächst keine Nebenbedingungen, wie z. B. Einschränkungen der Winkel aufgrund der mechanischen Konstruktion, berücksichtigen. Deshalb wird in dieser Arbeit die Verwendung der inversen Kinematik vermieden und an dieser Stelle nicht näher auf die Einzelheiten eingegangen.

Häufig ist die inverse Kinematik in der Steuereinheit des Roboters implementiert. Es ist dann möglich Translation und Rotation zu definieren und die inverse Kinematik wird intern berechnet: Existiert eine Lösung, so wird die entsprechende Position angefahren, andernfalls wird eine Fehlermeldung produziert.

Die Problematik der direkten und inversen Kinematik wurde vorgestellt. Aufgrund der Schwierigkeiten der inversen Kinematik im Vergleich zur direkten wird für die späteren Experimente (vgl. Kapitel 6) auf den Einsatz der inversen Kinematik verzichtet und stattdessen ausschließlich direkte Kinematik verwendet. Der Optimierungsprozess für die Auswahl der Ansichten wird nicht direkt Position und Orientierung der Kamera bestimmen, sondern die entsprechenden Gelenkeinstellungen. Dies hat zwei Vorteile: Erstens muss das Problem der inversen Kinematik nicht gelöst werden. Zweitens werden nur Positionen und Orientierungen für

die Ansichtenauswahl in Betracht gezogen, die auch erreichbar sind. Ein Nachteil ist, dass unter Umständen einige Positionen und Orientierungen mehrfach ausgewertet werden. Dies ist der Fall, wenn Mehrdeutigkeiten, wie in Bild 2.7, auftreten. Ferner muss natürlich bei einem Wechsel des Roboters die Denavit-Hartenberg-Matrix erneut aufgestellt werden.

2.5 Triangulation und Bündelausgleich

Bisher wurde gezeigt, wie man mit Hilfe der Kamerakalibrierung die intrinsischen Kameraparameter und mittels Kamerakalibrierung, Hand-Auge-Kalibrierung und direkter Kinematik die extrinsischen Parameter berechnen kann. Damit sind alle Kameraparameter bekannt. Um eine 3-D-Rekonstruktion, z. B. durch Triangulation (Abschnitt 2.5.2), durchführen zu können benötigt man aber noch korrespondierende Punkte, also eine Zuordnung, welcher Pixel im ersten Bild zu welchem im zweiten gehört. Andernfalls weiß man nicht, welche Geraden man schneiden muss (vgl. Abschnitt 1.1.2 und Bild 1.4). Diese Problematik wird im folgenden Abschnitt diskutiert. Danach wird auf die Triangulation eingegangen. Der letzte Abschnitt behandelt den Bündelausgleich als Verfahren zur abschließenden nichtlinearen Optimierung des Rekonstruktionsergebnisses.

2.5.1 Punktdetektion und Punktverfolgung

Für die Berechnung einer 3-D-Rekonstruktion werden in dieser Arbeit ausschließlich Punktmerkmale verwendet. Es existiert eine Vielzahl von Algorithmen zur Punktverfolgung. Häufig enthalten sie die Annahme, dass die Änderung im Bild klein ist, d. h. dass sich die Punktmerkmale zwischen zwei Bildern nur um einen kleinen Betrag verschieben. In diesem Fall können differentielle Methoden eingesetzt werden, wie in [Bur82, Woo83, Ana89, För91, Tom91].

In dieser Arbeit wird der Algorithmus aus [Zin04, Zin11] verwendet. Dieser Ansatz erweitert den von [Hei04], indem auch Beleuchtungsänderungen berücksichtigt werden. In [Hei04] wird der Tomasi-Kanade-Ansatz [Tom91] zusammen mit der Erweiterung von Shi [Shi94], die eine affine Modifikationen des Merkmalsfensters erlaubt, verwendet. Auf den Grundgedanken der Punktverfolgung aus [Tom91] wird kurz eingegangen, da in Abschnitt 3.2 bzw. Kapitel 4 von verwechselten Bildkoordinaten der verfolgten Punkte ausgegangen wird.

Bei der Punktverfolgung können keine einzelnen Pixel verfolgt werden, wenn sich ihr Grauwert nicht stark von den Grauwerten ihrer Nachbarn unterscheidet. Insbesondere innerhalb homogener Flächen lässt sich die Bewegung einzelner Pixel nicht detektieren, da alle Pixel den gleichen oder zumindest einen sehr ähnlichen Grauwert haben. Es werden also zunächst geeignete Punkte, die als Ecken bezeichnet werden, detektiert, um sie anschließend verfolgen zu können. Tomasi und Kanade entwickelten ein abgestimmtes Vorgehen für diese zwei Schritte [Tom91]:

Für jeden Bildpunkt wird zunächst die Strukturmatrix S für eine quadratische Umgebung mit Seitenlänge $2k + 1$ berechnet:

$$S(x,y) := \sum_{i,j=-k}^k \nabla f(x+i,y+j) \cdot \nabla f^T(x+i,y+j), \quad (2.51)$$

dabei ist ∇f der Gradient der Grauwerte. Der Gradient kann z. B. mit Hilfe des Sobel-Operators [Jäh05] berechnet werden.

Die Punktdetektion basiert auf der Eigenwertanalyse der Matrix S . Da S symmetrisch, reell und positiv semidefinit ist, kann S durch eine Rotation des Koordinatensystems diagonalisiert werden [Tru98] oder anders ausgedrückt: die Matrix S ist ähnlich zu einer reellen Diagonalmatrix. In der Diagonalmatrix befinden sich die beiden (positiven) Eigenwerte λ_1, λ_2 von S . O. B. d. A. kann S so diagonalisiert werden, dass $\lambda_1 \geq \lambda_2$ ist. Die Bedeutung des kleineren Eigenwerts λ_2 lässt sich in drei Spezialfällen verdeutlichen:

1. Besitzen die Bildpunkte in einer Umgebung von (x, y) alle den gleichen Grauwert $f(x, y)$, so gilt: $\lambda_1 = \lambda_2 = 0$.
2. Verläuft durch den Bildpunkt eine perfekte Kante, so gilt: $\lambda_1 > 0, \lambda_2 = 0$.
3. Ist der Bildpunkt (x, y) eine Ecke eines hellen Quadrats auf dunklem Hintergrund, so gilt: $\lambda_1 \geq \lambda_2 > 0$. Je größer die Eigenwerte, desto stärker ist der Kontrast in diesem Bildausschnitt [Tru98].

Entscheidend für die Punktdetektion ist die Größe des Eigenwerts λ_2 . Ecken befinden sich also dort, wo der Eigenwert λ_2 groß wird.

Die Grundidee der Verfolgung besteht darin, die „Ähnlichkeit“ von Merkmalen zu vergleichen, d. h. die Summe der quadrierten Abweichungen einer quadratischen Umgebung mit Seitenlänge $2k + 1$ im Bild zum Zeitpunkt τ und im nächsten Bild zum Zeitpunkt $\tau + 1$:

$$\epsilon(\mathbf{d}) = \sum_{i,j=-k}^k (f(x_n + i, y_n + j, \tau) - f(x_n + i + d_x, y_n + j + d_y, \tau + 1))^2. \quad (2.52)$$

Dabei ist $\mathbf{d} = (d_x \ d_y)$ die Verschiebung und $(x_n \ y_n)$ sind die Koordinaten des detektierten Punkts. Man geht davon aus, dass man einen Punkt (mit seiner Umgebung) im nächsten Bild wiedergefunden hat, wenn $\epsilon(\mathbf{d})$ minimal wird. Da ∇f eine Annäherung an den Gradienten von f ist, kann folgende Umformung mit Hilfe einer Taylor-Reihe gemacht werden:

$$f(x_n^* + d_x, y_n^* + d_y, \tau + 1) \approx f(x_n^*, y_n^*, \tau + 1) + \nabla f^T(x_n^*, y_n^*, \tau + 1) \cdot \mathbf{d}. \quad (2.53)$$

Mit den Abkürzungen: $x_n^* = x_n + i$ bzw. $y_n^* = y_n + j$. Die Approximation (2.53) in (2.52) eingesetzt und $\epsilon(\mathbf{d})$ nach $\mathbf{d}$ abgeleitet ergibt nach einigen Umformungen das lineare Gleichungssystem:

$$S\mathbf{d} = \sum_{i,j=-k}^k (f(x_n^*, y_n^*, \tau) - f(x_n^*, y_n^*, \tau + 1)) \nabla f(x_n^*, y_n^*, \tau + 1). \quad (2.54)$$

Die Lösung des Gleichungssystems (2.54) liefert die Verschiebung $\mathbf{d}$ im Bild der detektierten Punkte. Bildpunkte, die von demselben 3-D-Punkt stammen, und die man über zwei oder mehr Bilder verfolgt, nennt man korrespondierende Punkte.

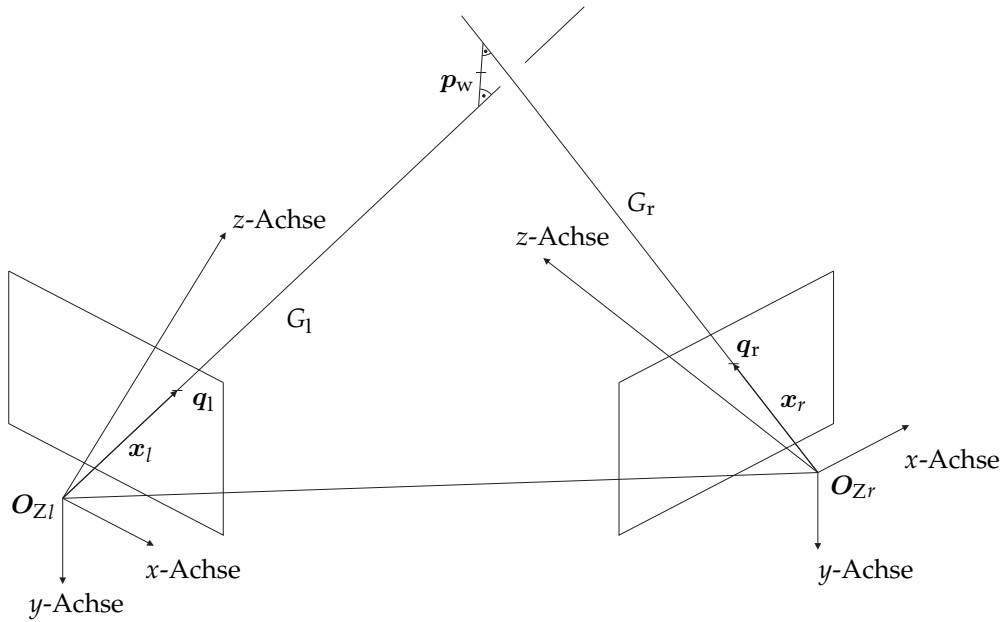


Bild 2.8 Die Sichtstrahlen G_l und G_r treffen sich aufgrund verrauschter Daten nicht. Der Punkt $\hat{p}_w$ wird deshalb so bestimmt, dass er von beiden Geraden minimalen Abstand hat.

2.5.2 Triangulation

Mit Hilfe von korrespondierenden Punkten und den zugehörigen ex- und intrinsischen Kameraparametern kann man die entsprechenden Geradengleichungen für die Sichtstrahlen aufstellen. Aufgrund von Rauschen in den gemessenen Daten werden die Geraden jedoch häufig keinen Schnittpunkt haben, sondern windschief sein. Für diesen Fall werden drei Möglichkeiten zur Berechnung vorgestellt.

2.5.2.1 Geometrischer Ansatz zur Triangulation

Dieser Ansatz basiert auf geometrischen Überlegungen [Tru98]. Die Idee ist, dass der zu rekonstruierende Punkt minimalen Abstand zu beiden Sichtstrahlen hat.

O. B. d. A. sei das Weltkoordinatensystem mit dem Kamerakoordinatensystem der linken Kamera identisch. Seien ferner x_l bzw. x_r die Vektoren, die vom optischen Zentrum O_{Zl} der linken bzw. O_{Zr} der rechten Kamera zu den Punkten q_l bzw. q_r auf der linken bzw. rechten Bildebene zeigen (vgl. Bild 2.8).

Der Sichtstrahl der linken Kamera hat die Geradengleichung

$$G_l(v_l) = v_l x_l \quad \text{mit } v_l \in \mathbb{R}. \quad (2.55)$$

Das Koordinatensystem der rechten Kamera lässt sich durch die Translation t und Rotation mit R in das Koordinatensystem der linken Kamera überführen. Für den Sichtstrahl G_r der rechten Kamera im Koordinatensystem der linken Kamera gilt:

$$G_r(v_r) = -R^T t + v_r R^T x_r \quad \text{mit } v_r \in \mathbb{R}. \quad (2.56)$$

Die kürzeste Verbindungsstrecke G_a steht auf beiden Geraden senkrecht. Somit ist ihr Richtungsvektor $x_a = x_l \times x_r$, der Stützpunkt sei vorübergehend mit a bezeichnet. Somit ist die Verbindungsgerade:

$$G_a(v_a) = a + v_a x_a \quad \text{mit } v_a \in \mathbb{R}. \quad (2.57)$$

Nun müssen Punkte existieren, die sowohl auf der Geraden G_l und G_a sowie auf G_r und G_a liegen. Durch Gleichsetzen der jeweiligen Geradengleichungen ergeben sich nach einigen trivialen Äquivalenzumformungen die Schnittpunkte von G_l mit G_a bzw. G_r mit G_a . Der gesuchte 3-D-Punkt ist aus dem arithmetischen Mittel der beiden Schnittpunkte bestimmbar.

2.5.2.2 Lineare Schätzung der 3-D-Koordinaten

Gleichung (2.1) beschreibt die Projektion eines 3-D-Punkts auf die Bildebene. Statt des Zeichens $\sim$ kann man auch einen Skalierungsfaktor ν verwenden:

$$\nu \underline{q} = \underline{P} \underline{p}_w. \quad (2.58)$$

O. B. d. A. kann der Skalierungsfaktor festgelegt werden: $\nu = {}_3 \underline{p} \underline{p}_w$, dabei bezeichnet ${}_3 \underline{p}$ die 3. Zeile der Projektionsmatrix $\underline{P}$. Daraus folgt für die 3. Komponente des Bildpunkts in homogenen Koordinaten: $q_3 = 1$. Aus der Projektionsgleichung (2.58) wird damit:

$${}_3 \underline{p} \underline{p}_w \underline{q} = \underline{P} \underline{p}_w. \quad (2.59)$$

Dies kann in ein Gleichungssystem überführt werden,

$$(\underline{q} {}_3 \underline{p} - \underline{P}) \underline{p}_w = \underline{0}_3, \quad (2.60)$$

das zwei Gleichungen liefert (die letzte Zeile der linken Seite ist Null, da $q_3 = 1$).

Seien m Bilder gegeben, dann lassen sich $2m$ lineare Gleichungen aufstellen, die sich in Matrix-Vektor-Form darstellen lassen. Der Nullraum der Matrix legt dann die Koordinaten des 3-D-Punkts fest. Der Nullraum wird in dieser Arbeit durch eine Singulärwertzerlegung [Gol96, Tre97, Sch06b] berechnet. Dazu wird die Matrix $(\underline{q} {}_3 \underline{p} - \underline{P})$ zerlegt:

$$(\underline{q} {}_3 \underline{p} - \underline{P}) = \underline{U}_l \underline{D}_s \underline{V}_r^T \quad (2.61)$$

Der letzte Spaltenvektor der Matrix $\underline{V}_r$ spannt dabei den Nullraum auf und ist somit die gesuchte Lösung für $\underline{p}_w$.

Diese Methode reduziert ein formales Fehlermaß, welches keine direkte geometrische Bedeutung besitzt [Hei04].

2.5.2.3 Nichtlineare Optimierung des Rückprojektionsfehlers

Der Rückprojektionsfehler ist der Abstand im Bild zwischen dem verfolgten Bildpunkt und der Projektion des rekonstruierten 3-D-Punkts [Hei04]. Für den j -ten rekonstruierten 3-D-Punkt $\hat{\underline{p}}_{wj}$, der im i -ten Bild aufgenommen wurde gilt in homogenen Koordinaten:

$$\hat{\underline{q}}_{ij} = \underline{P}_i \hat{\underline{p}}_{wj}. \quad (2.62)$$

Nach der Umrechnung in kartesische Koordinaten gilt für den Bildpunkt

$$\underline{q}_{ij} = \begin{pmatrix} \frac{{}_1 \underline{p}_i \hat{\underline{p}}_{wj}}{{}_3 \underline{p}_i \hat{\underline{p}}_{wj}} \\ \frac{{}_2 \underline{p}_i \hat{\underline{p}}_{wj}}{{}_3 \underline{p}_i \hat{\underline{p}}_{wj}} \end{pmatrix}, \quad (2.63)$$

dabei bezeichnet ${}_k \underline{p}$ die k -te Zeile der Projektionsmatrix $\underline{P}$.

Aus der Punktverfolgung ist die Position $\mathbf{q}_{ij}$ des 3-D-Punkts $\mathbf{p}_{wj}$ im Bild i bekannt. Der Rückprojektionsfehler des j -ten Punkts im i -ten Bild ist somit

$$\epsilon_{rp\,ij} := \mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij}. \quad (2.64)$$

Der totale Rückprojektionsfehler aller n 3-D-Punkte in allen m Bildern ist dann

$$\epsilon_{rp} = \left((\mathbf{q}_{00} - \hat{\mathbf{q}}_{00})^T, \dots, (\mathbf{q}_{mn} - \hat{\mathbf{q}}_{mn})^T \right)^T. \quad (2.65)$$

Der quadrierte Rückprojektionsfehler ist wie folgt definiert:

$$\bar{\epsilon}_{rp} := \epsilon_{rp}^T \epsilon_{rp}. \quad (2.66)$$

Die nichtlineare Optimierung minimiert die Summe der quadratischen Abstände zwischen den berechneten Bildkoordinaten $\hat{\mathbf{q}}_{ij}$ und den detektierten $\mathbf{q}_{ij}$ in Abhängigkeit der Koordinaten des Weltpunkts $\mathbf{p}_{wj}$. Die Zielfunktion lautet somit:

$$\mathbf{p}_{wj} = \operatorname{argmin}_{\hat{\mathbf{p}}_{wj} \in \mathbb{P}^3} \sum_{i=1}^m \left(\frac{1\mathbf{p}_i \hat{\mathbf{p}}_{wj}}{3\mathbf{p}_i \hat{\mathbf{p}}_{wj}} - q_{xj} \right)^2 + \left(\frac{2\mathbf{p}_i \hat{\mathbf{p}}_{wj}}{3\mathbf{p}_i \hat{\mathbf{p}}_{wj}} - q_{yj} \right)^2. \quad (2.67)$$

Die Optimierung der Zielfunktion kann mit einem üblichen numerischen Verfahren realisiert werden. In der Praxis wird häufig das Verfahren nach Levenberg-Marquardt [Pre02, Sch06b] verwendet [Har03, Hei04, Vog06, Sch06a, Sch08], welches auch in dieser Arbeit benutzt wird. Für die nichtlineare Optimierung wird eine gute Startlösung benötigt. Diese lässt sich relativ einfach aus der linearen Schätzung gewinnen.

Ein Vorteil der linearen Schätzung und der nichtlinearen Optimierung ist, dass aus beliebig vielen Bildern trianguliert werden kann. Der geometrische Ansatz sieht hingegen genau zwei Bildern vor. Eine Erweiterung auf 3 Bilder wäre möglich, jedoch sehr aufwändig.

2.5.3 Bündelausgleich

Der Bündelausgleich (engl. Bundle Adjustment) bezeichnet eine nichtlineare Optimierung, die sowohl die 3-D-Koordinaten der Punkte als auch die Kameraparameter optimiert. Ziel ist es, den Rückprojektionsfehler zu minimieren. Das Verfahren wurde in der Photogrammetrie entwickelt [Sla80] und findet im Rechnersehen einen weit verbreiteten Einsatz [Har94, Hey97, Sze98, Fit98, McL99, Shu99, Fua99, Ris99]; in [Har03] wird vorgeschlagen, jede Rekonstruktion zum Schluss durch den Bündelausgleich noch zu optimieren.

Der Bündelausgleich optimiert den Rückprojektionsfehler aus Gleichung (2.66) bzw. (2.67). Der einzige Unterschied zu Gleichung (2.67) ist, dass nun auch die Kameraparameter der Projektionsmatrix $\hat{\mathbf{P}}$ optimiert werden und dass über alle 3-D-Punkte $\hat{\mathbf{p}}_{wj}$ gleichzeitig optimiert werden muss. Die Optimierungsgleichung für die n Weltpunkte $\mathbf{p}_{wj}$ aus den m Bildern mit den Projektionsmatrizen $\mathbf{P}_i$ lautet somit:

$$\min_{\hat{\mathbf{P}}_i, \hat{\mathbf{p}}_{wj}} \sum_{i=1}^m \sum_{j=1}^n \left(\frac{1\hat{\mathbf{p}}_i \hat{\mathbf{p}}_{wj}}{3\hat{\mathbf{p}}_i \hat{\mathbf{p}}_{wj}} - q_{xj} \right)^2 + \left(\frac{2\hat{\mathbf{p}}_i \hat{\mathbf{p}}_{wj}}{3\hat{\mathbf{p}}_i \hat{\mathbf{p}}_{wj}} - q_{yj} \right)^2. \quad (2.68)$$

Jede Projektionsmatrix $\hat{P}_i$ hat 11 Freiheitsgrade (vgl. Abschnitt 2.2.1). Bei m Bildern und n Punkten sind somit $11m + 3n$ Parameter zu optimieren. Für die Minimierung von (2.66) muss eine $(11m + 3n) \times (11m + 3n)$ -Jacobi-Matrix aufgestellt und invertiert werden [Har03]. Bei einer großen Anzahl von Bildern und vielen detektierten und verfolgten Punkten wird dieser Prozess extrem zeitaufwändig.

Ein erster Schritt diesen Aufwand zu reduzieren ist, die Anzahl der Kameraparameter zu reduzieren: setzt man voraus, dass die intrinsischen Parameter für alle Aufnahmen konstant und exakt bestimmt worden sind, so ist die Zahl der zu optimierenden Parameter $6m + 3n$.

In [Tri00] werden neben der Parameterreduktion verschiedene Varianten für eine effiziente Berechnung analysiert. Zusätzlich werden auch unterschiedliche Parametrierungen von 3-D-Punkten oder Rotationen getestet. Einen Vergleich unterschiedlicher Repräsentationen der Rotation findet man auch in [Sch01, Sch06a].

Der eingebettete Bündelausgleich ist eine weitere Variante um den Berechnungsaufwand zu minimieren [Shu99, Har03]. Man nutzt dabei aus, dass die Schätzung der 3-D-Struktur bei festen Kameraparametern für jeden Weltpunkt unabhängig durchgeführt werden kann. Während eines Optimierungsschritts der Kameraparameter P_i werden also jedes mal n dreidimensionale Optimierungen für jeden Weltpunkt durchgeführt. Die zugehörige Optimierungsgleichung lautet hier:

$$\min_{\hat{P}_i} \sum_{i=1}^m \min_{\hat{P}_{wj}} \sum_{j=1}^n \left(\frac{1\hat{P}_i\hat{P}_{wj}}{3\hat{P}_i\hat{P}_{wj}} - q_{xj} \right)^2 + \left(\frac{2\hat{P}_i\hat{P}_{wj}}{3\hat{P}_i\hat{P}_{wj}} - q_{yj} \right)^2. \quad (2.69)$$

Nach [Har03] ist das Ergebnis des eingebetteten Bündelausgleichs identisch mit dem des simultanen, falls nur ein Minimum existiert. Durch den eingebetteten Ansatz wird jedoch u. U. die Konvergenzgeschwindigkeit reduziert.

Der Berechnungsaufwand lässt sich gem. [Sch08] durch ein schrittweises Vorgehen noch weiter reduzieren: Abwechselnd werden die Kameraparameter und die 3-D-Punkte optimiert. Da jeder Punkt unabhängig von den anderen, mit festen Kameraparametern, optimiert wird, sind für diesen Punkt nur 3 Parameter zu bestimmen (das entspricht exakt dem Vorgehen bei der nichtlinearen Optimierung des Rückprojektionsfehlers bei der Triangulation, vgl. Abschnitt 2.5.2.3). Analog geht man bei der Optimierung der Kameraparameter vor: Die Parameter jeder Kamera werden bei festen Punkten optimiert. Da die Kameras so separat optimiert werden, ist die maximale Anzahl der gleichzeitig zu bestimmenden Parameter in einem Schritt 11. In praktischen Fällen zeigt sich aber, dass dieses Verfahren u. U. in ein anderes lokales Minimum konvergiert als beim eingebetteten Bündelausgleich bzw. bei der ursprünglichen Formulierung des Problems.

Das Optimierungsproblem (2.69) kann mit üblichen Optimierungsverfahren aus der Numerik gelöst werden, wie zum Beispiel Gauß-Newton-Verfahren oder das Verfahren nach Levenberg-Marquardt [Pre02, Sch06b]. Letzteres wird auch in dieser Arbeit verwendet.

Kapitel 3

Grundlagen für die Auswahl von Ansichten

Im vorherigen Kapitel wurde gezeigt, wie man eine 3-D-Rekonstruktion bei bekannten Kameraparametern und mit Hilfe von korrespondierenden Punkten berechnen kann. Bei dieser Art der Rekonstruktion ist es möglich, zunächst alle Aufnahmen zu akquirieren und im Anschluss daran die 3-D-Geometrie zu rekonstruieren. An diesem passiven Verfahren, im Sinne des Marr-Paradigmas [Mar82, Den97], kann man nicht festhalten, wenn man Ansichten gezielt auswählen möchte, um ein besseres Rekonstruktionsergebnis zu erzielen. Stattdessen müssen sich die Schritte „Bildaufnahme“ und „Planung der nächsten Ansicht“ abwechseln, denn nur so lassen sich aus den bisherigen Aufnahmen Informationen gewinnen, auf deren Basis die nächste beste Ansicht ausgewählt werden kann. Wie bereits in Kapitel 1 angedeutet, lassen sich die Algorithmen für die Ansichtenauswahl in geometrische und probabilistische Ansätze unterscheiden.

Geometrische Ansätze verwenden Gütemaße mit geometrischer Bedeutung. Das kann z. B. die Abweichung in der 3-D-Position, die Abweichung der Normalen der Oberflächen oder Unterschiede in den Krümmungsradien sein. Fast alle in Abschnitt 1.2, Seite 9, erwähnten Methoden optimieren entweder nur hinsichtlich der Vollständigkeit oder zusätzlich über ein geometrisches Gütemaß die Genauigkeit. Aus diesem Grund werden im Abschnitt 3.1 zwei Verfahren aus der Literatur näher vorgestellt. Ziel ist zum einen die dortigen Ideen plausibel zu machen und zum anderen eine Abgrenzung zu den neuen Methoden dieser Arbeit zu ermöglichen.

Im Abschnitt 3.2 werden eigene Ergebnisse für Stereosysteme präsentiert. Dabei wird der Einfluss der einzelnen veränderbaren Kameraparameter oder die Kombination mehrerer veränderbarer Parameter auf das Gütemaß untersucht. Die Analysen werden zunächst an einem 2-D-Modell und später an einem 3-D-Modell durchgeführt. Dabei lassen sich Analogien zwischen beiden Modellen herstellen. Darüber hinaus dienen die Betrachtungen auch zur Plausibilisierung von weiteren Ergebnissen in der Ansichtenplanung.

Im Gegensatz zu den geometrischen Ansätzen wird im Abschnitt 3.3 ein probabilistischer Ansatz aus der Literatur vorgestellt. In diesem ist die Schätzung der 3-D-Geometrie durch Wahrscheinlichkeitsverteilungen gegeben. Als Gütemaß wird die Kovarianzmatrix dieser Verteilung, die den erwarteten mittleren quadratischen Schätzfehler repräsentiert, verwendet. Durch die Ansichtenplanung soll dieses optimiert werden. Dieser Abschnitt gibt einen Überblick über prinzipielle Arbeitsweisen von probabilistischen Methoden zur Ansichtenplanung für die 3-D-Rekonstruktion. Ein eigener probabilistischer Ansatz wird in Kapitel 4 vorgestellt. Zum Abschluss dieses Abschnitts erfolgt noch eine Diskussion über Gemeinsam-

keiten und Unterschiede des Algorithmus aus der Literatur und des in Kapitel 4 vorgestellten eigenen Ansatzes.

3.1 Allgemeine geometrische Ansätze

Geometrische Ansätze verwenden ein Gütemaß mit geometrischer Bedeutung als Optimierungskriterium. Wie bereits oben erwähnt gibt es dabei unterschiedliche Möglichkeiten, wie z. B. die Abweichung in der 3-D-Position, die Abweichung der Normalen der Oberflächen oder der Unterschied in den Krümmungsradien. Zwei Algorithmen aus der Literatur werden im Folgenden ausführlich dargestellt.

3.1.1 Structure-from-controlled-motion

Der erste Algorithmus, der hier vorgestellt wird, basiert auf den Arbeiten von Marchand und Chaumette [Mar94, Mar96, Cha94, Cha96]. Für die 3-D-Rekonstruktion wird ein „Structure-from-controlled-motion“ Ansatz verwendet. Dabei wird davon ausgegangen, dass die Szene nur aus geometrischen Grundkörpern, wie Kugeln, Zylindern, geraden Linien oder einzelnen Punkten, besteht. Der Ansatz erlaubt es, einen direkten Zusammenhang zwischen Kamerabewegung und Rekonstruktionsgüte herzuleiten und für die oben genannten Grundkörper einfache Regeln für die Rekonstruktion aufzustellen.

Ausgangspunkt ist folgende Überlegung: Bewegt sich die Kamera mit der Geschwindigkeit $\mathbf{U} = (\mathbf{v}, \boldsymbol{\omega})$ (Translationsgeschwindigkeit $\mathbf{v}$; Winkelgeschwindigkeit $\boldsymbol{\omega}$), so gilt für die Geschwindigkeit $\dot{\mathbf{p}}_c$ des 3-D-Punkts $\mathbf{p}_c$ [Hee92, Oli01]:

$$\dot{\mathbf{p}}_c = -\mathbf{v} - \boldsymbol{\omega} \times \mathbf{p}_c. \quad (3.1)$$

Aus der Gleichung (3.1) und der Gleichung für die perspektivische Projektion (2.3) lässt sich durch Ableiten und Maximum-Likelihood-Schätzung (unter der Annahme: Kovarianz $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$) eine Beziehung für die inverse Tiefe des Punkts $\mathbf{p}_c$ herleiten [May79, Hor87, Mat89]:

$$\frac{1}{p_{cz}} = \frac{(\dot{q}_x - \alpha_x)(q_x v_z - v_x) + (\dot{q}_y - \alpha_y)(q_y v_z - v_y)}{(q_x v_z - v_x)^2 + (q_y v_z - v_y)^2}, \quad (3.2)$$

mit

$$\alpha_x := q_x q_y \omega_x - (1 + q_x^2) \omega_y + q_y \omega_z, \quad \alpha_y := (1 + q_y^2) \omega_x - q_x q_y \omega_y + q_x \omega_z.$$

Um zweidimensionale geometrische Grundfiguren (z. B. Punkte, Linien, Kreise) oder dreidimensionale Grundkörper (z. B. Kugeln, Zylinder, Ringe) zu detektieren, wird ein einheitlicher Formalismus eingeführt. Ein Objekt O_w der Szene lässt sich durch seine Parameter $\mathbf{O}_w$ und eine spezielle Funktion h implizit darstellen:

$$h\left(\mathbf{q}, \frac{1}{p_{cz}}, \mathbf{O}_w\right) = 0, \quad \forall \mathbf{p}_c \in O_w. \quad (3.3)$$

Der Satz über implizite Funktionen (vgl. [Heu04, Seite 292], [Mey01, Seite 398] oder [Hof05, Seite 514]) garantiert, unter bestimmten Voraussetzungen [Cha96], die Existenz einer eindeutigen Funktion h_{exp} , so dass Gleichung 3.3 explizit dargestellt werden kann:

$$1/p_{cz} = h_{\text{exp}}(\mathbf{q}, \mathbf{O}_{wL}). \quad (3.4)$$

Dabei ist $\mathbf{O}_{wL}$ eine Repräsentation der Objektparameter $\mathbf{O}_w$ für die Darstellung in der expliziten Form (3.4).

In der Bildebene lassen sich alle Punkte, die zum Bild des Objekts $\mathbf{O}_b$ gehören, analog zu Gleichung (3.3), durch eine Funktion h_{bild} der Form

$$h_{\text{bild}}(\mathbf{q}, \mathbf{O}_b) = 0, \quad \forall \mathbf{q} \in \mathbf{O}_b \quad (3.5)$$

beschreiben. Dabei sind $\mathbf{O}_b$ die Parameter der Projektion des Objekts. Die Parameter $\mathbf{O}_b$ sind natürlich von der Kamerabewegung abhängig und ändern sich bei einer Bewegung entsprechend. Die zeitliche Änderung der Objektparameter im Bild $\mathbf{O}_b$ lässt sich dann durch eine Funktion ψ beschreiben [Esp92]:

$$\dot{\mathbf{O}}_b =: \psi(\mathbf{O}_b, \mathbf{O}_{wL}, U). \quad (3.6)$$

Im Kamerabild können die Parameter der Projektion $\mathbf{O}_b$ und deren zeitliche Änderung $\dot{\mathbf{O}}_b$ gemessen werden. Zusätzlich ist natürlich die Kamerabewegung U bekannt und die Funktion ψ ist für den jeweiligen Grundkörper festgelegt [Esp92]. Damit können die Parameter $\mathbf{O}_{wL}$ berechnet werden.

Da die Kamerabewegung kontrolliert werden kann, stellt sich nun die Frage, welche Kamerabewegung die Rekonstruktionsgüte erhöht. Hierzu werden zwei mögliche Fehlerquellen adressiert: zum einen der Diskretisierungsfehler, zum anderen die Unsicherheit in der Schätzung von $\mathbf{O}_{wL}$, die durch Rauschen in der Messung von $\mathbf{O}_b$ entstehen.

Der Diskretisierungsfehler entsteht dadurch, dass $\dot{\mathbf{O}}_b = \Delta \mathbf{O}_b / \Delta \tau$ gesetzt wird. Diese Bedingung ist nur dann erfüllt, falls $\ddot{\mathbf{O}}_b = \ddot{\mathbf{O}}_b = \dots = 0, \forall t$, d. h. $\mathbf{O}_b$ ändert sich nur linear. Nach der Kettenregel abgeleitet gilt:

$$\ddot{\mathbf{O}}_b = \frac{\partial \psi}{\partial \mathbf{O}_b} \dot{\mathbf{O}}_b + \frac{\partial \psi}{\partial \mathbf{O}_{wL}} \dot{\mathbf{O}}_{wL} + \frac{\partial \psi}{\partial U} \dot{U}. \quad (3.7)$$

Eine hinreichende (aber nicht notwendige) Bedingung für $\ddot{\mathbf{O}}_b = 0$ ist

$$\dot{\mathbf{O}}_b = \dot{\mathbf{O}}_{wL} = 0. \quad (3.8)$$

In diesem Fall gilt stets [Cha96]: $\psi(\mathbf{O}_b, \mathbf{O}_{wL}, U) = 0$. Dies kann folgendermaßen plausibilisiert werden: Ändern sich die Parameter $\mathbf{O}_b$ nur linear, so ändern sich die Bewegungsparameter U nicht. Weiterhin kann gezeigt werden, dass die Bedingung $\dot{\mathbf{O}}_b = 0$ die Bedingung $\dot{\mathbf{O}}_{wL} = 0$ impliziert (außer bei Punkten und Linien). Das bedeutet, dass die Projektion des 3-D-Objekts während der Bewegung stets an der gleichen Stelle im Bild erscheinen sollte.

Der zweite Aspekt, die Auswirkungen fehlerhafter Messungen zu reduzieren, besteht darin, die Varianz der Parameter $\mathbf{O}_w$ zu minimieren, die sich aus den Varianzen σ^2 der gemessenen Größen $\mathbf{O}_b, \dot{\mathbf{O}}_b, U$ ergibt. Durch Ableiten und Nullsetzen der Funktion für die Varianz von $\mathbf{O}_w$ erhält man ein nichtlineares Gleichungssystem, dessen Lösung die Kameraparameter für die Bewegung enthält, die die Varianz der Objektparameter minimiert.

Es ist sehr aufwändig eine analytische Lösung dieses nichtlinearen Gleichungssystems zu finden [Cha96]. Für verschiedene einfache Körper lassen sich jedoch Lösungen angeben, bei denen die Effekte des Messfehlers minimiert werden:

Punkte Die Kamera sollte sich auf einer Kugel bewegen, deren Mittelpunkt der zu rekonstruierende Punkt ist und es soll gelten $v_z = \omega_z = 0$.

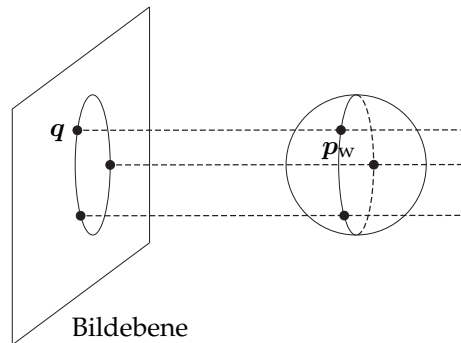


Bild 3.1 Aus dem Bild der Silhouette lässt sich die Oberfläche der Kugel rekonstruieren.

Kugeln Die Kamera sollte sich auf einer Kugel bewegen, deren Mittelpunkt mit dem Mittelpunkt der zu rekonstruierende Kugel identisch ist und es soll gelten $v_z = \omega_z = 0$.

Gerade Linien Gerade Linien sollten zentriert und vertikal bzw. horizontal im Bild liegen und es soll gelten $v_y = v_z = \omega_x = 0$ bzw. $v_x = v_z = \omega_y = 0$.

Zylinder Die beiden Ränder des Zylinders sollten vertikal bzw. horizontal im Bild liegen und gleichen Abstand zur Bildmitte haben. Die Kamera sollte auf einem Kreis um den Zylinder herum bewegt werden, so dass die Normale der Kreisebene parallel zur Längsachse des Zylinders verläuft.

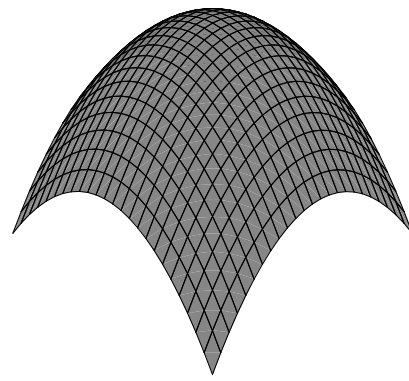
Durch diese Regeln wird der Diskretisierungsfehler und der Einfluss verrauschter Messungen minimiert. Die Parameter dieser vier geometrischen Formen können somit mit hoher Genauigkeit bestimmt werden. Nachteil dieser Methode ist, dass die Szene aus genau diesen Formen aufgebaut sein muss, was natürlich bei komplexen Szenen oder Objekten nicht der Fall ist.

3.1.2 Shape-from-occluding-contour

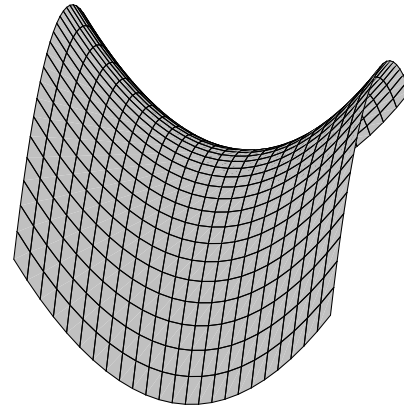
Ein zweiter Ansatz, der sich vor allem für gekrümmte Flächen eignet, verwendet ein Verfahren namens „Shape-from-occluding-contour“ [Kut94b]. Dabei werden die Umrisse des Objekts im Bild analysiert (vgl. Bild 3.1).

Es wird davon ausgegangen, dass die Oberfläche eine glatte, orientierte Fläche im $\mathbb{R}^3$ ist. Dann lässt sich die lokale Krümmung der Oberfläche durch die zweite Fundamentalform [Bär01] ausdrücken. Während die erste Fundamentalform die innere Geometrie beschreibt, mit deren Hilfe man z. B. die Länge einer Kurve auf der Oberfläche berechnen kann, beschreibt die zweite Fundamentalform die Krümmung der Fläche. Die zweite Fundamentalform besitzt zwei Extrema, ein Maximum k_{n1} und ein Minimum k_{n2} , in zwei zueinander orthogonalen Richtungen. Es werden vier mögliche Oberflächentypen unterschieden: elliptisch, hyperbolisch, parabolisch und planar. Diese Typen lassen sich anhand der Vorzeichen von k_{n1}, k_{n2} unterscheiden (vgl. Bild 3.2):

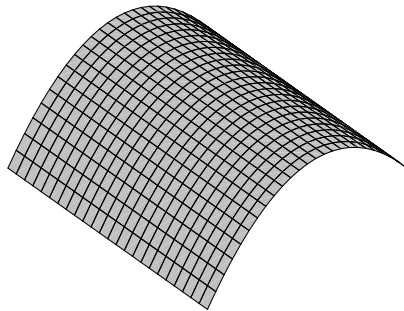
Während der Rekonstruktion wird ein Punkt auf der Konturlinie im Bild beobachtet. Durch kleine Änderungen der Blickrichtung der Kamera wird die Krümmung an dem besagten Punkt größer, kleiner oder bleibt konstant. Im letzteren Fall ist die Krümmung im Bild gleich den Extrema k_{n1}, k_{n2} und der untersuchte Punkt



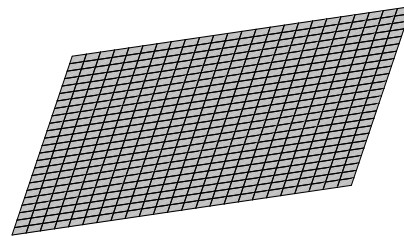
(a) Elliptische Oberfläche
($k_{n1} \cdot k_{n2} > 0$)



(b) Hyperbolische Oberfläche
($k_{n1} \cdot k_{n2} < 0$)



(c) Parabolische Oberfläche
($k_{n1} \neq 0, \quad k_{n2} = 0$)



(d) Planare Oberfläche
($k_{n1} = 0, \quad k_{n2} = 0$)

Bild 3.2 Glatte Flächen lassen sich in diese vier Kategorien einteilen.

ist ein Nabelpunkt einer lokalen elliptischen Oberfläche. Falls die Krümmung nicht konstant bleibt wird in die Richtung weiterbewegt, in der die Krümmung größer wird, solange bis das Maximum erreicht ist. Das Maximum im Bild ist dann gleich dem Maximum der zweiten Fundamentalform k_{n1} . Aus der Änderung des Winkels zwischen initialer und finaler Position sowie aus den zugehörigen Krümmungen kann auch das Minimum k_{n2} berechnet werden [Kut94b].

Probleme bei diesem Ansatz sind die Verfolgung des Punkts auf der Konturlinie, die Auswahl geeigneter Punkte für die Rekonstruktion und die Extrapolation der rekonstruierten Oberfläche auf benachbarte Punkte. Darüber hinaus kommen bei planaren Oberflächen Unstetigkeiten bei der Verfolgung des Objektpunkts im Bild vor, die gesondert zu behandeln sind. Schwierig sind auch hyperbolische Flächen, bei denen die möglichen Blickwinkel, aus denen rekonstruiert werden kann, eingeschränkt sind. Es muss zunächst immer der Typ der Oberfläche bestimmt werden und dann über Fallunterscheidungen die geeignete Strategie gewählt werden. Nach diesem Schritt kann jedoch die oben kurz dargestellte einfache Strategie der Änderung des Blickwinkels verwendet werden, um das Maximum in der Krümmung im Bild zu suchen und mit hoher Genauigkeit zu messen.

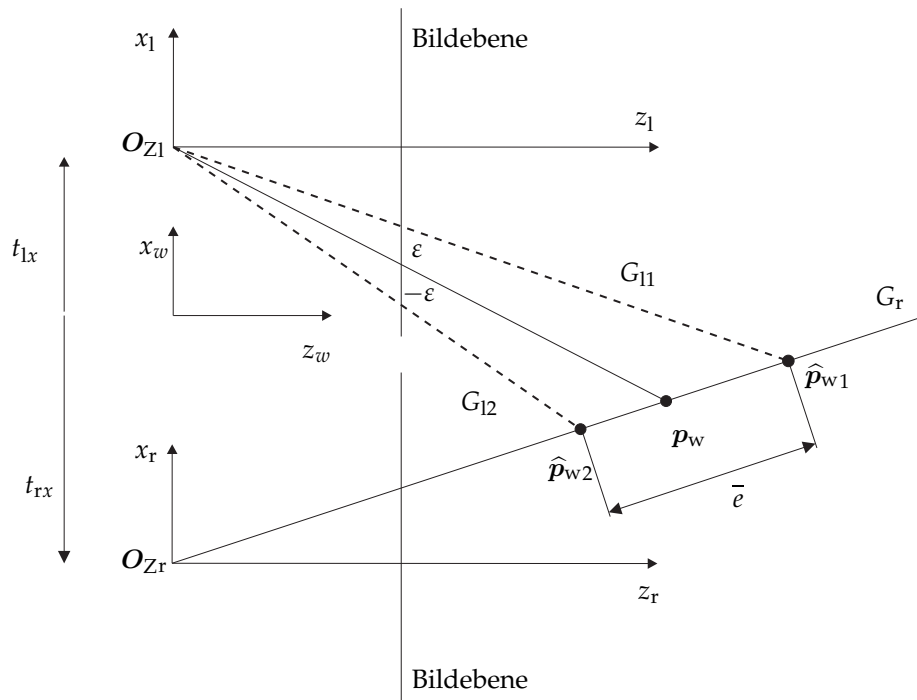


Bild 3.3 Normiertes Stereosystem. Durch die Abweichung $\pm\epsilon$ auf der Bildebene kann der tatsächliche 3-D-Punkt p_w nicht exakt rekonstruiert werden, stattdessen erhält man $\hat{p}_{w1}$ bzw. $\hat{p}_{w2}$, deren quadrierter Abstand als Rekonstruktionsgüte e definiert wird.

3.2 Ein geometrischer Ansatz für Stereosysteme

Neben den in Abschnitt 3.1 vorgestellten Algorithmen soll nun ein eigenes Verfahren vorgestellt werden. Hierbei handelt es sich um eine systematische Untersuchung des Einflusses der einzelnen, veränderbaren Kameraparameter auf die Rekonstruktionsgüte in einem Stereokamerasystem [Wen04, Wen07a].

Die veränderbaren Kameraparameter sind Brennweite, Rotation und Translation. Alle weiteren Parameter (Hauptpunkt, Skew, Verzerrungsparameter) können nicht (zielgerichtet) verändert und somit auch nicht optimiert werden.

Analysiert werden zunächst die Parameter mit einem zweidimensionalen Stereosystem (d. h. alle y -Komponenten sind auf Null gesetzt) und anschließend mit einem allgemeineren 3-D-Modell. Für das einfache 2-D-Modell wird zu Beginn das Gütemaß definiert. Anschließend erfolgt eine Analyse des Einflusses der einzelnen Parameter und eine Diskussion der jeweiligen Ergebnisse. In diesem Fall können z. T. analytische Untersuchungen durchgeführt werden, die allgemeingültige Schlüsse (für dieses Modell) zulassen.

Im anschließenden 3-D-Modell werden die Modellbedingungen aus dem zweidimensionalen Fall erweitert, um realistischere Rahmenbedingungen zu untersuchen. Eine analytische Auswertung der entsprechenden Fehlerfunktion ist dann allerdings nicht mehr möglich. Deshalb werden typische Konstellationen mit Hilfe numerischer Simulationen analysiert. Darüber hinaus werden die Ergebnisse mit denen aus dem zweidimensionalen Modell verglichen und Analogien aufgezeigt.

3.2.1 Zweidimensionales Modell

Das Stereosystem besteht aus zwei Kameras, die einen 3-D-Punkt beobachten (vgl. Bild 3.3). In Blickrichtung der positiven z -Achse des Weltkoordinatensystems wird die linke Kamera und ihre Parameter mit dem Subskript l bezeichnet; analog wird für die rechte Kamera und ihre Parameter das Subskript r verwendet.

Ausgangspunkt ist ein normalisiertes Stereosystem, d. h. die Orientierung der Kameras ist gleich, sie besitzen die gleiche Brennweite und die Translation zwischen den Kameras verläuft parallel zur x -Richtung.

Die Untersuchungen dienen der Analyse des Einflusses der veränderbaren Kameraparameter. Dabei werden die Parameter zunächst einzeln optimiert. Auf die Optimierung aller Parameter gleichzeitig wird in Abschnitt 3.2.1.6 eingegangen. In diesem 2-D-Modell wird nur ein einzelner 3-D-Punkt rekonstruiert. Da die Fehlerfunktion bei mehreren Punkten aus der Summe der einzelnen Fehler besteht, bringt eine Betrachtung mehrerer Punkte zunächst keinen Informationsgewinn. Zusätzlich wird hier angenommen, die Bildebene sei unbegrenzt, d. h. nicht wie bei einer realen Kamera durch die Ausmaße des optischen Sensors eingeschränkt. Diese Einschränkungen erlauben teilweise eine analytische Betrachtung der Fehlerfunktion. Im 3-D-Modell (Abschnitt 3.2.2) werden diese Beschränkungen aufgehoben, das heißt, dort hat man eine begrenzte Bildebene und es werden mehrere 3-D-Punkte rekonstruiert. Gerade deshalb wird später auf Analogien zwischen dem Verhalten im 2-D- und 3-D-Modell explizit hingewiesen.

3.2.1.1 Fehlermodell und Gütefunktion im 2-D-Modell

Im zweidimensionalen Modell wird angenommen, dass in einem Bild ein Punkt mit Hilfe der Punktdetektion ausgewählt wird, der dann mit Hilfe eines Punktverfolgers im zweiten Bild wiedergefunden wird (vgl. Abschnitt 2.5.1). Die Koordinaten des Punkts im ersten Bild sind somit exakt, die Koordinaten im zweiten Bild sind aber verrauscht, da der Punktverfolger (aufgrund verrauschter Bilddaten) nicht fehlerfrei die Koordinaten schätzen kann. Der Fehler tritt unter dieser Modellannahme nur im zweiten Bild auf. Diese Einschränkung erleichtert die analytischen Betrachtungen erheblich, sie wird aber im später beschriebenen, allgemeineren 3-D-Modell aufgehoben.

Das Rauschen der Koordinaten wird nicht als stochastisches Rauschen modelliert. Stattdessen wird angenommen, dass der in der Bildebene maximal auftretende Fehler ε sei. Dabei wird untersucht, wie sich der Rekonstruktionsfehler e verhält, wenn dieser größte anzunehmende Fehler ε auftritt („worst-case“-Szenario).

Die Rekonstruktionsgüte für einen einzelnen Punkt wird folgendermaßen gemessen: O. B. d. A. wird angenommen, dass der Fehler ε in der linken Kamera auftritt. Die Bildebene ist im 2-D-Modell eindimensional. Der Fehler ε kann deshalb in positiver oder negativer Richtung auftreten. Für diese beiden Fälle entsteht jeweils ein rekonstruierter Punkt $\hat{p}_{w1}$ bzw. $\hat{p}_{w2}$ (vgl. Bild 3.3). Die Rekonstruktionsgüte e wird als quadratischer Abstand der Punkte $\hat{p}_{w1}$ und $\hat{p}_{w2}$ auf dem Sichtstrahl G_r der rechten Kamera definiert. Die Rekonstruktionsgüte ist definiert als Summe der Abstände der beiden fehlerhaft rekonstruierten Punkte zum realen, was sich in diesem Modell noch vereinfachen lässt:

$$e(f_l, t_l, t_r, \gamma_l) = \|\hat{p}_{w1} - p_w\|^2 + \|p_w - \hat{p}_{w2}\|^2 = \|\hat{p}_{w1} - \hat{p}_{w2}\|^2. \quad (3.9)$$

Dabei bezeichnet f die Brennweite, t die Translation bezüglich des Weltkoordinatensystems und γ den Rotationswinkel um das optische Zentrum O_Z der jeweiligen Kamera und die Subskripte l bzw. r kennzeichnen die Kameraparameter der linken bzw. rechten Kamera. Die Einheit von e ist demnach Längeneinheiten zum Quadrat. Es lässt sich leicht zeigen, dass die Rotation oder eine Brennweitenänderung der rechten Kamera keinen Einfluss auf den Sichtstrahl G_r besitzt, da dieser nur durch den ursprünglichen 3-D-Punkt und das optische Zentrum definiert ist. Die Fehlerfunktion e ist also unabhängig von der Rotation γ_r und der Brennweite f_r der rechten Kamera. Für das zweidimensionale Modell ist in Bild 3.3 eine typische Konfiguration dargestellt.

Die Fehlerfunktion e kann analytisch hergeleitet werden (vgl. Anhang B). Für den Spezialfall aus Bild 3.3 (hier gilt $\gamma_l = 0$, $t_{lz} = 0$ und $t_{rz} = 0$) ergibt sich die Fehlerfunktion in Längeneinheiten zum Quadrat:

$$e = \frac{4f_l^2 \varepsilon^2 p_{wz}^2 (t_{rx} - t_{lx})^2 ((t_{rx} - p_{wx})^2 + p_{wz}^2)}{((t_{rx} - t_{lx})^2 f_l^2 - \varepsilon^2 p_{wz}^2)^2}. \quad (3.10)$$

3.2.1.2 Optimierung der Brennweite

Zunächst wird der Einfluss der Brennweite f_l untersucht. Um die Extrema der Funktion e aus Gleichung (3.9) zu finden wird die Ableitung nach f_l betrachtet:

$$\frac{\partial e}{\partial f_l} = - \frac{8f_l p_{wz}^2 \varepsilon^2 (t_{rx} - t_{lx})^2 ((t_{rx} - p_{wx})^2 + p_{wz}^2) ((t_{rx} - t_{lx})^2 f_l^2 + p_{wz}^2 \varepsilon^2)}{((t_{rx} - t_{lx})^2 f_l^2 - p_{wz}^2 \varepsilon^2)^3}. \quad (3.11)$$

Es lässt sich zeigen¹, dass für

$$f_l \in]0, p_{wz} \varepsilon / (t_{lx} - t_{rx})[\quad (3.12)$$

der Punkt $\hat{p}_{w1}$ hinter der Kamera liegt. Dieser Fall ist also ohne praktische Bedeutung. Für

$$f_l \in]p_{wz} \varepsilon / (t_{lx} - t_{rx}), \infty[\quad (3.13)$$

ist die erste Ableitung stets negativ. Das heißt, die Fehlerfunktion ist streng monoton fallend. Daraus folgt, die Brennweite f_l sollte so groß wie möglich sein. Natürlich ist in praktischen Anwendungen die maximale Brennweite durch das Objektiv vorgegeben und selbstverständlich müssen die zu rekonstruierenden Punkte im Sichtbereich der Kamera liegen.

3.2.1.3 Optimierung der Stereobasis

In der Anfangskonfiguration sind die Kameras gleich weit von dem zu rekonstruierenden Punkt entfernt und liegen symmetrisch zum Weltkoordinatensystem, das gem. Bild 3.4 festgelegt ist. Als Stereobasis wird der Abstand der beiden Kameras bezeichnet:

$$s = \|t_l - t_r\|. \quad (3.14)$$

Für die Analyse der Stereobasis s werden deshalb nur Translationen in x -Richtung zugelassen.

¹Es wird zusätzlich angenommen: Der 3-D-Punkt liegt vor der Kamera, d.h. $p_{wz} > 0$, für den Fehler gilt $\varepsilon > 0$, und $t_{lx} > t_{rx}$, d.h. linke und rechte Kamera sind nicht vertauscht.

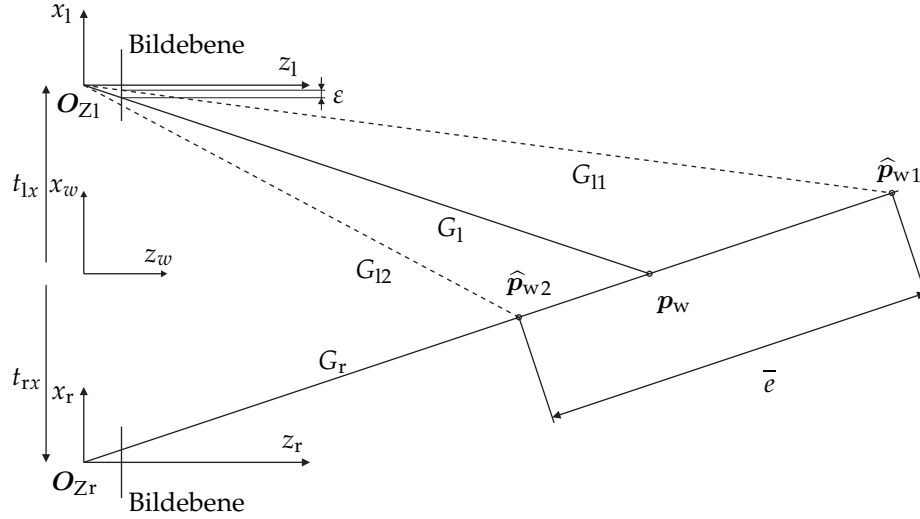


Bild 3.4 Maßstabsgetreue Anfangskonfiguration für die numerische Simulation für die Optimierung der Rotation und der Translation.

Durch Ableiten der Funktion e nach t_{lx} und t_{rx} und Nullsetzen der partiellen Ableitungen ergibt sich ein nichtlineares Gleichungssystem mit Polynomen vom Grad 5 in den gesuchten Komponenten t_{lx} und t_{rx} . Dieses ist im Allgemeinen analytisch nicht lösbar [Jac64]. Deshalb werden zunächst die Grenzfälle $t_{lx} \rightarrow \infty$ und $t_{rx} \rightarrow -\infty$ betrachtet. Anschließend wird mit Hilfe eines Gradientenabstiegs eine konkrete Konstellation analysiert.

Eine allgemein bekannte Faustregel besagt, dass die Stereobasis so groß wie möglich sein sollte. Diese Regel soll nun etwas genauer betrachtet werden. Es gilt:

$$\lim_{t_{lx} \rightarrow \infty} e = 0, \quad (3.15)$$

aber für die Translation in x-Richtung der rechten Kamera gilt:

$$\lim_{t_{rx} \rightarrow -\infty} e = \frac{4\epsilon^2 p_{wz}^2}{f_1^2}. \quad (3.16)$$

Die Asymmetrie erklärt sich dadurch, dass die Koordinaten des Punkts auf der rechten Bildebene exakt sind, auf der linken aber verrauscht. Es zeigt sich, selbst mit unendlich großer Stereobasis verbleibt ein systematischer Fehler.

Nun soll eine konkrete Konfiguration betrachtet und mit Hilfe eines Gradientenabstiegs analysiert werden. Alle Längenangaben sind im Folgenden in der gleichen Einheit gemessen und sind deshalb nicht explizit angegeben. Der 3-D-Punkt sei $p_w = (0 \ 0 \ 15)$, die Brennweite $f_1 = 1$, der maximale Fehler $\epsilon = 1/4$. In Bild 3.4 findet man eine maßstabsgetreue Zeichnung der Ausgangskonfiguration.

Bild 3.5 zeigt für verschiedene Initialisierungen von t_{lx}, t_{rx} den Verlauf der beiden Parameter beim Gradientenabstieg. Es lässt sich beobachten, dass t_{rx} in die Nähe von Null konvergiert, während die Translation t_{lx} in jedem Schritt größer wird. Die Werte scheinen gegen eine waagerechte Asymptote zu konvergieren.

Die Ableitung der Funktion e nach t_{lx} ist:

$$\frac{\partial e}{\partial t_{lx}} = - \frac{8f_1^2 \epsilon^2 p_{wz}^2 (t_{lx} - t_{rx}) (p_{wz}^2 + (p_{wx} - t_{rx})^2) (f_1^2 (t_{rx} - t_{lx})^2 + \epsilon^2 p_{wz}^2)}{((t_{rx} - t_{lx})^2 f_1^2 - \epsilon^2 p_{wz}^2)^3}. \quad (3.17)$$

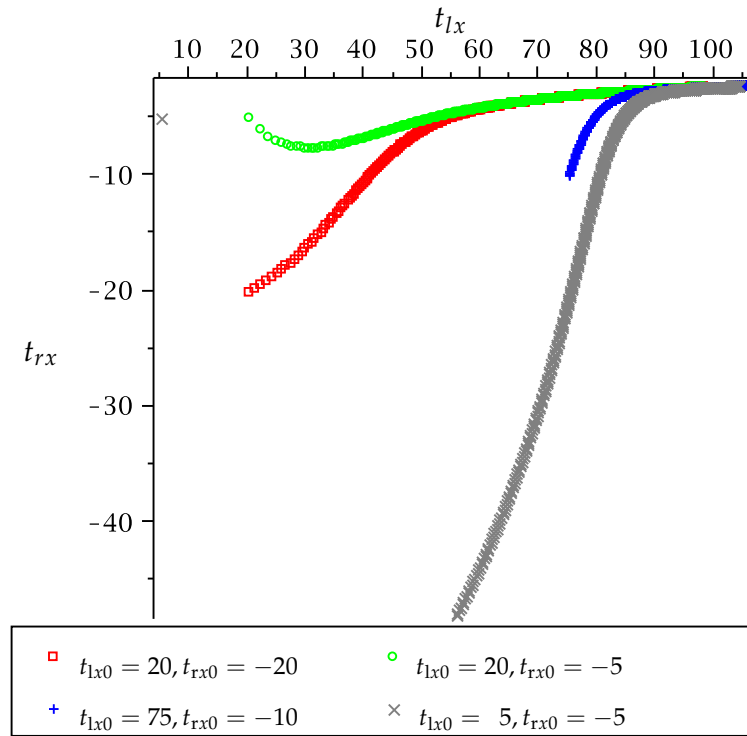


Bild 3.5 Optimierung der Stereobasis: Anhand der Trajektorien ist zu sehen, dass t_{lx} während des Gradientenabstiegs stets größer wird, während t_{rx} gegen Null konvergiert. Dieses Verhalten lässt sich bei allen vier Initialisierungen von t_{lx0}, t_{rx0} beobachten.

Für die Ableitung (3.17) gilt: sie ist stets negativ, falls²

$$(t_{rx} - t_{lx})f_1 > \varepsilon p_{wz}. \quad (3.18)$$

Ähnlich wie bei der Brennweite lässt sich zeigen, dass sich die beiden Sichtstrahlen hinter der Kamera treffen, falls die Bedingung (3.18) nicht erfüllt ist. Ist sie hingegen erfüllt, so ist die partielle Ableitung nach t_{lx} stets negativ. Das bedeutet, je größer die Translation der linken Kamera in x -Richtung t_{lx} wird, desto kleiner wird der Rekonstruktionsfehler e .

Leider ist eine einfache analytische Betrachtung für die Translation der rechten Kamera nicht möglich. Offensichtlich ist die partielle Ableitung nach t_{rx} nicht überall negativ, denn das würde dem Verhalten im Gradientenabstieg widersprechen. Es lässt sich jedoch anhand Bild 3.5 vermuten, dass beim Gradientenabstieg t_{rx} gegen p_{wx} konvergiert.

²Es wird zusätzlich angenommen: der 3-D-Punkt liegt vor der Kamera, d. h. $p_{wz} > 0$, für den Fehler gilt $\varepsilon > 0$, und $t_{lx} > t_{rx}$, d. h. linke und rechte Kamera sind nicht vertauscht.

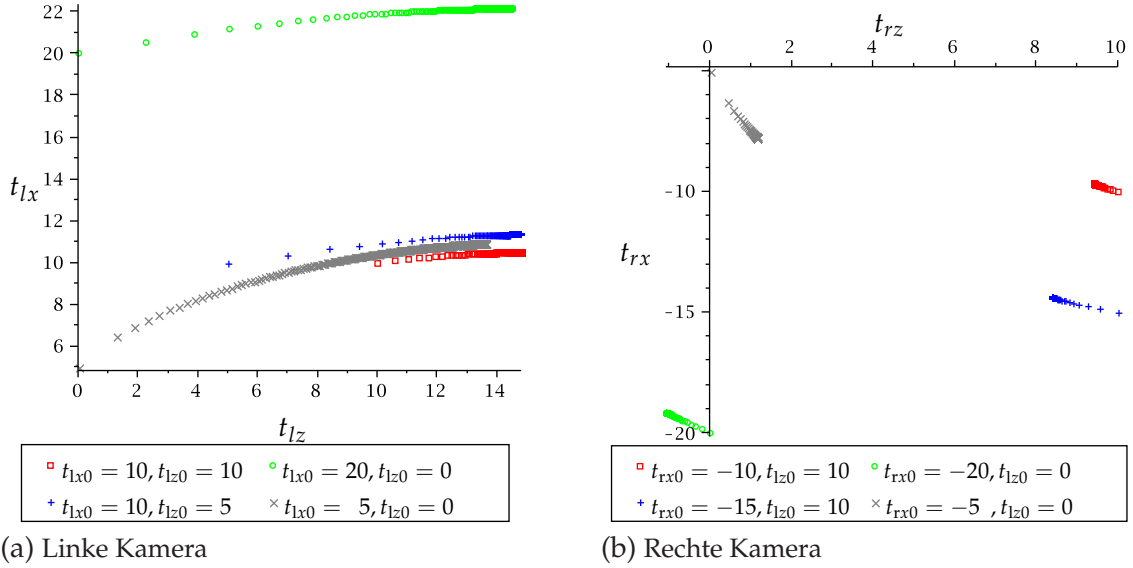


Bild 3.6 Im Bild wird das Verhalten der Translationsparameter von (a) linker bzw. (b) rechter Kamera beim Gradientenabstieg für verschiedene Startwerte dargestellt. Bei der linken Kamera wird für verschiedene Initialisierungen von t_{lx0}, t_{lz0} der Parameter t_{lz} schnell größer, während der Parameter t_{lx} nur langsam wächst. Bei der rechten Kamera lässt sich bei den verschiedenen Initialisierungen von t_{rx0}, t_{rz0} kein einheitliches Verhalten feststellen.

3.2.1.4 Optimierung der Translation in x - und z -Richtung

Lässt man Translationen in z -Richtung zu, muss das Modell des normalisierten Stereosystems erweitert werden. Die entsprechende Fehlerfunktion³ lautet:

$$e = \frac{4f_1^2 \varepsilon^2 (p_{wz} - t_{lz}^4 ((p_{wx} - t_{rx})^2 - (p_{wz} - t_{rz})^2) \eta_1^2}{(f_1^2 \eta_1^2 - \varepsilon^2 (p_{wz} (p_{wz} - t_{lz} - t_{rz}) + t_{lz} t_{rz})^2)^2} \quad (3.19)$$

mit der Hilfsvariablen

$$\eta_1 := t_{lz} (p_{wx} - t_{rx}) - t_{rz} (p_{wx} - t_{lx}) + p_{wz} (t_{rx} - t_{lx}). \quad (3.20)$$

Für die Analyse dieser Gleichungen wird erneut ein Gradientenabstieg für eine konkrete Konfiguration durchgeführt. Der 3-D-Punkt sei $\mathbf{p}_w = (0 \ 0 \ 15)^T$, die Brennweite $f_1 = 1$, der maximale Fehler $\varepsilon = 1/4$. In Bild 3.6 ist für verschiedene Initialisierungen der Verlauf der Translationen während des Gradientenabstiegs dargestellt. Zusätzlich zeigt Bild 3.7a den Fehler der linken Kamera bei fester rechter Kamera mit $\mathbf{t}_r = (-5 \ 0 \ 0)^T$ in Abhängigkeit der Translation in x -Richtung t_{lx} und z -Richtung t_{lz} . Analog zeigt Bild 3.7b den Fehler in Abhängigkeit der Translation der rechten Kamera bei fester linker Kamera mit $\mathbf{t}_l = (5 \ 0 \ 0)^T$.

Es zeigt sich beim Gradientenabstieg (in Bild 3.6a), dass die Translation der linken Kamera in x -Richtung relativ kleine Schritte macht, während die in z -Richtung wesentlich größer sind. Daraus lässt sich schließen, dass eine Verschiebung der Kamera zum Objekt hin wesentlich stärker zur Fehlerreduktion beiträgt als die Vergrößerung der Stereobasis. Trotzdem lässt sich anhand von Bild 3.6a feststellen,

³Wenn die Translationen in z -Richtung Null werden, erhält man die ursprüngliche Fehlerfunktion e wie in Gleichung (3.9). Die allgemeine Herleitung der Fehlerfunktion findet sich in Anhang B.

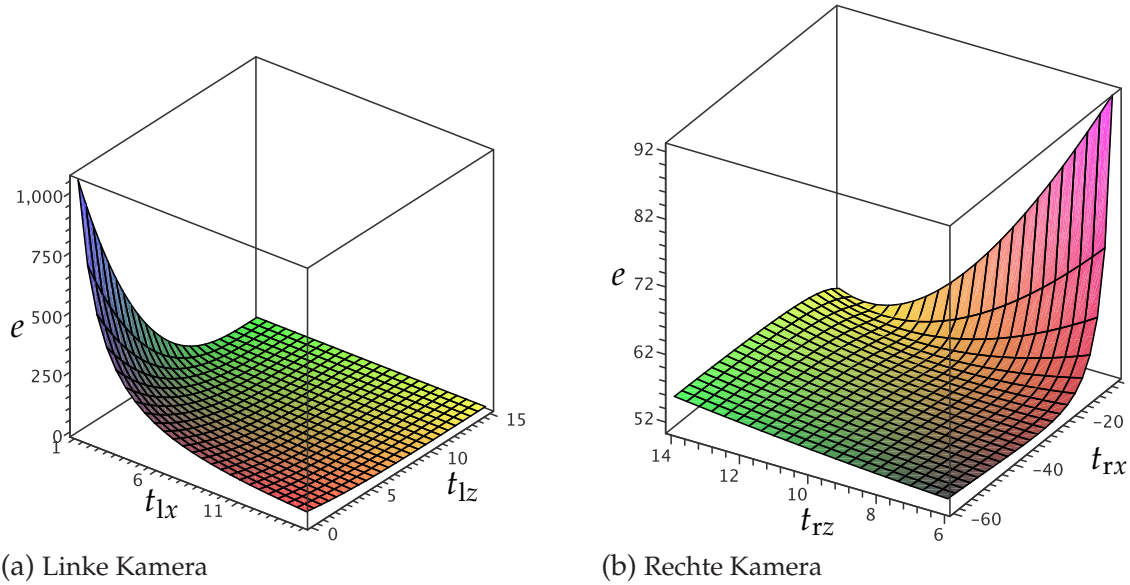


Bild 3.7 Fehler in Abhängigkeit der Translation (a) der linken Kamera bzw. (b) der rechten Kamera.

dass die Translationsparameter der linken Kamera nicht gegen ein einheitliches Minimum konvergieren. Dies liegt auch nahe, wenn man Bild 3.7a betrachtet. Dort zeigt sich, dass sich der Fehler mit größerer Translation in x und z Richtung asymptotisch verringert, jedoch kein Minimum vorhanden ist.

Generell sind die Bewegungen der rechten Kamera kleiner als die der linken. Für die unterschiedlichen Initialisierungen erkennt man im Bild 3.6b im Gegensatz zur linken Kamera keine einheitliche Bewegung der rechten Kamera. Anhand von Bild 3.7b sieht man auch, dass die Fehlerfunktion nicht asymptotisch bei kleineren Translationen in x - bzw. größeren in z -Richtung gegen ein Minimum konvergiert. Hält man t_{rz} fest, z. B. bei $t_{rz} = 8$ so zeigt sich, dass der Fehler zunächst für kleiner werdende t_{rx} abnimmt, bevor er anschließend wieder anwächst (vgl. Bild 3.8). Entscheidend ist in diesem Fall der Winkel zwischen den Sichtstrahlen: bei ca. 90° wird der Rekonstruktionsfehler e minimal.

3.2.1.5 Optimierung der Rotation

Als letzter noch verbliebener Parameter wird nun der Rotationswinkel γ_1 der linken Kamera untersucht (der Rotationswinkel der rechten hat keinen Einfluss, wie bereits in Abschnitt 1.1.2 diskutiert und in Bild 1.5 skizziert wurde). Durch die Rotation handelt es sich nicht mehr um ein normalisiertes Stereosystem. Trotzdem hat die Rotation natürlich einen Einfluss auf die Rekonstruktionsgüte, der hier analysiert werden soll.

Die Fehlerfunktion lautet (Herleitung siehe Anhang B):

$$e = \frac{4\varepsilon^2\eta_2((t_{lx} - p_{wx})\sin(\gamma_1) - p_{wz}\cos(\gamma_1))^4((t_{rx} - p_{wx})^2 + p_{wz}^2)}{(\eta_2^2 - \varepsilon^2(\eta_3 + \eta_4)^2)^2}, \quad (3.21)$$

mit den Hilfsvariablen

$$\eta_2 = (t_{rx} - t_{lx})^2 f_1^2 p_{wz}^2,$$

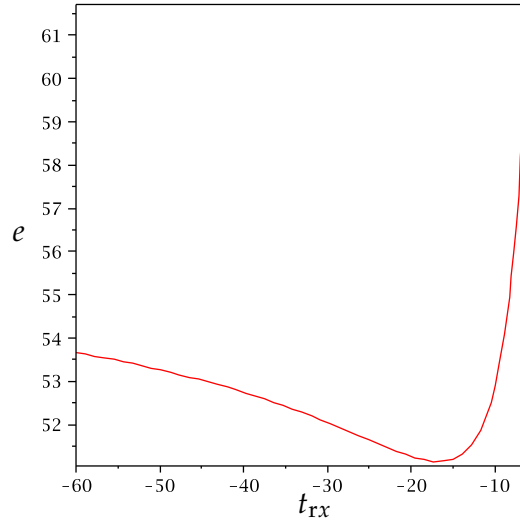


Bild 3.8 Fehlerverlauf bei Verschiebung der rechten Kamera in x -Richtung, bei fester linker Kamera $t_l = (5 \ 0 \ 0)^T$ und $t_{rz} = 10$.

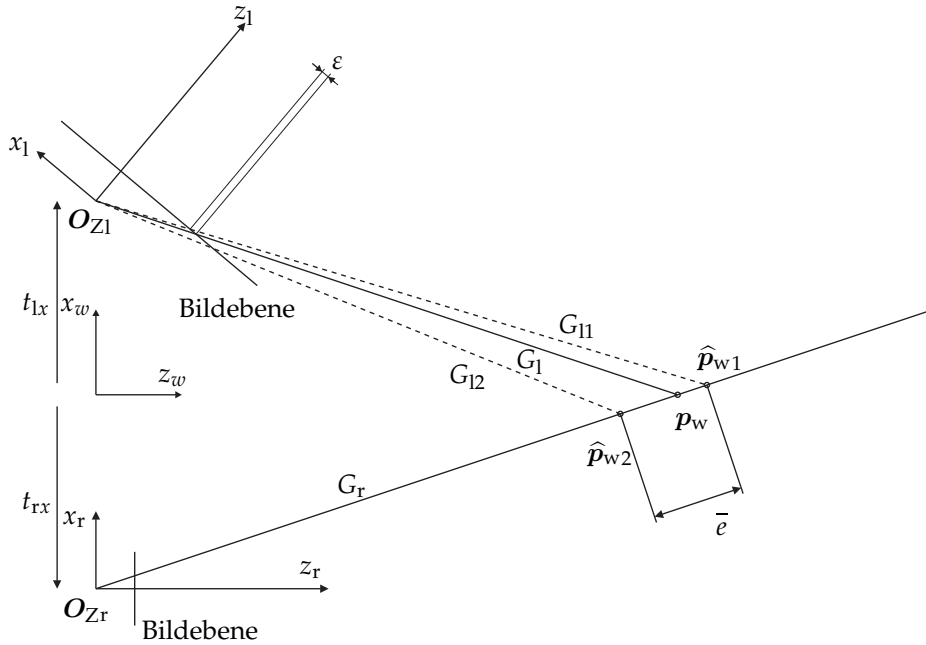


Bild 3.9 Die linke Kamera ist um 50° gedreht. Der Rekonstruktionsfehler ist dadurch gegenüber Bild 3.4 wesentlich kleiner geworden.

$$\eta_3 = (p_{wz} \cos(\gamma_1) + p_{wx} \sin(\gamma_1))^2,$$

$$\eta_4 = -(t_{lx} + t_{rx})(p_{wx} \sin^2(\gamma_1) + p_{wz} \cos(\gamma_1) \sin(\gamma_1)) + t_{lx} t_{rx} \sin^2(\gamma_1).$$

Gleichung (3.21) kann nach γ_1 abgeleitet werden. Die analytische Bestimmung der Nullstellen ist jedoch aufgrund der trigonometrischen Funktionen sehr aufwändig. Deshalb wird nur ein Fall exemplarisch erläutert und diskutiert: Sei der 3-D-Punkt $p_w = (0 \ 0 \ 15)$, die Brennweite $f_1 = 1$, der maximale Fehler $\varepsilon = 1/4$ und die Translationen der beiden Kameras $t_l = (5 \ 0 \ 0)^T$ bzw. $t_r = (-5 \ 0 \ 0)^T$. Dies entspricht für $\gamma_1 = 0$ der Anfangskonfiguration aus Bild 3.4. Es existieren im Intervall $[-90^\circ, 90^\circ]$ zwei Nullstellen der Ableitung von e , nämlich bei $\gamma_{l1} \approx 1.249 \approx 71,6^\circ$ und bei $\gamma_{l2} \approx 0.244 \approx -14,0^\circ$. Bei γ_{l2} befindet sich jedoch das

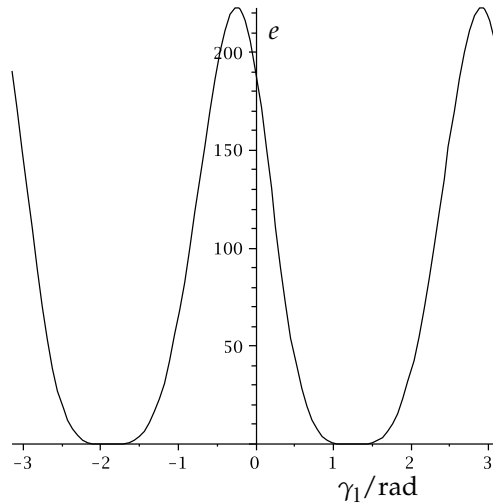


Bild 3.10 Der Rekonstruktionsfehler e in Abhängigkeit des Rotationswinkels der linken Kamera γ_1 .

Maximum der Fehlerfunktion (vgl. Bild 3.10). Also muss die Kamera um ca. $71,6^\circ$ gegen den Uhrzeigersinn gedreht werden. Um anzudeuten, was bei der Rotation geschieht, ist in Bild 3.9 die Konfiguration bei einer Rotation um 50° dargestellt.⁴ Anhand der Zeichnung ist zu sehen, dass der Rekonstruktionsfehler durch die Rotation wesentlich geringer geworden ist. In Bild 3.10 ist der Rekonstruktionsfehler in Abhängigkeit des Winkels aufgetragen. Offensichtlich ist die Fehlerfunktion e periodisch mit Periode π .

Das Wegdrehen der Kameras scheint zunächst der Intuition zu widersprechen. Es kann jedoch bei näherer Betrachtung plausibel gemacht werden: Der Rekonstruktionsfehler ist die Länge des Schattens der Strecke $[-\varepsilon, \varepsilon]$ in der Bildebene der linken Kamera auf den Sichtstrahl G_r der rechten Kamera. Die Länge dieses Schattens auf dem rechten Sichtstrahl G_r wird natürlich umso kleiner, je spitzer der Winkel zwischen der Strecke $[-\varepsilon, \varepsilon]$ in der Bildebene und dem Sichtstrahl G_l wird. Der Fehler wird Null, sobald sich die Bildebene parallel zum Sichtstrahl befindet, d. h.

$$\lim_{\gamma_1 \rightarrow \gamma_G} e = 0, \quad (3.22)$$

mit dem Grenzwinkel γ_G :

$$\gamma_G = \begin{cases} \arctan(p_{wz}/(t_{lx} - p_{wx})) & \text{falls } t_{lx} \neq p_{wx} \\ \pi/2 & \text{sonst} \end{cases}. \quad (3.23)$$

In diesem Fall liegt die Bildebene jedoch parallel zum Sichtstrahl und das heißt, der Schnittpunkt liegt im Unendlichen. In praktischen Fällen ist diese Grenzwertüberlegung jedoch irrelevant, da bei großen Winkeln der optische Sensor in der Kamera sehr groß sein muss, um überhaupt den zu rekonstruierenden Punkt im Sichtbereich zu haben.

Es gibt noch zwei weitere Indizien, dass sich das Wegdrehen der Kamera für die Rekonstruktionsgüte positiv auswirkt: Zum einen wird in den Arbeiten zur Objektverfolgung [Zob04, Deu06], bei denen auch die 3-D-Position berechnet wird,

⁴Bei einer Zeichnung mit ca. 71° wäre der Rekonstruktionsfehler ohne Änderung der Skalierung nicht mehr sichtbar.

ein ähnliches Verhalten festgestellt. Zum anderen, wenn man den probabilistischen Ansatz aus Kapitel 4 auf dieses Modell anwendet und nur die Rotation optimiert, so ergibt sich als optimaler Rotationsparameter ebenfalls ein Wegdrehen der Kamera. Details hierzu werden in Anhang C erläutert.

3.2.1.6 Optimierung aller Parameter

Zum Abschluss wird nun diskutiert, wie sich das zweidimensionale Modell bei der Optimierung aller Parameter verhält. Aufgrund der bisherigen Ergebnisse kann festgestellt werden, dass kein eindeutiges globales Minimum der Fehlerfunktion e existieren kann. Zur Begründung:

Betrachtet werden soll hier zunächst die Kombination der Parameter Translation in x -Richtung und die Rotation der linken Kamera. Aus Abschnitt 3.2.1.5 ist bekannt, dass bei fixer Translation der Rotationswinkel γ_1 so eingestellt werden kann, dass der Fehler gegen Null konvergiert (vgl. Gleichung (3.22)), falls der Winkel γ_1 gegen den entsprechenden Grenzwinkel γ_G konvergiert. Da ein solcher Grenzwinkel für alle $t_{1x} \in \mathbb{R}$ existiert und der Wert Null für die Fehlerfunktion e aufgrund ihrer Konstruktion (Summe quadrierter Terme – vgl. Gleichung (3.9)) ein absolutes Minimum darstellt, folgt, dass hier kein eindeutiges Minimum existieren kann. Da kein eindeutiges globales Minimum existieren kann, wenn man die Parameter Translation und Rotation in x -Richtung alleine optimiert, kann auch kein globales Minimum existieren, wenn man alle Parameter optimiert. Dies war zu zeigen.

Es sei an dieser Stelle aber nochmals explizit darauf hingewiesen, dass diese Grenzwertbetrachtungen nur dann Gültigkeit besitzen, wenn von einer unbegrenzten Bildebene ausgegangen wird. Da dies in der Praxis aber nicht erfüllt werden kann, wird im nun folgenden 3-D-Modell diese Nebenbedingung bei der Optimierung der unterschiedlichen Parameter modelliert.

3.2.1.7 Zusammenfassung

In diesem Abschnitt wurden für das zweidimensionale Modell die einzelnen Parameter Brennweite, Translation in x -Richtung, Translation in x - und z -Richtung und Rotation sowie alle Parameter zusammen analysiert. Zusammenfassen lassen sich die Ergebnisse, bei einer einzelnen Betrachtung dieser Parameter, wie folgt:

Die Brennweite sollte so groß wie möglich sein. Diese Aussage konnte analytisch hergeleitet werden mit dem Ergebnis, dass der Fehler mit größer werdender Brennweite monoton fallend ist.

Bei der Translation in x -Richtung sollte die der linken Kamera möglichst groß sein. Die der rechten sollte dagegen einen Wert nahe der x -Koordinate des zu untersuchenden Punkts sein. Lässt sich die Translation in x - und z -Richtung ändern, so sollte die linke Kamera möglichst nahe an den Weltpunkt herangebracht werden. Bei der Translation der rechten Kamera ergab sich keine einfache Regel, wie diese zu positionieren ist.

Die Analyse der Rotation hat ein zunächst überraschendes Ergebnis geliefert: Die linke Kamera sollte sich wegdrehen. Dieses Verhalten scheint zunächst nicht plausibel, konnte aber bei näherer Betrachtung erklärt werden.

Bei der Analyse aller Parameter konnte gezeigt werden, dass die entsprechende Fehlerfunktion ohne eingeschränkte Bildebene unendlich viele globale Mini-

ma besitzt. Im nun folgenden 3-D-Modell wird unter anderem der eingeschränkte Sichtbereich durch eine endliche Bildebene mit berücksichtigt. Trotzdem sind die entsprechenden 2-D-Analysen wichtig, da viele Aussagen analytisch gezeigt werden konnten und sich diese nun vom 2-D-Modell auf das 3-D-Modell übertragen lassen.

3.2.2 Dreidimensionales Modell

Das zweidimensionale Modell wird nun erweitert, um realistischere Bedingungen zu untersuchen. Die Erweiterungen betreffen:

1. Es werden 3-D-Punkte auf die 2-D-Bildebene perspektivisch projiziert. Die y -Komponenten werden nicht mehr vernachlässigt, wie im 2-D-Modell.
2. Es wird angenommen, dass die Koordinaten der Bildpunkte des linken und rechten Bilds durch additives, mittelwertfreies, Gauß-Rauschen überlagert sind. Das Gauß-Rauschen hat Erwartungswert 0 und Varianz σ^2 und ist für die jeweiligen x - und y -Komponenten statistisch unabhängig. Dadurch werden die Sichtstrahlen im Allgemeinen windschief.
3. Die Bildebene sei quadratisch, also Breite b_x = Höhe b_y .
4. Es werden $n > 1$ Punkte betrachtet.

Die Berechnung des Erwartungswerts der Koordinaten eines 3-D-Punkts ist aufgrund des Rauschens und den daraus resultierenden meist windschiefen Geraden für die Triangulation (es wird das lineare Verfahren aus Abschnitt 2.5.2.2 verwendet) sehr aufwändig: Für $m = 2$ Bilder lässt sich für die Triangulation aus Gleichung (2.60) ein homogenes Gleichungssystem mit einer 4×4 Komponentenmatrix aufstellen. Sind die Bildpunkte unverrauscht, so besitzt das homogene Gleichungssystem nichttriviale Lösungen. Man kann das Gleichungssystem analytisch lösen und anschließend die Bildpunkte in der Lösung durch Addition des Rauschterms verrauschen. Die Weltkoordinaten des 3-D-Punkts sind dann durch eine Wahrscheinlichkeitsverteilung gegeben. Es ist aber sehr aufwändig die konkrete Verteilung zu berechnen, denn in der Lösung des homogenen Gleichungssystems für die Komponenten des 3-D-Punkts in Weltkoordinaten finden sich Produkte und Quotienten der Rauschterme. Die Verteilung ist im Allgemeinen keine Normalverteilung mehr. Sie kann mit Hilfe des Transformationssatzes ([Irl05, Seite 123]) für Dichten berechnet werden. Aufgrund des hohen Aufwands, den eine analytische Berechnung des Erwartungswerts verursacht, wird eine Monte-Carlo-Simulation [Dou01] benutzt.

Mit Hilfe der Monte-Carlo-Simulation wird der Erwartungswert des Triangulationsfehlers geschätzt und mit „sequentieller quadratischer Programmierung“ (SQP) optimiert [Han77, Pow78a, Pow78b]. SQP ist ein lokales Optimierungsverfahren für nichtlineare Probleme mit Nebenbedingungen. Als Nebenbedingungen wurden hier zum einen sinnvolle Bereiche für die zu optimierenden Parameter vorgegeben, d. h. Rotationswinkel zwischen $-\pi/2$ und $\pi/2$ und Brennweite größer Null. Zusätzlich wird sichergestellt, dass die projizierten 3-D-Punkte im Sichtbereich der Kamera liegen. Für die Untersuchungen wurde die Implementierung der Matlab Optimization Toolbox verwendet.

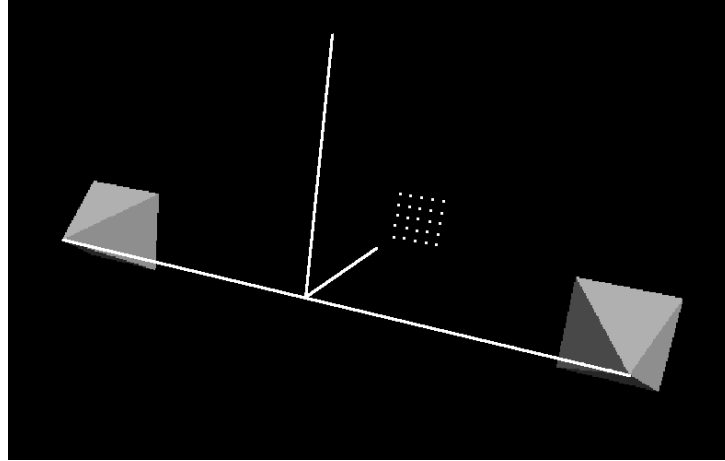


Bild 3.11 Anfangskonfiguration. Die Kameras sind als Pyramiden dargestellt. Die Spitze entspricht dem optischen Zentrum, die Grundfläche der Bildebene.

Die Ausgangssituation und das Weltkoordinatensystem sind in Bild 3.11 skizziert. Die Kameraparameter für die Ausgangssituation sind

$$\begin{aligned} t_l &= (300 \ 0 \ 0)^T, & t_r &= (-300 \ 0 \ 0)^T, \\ R_l &= R_r = I_{3 \times 3}, & f_l &= f_r = 10. \end{aligned}$$

Es werden $n = 25$ 3-D-Punkte betrachtet, die in einem regulären Gitter angeordnet sind, wie in Bild 3.11. Für die Koordinaten gilt:

$$p_w = (20i \ 20j \ 500)^T \quad i, j = -2, -1, \dots, 2. \quad (3.24)$$

Schließlich wird noch festgelegt, das Rauschen in den Bildkoordinaten habe eine Varianz von $\sigma_q^2 = 0,2$ und die Seitenlänge des optischen Sensors sei $b_x = 30$.

Wie im zweidimensionalen Modell sollen nun die Parameter zunächst einzeln analysiert werden. Zum Schluss werden alle veränderbaren Parameter zusammen optimiert, was ohne die Nebenbedingungen nicht sinnvoll ist, wie gezeigt wurde. Zunächst erfolgt nun die Definition des Fehlermaßes.

3.2.2.1 Fehlermodell im dreidimensionalen Fall

Die Rekonstruktionsgüte wird auch in diesem Fall durch die Abweichung von dem tatsächlichen zu dem rekonstruierten 3-D-Punkt gemessen:

$$e(f_l, f_r, t_l, t_r, R_l, R_r) = \frac{1}{n} \sum_{i=1}^n \|p_{wi} - \hat{p}_{wi}\|, \quad (3.25)$$

dabei bezeichnet n die Anzahl der Punkte, f die Brennweite, t den Translationsvektor, R die Rotationsmatrix, $\hat{p}_w$ die Schätzung der Koordinaten des triangulierten Punkts und p_w die ursprünglichen Koordinaten des 3-D-Punkts.

Da die 2-D-Punkte durch normalverteiltes Rauschen überlagert sind, ist der Erwartungswert des Fehlers e von Interesse. Der empirische Erwartungswert lässt sich durch eine Monte-Carlo-Simulation aus S Stichproben ermitteln:

$$\bar{e}(f_l, f_r, t_l, t_r, R_l, R_r) = \frac{1}{Sn} \sum_{i=1}^n \sum_{j=1}^S \|p_{wi} - \hat{p}_{wi}^{(j)}\|. \quad (3.26)$$

Der empirische Erwartungswert des Rekonstruktionsfehlers $\bar{e}$ in der oben beschriebenen Anfangskonfiguration bei $S = 1000$ Stichproben beträgt $\bar{e} = 14,1$ und die Varianz des Fehlers, definiert als

$$\sigma_e^2 = \frac{1}{n(S-1)} \sum_{i=1}^n \sum_{j=1}^S (\|p_{wi} - \hat{p}_{wi}^{(j)}\| - \bar{e})^2, \quad (3.27)$$

beträgt in der Ausgangskonfiguration $\sigma_e^2 = 42,87$. Die Varianz des Fehlers ist in diesem Fall deshalb so hoch, da schon eine kleine Änderung der Bildkoordinaten sehr große Fehler bewirkt. Ist der Fehler in der Bildebene konstant $\varepsilon = 0,2$ so beträgt der Rekonstruktionsfehler $e = 14,1$, bei einem konstanten Fehler in der Bildebene von $\varepsilon = 0,4$ beträgt der Rekonstruktionsfehler $e = 28,3$. Dadurch lässt sich die relativ hohe Varianz erklären.

3.2.2.2 Optimierung der Brennweite

Mit Hilfe der SQP wird die Brennweite optimiert. Der Algorithmus vergrößert die Brennweite in beiden Kameras. Die optimalen Brennweiten in der obigen Konfiguration sind:

$$f_l = f_r = 44,1. \quad (3.28)$$

Das ist die jeweils maximale Brennweite, in der die Punkte gerade noch im Bild liegen. Der erwartete Rekonstruktionsfehler reduziert sich von ursprünglich $\bar{e} = 14,1$ auf $\bar{e} = 3,20$, die Varianz von $\sigma_e^2 = 42,87$ auf $\sigma_e^2 = 2,17$.

Es lässt sich festhalten: auch im 3-D-Fall sollte die Brennweite möglichst groß sein. Natürlich gelten hier zwei Einschränkungen: Die zu rekonstruierenden Punkte müssen im Blickfeld der Kamera liegen und die Brennweite der Kamera ist natürlich durch die physikalischen Eigenschaften des Objektivs eingeschränkt.

3.2.2.3 Optimierung der Stereobasis (Translation in x - und y -Richtung)

In der zweidimensionalen Betrachtung ist die linke Kamera vom Objekt weg gewandert, um die Stereobasis zu vergrößern. Im aktuell betrachteten Fall wird deshalb erwartet, dass sich beide Kameras analog wie die linke im 2-D-Modell verhalten. Tatsächlich erhält man die optimale Lösung der Translationsvektoren

$$t_l = (1460 \ 9,39 \ 0)^T, \quad t_r = (-1459,7 \ -13,3 \ 0)^T. \quad (3.29)$$

Der Erwartungswert des Rekonstruktionsfehlers liegt bei $\bar{e} = 9,33$ und die Varianz bei $\sigma_e^2 = 20,42$. Mit der angegebenen Stereobasis ist gerade noch jeder Punkt sichtbar, d. h. er befindet sich noch im Bild.

Warum wird aber die Stereobasis nicht weiter vergrößert, indem eine Verschiebung in y -Richtung ausgenutzt wird? Falls eine andere Startposition der Kamera verwendet wird, nämlich

$$t_l = (300 \ 300 \ 0)^T, \quad t_r = (-300 \ -300 \ 0)^T, \quad (3.30)$$

ergibt sich eine Endposition mit den Werten:

$$t_l = (1444,1 \ 1459,2 \ 0)^T, \quad t_r = (-1455,4 \ -1460 \ 0)^T. \quad (3.31)$$

Der Fehler beträgt dabei $\bar{e} = 9,14$ (Varianz $\sigma_e^2 = 20,93$). Hier wird ebenfalls die Position in y -Richtung optimiert. Deshalb lässt sich vermuten, dass bei der Startposition (3.2.2) in y -Richtung die Funktion zu flach ist, so dass der SQP Algorithmus in diese Richtung nicht optimiert.

Zur Verifikation lässt sich zeigen, dass bei der Initialisierung

$$\mathbf{t}_l = (0 \ 300 \ 0)^T, \quad \mathbf{t}_r = (0 \ -300 \ 0)^T \quad (3.32)$$

ein zur Initialisierung (3.2.2) analoges Optimum gefunden wird:

$$\mathbf{t}_l = (8,28 \ 1460 \ 0)^T, \quad \mathbf{t}_r = (-12,1 \ -1458,7 \ 0)^T. \quad (3.33)$$

Der Fehler in diesem Fall ist $\bar{e} = 9,33$. Das Optimum wird analog zu der ursprünglichen Initialisierung gefunden. Ignoriert man die kleinen Abweichungen aufgrund von numerischen Ungenauigkeiten, ist das Problem symmetrisch, wie man es aufgrund des Modellaufbaus auch erwarten würde.

Abschließend lässt sich feststellen, dass die Stereobasis möglich groß sein sollte und dass die Kameras symmetrisch zum Objekt angeordnet werden sollten, um den 3-D-Fehler zu minimieren.

3.2.2.4 Optimierung der Translation in alle Richtungen

Im 2-D-Modell wurde die linke Kamera sehr nahe am Objekt positioniert. Im generalisierten 3-D-Modell erhält man analoge Ergebnisse. Die Translationsvektoren der linken bzw. rechten Kamera sind nach der Optimierung:

$$\mathbf{t}_l = (28,9 \ 1,76 \ 477,0)^T, \quad \mathbf{t}_r = (-30,2 \ -2,21 \ 476,6)^T. \quad (3.34)$$

Der Fehler in der Endkonfiguration ist $\bar{e} = 0,61$ und die Varianz $\sigma_e^2 = 0,11$.

Es zeigt sich, wie bereits im 2-D-Modell, dass die Stereobasis, also der Abstand der Kameras zueinander, einen viel geringeren Einfluss hat, als der Abstand Kamera - Objekt. Es ist daher zu bevorzugen die Kameras möglichst nahe an die zu rekonstruierenden Objektpunkte zu platzieren.

Die Unterschiede bei der Rekonstruktionsgüte sind, verglichen mit einer Optimierung der Stereobasis, erheblich: Der Fehler in der Anfangskonfiguration liegt bei $\bar{e} = 14,1$ und konnte bei Optimierung der Stereobasis auf $\bar{e} = 9,25$ (bzw. auf $\bar{e} = 9,14$ bei modifizierter Initialisierung) und bei der Optimierung der Translation in alle Richtungen auf $\bar{e} = 0,61$ reduziert werden.

3.2.2.5 Optimierung der Rotation

Auch hier ergibt sich, dass sich beide Kameras analog zu der linken Kamera im 2-D-Modell verhalten, d. h. sie drehen sich weg. Man erhält die Cardan-Winkel

$$\begin{aligned} \gamma_{lz} &= \gamma_{rz} = -0,82 \approx -47^\circ, \\ \gamma_{ly} &= -\gamma_{ry} = 0,73 \approx 42^\circ, \\ \gamma_{lx} &= -\gamma_{rx} = 0,04 \approx 2^\circ. \end{aligned} \quad (3.35)$$

Dabei steht das Subskript x, y bzw. z für die jeweilige Koordinatenachse um die gedreht wird (Reihenfolge der Rotation: zuerst um die z -, dann um die y - und

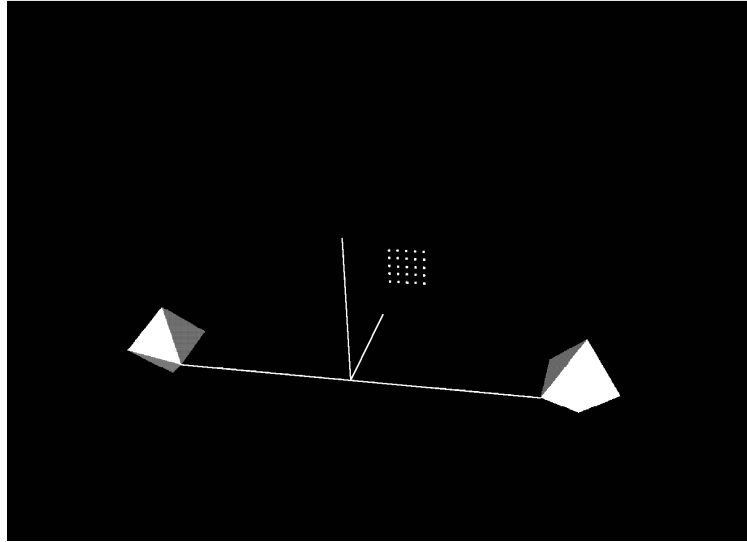


Bild 3.12 Endkonfiguration bei Optimierung der Rotation: Die Kameras haben sich weggedreht.

schließlich um die x -Achse). In Bild 3.12 ist die Endkonfiguration der Kameras, das Weltkoordinatensystem und die 3-D-Punkte eingezeichnet. Der Fehler beträgt nach der Optimierung $\bar{e} = 2,71$ mit Varianz $\sigma_e^2 = 1,90$.

Es zeigt sich, wie schon im 2-D-Fall, dass die Kameras sich nicht zum Objekt hin, sondern wegdrehen müssen um den 3-D-Fehler der Rekonstruktion zu minimieren. Auch hier lässt sich der Sachverhalt ähnlich interpretieren: Durch das Wegdrehen wird der „Schatten“ bei der Projektion auf den jeweils anderen Sichtstrahl kleiner.

3.2.2.6 Optimierung aller veränderbaren Parameter

Abschließend bleibt nun die Frage, welche Konfiguration optimal ist, falls alle veränderbaren Kameraparameter variabel sind. Diese Untersuchung ist deshalb notwendig, da die Seitenlänge des Sensors endlich ist. So kann z. B. nicht die Brennweite und die Stereobasis unabhängig voneinander optimiert werden, da sonst das zu rekonstruierende Objekt nicht mehr (komplett) im Bild liegen würde.

Startet man die Optimierung mit der initialen Konfiguration (3.31), so erhält man einen Fehlerwert von $\bar{e} = 1,64$. Dabei handelt es sich offensichtlich nur um ein lokales Minimum, da bei der Optimierung der Translation bereits ein Wert von $\bar{e} = 0,61$ erreicht wurde.

Deshalb wird der Optimierungsprozess an den jeweiligen Endpositionen der Optimierungen der einzelnen Parameter (3.28), (3.31), (3.34) und (3.35) gestartet.

Die niedrigsten Fehlerwerte erhält man bei der Startkonfiguration (3.31) mit der Endkonfiguration:

$$\begin{aligned}
 t_l &= (76,7 \quad 19,3 \quad 427,2)^T, & t_r &= (-55,3 \quad -3,4 \quad 401,0)^T, \\
 \gamma_{lz} &= \gamma_{rz} = 0 = 0^\circ, \\
 \gamma_{ly} &= \gamma_{ry} = 0 = 0^\circ, \\
 \gamma_{lx} &= \gamma_{rx} = 0 = 0^\circ, \\
 f_l &= 64,7, & f_r &= 73,3,
 \end{aligned} \tag{3.36}$$

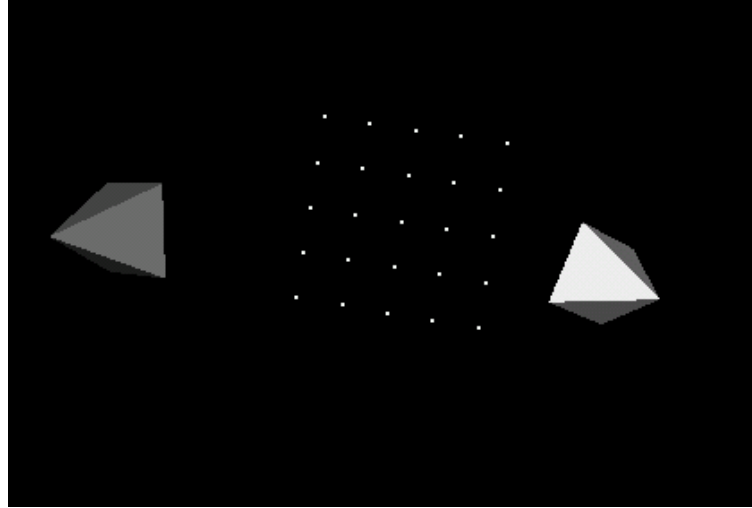


Bild 3.13 Endkonfiguration bei Optimierung aller Parameter

bzw. bei der Startkonfiguration (3.34) mit der Endkonfiguration:

$$\begin{aligned}
 t_l &= (97.4 \quad -16.5 \quad 381.8)^T, & t_r &= (-90.9 \quad -9.1 \quad 417.2)^T, \\
 \gamma_{lz} &= 1.38 \approx 79^\circ, & \gamma_{rz} &= 0.31 \approx 18^\circ \\
 \gamma_{ly} &= -0.57 \approx -33^\circ, & \gamma_{ry} &= 0.65 \approx 37^\circ \\
 \gamma_{lx} &= -0.15 \approx -9^\circ & \gamma_{rx} &= -0.1 \approx -7^\circ, \\
 f_l &= 100, & f_r &= 75.2.
 \end{aligned} \tag{3.37}$$

Der Fehler liegt bei den beiden Endkonfigurationen (3.36) bzw. (3.37) bei $\bar{e} = 0.47$ (Varianz: $\sigma_e^2 = 0.047$) bzw. $\bar{e} = 0.46$ (Varianz: $\sigma_e^2 = 0.041$).

Die als optimal gefundenen Endkonfigurationen positionieren die Kameras wieder sehr nahe am Objekt, so dass dieses gerade noch ins Bild passt. Im Falle der Endkonfiguration (3.37) sind die Kameras zum Objekt hingedreht (vgl. Bild 3.13), obwohl sie sich bei der getrennten Betrachtung der Rotation wegdrehen. Der Einfluss der Rotation scheint damit geringer zu sein als der durch die nahe Positionierung der Kamera an das Objekt. Dadurch, dass sich die Kameras zum Objekt drehen, können sie sogar noch näher an das Objekt herangebracht werden, ohne das Objekt aus dem Sichtbereich zu verlieren.

Zum Abschluss noch eine Anmerkung zu den konkurrierenden Größen Brennweite und Translation in Richtung des Objekts: Zwar wird ein Objekt im Bild größer, wenn die Brennweite erhöht wird bzw. wenn die Kamera näher am Objekt positioniert wird. Jedoch ist bei der Änderung der Brennweite der Winkel zwischen den Sichtstrahlen konstant, wohingegen beim Verschieben der Kamera in Richtung des Objekts sich der Winkel zwischen den Sichtstrahlen vergrößert. Das ist der Grund warum eine Änderung der beiden Parameter unterschiedliche Effekte nach sich zieht.

3.2.2.7 Zusammenfassung

Für den 3-D-Fall wurden die veränderbaren Parameter jeweils einzeln oder gemeinsam optimiert. Bei der Optimierung der einzelnen Parameter konnte stets das

Veränd. Parameter	Fehler	Varianz
Initiale Konfiguration	14,1	42,87
Brennweite	3,20	2,17
Stereobasis	9,14	20,93
Rotation	2,71	1,90
Translation	0,61	0,11
Alle Parameter	0,46	0,04

Tabelle 3.1 Übersicht über verschiedene veränderbare Parameter und die daraus resultierenden Fehler.

Verhalten der linken Kamera aus dem 2-D-Modell auf das Verhalten beider Kameras übertragen werden. In Tabelle 3.1 sind die Resultate für die verschiedenen modifizierbaren Parameter zusammengefasst.

Durch die Untersuchungen konnte der Einfluss der einzelnen Parameter genau analysiert werden. So lässt sich zeigen, dass eine Positionierung der Kameras nahe am Objekt einen viel größeren Einfluss auf die Rekonstruktionsgüte hat, als die Vergrößerung der Stereobasis bewirken würde. Andererseits hätte man wohl auch vermutet, dass eine Rotation der Kameras zum Objekt den Fehler eher reduziert hätte, als eine Rotation weg vom Objekt.

Dieses Modell vernachlässigt Verzerrungen der Linse (Abschnitt 2.1.2), da davon ausgegangen wird, dass die Bilder alle stets vorher entzerrt wurden. Möchte man dennoch Linsenverzerrungen berücksichtigen, so muss man dafür ein Modell für die Verzerrungen einführen. Wird dabei das gleiche physikalische Modell verwendet, mit dessen Hilfe man die Bilder auch entzerrt, so hat man dadurch nichts gewonnen. Linsenverzerrungen zu berücksichtigen macht also nur dann Sinn, wenn man direkt auf den nicht entzerrten Bildern arbeiten möchte.

3.3 Probabilistische Ansätze

Im Folgenden werden probabilistische Ansätze diskutiert. Die verwendeten Gütemaße basieren dabei auf probabilistischen Eigenschaften. Dazu wird das Problem der 3-D-Rekonstruktion als stochastisches Schätzproblem eines Systems umformuliert. In der Literatur sind allgemeine Schätzprobleme statischer oder dynamischer Systeme ausgiebig untersucht worden [May79, BS88, Kay93, Sta94, Den03a]. Standardverfahren aus der Schätztheorie lassen sich auch zur 3-D-Rekonstruktion einsetzen. Beispielsweise wird in [Mat89, Hun99, Yu05] eine 3-D-Rekonstruktion mit Hilfe des Kalman-Filters durchgeführt. Die zitierten Verfahren sind jedoch alle passiv im Sinne des Marr Paradigmas [Mar82, Den97]. Dies bedeutet, der Algorithmus verwendet lediglich die ihm zur Verfügung gestellten Daten zur Berechnung. Es gibt also keine aktive Komponente, die Ansichten gezielt auswählt, um die Rekonstruktionsgüte zu erhöhen. Im Kapitel 4 wird ein Verfahren entwickelt, welches ein probabilistisches Gütemaß optimiert, um eine möglichst akkurate 3-D-Rekonstruktion berechnen zu können. Zum Vergleich zu dem neuen Verfahren soll nun vorher ein Algorithmus aus der Literatur dargestellt werden. Am Ende dieses Abschnitts werden dann die Gemeinsamkeiten und Unterschiede zu dem in dieser Arbeit entwickelten Algorithmus herausgearbeitet.

Der Ansatz aus der Literatur stammt aus [Ola97, Ola98, Ola02]. In den Publikationen wird zunächst ein Gütemaß eingeführt, welches anschließend optimiert wird. Auf diese beiden Schritte soll nun kurz eingegangen werden.

Für die Entwicklung eines Gütemaßes wird zunächst davon ausgegangen, dass ein 3-D-Punkt $\mathbf{p}_w$ in mehrere Kamerabilder durch die Projektionsmatrizen $\mathbf{P}_i$ perspektivisch projiziert wird (vgl. Abschnitt 2.1.4.1). Dann lassen sich (unter bestimmten Voraussetzungen an die Projektionsmatrizen $\mathbf{P}_i$, s. u.) die Koordinaten des 3-D-Punkts $\mathbf{p}_w$ berechnen. In diesem Algorithmus wird, ähnlich wie bei der linearen Triangulation aus Abschnitt 2.5.2.2, vorgegangen: Für einen 3-D-Punkt $\mathbf{p}_w$ lässt sich aus Gleichung (2.1) für den Bildpunkt $\mathbf{q}_i$ im i -ten Bild in homogenen Koordinaten die Gleichung

$$\mathbf{A}_i \mathbf{p}_w = \mathbf{b}_i \quad (3.38)$$

aufstellen (vgl. [Fau93, Seite 231 ff.] und [Ola02]) mit

$$\mathbf{A}_i := \begin{pmatrix} q_{ix}p_{i,3,1} - p_{i,1,1} & q_{ix}p_{i,3,2} - p_{i,1,2} & q_{ix}p_{i,3,3} - p_{i,1,3} \\ q_{iy}p_{i,3,1} - p_{i,2,1} & q_{iy}p_{i,3,2} - p_{i,2,2} & q_{iy}p_{i,3,3} - p_{i,2,3} \end{pmatrix}, \quad (3.39)$$

und

$$\mathbf{b}_i := \begin{pmatrix} p_{i,1,4} - q_{ix}p_{i,3,4} \\ p_{i,2,4} - q_{iy}p_{i,3,4} \end{pmatrix}, \quad (3.40)$$

mit $p_{i,j,k}$ gleich der Komponente j,k der i -ten Projektionsmatrix $\mathbf{P}_i$. Betrachtet man nun m Bilder gleichzeitig, so ergibt sich ein Gleichungssystem

$$\mathbf{A} \mathbf{p}_w = \mathbf{b}, \quad (3.41)$$

mit der $2m \times 3$ Matrix $\mathbf{A}$, in der die einzelnen Matrizen $\mathbf{A}_1, \dots, \mathbf{A}_m$ untereinander angeordnet sind und dem $2i$ -dimensionalen Vektor $\mathbf{b}$, der analog aufgebaut ist. Falls $\text{rang}(\mathbf{A}) = 3$ ist, existiert eine eindeutige Lösung; die Koordinaten lauten:

$$\mathbf{p}_w = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}. \quad (3.42)$$

Dabei entspricht $(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ der Pseudoinversen $\mathbf{A}^+$ von $\mathbf{A}$ [Tre97].

Damit lässt sich nun die Triangulation als Funktion $\phi^+ : \mathbb{R}^{2m} \rightarrow \mathbb{R}^3$ beschreiben, die aus m Bildpunkten $\mathbf{q}_i, i = 1, \dots, m$ die Koordinaten des 3-D-Punkts $\mathbf{p}_w$ schätzt.

Der Schlüssel für die Ansichtenplanung ergibt sich nun durch die Betrachtung, wie sich das Rauschen bei der Bestimmung der 2-D-Bildpunkte auf die Unsicherheit der Schätzung des 3-D-Fehlers auswirkt. Zunächst wird davon ausgegangen, dass die 2-D-Punkte einem normalverteilten Rauschen unterliegen. Die folgende Definition und der folgende Satz sind für die weitere Argumentation wichtig:

Def: 3.1 (Hinreichende Statistik)

Den minimalen Satz an Parametern, der die statistischen Eigenschaften einer Zufallsvariable vollständig beschreibt, nennt man „hinreichende Statistik“ [Den03a].

Satz 3.2

Die Parameter Erwartungswert und Kovarianzmatrix stellen die hinreichende Statistik für eine Normalverteilung dar.

Beweis: [Den03a, Seite 18].

Unter der Annahme, dass die Koordinaten der Bildpunkte durch additives normalverteiltes Rauschen überlagert sind, stellt sich nun die Frage, welchen Einfluss dies auf die Schätzung der 3-D-Punktkoordinaten, im Sinne einer Fehlerfortpflanzung, hat. Diese Frage ist im Allgemeinen sehr schwer zu beantworten, da die Transformation der 2-D-Zufallsvariablen von der Funktion ϕ^+ abhängt. Das Problem vereinfacht sich wesentlich, wenn man eine lineare Approximation der Funktion ϕ^+ mit Hilfe einer Taylor-Reihen-Entwicklung durchführt [Fau93, Ola02]. Denn hier gelten die folgenden zwei Sätze:

Satz 3.3

Sei $\mathbf{X}$ ein m -dimensionaler Zufallsvektor mit Erwartungswert $E(\mathbf{X})$ und Kovarianz $\Sigma_{\mathbf{X}}$. Wird der Zufallsvektor $\mathbf{X}$ durch eine lineare Transformation $\mathbf{AX} + \mathbf{b}$ mit der regulären $m \times n$ Matrix $\mathbf{A}$ und dem m -dimensionalen Vektor $\mathbf{b}$ abgebildet, so ist der Erwartungswert des transformierten Zufallsvektors $E(\mathbf{AX} + \mathbf{b}) = \mathbf{A}E(\mathbf{X}) + \mathbf{b}$ und die Kovarianz $\Sigma_{\mathbf{AX}+\mathbf{b}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T$.

Beweis: [Irl05, Seite 135]

Satz 3.4

Sei $\mathbf{X}$ ein m -dimensionaler normalverteilter Zufallsvektor mit positiv definiter Kovarianzmatrix. Der Zufallsvektor $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, mit der $m \times n$ Matrix $\mathbf{A}$ und dem m -dimensionalen Vektor $\mathbf{b}$, ist ebenfalls normalverteilt, falls für die Matrix $\mathbf{A}$ gilt: $m \leq n$ und $\text{rang}(\mathbf{A}) = m$.

Beweis: [Rin03, Seite 313].

Aus den Sätzen 3.2, 3.3 und 3.4 sowie aus Definition 3.1 folgt, dass aus den Parametern der normalverteilten Schätzung eines 2-D-Bildpunkts die Parameter der Schätzung des zugehörigen 3-D-Punkts berechnet werden können. Dazu wird, wie oben bereits erwähnt, die Funktion ϕ^+ in einer Umgebung von $E(\mathbf{p}_w)$ mit Hilfe einer Taylor-Reihe entwickelt:

$$\phi^+(\mathbf{q}) = \phi^+(E(\mathbf{q})) + \mathbf{J}(\phi^+(E(\mathbf{q}))) (\mathbf{q} - E(\mathbf{q})) + \boldsymbol{\rho}(\mathbf{q} - E(\mathbf{q})), \quad (3.43)$$

mit der Jacobi-Matrix $\mathbf{J}$,

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \phi^+(E(\mathbf{q}_1))}{\partial q_{1x}} & \frac{\partial \phi^+(E(\mathbf{q}_2))}{\partial q_{2x}} & \cdots & \frac{\partial \phi^+(E(\mathbf{q}_m))}{\partial q_{mx}} \\ \frac{\partial \phi^+(E(\mathbf{q}_1))}{\partial q_{1y}} & \frac{\partial \phi^+(E(\mathbf{q}_2))}{\partial q_{2y}} & \cdots & \frac{\partial \phi^+(E(\mathbf{q}_m))}{\partial q_{my}} \\ \frac{\partial \phi^+(E(\mathbf{q}_1))}{\partial q_{1z}} & \frac{\partial \phi^+(E(\mathbf{q}_2))}{\partial q_{2z}} & \cdots & \frac{\partial \phi^+(E(\mathbf{q}_m))}{\partial q_{mz}} \end{pmatrix} \quad (3.44)$$

dem Restglied $\boldsymbol{\rho}$ und dem Erwartungswert $E(\mathbf{q})$ der 2-D-Bildpunkte des zu rekonstruierenden 3-D-Punkts $\mathbf{p}_w$.

Aus der Unsicherheit der Schätzung der 2-D-Bildpunkte $\Sigma_{\mathbf{q}}$ lässt sich somit nun die Unsicherheit des 3-D-Punkts $\Sigma_{\mathbf{p}_w}$ berechnen:

$$\Sigma_{\mathbf{p}_w} = \mathbf{J}(\phi^+(E(\mathbf{q}))) \Sigma_{\mathbf{q}} \mathbf{J}(\phi^+(E(\mathbf{q})))^T. \quad (3.45)$$

Deshalb ist nun ein Fehlermodell für die Bestimmung der 2-D-Bildpunkte notwendig. Dies lässt sich über das Doppelverhältnis⁵ herleiten, das für die eindimensionale projektive Gerade $\mathbb{P}^1$ bzw. die zweidimensionale projektive Ebene $\mathbb{P}^2$ wie folgt definiert ist [Fau93, Seiten 14 und 17]:

⁵engl. cross-ratio

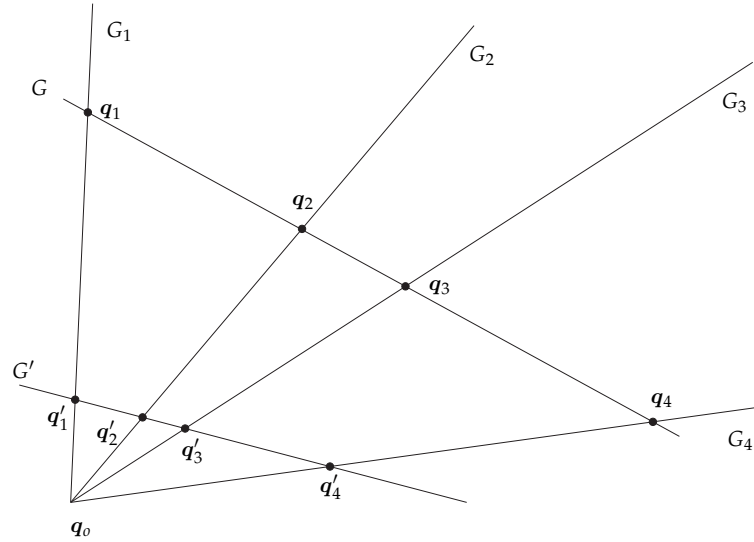


Bild 3.14 Das Doppelverhältnis der 4 Geraden $[G_1, G_2; G_3, G_4]$ ist das Doppelverhältnis der vier Punkte $[q_1, q_2; q_3, q_4]$, die beim Schnitt mit der Geraden G entstehen. Das Doppelverhältnis ist unabhängig von der Wahl der Geraden G , d. h. durch die Gerade G' wird das Doppelverhältnis $[q'_1, q'_2; q'_3, q'_4]$ festgelegt. Dabei gilt die Gleichheit $[q_1, q_2; q_3, q_4] = [q'_1, q'_2; q'_3, q'_4]$.

Def: 3.5 (Doppelverhältnis im $\mathbb{P}^1$)

Gegeben seien vier kollineare Punkte $q_1, q_2, q_3, q_4 \in \mathbb{P}^1$. Dann heißt

$$[q_1, q_2; q_3, q_4] := \frac{\|q_1 - q_3\| \|q_2 - q_4\|}{\|q_1 - q_4\| \|q_2 - q_3\|} \quad (3.46)$$

das Doppelverhältnis von q_1, q_2, q_3 und q_4 .

Def: 3.6 (Doppelverhältnis im $\mathbb{P}^2$)

Gegeben sei ein Geradenbüschel mit 4 Geraden G_1, G_2, G_3, G_4 , die sich im Punkt q_0 treffen. Dann ist das Doppelverhältnis der Geraden $[G_1, G_2; G_3, G_4]$ gleich dem Doppelverhältnis der vier Punkte $[q_1, q_2; q_3, q_4]$, die durch den Schnitt mit der Geraden G entstehen. Die Gerade G schneidet den Punkt q_0 nicht, ist sonst aber beliebig (vgl. Bild 3.14).

Die für diese Anwendung entscheidende Eigenschaft des Doppelverhältnisses wird in folgendem Satz zusammengefasst:

Satz 3.7

Das Doppelverhältnis ist gegenüber projektiven Transformationen invariant.

Beweis: siehe [Fau93, Seiten 14 und 559 ff.] oder auch [Bla47, Seiten 33 ff.].

Dies lässt sich direkt auf das Problem der Schätzung der 2-D-Unsicherheit der Punkte übertragen: Vier kollineare 3-D-Weltpunkte $p_{w1}, \dots, p_{w4}$ liegen auf einer Geraden. Diese Gerade kann als Gerade im $\mathbb{P}^2$ aufgefasst werden. Wird die Gerade durch eine perspektivische Projektion in eine andere (Bild-)ebene projiziert, so ändert sich das Doppelverhältnis nicht, da die perspektivische Projektion eine projektive Transformation darstellt (vgl. Kapitel 2.1.4). In Bild 3.15 wird der Sachverhalt noch einmal verdeutlicht.

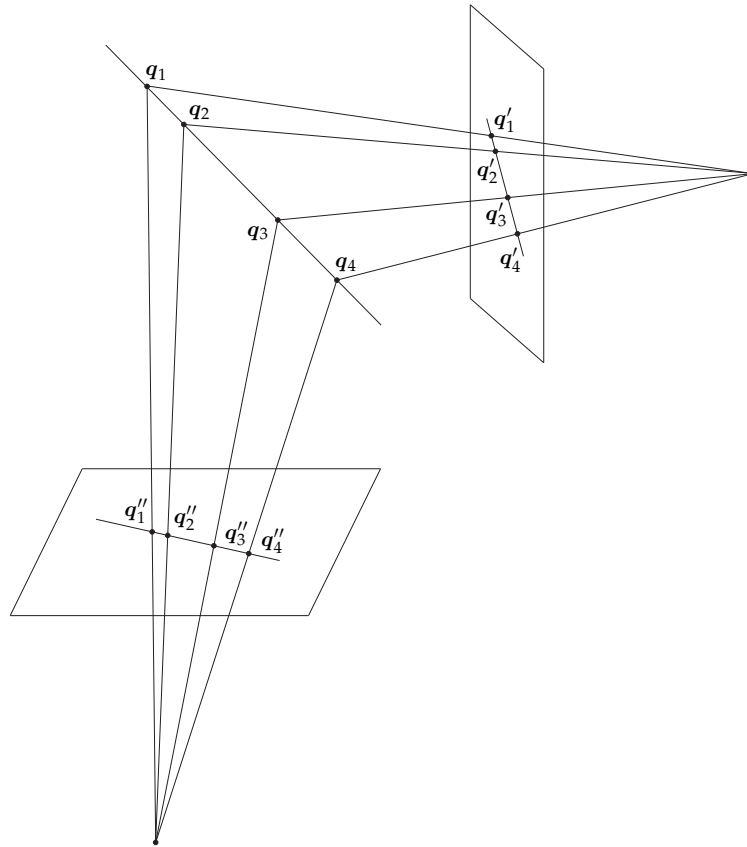


Bild 3.15 Das Doppelverhältnis ist invariant gegenüber projektiven Transformationen, d. h. hier gilt $[q_1, q_2; q_3, q_4] = [q'_1, q'_2; q'_3, q'_4] = [q''_1, q''_2; q''_3, q''_4]$.

Das Doppelverhältnis für vier kollineare Punkte ist also auf jeder Bildebene identisch und damit in jedem Bild, das diese vier Punkte enthält. In realen Experimenten wird das Doppelverhältnis jedoch durch Rauschen nicht stets gleich sein. In [Ola02] wird experimentell nachgewiesen, dass die Streuung des Doppelverhältnisses vom Blickwinkel abhängt. Dazu wurden vier äquidistante Punkte eines Kalibrieramusters untersucht. Bei Winkeln kleiner 30° ist der Fehler relativ klein, zwischen 30° und 60° wächst der Fehler mit steigendem Winkel stark an. Für Winkel größer als 80° konnte Olague und Mohr in ihrer Publikation [Ola02] keine Messungen mehr durchführen, sie nehmen jedoch an, dass der Fehler bei Annäherung an 90° gegen unendlich divergiert.

Aus der Streuung des Doppelverhältnisses kann die Streuung Σ_q geschätzt werden [Ola02]. Dies geschieht wiederum durch die Verwendung von Satz 3.3. Dazu wird die Berechnung des Doppelverhältnisses als Funktion aufgefasst, die dann durch eine Taylor-Reihen-Approximation erster Ordnung zur Berechnung der Kovarianz verwendet werden kann.

Die Streuung in Abhängigkeit des Blickwinkels wird dann durch eine modifizierte Hyperbolicus-Funktion approximiert. Damit ist ein direkter Zusammenhang zwischen Blickwinkel auf eine Ebene und der daraus resultierenden Unsicherheit der 3-D-Rekonstruktion dieser Punkte in Form ihrer Kovarianz hergestellt.

Um nun Ansichten planen zu können, wird ein Gütemaß benötigt, welches auf der Kovarianzmatrix Σ_{p_w} beruht. In [Ola02] werden hier diverse Definitionen für Metriken vorgestellt:

- Der maximale Eigenwert der Kovarianz Σ_{p_w} wird betrachtet.
- Die Spur der Kovarianz Σ_{p_w} wird betrachtet.
- Der maximale Wert in der Diagonalen der Kovarianz Σ_{p_w} wird betrachtet.

Tatsächlich verwendet wird in [Ola02] der maximale Wert in der Diagonalen der Kovarianz Σ_{p_w} .

Nachdem das Kriterium definiert ist, muss dieses mit einem geeigneten Algorithmus optimiert werden. In [Ola02] wird hierzu ein genetischer Algorithmus [Gol89, Koz92, Kin94] vorgeschlagen.

Für die Optimierung wird angenommen, dass die Kamerapositionen sich auf einer Halbkugel befinden, mit dem zu rekonstruierenden Objekt in der Mitte. Dadurch lässt sich eine Kameraposition durch zwei Raumwinkel beschreiben. Die Anzahl der Kameras wird als konstant angenommen. Der genetische Algorithmus arbeitet dann wie folgt:

1. Eine zufällige initiale Population von Kamerapositionen wird erzeugt.
2. Jedes Mitglied der Population wird mit dem oben definierten Gütekriterium evaluiert.
3. Es werden „gute“ Mitglieder mit Hilfe der „tournament selection“ [Bli95] selektiert.
4. Jeweils zwei der verbleibenden Mitglieder werden mit einer Wahrscheinlichkeit von $p_k = 0,7$ gekreuzt. Zusätzlich findet mit der Wahrscheinlichkeit $p_m = 0,005$ bei jeder Position eine Mutation statt. Diese Operationen erschaffen neue Mitglieder, die zur Population hinzugefügt werden.
5. Der Algorithmus wird ab Schritt 2 wiederholt, bis er gegen einen Lösung konvergiert.

Damit wird eine Lösung gefunden, die die mittlere Unsicherheit über alle 3-D-Punkte gemessen minimiert [Ola02].

Der oben vorgestellte Algorithmus hat mit dem probabilistischen Ansatz aus dieser Arbeit (der in Kapitel 4 vorgestellt wird) einige Ideen gemeinsam, aber natürlich auch Unterschiede:

Zu den Gemeinsamkeiten gehört, dass die 3-D-Struktur durch eine Punktemenge von 3-D-Punkten repräsentiert wird. Die Koordinaten eines einzelnen Punkts werden dabei durch eine Normalverteilung repräsentiert. Für die Schätzung der Parameter der Normalverteilung für die 3-D-Punkte werden zwei Annahmen gemacht: Zum einen unterliegen die Bildpunktkoordinaten ebenfalls normalverteiltem Rauschen, zum anderen haben die Kameraparameter einen Einfluss auf die Parameter der Normalverteilung des rekonstruierten Punkts im Sinne der Fehlerfortpflanzung. Diese Fehlerfortpflanzung wird auch im Kapitel 4 verwendet.

Die beiden Algorithmen unterscheiden sich jedoch in folgenden Punkten: Im Ansatz von Olague und Mohr [Ola97, Ola98, Ola02] wird mittels Doppelverhältnis die Kovarianz des Rauschens in den 2-D-Koordinaten bestimmt. Dies ist in dem Algorithmus aus Kapitel 4 nicht möglich, da für das Doppelverhältnis vier kollineare Punkte benötigt werden. Dort wird jedoch ein Punktverfolger verwendet, der gut zu verfolgende Punkte auf der Oberfläche des zu rekonstruierenden Objekts sucht. Eine Kollinearität zwischen diesen Punkten lässt sich im Allgemeinen nicht erzwingen. In Kapitel 4 wird deshalb davon ausgegangen, dass die Kovarianz eines 2-D-Punkts stets konstant ist.

Ein weiterer Unterschied ist die Vorgehensweise bei der Optimierung. Olague und Mohr gehen davon aus, dass m Kameras zur Verfügung stehen, deren Positionen simultan optimiert werden, um die Lage der Marker möglichst gut vermessen zu können. Im Gegensatz dazu arbeitet der Algorithmus aus Kapitel 4 iterativ, d. h. es werden nicht die Positionen aller Kameras gleichzeitig optimiert, sondern es wird zunächst aus der bisherigen Schätzung der 3-D-Koordinaten die neue Ansicht ermittelt, die ein Gütemaß (siehe Abschnitt 4.3) optimiert. Von dieser optimalen Ansicht wird eine Aufnahme gemacht und in die Schätzung integriert. Der Algorithmus wird schließlich wiederholt, bis ein Abbruchkriterium erreicht ist.

Durch das iterative Vorgehen müssen darüber hinaus keine a priori Annahmen über die Objektgeometrie gemacht werden. Bei dem Ansatz von Olague und Mohr sind die Ebenen, wo die Marker angebracht werden, bereits grob bekannt. Nur so lässt sich dann der Zusammenhang zwischen Blickwinkel auf die Ebene und dem Fehler bei der Lokalisierung der zweidimensionalen Bildpunkte ausnutzen. Der in Kapitel 4 vorgestellte Algorithmus basiert zunächst nur auf Punktwolken. Dadurch sind keine Informationen über die Normalen der Objektoberfläche bekannt. Somit kann kein Blickwinkel auf die Ebene berechnet werden.

Der Algorithmus aus Kapitel 4 hat gegenüber dem von Olague und Mohr teilweise mit schwierigeren Ausgangsbedingungen klar zu kommen. Zum einen wird durch die Verwendung eines Punktverfolgers, wie er in Abschnitt 2.5.1 beschrieben ist, das Problem der Ansichtenplanung zur 3-D-Rekonstruktion dadurch erschwert, dass während der Punktverfolgung Punkte nicht immer verfolgt werden können, weil sie z. B. außerhalb des Sichtbereichs der Kamera liegen. Das Wiederfinden solcher Punkte ist dann ein aufwändiger Prozess (vgl. Kapitel 5) und ist auch nur teilweise erfolgreich, wie in den Experimenten in Kapitel 6 erläutert wird. Zum anderen kann die Unsicherheit in der Schätzung der 2-D-Punkte, die vom verwendeten Punktverfolger abhängt, nicht so einfach ermittelt werden, wie bei der Verwendung vordefinierter Marker. Deshalb wird die Kovarianz dort als konstant und damit unabhängig vom Blickwinkel angenommen. Dies ist natürlich eine schlechtere Approximation der Realität gegenüber der Blickwinkelabhängigkeit von Olague und Mohr. Wie der Blickwinkel geschätzt werden könnte, ist in Abschnitt 7.1 skizziert.

Kapitel 4

Probabilistische 3-D-Rekonstruktion aus einer Folge von Einzelbildern

In diesem Kapitel wird ein probabilistischer Ansatz für die Ansichtenplanung entwickelt. Dazu werden Ideen aus Arbeiten zur Ansichtenplanung für andere Fragestellungen, insbesondere der Objekterkennung [Den02, Dei05, Dei06a, Dei06b, Der06] und der Objektverfolgung [Zob04, Deu06], sowie die Vorarbeiten [Wen06, Wen07b, Der08] verwendet.

Der hier vorgestellte Ansatz unterscheidet sich von dem in Abschnitt 3.2 in mehreren Punkten: Zum einen wird hier kein Stereosystem betrachtet, sondern es werden iterativ Ansichten einer statischen Szene ausgewählt und in eine probabilistische Schätzung integriert. Die Anzahl der verwendeten Ansichten für eine 3-D-Rekonstruktion ist nicht auf zwei, wie im Stereosystem, begrenzt, sondern kann theoretisch beliebig groß werden. Für die Experimente in Abschnitt 6.3 und 6.4 wurden typischerweise zwischen 10 und 15 Aufnahmen verwendet. Zum anderen wird bei diesem Ansatz nicht direkt ein geometrisches Gütemaß optimiert, sondern die Kriterien zur Ansichtenauswahl basieren auf einer probabilistischen Schätzung der 3-D-Koordinaten. Diese Schätzung wird dann in jedem Iterationsschritt entsprechend eines Gütemaßes verbessert. Drei unterschiedliche Gütemaße werden hier vorgestellt und später in den Experimenten evaluiert.

Als Vorbereitung für den Algorithmus zur Ansichtenauswahl muss zunächst die 3-D-Rekonstruktion probabilistisch durchgeführt werden. Hierbei bieten sich Algorithmen zur Zustandsschätzung an. Auf unterschiedliche Zustandsschätzer für statische und dynamische Systeme wird in Abschnitt 4.1 eingegangen.

Der Abschnitt 4.2 befasst sich deshalb mit der Übertragung des Problems der 3-D-Rekonstruktion auf das einer Zustandsschätzung. Anschließend wird dieses Zustandsschätzproblem mit Hilfe des Kalman-Filters gelöst.

Damit ist die 3-D-Rekonstruktion als Zustandsschätzproblem formuliert. Die nächste beste Ansicht soll nun basierend auf dieser Zustandsschätzung gewählt werden. Dabei muss noch definiert werden, was die beste Ansicht auszeichnet. Dazu werden im Abschnitt 4.3 verschiedene Gütemaße vorgestellt, die Ansichten bewerten können. Die Ansicht mit der jeweils besten Bewertung ist die nächste beste Ansicht. Die Gütemaße basieren dabei auf der Kovarianzmatrix der probabilistischen Schätzung der 3-D-Rekonstruktion.

In Abschnitt 4.4 werden Nebenbedingungen formuliert, die bei der Planung der nächsten besten Ansicht zu berücksichtigen sind. Insgesamt drei Arten von Nebenbedingungen werden in dieser Arbeit berücksichtigt: das eingeschränkte Sichtfeld der Kamera, Selbstverdeckungen sowie Erreichbarkeit einer Ansicht durch einen Roboterarm (worauf aber erst im Abschnitt 5.1 eingegangen wird). Es wird gezeigt,

wie diese Nebenbedingungen in die Optimierungsaufgabe mit einbezogen werden können, so dass ein geschlossenes Optimierungskriterium zur Bestimmung der nächsten besten Ansicht hergeleitet werden kann.

4.1 Zustandsschätzung statischer und dynamischer Systeme

In der Literatur ist eine ganze Reihe von verschiedensten Zustandsschätzern bekannt [BS88, Kay93, Sta94, Den03a]. Ein Zustandsschätzer schätzt den Zustand z_τ eines Systems aufgrund von Beobachtungen o_τ des Zustands. Man unterscheidet dabei grundsätzlich zwei mögliche Ansätze [BS88]:

- Ein deterministischer Ansatz: Der Zustand ist unbekannt, besitzt aber einen festen Wert. Diese Vorgehensweise ist auch als deterministischer Fischer-Ansatz bekannt [Zob04].
- Ein probabilistischer Ansatz: Der Zustand lässt sich als Wahrscheinlichkeitsverteilung mit einer bestimmten Dichtefunktion auffassen. Diese Vorgehensweise ist auch als statistischer Bayes-Ansatz bekannt [BS95].

Im Folgenden wird ausschließlich der Bayes-Ansatz verwendet. Zunächst werden nun die elementaren Begriffe „Zustandsschätzung“ und „Beobachtung“ definiert:

Def: 4.1 (Zustandsschätzer)

Ein Zustandsschätzer zur Schätzung des Zustands eines Systems ist eine Funktion $\hat{z}_\tau(\langle \hat{z} \rangle_{\tau-1}, \langle o \rangle_\tau)$, die zum Zeitpunkt τ einen Schätzwert $\hat{z}_\tau$ des wahren Zustands z_τ liefert. Zur Berechnung werden die Folge der vorherigen Zustandsschätzwerte $\langle \hat{z} \rangle_{\tau-1} = \hat{z}_0, \hat{z}_1, \dots, \hat{z}_{\tau-1}$ und die Folge der Beobachtungen $\langle o \rangle_\tau = o_0, o_1, \dots, o_\tau$ verwendet.

Def: 4.2 (Beobachtung)

Die direkte oder indirekte Messung eines Zustands nennt man Beobachtung. Die Beobachtung lässt sich mit Hilfe der Beobachtungsfunktion b darstellen:

$$o_\tau = b(z_\tau, v_\tau). \quad (4.1)$$

Dabei charakterisiert v_τ das Beobachtungsrauschen.

Das Beobachtungsrauschen charakterisiert dabei Ungenauigkeiten in der Messung. Häufig ist der Zustand nur indirekt oder partiell messbar. In diesem Fall gilt für die Dimensionen von o und z : $\dim(o) < \dim(z)$. Zusätzlich muss in diesem Fall sichergestellt werden, dass das System vollständig beobachtbar ist. Das bedeutet, der Zustand muss durch eine oder mehrere Beobachtungen vollständig bestimmbar sein [Lun06]. Bei der 3-D-Rekonstruktion ist diese Bedingung erfüllt, sobald ein 3-D-Punkt von zwei unterschiedlichen Positionen aus beobachtet wurde und nicht die degenerierten Fälle aus Bild 1.5 oder Bild 1.6 auftreten.

Bei Zustandsschätzproblemen unterscheidet man die Zustandsschätzung von statischen oder dynamischen Systemen. Im ersten Fall ist der Zustand nicht von

der Zeit τ abhängig, d. h. es gilt $z_\tau = z$. Man beachte, dass dort die Zustandsschätzung u. U. zeitabhängig ist, wenn sukzessive weitere Beobachtungen gemacht werden, die die Schätzung verbessern.

Im Folgenden werden nun zunächst einige einfache Zustandsschätzer für statische Systeme erläutert. Dies dient dem tieferen Verständnis für die darauf folgenden Zustandsschätzer von dynamischen Systemen.

4.1.1 Zustandsschätzer für statische Systeme

4.1.1.1 Der Maximum-Likelihood-Schätzer

Der Maximum-Likelihood-Schätzer (kurz: ML-Schätzer) ist definiert als

$$\hat{z}_\tau = \underset{z}{\operatorname{argmax}} p(\langle o \rangle_\tau | z). \quad (4.2)$$

Hier wird die Likelihood-Funktion $p(\langle o \rangle_\tau | z)$ betrachtet. Anschaulich beschreibt die Likelihood-Funktion die Wahrscheinlichkeit, dass der Zustand z die Folge von Beobachtungen $\langle o \rangle_\tau$ generiert hat. Diese Wahrscheinlichkeit wird beim Maximum-Likelihood-Schätzer maximiert.

4.1.1.2 Der Maximum-A-Posteriori-Schätzer

Der Maximum-A-Posteriori-Schätzer (kurz: MAP-Schätzer) maximiert die a posteriori Wahrscheinlichkeit:

$$\hat{z} = \underset{z}{\operatorname{argmax}} p(z | \langle o \rangle_\tau). \quad (4.3)$$

Nach der Bayes-Formel [Rin03, Seite 183] lässt sich dies umformen zu¹:

$$\hat{z} = \underset{z}{\operatorname{argmax}} p(z | \langle o \rangle_\tau) = \underset{z}{\operatorname{argmax}} \frac{p(\langle o \rangle_\tau | z) p(z)}{p(\langle o \rangle_\tau)} = \underset{z}{\operatorname{argmax}} p(\langle o \rangle_\tau | z) p(z). \quad (4.4)$$

Der ML- und der MAP-Schätzer unterscheiden sich formal durch den Faktor der a priori Wahrscheinlichkeit $p(z)$. Im Falle einer gleichverteilten a priori Wahrscheinlichkeit $p(z)$ geht der MAP-Schätzer in den ML-Schätzer über.

4.1.1.3 Minimierung des mittleren quadratischen Fehlers

Ein wichtiger Schätzer ist der, der den mittleren quadratischen Schätzfehler minimiert. Der Schätzer ist unter den Namen MMSE-Schätzer (engl: minimum mean square error) oder auch Minimum-Varianz-Schätzer bekannt. Für die Zufallsvariable z mit der zugehörigen Dichte $p(z)$ ist der mittlere quadratische Schätzfehler definiert als

$$E((\hat{z}_\tau - z)^T (\hat{z}_\tau - z) | \langle o \rangle_\tau). \quad (4.5)$$

Gesucht wird der Schätzwert $\hat{z}_\tau$, der die erwartete Abweichung zwischen $\hat{z}_\tau$ und dem wahren Zustand z_τ minimiert, wenn die Folge von Beobachtungen $\langle o \rangle_\tau$ gegeben ist:

$$\hat{z}_\tau = \underset{\hat{z}_\tau}{\operatorname{argmin}} E((\hat{z}_\tau - z)^T (\hat{z}_\tau - z) | \langle o \rangle_\tau). \quad (4.6)$$

¹Man beachte: Der Nenner im 2. Schritt ist unabhängig von z und damit für die Maximierung unerheblich.

Durch Ableiten von (4.6) nach $\hat{z}_\tau$ und Nullsetzen erhält man den MMSE-Schätzer [Den03a]

$$\hat{z}_\tau = \int z p(z | \langle o \rangle_\tau) dz. \quad (4.7)$$

Der beste Schätzwert $\hat{z}_\tau$ ist also der Erwartungswert der a posteriori Dichte. Dieser Schätzer minimiert die Varianz. Die Minimierung der Varianz ist der Grundgedanke für das Kalman-Filter, welches den Zustand eines dynamischen Systems schätzt und im Anschluss vorgestellt wird.

4.1.2 Zustandsschätzer für dynamische Systeme

Betrachtet man dynamische Systeme, so ist der Zustand von der Zeit abhängig und nicht mehr konstant, im Gegensatz zu statischen Systemen. Es werden hier ausschließlich autonome zeitdiskrete dynamische Systeme betrachtet. Der Zustandsübergang von Zeitpunkt τ zum Zeitpunkt $\tau + 1$ lässt sich wie folgt definieren:

$$z_{\tau+1} = g_\tau(z_\tau, z_{\tau-1}, \dots, z_0, w_\tau). \quad (4.8)$$

Die Zustandsübergangsfunktion g_τ beschreibt, wie sich das System im zeitlichen Verlauf verhält. Der neue Zustand $z_{\tau+1}$ ist dabei von den vorherigen Zuständen sowie einem Zustandsübergangsrauschen w_τ abhängig. Die Zustandsübergangsfunktion muss für das dynamische System bekannt sein, um den Zustand des Systems schätzen zu können [Den03a, Seite 42].

Diese Arbeit beschränkt sich auf autonome zeitdiskrete dynamische Systeme, für die die sogenannte Markov-Eigenschaft erfüllt ist: Die Markov-Eigenschaft, auch als Gedächtnislosigkeit bekannt, besagt, dass für den Zustandsübergang zum Zeitpunkt $\tau + 1$ die Zustände vor dem Zustand zum Zeitpunkt τ nicht relevant sind, d. h. :

$$z_{\tau+1} = g_\tau(z_\tau, w_\tau). \quad (4.9)$$

Die Zustandsübergangsfunktion g_τ legt die Verbunddichte $p(z_{\tau+1} | z_\tau)$ fest. Die Verbunddichte lässt sich je nach Zustandsübergangsfunktion mehr oder weniger einfach bestimmen. Im Falle einer linearen Zustandsübergangsfunktion ist die Bestimmung der Verbunddichte besonders einfach.

Die Beobachtung im Falle dynamischer Systeme ist gegeben durch:

$$o_\tau = b_\tau(z_\tau, v_\tau). \quad (4.10)$$

Die Beobachtungsfunktion b_τ kann zeitabhängig sein. Dies ist z. B. dann der Fall, wenn ein 3-D-Punkt von einer Kamera beobachtet wird, die sich bewegt. Dann ist die Beobachtungsfunktion b_τ von den Kameraparametern abhängig, die sich zwischen zwei Aufnahmen ändern. Darüber hinaus ist v_τ das Beobachtungsrauschen. Das Beobachtungsrauschen ist stets unabhängig vom Zustandsübergangsrauschen, da die Beobachtung keinen Einfluss auf den Zustand und den Zustandsübergang hat.

Von Interesse ist wieder die a posteriori Zustandsschätzung zum Zeitpunkt $\tau + 1$, also nachdem die Beobachtung gemacht wurde. Zunächst gilt für die a priori Wahrscheinlichkeit und den Zustand $z_{\tau+1}$ [Den03a, Seite 44]:

$$p(z_{\tau+1} | \langle o \rangle_\tau) = \int_{z_\tau} p(z_{\tau+1} | z_\tau) p(z_\tau | \langle o \rangle_\tau) dz_\tau. \quad (4.11)$$

Mit Hilfe der Bayes-Formel lässt sich die a posteriori Wahrscheinlichkeit, ähnlich zum MAP-Schätzer (vgl. Abschnitt 4.1.1.2), bestimmen [Den03a]:

$$p(z_{\tau+1} | \langle o \rangle_{\tau+1}) = \frac{p(o_{\tau+1} | z_{\tau+1}) \int_{z_\tau} p(z_{\tau+1} | z_\tau) p(z_\tau | \langle o \rangle_\tau) dz_\tau}{p(o_{\tau+1})}. \quad (4.12)$$

Die Berechnung der a posteriori Wahrscheinlichkeit ist in praktischen Fällen häufig schwierig, da sich das Integral in Gleichung (4.12) im Allgemeinen nicht geschlossen lösen lässt [Zob04]. Aus diesem Grund werden nun im Folgenden Einschränkungen gemacht und die Lösung für diesen Spezialfall diskutiert.

4.1.2.1 Lineare dynamische Systeme

In linearen dynamischen Systemen gilt, dass sowohl der Zustandsübergang als auch die Beobachtungsfunktion linear sind. Zusätzlich wird angenommen, dass die Rauschprozesse w_τ und v_τ additiv sind. Für den Zustandsübergang gilt

$$z_{\tau+1} = g_\tau(z_\tau, w_\tau) = G_\tau z_\tau + w_\tau, \quad (4.13)$$

mit der Zustandsübergangsmatrix $G_\tau \in \mathbb{R}^{N \times N}$ und dem Zustandsübergangsrauschen $w_\tau \in \mathbb{R}^N$. Für die Beobachtungsfunktion gilt

$$o_\tau = b_\tau(z_\tau, v_\tau) = B_\tau z_\tau + v_\tau, \quad (4.14)$$

mit der Beobachtungsmatrix $B_\tau \in \mathbb{R}^{N \times M}$ und dem Beobachtungsrauschen v_τ .

Zusätzlich wird bei der Untersuchung von linearen dynamischen Systemen häufig die Annahme getroffen, dass die Rauschterme w_τ und v_τ durch ein mittelwertfreies normalverteiltes Rauschen mit Kovarianzmatrix Q_τ bzw. $\mathfrak{R}_\tau$ gegeben sind:

$$w_\tau \sim \mathcal{N}(0, Q_\tau), \quad v_\tau \sim \mathcal{N}(0, \mathfrak{R}_\tau). \quad (4.15)$$

Ist die initiale Zustandsschätzung z_0 normalverteilt, so sind sowohl alle folgenden Zustandsschätzungen z_τ als auch die Beobachtung o_τ normalverteilt. Dies ist eine direkte Folgerung aus der linearen Zustandsübergangs- bzw. Beobachtungsfunktion aus den Gleichungen (4.13) bzw. (4.14) und Satz 3.4. In diesem Fall spricht man von linearen dynamischen Gauß-Systemen.

Für genau diesen Fall existiert ein spezieller Zustandsschätzer mit dem Namen Kalman-Filter [Kal60], der nun im nächsten Abschnitt vorgestellt wird.

4.1.2.2 Das Kalman-Filter

Das Kalman-Filter [Kal60, Wel06] ist das optimale Filter für lineare dynamische Gauß-Systeme. Optimal heißt, es minimiert den erwarteten quadratischen Schätzfehler, d. h. der Schätzwert ist der MMSE-Schätzer (vgl. Abschnitt 4.1.1.3).

Das Kalman-Filter liefert für jeden Zeitschritt τ den Erwartungswert $\hat{z}_\tau = E(z_\tau)$ und die Kovarianz Σ_τ des aktuellen Zustands. Da, wie oben festgestellt, alle Zustände normalverteilt sind und Erwartungswert und Kovarianz eine hinreichende Statistik für eine Normalverteilung sind (vgl. Definition 3.1 und Satz 3.2), sind also alle statistischen Eigenschaften des geschätzten Zustands $\hat{z}_\tau$ festgelegt.

Das Kalman-Filter arbeitet in zwei Schritten, die zu jedem Beobachtungszeitpunkt durchgeführt werden:

Vorhersage Im Vorhersageschritt wird der aktuelle Zustand z_τ aus dem vorherigen $z_{\tau-1}$ mit Hilfe des Zustandsübergangsmodells (4.13) bestimmt. Da der geschätzte Zustand $\hat{z}_{\tau-1}$ normalverteilt ist mit $\mathcal{N}(\hat{z}_{\tau-1}, \Sigma_{\tau-1})$ gilt gem. Satz 3.3 und Satz 3.4 für Mittelwert und Kovarianz des geschätzten Zustands $\hat{z}_\tau^-$

$$\hat{z}_\tau^- = G_\tau \hat{z}_{\tau-1} \quad (4.16)$$

$$\Sigma_\tau^- = G_\tau^T \Sigma_{\tau-1} G_\tau + Q_\tau. \quad (4.17)$$

Das Superskript $-$ markiert, dass dies die a priori Größen sind, d. h. vor dem Auffrischungsschritt.

Auffrischung Der vorhergesagte Zustandsschätzwert $\hat{z}_\tau^-$ wird anhand einer Beobachtung des tatsächlichen Zustands z_τ durch folgende Gleichungen korrigiert:

$$\mathfrak{K}_\tau = \Sigma_\tau^- B_\tau^T (B_\tau \Sigma_\tau^- B_\tau^T + \mathfrak{R}_\tau)^{-1}, \quad (4.18)$$

$$\hat{z}_\tau = \hat{z}_\tau^- + \mathfrak{K}_\tau (\mathbf{o}_\tau - B_\tau \hat{z}_\tau^-), \quad (4.19)$$

$$\Sigma_\tau = (I - \mathfrak{K}_\tau B_\tau) \Sigma_\tau^-. \quad (4.20)$$

Hier bezeichnet I die Einheitsmatrix und $\mathfrak{K}_\tau$ die Kalman-Verstärkungsmatrix². Sie gewichtet bei der Berechnung des a posteriori Erwartungswerts die Abweichung zwischen der tatsächlichen Beobachtung $\mathbf{o}_\tau$ und der erwarteten Beobachtung $B_\tau \hat{z}_\tau^-$.

Zur Plausibilisierung lässt sich zeigen: Geht die Kovarianz des Beobachtungsrauschens gegen Null, die Messungen sind also weniger verrauscht, es gilt

$$\lim_{\|\mathfrak{R}\| \rightarrow 0} \mathfrak{K} = B^{-1}. \quad (4.21)$$

Dies gilt nur, falls B invertierbar ist, was sie im Falle eines nur partiell beobachtbaren Systems nicht sein kann (weil die Matrix dann nicht quadratisch ist). Damit würde in Gleichung (4.19) der a priori Zustand keine Rolle mehr spielen, der aktuelle Zustand ergäbe sich rein aus der Beobachtung.

Auf der anderen Seite, wenn die a priori Schätzung bereits sehr sicher ist, wird die Kalman-Verstärkungsmatrix sehr klein:

$$\lim_{\|\Sigma_\tau^-\| \rightarrow 0} \mathfrak{K} = 0. \quad (4.22)$$

Dies bedeutet, dass der Einfluss einer neuen Beobachtung sehr klein wird.

Auf die Herleitung der Formeln (4.18)-(4.20) wird an dieser Stelle verzichtet und auf einschlägige Literatur verwiesen [Kal60, BS88, Chu91, Den03a, Wel06].

Festzuhalten sind einige wichtige Eigenschaften:

1. Das Kalman-Filter ist erwartungstreu.
2. Es minimiert den erwarteten mittleren quadratischen Schätzfehler (MMSE-Schätzer).
3. Es ist das optimale Filter für lineare dynamische Gauß-Systeme in dem Sinne, dass es den erwarteten mittleren quadratischen Schätzfehler minimiert.
4. Für alle nichtlinearen dynamischen Gauß-Systeme ist es das optimale lineare Filter.

²engl. Kalman gain matrix

4.1.2.3 Das erweiterte Kalman-Filter

Möchte man die 3-D-Rekonstruktion als Zustandsschätzproblem formulieren, so wird man feststellen, dass die 3-D-Rekonstruktion nicht als lineares System aufgefasst werden kann, da die Beobachtungsfunktion $\mathbf{b}$ eine nichtlineare Funktion ist: Die Beobachtungsfunktion $\mathbf{b}$ realisiert die perspektivische Projektion (vgl. Abschnitt 2.1.4.1), und diese ist nichtlinear. In diesem Fall kann man sich mit dem erweiterten Kalman-Filter (kurz: EKF) behelfen.

Für das Kalman-Filter werden im nichtlinearen Fall die Zustandsübergangsfunktion $\mathbf{g}$ und die Beobachtungsfunktion $\mathbf{b}$ mit Hilfe einer Taylor-Reihe approximiert. Bricht man die Entwicklung nach dem ersten Ableitungsterm ab, so spricht man von einem EKF erster Ordnung, bricht man nach dem zweiten Term ab, so spricht man von einem EKF zweiter Ordnung [Den03a]. Die weiteren Ausführungen beschränken sich auf das EKF erster Ordnung. In diesem Fall ist die approximierten Zustandsübergangs- bzw. Beobachtungsfunktion in ihren Komponenten linear.

Die Formeln für den Vorhersage- und den Auffrischungsschritt unterscheiden sich nur geringfügig. Es gilt für den Vorhersageschritt:

$$\hat{\mathbf{z}}_{\tau}^{-} = \mathbf{g}_{\tau}(\hat{\mathbf{z}}_{\tau-1}, \mathbf{0}) \quad (4.23)$$

$$\Sigma_{\tau}^{-} = \mathbf{G}_{\tau}^T \Sigma_{\tau-1} \mathbf{G}_{\tau} + \mathbf{W}_{\tau} \mathbf{Q}_{\tau} \mathbf{W}_{\tau}^T, \quad (4.24)$$

mit den Jacobi-Matrizen³

$$\mathbf{G}_{\tau[i,j]} = \frac{\partial g_{\tau i}}{\partial z_j}(\hat{\mathbf{z}}_{\tau-1}, \mathbf{0}), \quad (4.25)$$

$$\mathbf{W}_{\tau[i,j]} = \frac{\partial g_{\tau i}}{\partial w_j}(\hat{\mathbf{z}}_{\tau-1}, \mathbf{0}). \quad (4.26)$$

Für die Auffrischung gelten folgende Formeln:

$$\mathbf{K}_{\tau} = \Sigma_{\tau}^{-} \mathbf{B}_{\tau}^T (\mathbf{B}_{\tau} \Sigma_{\tau}^{-} \mathbf{B}_{\tau}^T + \mathbf{V}_{\tau} \mathbf{R}_{\tau} \mathbf{V}_{\tau}^T)^{-1}, \quad (4.27)$$

$$\hat{\mathbf{z}}_{\tau} = \hat{\mathbf{z}}_{\tau}^{-} + \mathbf{K}_{\tau} (\mathbf{o}_{\tau} - \mathbf{b}_{\tau}(\hat{\mathbf{z}}_{\tau}^{-}, \mathbf{0})), \quad (4.28)$$

$$\Sigma_{\tau} = (\mathbf{I} - \mathbf{K}_{\tau} \mathbf{B}_{\tau}) \Sigma_{\tau}^{-}, \quad (4.29)$$

mit den Jacobi-Matrizen⁴

$$\mathbf{B}_{\tau[i,j]} = \frac{\partial b_{\tau i}}{\partial z_j}(\hat{\mathbf{z}}_{\tau-1}, \mathbf{0}), \quad (4.30)$$

$$\mathbf{V}_{\tau[i,j]} = \frac{\partial b_{\tau i}}{\partial v_j}(\hat{\mathbf{z}}_{\tau-1}, \mathbf{0}). \quad (4.31)$$

Im Falle eines additiven Zustandsübergangsrauschens wird die Jacobi-Matrix $\mathbf{W}$ zur Einheitsmatrix. Analog wird die Jacobi-Matrix $\mathbf{V}$ zur Einheitsmatrix, falls das Beobachtungsrauschen additiv ist.

³Die Jacobi-Matrix $\mathbf{G}$ im EKF entspricht der Zustandsübergangsmatrix beim Kalman-Filter (Abschnitt 4.1.2.2). Deshalb wurde kein neuer Bezeichner eingeführt.

⁴Die Jacobi-Matrix $\mathbf{B}$ im EKF entspricht der Beobachtungsmatrix beim Kalman-Filter (Abschnitt 4.1.2.2). Deshalb wurde kein neuer Bezeichner eingeführt.

4.1.2.4 Anmerkungen zum erweiterten Kalman-Filter

Man beachte, dass durch die Linearisierung der Zustandsübergangs- bzw. Beobachtungsfunktion die Normalverteilungseigenschaft der Zustandsgrößen erhalten bleibt. Dies ist im Allgemeinen bei nichtlinearen Abbildungen von Zufallsvektoren nicht gegeben. Dies lässt sich relativ einfach an einem Beispiel zeigen. Dazu hilft folgender Satz:

Satz 4.3

Die Dichte $r(y)$ des Quotienten zweier Zufallsvariablen $Y = X/Z$ mit der Verbunddichte $r(x,z)$ ist gegeben durch

$$r(y) = \int_{-\infty}^{\infty} |z| r(yz, z) dz. \quad (4.32)$$

Beweis: [Pap02, Seite 173]

Geht man von einem 2-D-Modell, wie in Abschnitt 3.2.1, mit Brennweite $f = 1$ aus, so ergibt sich die Wahrscheinlichkeitsdichte der Koordinate des projizierten Punkts aus dem Satz 4.3. Geht man von zwei unabhängigen Normalverteilungen der Koordinaten $p_{wx} \sim \mathcal{N}(0,1)$ und $p_{wz} \sim \mathcal{N}(0,1)$ aus, so ergibt sich nach einigen Umformungen [Pap02, Seite 207] die Dichte für den Quotienten

$$r(y) = \frac{1}{\pi} \cdot \frac{1}{y^2 + 1}. \quad (4.33)$$

Die Dichtefunktion aus (4.33) ist die Dichte einer Cauchy-Verteilung. Wichtig sind hierbei zwei Punkte:

- Die transformierte Verteilung von Y ist keine Normalverteilung mehr.
- Der Satz 3.3 für Erwartungswert und (Ko-)Varianz bei linearen Transformationen kann nicht angewandt werden. Vielmehr kann überhaupt kein ähnlicher Satz für nichtlineare Transformationen existieren, denn die im Beispiel berechnete Verteilung ist eine Cauchy-Verteilung, die bekanntlich weder Erwartungswert noch Varianz besitzt [Rin03, Seite 343].

Diese Punkte sollte man sich vor Augen halten, wenn man nichtlineare Transformationen (in dieser Arbeit also insbesondere die perspektivische Projektion aus Abschnitt 2.1.4.1) von Zufallsvariablen durch lineare Funktionen approximiert. In diesem Zusammenhang wurde schon in Abschnitt 2.1.4.3 darauf eingegangen, dass die Linearisierung der perspektivischen Projektion mittels einer Taylor-Reihe auf die paraperspektivische Projektion hinaus läuft.

Im folgenden Abschnitt wird die 3-D-Rekonstruktion als Zustandsschätzproblem formuliert. Dieses Zustandsschätzproblem wird mit Hilfe des erweiterten Kalman-Filters gelöst.

4.2 Integration neuer Ansichten mit dem erweiterten Kalman-Filter

Mit Hilfe des Kalman-Filters wird nun eine 3-D-Rekonstruktion berechnet. Dazu werden die Begriffe aus der Schätztheorie auf das Problem übertragen.

4.2.1 3-D-Rekonstruktion mit Hilfe des erweiterten Kalman-Filters

Die wichtigste Größe ist der Zustand z des Systems. Der Zustand besteht hier aus einer Liste der n rekonstruierten 3-D-Punkte p_{wi} ($i = 1, \dots, n$):

$$z = \begin{pmatrix} p_{w1} \\ p_{w2} \\ \vdots \\ p_{wn} \end{pmatrix} \in \mathbb{R}^{3n}. \quad (4.34)$$

Da in dieser Arbeit ausschließlich statische Szenen betrachtet werden, ist der Zustand über die Zeit konstant. Die Zustandsübergangsfunktion ist daher die Identität: $G = I$. Deshalb entfällt beim Zustand z der Index τ , nicht jedoch bei der Zustandsschätzung $\hat{z}_\tau$, die ja mit jeder Beobachtung aufgefrischt wird. Da kein Zustandsübergang stattfindet, kann dort kein Zustandsübergangsrauschen auftreten. Die Kovarianz des Zustandsübergangsrauschens ist daher Null: $Q = 0$.

Damit wird der Vorhersageschritt im Kalman-Filter irrelevant. Es bleiben also noch die Größen für die Auffrischung zu klären: Man kann die 2-D-Koordinaten q_i der 3-D-Punkte p_{wi} direkt beobachten. Da der Zustand aus den Weltkoordinaten von 3-D-Punkten besteht, muss die Beobachtung, in diesem Fall die 2-D-Koordinaten in der Bildebene, dem jeweiligen Weltpunkt zugeordnet werden; d. h. man benötigt Punktkorrespondenzen. In dieser Arbeit wird die Berechnung der Punktkorrespondenzen durch einen Punktverfolger (vgl. Abschnitt 2.5.1) realisiert. Prinzipiell lässt sich der Algorithmus auch mit Punktkorrespondenzen aus anderen Quellen (a priori Wissen, manuelle Auswahl, Verwendung von Markern, etc.) anwenden. Der Punktverfolger liefert in mehreren Bildern die Punktkoordinaten der 2-D-Punkte, die von einem 3-D-Weltpunkt stammen. Hat man m Bilder, also Beobachtungen, so sind die 2-D-Koordinaten $q_{i\tau}$, $\tau = 1, \dots, m$ eines verfolgten 3-D-Punkts p_{wi} die Beobachtungen $o_{i\tau}$ dieses Punkts.

Die Beobachtung in einem Zeitschritt lässt sich damit als Liste oder Vektor von 2-D-Punktkoordinaten schreiben:

$$o_\tau = \begin{pmatrix} q_{1\tau} \\ q_{2\tau} \\ \vdots \\ q_{n\tau} \end{pmatrix} \in \mathbb{R}^{2n}. \quad (4.35)$$

Die Beobachtung o_τ entsteht dabei durch die perspektivische Projektion der 3-D-Weltpunkte. Wie bereits angemerkt, ist die perspektivische Projektion eine nicht-lineare Abbildung. Deshalb wird das erweiterte Kalman-Filter verwendet. Dazu wird die Jacobi-Matrix B_τ der Beobachtungsfunktion b_τ (die sich aus Gleichung (2.1) durch Umrechnung in kartesische Koordinaten ergibt) berechnet, in der die Ableitungen nach dem Zustand z_τ , also nach den 3-D-Koordinaten p_w , stehen.

Das Beobachtungsrauschen modelliert die Unsicherheit bei der Bestimmung der 2-D-Koordinaten. Im realen Experiment wird dieses Rauschen im Wesentlichen durch die Ungenauigkeit des eingesetzten Punktverfolgers (vgl. Kapitel 5) verursacht. Da, wie im Abschnitt 3.3 erläutert, keine direkte Messung der Unsicherheit möglich ist, wird die Kovarianzmatrix des Beobachtungsrauschens durch eine Diagonalmatrix mit $\mathfrak{R} = \sigma_o^2 I$ mit gegebenem σ_o^2 approximiert. Ein geeignetes σ_o^2 ergibt

sich durch die Abschätzung, wie exakt der Punktverfolger die Pixelkoordinaten im Bild schätzt. Aufgrund der Erfahrungen mit dem hier verwendeten Punktverfolger wird für die Experimente $\sigma_o^2 = 8 \text{ Pixel}^2$ gesetzt.

Damit sind alle relevanten Größen des Kalman-Filters festgelegt und es kann eine 3-D-Rekonstruktion durchgeführt werden: Bei der Rekonstruktion werden die Beobachtungen der 2-D-Bildpunkte, die während der Kamerabewegung entstehen, mit Hilfe der Kalman-Filter-Gleichungen (4.18)-(4.20) zur Verbesserung der 3-D-Schätzung verwendet. Eine detaillierte Beschreibung der einzelnen Schritte des Algorithmus findet sich in Abschnitt 5.3

Im nächsten Abschnitt wird auf spezielle Eigenschaften des Kalman-Filters hingewiesen, die sich im Fall der 3-D-Rekonstruktion anwenden lassen. Dadurch kann zum einen die Berechnungskomplexität verringert werden, zum anderen ermöglicht die Modifikation eine elegante Behandlung der Nebenbedingungen der Sichtbarkeit, in Abschnitt 4.4.

4.2.2 Das sequentielle Kalman-Filter und eine weitere Modifikation

Laut [Chu91, Kapitel 7] ist der rechenintensivste Teil des Kalman-Filters die Berechnung der Kalman-Verstärkungsmatrix $\mathbf{K}$, da dort eine $2n \times 2n$ Matrix invertiert wird. Bei einer 3-D-Rekonstruktion, wie sie hier durchgeführt wird, ist die Anzahl der Punkte n am Anfang der Rekonstruktion typischerweise in der Größenordnung zwischen 60 und 120. Während der Rekonstruktion können weitere Punkte gefunden werden, z. B. wenn die Kamera eine Seite des Objekts betrachtet, die vorher noch nicht beobachtet wurde. So können zum Ende der Rekonstruktion leicht 400 Punkte oder mehr rekonstruiert werden. Es müssen also bei jeder Auffrischung relativ große Matrizen invertiert werden.

Ein erster Schritt zur Reduktion der Dimension ist die Verwendung des sequentiellen Kalman-Filters. Bedingung für dessen Anwendung ist, dass das Beobachtungsrauschen unkorreliert ist, d. h. die Kovarianz des Beobachtungsrauschens $\mathfrak{R}$ ist eine Diagonalmatrix [Chu91]. In diesem Fall lässt sich jede einzelne Komponente aus dem Beobachtungsvektor $\mathbf{o}_\tau$ in einem eigenen Schritt zur Auffrischung des Zustands verwenden. Die Beobachtungsfunktion $\mathbf{b}_\tau$ ist in diesem Fall eine Abbildung $\mathbb{R}^N \rightarrow \mathbb{R}$. Damit entspricht die Jacobi-Matrix $\mathbf{B}_\tau$ dem transponierten Gradienten der Beobachtungsfunktion:

$$\mathbf{B}_\tau = (\nabla \mathbf{b}_\tau)^T. \quad (4.36)$$

Die Dimension der Jacobi-Matrix ist somit $1 \times 3n$, und damit besteht der Term

$$(\mathbf{B}_\tau \boldsymbol{\Sigma}_\tau^- \mathbf{B}_\tau^T + \mathbf{V}_\tau \mathfrak{R}_\tau \mathbf{V}_\tau^T) \quad (4.37)$$

aus Gleichung (4.27) nur noch aus einem skalaren Wert, dessen Inverses schnell berechnet werden kann.

Diese Vereinfachung wird in der Literatur als sequentielles Kalman-Filter bezeichnet und kann immer dann angewendet werden, wenn das Beobachtungsrauschen $\mathbf{v}$ eine diagonale Kovarianzmatrix $\mathfrak{R}_\tau$ besitzt. Das sequentielle Kalman-Filter reduziert die Komplexität der Berechnungen erheblich. Diese Vereinfachung basiert nicht auf einer Approximation, sondern es gilt: Das Ergebnis ist identisch mit dem des nichtsequentiellen Kalman-Filters.

Für die 3-D-Rekonstruktion kann noch eine weitere Vereinfachung gemacht werden. Das sequentielle Kalman-Filter erlaubt es, jede einzelne Komponente des Beobachtungsvektors $\mathbf{o}$ als eigene Beobachtung aufzufassen und somit sukzessive den Zustand mit der Beobachtung aufzufrischen. Man kann auch Teilbeobachtungen für die Auffrischung zusammenfassen und gemeinsam integrieren. Im Fall der 3-D-Rekonstruktion bietet es sich an, die x - und y -Komponente eines beobachteten 3-D-Punkts nicht einzeln zur Auffrischung zu verwenden, sondern die Auffrischung in einem Schritt durchzuführen.

Die Beobachtungsfunktion $\mathbf{b}_\tau : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ für den i -ten 3-D-Punkt, ist nur von drei Komponenten, der x -, y -, z - Koordinate dieses i -ten Punkts, abhängig. Das folgt unmittelbar, denn die Beobachtungsfunktion besteht aus der perspektivischen Projektion des 3-D-Punkts auf die Bildebene und diese ist natürlich unabhängig von der Lage der anderen 3-D-Punkte.

Daraus lassen sich zwei Dinge folgern: Zum einen sind in der Jacobi-Matrix $\mathbf{B}_\tau$ der Beobachtungsfunktion höchstens vier Einträge von Null verschieden. Zum anderen lässt sich für die Schätzung der Koordinaten eines 3-D-Punkts ein eigenes Kalman-Filter verwenden. Beide Eigenschaften werden nun bewiesen.

Für den Gradienten $\mathbf{B}_\tau$ der Beobachtungsfunktion $\mathbf{b}_\tau = (b_{\tau x} \ b_{\tau y})^T$ gilt:

$$\mathbf{B}_\tau^T = \begin{pmatrix} \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{w1x}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{w1x}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{w1y}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{w1y}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{w1z}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{w1z}} \\ \vdots & \vdots \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wix}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wix}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wiy}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wiy}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wiz}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wiz}} \\ \vdots & \vdots \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wnx}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wnx}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wny}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wny}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wnz}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wnz}} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wix}} & 0 \\ 0 & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wiy}} \\ \frac{\partial b_{\tau x}(\mathbf{p}_{wi})}{\partial p_{wiz}} & \frac{\partial b_{\tau y}(\mathbf{p}_{wi})}{\partial p_{wiz}} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} =: \begin{pmatrix} 0 \\ \vdots \\ \mathbf{B}_{\tau i}^{*T} \\ \vdots \\ 0 \end{pmatrix}.$$

Dabei bezeichnet p_{wix} die x -Komponente des i -ten 3-D-Punkts in Weltkoordinaten, p_{wiy} bzw. p_{wiz} analog dessen y - bzw. z -Koordinate und $b_{\tau x}$ bzw. $b_{\tau y}$ steht für die x - bzw. y -Koordinate der Beobachtungsfunktion zum Zeitpunkt τ . Der Block der partiellen Ableitungen, die nicht Null sind, wird als $\mathbf{B}_{\tau i}^*$ definiert.

An dieser Stelle wird, neben der Bedingung einer diagonalen Beobachtungsmatrix, noch eine zweite Annahme gemacht: Es wird davon ausgegangen, dass die initiale Kovarianzmatrix des Zustands Σ_0 eine Blockdiagonalstruktur mit 3×3

Blöcken besitzt. Diese Bedingung lässt sich folgendermaßen interpretieren: Die initiale Schätzung und damit die Kovarianz eines 3-D-Punkts hängt ausschließlich von den Beobachtungen dieses 3-D-Punkts ab und nicht von den Beobachtungen der anderen Punkte. Fasst man die Koordinaten der 3-D-Punkte als Zufallsvariablen auf, so sind die drei Koordinaten eines 3-D-Punkts im Allgemeinen miteinander korreliert. Die Schätzung eines 3-D-Punkts hängt nicht von den Beobachtungen der anderen ab. So ist der Zufallsvektor, der aus den Koordinaten eines 3-D-Punkts besteht, unabhängig von den Zufallsvektoren, die aus den Koordinatentripel der anderen 3-D-Punkte bestehen. Somit ergibt sich für die Kovarianzmatrix des initialen Zustands Σ_0 eine Blockdiagonalstruktur. Eine initiale Schätzung des Zustands kann z. B. durch eine Triangulation (vgl. Abschnitt 2.5) geschätzt werden.

Damit ist der Induktionsanfang bewiesen. Nun folgt der Induktionsschluss, um zu zeigen, dass die Eigenschaft der Blockdiagonalität bei der Auffrischung vom Zeitpunkt τ nach $\tau + 1$ erhalten bleibt (unter der Voraussetzung, dass die Beobachtungen unkorreliert sind). Es wird angenommen, dass die x - und die y -Beobachtung zeitgleich zur Auffrischung verwendet werden. Es gilt folgende Konklusion:

Sei die Kovarianzmatrix Σ_τ des Zustands z_τ zum Zeitpunkt τ eine Blockdiagonalmatrix mit der Struktur

$$\Sigma_\tau = \begin{pmatrix} \Sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \Sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix}, \quad (4.38)$$

mit den 3×3 Kovarianzmatrizen $\Sigma_{\tau,i}$, die die Unsicherheit des i -ten Punkts in seinen drei Koordinaten charakterisiert. Es gilt für die Kalman-Verstärkungsmatrix:

$$\begin{aligned} \mathfrak{K} &= \begin{pmatrix} \Sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \Sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ B_{\tau i}^{*T} \\ \vdots \\ 0 \end{pmatrix} \\ &= \left(\begin{pmatrix} 0 & \dots & B_{\tau i}^* & \dots & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \Sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ B_{\tau i}^{*T} \\ \vdots \\ 0 \end{pmatrix} + \mathfrak{R} \right)^{-1} \\ &= \begin{pmatrix} 0 & \dots & \left(\frac{(\Sigma_{\tau,i} B_{\tau i}^{*T})}{B_{\tau i}^* \Sigma_{\tau,i} B_{\tau i}^{*T} + \mathfrak{R}} \right)^T & \dots & 0 \end{pmatrix}^T. \end{aligned} \quad (4.39)$$

Die Bruchschreibweise in der Gleichung wurde aus Platzgründen verwendet und bedeutet eine Multiplikation des Zählers mit dem Inversen des Nenners von rechts.

Für die a posteriori Kovarianz $\Sigma_{\tau+1}^{(i)}$ gilt nach der Auffrischung mit der Beobachtung des i -ten Punkts:

$$\begin{aligned} \Sigma_{\tau+1}^{(i)} &= \left(I - \begin{pmatrix} 0 \\ \vdots \\ \frac{(\Sigma_{\tau,i} B_{\tau i}^{*T})}{B_{\tau i}^* \Sigma_{\tau,i} B_{\tau i}^{*T} + \mathfrak{R}} \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 0 & \dots & B_{\tau i}^* & \dots & 0 \end{pmatrix} \right) \\ &\quad \begin{pmatrix} \Sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \Sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix} = \\ &\quad \left(I - \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{(\Sigma_{\tau,i} B_{\tau i}^{*T})}{B_{\tau i}^* \Sigma_{\tau,i} B_{\tau i}^{*T} + \mathfrak{R}} B_{\tau i}^* & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{\tau,1} & 0 & \dots & 0 \\ 0 & \Sigma_{\tau,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix} \right) = \\ &\quad \begin{pmatrix} \Sigma_{\tau,1} & 0 & 0 & \dots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & \dots & \Sigma_{\tau+1,i} & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Sigma_{\tau,n} \end{pmatrix}. \end{aligned} \quad (4.40)$$

Die Blockdiagonalstruktur bleibt also erhalten, wenn die Beobachtung des i -ten Punkts in die Schätzung integriert wird. Da über $i = 1, \dots, n$ keine Annahmen gemacht wurden, bleibt die Blockdiagonalstruktur für jede Auffrischung mittels der Beobachtung eines 3-D-Punkts erhalten. Damit ist per Induktion bewiesen, dass die Blockdiagonalstruktur erhalten bleibt, falls die initiale Kovarianz Σ_0 als Blockdiagonalmatrix initialisiert wurde. Dass diese Annahme sinnvoll ist, wurde oben bereits erläutert.

Ferner lässt sich das Kalman-Filter für eine 3-D-Rekonstruktion von n Punkten als n Kalman-Filter für jeweils einen 3-D-Punkt auffassen. Auch in diesem Fall gilt: Das Ergebnis der n Kalman-Filter liefert exakt die gleichen Ergebnisse wie ein erweitertes Kalman-Filter, dass alle Punkte gleichzeitig schätzt. Außer den Annahmen „unkorreliertes Beobachtungsrauschen“ und „Blockdiagonalstruktur der initialen Kovarianz Σ_0 “ sind keine weiteren Annahmen oder Approximationen notwendig, um diese Behauptung zu beweisen.

Die bisherigen Erkenntnisse lassen sich nun zusammenfassen: Das Problem der 3-D-Rekonstruktion lässt sich als Zustandsschätzproblem formulieren und auf das Kalman-Filter übertragen. Man erhält dabei die 3-D-Koordinaten der rekonstruierten Punkte als Wahrscheinlichkeitsverteilung, und zwar als eine Normalverteilung. Dabei werden die Parameter Erwartungswert und Kovarianz geschätzt. Da

diese eine hinreichende Statistik für die Normalverteilung bilden (vgl. Satz 3.2), ist damit die Dichte der Normalverteilung vollständig beschrieben.

Durch die Verwendung des sequentiellen Kalman-Filters ist es weiterhin möglich, die Berechnungskomplexität deutlich zu reduzieren. Einzige Annahme hierfür ist, dass das Beobachtungsrauschen unkorreliert ist, d. h. die Beobachtungsrauschmatrix $\mathfrak{R}$ muss eine Diagonalmatrix sein. Dadurch kann die Matrixinversion einer $3n \times 3n$ Matrix durch n Inversionen einer 2×2 Matrix ersetzt werden. Für 2×2 Matrizen existieren darüber hinaus relativ einfache Formeln für die Inversion.

Darüber hinaus ist es sinnvoll, die initiale Kovarianzmatrix des Zustands Σ_0 als Blockdiagonalmatrix zu definieren, da jede initiale Schätzung eines 3-D-Punkts unabhängig von der Schätzung der anderen 3-D-Punkte ist. Dadurch kann nicht nur die aufwändige Matrixinversion gespart werden, sondern man kann das Problem auf mehrere Kalman-Filter, die hintereinander geschaltet sind, übertragen. Diese Umrechnung spart nicht nur weitere Berechnungskomplexität, sondern erlaubt auch eine effiziente und einfache Speicherverwaltung. So lassen sich die entsprechenden Blöcke einzeln auffrischen, d. h. die Matrixmultiplikationen mit den dünn besetzten Matrizen lassen sich sehr effizient implementieren.

Zusätzlich ermöglicht diese Vereinfachung zum einen eine sehr elegante Weise, ein Optimierungskriterium auszuwerten [Wen06]. Zum anderen macht es die Berücksichtigung der Nebenbedingungen der Sichtbarkeit in Abschnitt 4.4 erst möglich. Auf diese beiden Punkte wird nun im Folgenden eingegangen. Verschiedene Optimierungskriterien für eine Ansichtenauswahl zur 3-D-Rekonstruktion werden im nächsten Abschnitt dargestellt.

4.3 Auswahlkriterien für neue Ansichten

Für die Ansichtenauswahl wird ein Gütemaß benötigt, welches es erlaubt, die Genauigkeit der Rekonstruktion zu verbessern. Die Genauigkeit ist umso größer, je kleiner der Abstand zwischen dem geschätzten und dem tatsächlichen Zustand ist:

$$z - \hat{z}_\tau. \quad (4.41)$$

Da der wahre Zustand des Systems z zum Einen nicht bekannt ist und zum Anderen als Dichte aufgefasst wird (Bayes-Ansatz), lässt sich die Güte nicht einfach als Abstand zwischen tatsächlichem Zustand z und geschätztem Zustand $\hat{z}_\tau$ berechnen. Stattdessen kann der erwartete quadratische Schätzfehler

$$E((z - \hat{z}_\tau)^T(z - \hat{z}_\tau)) \quad (4.42)$$

(vgl. Gleichung (4.5)) bestimmt werden. Die Kovarianzmatrix des Fehlers (4.41)

$$E((z - \hat{z}_\tau)(z - \hat{z}_\tau)^T). \quad (4.43)$$

ist identisch mit der a posteriori Kovarianz Σ_τ aus den Kalman-Filter-Gleichungen (4.27) und (4.29) [Wel06], d. h. :

$$\Sigma_\tau = E((z - \hat{z}_\tau)(z - \hat{z}_\tau)^T). \quad (4.44)$$

Die Spur der Kovarianzmatrix Σ_τ entspricht genau dem erwarteten quadratischen Schätzfehler (4.42), der minimiert werden soll.

Es ist nun in der Theorie des optimalen experimentellen Designs [Puk93] üblich, nicht nur die Spur, also direkt den erwarteten quadratischen Schätzfehler, zu betrachten, sondern allgemein die Kovarianzmatrix zu „minimieren“. Dabei gibt es verschiedene Möglichkeiten die „Größe“ der Kovarianz zu messen, von denen drei im Folgenden vorgestellt werden. Wichtig bleibt festzuhalten, dass eine Minimierung der Kovarianz im Sinne des jeweiligen Kriteriums immer auch die Spur der Matrix verkleinern sollte.

Aus den Gleichungen (4.27) und (4.29) ist zunächst eine entscheidende Eigenschaft abzulesen: Die a posteriori Kovarianz kann berechnet werden, ohne die konkreten Werte der Beobachtung $\mathbf{o}_\tau$ zu kennen. Die Kovarianz Σ_τ ist nur von der a priori Kovarianz $\Sigma_{\tau-1}$, von der Kovarianz des Beobachtungsrauschens $\mathfrak{R}$, und der Jacobi-Matrix der Beobachtungsfunktion $\mathbf{B}_\tau$ abhängig. Die a priori Kovarianz $\Sigma_{\tau-1}$ ist aus dem vorherigen Zeitschritt bekannt und die Kovarianz des Beobachtungsrauschens $\mathfrak{R}$ wird als bekannt und konstant angenommen mit $\mathfrak{R} = \sigma_o^2 \mathbf{I}$. Die einzige Größe, die in jedem Schritt zunächst nicht fest vorgegeben ist, ist also die Jacobi-Matrix der Beobachtungsmatrix $\mathbf{B}_\tau$. Diese hängt von der a priori Zustandsschätzung $\hat{\mathbf{z}}_{\tau-1}^-$ ab, die in dem statischen Fall der 3-D-Rekonstruktion gleich $\hat{\mathbf{z}}_{\tau-1}$ ist, und von den Kameraparametern.

Es bleibt festzuhalten: Die Kovarianzmatrix Σ_τ repräsentiert die Kovarianz des Fehlers der 3-D-Rekonstruktion, und die Spur der Kovarianz ist gleich dem erwarteten quadratischen Schätzfehler, der für eine möglichst genaue Rekonstruktion minimiert werden muss. Ferner haben die Kameraparameter einen direkten Einfluss auf die Jacobi-Matrix der Beobachtungsfunktion $\mathbf{B}_\tau$ und somit auf die Kovarianz Σ_τ . Die Frage ist daher: „Wie müssen die Kameraparameter gewählt werden, so dass die Unsicherheit in der Schätzung der 3-D-Rekonstruktion möglichst klein wird?“ Zwei Aspekte sind hier zu beachten:

Erstens, welche Kameraparameter können überhaupt (zielgerichtet) verändert werden? Die Position und Ausrichtung der Kamera lässt sich mithilfe des Roboterarms, auf den sie montiert ist, zielgerichtet verändern. Dadurch ist die Rotationsmatrix $\mathbf{R}$ und der Translationsvektor $\mathbf{t}$ für die Transformation von Welt- in Kamerakoordinaten festgelegt, wie in Definition und Satz 2.2. Bei den intrinsischen Parametern ist nur die Brennweite f direkt beeinflussbar. Die beeinflussbaren Parameter werden im Folgenden durch den Vektor $\mathbf{c}$ dargestellt, der im Allgemeinen die Rotationsparameter⁵, den Translationsvektor und die Brennweite enthält.

Zweitens, wie lässt sich die „Größe“ der Kovarianzmatrix Σ_τ messen, da im $\mathbb{R}^{3n \times 3n}$ keine Ordnungsrelation existiert? In der Literatur [Puk93, Wen07b] sind diverse Gütemaße bekannt, von denen drei in den nächsten Abschnitten vorgestellt werden.

4.3.1 Die Entropie als Gütemaß (D-Optimalität)

Die Einführung der D-Optimalität basiert auf informationstheoretischen Überlegungen: Eines der wichtigsten Maße in der Informationstheorie ist die Entropie H ,

⁵Es spielt hier keine Rolle, ob die Rotation durch Cardan-Winkel, Achse-Winkel-Darstellung oder Quaternionen repräsentiert wird. Details zur Umsetzung finden sich in Kapitel 5

die, wie unten gezeigt wird, direkt mit der Determinante, deshalb D-Optimalität, zusammenhängt. Die Entropie ist folgendermaßen definiert:

Def: 4.4 (Entropie)

Sei X eine Zufallsvariable mit Dichte $r(x)$. Dann heißt

$$H(X) := - \int r(x) \log r(x) dx \quad (4.45)$$

die Entropie von X .

Für die Entropie gilt

$$H(X) \geq 0. \quad (4.46)$$

Die Entropie ist ein Maß für den Informationsgehalt der bisher gesammelten Beobachtungen. Sie kann zusätzlich als „Unsicherheit“ der Schätzung aufgefasst werden. Die Unsicherheit einer Schätzung ist andererseits ein Maß, welches durch die Kovarianz ausgedrückt werden kann. Auf diesen Zusammenhang wird später noch eingegangen.

Ist eine Zufallsvariable X gleichverteilt, so ist die Unsicherheit am größten; sie nimmt den Wert Null an, wenn man sich über X absolut sicher ist, d. h. die Dichte $r(x)$ ist ein Dirac-Impuls.

Bei der Untersuchung von zwei Zufallsvariablen X und Y ist die sogenannte bedingte Entropie von Interesse:

Def: 4.5

Seien X und Y zwei Zufallsvariablen mit der Verbunddichte $r(X,Y)$, dann heißt

$$H(X|Y) := - \int r(y) \int r(x|y) \log r(x|y) dx dy \quad (4.47)$$

die bedingte Entropie von X gegeben Y .

Die bedingte Entropie ist der Erwartungswert der Entropie der bedingten Dichte $r(x|y)$. Die bedingte Entropie gibt die Unsicherheit an, die über X verbleibt, wenn die Zufallsvariable Y bekannt ist. Aus dieser Argumentation folgt, dass die Unsicherheit über X durch zusätzliche Information, also durch die Kenntnis von Y , nicht größer werden kann, d. h.

$$H(X|Y) \leq H(X), \quad (4.48)$$

wobei das Gleichheitszeichen genau dann gilt, wenn X und Y statistisch unabhängig sind.

Die bedingte Entropie kann auf das Problem der Integration neuer Ansichten mit dem Kalman-Filter angewendet werden:

$$H_{c_\tau}(z|o_\tau) = - \int r_{c_\tau}(o_\tau) \int r(z|\langle o \rangle_\tau) \log r(z|\langle o \rangle_\tau) dz do_\tau. \quad (4.49)$$

Dabei steht das Subskript c_τ für die gezielt änderbaren Parameter Rotation, Translation und Brennweite, welche die Dichte $r_{c_\tau}(o_\tau)$ direkt beeinflussen.

Die optimalen Kameraparameter c^* sind die, die die Unsicherheit am stärksten reduzieren [Den03b]:

$$c^*_\tau = \underset{c_\tau}{\operatorname{argmin}} H_{c_\tau}(z|o_\tau). \quad (4.50)$$

Die Berechnung des Integrals in Formel (4.49) ist im Allgemeinen schwierig, was die Auswertung des Optimierungskriteriums (4.50) ebenfalls erschwert [Den03b]. Da das Gütemaß (4.50) im Rahmen des Kalman-Filter-Ansatzes verwendet wird, sind sowohl die Beobachtungen als auch der Zustand normalverteilt. Dies vereinfacht das Problem erheblich: Für einen normalverteilten Zufallsvektor $\mathbf{X} \in \mathbb{R}^N$ mit Kovarianzmatrix Σ ist bekannt, dass sich seine Entropie geschlossen darstellen lässt [Cov91]:

$$H(\mathbf{X}) = \frac{N}{2} + \frac{1}{2} \log((2\pi)^N \det(\Sigma)). \quad (4.51)$$

Das innere Integral in (4.49) entspricht genau der a posteriori Entropie $H(z_\tau)$ und mit Gleichung (4.51) ergibt sich:

$$\int r(z|\langle \mathbf{o} \rangle_\tau) \log r(z|\langle \mathbf{o} \rangle_\tau) dz = H(z_\tau) = \frac{3n}{2} + \frac{1}{2} \log((2\pi)^{3n} \det(\Sigma_\tau)). \quad (4.52)$$

Man beachte: die Beobachtung $\mathbf{o}_\tau$ hat keinen Einfluss auf die a posteriori Entropie $H(z_\tau)$. Ferner können bei der Optimierung (4.50) konstante Terme vernachlässigt werden. Somit ergibt sich im konkreten Fall für das Gütemaß (4.50)

$$\mathbf{c}_\tau^* = \operatorname{argmin}_{\mathbf{c}_\tau} \int r_c(\mathbf{o}_\tau) \log(\det(\Sigma_\tau)) d\mathbf{o}_\tau. \quad (4.53)$$

Der Term $\log(\det(\Sigma_\tau))$ in (4.53) ist von $\mathbf{o}_\tau$ unabhängig und kann deshalb vor das Integral gezogen werden. Das verbleibende Integral integriert eine Wahrscheinlichkeitsdichte über den gesamten Integrationsbereich und ist somit gleich eins. Das Gütemaß (4.53) lautet damit

$$\mathbf{c}_\tau^* = \operatorname{argmin}_{\mathbf{c}_\tau} \log(\det(\Sigma_\tau)). \quad (4.54)$$

Durch die in Abschnitt 4.2.2 diskutierten Annahmen folgt, dass die Kovarianzmatrix des Zustands Σ_τ eine Blockdiagonalstruktur besitzt. Die Determinante einer Blockdiagonalmatrix ist gleich dem Produkt der Determinanten der Blöcke. Es ergibt sich damit das abschließende Optimierungskriterium:

$$\mathbf{c}_\tau^* = \operatorname{argmin}_{\mathbf{c}_\tau} \log \left(\prod_{i=1}^n \det(\Sigma_{\tau,i}) \right) = \operatorname{argmin}_{\mathbf{c}_\tau} \sum_{i=1}^n \log(\det(\Sigma_{\tau,i})). \quad (4.55)$$

Für eine Auswertung des Gütemaßes für ein konkretes $\mathbf{c}_\tau$ ist also nur die Summe der Determinanten der 3×3 Blöcke zu berechnen, da die Kovarianzmatrix ja eine Blockdiagonalmatrix mit 3×3 Blöcken ist (vgl. Abschnitt 4.2.2).

Das Gütemaß (4.55) minimiert die Entropie und maximiert somit den Informationsgewinn einer neuen Beobachtung. Da es alleine vom Logarithmus der Determinante abhängt, wird es auch als D-Optimalitätskriterium bezeichnet. In [Puk93] wird das D-Kriterium ohne Logarithmus betrachtet:

$$\mathbf{c}_\tau^{**} = \operatorname{argmin}_{\mathbf{c}_\tau} (\det(\Sigma_\tau)). \quad (4.56)$$

Da der Logarithmus eine monotone Funktion ist, existiert das Minimum an der gleichen Stelle. Im Abschnitt 4.4 wird deutlich, dass sich die Sichtbarkeitsbedingungen nur dann einfach in das Gütemaß integrieren lassen, wenn das Gütemaß (4.55) anstelle von (4.56) verwendet wird. Zuvor werden nun noch zwei weitere Gütemaße vorgestellt: die modifizierte E- und die T-Optimalität.

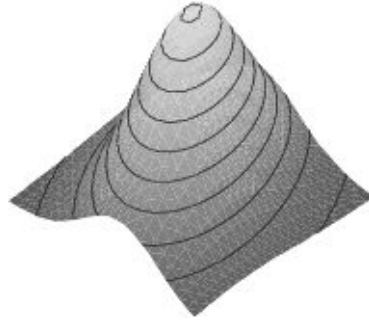


Bild 4.1 Die Höhenlinien einer 2-D-Normalverteilung sind Ellipsen.

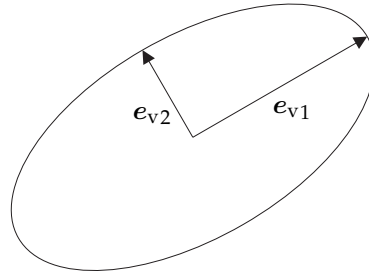


Bild 4.2 Die Kovarianzmatrix kann als Ellipse visualisiert werden. Die Eigenvektoren e_{v1} und e_{v2} definieren dabei die Richtung der Halbachsen, die Eigenwerte bestimmen die jeweilige Länge.

4.3.2 Modifizierte E-Optimalität

Die E-Optimalität nach Pukelsheim [Puk93] minimiert den größten Eigenwert der Kovarianzmatrix Σ_τ . Da die Kovarianz Σ_τ reell und positiv definit ist, existieren nur reelle Eigenwerte [Nip02]. Um dieses Gütemaß interpretieren zu können, wird zunächst nur ein einzelner 3-D-Punkt betrachtet. Anschließend lässt sich die E-Optimalität für das Problem der Ansichtenauswahl für die 3-D-Rekonstruktion sinnvoll modifizieren.

Die Höhenlinie im Graph einer 2-D-Normalverteilung sind Ellipsen (Bild 4.1). Die Halbachsen dieser Ellipsen, werden durch die Richtung der Eigenvektoren e_{v1} bzw. e_{v2} der Kovarianzmatrix Σ_τ festgelegt sind. Da die Kovarianz symmetrisch ist, ist sichergestellt, dass die beiden Eigenvektoren orthogonal zueinander stehen [Nip02]. Das Verhältnis der Längen der Halbachsen ist durch die zum jeweiligen Eigenvektor gehörenden Eigenwerte λ_1 bzw. λ_2 festgelegt. Ein Beispiel ist in Bild 4.2 dargestellt. Die Ellipse deren Halbachsen den Eigenwerten entsprechen heißt Kovarianzellipse.

Geht man von einem 3-D-Punkt aus, so ist $\Sigma_\tau \in \mathbb{R}^{3 \times 3}$. Im Dreidimensionalen lässt sich analog zur Kovarianzellipse ein Kovarianzellipsoid definieren. Die Eigenvektoren der Kovarianzmatrix entsprechen den Richtungen und die zugehörigen Eigenwerte den Längen der Halbachsen. Die E-Optimalität nach Pukelsheim minimiert den größten Eigenwert, d. h. bei der 3-D-Rekonstruktion wird die Unsicherheit genau in die Raumrichtung minimiert, in der sie aktuell am größten ist.

Betrachtet man mehr als einen Punkt, so lässt sich das E-Kriterium direkt anwenden: Es wird der maximale Eigenwert der gesamten Kovarianz Σ_τ verwendet. Dies erscheint für das Ziel einer 3-D-Rekonstruktion weniger sinnvoll, da die Optimierung dann einzig von dem 3-D-Punkt abhängt, dessen Unsicherheit am

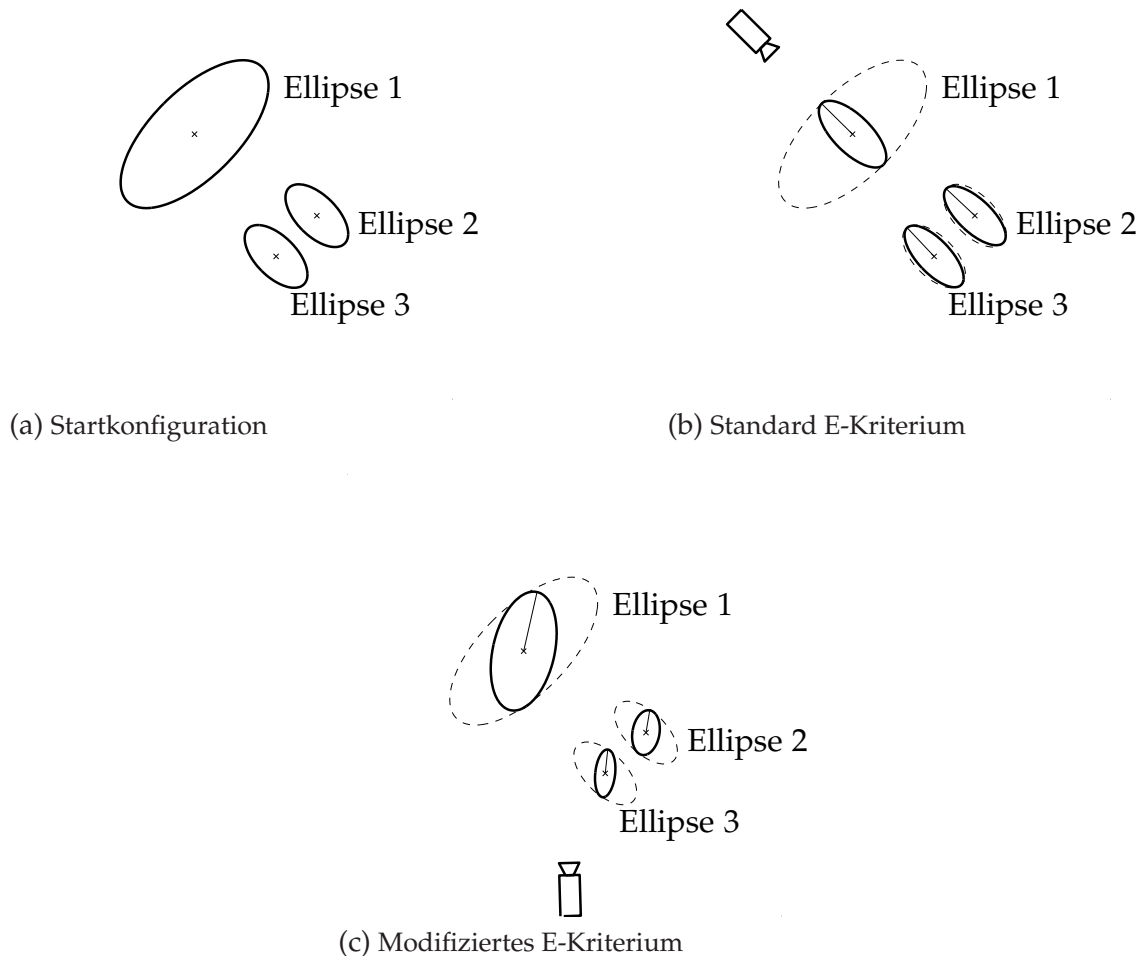


Bild 4.3 Die Unsicherheit in der Startkonfiguration (a) wird durch eine neue Aufnahme reduziert. Beim Standard E-Kriterium (b) ist ausschließlich der größte Eigenwert (hier bei Ellipse 1) ausschlaggebend. Die Unsicherheit der Ellipsen 2 und 3 wird dagegen kaum reduziert. Beim modifizierten E-Kriterium (c) wurde zwar die Unsicherheit der Ellipse 1 nicht ganz so stark reduziert, wie beim Standard E-Kriterium, dafür nimmt aber auch die Unsicherheit in der Ellipse 2 und 3 wesentlich ab. In den Bildern (b) und (c) wurde die jeweils optimale Kameraposition eingezeichnet (Annahme: Die Kamera bewegt sich auf einem Kreis um den Schwerpunkt der Erwartungswerte).

größten ist. Ein einziger Punkt hat somit exklusiven Einfluss auf die Planung der nächsten Ansicht. Dies hat u. U. zur Folge, dass nur die Schätzung der Koordinaten dieses Punkts wesentlich verbessert wird, während die Koordinatenschätzung anderer Punkte keine wesentliche Reduktion der Unsicherheit erfährt. Der Sachverhalt ist im zweidimensionalen in Bild 4.3 angedeutet.

Daher liegt es nahe, für die Planung der nächsten Ansicht alle Punkte zu berücksichtigen. Das Eigenwertkriterium wird deshalb modifiziert, so dass es den Mittelwert der maximalen Eigenwerte der Kovarianzblöcke optimiert. Die nächste beste Ansicht im Sinne des modifizierten E-Kriteriums ist die, die den Mittelwert

der maximalen Eigenwerte aller Kovarianzblöcke der 3-D-Punkte minimiert:

$$\mathbf{c}^*_{\tau} := \operatorname{argmin}_{\mathbf{c}_{\tau}} \frac{1}{n} \sum_{i=1}^n \max_{j=1,2,3} \lambda_j(\Sigma_{\tau,i}), \quad (4.57)$$

dabei bezeichnet $\lambda_j(\Sigma_{\tau,i})$ den j -ten Eigenwert des i -ten Blocks in der Kovarianz des Zustands. Der konstante Faktor $1/n$ hat dabei keinen Einfluss auf die Minimierung und kann vernachlässigt werden.

Das modifizierte E-Kriterium minimiert also die Summe der maximalen Ausdehnung der einzelnen Kovarianzblöcke.

Für die Berechnung der einzelnen Eigenwerte und Eigenvektoren kann im übrigen folgender Satz benutzt werden:

Satz 4.6 (Eigenwerte und Eigenvektoren von Blockdiagonalmatrizen)

Sei $\mathbf{A}_1 \in \mathbb{R}^{N \times N}$ eine Matrix und $\mathbf{e}_v$ ein Eigenvektor von $\mathbf{A}_1$ und λ der zugehörige Eigenwert. Sei ferner $\mathbf{A}_2 \in \mathbb{R}^{M \times M}$ eine Matrix. Es gilt $\mathbf{e}_v^* = (\mathbf{e}_v \ 0)^T$ ist ein Eigenvektor der Matrix

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix} \quad (4.58)$$

und λ der zugehörige Eigenwert.

Beweis: Da $\mathbf{e}_v$ Eigenvektor mit Eigenwert λ von $\mathbf{A}_1$ ist, gilt

$$\mathbf{A}_1 \mathbf{e}_v = \lambda \mathbf{e}_v. \quad (4.59)$$

Es gilt aber auch

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix} \begin{pmatrix} \mathbf{e}_v \\ \mathbf{0} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{e}_v \\ \mathbf{0} \end{pmatrix}. \quad (4.60)$$

4.3.3 T-Optimalität

Die T-Optimalität nach Pukelsheim [Puk93] ignoriert sämtliche Kovarianzen in der Kovarianzmatrix Σ_{τ} und bewertet ausschließlich die Varianzen. Die Varianzen der einzelnen Punktkoordinatenschätzungen befinden sich in der Kovarianzmatrix Σ_{τ} des Zustands $\mathbf{z}$ auf der Diagonalen. Deshalb ist das Optimalitätskriterium die Spur (engl.: trace, deshalb T-Optimalität) der Kovarianzmatrix Σ_{τ} :

$$\mathbf{c}^*_{\tau} := \operatorname{argmin}_{\mathbf{c}_{\tau}} \operatorname{spur}(\Sigma_{\tau}). \quad (4.61)$$

Die T-Optimalität hat den Vorteil, dass sie sehr einfach auszuwerten ist. Darüber hinaus ist sie linear. Dafür muss man den Kompromiss eingehen, dass sie nicht die kompletten Informationen, die über die aktuelle Zustandsschätzung bekannt sind, verwendet, sondern sämtliche Kovarianzbedingungen ignoriert.

Für die T-Optimalität (4.61) ist leicht zu erkennen, dass sie in die Form

$$\mathbf{c}^*_{\tau} := \operatorname{argmin}_{\mathbf{c}_{\tau}} \sum_{i=1}^n \operatorname{spur}(\Sigma_{\tau,i}) \quad (4.62)$$

gebracht werden kann. Dies erlaubt eine gemeinsame Schreibweise der drei vorgestellten Gütemaße:

$$\mathbf{c}^*_{\tau} := \operatorname{argmin}_{\mathbf{c}_{\tau}} \sum_{i=1}^n \Gamma_{\mathbf{g}}(\Sigma_{\tau,i}), \quad (4.63)$$

dabei ist $\mathbf{g} = \{D, E, T\}$ und $\Gamma_{\mathbf{g}}$ steht für die (D)eterminante, den maximalen (E)igenwert bzw. die Spur (=T)race) von $\Sigma_{\tau,i}$.

Damit ist es möglich Ansichten zu bewerten und jene zu identifizieren, die die Unsicherheit im Sinne des jeweils verwendeten Gütemaßes am stärksten reduziert.

4.4 Berücksichtigung von Nebenbedingungen für die Sichtbarkeit

Mit Gleichung (4.63) ist ein allgemeines Optimierungskriterium gegeben. Bei der Optimierung sind Nebenbedingungen zu berücksichtigen, da es möglich ist, dass keine Beobachtung $\mathbf{o}$ gemacht werden kann. Ohne Beobachtung $\mathbf{o}$ lässt sich der Auffrischungsschritt nicht durchführen, da der Erwartungswert $\hat{\mathbf{z}}_{\tau}$ nicht mit Hilfe der Beobachtung aufgefrischt und somit die Unsicherheit nicht minimiert werden kann.

Dies ist kein Widerspruch: Man kann ohne Beobachtung zwar berechnen, wie sich die Kovarianz im Falle einer Beobachtung verhalten würde (anhand der Gleichungen (4.27) und (4.29)), nicht jedoch, wie der Erwartungswert der Schätzung $\hat{\mathbf{z}}_{\tau}$ sich verhält (Gleichung (4.28)). Deshalb ist es wichtig nur solche Ansichten zu wählen, in denen eine Beobachtung möglich ist. Diese Nebenbedingung wird nun in das Optimierungskriterium (4.63) integriert.

Zunächst ein paar Beispiele, warum u. U. keine Beobachtung möglich ist:

Außerhalb des Sichtbereichs Befindet sich ein Punkt außerhalb des Sichtbereichs der Kamera, so kann er nicht beobachtet werden.

Verdeckung Ein 3-D-Punkt kann durch ein anderes Element der Szene (Selbstverdeckung) oder durch weitere Gegenstände, z. B. dem Roboterarm, verdeckt sein.

Punktkorrespondenz Der Punktverfolger kann Punkte während der Verfolgung verlieren. Dadurch stehen die Punktkorrespondenzen nicht mehr zur Verfügung, d. h. es ist nicht mehr zuordenbar, welcher 3-D-Punkt welche 2-D-Beobachtung generiert. Somit steht keine Beobachtung für diesen 3-D-Punkt zur Verfügung.

Ursachen von verlorenen Punktkorrespondenzen sind z. B. eine sich ändernde Beleuchtung oder Glanzlichter. Solche Effekte sind sehr schwierig vorherzusagen und werden deshalb in dieser Arbeit nicht explizit modelliert.

4.4.1 Integration der Nebenbedingung in das Optimierungskriterium

Für die Integration der Nebenbedingungen sind zwei Fragen wichtig. Erstens, wie verhält sich der Zustandsschätzer, wenn er keine Beobachtung integrieren kann? Zweitens, wie lässt sich das in das Optimierungskriterium einbringen?

Die erste Frage ist einfach zu beantworten: Die Auffrischung kann nicht durchgeführt werden, da keine neuen Informationen verfügbar sind. Da das Kalman-Filter für den vorherigen Zeitschritt schon die optimale Schätzung bestimmt hat und keine zusätzlichen Informationen gewonnen werden konnten, sind die neuen Schätzwerte gleich den bisherigen⁶. Es gilt für den Schätzwert $\hat{z}_\tau = \hat{z}_{\tau-1}$ und für die neue Kovarianz $\Sigma_\tau = \Sigma_{\tau-1}$.

Zur Beantwortung der zweiten Frage muss man etwas weiter zurückgehen. Analog zu Gleichung (4.53) für das D-Kriterium lassen sich auch das modifizierte E- und das T-Kriterium in folgender Form schreiben:

$$c^*_\tau = \operatorname{argmin}_{c_\tau} \int r_c(o_\tau) \Gamma_g(\Sigma_\tau) d o_\tau. \quad (4.64)$$

Das Integral in Gleichung (4.64) berechnet den Erwartungswert von Γ_g , indem über alle Beobachtungen o_τ integriert wird. Falls Σ_τ unabhängig von o_τ ist, lässt sich das Argument aus Abschnitt 4.3.1 verwenden: Der Term $\Gamma_g(\Sigma_\tau)$ hängt nicht von o_τ ab. Er kann aus dem Integral gezogen werden. Anschließend wird nur noch eine Wahrscheinlichkeitsdichte über den gesamten Definitionsbereich integriert. Das Integral ist daher gleich Eins. In diesem Fall würde sich das Optimierungskriterium (4.63) ergeben. Der Term $\Gamma_g(\Sigma_\tau)$ hängt davon ab, ob eine Beobachtung gemacht werden kann oder nicht.

Für das weitere Vorgehen wird angenommen, dass jeder 3-D-Weltpunkt durch seinen eigenen Kalman-Filter rekonstruiert wird (vgl. Abschnitt 4.2.2). Ferner wird für jeden Punkt nun berücksichtigt, ob er beobachtet werden kann. In diesem Fall hängt zwar die a posteriori Kovarianz $\Sigma_{\tau,i}$ nicht direkt von der Beobachtung $o_{\tau,i}$ ab, sehr wohl aber davon, ob die Beobachtung zustande kommt. Sei O die Menge von Beobachtungen, die sichtbar sind. Damit gilt für die a posteriori Kovarianz

$$\Sigma_{\tau,i}^+ = \begin{cases} \Sigma_{\tau,i} = (I - \mathfrak{K}_{\tau,i} B_\tau) \Sigma_{\tau-1,i} & \text{falls } o_{\tau,i} \in O \\ \Sigma_{\tau-1,i} & \text{sonst} \end{cases}. \quad (4.65)$$

Um ein allgemeines Gütekriterium herzuleiten, wird über die Anzahl der Punkte n ein Induktionsbeweis geführt. Zunächst gilt für $n = 1$: Die Kovarianz Σ_τ aus Gleichung (4.64) ist nicht mehr konstant für alle $o_{\tau,i}$, daher muss das Integral aus (4.64) aufgespalten werden:

$$\begin{aligned} c^*_\tau = \operatorname{argmin}_{c_\tau} & \int_{o_\tau \in O} r_c(o_\tau) \Gamma_g(\Sigma_\tau) d o_\tau + \int_{o_\tau \notin O} r_c(o_\tau) \Gamma_g(\Sigma_{\tau-1}) d o_\tau = \\ & \operatorname{argmin}_{c_\tau} p \Gamma_g(\Sigma_\tau) + (1 - p) \Gamma_g(\Sigma_{\tau-1}), \end{aligned} \quad (4.66)$$

da $\Gamma_g(\Sigma_\tau)$ nach der Entscheidung, ob die Beobachtung sichtbar ist, nicht mehr von o_τ abhängt. Die Wahrscheinlichkeit p ist definiert als Wahrscheinlichkeit, ob der Punkt sichtbar ist:

$$p := \int_{o_\tau \in O} r_c(o_\tau) d o_\tau. \quad (4.67)$$

⁶Dies gilt nur dann, wenn das Kalman-Filter ohne Zustandsübergang verwendet wird. Andernfalls muss der Zustandsübergang berechnet werden. Dadurch ändert sich nicht nur der Erwartungswert der Schätzung, sondern auch die Unsicherheit nimmt durch das Zustandsübergangsrauschen zu.

Geht man nun von zwei Punkten aus, so gibt es für die Wahrscheinlichkeiten p_1 und p_2 insgesamt vier Möglichkeiten für die Sichtbarkeit: $(o_{\tau,1} \in O, o_{\tau,2} \in O)$, $(o_{\tau,1} \notin O, o_{\tau,2} \in O)$, $(o_{\tau,1} \in O, o_{\tau,2} \notin O)$, $(o_{\tau,1} \notin O, o_{\tau,2} \notin O)$. Es wird angenommen, dass die Wahrscheinlichkeiten p_1 und p_2 unkorreliert sind. Die Unkorreliertheit folgt aus den oben genannten Annahmen: unkorreliertes Beobachtungsrauschen und Blockdiagonalmatrix des Zustands. Für das Optimierungskriterium gilt damit:

$$\begin{aligned} c_{\tau}^* = \operatorname{argmin}_{c_{\tau}} \quad & p_1 p_2 \Gamma_g(\operatorname{diag}(\Sigma_{\tau,1}, \Sigma_{\tau,2})) + \\ & (1 - p_1) p_2 \Gamma_g(\operatorname{diag}(\Sigma_{\tau-1,1}, \Sigma_{\tau,2})) + \\ & p_1 (1 - p_2) \Gamma_g(\operatorname{diag}(\Sigma_{\tau,1}, \Sigma_{\tau-1,2})) + \\ & (1 - p_1) (1 - p_2) \Gamma_g(\operatorname{diag}(\Sigma_{\tau-1,1}, \Sigma_{\tau-1,2})). \end{aligned} \quad (4.68)$$

Da für alle drei möglichen Gütemaße $g = \{D, E, T\}$ die Beziehung

$$\Gamma_g(\operatorname{diag}(\Sigma_{\tau,1}, \Sigma_{\tau,2})) = \Gamma_g(\Sigma_{\tau,1}) + \Gamma_g(\Sigma_{\tau,2}) \quad (4.69)$$

gilt, lässt sich (4.68) durch einfache Umformungen vereinfachen:

$$\begin{aligned} c_{\tau}^* = \operatorname{argmin}_{c_{\tau}} \quad & p_1 \Gamma_g(\Sigma_{\tau,1}) + (1 - p_1) \Gamma_g(\Sigma_{\tau-1,1}) + \\ & p_2 \Gamma_g(\Sigma_{\tau,2}) + (1 - p_2) \Gamma_g(\Sigma_{\tau-1,2}). \end{aligned} \quad (4.70)$$

Per Induktion lässt sich allgemein für $n > 2$ Punkte zeigen:

$$c_{\tau}^* = \operatorname{argmin}_{c_{\tau}} \quad \sum_{i=1}^n p_i \Gamma_g(\Sigma_{\tau,i}) + (1 - p_i) \Gamma_g(\Sigma_{\tau-1,i}). \quad (4.71)$$

Mit (4.71) ist in das ursprüngliche Optimierungskriterium (4.63) die Sichtbarkeitsnebenbedingung integriert worden. Um nun mit der Optimierungsgleichung (4.71) die nächste beste Ansicht zu wählen, wird nun geklärt, wie man die Wahrscheinlichkeit p_i berechnet, mit der ein Punkt p_{wi} im Kamerabild sichtbar ist.

4.4.2 Berechnung der Wahrscheinlichkeit für die Sichtbarkeit

Aus der Aufzählung am Anfang von Abschnitt 4.4 (Seite 91) werden zwei Effekte genauer betrachtet, weshalb ein Punkt nicht sichtbar sein kann: Der Punkt ist außerhalb des Sichtbereichs oder der Punkt ist durch einen anderen Teil der Szene verdeckt (Selbstverdeckung). Es wird davon ausgegangen, dass diese beiden Effekte statistisch unabhängig sind. Durch die Unabhängigkeitsannahme kann die Sichtbarkeit durch das eingeschränkte Sichtfeld der Kamera und durch Selbstverdeckungen getrennt voneinander untersucht werden.

4.4.2.1 Punkt außerhalb des Sichtbereichs

Um die Wahrscheinlichkeit zu bestimmen, mit der der Punkt im Sichtbereich der Kamera liegt, kann man folgendermaßen vorgehen: Jeder 3-D-Punkt ist durch die Parameter einer 3-D-Normalverteilung gegeben. Diese Normalverteilung lässt sich auf die 2-D-Bildebene projizieren. Der Sichtbereich der Kamera ist ein Rechteck

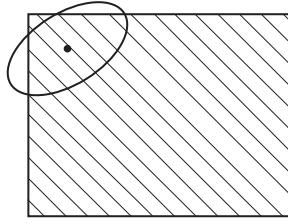


Bild 4.4 Die Wahrscheinlichkeit, dass ein Punkt im Sichtbereich liegt, lässt sich durch die Integration über die Rechtecksfläche berechnen. Die Unsicherheit des 2-D-Punkts ist durch seine Kovarianzellipse angedeutet.

in der Bildebene. Um die Wahrscheinlichkeit zu ermitteln, muss nun die projizierte Wahrscheinlichkeitsdichte über die Rechtecksfläche integriert werden (vgl. Bild 4.4).

Die genannte Vorgehensweise scheint zunächst relativ einfach, jedoch ist die praktische Umsetzung nicht ganz trivial. Zwei Probleme treten dabei auf:

1. Werden die normalverteilten Koordinaten eines 3-D-Punkts auf die Bildebene perspektivisch projiziert, so sind die Bildkoordinaten nicht mehr normalverteilt (vgl. Abschnitt 4.1.2.4).
2. Geht man von einer affinen Projektion aus, z. B. durch Verwendung des paraperspektivischen Projektionsmodells (vgl. Abschnitt 2.1.4.2), so sind die Bildkoordinaten zwar weiterhin normalverteilt, jedoch muss eine 2-D-Gauß-Dichte integriert werden, was nicht geschlossen möglich ist.

Im ersten Fall kann der Transformationssatz für Dichten [Irl05, S. 123] verwendet werden, um die Dichtefunktion der projizierten Normalverteilung zu bestimmen. Dies ist aufgrund der auftretenden Integrale sehr aufwändig.

Eine Alternative ist eine Monte-Carlo-Integration [Dou01]: Aus der 3-D-Normalverteilung werden S Stichproben gezogen. Jede Stichprobe wird perspektivisch auf die Bildebene projiziert. Gezählt wird die relative Häufigkeit, d. h. wie oft eine Stichprobe im Bild liegt, dividiert durch die Anzahl der Stichproben S .

Die zweite Möglichkeit wird in [Zob04] benutzt: Mit Hilfe der bereits aus dem Kalman-Filter bekannten Größen kann man den Erwartungswert der 3-D-Normalverteilung und die Kovarianz in der Bildebene bestimmen:

$$E(\mathbf{o}_\tau) = \mathbf{b}(\hat{\mathbf{z}}_\tau), \quad \Sigma_{\mathbf{o}_\tau} = \mathbf{B} \Sigma_\tau \mathbf{B}^\top. \quad (4.72)$$

Vernachlässigt man zusätzlich noch die Nebendiagonalelemente und geht von einer unkorrelierten Beobachtung aus, so lässt sich die Wahrscheinlichkeit p_{sb} für einen Punkt folgendermaßen ermitteln:

$$p_{sb} = \int_0^{b_x} r(\mathbf{o}_\tau) d\mathbf{o}_{\tau,x} \cdot \int_0^{b_y} r(\mathbf{o}_\tau) d\mathbf{o}_{\tau,y}, \quad (4.73)$$

dabei ist b_x bzw. b_y die Bildbreite bzw. -höhe in Pixeln und das Subskript sb steht für Sichtbereich. Die Integrale lassen sich nach vorheriger Transformation auf Varianz $\sigma^2 = 1$ mit Hilfe der Gauß-Fehlerfunktion auswerten [Zob04, Den03b].

Alternativ lassen sich entsprechende Tabellen für die Auswertung des Integrals erzeugen, so dass der Zeitaufwand geringer als bei der Monte-Carlo-Simulation ist, aber trotzdem die Nebendiagonalelemente in der Kovarianz der Normalverteilung berücksichtigt werden [Deu06].

4.4.2.2 Selbstverdeckungen

Ein Punkt ist genau dann sichtbar, wenn er nicht durch eine Oberfläche verdeckt wird. Zum Beispiel ist die Rückseite eines Objekts von vorne nicht sichtbar. Wurde ein Punkt auf der Rückseite rekonstruiert, so wird bei der Ansichtenplanung versucht, die Schätzung dieses Punkts zu verbessern. Deshalb ist es wichtig, dass der Algorithmus beachtet, von welchen Positionen aus welcher Punkt sichtbar ist. Der Algorithmus muss entscheiden, ob es die Unsicherheit eher reduziert, Punkte auf der Rückseite oder auf der Vorderseite zu beobachten.

Die Frage ist: Wie kann man die Wahrscheinlichkeit berechnen, mit der ein Punkt verdeckt ist? Hat man nur die 3-D-Rekonstruktion als Menge von 3-D-Punkten, so lassen sich keine Aussagen über Selbstverdeckungen machen. Man benötigt daher eine Oberfläche und nicht nur einzelne Punkte. Dazu wird die Oberfläche durch ein Dreiecksnetz dargestellt, wobei die Ecken eines jeden Dreiecks in den rekonstruierten Punkten liegen.

Das Verfahren basiert auf der Delauny-Triangulation [For97]. Eine Beschreibung der Delauny-Triangulation würde über den Rahmen dieser Arbeit hinausgehen, deshalb wird an dieser Stelle auf die Arbeit von Scholz verwiesen [Sch08, Seite 78 ff.].

Mit Hilfe dieses Dreiecksnetzes lässt sich nun die Wahrscheinlichkeit p_{sv} bestimmen, ob ein gegebener Punkt p_{wi} verdeckt ist. Zunächst gilt, wenn der Sichtstrahl von p_{wi} zum optischen Zentrum der Kamera O_Z mindestens ein Dreieck aus dem Dreiecksnetz schneidet, so ist der Punkt offensichtlich verdeckt, andernfalls ist er nicht verdeckt. Die Berechnung, ob sich der Schnittpunkt des Sichtstrahls mit der Ebene des Dreiecks innerhalb oder außerhalb des Dreiecks schneidet, lässt sich mit baryzentrischen Koordinaten relativ einfach feststellen. Eine Beschreibung von baryzentrischen Koordinaten findet sich in einschlägiger Literatur, z. B. [Sla02].

Bei einem einzelnen Dreieck muss für die Berechnung von p_{sv} eine zwölfdimensionale Wahrscheinlichkeitsverteilung ausgewertet werden, denn der zu untersuchende 3-D-Punkt p_{wi} und die drei Punkte, die das jeweilige Dreieck aufspannen, sind jeweils durch dreidimensionale, voneinander unabhängige Normalverteilungen gegeben. Für diese Wahrscheinlichkeit muss die Verteilung des Schnittpunkts des Sichtstrahls mit der Ebene, in der sich das Dreieck befindet, berechnet werden und anschließend über die Dreiecksfläche integriert werden. Der Aufwand ist dabei extrem hoch, so dass diese Berechnungsmöglichkeit in dieser Arbeit nicht weiter verfolgt wird.

Die Wahrscheinlichkeit lässt sich auch mit einer Monte-Carlo-Integration berechnen: Für jeden 3-D-Punkt, also sowohl für den zu untersuchenden als auch für die, die das Dreiecksnetz aufspannen, werden jeweils S Stichproben, mit den entsprechenden Verteilungsparametern, die ja aufgrund des Kalman-Filter-Ansatzes bekannt sind, gezogen. Bei einer zu analysierenden Kameraposition wird gezählt, wie häufig der jeweilige Sichtstrahl mindestens ein Dreieck schneidet. Die Wahrscheinlichkeit für einen Punkt ist gegeben durch die relative Häufigkeit.

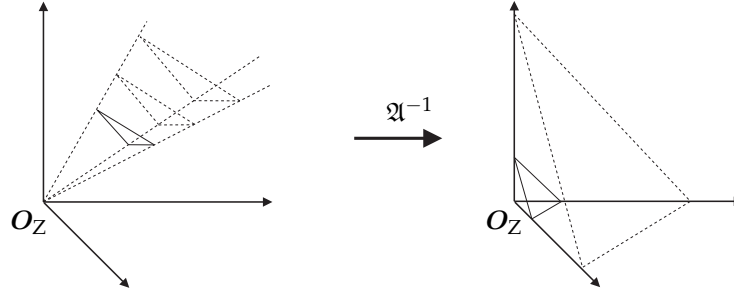


Bild 4.5 Durch eine affine Transformation lässt sich das Dreieck auf die Grundfläche des Standardsimplex abbilden, wodurch die Festlegung der Integrationsgrenzen wesentlich vereinfacht wird.

Variante Eine direkte Integration der 12-dimensionalen Verteilung ist extrem aufwändig. Man kann alternativ die vereinfachende Annahme machen, dass die 3-D-Punkte, die das Dreiecksnetz aufspannen, fest sind, also nicht mehr normalverteilt. Dann muss nur noch eine dreidimensionale Verteilung ausgewertet werden. Dies kann relativ einfach durch eine Integration in baryzentrischen Koordinaten durchgeführt werden, deren Grundidee kurz vorgestellt wird:

Der zu untersuchende 3-D-Punkt $\mathbf{p}_{wi}$ ist verdeckt, wenn er im Schatten des Dreiecks liegt (vgl. linker Teil von Bild 4.5). O. B. d. A. wird dabei das optische Zentrum O_Z der Kamera in den Ursprung gelegt. Die Wahrscheinlichkeit p_{sv} ist die Integration über den Schatten des Dreiecks. Die Integrationsgrenzen sind in diesem Fall analytisch schwer zu beschreiben. Durch eine affine Transformation $\mathfrak{A}^{-1}$ können die Ecken des Dreiecks auf den Ecken des Standardsimplex $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ und $(0,0,1)$ abgebildet werden. Dadurch lassen sich die Integrationsgrenzen einfach darstellen und für die Wahrscheinlichkeit p_{sv} ergibt sich:

$$p_{sv} = \int_{\triangle\text{-Schatten}} r(\mathbf{p}_{wi}) d\mathbf{p}_{wi} = \quad (4.74)$$

$$\int_{1.\text{Oktant}} r(\mathfrak{A}\mathbf{p}_{wi}) \cdot \det(\mathfrak{A}) d\mathbf{p}_{wi} - \int_{\text{Standardsimplex}} r(\mathfrak{A}\mathbf{p}_{wi}) \cdot \det(\mathfrak{A}) d\mathbf{p}_{wi} = \quad (4.75)$$

$$\begin{aligned} & \int_0^\infty \int_0^\infty \int_0^\infty r(\mathfrak{A}\mathbf{p}_{wi}) \cdot \det(\mathfrak{A}) dp_{wiz} dp_{wiy} dp_{wix} - \\ & \int_0^1 \int_0^{1-x} \int_0^{1-x-y} r(\mathfrak{A}\mathbf{p}_{wi}) \cdot \det(\mathfrak{A}) dp_{wiz} dp_{wiy} dp_{wix} \end{aligned} \quad (4.76)$$

Die Normalverteilungseigenschaft des Punkts $\mathbf{p}_w$ bleibt erhalten, da die Transformation $\mathfrak{A}$ affin ist (vgl. Satz 3.4). Die Integration findet also im dreidimensionalen Raum statt und nicht auf der Ebene, die durch das jeweilige Dreieck aufgespannt wird. Dadurch kann die aufwändige Berechnung der 2-D-Verteilung in der Ebene mit dem Transformationssatz für Dichten eingespart werden.

Da die Ecken der jeweiligen Dreiecke fest sind, ergibt sich die Gesamtwahrscheinlichkeit, dass ein Punkt sichtbar ist, aus dem Produkt der einzelnen Wahrscheinlichkeiten, dass der Punkt $\mathbf{p}_{wi}$ nicht vom j -ten Dreieck verdeckt ist. Die Ge-

samtwahrscheinlichkeit p_i , dass der i -te Punkt sichtbar ist, ist somit:

$$p_i = p_{sb} \cdot p_{sv}. \quad (4.77)$$

Damit ist nun also geklärt, wie man die Wahrscheinlichkeit p_i für die Sichtbarkeit des i -ten Punkts berechnen kann. Somit kann man das Gütemaß aus Gleichung (4.71) auswerten und die nächste optimale Ansicht planen.

4.5 Zusammenfassung

In diesem Kapitel wurden die theoretischen Grundlagen für ein probabilistisches Verfahren zur sukzessiven Verbesserung der Genauigkeit einer 3-D-Rekonstruktion mit Hilfe der Ansichtenplanung erarbeitet.

Die 3-D-Rekonstruktion wurde dazu zunächst als probabilistisches Zustandsschätzproblem formuliert. Dadurch können aus der Literatur bekannte Zustandsschätzer auf das Problem angewendet werden. In dieser Arbeit wird dazu ein Kalman-Filter verwendet. Es wurde explizit erläutert, wie sich das Kalman-Filter auf das konkrete Problem der 3-D-Rekonstruktion übertragen lässt. Dazu wird das Kalman-Filter, ein Zustandsschätzer für dynamische Systeme, auf ein statisches System angewandt und zusätzlich dafür eine sich zeitlich änderbare Beobachtungsfunktion b_τ berücksichtigt.

Durch den Einsatz des Kalman-Filters lassen sich die Koordinaten der zu rekonstruierenden Punkte durch eine mehrdimensionale Normalverteilung beschreiben. Mit Hilfe plausibler Annahmen konnte analytisch bewiesen werden, dass sich der Berechnungsaufwand mit Hilfe des sequentiellen Kalman-Filters und der in Abschnitt 4.2.2 bewiesenen Modifikation erheblich reduzieren lässt: Das Problem für die 3-D-Rekonstruktion von n Punkten lässt sich auf die n Rekonstruktionen eines einzelnen Punkts zurückführen, wodurch sich die Dimensionen der zu invertierenden Matrizen bei der Berechnung des Kalman-Filters deutlich reduzieren. Dabei muss für die Reduktion des Berechnungsaufwands keine Approximation durchgeführt werden, sondern das Ergebnis ist weiterhin exakt, d. h. es stimmt mit dem des unmodifizierten erweiterten Kalman-Filters, wie es in der Literatur verwendet wird, überein.

Auf der Basis der probabilistischen Verteilungen der einzelnen Punkte konnten Gütemaße abgeleitet werden, die D-, modifizierte E- und die T-Optimalität. Alle Gütemaße basieren im Wesentlichen darauf, die „Größe“ der Kovarianzmatrix Σ_τ der Schätzung zu ermitteln und somit eine Ordnungsrelation im $\mathbb{R}^{N \times N}$ einzuführen. Die Kovarianzmatrix ist dabei die Kovarianz des Schätzfehlers und kann allgemein als Unsicherheit der Schätzung interpretiert werden.

Entscheidend für die Ansichtenplanung ist, dass die a posteriori Kovarianzmatrix Σ_τ des Zustands berechnet werden kann, ohne dass eine Beobachtung tatsächlich gemacht werden muss, also ohne den konkreten Wert o_τ zu kennen. Es lässt sich also vorhersagen, wie sich die Kovarianzmatrix verhalten wird, wenn von einer bestimmten Ansicht aus eine Beobachtung generiert wird. Man wird folglich die Ansicht wählen, die die Kovarianz im Sinne des jeweiligen Optimierungskriteriums am stärksten reduziert.

Bei der Berechnung der nächsten besten Ansicht sind Nebenbedingungen zu beachten. In diesem Kapitel wurde dabei explizit die Sichtbarkeit eines Punkts be-

rücksichtigt. Es gibt noch eine weitere wichtige Nebenbedingung, nämlich die Erreichbarkeit, also ob aufgrund von Beschränkungen des Roboterarms die optimale Position überhaupt erreichbar ist. Dies ist unter anderem Gegenstand von Kapitel 5. Die Sichtbarkeitsbedingungen wurden in die Gleichungen für die Optimierung integriert, so dass ein geschlossenes Optimierungskriterium für die nächste beste Ansicht entwickelt werden konnte.

Gegenstand des Abschnitts 4.4 war, wie die Wahrscheinlichkeiten für die Sichtbarkeit eines Punkts berechnet werden können. Zwei Möglichkeiten, warum ein Punkt nicht sichtbar ist, wurden dabei analysiert: Die Wahrscheinlichkeit, dass der Punkt außerhalb des Sichtbereichs liegt, und die, dass der Punkt verdeckt ist. Beide Wahrscheinlichkeiten sind voneinander unabhängig und somit konnten diese beiden Möglichkeiten getrennt voneinander betrachtet werden. Die einfachste Möglichkeit die beiden Wahrscheinlichkeiten zu berechnen, ist eine Monte-Carlo-Integration durchzuführen. Alternativ könnten vereinfachende Annahmen getroffen werden, um anschließend die Integrale numerisch zu berechnen.

Wie nun die Gleichung (4.71) in einem realen System optimiert wird, wird im Kapitel 5 beschrieben.

Kapitel 5

Ein System zur Realisierung des probabilistischen Ansatzes

Dieses Kapitel beschreibt wichtige Details für die praktische Realisierung des probabilistischen Ansatzes aus Kapitel 4.

Prinzipiell wird bei der Ansichtenplanung vorausgesetzt, dass die Kamera möglichst präzise an der optimalen Position positioniert werden kann. Für die Positionierung wird deshalb ein Roboterarm verwendet. Die Experimente in Kapitel 6 wurden mit unterschiedlichen Robotern durchgeführt. Eine kurze Beschreibung der Roboterarme findet sich im Abschnitt 5.1. Dort wird darüber hinaus erläutert, wie die unterschiedlichen Bewegungsmodelle während der Ansichtenplanung berücksichtigt werden können. Diese Vorgehensweise wird sowohl für den probabilistischen Ansatz als auch für die Optimierung der Stereoparameter eingesetzt.

In Kapitel 4 wurde bereits deutlich, dass die einzelnen Arbeitsschritte des Algorithmus sich in zwei Hauptkomponenten aufteilen lassen: die 3-D-Rekonstruktion und die Ansichtenplanung. Beide Arbeitsschritte werden abwechselnd aufgerufen. Im Abschnitt 5.2 werden diese beiden Hauptkomponenten, die den wesentlichen Anteil des Gesamtsystems bilden, erklärt.

In Abschnitt 5.3 werden schließlich alle Komponenten des Gesamtsystems noch einmal kurz erläutert sowie ihre Abhängigkeiten deutlich gemacht.

5.1 Beschreibung der Roboter

In diesem Abschnitt werden die verwendeten Roboter sowie die Berücksichtigung des Bewegungsmodells des Roboters für die Ansichtenplanung zur 3-D-Rekonstruktion beschrieben. Durch den modularen Aufbau des Gesamtsystems (vgl. Abschnitt 5.2 und 5.3) ist das System relativ leicht mit anderen Robotern zu betreiben. Für die Experimente in Kapitel 6 wurden insgesamt zwei Roboter verwendet: ein Drehteller der Firma Isel (Bild 5.1a) und ein Scorbot ER VII (Bild 5.1b).

Für die Ansichtenplanung muss das Bewegungsmodell des Roboters bekannt sein. Das Bewegungsmodell beschreibt die Positionen, die die Roboterspitze anfahren kann. Natürlich ist das im Falle des Drehtellers recht einfach: Hier kann die Roboterspitze, und damit die Kamera, auf der oberen Hälfte einer Kugel positioniert werden. Der Arbeitsbereich des Roboterarms ließe sich hier also über Kugelkoordinaten parametrieren. Bei dem Gelenkroboter Scorbot ER VII ist diese Aufgabe wesentlich schwieriger. Um zu überprüfen, ob eine 3-D-Position mit der gewünschten Ausrichtung der Roboterhand erreichbar ist, muss das Problem der inversen Kinematik gelöst werden, d. h. es muss geklärt werden, ob sich die ge-

Gelenk-Nr	θ	d	l	α
1	θ_1	436,7 mm	50 mm	90°
2	θ_2	35,2 mm	300 mm	0
3	θ_3	0	250 mm	0
4	θ_4	0	151 mm	α_4

Tabelle 5.1 Tabelle der Denavit-Hartenberg-Parameter für den Scrobot ER VII. Das 5. Gelenk ist im 4. mit berücksichtigt, da es ohne weitere Translation eine Rotation um die x -Achse ermöglicht.

Gelenk-Nr	θ	d	l	α
1	θ_1	0	0	α_1
2	0	1000mm	0	180°

Tabelle 5.2 Tabelle der Denavit-Hartenberg-Parameter für den Drehteller. Die zwei Rotationsgelenke treffen sich hier im Ursprung.

wünschte Konstellation im „dextrous workspace“ oder im „reachable workspace“ befindet (vgl. Abschnitt 2.4.2 und [Cra04]).

Wie bereits im Abschnitt 2.4.2 erläutert, existiert kein allgemeingültiger Lösungsansatz für das Problem der inversen Kinematik [Sie96]. Hingegen ist die Beschreibung des Bewegungsmodells mit Hilfe der direkten Kinematik (s. Abschnitt 2.4.1) relativ einfach: sind die konstanten Denavit-Hartenberg-Parameter bekannt und die Variablen auslesbar, so lässt sich mit der Denavit-Hartenberg-Matrix die Position und Orientierung der Hand bestimmen. Da für die einzelnen Drehgelenke üblicherweise die zulässigen Winkelbereiche im Datenblatt des Roboterarms angegeben sind, lässt sich über die direkte Kinematik mit Hilfe der Denavit-Hartenberg-Matrix das Bewegungsmodell des Roboters und damit der zulässige Arbeitsbereich der Roboterhand sehr einfach beschreiben.

In Bild 5.2 ist für den Scrobot ER VII ein Profil des Arbeitsbereichs visualisiert. Dazu wurden die zulässigen Winkelbereiche der einzelnen Gelenke äquidistant abgetastet und die Endposition mit Hilfe der Denavit-Hartenberg-Parameter berechnet. In diesem Fall besteht die Endposition nicht mehr aus äquidistanten Positionen, wie man ebenfalls in Bild 5.2 erkennt.

Für die Denavit-Hartenberg-Parameter ist eine tabellarische Auflistung der einzelnen Parameter für alle Gelenke üblich. Für den Scrobot ER VII sind die Parameter in Tabelle 5.1 aufgeführt. Für den Drehteller lässt sich zwar aufgrund des Aufbaus der Arbeitsbereich auf einer Halbkugel über Kugelkoordinaten direkt bestimmen, aber der Ansatz über die Denavit-Hartenberg-Matrix ist allgemeiner und wird deshalb in dieser Arbeit verwendet. Dazu sind in Tabelle 5.2 die Parameter für den Drehteller aufgelistet. Das zweite Gelenk ist eigentlich starr und dient nur dazu, die Hand-Auge-Kalibrierung vorzuinitialisieren.

Mit der Denavit-Hartenberg-Matrix kann man also die Bewegungsmodelle der Roboterarme Scrobot ER VII und Drehteller auf sehr einfache Art und Weise beschreiben.

Optimiert wird das Gütemaß (4.71) über eine vollständige Suche einer diskreten Abtastung des Arbeitsbereichs. Die dafür verwendete Zeit liegt beispielsweise bei 110 3-D-Punkten und 4100 möglichen Positionen, wie sie bei den Drehteller-Experimenten in Kapitel 6.3 verwendet wird, bei ca. 2s. Für die Verwendung der

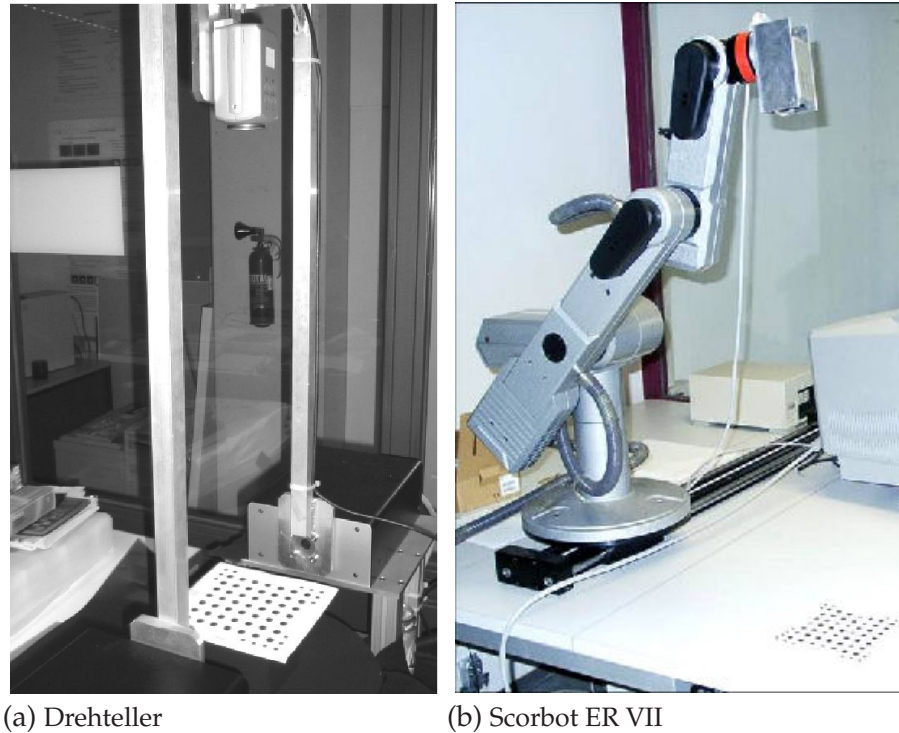


Bild 5.1 Zwei verschiedene Roboterarme, die in den Experimenten verwendet wurden.

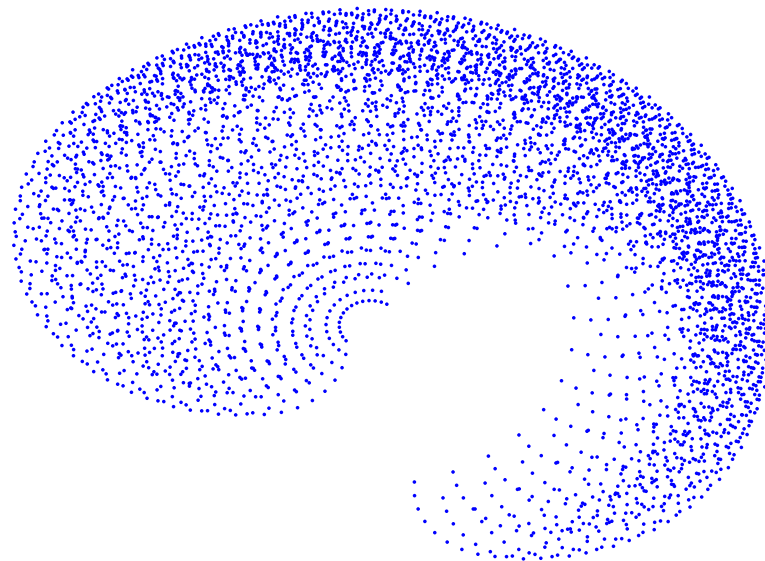


Bild 5.2 Das Profil des Arbeitsbereichs des Scorbot ER VII. Die Punkte visualisieren erreichbare Positionen bei einer diskreten Abtastung der Winkelbereiche.

vollständigen Suche sprechen folgende Gründe: Über die Denavit-Hartenberg-Matrix lässt sich der Arbeitsbereich sehr einfach beschreiben; nicht erreichbare Positionen werden überhaupt nicht berücksichtigt. Eine lokale Optimierung über die Winkel ist nicht unbedingt erfolgversprechend: Kann eine Startposition des Endeffektors über zwei oder mehr Gelenkeinstellungen erreicht werden (wie z. B. in Bild 2.7 angedeutet), so ist es möglich, dass von der einen Position das globale Optimum gefunden wird, von der anderen, bei lokaler Optimierung, aber nicht. Deshalb ist

entweder eine gute Startlösung (aber nicht in $\mathbb{R}^3$, sondern im Gelenkraum des Roboters) oder ein globales Optimierungsverfahren notwendig. In dieser Arbeit wird zur globalen Optimierung die vollständige Suche verwendet. So können alle möglichen Positionen auf einem diskreten Gitter sukzessive durchprobiert werden. Andere Optimierungsverfahren sind natürlich möglich. Eine kurze Diskussion findet sich in Abschnitt 7.1.

Prinzipiell genügt es bei diesem Ansatz, eine Liste von Aufnahmepositionen vorzuhalten. Dies ist ein weiterer Vorteil der vollständigen Suche, denn so ist der Ansatz in dieser Hinsicht äußerst flexibel: es ist überhaupt nicht notwendig auf spezielle Bewegungsmodelle des Roboters einzugehen.

Es wurde gezeigt, wie der vorgestellte Algorithmus mit verschiedenen Roboterarmen verwendet werden kann. Prinzipiell ist dazu eine Beschreibung des Arbeitsbereichs notwendig. Bei einfachen Arbeitsbereichen, wie z. B. einer Teilkugel, ist die Beschreibung sehr einfach. Bei Roboterarmen mit mehreren Gelenken bietet sich dagegen die Parametrierung des Arbeitsbereichs mit Hilfe der Denavit-Hartenberg-Parameter an. Da diese Parametrierung in der Robotik sehr weit verbreitet ist [Cra04], kann der Algorithmus mit geringem Aufwand an andere Roboterarme angepasst werden.

Diese Methodik ist sowohl bei der Optimierung der Stereoparameter, wie für den Algorithmus aus Abschnitt 3.2, als auch für die probabilistische Ansichtenauswahl aus Kapitel 4 möglich.

5.2 Softwarekomponenten

Der in Kapitel 4 vorgestellte Algorithmus arbeitet in zwei Phasen: die 3-D-Rekonstruktion und die Ansichtenplanung. Beim Start des Algorithmus wird zunächst eine, im Allgemeinen grobe, initiale Schätzung der 3-D-Geometrie erstellt. Darauf basierend wird anschließend die nächste beste Ansicht geplant. Von dieser wird anschließend eine Aufnahme gemacht. Diese Beobachtung wird nun mit Hilfe des Kalman-Filters integriert, um die bestehende Schätzung zu verbessern. Solange ein zuvor festgelegtes Abbruchkriterium nicht erreicht ist, wiederholen sich nun die Schritte Ansichtenplanung und 3-D-Rekonstruktion abwechselnd. Nun zu den Details, die in beiden Schritten ablaufen.

5.2.1 Initiale 3-D-Rekonstruktion

Für die initiale Schätzung werden mindestens zwei Bilder der Szene benötigt. Dazu werden manuell zwei Aufnahmepositionen festgelegt. Der Roboterarm fährt von der ersten Position zur zweiten. Während der Fahrt werden Bilder aufgenommen und die zugehörigen Kameraparameter gespeichert. Um Synchronisationsprobleme und Bewegungsunschärfe zu vermeiden, stoppt der Roboter kurz, um die Aufnahme zu machen, die Parameter auszulesen und alles zu speichern. Die Zwischenaufnahmen sind wichtig, da der Algorithmus zur Punktverfolgung aus Abschnitt 2.5.1 auf der Annahme basiert, dass sich die Punkte im Bild nicht zu stark bewegen (der Ansatz basiert auf einer Taylor-Reihen-Approximation, wodurch der systematische Fehler umso größer wird, je größer die Bewegung im

Bild ist). Deshalb ist es notwendig, die Bewegung zwischen zwei einzelnen Bildern möglichst klein zu halten.

Aus den Punktkorrespondenzen und den Kameraparametern pro Bild kann eine initiale Schätzung der 3-D-Rekonstruktion durch eine Triangulation (vgl. Abschnitt 2.5) berechnet werden. In den Experimenten wird dabei zunächst die lineare Triangulation (vgl. Abschnitt 2.5.2.2) berechnet. Diese dient als Startlösung für die anschließende nichtlineare Optimierung (vgl. Abschnitt 2.5.2.3). Die initiale Zustandsschätzung $\hat{z}_0$ ist somit gegeben.

Die initiale Unsicherheit der Schätzung wird zunächst durch eine diagonale Kovarianzmatrix $\Sigma_0^- = \sigma^2 \mathbf{I}$ repräsentiert. In den Experimenten wird dazu konstant $\sigma^2 = 8$ gesetzt, was in etwa der erwarteten Unsicherheit des Punktverfolgers entspricht. Anschließend werden alle 2-D-Punktkoordinaten der korrespondierenden Punkte mit Hilfe der Kalman-Filter-Gleichungen (4.18)-(4.20) dazu verwendet, die Schätzung aufzufrischen. Dadurch kann eine realistischere initiale Kovarianzmatrix Σ_0 berechnet werden, d. h. eine, die die tatsächliche Unsicherheit besser repräsentiert als die Diagonalmatrix. Schließlich haben die Aufnahmepositionen ja einen entscheidenden Einfluss auf die Kovarianzmatrix und die Kovarianzmatrix bei der Ansichtenplanung wiederum auf die Auswahl der Aufnahmeposition. Deswegen ist eine realistische Initialisierung der Kovarianzmatrix sehr wichtig.

Dies gilt insbesondere für das E-Kriterium: ist die Kovarianzmatrix eine Diagonalmatrix der Form $\Sigma_0 = \sigma^2 \mathbf{I}$, so liefern alle möglichen Aufnahmepositionen den gleichen Wert des E-Kriteriums. Zur Begründung: Die Kovarianz $\Sigma_0 = \sigma^2 \mathbf{I}$ eines einzelnen Punkts kann als Kugel aufgefasst werden (also der Spezialfall eines Ellipsoids). Durch die Auffrischung wird die Kugel zu einem Ellipsoiden verformt. Der größte Halbmesser ist aber immer noch gleich dem Radius der ursprünglichen Kugel, d. h. das E-Kriterium ändert sich nicht. Dies ist für einen zweidimensionalen Fall in Anhang C („Anmerkungen zum E-Kriterium“, Seite 178) noch einmal verdeutlicht.

Ein weiterer Aspekt, der während der 3-D-Rekonstruktion berücksichtigt wird, ist die Erstellung der Dreiecksnetze, wie sie zur Berechnung der Verdeckung benötigt werden. Da hierzu die Kamerakoordinaten benötigt werden, ist es sinnvoll, bei der 3-D-Rekonstruktion der Punktwolke die Dreiecksnetze mit zu erstellen.

Die 3-D-Koordinaten der rekonstruierten Punkte, die zugehörige Kovarianzmatrix sowie das berechnete Dreiecksnetz wird jeweils an die Komponenten zur Ansichtenplanung weitergereicht.

5.2.2 Planung der nächsten besten Ansicht

Die nächste beste Ansicht wird auf der Basis der Zustandsschätzung in Form des geschätzten Zustands $\hat{z}_\tau$ und der Kovarianzmatrix Σ_τ gewählt. Wie oben erwähnt, werden dazu die zur Verfügung stehenden Ansichten sukzessive durchprobiert. Für jede Ansicht müssen die folgenden Schritte durchgeführt werden:

Für jeden 3-D-Punkt müssen zunächst die Sichtbarkeitswahrscheinlichkeiten p_{sb} für den Sichtbereich und p_{sv} für die Selbstverdeckung ermittelt werden. Die beiden Sichtbarkeitswahrscheinlichkeiten können zusammen multipliziert werden, um die Gesamtwahrscheinlichkeit zu erhalten, ob ein 3-D-Punkt in der Ansicht sichtbar ist. Anschließend kann die a posteriori Kovarianzmatrix für den Fall, dass

der Punkt beobachtet wird, berechnet werden. Wiederholt man dies für alle Punkte, so lässt sich für diese Ansicht das Optimierungskriterium (4.71) bestimmen.

Dies wird für alle zur Verfügung stehenden Ansichten berechnet. Die Ansicht, für die das Optimierungskriterium (4.71) das globale Minimum liefert, ist die nächste beste Ansicht, die anschließend dem Algorithmus zur 3-D-Rekonstruktion aus Abschnitt 5.2.3 übergeben wird.

5.2.3 3-D-Rekonstruktion

Die Aufnahmeposition, die die Komponente zur Ansichtenplanung ausgibt, wird anschließend vom Roboter angefahren, wobei wieder zwischen den Positionen erneut Aufnahmen gemacht werden. Die Zwischenaufnahmen werden in dieser Phase noch aus einem weiteren Grund benötigt: Der Punktverfolger kann Punkte „verlieren“, d. h. er findet einen verfolgten Punkt im entsprechend nächsten Bild nicht mehr. In diesem Fall werden parallel zwei Wege beschritten:

Erstens wird versucht, den entsprechenden 3-D-Punkt im Bild wiederzufinden. Dazu werden alle bisher rekonstruierten 3-D-Punkte, deren Punktkorrespondenzen verloren gegangen sind, in das aktuelle Bild projiziert. An der Stelle der Projektion wird der Punktverfolger für diesen Punkt noch einmal initialisiert, um die Punktkorrespondenz wiederzufinden. Dies stellt einen entscheidenden Zeitfaktor bei der Berechnung dar, wenn die Anzahl der verlorenen Punkte bzw. die Anzahl der Bilder sehr groß ist. Näheres dazu in den Diskussionen bei den entsprechenden Experimenten in Kapitel 6.

Zweitens wird, falls der verlorene Punkt nicht (sofort) wiedergefunden wird, ein neuer Merkmalspunkt im Bild detektiert, der ab diesem Zeitpunkt weiter verfolgt wird. Sind mindestens zwei Aufnahmen (Beobachtungen) eines neu verfolgten 3-D-Punkts vorhanden, so wird mittels Triangulation eine initiale Schätzung berechnet und dieser Punkt in die bisherige Schätzung integriert. In diesem Fall ändert sich die Dimension des Zustands (und der Beobachtung) im Kalman-Filter. Da jedoch alle Koordinaten eines 3-D-Punkts mit einem eigenen Kalman-Filter geschätzt werden können (vgl. Abschnitt 4.2.2), stellt die Erweiterung kein Problem dar: Die Schätzung der 3-D-Koordinaten des neuen Punkts werden einfach unten an den Zustandsvektor $\hat{z}$ angehängt, und die Kovarianz wird durch einen 3×3 Block erweitert, in dem die initiale Kovarianz des neuen Punkts steht. Dadurch kann sich im Laufe der Rekonstruktion die Dimension des Zustandsvektors erheblich erhöhen, denn der verlorene Bildpunkt bleibt ebenfalls im Zustandsvektor erhalten.

Die gefundenen Punktkorrespondenzen an der Endposition werden mit Hilfe des Kalman-Filters in die aktuelle Schätzung integriert, um diese aufzufrischen. Wie schon bei der initialen Schätzung werden zum Abschluss dieses Schritts zusätzlich die Dreiecksnetze berechnet, wie sie für die Berechnung der Sichtbarkeit notwendig sind.

Die beiden Komponenten aus den Abschnitten 5.2.2 und 5.2.3 wechseln sich ab, bis ein Abbruchkriterium erreicht wird. Abbruchkriterien können sein: eine maximale Anzahl von Aufnahmen, oder eine gewünschte Genauigkeit, die erreicht werden soll (wobei im Allgemeinen nur der erwartete quadratische Schätzfehler Aussagen über die Genauigkeit erlaubt, da die 3-D-Struktur ja nicht bekannt ist).

5.3 Das Gesamtsystem

In Bild 5.3 ist noch einmal der komplette Ablauf der Ansichtenplanung zur 3-D-Rekonstruktion mit dem probabilistischen Ansatz aus Kapitel 4 graphisch dargestellt. Die einzelnen Schritte werden im Folgenden noch einmal ausführlich erläutert:

1. Zunächst bietet es sich an, die Denavit-Hartenberg-Matrix (vgl. Abschnitt 2.4) aufzustellen. Diese beschreibt das Bewegungsmodell, und damit die verfügbaren Aufnahmepositionen, die bei der Optimierung in Betracht gezogen werden können. Unter Umständen wird die Denavit-Hartenberg-Matrix benötigt, um die Hand-Auge-Kalibrierung durchzuführen, nämlich genau dann, wenn der Roboterarm keine direkten Daten über Position und Orientierung des Endeffektors zurück liefert. Diese Möglichkeit ist bei den zwei verwendeten Robotern vorhanden.
2. Die Kamerakalibrierung (vgl. Abschnitt 2.2) ist notwendig, um die in- und extrinsischen Parameter zu bestimmen. Die extrinsischen sind anschließend für die Hand-Auge-Kalibrierung notwendig.
3. Die Hand-Auge-Kalibrierung ist wichtig, damit aus der Position der Hand die Position der Kamera (ohne Kalibriermuster) bestimmt werden kann (vgl. Abschnitt 2.3). Damit sind die extrinsischen Parameter ermittelbar und die Parameter der Beobachtungsfunktion b des Kalman-Filter definiert.
4. Nun kann eine manuell definierte Sequenz von Bildern aufgenommen werden und
5. Mit Hilfe der Punktverfolgung können korrespondierende Punkte gefunden werden. Mit den Kameraparametern aus der Kalibrierung kann eine erste Schätzung der 3-D-Geometrie erstellt werden (vgl. Abschnitt 5.2.1).
6. Basierend auf der aktuellen Schätzung der 3-D-Geometrie wird die nächste beste Ansicht geplant (vgl. Abschnitt 5.2.2).
7. Die geplante Ansicht wird angefahren.
8. Die neue Ansicht wird zur Verbesserung der 3-D-Rekonstruktion verwendet. Dazu wird das Kalman-Filter benutzt (vgl. Abschnitt 5.2.3). Der Algorithmus stoppt hier, falls ein Abbruchkriterium erreicht wird.

Im folgenden Kapitel 6 wird der hier vorgestellte probabilistische Ansatz in realen Experimenten evaluiert. Zusätzlich erfolgt in diesem Kapitel die experimentelle Evaluation der geometrischen Überlegungen aus Abschnitt 3.2.

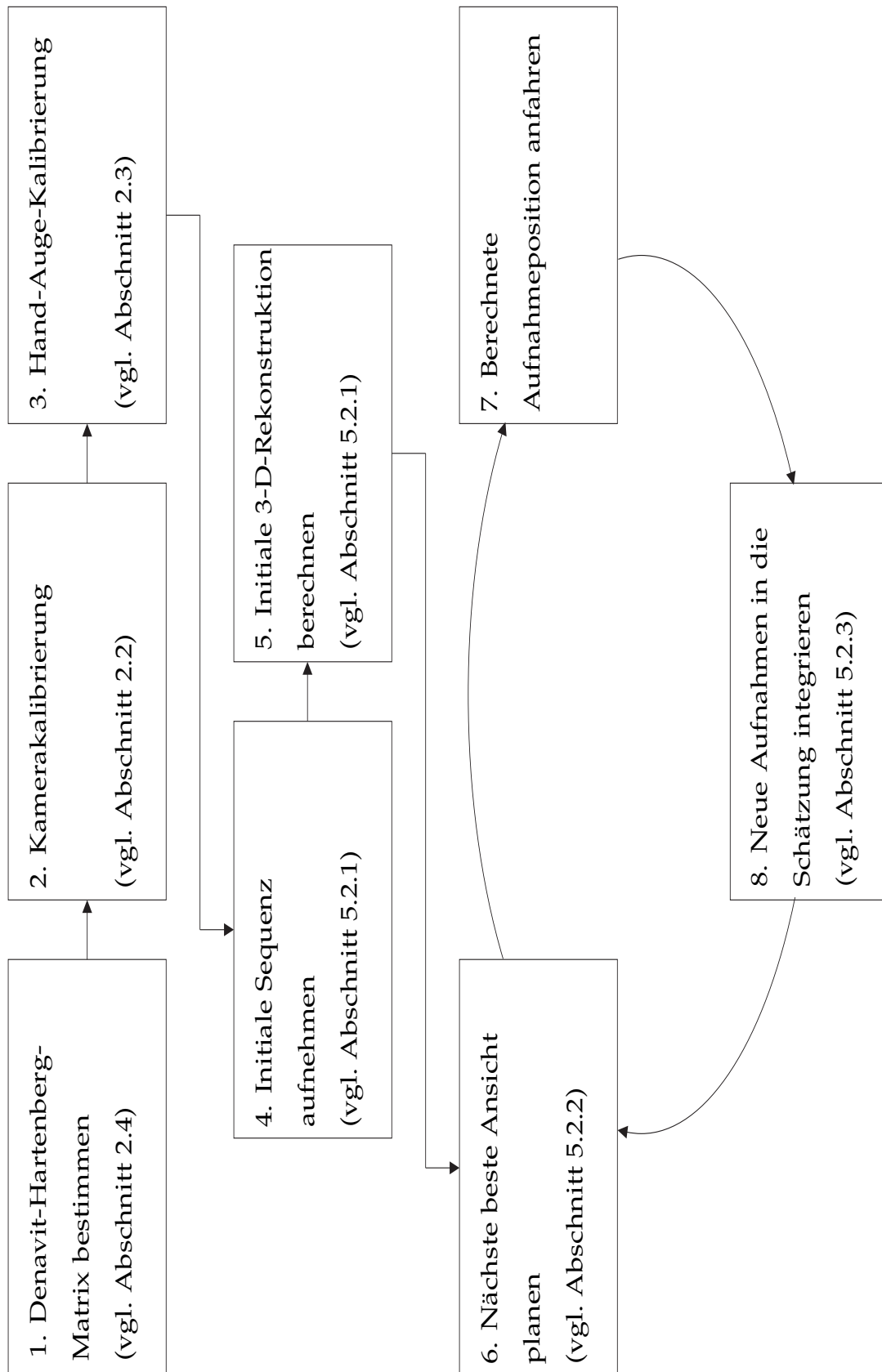


Bild 5.3 Überblick über das Gesamtsystem mit erforderlichen initialen Schritten. Die Schleife am Ende wird verlassen, sobald ein Abbruchkriterium erreicht wird.

Kapitel 6

Experimente

In diesem Kapitel werden die Aussagen über Stereosysteme aus Abschnitt 3.2 experimentell verifiziert. Darüber hinaus soll der probabilistische Ansatz aus Kapitel 4 in verschiedenen Versuchen überprüft werden.

Im Abschnitt 6.1 werden zunächst Gütekriterien vorgestellt, mit denen 3-D-Rekonstruktionen miteinander verglichen werden können. Grundidee ist dabei immer, den 3-D-Fehler zwischen der Rekonstruktion und der wahren 3-D-Geometrie zu messen.

Die Aussagen über Stereosysteme werden anschließend in Abschnitt 6.2 untersucht. Da hier Selbstverdeckungen nicht im Gütekriterium modelliert werden können, erfolgen die Experimente mit relativ einfachen Prüfkörpern.

In den Abschnitten 6.3 und 6.4 werden verschiedene Experimente für den probabilistischen Ansatz aus Kapitel 4 vorgestellt. Zunächst werden dazu in Abschnitt 6.3 Versuche für einfache Objekte ohne die Sichtbarkeitseinschränkungen gemäß Abschnitt 4.4 durchgeführt. So wird zunächst die allgemeine Idee des probabilistischen Ansatzes verifiziert. Anschließend werden im Abschnitt 6.4 Experimente an komplexeren Objekten vorgestellt. Hier kommt den Sichtbarkeitsbedingungen ein entscheidender Einfluss zu.

Das Kapitel schließt mit einer Diskussion der Ergebnisse in Abschnitt 6.5.

6.1 Gütekriterien der Rekonstruktionen

Um zu entscheiden, wie gut die Ansichtenplanung die Rekonstruktionsgenauigkeit erhöht, benötigt man ein Gütekriterium, das es erlaubt die Genauigkeit der Rekonstruktion zu bestimmen. Die Einführung von Gütekriterien wird hier nötig, da im Allgemeinen die „Ground-Truth“-Daten der zu rekonstruierenden Punkt nicht bekannt sind (Ausnahme: Kalibriermuster). Dies liegt daran, dass die rekonstruierten Punkte durch einen Punktverfolgungsalgorithmus bestimmt werden. Dabei ist im Allgemeinen nicht vorhersehbar, welche Punktmerkmale vom Punktverfolger gefunden und verfolgt werden können. Deswegen werden hier für verschiedene Kategorien von Objekten indirekte Gütemaße vorgeschlagen, die im Wesentlichen die Lage der rekonstruierten 3-D-Punkte zueinander bewerten. Zum Beispiel wird bei einem planaren Objekt bewertet, wie gut die rekonstruierten Punkte in einer Ebene liegen.

Anmerkung zu D-, modifizierter E- und T-Optimalität, als Gütemaß der Rekonstruktion: Für die Bewertung der Güte einer Ansicht eignen sich diese Kriterien, da sie a priori, d. h. ohne eine Beobachtung zu machen, berechnet werden können. Im Gegensatz dazu ist für die Bewertung der Güte einer 3-D-Rekonstruktion nicht die

Kovarianz, sondern vielmehr die Lage der rekonstruierten 3-D-Punkte interessant. Die Lage eines Punktes wird dabei durch seinen Erwartungswertvektor beschrieben. Die unten eingeführten Gütemaße bewerten deshalb den Erwartungswertvektor. Sie können jedoch im Vergleich zu D-, modifizierter E- und T-Optimalität nicht direkt optimiert werden, da die Berechnung des Erwartungswertvektors nur möglich ist, wenn die Beobachtung in die Schätzung integriert wird. Deshalb ist es im probabilistischen Ansatz nicht möglich die Schätzung des Erwartungswerts a priori zu optimieren. Dies ist der Grund warum für die Ansichtenplanung D-, modifizierte E- und T-Optimalität, für die Bewertung der Güte der Rekonstruktion jedoch die untenstehenden Gütemaße, definiert werden.

Für verschiedene Prüfkörper sind unterschiedliche Herangehensweisen für das Gütemaß nötig, die im Folgenden kurz vorgestellt werden.

6.1.1 Kalibriermuster

Das Kalibriermuster ist das einfachste Prüfobjekt: Das Korrespondenzproblem lässt sich sehr einfach lösen, wenn man ein asymmetrisches Kalibriermuster verwendet. In diesem Fall kann die Güte der 3-D-Rekonstruktion direkt abgelesen werden, da das Kalibriermuster das Weltkoordinatensystem festlegt und deshalb die Koordinaten der Kalibrierpunkte im Weltkoordinatensystem bekannt sind. Als Gütemaß der 3-D-Rekonstruktion kann hier also einfach der mittlere euklidische Abstand der rekonstruierten 3-D-Punkte zum jeweiligen Referenzpunkt des Kalibrierusters verwendet werden. Dieser Fehler wird mit e bezeichnet.

6.1.2 Objekte mit einer planaren Fläche

Besteht der Prüfkörper nur aus einer planaren Fläche, so sollten die rekonstruierten Punkte ebenfalls in einer planaren Fläche liegen. Hier kann die Güte der Rekonstruktion mit Hilfe einer Regressionsebene bestimmt werden: Durch die rekonstruierten Punkte wird eine Regressionsebene mit der Methode der kleinsten Fehlerquadratsumme gelegt. Dies wird folgendermaßen realisiert:

Es ist offensichtlich, dass die Regressionsebene durch den Schwerpunkt der 3-D-Punkte $\mathbf{p}_{wi}$, $i = 1, \dots, n$, verläuft. Zieht man den Schwerpunkt der 3-D-Punkte von allen Koordinaten ab, so verläuft die so modifizierte Regressionsebene durch den Ursprung. Die modifizierten 3-D-Punkte werden mit $\tilde{\mathbf{p}}_{wi}$ bezeichnet. Mit dem Normalenvektor $\mathbf{n}$ gilt

$$\mathbf{n} \tilde{\mathbf{p}}_{wi} = 0, \quad \forall \tilde{\mathbf{p}}_{wi}, \quad i = 1, \dots, n. \quad (6.1)$$

Um bei verrauchten Punktkoordinaten $\tilde{\mathbf{p}}_{wi}$ den optimalen Normalenvektor der Regressionsebene $\mathbf{n}_{\text{opt}} \neq \mathbf{0}$ zu finden, werden die Gleichungen aus (6.1) umgeformt zu dem Minimierungsproblem:

$$\mathbf{n}_{\text{opt}} = \underset{\mathbf{n} \in \mathbb{R}^3, |\mathbf{n}|=1}{\operatorname{argmin}} \left\| \underbrace{\begin{pmatrix} \tilde{\mathbf{p}}_{w1}^T \\ \tilde{\mathbf{p}}_{w2}^T \\ \vdots \\ \tilde{\mathbf{p}}_{wn}^T \end{pmatrix}}_{=:A} \mathbf{n} \right\|. \quad (6.2)$$

Die Nebenbedingung $|\mathbf{n}| = 1$ wird gesetzt, um den Trivialfall $\mathbf{n} = \mathbf{0}$ auszuschließen. Mit der Matrix $\mathbf{A}$, wie in (6.2) definiert, lässt sich das Problem in die Form

$$\mathbf{n}_{\text{opt}} = \underset{\mathbf{n} \in \mathbb{R}^3, |\mathbf{n}|=1}{\operatorname{argmin}} \mathbf{n}^T \mathbf{A}^T \mathbf{A} \mathbf{n} \quad (6.3)$$

bringen. Dieses Problem lässt sich mit Hilfe von Lagrange-Multiplikatoren sehr einfach lösen [Mey01]: Der Eigenvektor, der zum kleinsten Eigenwert der Matrix $\mathbf{A}$ gehört, minimiert (6.2) und der Eigenwert ist dabei der Wert des Minimums.

Bei Objekten, die nur aus einer planaren Fläche bestehen, muss also zur Bestimmung des Rekonstruktionsfehlers e_R nur das zugehörige Eigenwertproblem gelöst werden.

6.1.3 Objekte mit mehreren planaren Flächen

Bei Objekten, die aus mehreren planaren Flächen bestehen, funktioniert der Ansatz über Regressionsebenen nicht, da im Allgemeinen unbekannt ist, welche rekonstruierten Punkte zu welcher Ebene gehören. Deshalb muss das Ebenenmodell in die rekonstruierte Punktwolke einbeschrieben werden. Dazu ist es notwendig, die starre Transformation zwischen dem Weltkoordinatensystem der rekonstruierten Punktwolke und dem Modell zu schätzen. Dies lässt sich mit einem „Iterative Closest Point (ICP)“-Algorithmus [Zin02, Zin05] effizient realisieren:

Aus dem Ebenen-Modell wird eine Punktwolke generiert. Der ICP-Algorithmus berechnet eine starre Transformation zwischen den beiden Punktwolken, so dass der durchschnittliche Abstand zwischen je zwei Punkten der beiden Punktwolken minimal wird. Die starre Transformation besteht dabei aus Rotation und Translation und lässt sich anschließend direkt auf das Ebenenmodell anwenden.¹

Im so transformierten Ebenenmodell kann man nun den Abstand eines rekonstruierten Punkts $\tilde{\mathbf{p}}_{wi}$ zu allen Ebenen berechnen. Dabei wird angenommen, dass der Punkt zu jener Ebene gehört, zu der er minimalen Abstand hat.

Das arithmetische Mittel der Distanz aller Punkte zu ihrer jeweiligen am nächsten liegenden Ebene bildet dabei das Maß, mit dem die Güte der Rekonstruktion gemessen wird. Zur Unterscheidung zum 3-D-Fehler e wird dieser Fehler mit e_p bezeichnet.

6.1.4 Objekte mit nichtplanaren Flächen

Hierzu muss die Geometrie des Objekts bekannt sein, um wirklich einen 3-D-Fehler messen zu können. Die Geometrie ist z. B. bei einem definierten Prüfkörper bekannt oder kann durch andere Messverfahren, wie beispielsweise einer Streifenprojektion, bekannt sein.

Mithilfe der geometrischen Beschreibung des Objekts lässt sich eine Punktwolke aus dem Modell erstellen und mit einem ICP-Algorithmus zur Deckung bringen. So lässt sich die Rotation und Translation zwischen den beiden Punktwolken bestimmen. Anschließend kann man den Abstand zwischen den Punktwolken bestimmen. Dazu wird das in [Mun07] vorgeschlagene Verfahren verwendet:

¹Der Vollständigkeit wegen sei erwähnt, dass ICP-Algorithmen auch die Skalierung schätzen können. Durch die Kamerakalibrierung ist die Rekonstruktion metrisch und die Skalierung bekannt und somit braucht eine Skalierungsschätzung nicht durchgeführt werden.

Für jeden rekonstruierten Punkt wird der Abstand zu allen Punkten der Referenzwolke berechnet. Dabei wird angenommen, dass der rekonstruierte Punkt jenem Punkt der Referenzwolke entspricht, bei dem der Abstand minimal ist. Der mittlere Gesamtfehler für solche Objekte wird mit e_{np} bezeichnet.

6.2 Evaluierung des geometrischen Ansatzes mit Stereosystemen

In diesem Abschnitt werden die in Abschnitt 3.2 getroffenen Aussagen für die Kameraparameter in Stereosystemen experimentell überprüft. Analog zu Abschnitt 3.2 werden dabei zunächst die einzelnen, gezielt veränderbaren Parameter Brennweite, Stereobasis, Translation in alle Richtungen und Rotation untersucht. Zum Abschluss werden alle Parameter gleichzeitig optimiert.

Als Testobjekte werden verschiedene Testkörper verwendet: ein Kalibriermuster (siehe Bild 6.2) und ein L-förmiges Objekt (siehe Bild 6.9). Die Testkörper stellen unterschiedliche Anforderungen an den Algorithmus. Durch Verwendung des Kalibrierusters sind die in- und extrinsischen Kameraparameter sehr einfach zu ermitteln. Zusätzlich erlaubt es eine direkte Berechnung der Rekonstruktionsgüte, denn hier können die rekonstruierten mit den tatsächlichen Weltpunkten verglichen werden. Beim L-förmigen Objekt können diese Daten nicht mehr aus den Bildern generiert werden. Hier müssen die intrinsischen Kameraparameter in einem vorherigen Schritt ermittelt werden. Die extrinsischen Kameraparameter werden über eine Hand-Auge-Kalibrierung ermittelt und die Punktkorrespondenzen werden über einen Punktverfolger (vgl. Abschnitt 2.5.1) berechnet. Das L-förmige Objekt stellt ein Objekt dar, das aus zwei planaren Flächen besteht, die senkrecht aufeinander stehen. Damit kann gezeigt werden, dass die Vorhersagen nicht nur genau dann gelten, wenn die Punkte alle in einer Ebene liegen. Um das Problem der Selbstverdeckung zu umgehen, wird das L-förmige Objekt immer nur aus den Blickrichtungen betrachtet, aus denen immer genau zwei Seiten zu sehen sind.

6.2.1 Ergebnisse Kalibriermuster

6.2.1.1 Optimierung der Brennweite

Bei diesen Experimenten werden zunächst an der Position der linken Kamera alle Aufnahmen mit den entsprechenden Brennweiten f_l gemacht. Anschließend wurde der Roboterarm mittels Linearschlitten zur Position der rechten Kamera bewegt, um dort die Aufnahmen mit den verschiedenen Brennweiten f_r aufzuzeichnen. Die Ergebnisse sind in Tabelle 6.1 dargestellt, dabei ist die effektive Brennweite in x -Richtung angegeben (vgl. Abschnitt 2.1.2). Einen kurzen Ausschnitt aus der Bildsequenz sieht man in Bild 6.2.

Die Ergebnisse aus Tabelle 6.1 sind nicht symmetrisch. Dies hat im Wesentlichen zwei Gründe: Zum einen ist die Verschiebung zwischen linker und rechter Position entlang des Linearschlittens nicht perfekt symmetrisch bezüglich der Lage des Kalibrierusters, d. h. der Abstand zwischen dem Mittelpunkt des Kalibrierusters und linker bzw. rechter Kameraposition ist ungleich: Bezüglich des Weltkoordinatensystems, das durch das Kalibriermuster festgelegt wird, sind die Translati-

$f_r =$	760	814	910	1035	1172	1352	1575	1851	2225
$f_l = 760$	0.99	0.78	0.77	0.67	0.69	0.71	0.61	0.65	0.62
$f_l = 814$	1.40	0.91	0.73	0.76	0.70	0.74	0.67	0.67	0.60
$f_l = 910$	1.25	0.93	0.66	0.59	0.64	0.60	0.56	0.60	0.53
$f_l = 1035$	1.09	0.80	0.64	0.57	0.54	0.55	0.48	0.50	0.40
$f_l = 1172$	1.15	0.81	0.69	0.54	0.51	0.57	0.47	0.47	0.43
$f_l = 1352$	1.09	0.78	0.58	0.49	0.48	0.47	0.38	0.40	0.34
$f_l = 1575$	1.08	0.81	0.62	0.49	0.48	0.47	0.44	0.39	0.34
$f_l = 1851$	1.21	0.79	0.59	0.48	0.52	0.47	0.41	0.43	0.34
$f_l = 2225$	1.06	0.78	0.62	0.51	0.48	0.46	0.39	0.39	0.32

Tabelle 6.1 Der 3-D-Fehler der Rekonstruktion des Kalibrierusters in Abhängigkeit der Brennweiten f_l der linken bzw. f_r in Pixel der rechten Kamera: Große Brennweiten ergeben kleine Fehler.

onsvektoren von linker bzw. rechter Kameraanordnung, bei einer Verschiebung in y -Richtung:

$$\begin{aligned} t_l &= (177,7 \text{ mm} \quad 41,8 \text{ mm} \quad -782,8 \text{ mm}), \\ t_r &= (182,0 \text{ mm} \quad -30,5 \text{ mm} \quad -832,2 \text{ mm}). \end{aligned}$$

Zum anderen ist die Detektion der Punkte des Kalibrierusters auch fehlerbehaftet und dieser Fehler ist bei jeder Aufnahmen unterschiedlich, d. h. selbst bei perfekt symmetrischem Aufbau, würde sich nicht unbedingt eine symmetrische Tabelle der Fehler ergeben.

In Bild 6.1 ist die Abhängigkeit des 3-D-Fehlers noch einmal graphisch dargestellt. Man beachte dabei, dass die gewählten Brennweiten entgegen der Darstellung in Bild 6.1 nicht äquidistant sind. Die Aufnahmen wurden bei folgenden Zoomwerten des Stellmotors des Objektivs der Kamera gemacht: 40, 100, 200, 300, 400, 500, 600, 700, 800. Diese sind bis auf den ersten Wert äquidistant, jedoch ergeben sich nach der Kalibrierung nicht äquidistante Abstände für die effektiven Brennweiten in Pixeln. Bild 6.1 bestätigt, abgesehen von einigen Ausreißern, die Aussage, dass der 3-D-Fehler abnimmt, falls die Brennweite zunimmt. Der optimale Wert liegt dabei bei $e = 0.325 \text{ mm}$.

6.2.1.2 Optimierung der Translation

Bei der Translation wird der Arm des Scrobot ER VII so positioniert, dass die Kamera senkrecht auf das Kalibriermuster gerichtet ist, d. h. die Normale der Ebene des Kalibrierusters verläuft parallel zur optischen Achse der Kamera. Zunächst wird mit Hilfe des Linearschlittens die Kamera parallel zur x -Achse im Bild bewegt, so dass hier die Translation in x -Richtung, d. h. der Einfluss der Stereobasis, untersucht werden kann. Erreicht der Linearschlitten das andere Ende seines Definitionsbereichs, so wird er wieder in die Ausgangslage zurück gefahren. Anschließend wird eine Translation in z -Richtung, also auf das Kalibriermuster zu, durchgeführt. Auf diese Weise wird ein definiertes Gitter abgefahren. Jede Gitterposition wird durch eine Bildnummer repräsentiert, entsprechend der Tabelle 6.2. Da sich bei Annäherung an das Kalibriermuster der Sichtbereich verkleinert, ist

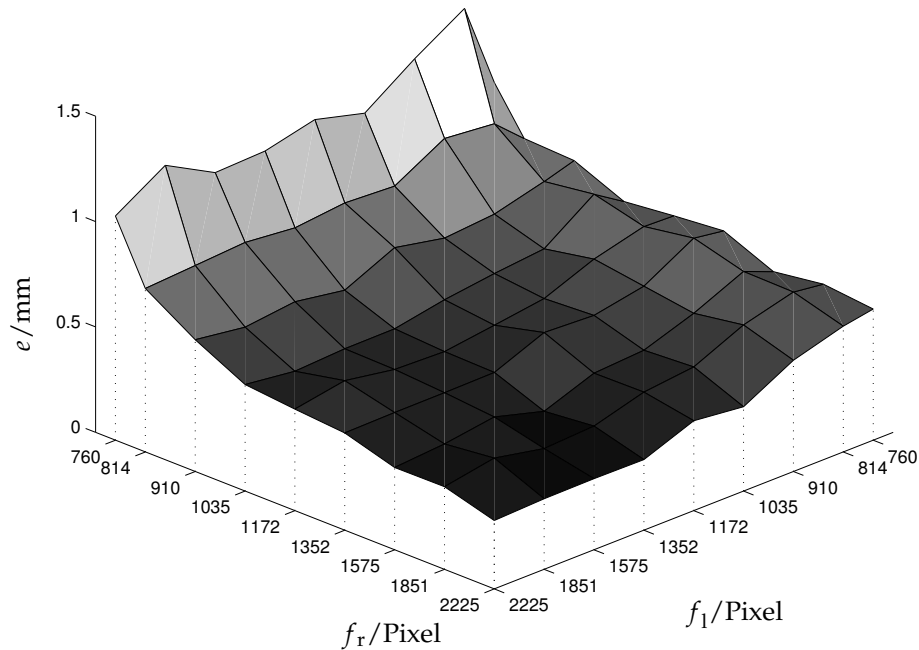


Bild 6.1 3-D-Fehler e in mm der Rekonstruktion des Kalibrierusters in Abhängigkeit der Brennweiten f_l der linken bzw. f_r der rechten Kamera. Man sieht: große Brennweiten ergeben kleine Fehler.

bei kleinen Abständen in z -Richtung das Kalibriermuster u. U. nicht mehr komplett im Bild. Solche Positionen werden nicht berücksichtigt; sie sind in Tabelle 6.2 durch einen „*“ gekennzeichnet.

In Bild 6.3 ist in Abhängigkeit der Bildnummer der 3-D-Fehler der Triangulation dargestellt. Dabei gilt: Ein dunkler Grauwert entspricht einem niedrigen Fehler. Bei der Analyse des Bilds ist festzustellen, dass eine Symmetrieachse existiert, nämlich die Hauptdiagonale. Dies ist leicht zu plausibilisieren, denn schließlich ist das Problem symmetrisch, da die Triangulation aus den Bildern Nummern i und j den gleichen Fehler wie die Triangulation aus den Bildern Nummer j und i ergibt. Die Werte auf der Hauptdiagonalen selbst sind hier auf Null gesetzt, da dies einer Triangulation von zwei identischen Kameraposition entspricht, was für die Triangulation ungeeignet ist.

Es fallen die waagerechten und senkrechten Kanten im Bild auf. Diese verlaufen genau an den Stellen, an denen der Linearschlitten wieder zurück auf die Ausgangsposition fährt. Durch das Zurückstellen ändert sich der Fehler entsprechend stark und es entsteht der optische Eindruck einer Kante.

Weiterhin sind die hellen Nebendiagonalen in Bild 6.3 auffällig. Betrachtet man die 1. Pixelzeile im Querschnitt (vgl. Bild 6.4), so erkennt man, dass der Fehler bis zur Bildnr. 21 abnimmt und danach wieder sprunghaft nach oben steigt. Zwischen Position 21 und 22 fährt der Linearschlitten des Scrobot ER VII wieder an seine Ausgangsposition zurück und positioniert die Kamera nun näher am Kalibriermuster (vgl. Tabelle 6.2). Für diese beiden Positionen gilt, dass die Stereobasis sehr gering ist, was einen hohen Fehler nach sich zieht. Überall dort, wo die Positionen der Kameras zu fast parallelen Sichtstrahlen führen, ist der Fehler sehr groß, d. h. dort, wo sich die zweite Kamera quasi in die Sichtstrahlen der ersten Kamera stellt.

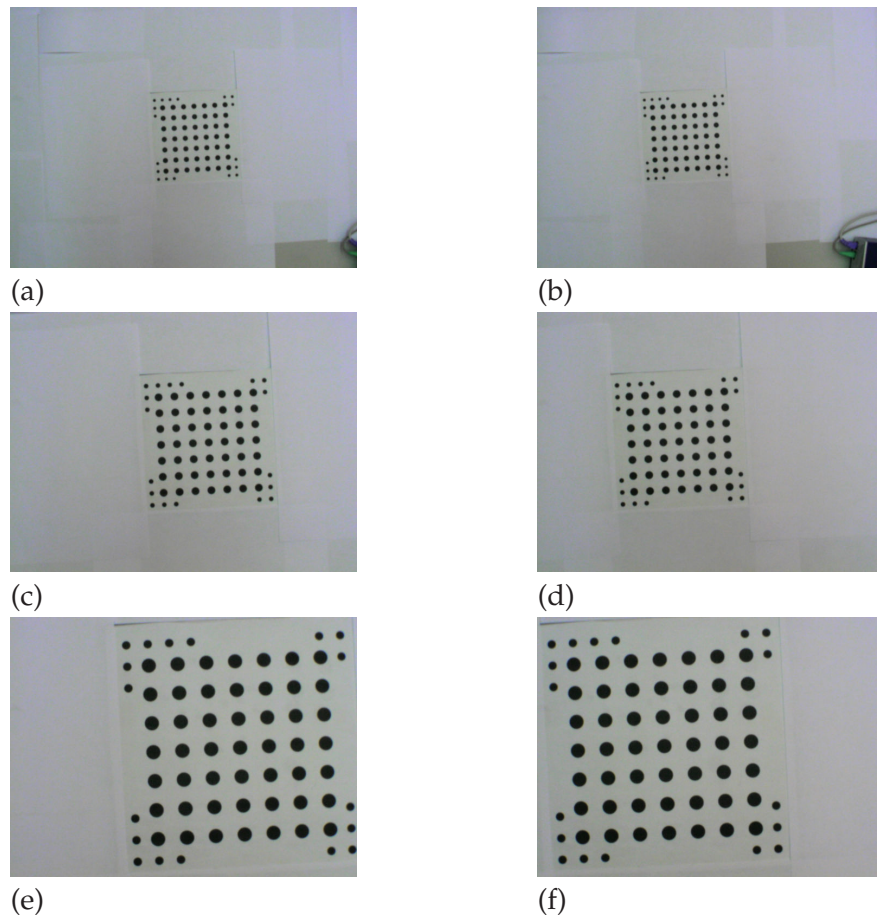


Bild 6.2 Kalibriermuster mit Brennweiten 760, 1172 bzw. 2225 Pixel in der 1., 2. bzw. 3. Zeile für die linke Kamera in der linken Spalte bzw. für die rechte Kamera in der rechten Spalte. Die Höhe des optischen Zentrums über dem Kalibriermuster beträgt links ca. 780 mm, rechts ca. 830 mm, der Abstand zwischen linker und rechter Kamera beträgt ca. 72 mm

Dies erklärt die Nebendiagonalen, die parallel zur Hauptdiagonalen laufen. Man erkennt an den waagerechten und senkrechten Kanten gelegentlich einen kleinen Versatz in den Diagonalen. Dieser ergibt sich durch den nicht perfekt symmetrischen Versuchsaufbau.

Der optimale Wert von $e = 0.225$ mm wird bei den Kamerapositionen 22 und 83 erreicht. Es wird also in dieser Konstellation nicht der volle z -Bereich ausgenutzt. Jedoch ist der Verlauf der Fehlerfunktion dort sehr flach, d. h. die Messungenauigkeiten übersteigen in diesen Bereichen die Maßnahmen, die durch die Optimierung erreicht werden. Dass sich hier nur kleine Unterschiede des Rekonstruktionsfehlers bei unterschiedlichen z -Werten ergeben, lässt sich in Bild 6.4 ablesen.

6.2.1.3 Optimierung der Rotation

Wie bereits bei der Untersuchung der Brennweite werden hier wiederum zwei feste Positionen des Roboterarms, an dem die Kamera befestigt ist, gewählt. Anschließend werden mit Hilfe des obersten Rotationsgelenks des Scorbot ER VII verschiedene Rotationswinkel eingestellt. An dieser Stelle sei kurz darauf hingewiesen, dass aufgrund der Vorrichtung zur Befestigung der Kamera an der Ro-

	z=85	z=80	z= 75	z= 70	z=65	z= 60	z= 55	z= 50	z=45
x=-20	1	22	43	64	85	*	*	*	*
x=-18	2	23	44	65	86	106	*	*	*
x=-16	3	24	45	66	87	107	126	*	*
x=-14	4	25	46	67	88	108	127	143	*
x=-12	5	26	47	68	89	109	128	144	*
x=-10	6	27	48	69	90	110	129	145	158
x=-8	7	28	49	70	91	111	130	146	159
x=-6	8	29	50	71	92	112	131	147	160
x=-4	9	30	51	72	93	113	132	148	161
x=-2	10	31	52	73	94	114	133	149	162
x=0	11	32	53	74	95	115	134	150	163
x=2	12	33	54	75	96	116	135	151	164
x= 4	13	34	55	76	97	117	136	152	165
x=6	14	35	56	77	98	118	137	153	166
x=8	15	36	57	78	99	119	138	154	167
x=10	16	37	58	79	100	120	139	155	168
x=12	17	38	59	80	101	121	140	156	*
x=14	18	39	60	81	102	122	141	157	*
x=16	19	40	61	82	103	123	142	*	*
x=18	20	41	62	83	104	124	*	*	*
x= 20	21	42	63	84	105	125	*	*	*

Tabelle 6.2 Zuordnung der x-z-Position zur Bildnummer. Die Positionsangaben der x- bzw. z- Komponente sind in cm gemessen. Liegt das Kalibriermuster nicht vollständig im Bild, so ist die entsprechende Position mit einem * gekennzeichnet.

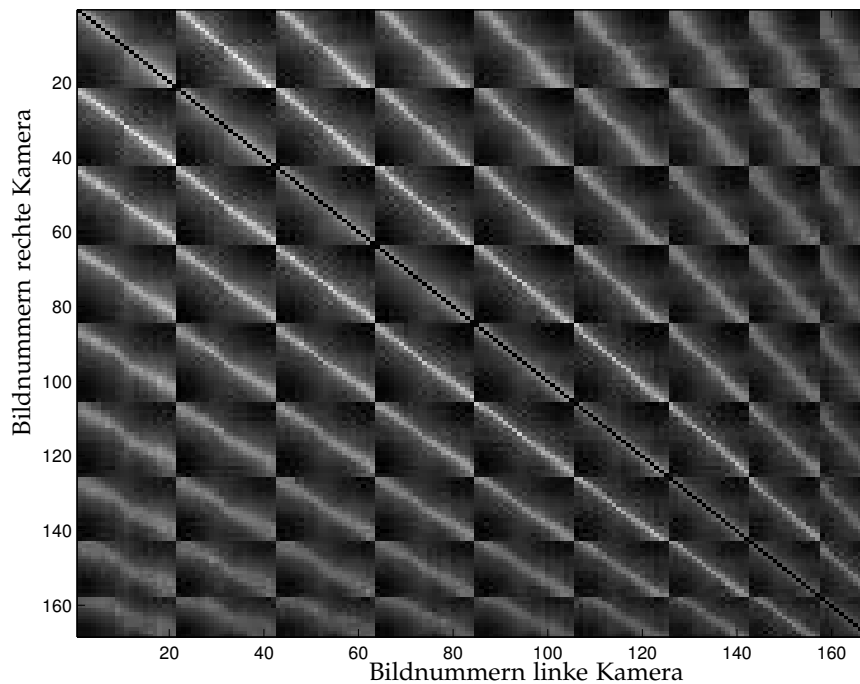


Bild 6.3 Rekonstruktion eines Kalibrierusters in Abhängigkeit der Translation. Die Graustufen im Bild signalisieren den 3-D-Fehler, wobei ein heller Grauwert einen großen Fehler bedeutet.

boterspitze die Drehachse des Roboters nicht exakt durch das optische Zentrum der Kamera geht. Über die gesamte Rotation von 12° ergibt sich dabei eine Verschiebung von ca. 4 mm. Da der Abstand von linker und rechter Kameraposition

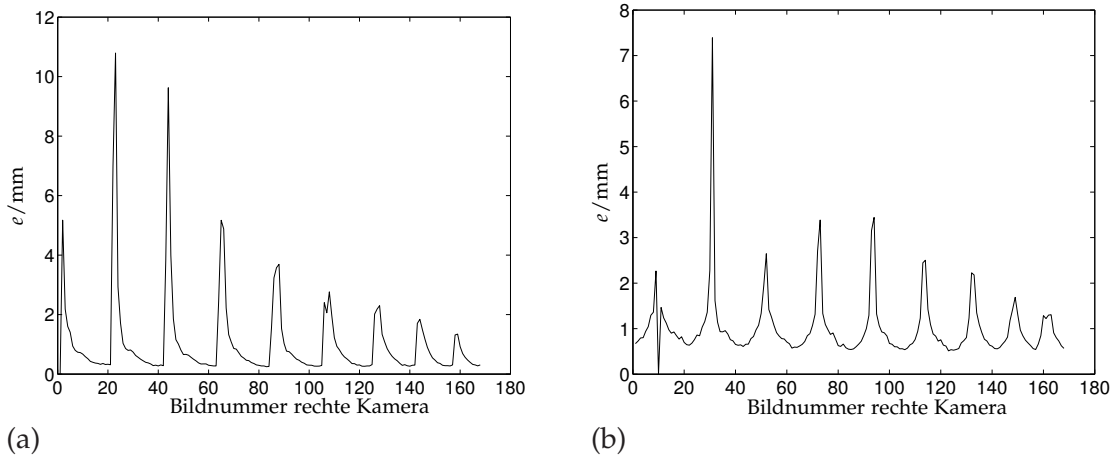


Bild 6.4 Rekonstruktionsfehler e in mm in Abhängigkeit der Bildnummer der rechten Kamera für die linke Kamera mit Bildnummer (a) 1 bzw. (b) 10. An der Stelle, wo die Bildnummer der linken und rechten Kamera identisch ist, wurde der Fehlerwert auf Null gesetzt.

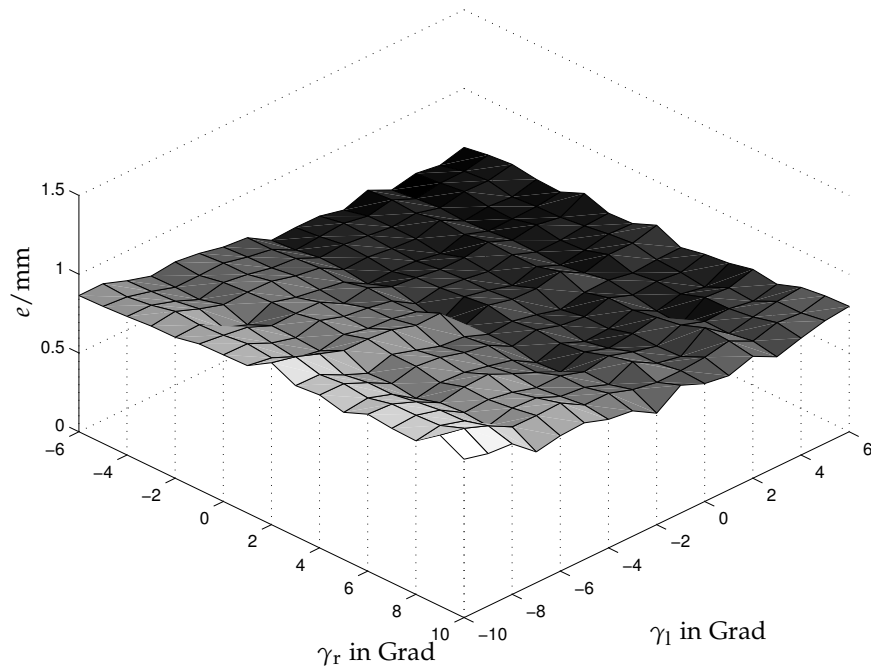


Bild 6.5 Darstellung des Rekonstruktionsfehlers e in mm bei der Rekonstruktion eines Kalibriermusters in Abhängigkeit des Rotationswinkels von linker und rechter Kamera. Die linke Kamera ist bei $+6^\circ$ vom Objekt weggedreht, die rechte bei -6° .

20 cm beträgt und sich die Kameras dabei ca. 85 cm vom Kalibriermuster entfernt befinden, wird dieser Translationsanteil vernachlässigt.

In Bild 6.5 ist die Abhängigkeit des Rotationswinkels der linken und rechten Kamera dargestellt. Dabei ist die linke Kamera bei $+6^\circ$ weggedreht, die rechte bei -6° . Das absolute Minimum befindet sich bei $\gamma_l = +3^\circ$ und $\gamma_r = -6^\circ$ mit einem Wert von $e = 0.577$ mm (vgl. Tabelle 6.3). Anhand der Abbildung ist zu erkennen, dass sich in dem Bereich, wo beide Kameras weggedreht sind, im Allgemeinen ein niedrigerer Rekonstruktionsfehler ergibt als bei anderen Konstellationen.

$\gamma_r =$	-6°	-5°	-4°	-3°	-2°	-1°	0°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
$\gamma_l = -10^\circ$	0.865	0.870	0.879	0.892	0.895	0.926	0.931	0.935	1.028	0.956	0.970	0.934	0.995	0.979	0.974	1.083	1.005
$\gamma_l = -9^\circ$	0.866	0.856	0.877	0.912	0.900	0.907	0.927	0.911	0.992	0.953	0.968	0.951	0.987	0.989	0.997	1.053	0.966
$\gamma_l = -8^\circ$	0.811	0.818	0.829	0.855	0.808	0.885	0.859	0.837	0.937	0.877	0.902	0.902	0.928	0.948	0.954	0.996	0.955
$\gamma_l = -7^\circ$	0.766	0.770	0.766	0.807	0.753	0.821	0.818	0.809	0.885	0.814	0.821	0.801	0.847	0.859	0.848	0.915	0.830
$\gamma_l = -6^\circ$	0.784	0.772	0.800	0.812	0.822	0.824	0.854	0.839	0.883	0.859	0.858	0.828	0.874	0.868	0.888	0.913	0.858
$\gamma_l = -5^\circ$	0.763	0.776	0.777	0.809	0.780	0.822	0.806	0.802	0.870	0.813	0.840	0.819	0.881	0.840	0.878	0.919	0.867
$\gamma_l = -4^\circ$	0.730	0.741	0.745	0.768	0.755	0.741	0.786	0.791	0.823	0.764	0.798	0.786	0.810	0.795	0.814	0.834	0.815
$\gamma_l = -3^\circ$	0.716	0.716	0.731	0.745	0.729	0.767	0.757	0.762	0.826	0.750	0.770	0.749	0.745	0.798	0.798	0.834	0.795
$\gamma_l = -2^\circ$	0.626	0.639	0.664	0.661	0.644	0.700	0.666	0.627	0.672	0.693	0.691	0.661	0.673	0.750	0.743	0.783	0.707
$\gamma_l = -1^\circ$	0.633	0.650	0.673	0.667	0.650	0.684	0.692	0.710	0.730	0.702	0.721	0.710	0.728	0.776	0.813	0.756	0.810
$\gamma_l = 0^\circ$	0.632	0.661	0.656	0.648	0.642	0.662	0.674	0.677	0.694	0.670	0.674	0.648	0.677	0.728	0.730	0.756	0.748
$\gamma_l = 1^\circ$	0.597	0.633	0.637	0.636	0.616	0.663	0.649	0.634	0.688	0.629	0.656	0.655	0.697	0.692	0.699	0.767	0.724
$\gamma_l = 2^\circ$	0.649	0.661	0.651	0.679	0.663	0.688	0.682	0.681	0.722	0.663	0.703	0.700	0.726	0.751	0.808	0.797	0.773
$\gamma_l = 3^\circ$	0.577	0.596	0.612	0.603	0.612	0.628	0.626	0.616	0.640	0.669	0.648	0.620	0.609	0.702	0.697	0.724	0.697
$\gamma_l = 4^\circ$	0.609	0.621	0.641	0.619	0.614	0.651	0.626	0.637	0.683	0.639	0.660	0.653	0.706	0.708	0.729	0.766	0.742
$\gamma_l = 5^\circ$	0.586	0.622	0.637	0.607	0.619	0.624	0.643	0.666	0.682	0.650	0.643	0.624	0.693	0.681	0.753	0.754	0.778
$\gamma_l = 6^\circ$	0.626	0.652	0.666	0.645	0.638	0.702	0.651	0.659	0.722	0.654	0.684	0.702	0.736	0.716	0.769	0.789	0.796

Tabelle 6.3 Der 3-D-Fehler e in Abhängigkeit der Rotation der linken und rechten Kamera.

6.2.1.4 Optimierung aller Parameter

Zuletzt werden nun alle gezielt veränderbaren Kameraparameter gleichzeitig optimiert, um die global beste Stereoansicht zu finden. Dazu muss das Bewegungsmodell des verwendeten Roboters berücksichtigt werden, um die Kamera an der entsprechenden Position zu positionieren. Um das Bewegungsmodell zu berücksichtigen, wird wie in Kapitel 5 vorgegangen und der Aktionsraum des Roboters diskret abgetastet. Für dieses Experiment wurde der Scrobot ER VII verwendet. Für die entsprechenden Drehgelenke wurde hierbei eine Schrittweite von $0.2 \text{ rad} \approx 11.5^\circ$ verwendet, für den Linearschlitten eine Schrittweite von 20 mm.

Um sicherzustellen, dass das Kalibrieremuster in den entsprechenden Bildern komplett im Bild liegt, wurde eine initiale Triangulation durchgeführt und die entsprechenden 3-D-Punkte rekonstruiert. Von jeder generierten Position wird nun eine Rückprojektion dieser initialen Schätzung durchgeführt und überprüft, ob die entsprechenden Punkte im Bild liegen. Die verwendete Auflösung beträgt hierbei 640×480 Bildpunkte. Zusätzlich wird noch ein Sicherheitsabstand von 80 Pixeln vom Rand weg berücksichtigt.

Im Experiment sind insgesamt 20477 gültige Parameter gefunden worden, d. h. Positionen, die erreichbar sind und bei denen die Projektion des Kalibrieremusters innerhalb des Sicherheitsabstands liegt. Für die Bestimmung des global besten Stereopaars müsste man bei einer vollständigen Suche somit 209643526 Vergleiche durchführen². Bei einer Geschwindigkeit von ca. 30 Positionsvergleichen pro Sekunde würde das Programm somit fast 81 Tage benötigen. Daher scheidet eine vollständige Suche aus.

Stattdessen wird ein spezieller Suchalgorithmus verwendet. Dabei wird angenommen, dass sich die entsprechenden optimalen Kameraparameter von linker und rechter Kamera unabhängig voneinander finden lassen. Dies wird in den theoretischen Überlegungen aus den Abschnitten 3.2.1 und 3.2.2 bestätigt: Im 2-D-Fall wurden die Kameraparameter der einen Kamera konstant gehalten und nur die der anderen Kamera optimiert. Die Optimierung liefert dabei ähnliche Ergebnisse, unabhängig davon, ob nur die Parameter einer Kamera, wie im 2-D-Modell, oder ob beide Kameras optimiert werden, wie im 3-D-Modell. Bei der Optimierung werden zunächst $n = 25$ Parametersätze zufällig gezogen. Die Wahrscheinlichkeit, mit der ein Parametersatz gezogen wird, ist jeweils gleich hoch. O. B. d. A. wird angenommen, dass dieser Parametersatz die Kameraparameter der linken Kamera definiert.³

Für jeden der $n = 25$ Parametersätze der linken Kamera wird der optimale Parametersatz für die rechte Kamera gesucht, indem für die rechte Kamera alle verbleibenden Parametersätze durchprobiert werden. Der optimale Parametersatz der rechten Kamera ist der, der den geringsten Fehler produziert. So ergibt sich für jeden der 25 Parametersätze der linken Kamera jeweils genau ein optimaler Parametersatz der rechten Kamera. Somit hat man 25 Parametersätze der rechten Kamera, die aus der Sicht des jeweils zugehörigen der linken Kamera optimal sind. Insgesamt ergeben sich so 25 Paare von Parametersätzen für linke und rechte Kamera. Nun wird für jeden dieser 25 Parametersätze der rechten Kamera der je-

²Berechnet über Binomialkoeffizient $\binom{20477}{2}$

³die Bezeichnungen links und rechts werden aus Konsistenzgründen beibehalten; prinzipiell haben die Begriffe links und rechts keine Bedeutung, da die Positionen vertauscht werden können

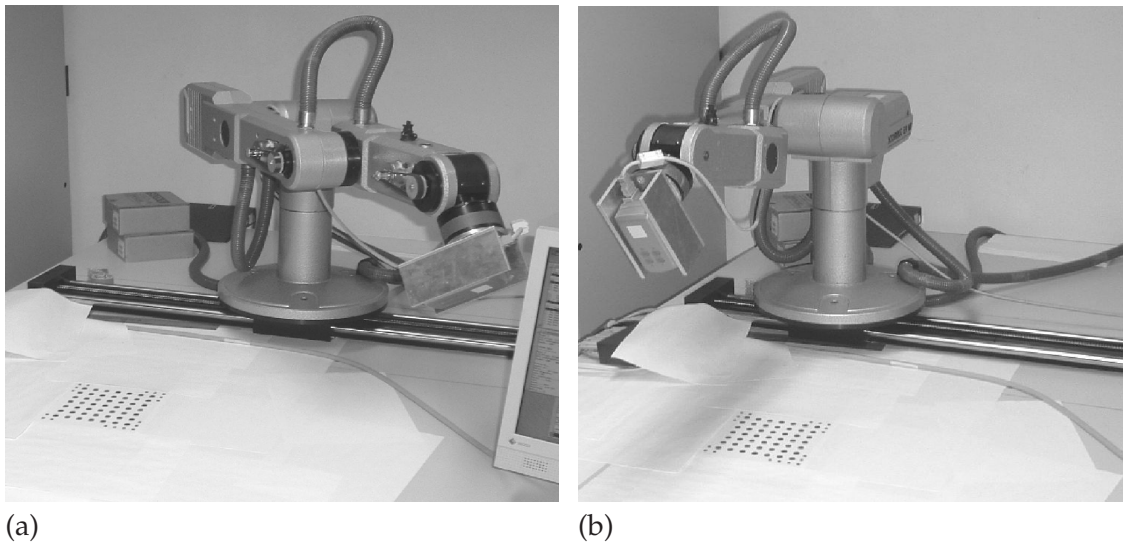


Bild 6.6 Optimale Kamerapositionen des Scorbot ER VII für die Rekonstruktion des Kalibrieramusters

weils optimale Parametersatz der linken Kamera gesucht. Diese Schritte wechseln sich solange ab, bis sich keine Änderung mehr ergibt. Die Berechnungsdauer für die optimale Ansicht beträgt in diesem Fall nur mehr 3 Stunden und 51 Minuten anstatt 81 Tagen bei der vollständigen Suche.

Die optimalen Kamerapositionen sind in Bild 6.6 und 6.7 verdeutlicht. Dabei zeigt Bild 6.6a und Bild 6.6b die tatsächlichen Positionen als Foto des Scorbot, während Bild 6.7 die Positionen in einer 3-D-Grafik inklusive der rekonstruierten Punkte darstellt. Die zugehörigen Aufnahmen aus diesen beiden Positionen sind in Bild 6.8 dargestellt.

Die als optimal errechneten Aufnahmepositionen sind nicht symmetrisch zum Kalibrieramuster, wie man es aufgrund der Simulationen in Abschnitt 3.2.1.4 erwarten würde. Dies kann erklärt werden: Bei der Optimierung wird das Roboterbewegungsmodell mit berücksichtigt. So kann es zwar prinzipiell noch bessere Stereopositionen geben, diese sind aber mit dem benutzen Roboterarm (hier der Scorbot ER VII) nicht erreichbar. Deshalb werden sie bei der Optimierung nicht berücksichtigt, da sie offenbar ohne Belang sind.

Wie in Abschnitt 3.2.1.4 gezeigt, wird bei einem Winkel zwischen den Sichtstrahlen von 90° der Rekonstruktionsfehler am geringsten. Bei der gewählten Stereoansicht beträgt der Winkel zwischen den z-Achsen der beiden Kamerapositionen ca. 70° .

Der mittlere 3-D-Fehler e liegt bei dieser Konstellation bei $e = 0.056 \text{ mm}$ und ist um mehr als Faktor 4 kleiner als bei der besten Optimierung einzelner Parameter. Das passt zu den Erwartungen aus Abschnitt 3.2.2.6, dass die Optimierung aller Parameter deutlich bessere Ergebnisse liefert als die Optimierung einzelner Parameter.

6.2.2 L-förmiges konkaves Objekt

Um die Ergebnisse des Kalibrieramusterexperiments weiter zu untermauern, werden die Experimente mit einem weiteren Objekt verifiziert. Im Unterschied zu

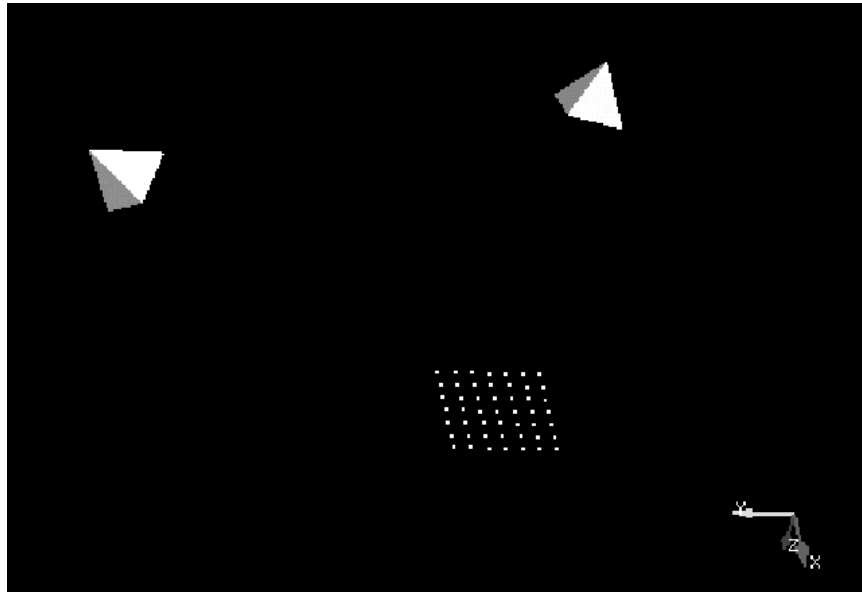
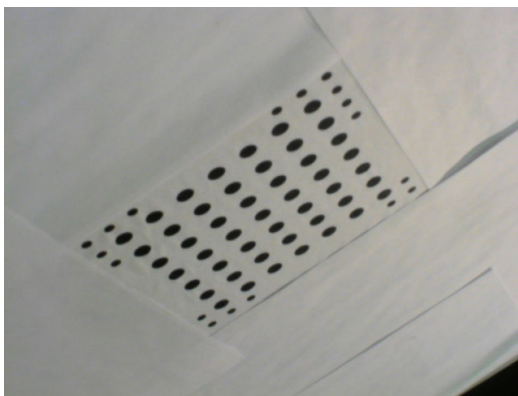
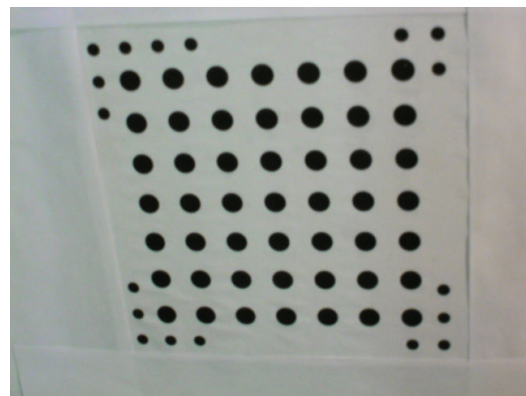


Bild 6.7 Rekonstruktion der 49 Punkte des Kalibrieramusters mit den beiden Kamerapositionen.



(a)



(b)

Bild 6.8 Kalibrieramuster aus den optimalen Ansichten, wie sie vom Scorbot ER VII erreicht werden können.

den Experimenten 6.3.3 und 6.4.1 ist es bei den Stereoaufnahmen nicht möglich, durch eine entsprechende Punktverfolgung korrespondierende Punkte zu extrahieren: Die Punktverfolgung, wie in Abschnitt 2.5.1, basiert auf einer Approximation durch eine Taylor-Reihe. Die Approximation versagt, wenn die Distanz zwischen Bildpunkten in den beiden aufeinanderfolgenden Bildern zu groß wird. Bei Stereoaufnahmen ist die Distanz zwischen den Bildpunkten in der Regel groß, von daher kann hier der Algorithmus zur Punktverfolgung nicht zuverlässig benutzt werden. Dies ist der Grund, warum hier mit einer Simulation gearbeitet wird.

Um die Simulation so realistisch wie möglich zu halten, wird das reale Objekt aus Bild 6.9, das auch in den Experimenten 6.3.3 und 6.4.1 verwendet wird, in der Simulation nachgebildet. Das Objekt besteht aus zwei Quadern mit den Maßen 10 cm x 10 cm x 1 cm, die an einer Kante senkrecht aufeinander stehen, so dass sich ein „L“-Profil ergibt.

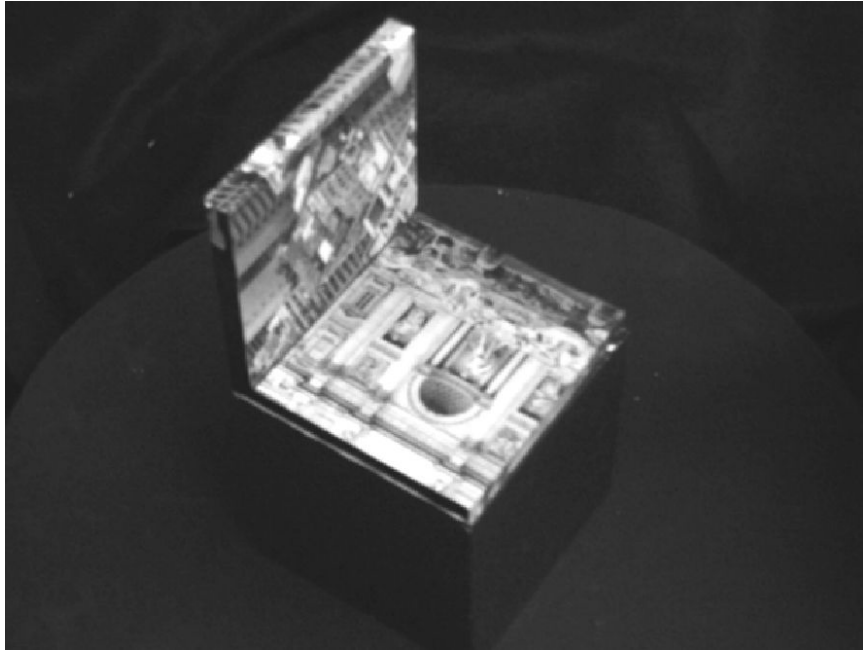


Bild 6.9 L-förmiges konkaves Objekt.

Für die Simulation werden aus dem „L“-förmigen Objekt auf jeder Seite 5 mal 5 Punkte der Oberfläche ins Bild projiziert, anschließend verrauscht und wieder zurückprojiziert. Als Fehlermaß wird dabei sowohl der mittlere euklidische Abstand der originalen zu den rekonstruierten 3-D-Punkten e als auch das in Abschnitt 6.1.3 vorgestellte Fehlermaß e_p verwendet. So lässt sich zeigen, dass die beiden Fehlermaße korrelieren und das im obigen Abschnitt definierte Fehlermaß tatsächlich ein Maß für die Güte der Rekonstruktion darstellt.

6.2.2.1 Optimierung der Brennweite

Dieses Experiment verwendet im Prinzip den gleichen Aufbau wie Experiment 6.2.1.1: Es werden zwei fixe Kamerapositionen für die Triangulation gewählt und anschließend die Brennweite der linken Kamera f_l bzw. der rechten Kamera f_r erhöht. Die Ergebnisse bezüglich des 3-D-Fehlers sind in Tabelle 6.4 und bezüglich des Fehlermaßes e_p in Tabelle 6.5 dargestellt, dabei ist die effektive Brennweite (vgl. Abschnitt 2.1.2) auf der waagerechten Achse angetragen.

Bild 6.10 bestätigt erneut die Aussage, dass der 3-D-Fehler mit größer werden der Brennweite abnimmt. Das Optimum stellt sich ein, wenn die Brennweiten der linken und rechten Kamera am Größten sind, dabei ist dann $e = 0.0617$ mm bzw. $e_p = 0.0416$ mm.

6.2.2.2 Optimierung der Translation

Hier die Ergebnisse für das äquivalente Experiment zu Experiment 6.2.1.2. Da in der Simulation die Einschränkung der Sichtbarkeit durch die begrenzte Bildebene, also dem Bildsensor, keine Rolle spielt, können hier alle Positionen auf einem rechteckigen Gitter zur Analyse verwendet werden.

Ebenso wie in Abschnitt 6.2.1.2 wird hier das Ergebnis in der gleichen Form repräsentiert: Für die jeweilige Kameraposition der linken und rechten Kamera

$f_r =$	760	814	910	1035	1172	1352	1575	1851	2225
$f_l = 760$	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.06	0.06
$f_l = 814$	0.08	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.06
$f_l = 910$	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.05	0.05
$f_l = 1035$	0.07	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.05
$f_l = 1172$	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.04
$f_l = 1352$	0.07	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04
$f_l = 1575$	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.04
$f_l = 1851$	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.03
$f_l = 2225$	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.03

Tabelle 6.4 Der 3-D-Fehler e der Rekonstruktion des L-förmigen konkaven Objekts in Abhängigkeit der Brennweiten in Pixel f_l der linken bzw. f_r der rechten Kamera.

$f_r =$	760	814	910	1035	1172	1352	1575	1851	2225
$f_l = 760$	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04
$f_l = 814$	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04
$f_l = 910$	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
$f_l = 1035$	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03
$f_l = 1172$	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03
$f_l = 1352$	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
$f_l = 1575$	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
$f_l = 1851$	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.02
$f_l = 2225$	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02

Tabelle 6.5 Das Fehlermaß e_p der Rekonstruktion des L-förmigen konkaven Objekts in Abhängigkeit der Brennweiten in Pixel f_l der linken bzw. f_r der rechten Kamera.

ist der Fehler als Grauwert dargestellt. Dabei beziehen sich die Fehlergrafiken Bild 6.12 und Bild 6.13 wiederum auf den Fehler e bzw. e_p .

Man erkennt, dass die Fehlergrafik 6.12 im Wesentlichen die gleichen Charakteristiken wie Haupt- und Nebendiagonalen sowie Kanten, wie Bild 6.3 aufweist. Darüber hinaus ergeben sich ebenfalls die gleichen Charakteristiken beim Vergleich der Bilder bei 6.12 und 6.13, mit den unterschiedlichen Fehlermaßen. Dies lässt sich anhand des Vergleichs der Bilder 6.14 und 6.15 bestätigen. Somit kann gefolgert werden, dass die Aussagen des Kalibrierungsmusters auch beim L-förmigen Objekt gelten und ebenso, dass die Güte der Rekonstruktion auch über das Fehlermaß e_p bewertet werden kann.

Das Optimum befindet sich hier bei der Position 190 und 209, was der nächsten Position und der mit der größten Stereobasis entspricht, genau so wie vorhergesagt. Die Position ist dabei optimal, unabhängig davon, ob das Fehlermaß e oder e_p verwendet wird. Natürlich unterscheidet sich der Wert des Fehlers, dieser ist $e = 0,0169$ mm bzw. $e_p = 0,0184$ mm.

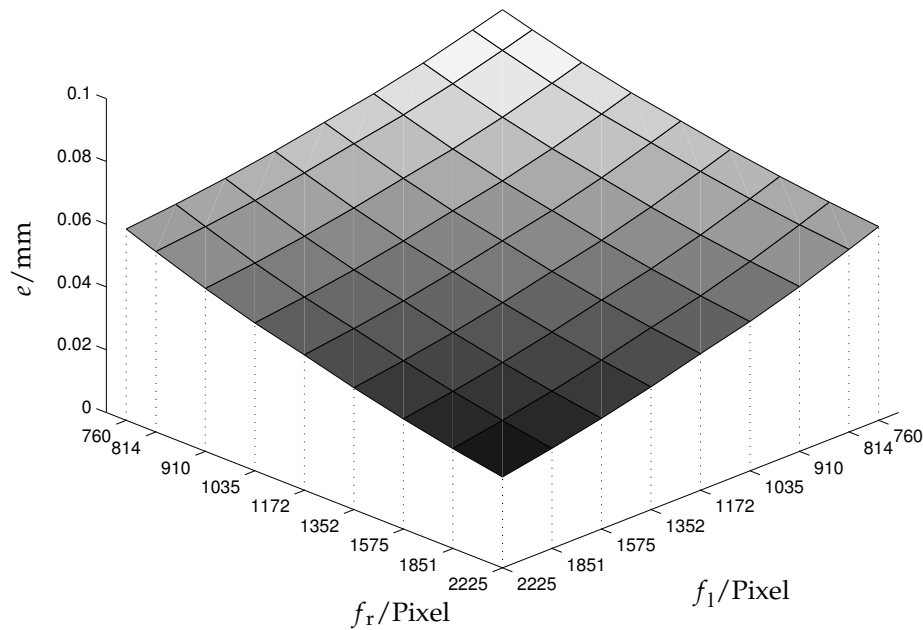


Bild 6.10 3-D-Fehler e in mm der Rekonstruktion des Objekts in Abhängigkeit der Brennweiten f_l der linken bzw. f_r der rechten Kamera. Man sieht: große Brennweiten ergeben kleine Fehler.

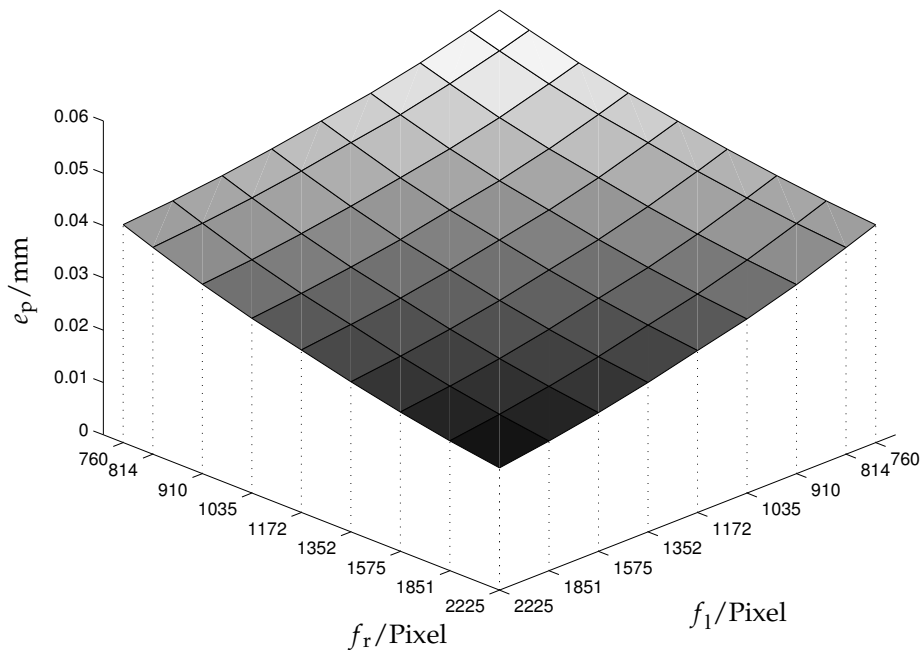


Bild 6.11 Das Fehlermaß e_p für verschiedenen Brennweiten f_l der linken bzw. f_r der rechten Kamera des L-förmigen konkaven Objekts.

6.2.2.3 Optimierung der Rotation

Dieses Experiment orientiert sich am Rotationsexperiment für das Kalibriermuster. Es wird erneut das L-förmige konkave Objekt aus zwei fixen Kamerapositionen betrachtet und die Kameras von dort jeweils um verschiedene Winkel im Bereich von $\pm 10^\circ$ gedreht. Im Gegensatz zum Kalibriermusterexperiment kann hier in der Simulation tatsächlich nur um das optische Zentrum rotiert werden,

	z=85	z=80	z= 75	z= 70	z=65	z= 60	z= 55	z= 50	z=45	z=40
x=-20	1	22	43	64	85	106	127	148	169	190
x=-18	2	23	44	65	86	107	128	149	170	191
x=-16	3	24	45	66	87	108	129	150	171	192
x=-14	4	25	46	67	88	109	130	151	172	193
x=-12	5	26	47	68	89	110	131	152	173	194
x=-10	6	27	48	69	90	111	132	153	174	195
x=-8	7	28	49	70	91	112	133	154	175	196
x=-6	8	29	50	71	92	113	134	155	176	197
x=-4	9	30	51	72	93	114	135	156	177	198
x=-2	10	31	52	73	94	115	136	157	178	199
x=0	11	32	53	74	95	116	137	158	179	200
x=2	12	33	54	75	96	117	138	159	180	201
x=4	13	34	55	76	97	118	139	160	181	202
x=6	14	35	56	77	98	119	140	161	182	203
x=8	15	36	57	78	99	120	141	162	183	204
x=10	16	37	58	79	100	121	142	163	184	205
x=12	17	38	59	80	101	122	143	164	185	206
x=14	18	39	60	81	102	123	144	165	186	207
x=16	19	40	61	82	103	124	145	166	187	208
x=18	20	41	62	83	104	125	146	167	188	209
x=20	21	42	63	84	105	126	147	168	189	210

Tabelle 6.6 Zuordnung der x - z -Position zur Bildnummer. Die Positionsangaben der x - bzw. z -Komponente sind in cm gemessen. Im Gegensatz zum Kalibriermustereperiment kann hier das Objekt aus jeder Position gesehen werden, so dass dort keine Positionen ausgelassen werden müssen.

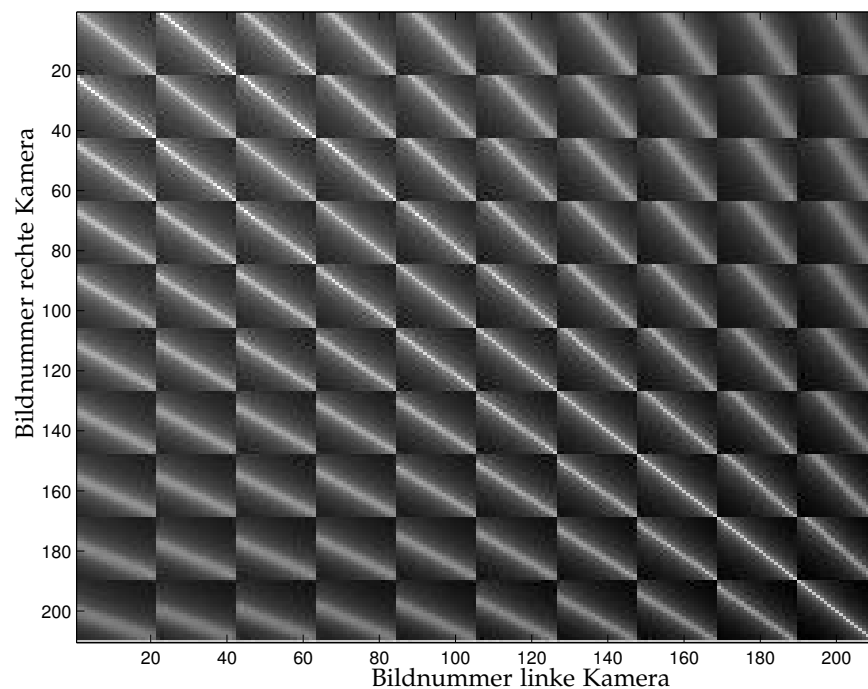


Bild 6.12 Rekonstruktionsfehler e des Objekts in Abhängigkeit der Translation, wobei ein heller Grauwert einen großen Fehler bedeutet.

d. h. ohne den geringfügigen Translationsanteil, der sich durch den mechanischen Aufbau ergibt.

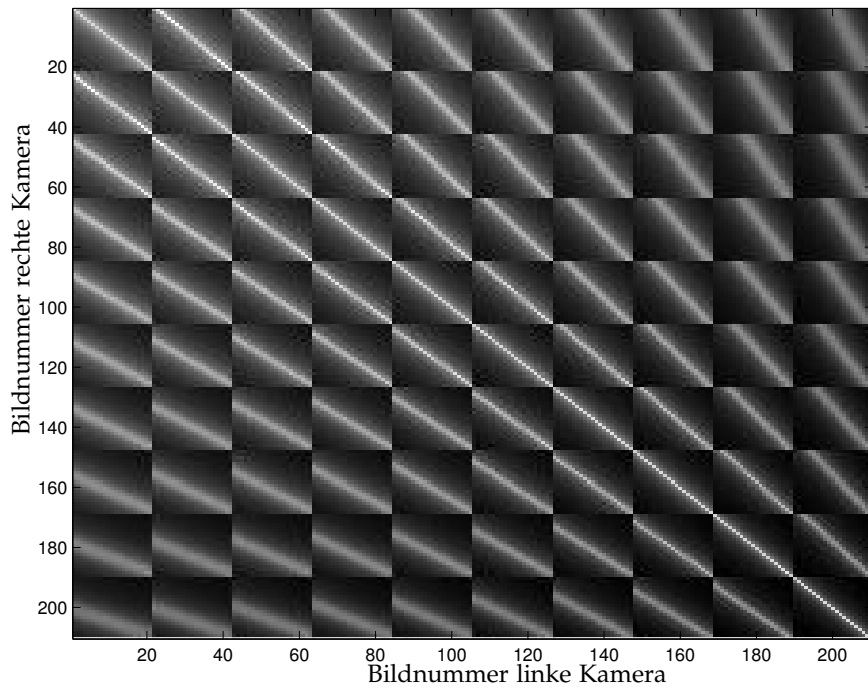


Bild 6.13 Rekonstruktionsfehler e_p des Objekts in Abhängigkeit der Translation, wobei ein heller Grauwert einen großen Fehler bedeutet.

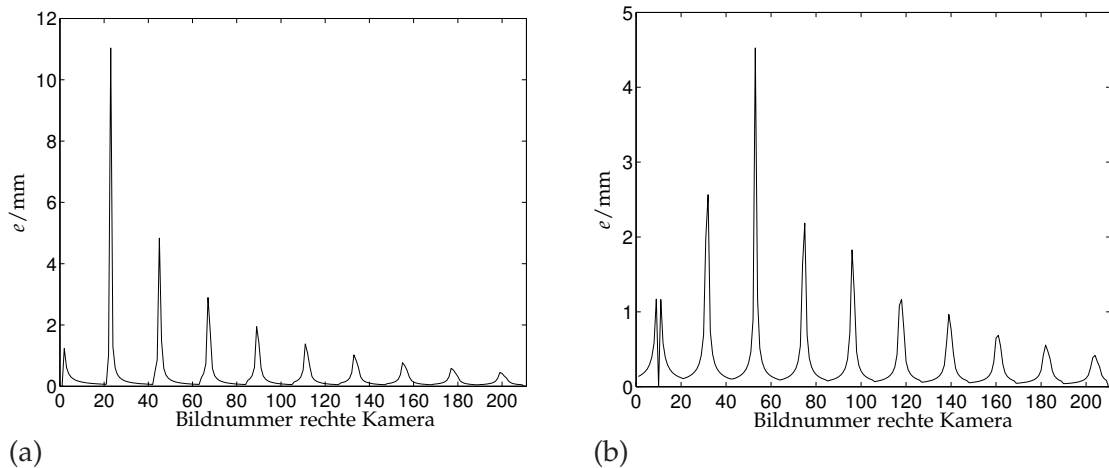


Bild 6.14 Rekonstruktionsfehler e in mm in Abhängigkeit der Bildnummer der rechten Kamera für die linke Kamera mit Bildnummer (a) 1 bzw. (b) 10. An der Stelle, wo die Bildnummer der linken und rechten Kamera identisch ist, wurde der Fehlerwert auf Null gesetzt.

In Bild 6.16 bzw. 6.17 ist die Abhängigkeit des Rotationsfehlers e bzw. e_p der linken und rechten Kamera dargestellt. Dabei ist die linke Kamera bei $+10^\circ$ weggedreht, die rechte bei -10° . Das absolute Minimum befindet sich bei $\gamma_l = 10^\circ$ und $\gamma_r = -10^\circ$ mit einem Wert von $e = 0,2292$ bzw. $e_p = 0,1363$ (vgl. Tabelle 6.3). Anhand der Abbildung ist zu erkennen, dass sich in dem Bereich, in dem beide Kameras weggedreht sind, im Allgemeinen ein niedrigerer Rekonstruktionsfehler ergibt als bei anderen Konstellationen.

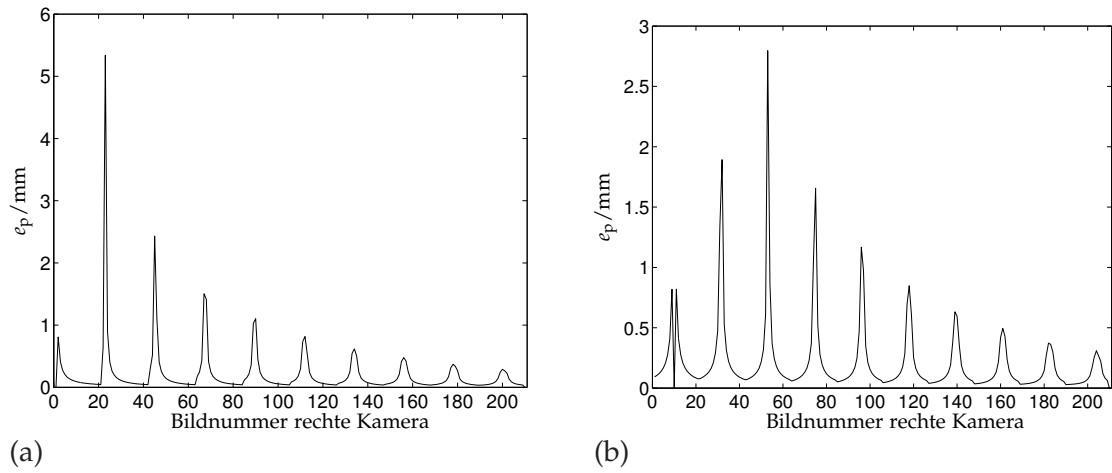


Bild 6.15 Rekonstruktionsfehler e_p gem. Abschnitt 6.1.3 in mm in Abhängigkeit der Bildnummer der rechten Kamera für die linke Kamera mit Bildnummer (a) 1 bzw. (b) 10. An der Stelle, wo die Bildnummer der linken und rechten Kamera identisch ist, wurde der Fehlerwert auf Null gesetzt.

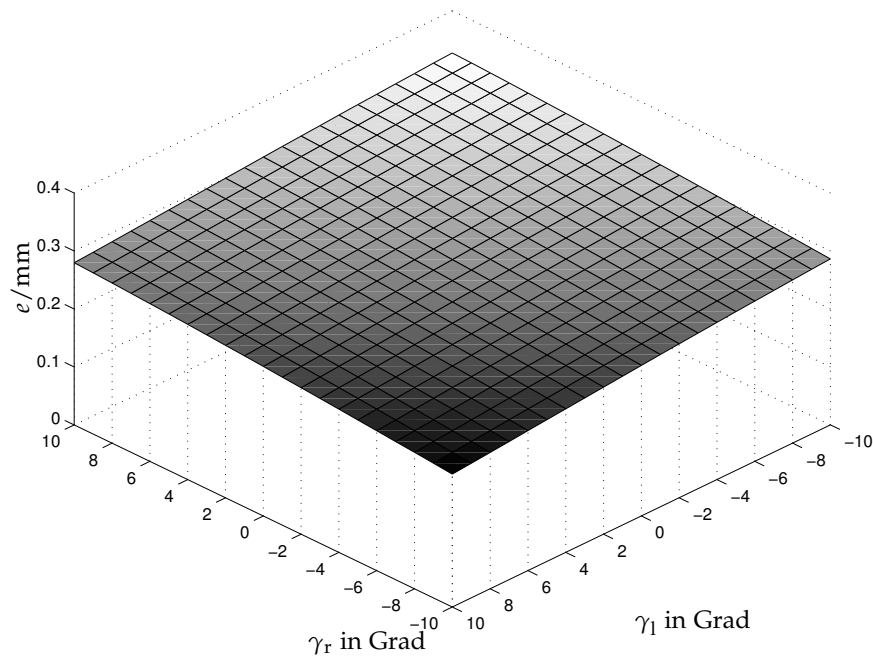


Bild 6.16 Darstellung des Rekonstruktionsfehlers e in mm bei der Rekonstruktion des Objekts in Abhängigkeit des Rotationswinkels von linker und rechter Kamera. Die linke Kamera ist bei $+10^\circ$ vom Objekt weggedreht, die rechte bei -10° .

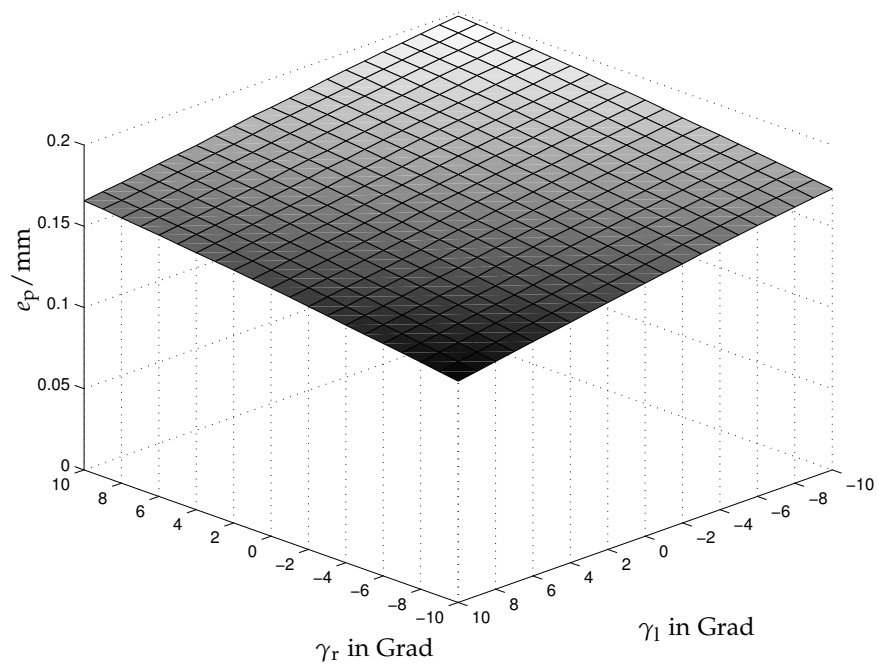


Bild 6.17 Darstellung des Rekonstruktionsfehlers e_p in mm bei der Rekonstruktion des Objekts in Abhängigkeit des Rotationswinkels von linker und rechter Kamera. Die linke Kamera ist bei $+10^\circ$ vom Objekt weggedreht, die rechte bei -10° .

$\gamma_r =$	-10°	-9°	-8°	-7°	-6°	-5°	-4°	-3°	-2°	-1°	0°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
$\gamma_l = -10^\circ$	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.33	0.33
$\gamma_l = -9^\circ$	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.33
$\gamma_l = -8^\circ$	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32
$\gamma_l = -7^\circ$	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32
$\gamma_l = -6^\circ$	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32
$\gamma_l = -5^\circ$	0.27	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.32	0.32
$\gamma_l = -4^\circ$	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31
$\gamma_l = -3^\circ$	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31
$\gamma_l = -2^\circ$	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31
$\gamma_l = -1^\circ$	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31
$\gamma_l = 0^\circ$	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31
$\gamma_l = 1^\circ$	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30
$\gamma_l = 2^\circ$	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30
$\gamma_l = 3^\circ$	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30
$\gamma_l = 4^\circ$	0.25	0.25	0.25	0.25	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.29
$\gamma_l = 5^\circ$	0.24	0.25	0.25	0.25	0.25	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29
$\gamma_l = 6^\circ$	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29
$\gamma_l = 7^\circ$	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29
$\gamma_l = 8^\circ$	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28
$\gamma_l = 9^\circ$	0.23	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28
$\gamma_l = 10^\circ$	0.23	0.23	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.27	0.28	0.28	0.28

Tabelle 6.7 Der 3-D-Fehler e in Abhängigkeit der Rotation der linken und rechten Kamera.

$\gamma_r =$	-10°	-9°	-8°	-7°	-6°	-5°	-4°	-3°	-2°	-1°	0°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
$\gamma_l = -10^\circ$	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.20	0.20	0.20
$\gamma_l = -9^\circ$	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.20	0.20	0.20
$\gamma_l = -8^\circ$	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
$\gamma_l = -7^\circ$	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19
$\gamma_l = -6^\circ$	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19
$\gamma_l = -5^\circ$	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19
$\gamma_l = -4^\circ$	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19
$\gamma_l = -3^\circ$	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19	0.19
$\gamma_l = -2^\circ$	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19
$\gamma_l = -1^\circ$	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = 0^\circ$	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = 1^\circ$	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = 2^\circ$	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18
$\gamma_l = 3^\circ$	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18
$\gamma_l = 4^\circ$	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18
$\gamma_l = 5^\circ$	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 6^\circ$	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 7^\circ$	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 8^\circ$	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17
$\gamma_l = 9^\circ$	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17
$\gamma_l = 10^\circ$	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17

Tabelle 6.8 Der Rekonstruktionsfehler e_p in Abhängigkeit der Rotation der linken und rechten Kamera.

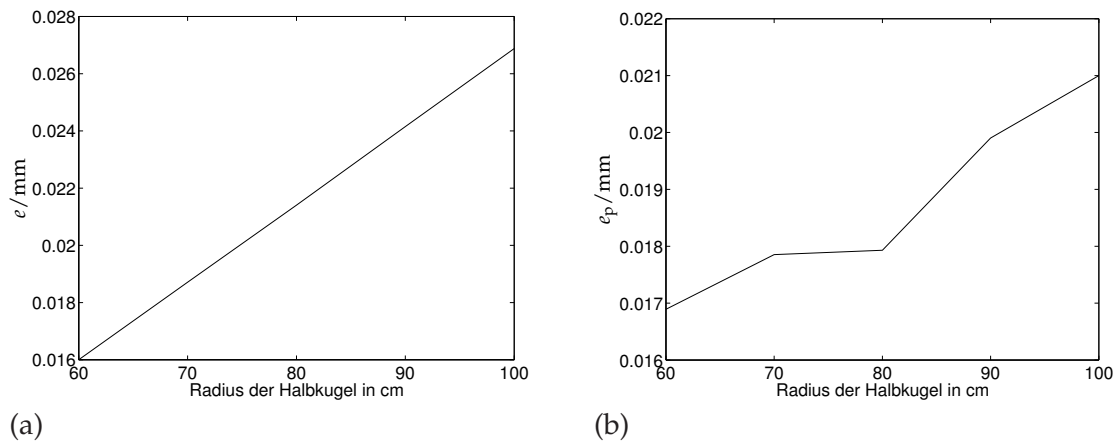


Bild 6.18 Fehler der Stereorekonstruktion des L-förmigen Objekts in Abhängigkeit des Radius bei Kameraanordnung auf einer Halbkugel mit dem direkt berechneten 3-D-Fehler e in Bild (a) bzw. e_p in Bild (b)

6.2.2.4 Optimierung aller Parameter

Die Optimierung aller Parameter mit den vom Scrobot ER VII erreichbaren Positionen liefert als bestes Ergebnis eine Rekonstruktion mit dem Fehler $e = 0.00815$ mm bzw. $e_p = 0.01596$ mm. Man sieht hier, dass die Optimierung über alle Parameter ein deutlich besseres Ergebnis hervorbringt als die Optimierung einzelner Parameter. Das Ergebnis ist also konsistent zu den Vorhersagen und Erwartungen.

Die Optimierung aller Parameter soll hier noch einen weiteren Punkt verdeutlichen: Durch die in Abschnitt 5.1 vorgestellte Methode kann die Bewegungsfreiheit des Scrobot ER VII vollständig genutzt werden. Im Gegensatz zu vielen Ansätzen in der Literatur, die von einer Halb- oder Teilkugel ausgehen, auf der sich die Kamerapositionen befinden, können durch den erweiterten Raum bessere Kamerapositionen gefunden werden. Zum Vergleich sind in Bild 6.18 die Fehler bei einer halbkugelförmigen Anordnung der Kameras für verschiedene Radien dargestellt. Neben der Positionierung auf der Halbkugel kann der Algorithmus auch noch aus neun verschiedenen Brennweiten wählen. Dabei wird stets die größte Brennweite als optimal ausgewählt. Dies ist ebenfalls eine Bestätigung für die Vorhersage aus Abschnitt 3.2.1.2.

Trotzdem ist zu erkennen, dass durch die Einschränkung der Kamerapositionen auf die Halbkugel das Rekonstruktionsergebnis, unabhängig davon, welches Gütemaß gewählt wird, doch teilweise um einiges schlechter ist als wenn man den gesamten möglichen Arbeitsbereich des zur Verfügung stehenden Roboters ausnutzt.

Aus den beiden Grafiken 6.18a und 6.18b wird nochmal klar, dass das in Abschnitt 6.1.3 definierte Fehlermaß sich vom tatsächlichen 3-D-Fehler nicht durch eine lineare Transformation unterscheidet. Trotzdem ist es ein Maß für die Güte bei Objekten, die aus mehreren planaren Flächen bestehen, da die Abweichung zu der Regressionsebene gemessen wird.



Bild 6.19 L-förmiges konvexes Objekt.

	e in mm	e_p in mm
Brennweite	0.0289	0.0249
Translation	0.0169	0.0156
Rotation	0.2291	0.1304
alle	0.0062	0.0143

Tabelle 6.9 Die Tabelle zeigt die Fehler e bzw. e_p für die Optimierung der einzelnen bzw. aller Parameter für das L-förmige konvexe Objekt.

6.2.3 L-förmiges konvexes Objekt

Das L-förmige Objekt kann auch gedreht werden, so dass es die geometrische Form eines Prismas (ohne Stirnseiten) hat. In diesem Fall wird vom L-förmigen konvexen Objekt gesprochen, das in Bild 6.19 dargestellt ist. Die Verwendung des L-förmigen Objekts mit der konvexen Seite bringt keine neuen oder weiteren Erkenntnisse, denn Eigenschaften wie konkave oder konvexe Bereiche sind bei dem in dieser Arbeit vorgestellten datengetriebenen Algorithmus nicht relevant. Trotzdem sind aus Gründen der Vollständigkeit die Ergebnisse im Anhang D abgedruckt. Da sich die Ergebnisse zwischen dem L-förmigen konkaven und L-förmigen konvexen Objekt im Grundsatz nicht unterscheiden, wurde dort auf eine Diskussion der Ergebnisse verzichtet. Die Tabelle 6.9 zeigt die Fehler e und e_p für die Optimierung der einzelnen bzw. aller Parameter.

6.3 Experimente zum probabilistischen Ansatz an einfachen Objekten

In diesem Abschnitt werden verschiedene Objekte mithilfe des theoretischen Modells aus Kapitel 4 rekonstruiert. Wie bereits im Abschnitt 6.2 werden dabei so-

wohl Aufnahmen realer Objekte als auch Simulationen eingesetzt. Dabei werden die Simulationen so nah wie möglich an die realen Experimente angelehnt, um die Ergebnisse miteinander vergleichen zu können.

Zur Erinnerung: der Ansatz zur probabilistischen Ansichtenplanung für die 3-D-Rekonstruktion aus Kapitel 4 unterscheidet sich grundlegend von dem geometrischen Ansatz aus Abschnitt 3.2. Deshalb können die Ergebnisse nicht direkt miteinander verglichen werden. Ergeben sich aber Kamerapositionen, die mit Hilfe des anderen Ansatzes ebenfalls plausibel erscheinen, wird explizit im Folgenden darauf hingewiesen.

Der probabilistische Ansatz benötigt zunächst eine initiale Schätzung, die anschließend sukzessive verbessert wird, wie bereits in Kapitel 5.3 dargestellt. Dazu wird stets von zwei zufällig gewählten aber konstanten Ausgangspositionen eine initiale 3-D-Rekonstruktion durchgeführt. Auf Basis der initialen Schätzung wird jeweils mit Hilfe der drei in Abschnitt 4.3 vorgestellten Gütemaße D-, modifizierte E- und T-Optimalität die initiale Schätzung verfeinert. Damit lassen sich im realen Experiment die drei Gütemaße direkt miteinander vergleichen.

Um für den Vergleich die Rahmenbedingungen, wie z. B. Beleuchtung oder Wiederholgenauigkeit des Roboters, konstant zu lassen, wurden die Bilder für die Experimente mithilfe eines Drehtellers einmalig aufgenommen und als konstante Bilddatenbank für alle Vergleichsmessungen herangezogen. Durch die Verwendung eines Drehtellers mit Schwenkarm liegen die Kamerapositionen auf einer Halbkugel mit dem Objekt im Mittelpunkt. Dabei wurden für jede Armposition 100 Aufnahmen mit je $3,6^\circ$ Drehung durchgeführt. Der Schwenkarm bewegte sich dabei von 90° auf 0° gegenüber der Horizontalen in 41 äquidistanten Schritten. So ergeben sich insgesamt 4100 Aufnahmen.

Um die Leistungsfähigkeit des probabilistischen Ansatzes zu analysieren, wurde neben der Untersuchung der drei Optimalitätskriterien zusätzlich noch eine 3-D-Rekonstruktion bei gleichmäßiger Abtastung und mit zufällig ausgewählten Kamerapositionen durchgeführt.

Gleichmäßige Abtastung bedeutet hier, dass nach den zwei initialen Positionen ein konstanter Armwinkel beibehalten und nur noch der Drehteller mit konstanter Schrittweite gedreht wird. Die Größe der Schrittweite ergibt sich dabei aus der gewünschten Gesamtzahl der Bilder, die in den folgenden Experimenten üblicherweise zwischen 10 und 15 Aufnahmepositionen besteht.

Zufällig ausgewählt bedeutet hier, dass aus den 4100 möglichen Aufnahmepositionen zufällig eine gezogen wird. Die Wahrscheinlichkeit mit der eine Position gezogen wird ist für jede Position gleich, d. h. $1/4100$.

In den folgenden Unterkapiteln werden nun die experimentellen Ergebnisse von verschiedenen Objekten dargestellt. Dabei wird die Komplexität der Objekte von Experiment zu Experiment immer größer. Es beginnt mit einer Simulation eines Kalibrierusters, anschließend wird ein Mousepad rekonstruiert, also nochmal ein planares Objekt. Dieses Experiment basiert auf realen Bilddaten und Kamerapositionen, es ist also keine Simulation mehr. Anschließend wird ein L-förmiges konkaves Objekt rekonstruiert, welches im Abschnitt 6.2.2 Grundlage für eine Simulation war. Zum Abschluss der Experimente zum probabilistischen Ansatz an einfachen Objekten, also ohne Selbstverdeckung, wird die Vorderseite einer Weihnachtsmann-Figur rekonstruiert und bewertet.

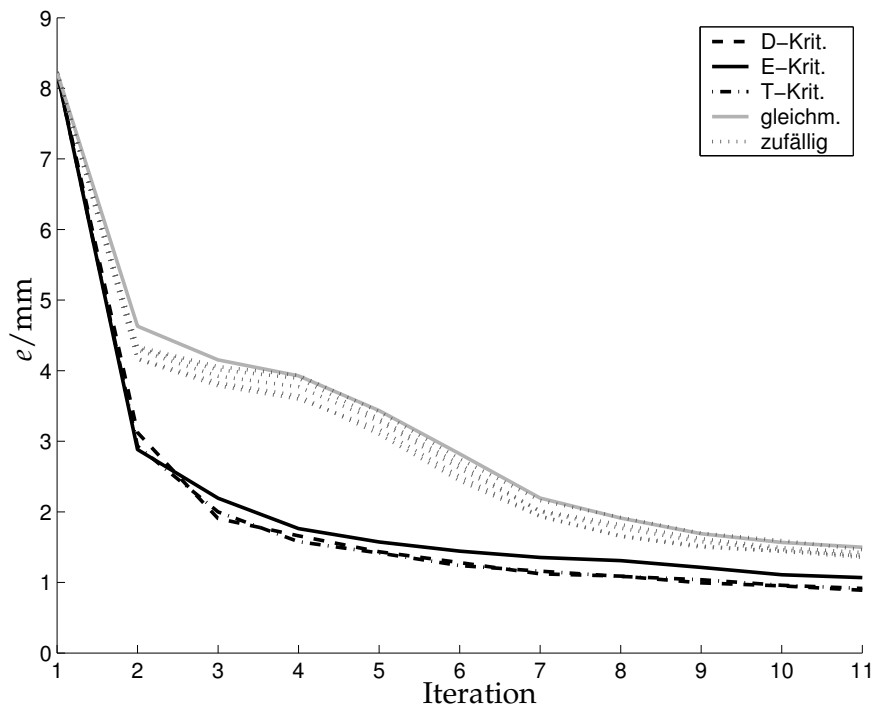


Bild 6.20 Rekonstruktion eines simulierten Kalibriermusters. Die Grafik zeigt den Fehler e der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei zufälliger Auswahl der Ansichten.

6.3.1 Ergebnisse Kalibriermuster

Bei der Rekonstruktion des Kalibriermusters kann, wie bereits oben erwähnt, der 3-D-Fehler direkt bestimmt werden. Die hier gezeigten Ergebnisse stammen aus einer Simulation, d. h. es werden keine realen Bilddaten verwendet, sondern die 3-D-Punkte direkt in die berechneten Kamerapositionen projiziert, verrauscht und anschließend eine 3-D-Rekonstruktion durchgeführt.

Der Vorteil der Simulation ist, dass hier direkt die Leistungsfähigkeit der in dieser Arbeit entwickelten Algorithmen beurteilt werden kann, während die weiteren nötigen Größen für die 3-D-Rekonstruktion, z. B. Hand-Auge-Transformation (vgl. Abschnitt 2.3) und Positionier- und Wiederholgenauigkeit des Roboterarms, exakt sind. Ein weiterer Vorteil bei der Verwendung des Kalibriermusters besteht darin, dass die Anzahl der zu rekonstruierenden Punkte stets identisch ist. Wie bei den weiter unten folgenden Experimenten erläutert, ist dies bei der Verwendung eines Algorithmus zur Punktverfolgung (vgl. Abschnitt 2.5.1) nicht gegeben.

Die zwei initialen Positionen für die Rekonstruktion sind für alle Algorithmen gleich. Dabei wurden zwei Positionen gewählt, die relativ nahe beieinander liegen. Deshalb ist die initiale Schätzung noch relativ ungenau. Diese wird anschließend durch die weiteren Ansichten in jedem Iterationsschritt verbessert. Dabei schneiden die optimierten Ansichten mit D-, modifiziertem E-, und T-Kriterium gegenüber der gleichmäßig abgetasteten (d. h. Arm auf 45° und Schrittweite des Tellers 36° , so dass sich ein geschlossener Kreis ergibt) und den fünf Versuchen mit zufälliger Ansichtenauswahl (d. h. für jeden Iterationsschritt wird zufällig eine Kameraposition aus den 4100 möglichen gezogen) deutlich besser ab, wie in Bild 6.20 zu sehen ist.

Für das modifizierte E-Kriterium sind in Bild 6.21 die optimalen Kamerapositionen nach jedem Iterationsschritt dargestellt. Man sieht in der Mitte das Kalibrieremuster mit 10 mal 10 Punkten und außen herum die Kamerapositionen. Die Kamerapositionen befinden sich auf einer Halbkugel, um ähnliche Bedingungen, wie bei den unten folgenden Drehteller-Experimenten, vorzufinden. Zur besseren Orientierung ist zusätzlich das originale Kalibrieremuster mit seinen 10 mal 10 Punkten eingezeichnet. Die Kamerapositionen aus dem D- und T-Kriterium sind sehr nahe an den jeweiligen des modifizierten E-Kriteriums und sind deshalb hier nicht aufgeführt. Es fällt auf, dass die Kamerapositionen meist sehr weit auseinander liegen. Die Optimierung liefert hier plausible Positionen: Bereits in Abschnitt 1.1.2 konnte anhand Bild 1.7 erläutert werden, dass eine fehlerbehaftete 2-D-Koordinate im Bild bei einem rechten Winkel zwischen den Sichtstrahlen die geringste Auswirkung auf die triangulierte 3-D-Koordinate hat. Dieser Sachverhalt zeigte sich bereits für Stereoaufnahmen in Abschnitt 3.2.1.4. Eine entscheidende Erkenntnis ist, dass über die definierten Fehlermaße implizit der Winkel optimiert wird.

Die Betrachtung der verschiedenen Kamerapositionen gibt also der Vermutung recht, dass hier möglichst rechte Winkel eine bessere Rekonstruktion erlauben. Dabei hat der geschlossene mathematische Ansatz aus Abschnitt 4.3 den Vorteil, dass sich beliebig komplexe Szenen oder Objekte ohne Vorkenntnisse über die Geometrie (im Gegensatz zu den Arbeiten [Mar99a, Mar99b]) rekonstruieren lassen. Dabei sorgt der Algorithmus intern für die entsprechende Gewichtung der einzelnen Punkte und geeignete Auswahl der Ansichten.

Bevor nun zur Rekonstruktion komplexerer Objekte übergegangen wird, soll noch kurz die Abhängigkeit der Rekonstruktionsgüte von den zwei initialen Aufnahmepositionen diskutiert werden. Es erscheint plausibel, dass das Ergebnis insgesamt von der initialen Schätzung abhängig ist: je nachdem wie gut oder schlecht die initialen Positionen gewählt sind, umso kleiner oder größer ist der Fehler in der initialen Schätzung. Trotzdem sind die Charakteristiken der Fehlerkurven aber ähnlich, wie in Bild 6.22 bzw. 6.23 für zwei andere Ausgangskonfigurationen bei diesem Kalibrieremusterexperiment zu sehen.

Nach dieser Simulation und den Plausibilitätsbetrachtungen der Ansichten wird nun zur Rekonstruktion komplexerer Objekte übergegangen. Insbesondere werden nun reale Objekte aufgenommen, dabei wird nun das Gesamtsystem inkl. Punktverfolgung und die Einbeziehung des Bewegungsmodells des Roboters verwendet.

6.3.2 Ergebnisse Mousepad

Beim Mousepad handelt es sich zunächst noch einmal um ein planes Objekt, welches auf dem Drehteller positioniert wurde, wie in Bild 6.24 dargestellt. Als Erschwernis kommt nun hinzu, dass verschiedene Punkte auf der Oberfläche mittels Punktverfolgung über die Rekonstruktionssequenz hinweg verfolgt werden müssen, um sie dann rekonstruieren zu können. Durch die Punktverfolgung kommt es zu verschiedenen Besonderheiten: Manche Punkte werden u. U. nicht in jeder Ansicht gefunden. Es wird zwar versucht, solche Punkte in einer späteren Ansicht wiederzufinden, jedoch klappt dies nicht immer. Der Punktverfolger ist so eingestellt, dass er die vorgegebene Anzahl von Punkten konstant hält – andernfalls wären irgendwann nur sehr wenige bis gar keine Punkte mehr in der Rekonstruktion

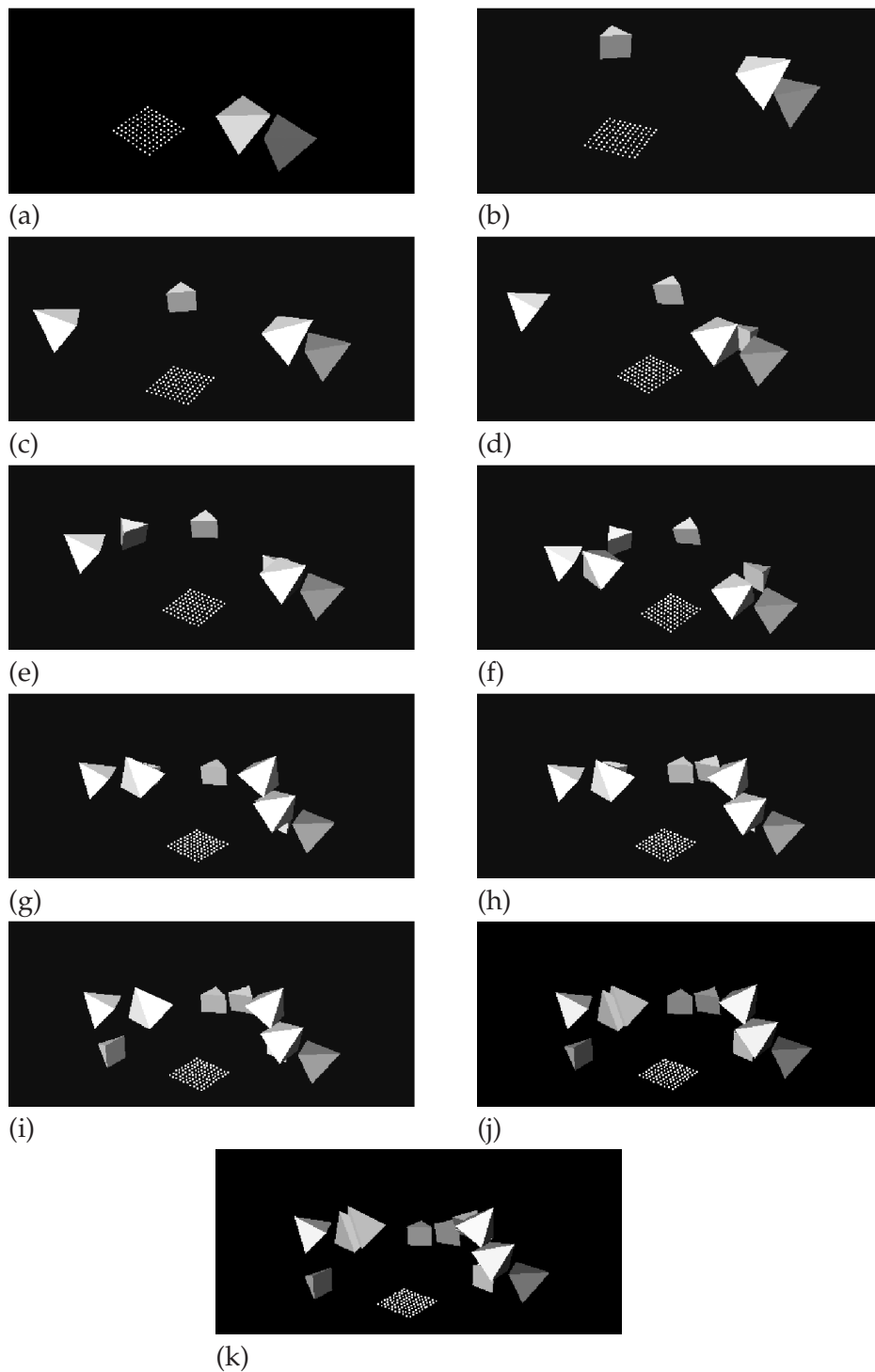


Bild 6.21 Kamerapositionen nach jedem Optimierungsschritt bei der Rekonstruktion des Kalibrieramusters mit Hilfe des modifizierten E-Kriteriums. Im ersten Bild sieht man das Kalibrieramuster sowie die zwei initialen Positionen. Im nächsten Bild kommt eine weitere Kameraposition hinzu: Wie oben erläutert wird der Fehler klein, wenn sich die Sichtstrahlen in einem 90° Winkel treffen. Im nächsten Bild kommt eine weitere Kamera hinzu, so dass die Sichtstrahlen wieder einen möglichst rechten Winkel bilden, usw. (Der Blickwinkel auf die Szene konnte aufgrund der Unzulänglichkeiten beim Rendering der 3-D-Szene nicht konstant gehalten werden und variiert deshalb etwas)

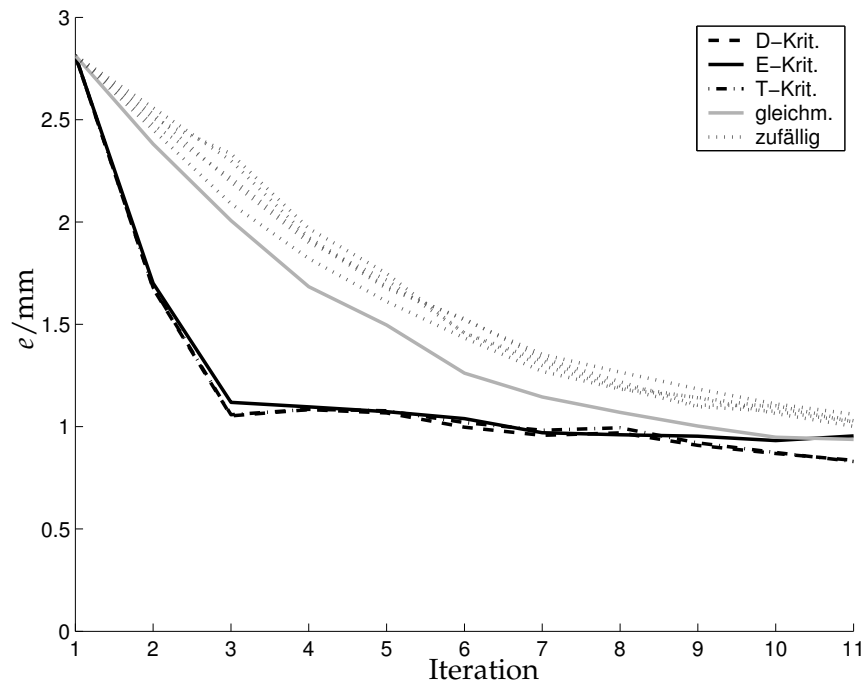


Bild 6.22 Rekonstruktion eines simulierten Kalibriermusters. Die Grafik zeigt den Fehler e der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei zufälliger Auswahl der Ansichten, mit einer andere initialen Schätzung als in Bild 6.20.

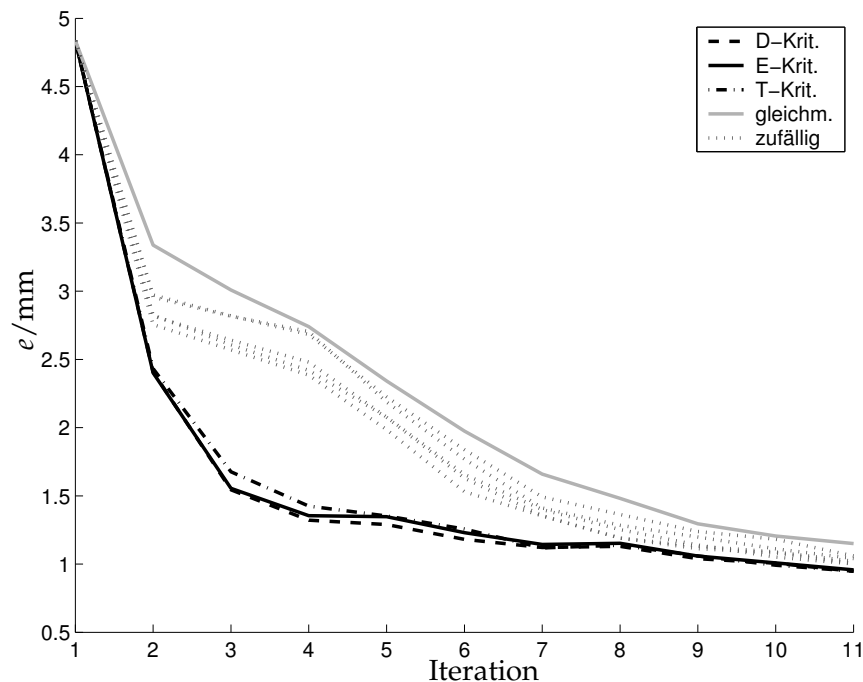


Bild 6.23 Rekonstruktion eines simulierten Kalibriermusters. Die Grafik zeigt den Fehler e der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei zufälliger Auswahl der Ansichten bei einer dritten initialen Schätzung gegenüber Bild 6.20 und 6.22.

zu finden. Die Koordinaten der neu gefundenen Punkte sind während der ersten Iterationsschritte noch sehr unpräzise. Dies ist der Grund, warum sich das Rekon-

struktionsergebnis manchmal bei einer weiteren Iteration plötzlich verschlechtern kann.

Bei diesem Experiment werden wieder die drei Optimierungskriterien D-, modifiziertes E- und T-Kriterium, eine gleichmäßige Abtastung und fünf Sequenzen mit zufälliger Auswahl von Ansichten verglichen. Die Ergebnisse sind in Bild 6.25 dargestellt. Dort sieht man den soeben erklärten Effekt, dass sich das Rekonstruktionsergebnis zwischen den Ansichten aufgrund neu gefundener Punkte durchaus einmal verschlechtern kann.

Ferner ist zu erwähnen, dass die drei Kriterien D, E und T durchaus andere Ansichten auswählen, schließlich werden die entsprechenden Kovarianzmatrizen durch ein anderes Maß bewertet. Zur Erinnerung: das modifizierte E-Kriterium minimiert den größten Eigenwert der jeweiligen 3×3 Blöcke der Kovarianz, das D-Kriterium minimiert die Entropie und das T-Kriterium minimiert die Spur der Kovarianz. Dies führt zu unterschiedlichen Ansichten und deshalb auch zu einem anderen Rekonstruktionsergebnis. In Bild 6.26 sind die Ansichten über den Verlauf der Optimierungssequenz für das modifizierte E-Kriterium dargestellt. Dabei fällt auf, dass auch hier von einem Iterationsschritt zum nächsten jeweils eine Kameraposition gewählt wird, die von der zuletzt gewählten relativ weit weg ist und sich zwischen den Sichtstrahlen grob ein rechter Winkel befindet. Dies entspricht der intuitiven Erwartung, und die gewählten Kamerapositionen erscheinen somit plausibel.

Vergleicht man die gewählten Kamerapositionen für die Simulation in Bild 6.21 mit denen aus Bild 6.26, so stellt man große Ähnlichkeiten fest. Dies bedeutet, dass die Ergebnisse aus der Simulation auf reale Experimente übertragen werden können. Manche Positionen liegen sehr nahe an der Position, die in der vorletzten Iteration auch schon gewählt wurde, z. B. ist die errechnete optimale Position in Bild 6.26(b) sehr nahe an der in Bild 6.26(d). Dies ist kein Widerspruch: Man betrachte hierzu die Rekonstruktion eines einzelnen 3-D-Punkts. Ausschlaggebend für die Kamerapositionierung mit Hilfe des modifizierten E-Kriteriums ist der größte Eigenwert und der zugehörige Eigenvektor der Kovarianzmatrix. Wird aus der berechneten optimalen Position nun die Beobachtung integriert, so wird dieser Eigenwert am stärksten minimiert. Das bedeutet in der nächsten Iteration ist ein anderer Eigenwert und sein zugehöriger Eigenvektor ausschlaggebend. Nach der Integration dieser 2. Beobachtung kann nun wiederum erneut der Eigenwert ausschlaggebend sein, dessen zugehöriger Eigenvektor fast parallel zu dem aus der ersten Iteration liegt. So ergibt sich entsprechend auch eine sehr ähnliche optimale Kameraposition. Das Verhalten ist nochmal in Bild 6.27 skizziert.

Im Vergleich von Bild 6.25 mit Bild 6.20 fällt auf, dass die Fehlerkurven nicht mehr monoton fallend sind, sondern gelegentlich ansteigen. Dies ist dann der Fall, wenn der Punktverfolger Punktkorrespondenzen verloren hat und diese durch neue ersetzt. In diesem Fall ist die Schätzung der 3-D-Position der neu gefundenen Punktkorrespondenz relativ unpräzise, was den Gesamtfehler verschlechtert. Da das verlieren von Punktkorrespondenzen nicht vorhersehbar ist, bleiben auch die geplanten Ansichten nicht von diesem Effekt verschont. So zeigt sich dieses Verhalten in Bild 6.25 für das modifizierte E-Kriterium zwischen der 3. und 4. Iteration.



Bild 6.24 Mousepad auf dem Drehteller

Da die Kamerapositionen beim Kalibriermuster und dem Mousepad ähnlich verteilt sind, lässt sich schließen, dass die exakte Lage der Punkte zueinander weniger entscheidend ist als die Oberflächengeometrie.

Es wurde gezeigt, dass der Algorithmus in seinen drei Varianten bei zwei sehr einfachen Objekten plausible Kamerapositionen auswählt und damit Ergebnisse erzielt, die besser als bei einer zufälligen oder gleichmäßigen Positionierung sind. Es stellt sich nun die Frage, welche Ergebnisse erreicht werden können, wenn die Geometrie komplizierter ist. Dies sollen nun die nächsten beiden Experimente zeigen, bei denen die Komplexität sukzessive erhöht wird.

Zu guter Letzt, soll an diesem Objekt noch einmal die Abhängigkeit der Fehlerkurven von der initialen Schätzung diskutiert werden. Dazu sind in Bild 6.28 und 6.29 die Fehlerkurven für die optimalen Ansichten, die gleichmäßige Abtastung sowie die zufällige Ansichtenauswahl für jeweils zwei unterschiedliche Startpositionen dargestellt. Auch hier im Experiment mit echten Bilddaten zeigt sich die Abhängigkeit des Fehlers von der initialen Schätzung. Es zeigt sich aber auch, dass die Fehlerkurven wiederum immer die selben Charakteristiken beim Vergleich zwischen geplanten und ungeplanten Ansichten aufweisen.

6.3.3 L-förmiges konkaves Objekt

In diesem Experiment wird das Objekt aus Bild 6.9 rekonstruiert, welches auch schon als Grundlage der Simulation in Abschnitt 6.2.2 diente. Das Objekt ist aus Aluminium gefertigt und hat eine Kantenlänge von 10 cm. Die Ebenen des Objekts sind plan mit einer maximalen Abweichung von weniger als 0,01 mm.

Da die Aluminiumoberfläche keine geeigneten Punktmerkmale zur Punktverfolgung bietet, sind die großen Seiten des Objekts mit einer matten Folie beklebt. Diese Klebefolie wurde zuvor mit verschiedenen Fotos von Gebäuden bedruckt, so dass sich dadurch geeignete Punktmerkmale finden lassen. Durch die Beklebung

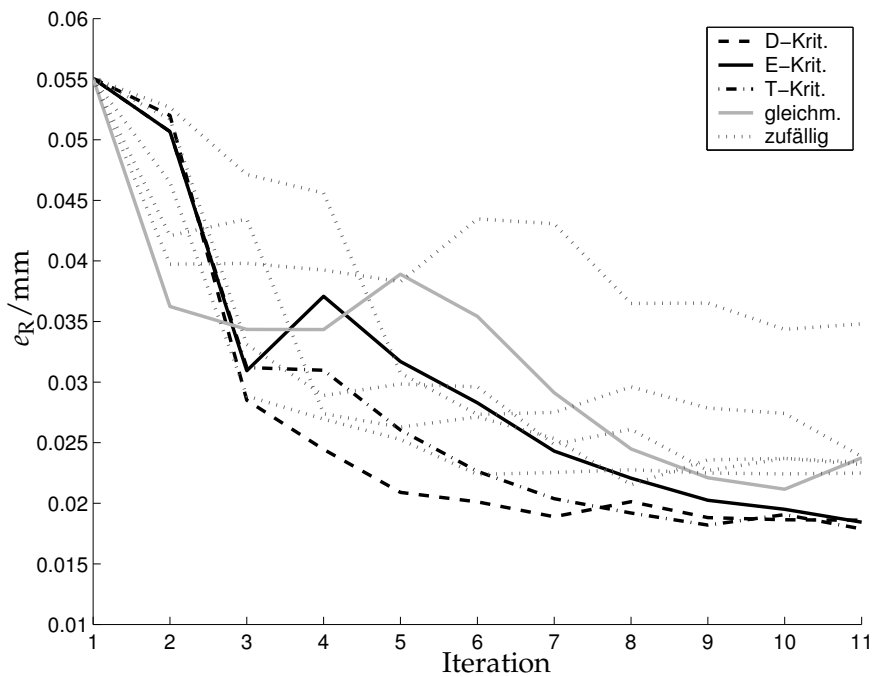


Bild 6.25 Rekonstruktion eines Mousepads. Die Grafik zeigt den Fehler e_R der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtsauswahl.

der Oberfläche ändert sich die Fertigungsgenauigkeit kaum, und wie man sieht, liegt sie mindestens ein bis zwei Zehnerpotenzen unter dem Rekonstruktionsergebnis. Es ist also nicht zu erwarten, dass dadurch die Aussagen zur Bewertung der Rekonstruktionsergebnisse beeinträchtigt werden.

In diesem Abschnitt wird der einfache probabilistische Ansatz, d. h. ohne Berücksichtigung der Sichtbarkeit, analysiert, weshalb hier nur die „Innenseite“ des Objekts rekonstruiert wird (in Abschnitt 6.4.1 wird das gesamte Objekt rekonstruiert und die auftretenden Selbstverdeckungen mit berücksichtigt). Die Stirnseiten sind deshalb im Gegensatz zu den großen Flächen homogen schwarz. Deshalb kann der Punktverfolger hier keine Merkmale finden und damit ergeben sich auch keine Verdeckungen. Da sich das zu rekonstruierende Objekt aus zwei Ebenen zusammensetzt, wird als Fehlermaß e_p aus Abschnitt 6.1.3 verwendet.

Wie bei den vorherigen Experimenten werden zunächst zwei initiale Aufnahmen gemacht. Anschließend wird die nächste Kameraposition gemäß dem gewählten Algorithmus angefahren und dort ein Bild gemacht. Da hier nur die „Innenseite“ des Objekts rekonstruiert werden soll, muss die Menge der gültigen Kamerapositionen vorher eingeschränkt werden. Da die Kamera mithilfe eines Drehtellers mit Schwenkarm bewegt wird, liegen die Kamerapositionen auf einer Halbkugel. Der mögliche Winkel des Drehtellers ist dabei auf den Bereich $[-90^\circ; +90^\circ]$ beschränkt. Dieser Bereich gibt die Schrittweite für die gleichmäßige Abtastung vor.

Die Ergebnisse sind in Bild 6.30 visualisiert. Der Verlauf des Fehlers zeigt deutlich, dass sich bei den drei Optimierungsvarianten der Fehler nach 4 bzw. 6 Iterationsschritten deutlich absenkt und dann auf niedrigem Niveau verbleibt. Eine

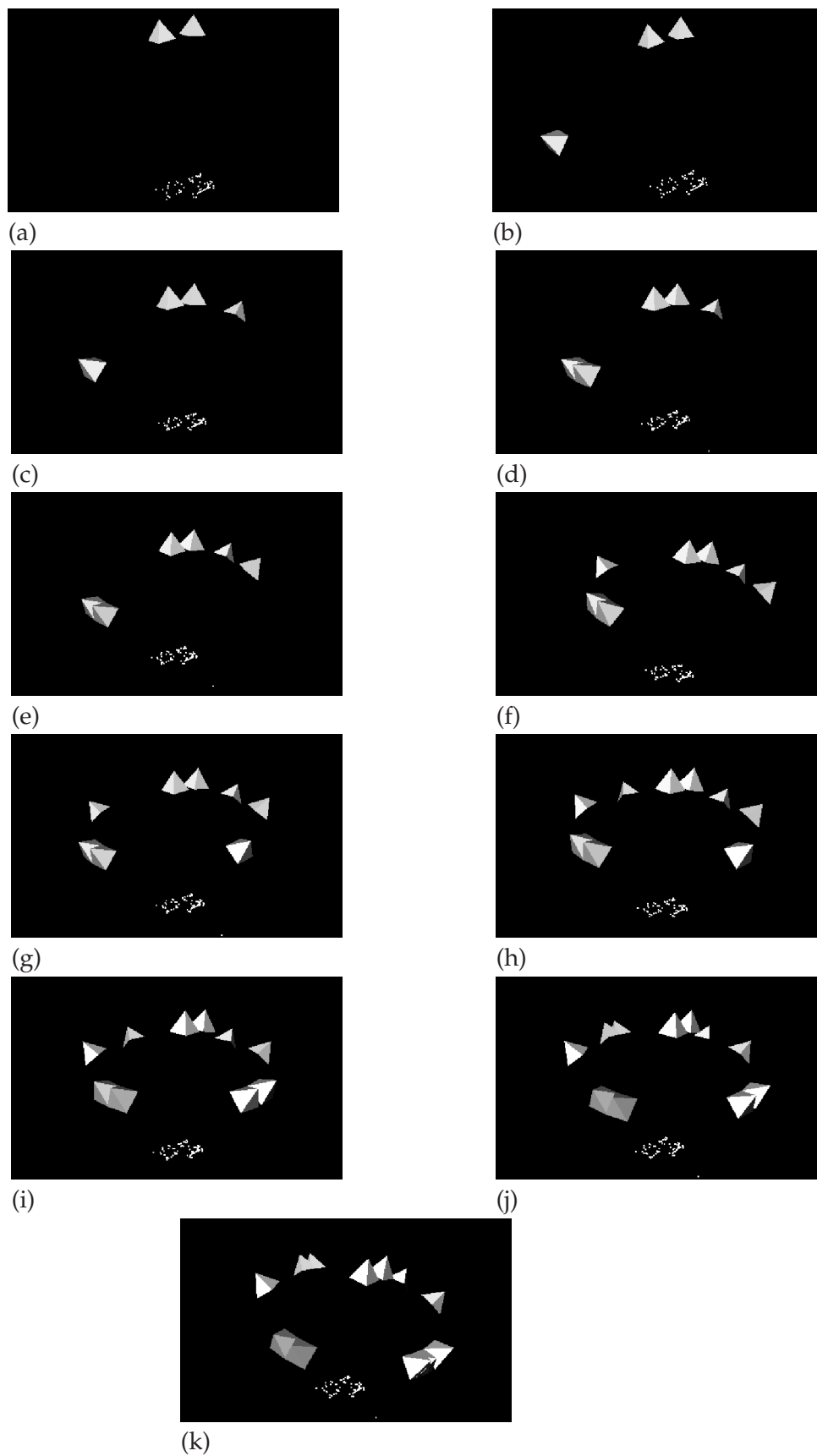
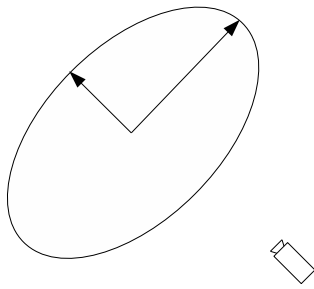
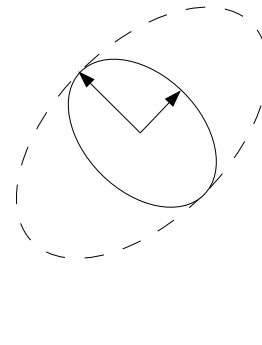


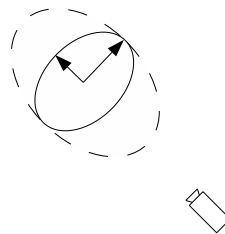
Bild 6.26 Kamerapositionen nach jedem Optimierungsschritt bei der Rekonstruktion eines Mousepads mit Hilfe des modifizierten E-Kriteriums.



(a) Startkonfiguration mit nächster bester Ansicht



(b) Schätzung nach erster Iteration und nächste beste Ansicht



(c) Schätzung nach zweiter Iteration und nächste beste Ansicht

Bild 6.27 Die Unsicherheit der Schätzung ist durch die Kovarianzellipse visualisiert. Gemäß dem modifizierten E-Kriterium wird in der Startkonfiguration (a) die Unsicherheit durch eine Aufnahme an der angedeuteten optimal reduziert. Nach Integration der Aufnahme aus dieser Position ergibt sich die Kovarianzellipse wie in (b) und erneut die optimale Position wie angedeutet. Nach der 2. Iteration ergibt sich nun die Kovarianzellipse (c) und die nun optimale Ansicht ist identisch mit der aus der Startkonfiguration.

Sequenz mit zufälliger Ansichtenauswahl kann zwar zunächst mithalten, stabilisiert sich aber nicht über die ganze Iterationssequenz.

Die gleichmäßige Positionierung der Kamera hingegen kann bei diesem Testobjekt keine besonders guten Resultate erzielen, auch wenn die Tendenz des Fehlers über die Sequenz abnehmend ist. Dies kann daran liegen, dass der Polarwinkel von 45° hier für den Punktverfolger nicht optimal ist, jedoch einen Kompromiss für die Rekonstruktion der beiden zu rekonstruierenden Ebenen darstellt (die ja einen rechten Winkel zueinander bilden).

Dazu kommt nun, dass sich die Kamera nur in einem Halbkreis und keinem kompletten Kreis bewegt. Deshalb fehlen die Aufnahmen von „gegenüber“, d. h. es gibt nur wenige Paare von Ansichten, bei denen Sichtstrahlen möglichst einen

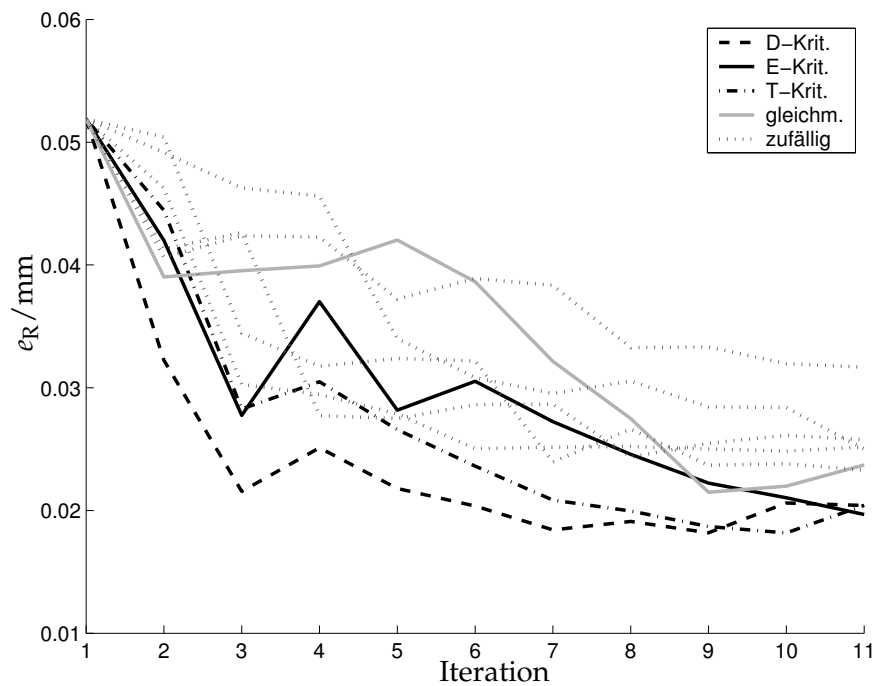


Bild 6.28 Rekonstruktion eines Mousepads. Die Grafik zeigt den Fehler e_R der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtsauswahl mit einer anderen initialen Schätzung wie in Bild 6.25.

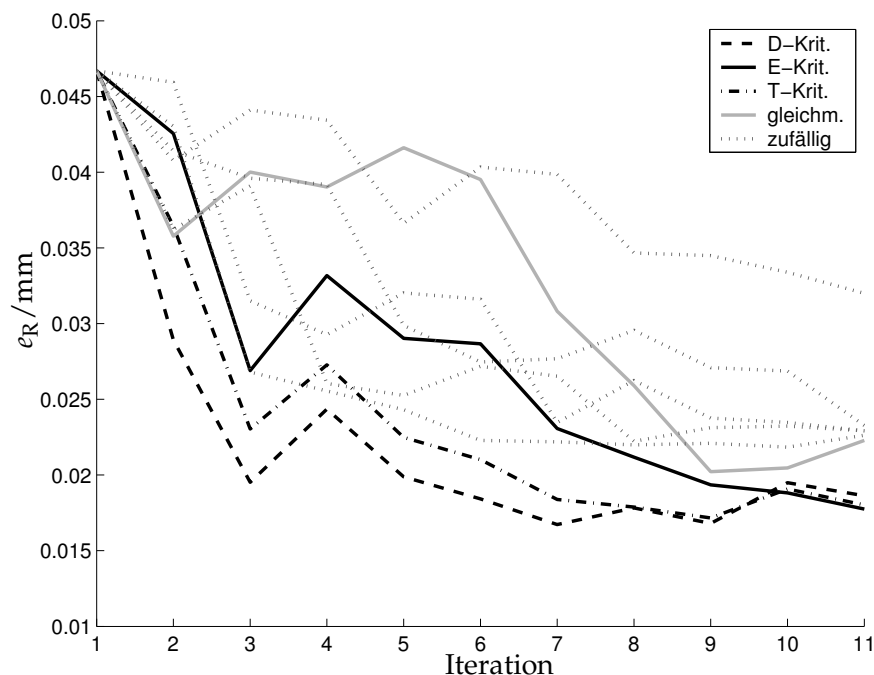


Bild 6.29 Rekonstruktion eines Mousepads. Die Grafik zeigt den Fehler e_R der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtsauswahl, bei einer dritten initialen Schätzung gegenüber Bild 6.25 und 6.28.

rechten Winkel bilden könnten. Aus diesem Grund ist das Rekonstruktionsergebnis bei der gleichmäßigen Auswahl entsprechend schlecht.

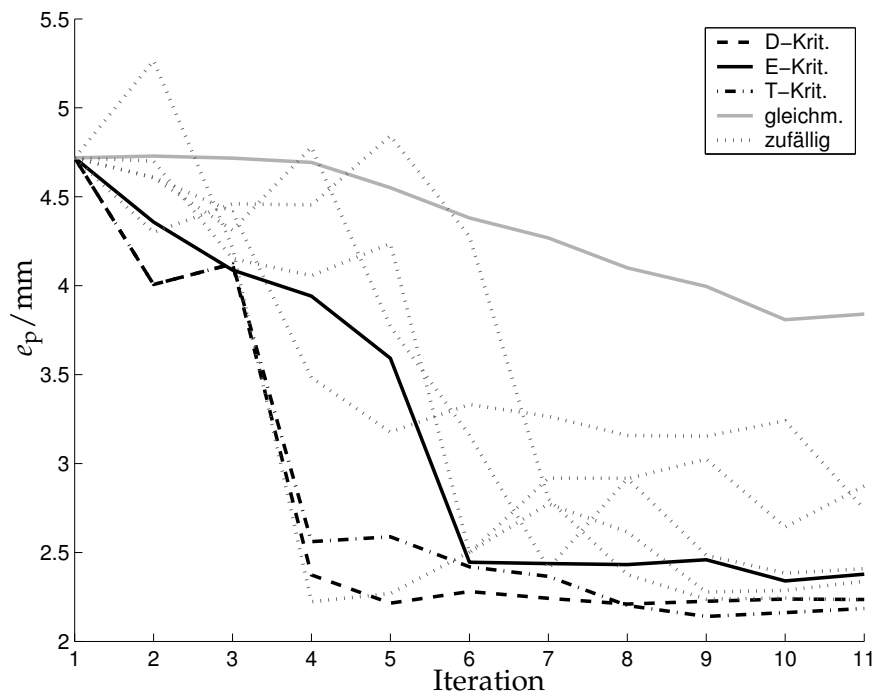


Bild 6.30 Rekonstruktion des L-förmigen konkaven Objekts. Die Grafik zeigt den Fehler e_p der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtenauswahl.



Bild 6.31 Weihnachtsmann

6.3.4 Weihnachtsmann

Das letzte Experiment für den einfachen Ansatz ohne Selbstverdeckung ist die Rekonstruktion eines Keramik-Weihnachtsmanns (siehe Bild 6.31). Der mögliche Winkel des Drehtellers wird in diesem Experiment auf den Bereich $[-60^\circ; +60^\circ]$ beschränkt, um Selbstverdeckungen zu vermeiden.

Das Fehlermaß lässt sich hier über den Abstand zur Referenzpunktewolke erstellen. Die Referenzpunktewolke wurde mithilfe einer Streifenprojektion erstellt

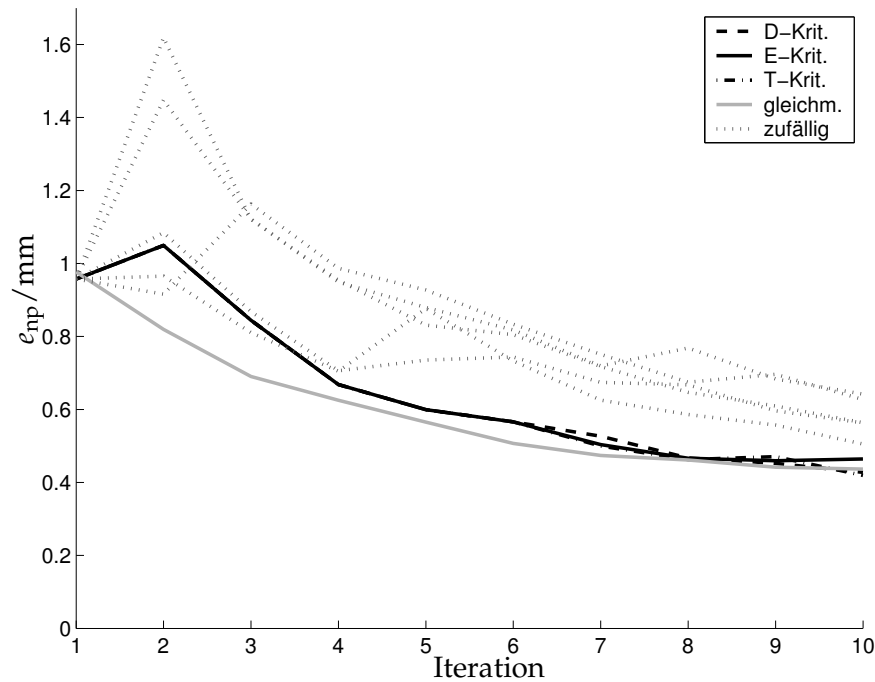


Bild 6.32 Rekonstruktion des Weihnachtsmanns. Die Grafik zeigt den Fehler e_{np} der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtenauswahl.

und hat eine Genauigkeit von ca. $50 \mu\text{m}$ [Tru11]. Dies ist als Referenz vollkommen ausreichend, da die hier erreichbare Genauigkeit, wie die vorherigen Experimente gezeigt haben, um Größenordnungen über der Messgenauigkeit der Streifenprojektion liegt.

Die Kamerapositionen für die gleichmäßige Abtastung wurden diesmal so gewählt, dass sie in einer horizontalen Ebene liegen. Damit sind die Sichtstrahlen in etwa orthogonal zur Oberfläche der zu rekonstruierenden Punkte. Dadurch sind für die 3-D-Rekonstruktion sehr gute Voraussetzungen erfüllt.

Aus diesem Grund schneidet hier auch die gleichmäßige Abtastung sehr gut ab, wie man anhand Bild 6.32 sehen kann. Die gleichmäßige Abtastung führt hier in jeder Iteration zu einer systematischen Verbesserung der Rekonstruktionsgenauigkeit, die von den zufälligen Kamerapositionen nicht erreicht wird. Die mit den drei Optimierungsansätzen gewählten Kamerapositionen können erst nach ungefähr sieben Iterationsschritten eine Rekonstruktionsgenauigkeit erreichen, welche mit der gleichmäßigen Abtastung vergleichbar gut abschneidet.

Anhand dieses Experiments und jenem aus Abschnitt 6.4.1 sieht man, dass die gleichmäßige Abtastung sowohl gute als auch weniger gute Ergebnisse liefern kann, während die optimierten Kamerapositionen jedes mal stabil zu einer großen Rekonstruktionsgüte führen. An dieser Stelle ist noch einmal darauf hinzuweisen, dass für die Berechnung der optimalen Kameraposition der zu erwartende Fehler minimiert wird, d. h. hier wird mit statistischen Methoden eine Vorhersage gemacht. Dies bedeutet dabei aber nicht, dass es keine andere Kameraposition gibt, die nach der Integration dieser Ansicht den tatsächlichen Fehler nicht stärker minimiert.

6.4 Experimente zum probabilistischen Ansatz an komplexen Objekten

In diesem Abschnitt werden verschiedene Objekte mithilfe des theoretischen Modells aus Kapitel 4 rekonstruiert. Dabei werden Aufnahmen realer Objekte sowie gerendertes Bildmaterial verwendet. Im Gegensatz zu Abschnitt 6.3 werden hier nun Objekte rekonstruiert, bei denen Selbstverdeckungen auftreten. Um die Selbstverdeckungen zu berücksichtigen, findet dabei der modifizierte Ansatz aus Abschnitt 4.4 Anwendung.

Das Vorgehen bei den Experimenten aus diesem Abschnitt ist somit identisch zu Abschnitt 6.3: Zunächst wird eine initiale Schätzung von zwei zufällig gewählten, aber konstanten, Ausgangspositionen durchgeführt. Auf Basis der initialen Schätzung wird jeweils mit Hilfe der drei in Abschnitt 4.4 vorgestellten Gütemaße D-, modifizierte E- und T-Optimalität die initiale Schätzung verfeinert. Damit lassen sich im realen Experiment die drei Gütemaße direkt miteinander vergleichen.

Auch bei der Rekonstruktion dieser Objekte wurden die Rahmenbedingungen, wie z. B. Beleuchtung oder Wiederholgenauigkeit des Roboters, durch die einmalige Aufnahme des jeweiligen Objekts aus allen möglichen Kamerapositionen eines Drehtellers konstant gehalten. Die Abtastung ist identisch zu jener aus Abschnitt 6.3: Die Kamerapositionen liegen auf einer Halbkugel mit dem Objekt im Mittelpunkt. Für jede Armposition werden 100 Aufnahmen mit je $3,6^\circ$ Drehung durchgeführt. Der Schwenkarm bewegt sich dabei von 90° auf 0° gegenüber der Horizontalen in 41 äquidistanten Schritten. So ergeben sich insgesamt 4100 Aufnahmen.

Weiterhin wird die Leistungsfähigkeit des probabilistischen Ansatzes zusätzlich mit der 3-D-Rekonstruktion aus gleichmäßiger Abtastung und mit fünf zufälligen Folgen von Kamerapositionen verglichen.

In den folgenden Unterkapiteln wird die Komplexität der Objekte von Experiment zu Experiment stets größer. Es beginnt mit dem schon bekannten L-förmigen Objekt, wobei nun das konvexe „Äußere“ des Objekts betrachtet wird und nicht das konkave „Innere“. Im anschließenden Experiment wird ein Tetraeder rekonstruiert, was deshalb interessant ist, da die Flächen im Gegensatz zu den bisherigen Objekten nicht mehr im rechten Winkel aufeinander stehen. Zum Abschluss wird ein von [Mun07] vorgeschlagener Prüfkörper rekonstruiert, der verschiedene geometrische Formen aufweist und somit für die 3-D-Rekonstruktion ein sehr anspruchsvolles Objekt darstellt. Dieses Modell lag nicht real vor, sondern nur die geometrische 3-D-Beschreibung, aus der, zur Durchführung des Experiments, die einzelnen Objektansichten gerendert werden.

6.4.1 L-förmiges konvexes Objekt

Das L-förmige Objekt diente bereits mit seiner Innenseite für das Experiment in Abschnitt 6.3.3. Dort wurde, da Selbstverdeckungen nicht berücksichtigt werden konnten, nur die Innenseite des Objekts rekonstruiert. In diesem Experiment wird das Objekt nun erneut gedreht betrachtet, wie bereits in Abschnitt 6.2.3. So ergibt sich die Oberfläche eines Prismas (vgl. Bild 6.19).

	Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13	14
E-Kriterium	sichtbar	60	61	32	89	83	71	142	140	127	82	85	85	85	85
	gesamt	60	62	98	116	156	160	183	183	183	183	183	183	183	183
D-Kriterium	sichtbar	60	60	40	100	75	72	99	77	94	94	107	110	60	103
	gesamt	60	62	125	138	163	180	180	194	198	204	204	204	206	208
T-Kriterium	sichtbar	60	60	36	83	67	97	90	104	104	104	104	104	104	104
	gesamt	60	62	123	139	185	185	209	233	233	233	233	233	233	233
gleichmäßig	sichtbar	60	59	55	45	52	56	64	59	65	70	53	64	80	90
	gesamt	60	60	61	76	99	107	114	115	118	121	121	134	134	134
zufällig 1	sichtbar	60	60	58	61	65	66	80	78	69	110	113	113	96	105
	gesamt	60	62	62	162	173	173	173	173	173	173	173	173	173	173
zufällig 2	sichtbar	60	60	62	61	56	75	74	79	64	144	106	133	105	120
	gesamt	60	61	62	68	102	124	168	168	185	185	185	185	185	185
zufällig 3	sichtbar	60	60	50	55	50	72	104	90	104	61	112	119	78	92
	gesamt	60	60	75	143	154	164	164	164	164	196	197	197	215	215
zufällig 4	sichtbar	60	55	67	71	113	51	125	114	121	92	89	132	130	147
	gesamt	60	106	138	173	173	173	186	186	186	186	218	218	218	218
zufällig 5	sichtbar	60	56	69	40	71	63	65	78	65	105	101	113	119	77
	gesamt	60	62	76	125	172	179	180	181	186	201	201	201	201	201

Tabelle 6.10 Die Tabelle zeigt für alle Rekonstruktionen des L-förmigen konvexen Objekts die Anzahl der sichtbaren Punktmerkmale und die Gesamtzahl der rekonstruierten Punkte in jedem Iterationsschritt.

Für die initiale Schätzung der 3-D-Rekonstruktion wurden wiederum zwei Ansichten gewählt, die relativ nahe beieinander liegen und die zusätzlich den Blick auf beide Seiten des Objekts ermöglichen.

Der Vergleich der drei Optimierungskriterien sowie der gleichmäßigen und der zufälligen Abtastungen sind in Bild 6.33 dargestellt. Da das verwendete Objekt aus zwei rechtwinklig zueinander stehenden Ebenen besteht, wird das verwendete Fehlermaß e_p aus Abschnitt 6.1.3 verwendet.

In Bild 6.33 ist der jeweilige Fehler für die verschiedenen Algorithmen dargestellt. Bei allen Kurven fällt auf, dass es irgendwann im Verlauf des Experiments zu einem extrem starken Fehleranstieg kommt. Dies liegt daran, dass das Objekt aus zwei Seiten besteht und dass beide Seiten gleichzeitig nur von einer verhältnismäßig geringen Anzahl von Ansichten auf der Halbkugel sichtbar sind. Deshalb passiert es im Verlauf des Experiments, dass eine Seite nicht mehr gesehen wird. Auf dieser Seite gehen dadurch alle bereits gefundenen Punktkorrespondenzen verloren. Hier sind nun zwei Dinge zu beachten: zum einen werden, wie bereits oben erwähnt, die verlorengegangenen Punktkorrespondenzen durch neue Punktmerkmale ersetzt. Dadurch ändert sich die Gesamtzahl der rekonstruierten Punkte. Die Koordinaten der neuen Punktmerkmale sind kurz nach ihrer Entdeckung nur durch wenige 2-D-Beobachtungen geschätzt und daher entsprechend ungenau. Deshalb steigt die Fehlerkurve sprunghaft an.

Zum anderen kann bei verlorengegangener Punktkorrespondenz keine 2-D-Beobachtung im Sinne des Kalman-Filters erfolgen. Folglich lässt sich die 3-D-Schätzung der nicht mehr gefundenen Punkte nicht weiter verbessern. Deshalb ist es wichtig, dass diese Punkte wiedergefunden werden, wenn sich die Kamera wieder in einer Position befindet, von wo aus das Punktmerkmal sichtbar ist. In Tabelle 6.10 sind deshalb für alle Rekonstruktionen die Gesamtzahl der 3-D-Punkte sowie die Zahl der sichtbaren Punktmerkmale für jeden Iterationsschritt ausgewiesen.

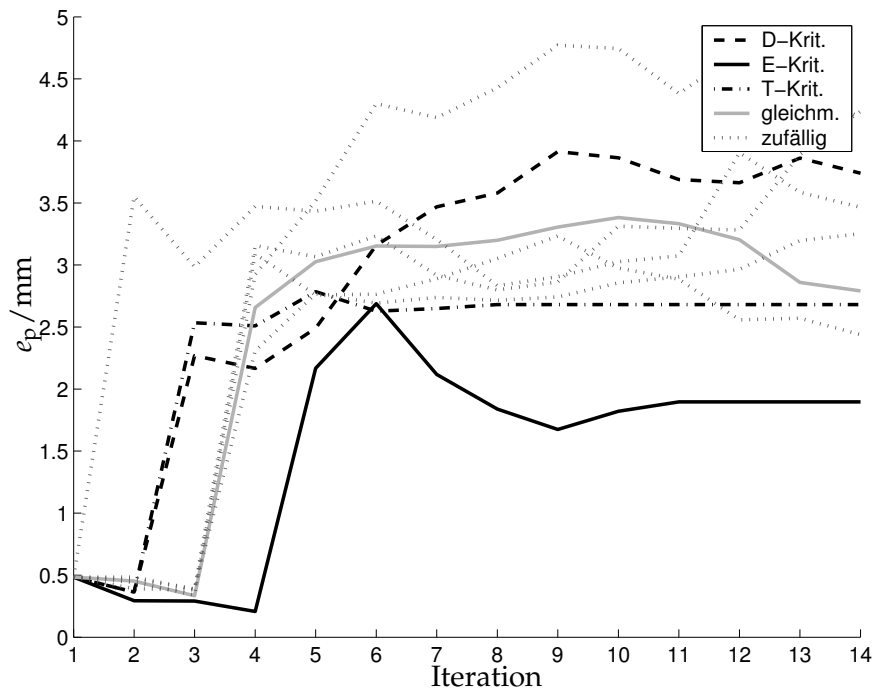


Bild 6.33 Rekonstruktion des L-förmigen konvexen Objekts. Die Grafik zeigt den Fehler e_p der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtenauswahl.

Um ein verlorenes Punktmerkmal wiederzufinden und die Punktkorrespondenz wiederherzustellen, wird die bisherige 3-D-Schätzung des entsprechenden Punkts in die Bildebene projiziert. Anschließend wird mithilfe des Punktdetektors in einer nahen Umgebung erneut das Punktmerkmal gesucht. Da es sich bei dem verwendeten Punktverfolger um einen differentiellen Ansatz handelt, der hier neu aufgesetzt wird, ist der Versuch, die Punktkorrespondenz über mehrere Bilder wiederherzustellen, nicht immer erfolgreich. Bei diesem und den folgenden Experimenten liegt die Wahrscheinlichkeit, die Punkte wiederzufinden, bei ca. 65 % bis 80 %. Dies bedeutet andersherum, es gibt Punktmerkmale, die einige Zeit lang verfolgt werden können und von denen eine 3-D-Schätzung gemacht wird. Diese gehen jedoch irgendwann im Verlauf des Experiments verloren und können nicht wiedergefunden werden.

Damit ist klar, dass sich in der 3-D-Rekonstruktion Punkte befinden, die keine weitere Beobachtung generieren. Somit lässt sich die Schätzung der 3-D-Position dieser Punkte nicht weiter verbessern. Aufgrund der angenommenen statistischen Unabhängigkeit der einzelnen Punkte lassen sich zumindest die anderen Punkte noch weiter verbessern. Jedoch bleiben die 3-D-Koordinaten der Punkte ohne Punktkorrespondenz damit unverändert, während die anderen Punkte präziser geschätzt werden können. Dies führt nach einigen Iterationen zu der Situation, dass die Punkte, die nicht weiter verfolgt wurden, maßgeblich zum Gesamtfehler beitragen. Damit erklärt sich weshalb die Fehlerkurven sich auf einem gewissen Niveau stabilisieren und nicht gegen Null konvergieren, wie es gemäß der Kalman-Filter-Theorie sein sollte und wonach es im simulierten Experiment, d. h. ohne den Einfluss eines fehleranfälligen Punktverfolgers, aussieht (vgl. Abschnitt 6.3.1).

Ferner kommt noch ein weiteres Problem hinzu, das sich aus der fehlenden Beobachtung aufgrund verloren gegangener Punktkorrespondenz ergibt. Wie in Abschnitt 4.2 dargelegt wurde, ist jeder 3-D-Punkt durch eine Normalverteilung mit Erwartungsvektor und Kovarianzmatrix beschrieben. Kann nun keine Beobachtung erfolgen, so lässt sich weder der Erwartungsvektor noch die Kovarianzmatrix aktualisieren. Die einzelnen Kovarianzmatrizen sind gemäß Abschnitt 4.3 entscheidend für die Auswahl der neuen Ansicht. Da dort alle Punkte berücksichtigt werden, haben Punkte mit entsprechend „großer“ Kovarianzmatrix einen entsprechend großen Einfluss auf die Bestimmung der nächsten optimalen Ansicht. Dies kann dazu führen, dass der entsprechende Algorithmus eine Ansicht vorschlägt, aber die Punkte, die maßgeblich zur Bestimmung dieser Ansicht als Optimum beigetragen haben, dort nicht gefunden werden.

Im ungünstigsten Fall mündet dies in einer Situation, in der stets die gleiche optimale Kameraposition berechnet wird, diese Ansicht jedoch das Rekonstruktionsergebnis nicht wesentlich verbessern kann, da die entscheidenden Punkte nicht detektiert werden können. Da der hier vorgestellte Algorithmus keine Vorhersage macht, ob ein Punktmerkmal vom Punktverfolger detektiert werden kann, sondern diese Daten nur als 2-D-Beobachtung des Kalman-Filter verwendet, kann keine Optimierung diesbezüglich durchgeführt werden. Um sich bei diesen und den folgenden Experimenten in Abschnitt 6.4.2 und Abschnitt 6.4.3 zu behelfen wird ein Kunstgriff vorgenommen: Unterscheidet sich bei der Berechnung der nächsten besten Ansicht diese nicht von der aktuellen, liegt offensichtlich die oben genannte Situation vor, dass der Algorithmus die 3-D-Schätzung von Punkten verbessern will, die aber in dieser Ansicht nicht sichtbar sind. Aus diesem Grund werden alle Punkte, die in der aktuellen Ansicht nicht detektiert wurden, aus der Optimierung der nächsten Ansicht herausgelassen. Mit Hilfe dieser Heuristik lässt es sich vermeiden, dass sich die Optimierung in eine Sackgasse manövriert und nicht mehr weiter optimiert.

In Bild 6.33 fällt ferner auf, dass es hier signifikante Unterschiede zwischen D-, modifizierter E- und T-Optimalität gibt. Dabei schneidet die, gemäß Abschnitt 4.3.2 geometrisch motivierte, modifizierte E-Optimalität am besten ab und liefert gegenüber D- und T-Optimalität sowie der gleichmäßigen Abtastung und der zufälligen Auswahl doch deutlich bessere Ergebnisse. Interessant ist in diesem Zusammenhang auch, dass die informationstheoretisch motivierte D-Optimalität schlechter abschneidet als der heuristische Ansatz der T-Optimalität.

6.4.2 Tetraeder

In diesem Experiment wird ein regelmäßiger Tetraeder rekonstruiert. In Bild 6.34 ist ein Bild des Tetraeders aus dem Experiment dargestellt. Der Tetraeder besteht aus vier gleichseitigen Dreiecken als Seiten, wobei nur drei Seiten rekonstruiert werden, da die 4. Seite als Standfläche verwendet wird. Der Tetraeder ist aus Karton hergestellt, und daher nicht ganz so exakt gearbeitet, wie die anderen Objekte in dieser Arbeit. Trotzdem ist die Genauigkeit für das Experiment ausreichend, da es hier ja aufgrund des in Abschnitt 6.1.3 vorgestellten Fehlermaßes nur auf planare Ebenen ankommt.

Für die Auswertung wird zunächst wieder eine initiale 3-D-Rekonstruktion aus zwei immer gleichbleibenden Ansichten gewählt. Die Ansichten sind manuell so

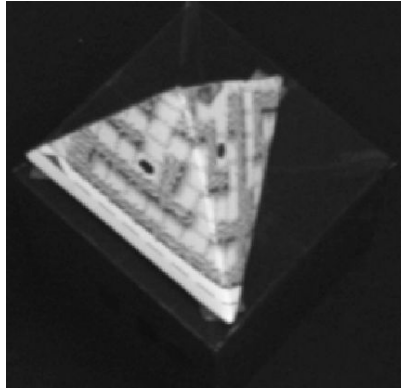


Bild 6.34 Tetraeder

gewählt, dass von dort alle drei Seiten sichtbar sind und Punktkorrespondenzen gefunden werden können.

Das Ergebnis des Experiments ist in Bild 6.35 dargestellt. Hier fällt zunächst auf, dass ein sprunghafter Anstieg, wie beim Experiment 6.4.1 mit dem L-förmigen konvexen Objekt, bei allen Kurven sehr unterschiedlich stark ausgeprägt ist. Dies liegt an der speziellen Form des Tetraeders. Im Gegensatz zum L-förmigen Objekt sind die Ansichten, aus denen nur eine Seite sichtbar ist, seltener. Insbesondere wenn die Kamera sich auf dem Schwenkarm des Drehtellers in der Horizontalen befindet, ist beim L-förmigen Objekt maximal eine Seite sichtbar (im ungünstigen Fall gar keine, da die Kamera genau seitlich auf das Objekt zeigt). Beim Tetraeder ist in dieser Stellung mindestens eine Seite sichtbar, teilweise auch zwei Seiten. Befindet sich der Schwenkarm also in der Horizontalen und wird der Drehteller einmal um 360° gedreht, so verschwinden beim L-förmigen Objekt zweimal unweigerlich alle Punktkorrespondenzen. Beim Tetraeder ist lange bevor die Punktkorrespondenzen auf einer Seite komplett nicht mehr sichtbar sind die nächste Seite bereits sichtbar, weshalb dort Punktkorrespondenzen gefunden werden können. Dieses Verhalten gilt nicht nur, wenn der Schwenkarm sich in der Horizontalen befindet, sondern für entsprechend andere Winkel analog. Deshalb ist hier nicht in jedem Fall ein so extrem sprunghafter Anstieg der Fehlerkurven zu erwarten, was Bild 6.35 in Kombination mit Tabelle 6.11 bestätigt.

Weiter sieht man in Abbildung 6.35, dass erneut die drei Optimierungskriterien wesentlich unterschiedliche Fehlerkurven produzieren. Dabei liefert, wie auch schon im Experiment 6.4.1 mit dem L-förmigen Objekt, das modifizierte E-Kriterium als Optimierungskriterium bessere Ergebnisse als die anderen beiden.⁴ Auch in anderen Veröffentlichungen zur Ansichtenplanung für die 3-D-Rekonstruktion wird ein Kriterium auf Basis der Eigenwerte verwendet. Jedoch beruht z. B. das Verfahren von Trummer oder Munkelt in [Tru11, Mun11] auf dem Standard E-Kriterium aus der Literatur [Puk93]. Dies hat jedoch, wie bereits in Abschnitt 4.3.2 erläutert, den Nachteil, dass allein der 3-D-Punkt mit größter Unsicherheit für die Bestimmung der nächsten besten Ansicht ausschlaggebend ist. Ferner zeigt sich auch hier, dass das heuristische T-Kriterium wesentlich besser als das D-Kriterium abschneidet. Letzteres liefert sogar nochmals deutlich schlechtere Werte als die gleichmä-

⁴Anmerkung: Dies gilt zumindest für den hier betrachteten Anwendungsfall der 3-D-Rekonstruktion. Eine Generalisierung auf andere Anwendungen, wie z. B. die 3-D-Objektverfolgung, kann hier nicht gemacht werden.

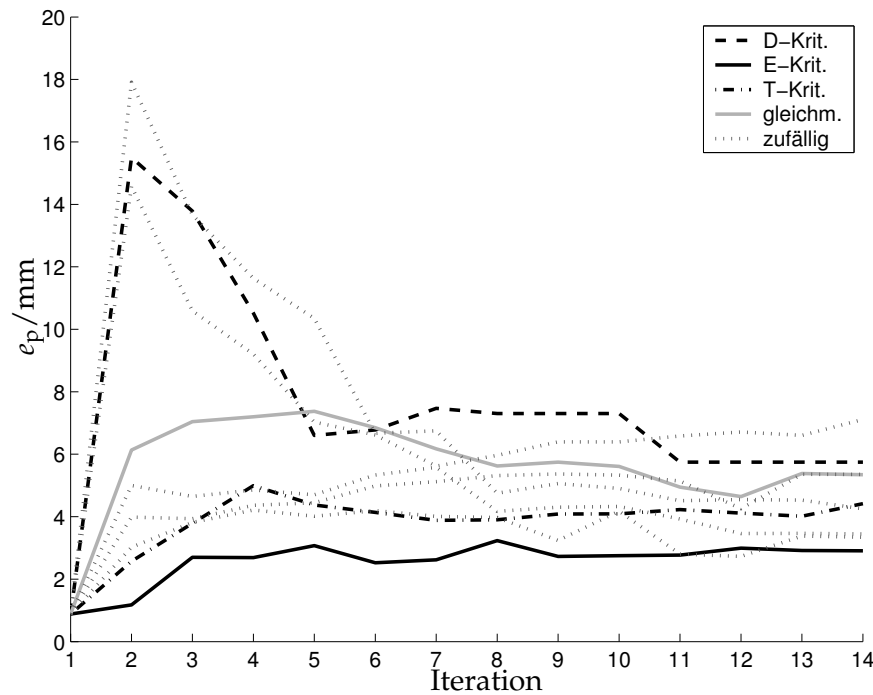


Bild 6.35 Rekonstruktion des Tetraeders. Die Grafik zeigt den Fehler e_p der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtenauswahl.

	Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13	14
E-Kriterium	sichtbar	59	59	58	69	59	64	84	73	97	56	61	63	90	98
	gesamt	60	84	106	111	129	165	170	184	184	184	184	188	189	189
D-Kriterium	sichtbar	59	63	64	70	93	85	61	105	105	105	110	110	110	110
	gesamt	60	126	145	196	196	196	216	221	221	221	221	221	221	221
T-Kriterium	sichtbar	59	59	56	65	72	82	84	63	88	91	80	80	97	92
	gesamt	60	95	114	163	164	172	172	176	176	176	176	176	176	176
gleichmäßig	sichtbar	59	52	55	57	47	56	63	67	58	51	65	68	87	78
	gesamt	60	91	99	104	111	138	151	154	155	169	185	185	185	185
zufällig 1	sichtbar	59	58	57	59	52	57	71	69	72	96	78	98	78	105
	gesamt	60	97	101	137	150	163	171	171	173	173	173	173	173	173
zufällig 2	sichtbar	59	53	78	61	63	69	59	77	57	91	79	93	112	112
	gesamt	60	91	102	107	122	126	178	181	186	189	189	189	189	189
zufällig 3	sichtbar	59	59	62	52	46	77	63	84	82	52	100	101	65	88
	gesamt	60	71	118	120	136	149	166	166	166	178	181	185	185	185
zufällig 4	sichtbar	59	52	62	59	82	77	96	81	100	68	97	110	89	111
	gesamt	60	114	155	198	201	201	201	201	201	201	227	232	232	232
zufällig 5	sichtbar	59	54	70	64	66	73	79	60	70	109	118	91	73	77
	gesamt	60	107	134	155	157	188	188	188	192	192	192	192	192	192

Tabelle 6.11 Die Tabelle zeigt für alle Rekonstruktionen des Tetraeders die Anzahl der sichtbaren Punktmerkmale und die Gesamtzahl der rekonstruierten Punkte in jedem Iterationsschritt.

ßige Abtastung oder die zufälligen Ansichten. Im Gegensatz dazu liefert das T-Kriterium Ansichten, die eine genauere Rekonstruktion ermöglichen, als die gleichmäßige Abtastung.

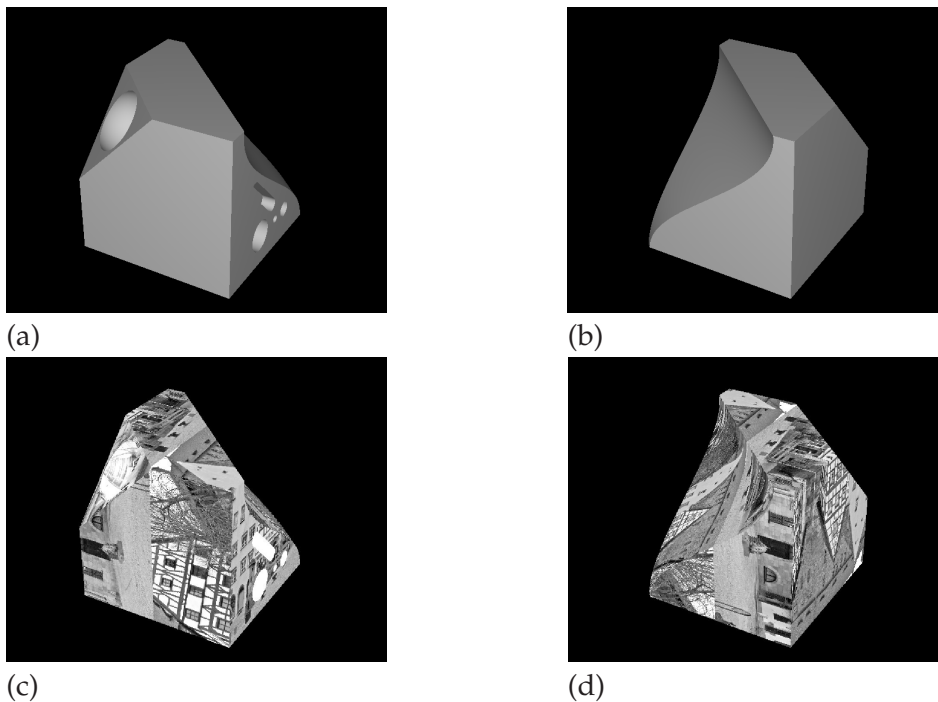


Bild 6.36 Verschiedene Ansichten des Prüfkörpers für die 3-D-Rekonstruktion, mit und ohne Textur

6.4.3 Prüfkörper

Der Prüfkörper ist ein Objekt, das in [Mun07] als Referenzkörper für die Bewertung von Algorithmen zur 3-D-Rekonstruktion vorgeschlagen wird. Der Körper weist verschiedene geometrische Formen auf und stellt somit für die 3-D-Rekonstruktion ein sehr anspruchsvolles Objekt dar. Dieses Modell wird mithilfe seiner geometrischen 3-D-Beschreibung gerendert. Der Prüfkörper ist in der Abbildung 6.36a und 6.36b dargestellt. Das menschliche Gehirn erkennt aufgrund der leicht unterschiedlichen Schattierung die 3-D-Geometrie sehr schnell. Der Punktdetektions- und -verfolgungsalgorithmus benötigt dagegen Punktmerkmale, die sich gut verfolgen lassen. Deshalb ist das Objekt für das Experiment mit verschiedenen Bildern „beklebt“. Das Objekt ist daher aus den gleichen zwei Ansichten in Bild 6.36c und 6.36d mit der im Experiment verwendeten Textur dargestellt.

Die Ergebnisse der 3-D-Rekonstruktion zeigt Bild 6.37 in Kombination mit Tabelle 6.12. Hier ist auffällig, dass der Fehler insgesamt bei allen Varianten doch deutlich größer ist als bei den vorangegangenen Experimenten. Das liegt zum einen daran, dass dieses Objekt aufgrund seiner verschiedenen Merkmale, wie z. B. die sinusförmige Fläche, die abgeschnittene Ecke mit der eingelassenen Halbkugel und den Bohrlöchern, viel komplexer ist als die vorherigen Objekte.

Darüber hinaus soll nochmal daran erinnert werden, dass das Fehlermaß hier ein anderes ist: Da es sich hier nicht mehr um N-planare Flächen handelt, kann das Fehlermaß aus Abschnitt 6.1.3 nicht benutzt werden. Stattdessen wird hier das Fehlermaß für Objekte mit nichtplanaren Flächen aus Abschnitt 6.1.4 verwendet. Die zugehörige Referenzpunktwolke, die als Ground-Truth-Daten dient, besteht aus 815960 Punkten.

Aufgrund des anderen Fehlermaßes können die absoluten Größenordnungen an den Fehlerkurven hier nicht direkt verglichen werden. Sehr wohl kann man

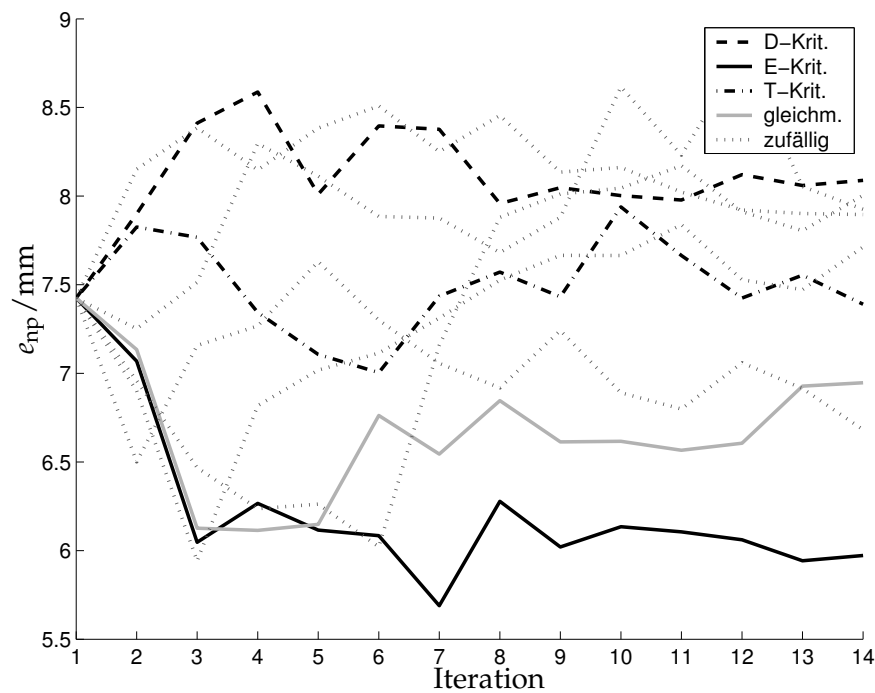


Bild 6.37 Rekonstruktion des Prüfkörpers. Die Grafik zeigt den Fehler e_{np} der Rekonstruktion bei jedem Iterationsschritt für die verschiedenen Gütekriterien sowie bei gleichmäßiger Abtastung und bei fünf Bildsequenzen mit zufälliger Ansichtsauswahl.

	Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13	14
E-Kriterium	sichtbar	49	52	61	42	69	77	59	103	64	74	84	70	79	89
	gesamt	72	130	139	186	217	217	234	238	238	241	241	241	269	269
D-Kriterium	sichtbar	49	43	53	81	61	45	35	74	73	109	91	81	79	72
	gesamt	72	153	234	241	244	288	344	360	360	360	360	360	360	360
T-Kriterium	sichtbar	49	53	54	41	75	49	62	84	41	65	62	72	64	68
	gesamt	72	138	176	207	221	221	233	239	240	260	260	262	262	262
gleichmäßig	sichtbar	49	58	61	47	30	41	48	48	54	48	38	47	51	57
	gesamt	72	97	105	107	114	138	149	158	169	177	190	210	219	225
zufällig 1	sichtbar	49	52	39	54	51	55	55	68	52	53	66	62	53	100
	gesamt	72	131	157	199	211	239	242	248	248	265	301	301	301	309
zufällig 2	sichtbar	49	53	55	56	61	57	88	94	99	65	86	85	91	101
	gesamt	72	128	154	230	257	299	303	305	305	314	316	317	317	321
zufällig 3	sichtbar	49	32	57	52	85	88	43	60	79	84	113	70	84	102
	gesamt	72	113	201	310	322	323	344	358	358	358	358	358	359	359
zufällig 4	sichtbar	49	48	74	56	76	87	44	62	71	60	64	61	53	61
	gesamt	72	151	208	208	226	226	292	306	337	337	337	338	348	350
zufällig 5	sichtbar	49	57	48	63	84	82	58	47	54	60	106	57	89	97
	gesamt	72	105	119	146	150	150	199	232	245	254	260	268	270	275

Tabelle 6.12 Die Tabelle zeigt für alle Rekonstruktionen des Prüfkörpers die Anzahl der sichtbaren Punktmerkmale und die Gesamtzahl der rekonstruierten Punkte in jedem Iterationsschritt.

aber die Tendenz vergleichen. Dabei zeigt sich erneut, dass das modifizierte E-Kriterium die besten Ergebnisse liefert. Dagegen liefert das T-Kriterium bei diesem komplexeren Objekt schlechtere Ergebnisse als die gleichmäßige Abtastung und teilweise sogar schlechtere als die zufälligen Abtastungen. Dies lässt den Schluss zu, dass die heuristisch motivierte T-Optimalität zwar bei relativ einfachen Objekten durchaus gute Ergebnisse liefert, bei komplexeren Objekten jedoch versagt. Bei

diesem Objekt und auch bei den einfacheren, die nur aus planaren Flächen bestehen, liefert die geometrisch motivierte modifizierte E-Optimalität stabil gute Rekonstruktionsergebnisse. Nur bei den einfachen Objekten ohne Selbstverdeckung konnte manchmal ein anderes Kriterium etwas bessere Ergebnisse liefern. Dafür versagen die anderen Kriterien bei den Objekten mit Selbstverdeckungen völlig und sind häufig schlechter als die gleichmäßige Abtastung oder gar die zufällige Ansichtenauswahl.

6.5 Zusammenfassung und Diskussion der Ergebnisse

Das Experimentekapitel teilt sich in insgesamt vier Teile. In diesen werden die theoretischen Überlegungen aus Kapitel 3 bzw. Kapitel 4 in ein Gesamtsystem, wie in Kapitel 5 beschrieben, integriert. Dann erfolgen die Auswertungen anhand von Simulationen und echten Aufnahmen.

Zunächst werden im Abschnitt 6.1 Gütekriterien aufgestellt, mit deren Hilfe die eigentlichen Experimente in den folgenden Abschnitten bewertet werden können. Dabei sind vier verschiedene Fehlermaße vorgestellt worden, die je nach Objekt für die Bestimmung der Rekonstruktionsgüte eingesetzt werden. Das Kalibriermuster lässt dabei als einziges eine direkte Fehlerberechnung zu, während bei allen anderen, komplexeren Objekten andere Mechanismen herangezogen werden müssen. Für Objekte, die nur aus einer planaren Fläche bestehen, wird die Regressionsebene berechnet. Dieses Problem wurde als Minimierungsproblem beschrieben und die Güte der Rekonstruktion ist das Ergebnis der Minimierungsaufgabe. Bei Objekten, die aus mehreren planaren Flächen bestehen, werden die planaren Flächen zunächst als Referenzpunktewolke diskretisiert. Mit Hilfe eines Iterative-Closest-Point-Algorithmus wird die Transformation, bestehend aus Translation und Rotation, zwischen dieser Referenzwolke und der Punktewolke der rekonstruierten Punkte bestimmt. Mit Hilfe der Transformationsparameter lassen sich die Ebenen, die das Objekt beschreiben, nun so transformieren, dass sie bestmöglich zu den rekonstruierten Punkten passen. Der mittlere Abstand der rekonstruierten Punkte zur jeweils nächsten Ebene beschreibt die Güte der Rekonstruktion. Ähnlich wird vorgegangen, falls das Objekt keine planaren Oberflächen aufweist. Hier wird zunächst mit Hilfe eines ICP-Algorithmus die Transformation vom originalen Koordinatensystem in das der Rekonstruktion berechnet. Anschließend wird für jeden rekonstruierten Punkt der Abstand zu allen Punkten aus der Referenzwolke berechnet und der minimale Abstand als tatsächlicher Abstand angenommen. Die Verfahren mit dem Ansatz über eine Ebene haben den Nachteil, dass rekonstruierte Punkte sich in der Ebene bewegen können, ohne dass sich der Fehler ändert. Analog gilt diese Aussage auch für das Verfahren beim Abgleich mit einer Referenzpunktewolke.

Im folgenden Abschnitt 6.2 werden die theoretischen Vorhersagen aus Kapitel 3 experimentell überprüft. Dabei geht es in Stereosystemen um die Analyse, welche gezielt veränderbaren Parameter das Rekonstruktionsergebnis wie stark beeinflussen. Die einzelnen Parameter Brennweite, Translation und Rotation werden dabei zunächst einzeln betrachtet und anschließend ein globales Optimum über alle Parameter gesucht.

Der Zusammenhang zwischen Brennweite und Rekonstruktionsergebnis entsprach dabei den Erwartungen: Eine große Brennweite verbessert das Rekonstruktionsergebnis. Bei den Betrachtungen zur Translation wurde zum einen nur eine Translation parallel zur Bildebene und zum anderen die Translation auf das Objekt zu betrachtet. Dort zeigte sich, dass prinzipiell eine große Stereobasis, was einer großen Translation parallel zur Bildebene entspricht, wichtig für ein gutes Ergebnis ist. Wird eine Translation auf das Objekt zu erlaubt, so kann die Translation nicht in beide Richtungen beliebig groß werden, da das Objekt irgendwann nicht mehr im Bild sichtbar ist. Stehen diese beiden Translationsrichtungen in Konkurrenz, so zeigt sich, dass eine Kamera, die möglichst nahe am Objekt positioniert ist, deutliche Vorteile bringt gegenüber der alleinigen Vergrößerung der Stereobasis.

Bezüglich der Rotation konnte das überraschende Ergebnis aus den theoretischen Überlegungen in Abschnitt 3.2.1.5 bestätigt werden. Bei einem Wegdrehen der Kameras reduziert sich der Rekonstruktionsfehler. Dies konnte nun auch in den realen Experimenten beobachtet werden.

Zum Abschluss des Abschnitts über die Stereoparameter wurden alle Parameter und ihre Wechselwirkung im Hinblick auf die Rekonstruktionsgenauigkeit analysiert. So kann nicht gleichzeitig die Brennweite vergrößert und der Abstand zwischen Kamera und Objekt beliebig unabhängig voneinander optimiert werden, da sonst das Objekt nicht mehr komplett im Bild liegt. Hier zeigt sich beispielsweise, dass die Rotation alleine genommen einen großen Einfluss auf die Rekonstruktionsgüte hat und dass die Kameras weggedreht werden sollten. Beim Zusammenspiel aller Parameter ist der Effekt der Rotation vergleichsweise gering, so dass hier die Kameras zum Objekt hin rotiert werden.

In den letzten beiden Abschnitten des Experimentekapitels geht es um die sukzessive 3-D-Rekonstruktion aus Kapitel 4 und 5.

In den Versuchsreihen werden die drei in Abschnitt 4.3 vorgestellten Optimalitätskriterien untereinander und mit jeweils einer gleichmäßigen Abtastung und zufällig ausgewählten Ansichten verglichen. Dabei werden die Objekte stetig komplexer. Begonnen wird in Abschnitt 6.3 mit Objekten, die keine Selbstverdeckungen aufweisen. Bei diesen Experimenten schneiden die drei Optimalitätskriterien in der Regel ähnlich gut ab. Dabei liefern sie fast durchgängig bessere Werte als die gleichmäßige Abtastung und die zufällige Auswahl. Nur bei der Rekonstruktion des Weihnachtsmanns (Abschnitt 6.3.4) schneidet die gleichmäßige Abtastung in den ersten paar Iterationen geringfügig besser ab.

Im Abschnitt 6.4 wird die Komplexität der Objekte sukzessive weiter erhöht und dabei dürfen nun auch Selbstverdeckungen auftreten. Dabei ergibt sich zu den Rekonstruktionsergebnissen aus dem Abschnitt 6.3 ohne Selbstverdeckungen doch eine unterschiedliche Charakteristik in den Fehlerkurven. So gibt es gelegentlich deutliche Sprünge des Rekonstruktionsfehlers nach oben. Dies kann dadurch erklärt werden, dass jetzt nicht mehr alle Punktmerkmale über die gesamte Zeit verfolgt werden können. Deshalb werden Punktmerkmale vom Punktverfolger verloren, dafür werden neue gefunden und diese weiter verfolgt. Diese neuen Punktmerkmale sind zunächst sehr ungenau und deshalb kann die Kurve der Rekonstruktionsgenauigkeit sprunghaft ansteigen.

Ein anderer Aspekt der Selbstverdeckungen bezüglich der Punktverfolgung ist, dass die verloren gegangenen Punktmerkmale zwar wieder gesucht werden, jedoch ist dies nicht immer erfolgreich. In diesem Fall sind Punkte in der 3-D-Rekon-

struktion enthalten, deren Positionsbestimmung durch hinzufügen weiterer Aufnahmen nicht weiter verbessert werden kann. Die Kurven verharren dann auf einem gewissen Niveau. Dieser Effekt war bei den Versuchen ohne Selbstverdeckungen dagegen nicht feststellbar. Zwar gehen dort Punktkorrespondenzen verloren und werden teilweise nicht wiedergefunden, jedoch in weitaus geringerem Umfang als das bei den Experimenten mit Selbstverdeckung der Fall ist.

Bei den Experimenten mit Selbstverdeckungen zeigte sich, dass D-, E-, und T-Optimalität sehr unterschiedliche Resultate liefern. Das geometrisch motivierte E-Kriterium lieferte dabei die besten Ergebnisse. Im Gegensatz dazu lieferte die T-Optimalität schon wesentlich schlechtere Werte, während die auf der Entropie basierende D-Optimalität häufig noch schlechter abschnitt. Teilweise war die D-Optimalität sogar schlechter als Ansätze ohne Optimierung. Es ist also festzuhalten, dass das geometrisch motivierte E-Kriterium für die Rekonstruktionsgenauigkeit stabil, und verglichen mit den anderen Kriterien, am besten zur Reduktion von e_p bzw. e_{np} geeignet ist.

Kapitel 7

Zusammenfassung und Ausblick

Die in dieser Arbeit beschriebenen Verfahren zur Ansichtenauswahl für die 3-D-Rekonstruktion statischer Szenen liefern Ansätze, wie durch gezielte Manipulation der Kameraparameter das Rekonstruktionsergebnis verbessert werden kann. In Abschnitt 7.1 werden einige Ansätze vorgestellt, die eine sinnvolle Ergänzung und Weiterentwicklung dieser Arbeit darstellen. Abschließend wird in Abschnitt 7.2 ein Überblick über die wichtigsten Thesen und Ergebnisse der Arbeit gegeben.

7.1 Ausblick

Varianten für die Punktverfolgung In dieser Arbeit wird sowohl ein geometrisch basierter Ansatz für Stereoaufnahmen als auch ein probabilistischer Ansatz für die Ansichtenplanung zur 3-D-Rekonstruktion vorgestellt. Beide Ansätze verarbeiten dabei letztendlich Punktkorrespondenzen, die durch einen von Zinßer [Zin11] erweiterten Kanade-Lukas-Tomasi (KLT) Punktverfolger berechnet wurden. Dieser Algorithmus zur Punktverfolgung eignet sich prinzipiell auch für handgeführte Kameras. Das bedeutet im Umkehrschluss, dass das Problem der Punktverfolgung in einer Allgemeinheit gelöst wird, die für dieses spezielle Problem gar nicht nötig ist. Schließlich bewegt sich die Kamera auf einer bekannten Bahn, und somit lässt sich der Suchbereich für die Punktkorrespondenzen mithilfe der Epipolargeometrie einschränken. Eine mögliche Umsetzung dieser Idee wird in der Arbeit von Trummer [Tru11] vorgestellt.

Prinzipiell lässt sich aufgrund des modularen Aufbaus (vgl. Kapitel 5) des probabilistischen Ansatzes ein gänzlich anderer Punktverfolger einsetzen. In diesem Gebiet sind in den letzten Jahren verschiedene neue Algorithmen erforscht worden. An dieser Stelle werden exemplarisch zwei genannt, die inzwischen eine relativ weite Verbreitung im Bereich Rechnersehen genießen: SIFT (Scale Invariant Feature Transform) [Low04] und SURF (Speeded Up Robust Features) [Bay06]. Diese Algorithmen versprechen dabei eine höhere Unempfindlichkeit gegenüber Transformationen und Rauschen im Bild. Was dabei für den hier vorgestellten Anwendungsfall noch wichtiger ist, ist, dass diese Algorithmen ein zeitweise verdecktes Objekt mit sehr hoher Wahrscheinlichkeit wiederfinden können.

Solche Punktverfolger sind auch bei Stereokamerasystemen interessant. Denn hier kann man die Punktkorrespondenzen, wie bereits mehrfach erwähnt, mithilfe des hier verwendeten differentiellen Verfahrens schlecht schätzen. Ein Punktverfolger, der auf SIFT- oder SURF-Merkmalen basiert, verspricht wesentlich robustere Ergebnisse. Dagegen wird SIFT- und SURF-Merkmalen häufig angelastet, dass sie weniger präzise als ein KLT-Punktverfolger seien. In der Literatur finden sich

keine eindeutigen Aussagen. Vielmehr kommen unterschiedliche Arbeiten je nach Einsatzgebiet, vgl. z. B. [Gov09, Kli07, Luó10], zu unterschiedlichen Ergebnissen. Daher stellt eine Untersuchung der Punktverfolger, im Kontext der hier vorgestellten Applikation, eine sinnvolle Ergänzung dieser Arbeit dar.

Ein anderer Aspekt, der auch dem Punktverfolger geschuldet ist, ist die affine Verzerrung des Merkmalsfensters während der Aufnahmen. Der eingesetzte Punktverfolger von Zinßer schätzt diese affine Transformation und gleicht sie somit aus. Es lässt sich auch anders auf diese Problematik eingehen. Statt nur die 3-D-Position eines Merkmals zu bestimmen, könnte zusätzlich die Richtung der Normalen an diesem Punkt geschätzt werden. Dadurch lässt sich bei den darauf folgenden Aufnahmen der Winkel zwischen der Normalen und dem Sichtstrahl berechnen. Dieser Winkel kann dazu verwendet werden, die Schätzung der affinen Verzerrung zu verbessern. Andererseits kann die Information über den Winkel auch dazu dienen, das Beobachtungsrauschen im probabilistischen Ansatz realistischer zu gestalten, anstatt immer von einem konstanten Rauschen auszugehen: Sind Sichtstrahl und Normale identisch, so ist das Rauschen minimal. Bei größer werdendem Winkel zwischen Normalen und Sichtstrahl wird das Beobachtungsrauschen entsprechend größer.

Erweiterung des Stereoansatzes Der hier vorgestellte Stereoansatz beinhaltet bislang vor allem die Optimierung der entsprechenden Kameraparameter. Hier sind gleich zwei Erweiterungen denkbar.

Zum einen werden hier nur zu einer initialen Schätzung der 3-D-Punktkoordinaten, die optimalen Stereoparameter berechnet. Die Vorgehensweise lässt sich in einen iterativen Algorithmus überführen, der, ähnlich wie beim probabilistischen Ansatz, zwischen 3-D-Rekonstruktion und Planung einer neuen Stereoansicht alterniert und so iterativ das Rekonstruktionsergebnis verbessert.

Zum anderen ist man prinzipiell nicht auf Stereoansichten beschränkt. Man kann hier stattdessen Multikamerasysteme einsetzen, die aus drei oder mehreren Kameras bestehen. Interessant wäre hier beispielsweise die Fragestellung, welche Beziehung zwischen Anzahl der Kameras und Rekonstruktionsgüte besteht. Wenn, wie in dieser Arbeit gezeigt, ein Winkel zwischen den Sichtstrahlen von 90° optimal ist, lässt sich möglicherweise keine Verbesserung bei mehr als 3 Kameras erreichen – unter der Voraussetzung eines nicht probabilistischen Fehlers, der sich durch die Hinzunahme von weiteren Aufnahmen „herausmittelt“.

Optimierungsverfahren für die direkte Kinematik Für diese Optimierungsaufgabe wurde die beste Ansicht durch eine vollständige Suche im vorher diskretisierten Parameterraum durchgeführt. Dadurch ist sichergestellt, dass im diskretisierten Lösungsraum tatsächlich das globale Optimum gefunden wird. Es lässt sich auch überlegen, wie mithilfe eines „intelligenteren“ mathematischen Ansatzes die Lösung schneller gefunden werden kann.

Dabei ergibt sich bei einer solchen Vorgehensweise das Problem, dass ähnliche und damit fast gleichwertige Kameraparameter zwar im $\mathbb{R}^3$ nahe beieinander liegen können, aber im Lösungsraum der direkten Kinematik u. U. weit auseinander liegen, was exemplarisch in Bild 2.7 skizziert ist. Dies hat zur Folge, dass es viele verschiedene lokale Minima gibt, jedoch das globale gesucht werden muss, da ansonsten der Ansatz unter seinen Möglichkeiten bleibt.

Eine Alternative dazu wäre, die Kameraparameter direkt zu optimieren und das Problem der inversen Kinematik zu lösen. Hierzu muss der Arbeitsbereich des Roboters bekannt sein und in die Optimierung mit eingearbeitet werden, so dass die dort errechnete Position auch tatsächlich angefahren werden kann. Darüber hinaus ist zu beachten, dass für solche Ansätze auch die Sichtbarkeitsbedingung als Nebenbedingung direkt mit integriert werden muss.

Berücksichtigung der Positionier- und Bewegungsungenauigkeit Der verwendete Roboter verursacht aufgrund seines mechanischen Aufbaus gewisse systematische und zufällige Fehler bei der Positionierung der Kameras. Diese Fehlerquelle wurde in dieser Arbeit vernachlässigt. Es gibt im probabilistischen Ansatz eine Möglichkeit, auch diese Fehlerquelle mit in den Ansatz zu integrieren. Die Rekonstruktion erfolgt dabei nicht im Weltkoordinatensystem des Roboters, sondern immer relativ zum aktuellen Kamerakoordinatensystem. Auf diese Art lässt sich die Bewegung der Kamera als Zustandsübergang im Kalman-Filter betrachten.

Für diesen Ansatz ist dann zu beweisen, inwiefern die gemachten Annahmen bezüglich der Unabhängigkeit der 3-D-Punkte noch gegeben sind. Andernfalls wäre die Zerlegung des Gesamtproblems für die Rekonstruktion in einzelne Kalman-Filter nicht mehr möglich, was die gesamte Berechnung der 3-D-Rekonstruktion um ein Vielfaches aufwändiger macht. Schließlich ist noch experimentell zu überprüfen, ob dieses sicherlich realistischere Modell zu einer signifikanten Verbesserung der 3-D-Rekonstruktion führt. So liegt z. B. die vom Hersteller angegebene Wiederholgenauigkeit des Scrobot ER VII bei 0,1mm. Die Wiederholgenauigkeit sagt zwar nichts über die absolute Positioniergenauigkeit aus, jedoch lässt sich aufgrund der Größenordnung vermuten, dass andere Faktoren, wie z. B. die Genauigkeit des verwendeten Punktverfolgers, die Güte der Kamerakalibrierung oder der Hand-Auge-Transformation, einen mindestens ebenso großen Einfluss haben.

Vollständigkeit Ein Aspekt, der in dieser Arbeit nicht behandelt wird, ist die Optimierung der Vollständigkeit der Rekonstruktion. Vollständigkeit hat dabei zwei unterschiedliche Aspekte: Zum einen lässt sich darunter verstehen, ob alle geometrischen Details des betrachteten Objekts gefunden werden. Dieser Aspekt ist in dem Ansatz nur mittelbar integriert. Nämlich dann, wenn die entsprechende Oberfläche an diesen Stellen verwertbare Punktmerkmale besitzt. Dort, wo keine Punktmerkmale zu finden sind, kann auch keine Rekonstruktion stattfinden.

Der zweite Aspekt ist die gezielte Suche nach noch nie beobachteten Bereichen des Objekts. Der vorgestellte Algorithmus konzentriert sich auf die Verbesserung der Schätzung der 3-D-Koordinaten von bekannten Punkten. Dies bedeutet im Umkehrschluss, dass der Algorithmus keine Ansicht von der Rückseite eines Objekts vorschlägt, wenn auf der Rückseite bislang keine Punkte detektiert worden sind. Dabei wird sofort offensichtlich, dass für die Verbesserung der Genauigkeit der bekannten 3-D-Punkte andere Ansichten nötig sind als für die Exploration einer Szene oder eines Objekts. Trotzdem sind natürlich Ansätze denkbar, die beide Gütekriterien, d. h. möglichst hohe Genauigkeit und Vollständigkeit, in entsprechender Gewichtung gleichzeitig optimieren, wie beispielsweise in [Tru11] vorgeschlagen.

Wegplanung im probabilistischen Ansatz Da im probabilistischen Ansatz die Beobachtung selbst keinen Einfluss auf die Kovarianz der Zustandsschätzung hat, sondern nur die Kovarianz des Beobachtungsrauschens, lässt sich folgern, dass die eigentliche Beobachtung für die Planung erstmal keine Rolle spielt. Dies lässt den naheliegenden Schluss zu, dass der iterative Ansatz aus Kapitel 5 sich nicht nur dazu verwenden lässt, die nächste beste Ansicht, sondern mehrere nächstbeste Ansichten im Voraus zu berechnen.

Damit ließe sich beispielsweise die Bewegung eines Roboterarms hinsichtlich der zurückgelegten Wegstrecke optimieren. Dabei ist das Rekonstruktionsergebnis nicht identisch, wenn zunächst andere Ansichten für die Rekonstruktion herangezogen werden. Außerdem spielt hier der Punktverfolger eine entscheidende Rolle: Es kann zwar zur Strategie einer konkreten Anwendung sein, dass möglichst kurze Wege zurückgelegt werden sollen, jedoch gehen auf diesen Wegen Punktmerkmale verloren und diese werden durch neue ersetzt. Dieser Vorgang ist dabei nicht vorhersagbar und damit werden die weiteren bereits vorhergesagten Schritte im Allgemeinen nicht mehr optimal sein.

7.2 Zusammenfassung

In dieser Arbeit geht es um die Auswahl von Ansichten für die 3-D-Rekonstruktion statischer Szenen aus Intensitätsbildern. Dafür werden zwei verschiedene Ansätze entwickelt, die eine möglichst akkurate 3-D-Rekonstruktion zum Ziel haben. Dabei handelt es sich um einen geometrischen Ansatz, der die Parameter eines Stereokamerasystems untersucht, sowie einen probabilistisch motivierten Ansatz, der iterativ die Genauigkeit der 3-D-Rekonstruktion verbessert.

In Kapitel 1 wird zunächst das Thema 3-D-Rekonstruktion durch die Vorstellung verschiedener Anwendungsbereiche motiviert. Um die in dieser Arbeit vorgestellten Verfahren besser einordnen zu können, wird kurz auf die mögliche Sensorik eingegangen. Dabei unterscheiden sich aktive Systeme, wie beispielsweise PMD-Kameras, Laserscanner oder Streifenprojektoren, von passiven, welche mittels Kameras Intensitätsbilder aufzeichnen, darin, wie aus den vom Sensor gewonnenen Informationen die 3-D-Rekonstruktion berechnet wird. Dies wirkt sich direkt darauf aus, wie Ansichten für eine möglichst akkurate 3-D-Rekonstruktion geplant werden können. Für die in dieser Arbeit gewählte 3-D-Rekonstruktion aus Intensitätsbildern wird anhand einiger einfacher Beispiele klar gemacht, dass mindestens zwei nichtdegenerierte Aufnahmen der Szene vorliegen müssen, um überhaupt eine Rekonstruktion berechnen zu können. Ferner wird gezeigt, dass die jeweilige Aufnahmeposition bei verrauschten Messdaten des Sensors einen nicht unerheblichen Einfluss auf das Rekonstruktionsergebnis hat.

Die mathematischen Grundlagen für die in dieser Arbeit benutzten und entwickelten Algorithmen werden in Kapitel 2 eingeführt. Bei der Definition der Kameraparameter ist dabei besonders wichtig, welche Parameter für die Beschreibung der Projektion gezielt verändert, d. h. für den Zweck der Ansichtenauswahl für die 3-D-Rekonstruktion optimiert, werden können. In der Arbeit werden als gezielt veränderbare Parameter alle extrinsischen (Rotation und Translation) und die Brennweite als intrinsischer Kameraparameter betrachtet.

Bei der Betrachtung des Zusammenhangs zwischen perspektivischen und paraperspektivischen Projektionsmodell ist insbesondere Satz 2.5 interessant. Hier wird gezeigt, dass die paraperspektivische Projektion eine Taylor-Reihen-Approximation 1. Ordnung für das perspektivische Projektionsmodell ist. Für das beim probabilistischen Ansatz verwendete erweiterte Kalman-Filter wird eine Taylor-Reihen-Approximation 1. Ordnung verwendet, also genau die Approximation, die auch sonst in der Literatur eine weite Verbreitung besitzt.

In den Abschnitten über Kamerakalibrierung und Hand-Auge-Transformation wird gezeigt, wie die unterschiedlichen Kameraparameter in einem realen System mittels Kalibriermuster bestimmt werden können. Entscheidend ist dabei, wie die auf einem Roboterarm montierte Kamera in eine definierte Position gebracht werden kann. Dabei zeigt sich, dass das Problem der direkten Kinematik sehr viel leichter zu lösen ist, als das der inversen. Deshalb wird im späteren Verlauf die Optimierung der Kamerapositionen so ausgelegt, dass nur die direkte Kinematik gelöst werden muss.

Im letzten Abschnitt aus Kapitel 2 wird kurz auf den verwendeten Punktverfolger sowie die Triangulation von 3-D-Punkten eingegangen. Da sich die Arbeit mit der 3-D-Rekonstruktion aus Intensitätsbildern beschäftigt, also keine aktiven Sensoren verwendet werden, müssen aus dem Bild Punktmerkmale extrahiert und im zweiten Bild bei einer Stereoaufnahme bzw. in den folgenden Bildern beim probabilistischen Ansatz wiedergefunden werden. Durch die jeweils korrespondierenden Punktpaare aus den Bildern kann die Position des jeweiligen Punkts im Weltkoordinatensystem berechnet werden.

In Kapitel 3 wird zunächst die Unterscheidung von Ansätzen in geometrische und probabilistische verdeutlicht. Ferner werden dort die Ideen für zwei geometrische und ein probabilistisches Verfahren aus der Literatur aufgezeigt, um diese von jenen in dieser Arbeit entwickelten Verfahren abzugrenzen und die jeweiligen Vor- und Nachteile einordnen zu können.

Bei den Verfahren aus der Literatur handelt es sich zum einen um den Ansatz von Marchand und Chaumette [Mar94, Mar96, Cha94, Cha96]. Dort werden die jeweiligen Parameter geometrischer Grundformen, wie Kugeln, Geraden und Zylinder, aufgrund von mathematischen Überlegungen aus den Messungen im Bild geschätzt. Dazu lassen sich für die jeweiligen Grundformen bestimmte Regeln für die Bewegung der Kamera ableiten, so dass die Parameter mit hoher Genauigkeit geschätzt werden können. Der Nachteil ist hier offensichtlich: Die Szene muss genau aus den verschiedenen geometrischen Grundformen aufgebaut sein. Diese Einschränkung gilt für die in dieser Arbeit entwickelten Algorithmen nicht.

Der zweite Ansatz von [Kut94b] verfolgt Punkte auf der Konturlinie zwischen Objekt und Hintergrund und berechnet darüber die Krümmungsradien an diesen Punkten. Über die Krümmung lassen sich die Oberflächenstruktur und deren Parameter bestimmen. Eines der Hauptprobleme ist dabei die Klassifikation, ob eine elliptische, hyperbolische oder parabolische Oberfläche vorliegt sowie die Behandlung von Unstetigkeitsstellen, die bei planaren Oberflächen entstehen.

Der in dieser Arbeit entwickelte geometrische Ansatz geht einen anderen Weg und untersucht mithilfe des eingeführten Fehlermodells die Auswirkungen der Stereoparameter auf das Rekonstruktionsergebnis. Die Idee des Fehlermodells ist, dass ein Merkmalspunkt aus einem Bild im anderen Bild mit Hilfe eines Punktdetektors wiedergefunden wird. Dabei werden zunächst in einem vereinfachten

zweidimensionalen Modell folgende Annahmen getroffen: Alle y -Komponenten sind null und die Koordinaten im rechten Bild sind nicht verrauscht, sondern exakt. Ferner wird die Nebenbedingung einer begrenzten Bildebene vernachlässigt, d. h. die Kamera hat ein unendlich großes Sichtfeld. In diesem stark vereinfachten Modell kann die Abhängigkeit der Brennweite sogar analytisch betrachtet werden. So lässt sich zeigen, dass die Fehlerfunktion bezüglich Brennweite monoton fallend ist. Die Translation in x lieferte für die linke Kamera eine möglichst große Translation, während bei der rechten Kamera der Sichtstrahl senkrecht auf die Bildebene treffen sollte, was einem Winkel zwischen den Sichtstrahlen von 90° entspricht. Bei zulässiger Translation in x - und z -Richtung lässt sich zeigen, dass die linke Kamera so nah wie möglich an den 3-D-Punkt positioniert werden sollte, während die Position der rechten Kamera, also der ohne fehlerhafte Messung auf der Bildebene, irrelevant ist. Bei der Rotation zeigt sich, dass das Wegdrehen der Kamera den Fehler minimiert. Dies scheint zunächst der Intuition zu widersprechen, man kann sich dies anhand von Skizzen (siehe Bild 3.4 und 3.9) leicht plausibel machen.

Neben dem zweidimensionalen Fehlermodell wurde zusätzlich ein dreidimensionales Modell analysiert. Dieses ist nicht nur aufgrund der weiteren Dimension komplexer, sondern berücksichtigt zusätzlich eine begrenzte Bildebene als Nebenbedingung. Außerdem wird nicht nur ein Punkt, sondern 25 Punkte rekonstruiert. Da eine analytische Betrachtung hier zu kompliziert ist, wurde stattdessen eine spezielle Konfiguration untersucht. Wird jeweils nur einer der Parameter Brennweite, Translation in x und y -Richtung, Translation in x -, y - und z -Richtung sowie Rotation analysiert, so können die Ergebnisse aus dem zweidimensionalen Modell direkt übertragen werden. Durch die Berechnung an der Beispielkonfiguration zeigt sich, wie stark der jeweilige Parameter zur Optimierung beiträgt. Dabei ist jetzt zu beachten, dass sich durch die endliche Bildebene jeder Parameter nicht mehr beliebig verändern darf. Es zeigt sich, dass die Ausgangskonfiguration mit einem Fehler von 14,1 LE (Längeneinheiten) durch eine Optimierung der Stereobasis auf 9,14 LE gesenkt werden kann. Die Stereobasis hat hier also einen relativ geringen Einfluss, denn durch Vergrößerung der Brennweite kann der Fehler auf 3,20 LE und durch Rotation auf 2,71 LE gesenkt werden. Ein deutlich besseres Ergebnis von 0,61 LE erreicht man, wenn man auch Translationen in z -Richtung, d. h. in Richtung auf die 3-D-Punkte zu, erlaubt. Bei der gleichzeitigen Optimierung aller veränderbaren Parameter kann dieses Ergebnis nur noch geringfügig auf 0,46 LE verbessert werden.

Als probabilistischer Ansatz aus der Literatur ist der Ansatz von Olague und Mohr [Ola97, Ola98, Ola02] näher erläutert. In diesem Ansatz wird durch eine Taylor-Reihen-Approximation ein linearer Zusammenhang zwischen den normalverteilten Weltkoordinaten des 3-D-Punkts und den Bildkoordinaten hergeleitet. Mit Hilfe des Doppelverhältnisses lässt sich dabei die Kovarianz in den Bildkoordinaten von vier kollinearen Punkten schätzen. Im Ansatz werden dabei die m Kamerapositionen gesucht, die insgesamt die Kovarianz minimieren. Als Metrik für die Größe der Kovarianz werden der maximale Eigenwert der Kovarianz, die Spur der Kovarianz und der maximale Wert der Diagonalen der Kovarianz betrachtet, was große Ähnlichkeit zu den in dieser Arbeit entwickelten Kriterien der D-, E- und T-Optimalität aufweist. Im Unterschied zu dem in Kapitel 4 verwendeten Ansatz benötigt der aus der Literatur jedoch immer mindestens vier kol-

lineare Punkte. Dies wird dadurch gewährleistet, dass die Punkte aufgrund von extra angebrachten Markern detektiert und verfolgt werden können und die Marker diese Kollinearität gewährleisten. Ferner unterscheidet sich der Algorithmus von Olague und Mohr dadurch, dass die m Kamerapositionen sich zum einen auf einer Halbkugel befinden müssen, zum anderen werden diese Kamerapositionen simultan optimiert. Der in der Arbeit entwickelte Algorithmus hingegen arbeitet iterativ, d. h. es wird immer, basierend auf der aktuellen Schätzung, die eine nächste beste Ansicht gesucht.

Der in dieser Arbeit entwickelte probabilistische Ansatz für die Ansichtenplanung basiert auf der Zustandsschätzung linearer dynamischer Systeme mithilfe des erweiterten Kalman-Filters. Da die perspektivische Projektion von 3-D-Koordinaten auf die 2-D-Bildebene nichtlinear ist, muss explizit darauf hingewiesen werden, dass sich dadurch Erwartungswert und Kovarianz nicht durch eine lineare Transformation berechnen lassen. Vielmehr wurde an einem Beispiel gezeigt, dass sich u. U. Verteilungen ergeben, die weder Erwartungswert noch Varianz besitzen. Die Linearität wird in diesem Fall dadurch hergestellt, dass die nichtlineare perspektivische Projektion durch eine Taylor-Reihen-Approximation 1. Ordnung linearisiert wird. Dies ist, wie bereits erwähnt, identisch mit dem paraperspektivischen Projektionsmodell, dass im Bereich der Literatur zur 3-D-Rekonstruktion eine weite Verbreitung findet.

Das Problem der 3-D-Rekonstruktion kann also auf ein Zustandsschätzproblem, welches mit dem Kalman-Filter zu lösen ist, übertragen werden. Durch Verwendung des sequentiellen Kalman-Filters und der Annahme, dass die 3-D-Koordinaten im initialen Zustand eines Punkts statistisch unabhängig von den 3-D-Koordinaten eines anderen Punkts sind, lässt sich die Schätzung der Normalverteilung eines 3-D-Punkts mit einem Kalman-Filter unabhängig von allen anderen 3-D-Punkten beschreiben. Bei n Punkten muss also nicht ein Kalman-Filter mit einer $3n \times 3n$ Kovarianzmatrix berechnet werden, sondern n einzelne Kalman-Filter mit 3×3 Kovarianzen, wodurch der Berechnungsaufwand drastisch reduziert werden kann.

Durch diese Aufteilung ist es möglich, die Sichtbarkeit eines Punktes geschlossen in den Ansatz zu integrieren: Ein nicht sichtbarer Punkt wird keine Beobachtung erzeugen, und damit wird das entsprechende Kalman-Filter dieses Punkts nicht aufgefrischt. Zur Sichtbarkeit wurden dabei zwei Bedingungen entwickelt und in den Ansatz integriert: Die Projektion des 3-D-Punkts liegt außerhalb der Bildebene oder der 3-D-Punkt ist durch ein anderes Element des zu rekonstruierenden Objekts verdeckt (Selbstverdeckung). Der erste Fall lässt sich relativ einfach berechnen: Die Normalverteilung, die die Lage des 3-D-Punkts beschreibt, wird in die Bildebene projiziert und über die Rechtecksfläche der Bildebene wird integriert. Im zweiten Fall wird ein Dreiecksnetz durch sämtliche 3-D-Punkte aufgespannt. Um die Sichtbarkeit eines Punkts von einer bestimmten Kameraposition aus zu bestimmen, wird berechnet, ob der Sichtstrahl vom Punkt zum optischen Zentrum der Kamera eines dieser Dreiecke durchstößt.

Die Kovarianz der 3-D-Punkte lässt sich also demnach berechnen, ohne tatsächlich die Beobachtung zu tätigen. Ferner lässt sich durch einzelne Betrachtung der 3-D-Koordinaten eines jeden zu rekonstruierenden Punkts noch die Sichtbarkeit mit in die Berechnung der Kovarianz integrieren. Da die Kovarianz den erwarteten quadratischen Schätzfehler beschreibt, liefert eine „kleine“ Kovarianz akkuratere

3-D-Positionen. Dies ist der Ansatz, um eine gezielte Ansichtenplanung durchzuführen: Man kennt somit die Abhängigkeit der Kovarianzmatrix zu den jeweiligen gezielt veränderbaren Kameraparametern. Nun benötigt man nur noch eine Metrik, um die „Größe“ der Kovarianz zu messen.

Dazu wurden drei Maße eingeführt: die D-Optimalität, die die Entropie berechnet, die modifizierte E-Optimalität, die die Summe der größten Eigenwerte der einzelnen Kovarianzen für jeden Punkt bestimmt, sowie die T-Optimalität, die die Spur der Kovarianzmatrix auswertet.

Zur experimentellen Überprüfung der theoretischen Überlegungen aus Kapitel 3 und 4 werden in Kapitel 5 die Komponenten eines Systems zur Umsetzung vorgestellt. Dazu gehört zum einen die Beschreibung der verwendeten Roboter zur Positionierung der Kamera. Insbesondere ist hier wichtig, wie der gültige Arbeitsbereich beschrieben werden kann, um diesen bei der Optimierung zu berücksichtigen. Es wird vorgeschlagen, die Achsenstellungen des Roboters zu optimieren, um mithilfe der direkten Kinematik die 3-D-Position und Orientierung zu bestimmen. Dies ist besonders für Roboterarme mit vielen verschiedenen Gelenken, wie dem verwendeten Scrobot ER VII, geeignet, lässt sich aber ebenso für den verwendeten Drehteller mit Schwenkarm einsetzen. Mithilfe der Denavit-Hartenberg-Matrix lässt sich somit der Arbeitsbereich des Roboters beschreiben. Durch die Kamerakalibrierung und der bekannten Hand-Auge-Transformation lassen sich die tatsächlichen Kameraparameter aus der Roboterposition berechnen. Durch eine kurze Aufnahmesequenz werden mithilfe eines Punktverfolgers korrespondierende Punktmerkmale extrahiert und damit eine initiale 3-D-Rekonstruktion berechnet. Auf Basis dieser Schätzung wird nun die nächste beste Ansicht geplant, vom Roboter angesteuert und diese Beobachtung wieder in die Schätzung integriert. Anschließend wird wiederum die nächste beste Ansicht geplant, und so weiter, bis ein Abbruchkriterium erreicht ist. In dieser Arbeit wird das Abbruchkriterium durch die Anzahl der geplanten Ansichten definiert.

Eine experimentelle Überprüfung der theoretischen Ergebnisse wird in Kapitel 6 präsentiert. Dazu werden zunächst die Gütekriterien erläutert, mit denen die Präzision der 3-D-Rekonstruktion gemessen wird. Bei einem Kalibrieremuster kann der 3-D-Fehler direkt bestimmt werden. Bei Objekten, die aus einer oder mehreren planaren Flächen bestehen, werden Ebenen durch die rekonstruierte Punktwolke gelegt. Der mittlere quadratische Fehler der Punkte zu der jeweiligen Regressionebene beschreibt die Güte der Rekonstruktion. Bei Objekten mit nichtplanaren Flächen wird die rekonstruierte Punktwolke mit einer Referenzpunktwolke registriert. Die Summe über den jeweils minimalen Abstand zwischen rekonstruiertem Punkt und allen Referenzpunkten gibt die Güte der Rekonstruktion an.

Für den geometrischen Ansatz wurde dabei ein Kalibrieremuster und in einer Simulation das L-förmige konkave Objekt rekonstruiert. Im Wesentlichen lassen sich, abgesehen von einem gewissen, nicht ganz unerheblichen Rauschen in den Messungen, die Ergebnisse der Theorie bestätigen: große Brennweite, Translation hin zum Objekt und Rotation vom Objekt weg minimiert den Fehler. Es zeigt sich, dass, wenn das Objekt einigermaßen groß in der Bildebene zu sehen ist, der Rotationsparameter nur sehr eingeschränkt optimiert werden kann. Zwar hat er bei den theoretischen Überlegungen eine große Rolle gespielt, jedoch konnte im Experiment nur eine Rotation um $\pm 6^\circ$ durchgeführt werden, um das Objekt noch komplett im Bild zu haben. Bei der Simulation des L-förmigen konkaven Objekts

ergaben sich analoge Aussagen. Dies bestätigt, dass die Vorhersagen nicht nur für das planare Kalibrieremuster sondern auch für andere Strukturen gelten.

Die Ergebnisse für den probabilistischen Ansatz sind in zwei Teile gegliedert. Im ersten Teil werden die Sichtbarkeitsbedingungen vernachlässigt und deshalb nur Objekte ohne Selbstverdeckungen rekonstruiert. Die Objekte werden dabei von Experiment zu Experiment komplexer. Die Ansichten wurden aus dem D-, E- oder T-Kriterium berechnet. Zum Vergleich erfolgte zusätzlich eine Rekonstruktion zum einen aus gleichmäßig um das Objekt herum verteilten Kamerapositionen, und zum anderen aus zufällig gewählten Ansichten. Die drei Optimalitätskriterien liefern hier häufig fast identische Rekonstruktionsergebnisse. Sie sind in der Regel besser, im Experiment mit dem Weihnachtsmann nur etwa gleich gut, wie die gleichmäßige Abtastung. Hingegen versagt die gleichmäßige Abtastung im Experiment mit dem L-förmigen konkaven Objekt, in welchem die zufälligen Ansichten sogar eine bessere Rekonstruktion erzielen.

Im zweiten Teil der Experimente zum probabilistischen Ansatz werden Selbstverdeckungen berücksichtigt. Auch hier werden die Objekte von Experiment zu Experiment komplexer. Angefangen mit dem L-förmigen konvexen Objekt, einem Tetraeder, bis schließlich hin zu einem in [Mun07] vorgeschlagenen Prüfkörper für die Evaluation einer 3-D-Rekonstruktion.

Hier zeigt sich, dass die drei Kriterien D-, E- und T-Optimalität deutlich unterschiedliche Ergebnisse liefern: Hier sind die Ergebnisse des geometrisch motivierten E-Kriteriums deutlich besser als die der anderen beiden. Bei den Experimenten wurde noch ein anderer Effekt deutlich: Dadurch, dass Punkte irgendwann verdeckt werden und wieder erscheinen, ergibt sich das Problem, dass diese wiedergefunden werden müssen. Dies stellt den Punktverfolger vor größere Herausforderungen als bei Objekten ohne Selbstverdeckung. Dies hat zur Folge, dass Punkte teilweise gar nicht wiedergefunden werden, und damit kann deren Positionsschätzung nicht verbessert werden. Der Rekonstruktionsfehler sinkt damit nicht gegen null sondern verbleibt auf einem gewissen Niveau.

Diese Arbeit liefert einen wichtigen Beitrag zum Themenkomplex Ansichtenplanung für die 3-D-Rekonstruktion. Dabei wurde für den geometrischen Ansatz für Stereosysteme eine systematische Untersuchung der Abhängigkeit der Stereo- und Kameraparameter auf die Rekonstruktionsgenauigkeit abgeleitet. Weiter wird ein probabilistischer Ansatz präsentiert, der die 3-D-Rekonstruktion als iteratives Zustandsschätzproblem formuliert. Mathematisch wurde hier bewiesen, dass durch geschickt gewählte Annahmen zum einen die Problemdimension extrem reduziert und zum anderen die Sichtbarkeitsbedingungen in den Optimierungsterm direkt integriert werden können. Alle aus der Theorie abgeleiteten Ergebnisse wurden anschließend experimentell überprüft. Dabei wurden Besonderheiten, die sich bei der praktischen Umsetzung der entwickelten Algorithmen ergaben, ausführlich diskutiert.

Anhang A

Mathematische Symbole

In diesem Anhang sind alle mathematischen Symbole und Notationen zusammengefasst. Zunächst folgen allgemeine Konventionen, anschließend eine Übersicht der verwendeten Formelbuchstaben.

- Skalare werden durch kleine kursive Buchstaben dargestellt, wie a, b, c .
- Vektoren werden durch fette kursive Buchstaben dargestellt, wie $\boldsymbol{x}$
- Das i -te Element eines Vektors $\boldsymbol{x}$ wird durch x_i dargestellt.
- Matrizen werden durch große fette kursive Buchstaben dargestellt, wie $\boldsymbol{X}$.
- Das Element in der i -ten Zeile und der j -ten Spalte der Matrix $\boldsymbol{X}$ wird durch x_{ij} dargestellt.
- Der Zeilenvektor der i -ten Zeile der Matrix $\boldsymbol{X}$ wird durch ${}_i\boldsymbol{x}$ bezeichnet
- Der Spaltenvektor der i -ten Spalte der Matrix $\boldsymbol{X}$ wird durch ${}^i\boldsymbol{x}$ bezeichnet
- Die Transponierten eines Vektors $\boldsymbol{x}$ bzw. einer Matrix $\boldsymbol{X}$ werden durch $\boldsymbol{x}^T$ und $\boldsymbol{X}^T$ dargestellt.
- Die Inverse einer Matrix $\boldsymbol{X}$ wird durch $\boldsymbol{X}^{-1}$ dargestellt.
- Die euklidische Norm eines Vektors wird durch $\|\boldsymbol{x}\|$ dargestellt.
- Die Frobenius-Norm einer Matrix wird durch $\|\boldsymbol{X}\|$ dargestellt.
- Die Determinante bzw. Spur einer Matrix $\boldsymbol{X}$ wird durch $\det(\boldsymbol{X})$ bzw. $\text{spur}(\boldsymbol{X})$ dargestellt.
- Ein homogener Vektor wird durch eine Unterstreichung gekennzeichnet, wie $\underline{\boldsymbol{x}}$.
- Die Dichte einer Zufallsvariablen x wird durch $p(x)$ dargestellt.
- Eine Schätzung eines Werts x wird durch $\hat{x}$ dargestellt. $\hat{x}$ wird ebenfalls für berechnete Werte verwendet, falls sie mit den richtigen Wert verglichen werden.
- Der Mittelwert einer Schätzung eines Werts x wird durch $\bar{x}$ dargestellt.
- Der Erwartungswert einer Zufallsvariable x bzw. eines Zufallsvektors $\boldsymbol{x}$ wird durch $E(x)$ bzw. $E(\boldsymbol{x})$ dargestellt.

- Die Varianz einer Zufallsvariable x wird durch σ_x^2 dargestellt.
- Die Kovarianzmatrix eines Zufallsvektors $\mathbf{x}$ wird durch Σ_x dargestellt.
- Eine Matrix $\mathbf{Y}$, die aus einer Matrix $\mathbf{X}$ und dem angehängten Vektor $\mathbf{x}$ aufgebaut ist, wird folgendermaßen notiert: $\mathbf{Y} = (\mathbf{X}|\mathbf{x})$.
- Eine Matrix $\mathbf{X} = \text{diag}(a, b, c, \dots)$ ist eine Diagonalmatrix mit den skalaren Einträgen $a, b, c, \dots$ auf der Diagonalen.
- Eine Matrix $\mathbf{X} = \text{diag}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots)$ ist eine Blockdiagonalmatrix mit den Blöcken $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ auf der Diagonalen.

Die folgende Tabelle listet die verwendeten Symbole und ihre Bedeutung auf. Ferner ist die Seitenzahl des ersten Auftretens notiert.

p_c	3-D-Punkt in Kamerakoordinaten	4
$\underline{q}$	Homogener 2-D-Punkt in Bildkoordinaten	15
$\underline{p}_w$	Homogener 3-D-Punkt in Weltkoordinaten	15
$\mathbf{P}$	3×4 Projektionsmatrix	15
$\mathbf{K}$	Keramatrix	15
$\mathbf{P}_M$	Projektionsmodellmatrix	15
$\mathbf{F}$	Matrix der extrinsischen Parameter	15
O_Z	Optisches Zentrum	16
f	Brennweite	16
ζ_x	Länge eines Pixels in horizontale Richtung	16
ζ_y	Länge eines Pixels in vertikale Richtung	16
u_p	Die x-Koordinate des Hauptpunkts	16
v_p	Die y-Koordinate des Hauptpunkts	16
$\beta_\angle$	Winkel zwischen den beiden Achsen des Bildkoordinatensystems	16
f_x	Effektive Brennweite in horizontaler Richtung	16
f_y	Effektive Brennweite in vertikaler Richtung	16
β	Skew im Sensorkoordinatensystem	16
$\mathbf{q}_d$	Verzerrter 2-D Bildpunkt	17
$\mathbf{q}$	2-D-Punkt in Bildkoordinaten	17
δ_x	Verschiebung in x-Richtung zur Entzerrung	17
δ_y	Verschiebung in y-Richtung zur Entzerrung	17
κ	Radialer Verzerrungsparameter	17
a	Abstand eines Bildpunkts vom Hauptpunkt	17
ϑ	Tangentialer Verzerrungsparameter	17

t	Translationsvektor	17
R	Rotationsmatrix	17
p_w	3-D-Punkt in Weltkoordinaten	18
$\underline{p}_c$	Homogener 3-D-Punkt in Kamerakoordinaten	18
0_3	Nullvektor der Länge drei	18
A	Matrix A eines Gleichungssystems $Ax = b$ oder einer linearen Transformation $Ax + b$	19
b	Vektor b eines Gleichungssystems $Ax = b$ oder einer linearen Transformation $Ax + b$	19
p_g	Schwerpunkt von 3-D-Punkten	20
ϕ	Funktion für die perspektivische Projektion	21
J	Jacobi-Matrix	21
ρ	Rest einer Taylor-Reihen-Entwicklung	21
o	Landau-Symbol	21
L	Bildhomographie	23
ν	Ein Skalierungsfaktor	23
$\tilde{t}$	3-D-Translationsvektor für die Kalibrierung nach Zhang	23
$\tilde{R}$	3-D-Rotationsmatrix für die Kalibrierung nach Zhang	23
Σ	Kovarianzmatrix	24
n	Anzahl der Punkte	24
i	Index	24
$I_{2 \times 2}$	Die 2×2 Einheitsmatrix	24
x	3-D-Vektor	24
0_2	Nullvektor der Länge zwei	24
C_a	Die absolute Konik	24
k^*	Vektor für die Nebenbedingungen bei der Bestimmung der intrinsischen Parameter	24
c	Vektor, der die Einträge der absoluten Konik enthält	24
K^*	Matrix für das Gleichungssystem zur Bestimmung der intrinsischen Parameter	25
U_l	Linke Matrix der Singulärwertzerlegung	25
D_s	Mittlere Matrix der Singulärwertzerlegung	25
V_r	Rechte Matrix der Singulärwertzerlegung	25
ζ	Ein Skalierungsfaktor	26
m	Anzahl der Bilder	26
j	Index	26
E	Position und Orientierung der Kamera	28

H	Position und Orientierung der Roboterhand	28
T	Eine euklidische Transformation	28
$I_{3 \times 3}$	Die 3×3 Einheitsmatrix	29
e_v	Eigenvektor	29
λ	Eigenwert	29
D	Denavit Hartenberg Matrix	31
N_g	Anzahl der Gelenke	32
θ	Drehwinkel zur Ausrichtung der z-Achsen in der Denavit-Hartenberg-Notation	32
d	Verschiebung entlang der z-Achse in der Denavit-Hartenberg-Notation	32
l	Verschiebung entlang der x-Achse in der Denavit-Hartenberg-Notation	33
α	Drehwinkel zur Ausrichtung der x-Achsen in der Denavit-Hartenberg-Notation	33
T	3-D-Translation in homogenen Koordinaten	33
p	3-D-Punkt	33
S	Strukturmatrix bei der Punktverfolgung	36
∇f	Gradient der Pixelwerte in einem Bild	36
f	Grauwert eines Bildes	37
τ	Zeit	37
ϵ	Fehler zwischen zwei Merkmalsfenstern	37
d	Eine Verschiebung im Bild	37
G	3-D-Gerade	38
a	Stützpunkt einer 3-D-Geraden	38
ϵ_{rp}	Rückprojektionsfehler	39
$\bar{\epsilon}_{rp}$	Durchschnittlicher Rückprojektionsfehler	40
$\mathbb{P}^3$	dreidimensionaler projektiver Raum	40
U	Bewegung der Kamera (Translation und Rotation)	44
v	Translationsgeschwindigkeit	44
ω	Winkelgeschwindigkeit	44
O_w	Ein 2-D- oder 3-D-Objekt	44
O_w	Parameter eines 3-D-Objekts	44
h	Funktion für die Beschreibung eines Objekts	44
h_{exp}	Explizite Formulierung der Parametrisierung eines Objekts	44
O_b	Projektion eines 2-D- oder 3-D-Objekts ins Bild	45

h_{bild}	Funktion für die Beschreibung eines Objekts im Bild	45
O_b	Parameter der Projektion eines 3-D-Objekts im Bild	45
ψ	Eine allgemeine Funktion	45
k_{n1}	Maximum der zweiten Fundamentalform einer Fläche im $\mathbb{R}^3$	46
k_{n2}	Minimum der zweiten Fundamentalform einer Fläche im $\mathbb{R}^3$	46
ε	Maximaler Fehler, der bei der Lösung des Korrespondenzproblems auftreten kann	48
e	3-D-Fehler der Rekonstruktion	48
γ	Rotationswinkel	49
s	Stereobasis	50
η	Diverse Hilfsvariable	52
b_x	Breite des Bildes in Pixeln	58
b_y	Höhe des Bildes in Pixeln	58
S	Anzahl der Samples einer Monte-Carlo-Simulation	59
ϕ^+	Funktion für die Triangulation	65
$\mathbf{X}$	Zufallsvektor	66
$\mathbf{Y}$	Zufallsvektor	66
$\mathbb{P}^1$	eindimensionale projektive Gerade	66
$\mathbb{P}^2$	zweidimensionale projektive Ebene	66
p	Wahrscheinlichkeit	69
z	Zustand eines Systems	72
o	Beobachtung	72
$\hat{z}$	Schätzer des Zustands eines Systems	72
$\langle \hat{z} \rangle$	Folge von Schätzwerten des Zustands eines Systems	72
$\langle o \rangle$	Folge von Beobachtungen	72
b	Beobachtungsfunktion	72
v	Beobachtungsrauschen	72
g	Zustandsübergangsfunktion	74
w	Zustandsübergangsrauschen	74
G	Zustandsübergangsmatrix	75
N	Dimension des Zustands beim Zustandsschätzer	75
B	Beobachtungsmatrix	75
M	Dimension der Beobachtung beim Zustandsschätzer	75
Q	Kovarianz des Zustandsübergangsrauschens	75
$\mathfrak{R}$	Kovarianz des Beobachtungsrauschens	75

$\mathfrak{K}$	Kalman Verstärkermatrix	76
W	Jacobi-Matrix der Zustandsübergangsfunktion abgeleitet nach den Rauschtermen	77
B	Jacobi-Matrix der Beobachtungsfunktion abgeleitet nach der Beobachtung	77
V	Jacobi-Matrix der Beobachtungsfunktion abgeleitet nach den Rauschtermen	77
G	Jacobi-Matrix der Zustandsübergangsfunktion abgeleitet nach den Zustand	77
r	Dichtefunktion einer Zufallsvariablen	78
y	Realisierung der Zufallsvariable Y	78
Y	Zufallsvariable	78
X	Zufallsvariable	78
Z	Zufallsvariable	78
x	Realisierung der Zufallsvariable X	78
z	Realisierung der Zufallsvariable Z	78
c	Kameraparameter, die direkt beeinflussbar sind	85
H	Entropie	85
c^*	Optimale Kameraparameter (Brennweite, Rotationsparameter, Translationsvektor)	86
Γ	Gütemaß einer Kovarianzmatrix	90
g	Spezielles Gütemaß, entweder D-, E- oder T-Optimalität	90
O	Menge der sichtbaren Beobachtungen	92
$\mathfrak{A}$	Eine affine Transformation	96
n	Normalenvektor einer Ebene	108
e_R	3-D-Fehlermaß der Rekonstruktion für Objekte bestehend aus einer planaren Fläche mithilfe der Regressionsebene	109
e_P	3-D-Fehlermaß der Rekonstruktion für plane Objekte	109
e_{np}	3-D-Fehlermaß der Rekonstruktion für nichtplanare Objekte	109

Anhang B

Herleitung der Fehlerfunktion für das 2-D-Modell aus Abschnitt 3.2

Die Fehlerfunktion e aus den Gleichungen (3.10), (3.19) und (3.21) wird hergeleitet. Die Herleitung erfolgt zunächst in allgemeiner Form und wird dann durch Einsetzen spezieller Bedingungen in die entsprechende Form aus dem Text vorne gebracht.

Alle Bezeichnungen ergeben sich gemäß Abbildung 3.3.

Der Punkt in Weltkoordinaten $\underline{p}_w$ hat in homogenen Koordinaten folgende Komponenten

$$\underline{p}_w = \begin{pmatrix} p_{wx} \\ p_{wz} \\ 1 \end{pmatrix}. \quad (\text{B.1})$$

Für die Abbildung werden die Kameramatrizen

$$\mathbf{K}_l = \begin{pmatrix} f_l & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{K}_r = \begin{pmatrix} f_r & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{B.2})$$

die Rotationsmatrizen

$$\mathbf{R}_l = \begin{pmatrix} \cos \gamma_l & \sin \gamma_l \\ -\sin \gamma_l & \cos \gamma_l \end{pmatrix} \quad \mathbf{R}_r = \begin{pmatrix} \cos \gamma_r & \sin \gamma_r \\ -\sin \gamma_r & \cos \gamma_r \end{pmatrix}, \quad (\text{B.3})$$

und die Translationvektoren

$$\underline{t}_l = (t_{lx} \ t_{lz})^T \quad \underline{t}_r = (t_{rx} \ t_{rz})^T \quad (\text{B.4})$$

definiert.

Der Bildpunkt $\underline{q}_l$ des linken Bilds ergibt sich dann gemäß Gleichung (2.6):

$$\underline{q}_l \sim \mathbf{P} \underline{p}_w \sim \mathbf{K} [\mathbf{R}^T, -\mathbf{R}^T \underline{t}] \underline{p}_w = \begin{pmatrix} f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz} \\ \sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz} \end{pmatrix}$$

Umrechnung in kartesische Koordinaten ergibt:

$$\underline{q}_l = \frac{f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz}}{\sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz}}$$

Analog gilt für den Bildpunkt $\underline{q}_r$ im rechten Bild:

$$\underline{q}_r = \frac{f_r \cos(\gamma_r) p_{wx} - f_r \sin(\gamma_r) p_{wz} - f_r \cos(\gamma_r) t_{rx} + f_r \sin(\gamma_r) t_{rz}}{\sin(\gamma_r) p_{wx} + \cos(\gamma_r) p_{wz} - \sin(\gamma_r) t_{rx} - \cos(\gamma_r) t_{rz}}$$

Für die Triangulation muss der Schnittpunkt der beiden Sichtstrahlen G_l und G_r der linken bzw. rechten Kamera berechnet werden. Der Sichtstrahl G wird aufgrund der Lage des Weltkoordinatensystems in der Form $x = G(z)$ aufgestellt.

Für den Sichtstrahl der rechten Kamera gilt: dieser Bildpunkt ist nicht verrauscht; es kann für die Berechnung des Sichtstrahls auch der Weltpunkt p_w verwendet werden. Es ergibt sich:

$$x = G_r(z) = \frac{p_{wx} - t_{rx}}{p_{wz} - t_{rz}}(z - t_{rz}) + t_{rx}. \quad (B.5)$$

Aufwändiger wird es für die Berechnung der Sichtstrahlen der linken Kamera. Zur Erinnerung, es gibt zwei Sichtstrahlen: einmal ist die Bildkoordinaten um $+\varepsilon$ und einmal um $-\varepsilon$ verrauscht. Für die Herleitung der Fehlerfunktion wird sich auf den ersteren Fall beschränkt, die entsprechende Gerade wird mit G_{l1} bezeichnet. Die Ausführungen für die Gerade G_{l2} sind völlig analog durchzuführen.

Die Gerade G_{l1} ist durch das optische Zentrum der linken Kamera O_{Zl} und dem verrauschten Bildpunkt festgelegt. Deshalb sind die Koordinaten des verrauschten Bildpunkts in Weltkoordinaten zu bestimmen. Zunächst gilt in Kamerakoordinaten:

$$p_{cl1} = \begin{pmatrix} q_l + \varepsilon \\ f_l \end{pmatrix}. \quad (B.6)$$

Man beachte, die z -Komponente des Punkts entspricht der Brennweite der Kamera. Für die Umrechnung in Weltkoordinaten kann nun Definition 2.2 benutzt werden und es gilt:

$$p_{wl+\varepsilon} = R_l \begin{pmatrix} q_l + \varepsilon \\ f_l \end{pmatrix} + t_l = \begin{pmatrix} \cos(\gamma_l) \left(\frac{f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz}}{\sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz}} + \varepsilon \right) + f_l \sin(\gamma_l) + t_{lx} \\ -\sin(\gamma_l) \left(\frac{f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz}}{\sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz}} + \varepsilon \right) + f_l \cos(\gamma_l) + t_{lz} \end{pmatrix}$$

Über die Formel

$$\frac{\Delta x}{\Delta z} \quad (B.7)$$

lässt sich die Steigung nun berechnen und in der Punkt-Steigungs-Form erhält man die Gerade G_{l1}

$$x = G_{l1}(z) = \frac{\cos(\gamma_l) \left(\frac{f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz}}{\sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz}} + \varepsilon \right) + f_l \sin(\gamma_l)}{-\sin(\gamma_l) \left(\frac{f_l \cos(\gamma_l) p_{wx} - f_l \sin(\gamma_l) p_{wz} - f_l \cos(\gamma_l) t_{lx} + f_l \sin(\gamma_l) t_{lz}}{\sin(\gamma_l) p_{wx} + \cos(\gamma_l) p_{wz} - \sin(\gamma_l) t_{lx} - \cos(\gamma_l) t_{lz}} + \varepsilon \right) + f_l \cos(\gamma_l)} (z - t_{lz}) + t_{lx}. \quad (B.8)$$

Für die Berechnung der Triangulation müssen nun die beiden Sichtstrahlen geschnitten werden. Man erhält den Schnittpunkt durch Gleichsetzen der Gleichungen (B.8) und (B.5). Anschließend löst man nach z auf. Für die x Komponente muss zuletzt noch z in eine der beiden Gleichungen eingesetzt werden. Man erhält so die beiden Koordinaten. Aufgrund der Größe der Gleichungen werden für die

Darstellung Hilfsvariablen η eingeführt, die fortlaufend nummeriert werden. Es wird definiert:

$$\begin{aligned}
\eta_1 &:= ((t_{lx} - p_{wx}) \sin \gamma_1 + (t_{lz} - p_{wz}) \cos \gamma_1) \\
&\quad ((t_{rx} - t_{lx}) p_{wz} \sin \gamma_1 + (t_{lx} - p_{wx}) t_{rz} \sin \gamma_1 + (t_{rz} - p_{wz}) t_{lx} \cos \gamma_1) \\
\eta_2 &:= p_{wz} f_1 (-t_{lx} p_{wz} + t_{rx} p_{wz} - t_{lz} t_{rx} - p_{wx} t_{rz} + t_{lx} t_{rz} + p_{wx} t_{lz}) \\
\eta_3 &:= ((t_{lx} - p_{wx}) \sin \gamma_1 + (t_{lz} - p_{wz}) \cos \gamma_1) \\
&\quad ((t_{rx} - p_{wx}) \sin \gamma_1 + (t_{rz} - p_{wz}) \cos \gamma_1) \\
\eta_4 &:= f_1 (-t_{lx} p_{wz} + t_{rx} p_{wz} - t_{lz} t_{rx} - p_{wx} t_{rz} + t_{lx} t_{rz} + p_{wx} t_{lz})
\end{aligned}$$

Die z-Komponente des Punkts $\hat{p}_{w1}$ lautet folgendermaßen:

$$\hat{p}_{w1z} = \frac{\eta_1 \varepsilon + \eta_2}{\eta_3 \varepsilon + \eta_4} \quad (\text{B.9})$$

Zur Beschreibung der x-Komponente wird weiter definiert:

$$\begin{aligned}
\eta_5 &:= ((p_{wz} - t_{lz}) \cos \gamma_1 + (p_{wx} - t_{lx}) \sin \gamma_1) \\
&\quad ((p_{wz} - t_{lz}) t_{rx} \cos \gamma_1 + (t_{lz} - t_{rz}) p_{wx} \cos \gamma_1 + (p_{wx} - t_{rx}) t_{lx} \sin \gamma_1) \\
\eta_6 &:= -f_1 p_{wx} (-t_{lx} p_{wx} + t_{rz} p_{wx} + p_{wz} t_{lx} - p_{wz} t_{rx} - t_{lx} t_{rz} + t_{lz} t_{rx}) \\
\eta_7 &:= ((p_{wz} - t_{rz}) \cos \gamma_1 + (p_{wx} - t_{rx}) \sin \gamma_1) \\
&\quad ((p_{wz} - t_{lz}) \cos \gamma_1 + (p_{wx} - t_{lx}) \sin \gamma_1) \\
\eta_8 &:= -f_1 (-t_{lx} p_{wx} + t_{rz} p_{wx} + p_{wz} t_{lx} - p_{wz} t_{rx} - t_{lx} t_{rz} + t_{lz} t_{rx})
\end{aligned}$$

Die x-Komponente des Punkts $\hat{p}_{w1}$ lautet folgendermaßen:

$$\hat{p}_{w1x} = \frac{(\eta_5 \varepsilon + \eta_6)}{(\eta_7 \varepsilon + \eta_8)} \quad (\text{B.10})$$

Für den zweiten Punkt $\hat{p}_{w2}$ aus Abbildung 3.3 gelten die entsprechenden Koordinaten aus (B.9) bzw. (B.10), nur dass ε durch $-\varepsilon$ ersetzt werden muss.

Nun müssen die Koordinaten in Gleichung (3.9) eingesetzt werden. Mit Hilfe des Computer-Algebra-Programms MAPLE lässt sich die Fehlerfunktion e in folgende Form bringen:

$$e = \frac{\eta_9}{\eta_{10}}, \quad (\text{B.11})$$

dabei werden folgende Hilfsvariablen benutzt:

$$\begin{aligned}
\eta_9 &:= 4f_1^2 \varepsilon^2 (p_{wz}^2 - 2p_{wz} t_{rz} + t_{rz}^2 + p_{wx}^2 - 2p_{wx} t_{rx} + t_{rx}^2) \\
&\quad (-p_{wx} \sin \gamma_1 - p_{wz} \cos \gamma_1 + t_{lx} \sin \gamma_1 + t_{lz} \cos \gamma_1)^4 \\
&\quad (t_{lx} p_{wz} - t_{lx} t_{rz} - p_{wx} t_{lz} + t_{rx} t_{lz} - t_{rx} p_{wz} + p_{wx} t_{rz})^2
\end{aligned}$$

$$\begin{aligned}
\eta_{10} := & \left(-\varepsilon (\cos \gamma_1)^2 p_{wz} t_{rz} + f_1 (\cos \gamma_1)^2 t_{lx} p_{wz} - f_1 (\cos \gamma_1)^2 t_{lx} t_{rz} \right. \\
& - \varepsilon (\cos \gamma_1)^2 t_{lz} p_{wz} - \cos \gamma_1 \varepsilon \sin \gamma_1 p_{wx} t_{rz} - t_{rx} \sin \gamma_1 \varepsilon \cos \gamma_1 p_{wz} \\
& + t_{rx} \sin \gamma_1 \varepsilon \cos \gamma_1 t_{lz} - p_{wx} \sin \gamma_1 \varepsilon \cos \gamma_1 t_{lz} + 2 \cos \gamma_1 \varepsilon \sin \gamma_1 p_{wx} p_{wz} \\
& + \cos \gamma_1 \varepsilon \sin \gamma_1 t_{lx} t_{rz} + \varepsilon (\cos \gamma_1)^2 t_{lz} t_{rz} - p_{wx} f_1 (\sin \gamma_1)^2 t_{lz} \\
& - t_{rx} \varepsilon (\sin \gamma_1)^2 p_{wx} + t_{rx} \varepsilon (\sin \gamma_1)^2 t_{lx} - p_{wx} f_1 (\cos \gamma_1)^2 t_{lz} \\
& + \varepsilon (\cos \gamma_1)^2 p_{wz}^2 + p_{wx}^2 \varepsilon (\sin \gamma_1)^2 - \cos \gamma_1 \varepsilon \sin \gamma_1 t_{lx} p_{wz} \\
& + t_{rx} f_1 (\cos \gamma_1)^2 t_{lz} - t_{rx} f_1 (\sin \gamma_1)^2 p_{wz} + t_{rx} f_1 (\sin \gamma_1)^2 t_{lz} \\
& - t_{rx} f_1 (\cos \gamma_1)^2 p_{wz} + f_1 (\sin \gamma_1)^2 p_{wx} t_{rz} + f_1 (\sin \gamma_1)^2 t_{lx} p_{wz} \\
& \left. - f_1 (\sin \gamma_1)^2 t_{lx} t_{rz} - p_{wx} \varepsilon (\sin \gamma_1)^2 t_{lx} + f_1 (\cos \gamma_1)^2 p_{wx} t_{rz} \right)^2 \\
& \left(\varepsilon (\cos \gamma_1)^2 p_{wz} t_{rz} + f_1 (\cos \gamma_1)^2 t_{lx} p_{wz} - f_1 (\cos \gamma_1)^2 t_{lx} t_{rz} \right. \\
& + \varepsilon (\cos \gamma_1)^2 t_{lz} p_{wz} + \cos \gamma_1 \varepsilon \sin \gamma_1 p_{wx} t_{rz} + t_{rx} \sin \gamma_1 \varepsilon \cos \gamma_1 p_{wz} \\
& - t_{rx} \sin \gamma_1 \varepsilon \cos \gamma_1 t_{lz} + p_{wx} \sin \gamma_1 \varepsilon \cos \gamma_1 t_{lz} - 2 \cos \gamma_1 \varepsilon \sin \gamma_1 p_{wx} p_{wz} \\
& - \cos \gamma_1 \varepsilon \sin \gamma_1 t_{lx} t_{rz} - \varepsilon (\cos \gamma_1)^2 t_{lz} t_{rz} - p_{wx} f_1 (\sin \gamma_1)^2 t_{lz} \\
& + t_{rx} \varepsilon (\sin \gamma_1)^2 p_{wx} - t_{rx} \varepsilon (\sin \gamma_1)^2 t_{lx} - p_{wx} f_1 (\cos \gamma_1)^2 t_{lz} \\
& - \varepsilon (\cos \gamma_1)^2 p_{wz}^2 - p_{wx}^2 \varepsilon (\sin \gamma_1)^2 + \cos \gamma_1 \varepsilon \sin \gamma_1 t_{lx} p_{wz} \\
& + t_{rx} f_1 (\cos \gamma_1)^2 t_{lz} - t_{rx} f_1 (\sin \gamma_1)^2 p_{wz} + t_{rx} f_1 (\sin \gamma_1)^2 t_{lz} \\
& - t_{rx} f_1 (\cos \gamma_1)^2 p_{wz} + f_1 (\sin \gamma_1)^2 p_{wx} t_{rz} + f_1 (\sin \gamma_1)^2 t_{lx} p_{wz} \\
& \left. - f_1 (\sin \gamma_1)^2 t_{lx} t_{rz} + p_{wx} \varepsilon (\sin \gamma_1)^2 t_{lx} + f_1 (\cos \gamma_1)^2 p_{wx} t_{rz} \right)^2
\end{aligned}$$

Durch weitere, per Hand durchgeführte Vereinfachungsschritte gelangt man schließlich auf folgende Form der Hilfsvariablen:

$$\begin{aligned}
\eta_9 := & 4f_1^2 \varepsilon^2 \left((p_{wz} - t_{rz})^2 + (p_{wx} - t_{rx})^2 \right) \\
& ((t_{lx} - p_{wx}) \sin \gamma_1 + (t_{lz} - p_{wz}) \cos \gamma_1)^4 \\
& (t_{lx} p_{wz} - t_{lx} t_{rz} - p_{wx} t_{lz} + t_{rx} t_{lz} - t_{rx} p_{wz} + p_{wx} t_{rz})^2
\end{aligned}$$

Ferner lässt sich der Nenner in die Form

$$\eta_{10} = (\eta_{11}^2 - \eta_{12}^2)^2 \quad (\text{B.12})$$

bringen. Dabei gilt:

$$\eta_{11} := f_1 (t_{lx} p_{wz} - t_{lx} t_{rz} - p_{wx} t_{lz} + t_{rx} t_{lz} - t_{rx} p_{wz} + p_{wx} t_{rz})$$

$$\begin{aligned}
\eta_{12} := & \left(-\varepsilon p_{wz} t_{rz} \cos^2 \gamma_1 \right. \\
& -\varepsilon t_{lz} p_{wz} \cos^2 \gamma_1 - \varepsilon p_{wx} t_{rz} \cos \gamma_1 \sin \gamma_1 - \varepsilon t_{rx} p_{wz} \cos \gamma_1 \sin \gamma_1 \\
& +\varepsilon t_{rx} t_{lz} \cos \gamma_1 \sin \gamma_1 - \varepsilon p_{wx} t_{lz} \cos \gamma_1 \sin \gamma_1 + 2\varepsilon p_{wx} p_{wz} \cos \gamma_1 \sin \gamma_1 \\
& +\varepsilon t_{lx} t_{rz} \cos \gamma_1 \sin \gamma_1 + \varepsilon t_{lz} t_{rz} \cos^2 \gamma_1 \\
& -\varepsilon t_{rx} p_{wx} \sin^2 \gamma_1 + \varepsilon t_{rx} t_{lx} \sin^2 \gamma_1 \\
& +\varepsilon p_{wz}^2 \cos^2 \gamma_1 + \varepsilon p_{wx}^2 \sin^2 \gamma_1 - \varepsilon t_{lx} p_{wz} \cos \gamma_1 \sin \gamma_1 \\
& \left. -\varepsilon p_{wx} t_{lx} \sin^2 \gamma_1 \right)
\end{aligned}$$

Durch die vereinfachten Annahmen in Abschnitt 3.2.1 entstehen die jeweiligen Fehlerfunktionen (3.10), (3.19) bzw. (3.21) für die Untersuchung der einzelnen Parameter.

Anhang C

Probabilistischer Ansatz für die Optimierung der Rotation im 2-D

Diese Überlegungen geben ein weiteres Indiz dafür, dass das Wegdrehen der Kamera, im Stereomodell aus Abschnitt 3.2, zu einer Reduktion des Fehlers führt. Die Analyse erfolgt analog zu den Untersuchungen im zweidimensionalen Fall aus Abschnitt 3.2.1.5. Zunächst werden nun die erforderlichen Größen definiert.

Sei der 3-D-Punkt $\mathbf{p}_w = (0 \ 0 \ 15)$, die Brennweite $f_1 = 1$ und die Translation der linken Kamera $\mathbf{t}_l = (5 \ 0 \ 0)^T$. Die rechte Kamera ist hier ohne Belang, wenn man davon ausgeht, dass die bisherige Schätzung durch eine (weitere) Beobachtung der linken Kamera aufgefrischt werden soll.

Im Gegensatz zu Abschnitt 3.2.1.5 braucht man statt des maximal auftretenden Fehlers ε die Kovarianzmatrix des Rauschens $\mathfrak{R}$. Im 2-D-Modell ist die Beobachtung nur Skalar, deswegen ist auch die Kovarianzmatrix des Beobachtungsrauschens nur eine Varianz. Sei

$$\mathfrak{R} = 0,25. \quad (\text{C.1})$$

Die Koordinaten des homogenen (eindimensionalen) Bildpunkts sind, wie in Abschnitt 2.1 eingeführt, gegeben durch:

$$\underline{\mathbf{q}} = \begin{pmatrix} \cos \gamma & \sin \gamma \\ -\sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} p_{wx} - t_x \\ p_{wz} - t_z \end{pmatrix}. \quad (\text{C.2})$$

Durch die Umrechnung in kartesische Koordinaten ergibt sich die Beobachtungsfunktion $\mathbf{b}$:

$$\mathbf{b} = \frac{(p_{wx} - t_x) \cos \gamma + (p_{wz} - t_z) \sin \gamma}{-(p_{wx} - t_x) \sin \gamma + (p_{wz} - t_z) \cos \gamma}. \quad (\text{C.3})$$

Und damit auch die Jacobi-Matrix der Beobachtungsfunktion $\mathbf{B}$:

$$\mathbf{B} = \begin{pmatrix} \frac{p_{wz} - t_z}{(-(p_{wx} - t_x) \sin \gamma + (p_{wz} - t_z) \cos \gamma)^2} & \frac{p_{wx} - t_x}{(-(p_{wx} - t_x) \sin \gamma + (p_{wz} - t_z) \cos \gamma)^2} \end{pmatrix} \quad (\text{C.4})$$

Schließlich sei noch festgelegt, dass für die a priori Kovarianzmatrix des Zustands Σ_τ gilt:

$$\Sigma_\tau = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (\text{C.5})$$

Dann lässt sich mit Hilfe von Gleichung (4.27) die Kalman-Verstärkungsmatrix $\mathfrak{K}$ berechnen¹:

$$\mathfrak{K}_\tau = \Sigma_{\tau-1} \mathbf{B}_\tau^T (\mathbf{B}_\tau \Sigma_{\tau-1} \mathbf{B}_\tau^T + \mathfrak{R}_\tau)^{-1} \quad (\text{C.6})$$

¹da das Beobachtungsrauschen additiv ist, wird $\mathbf{V}$ zur Einheitsmatrix, wie auf Seite 77 erläutert, und ist deshalb hier nicht aufgeführt



Bild C.1 Der Punkt wird mit Hilfe des Kalman-Filters analysiert. Die Unsicherheitsellipse der a priori Kovarianzmatrix (in diesem Fall die Einheitsmatrix) ist ein Kreis (gestrichelt). Unabhängig davon, von wo eine weitere Beobachtung gemacht wird, die Unsicherheit im Sinne der E-Optimalität bleibt immer gleich dem Radius des Kreises, da die Unsicherheit in die Richtung der z-Achse der Kamera nicht abnehmen kann.

und damit die a posteriori Kovarianzmatrix des Zustands nach Gleichung (4.29)

$$\Sigma_{\tau} = (\mathbf{I} - \mathbf{K}_{\tau} \mathbf{B}_{\tau}) \Sigma_{\tau-1}. \quad (\text{C.7})$$

Nun wird der Rotationswinkel gesucht, bei dem die a posteriori Kovarianz im Sinne des jeweiligen Gütemaßes (D-, mod. E- bzw. T-Optimalität) am kleinsten ist. Es ergibt sich dabei als optimaler Winkel für das D- und das T-Kriterium:

$$\gamma = 71,6^{\circ} \quad (\text{C.8})$$

Anmerkung zum E-Kriterium Für das E-Kriterium ergibt sich kein eindeutiges Minimum, da die a priori Kovarianzmatrix $\Sigma_{\tau-1}$ einem Kreis mit Radius 1 entspricht. Durch eine Beobachtung wird der Kreis zu einer Ellipse; die Unsicherheit wird orthogonal zur Blickrichtung eingeschränkt. Die größere Halbachse der Ellipse hat aber immer noch die Länge Eins, völlig unabhängig davon, von wo die Beobachtung gemacht wurde. Der Sachverhalt ist in Bild C.1 nochmal graphisch dargestellt. Dies gilt natürlich insbesondere auch dann, wenn die Kamera nur um ihr optisches Zentrum rotiert werden kann, was in Abschnitt 3.2.1.5 angenommen wurde.

Anhang D

L-förmiges konvexes Objekt

Hier sind die Ergebnisse der Simulation mit dem Stereoansatz eines L-förmigen konvexen Objekts abgedruckt, um sie mit den Ergebnissen mit dem L-förmigen konkaven Objekt aus Abschnitt 6.2.2 vergleichen zu können. Da sich die Ergebnisse nicht in ihrer Aussage von denen des L-förmigen konkaven Objekts unterscheiden, erfolgt hier der Abdruck weitestgehend kommentarlos.

$f_r =$	760	814	910	1035	1172	1352	1575	1851	2225
$f_l = 760$	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.06	0.06
$f_l = 814$	0.08	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.06
$f_l = 910$	0.08	0.07	0.07	0.07	0.06	0.06	0.06	0.05	0.05
$f_l = 1035$	0.07	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.05
$f_l = 1172$	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.04
$f_l = 1352$	0.07	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.04
$f_l = 1575$	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.04
$f_l = 1851$	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.03
$f_l = 2225$	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.03

Tabelle D.1 Der 3-D-Fehler e der Rekonstruktion des L-förmigen konvexen Objekts in Abhängigkeit der Brennweiten f_l der linken bzw. f_r in Pixel der rechten Kamera.

$f_r =$	760	814	910	1035	1172	1352	1575	1851	2225
$f_l = 760$	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05
$f_l = 814$	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.04	0.04
$f_l = 910$	0.06	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04
$f_l = 1035$	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04
$f_l = 1172$	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03
$f_l = 1352$	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03
$f_l = 1575$	0.05	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03
$f_l = 1851$	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
$f_l = 2225$	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02

Tabelle D.2 Das Fehlermaß e_p der Rekonstruktion des L-förmigen konvexen Objekts in Abhängigkeit der Brennweiten f_l der linken bzw. f_r in Pixel der rechten Kamera.

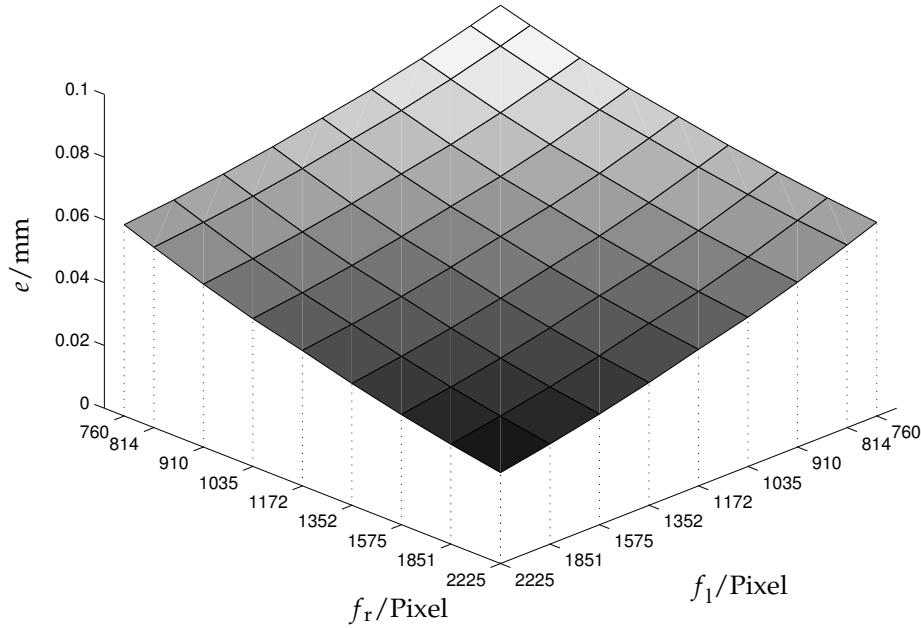


Bild D.1 3-D-Fehler e in mm der Rekonstruktion des Objekts in Abhängigkeit der Brennweiten f_l der linken bzw. f_r der rechten Kamera. Man sieht große Brennweiten ergeben kleine Fehler.

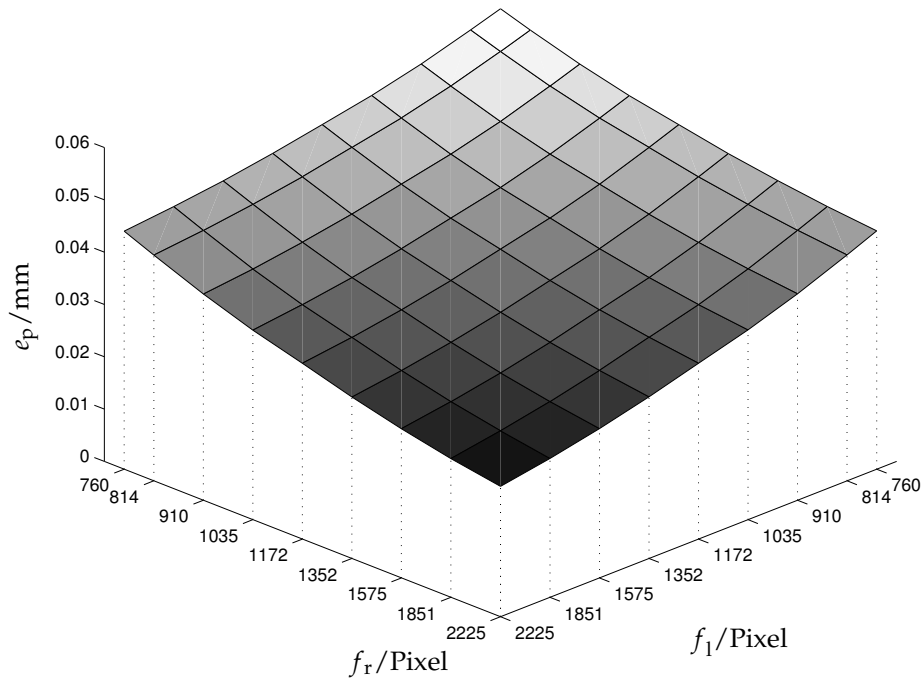


Bild D.2 Das Fehlermaß e_p für verschiedenen Brennweiten f_l der linken bzw. f_r der rechten Kamera des L-förmigen konvexen Objekts.

Die Optimierung aller Parameter über dem Raum der vom Scorbobot ER VII erreichbaren Positionen liefert als bestes Ergebnis eine Rekonstruktion mit dem Fehlermaß $e = 0.00624$ und dem Fehlermaß $e_p = 0.01429$ gem. 6.1.3.

	z=85	z=80	z= 75	z= 70	z=65	z= 60	z= 55	z= 50	z=45	z=40
x=-20	1	22	43	64	85	106	127	148	169	190
x=-18	2	23	44	65	86	107	128	149	170	191
x=-16	3	24	45	66	87	108	129	150	171	192
x=-14	4	25	46	67	88	109	130	151	172	193
x=-12	5	26	47	68	89	110	131	152	173	194
x=-10	6	27	48	69	90	111	132	153	174	195
x=-8	7	28	49	70	91	112	133	154	175	196
x=-6	8	29	50	71	92	113	134	155	176	197
x=-4	9	30	51	72	93	114	135	156	177	198
x=-2	10	31	52	73	94	115	136	157	178	199
x=0	11	32	53	74	95	116	137	158	179	200
x=2	12	33	54	75	96	117	138	159	180	201
x=4	13	34	55	76	97	118	139	160	181	202
x=6	14	35	56	77	98	119	140	161	182	203
x=8	15	36	57	78	99	120	141	162	183	204
x=10	16	37	58	79	100	121	142	163	184	205
x=12	17	38	59	80	101	122	143	164	185	206
x=14	18	39	60	81	102	123	144	165	186	207
x=16	19	40	61	82	103	124	145	166	187	208
x=18	20	41	62	83	104	125	146	167	188	209
x=20	21	42	63	84	105	126	147	168	189	210

Tabelle D.3 Zuordnung der x - z -Position zur Bildnummer. Die Positionsangaben der x - bzw. z - Komponente sind in cm gemessen. Im Gegensatz zum Kalibrierexperiment kann hier das Objekt aus jeder Position gesehen werden, so dass dort keine Positionen ausgelassen werden müssen.

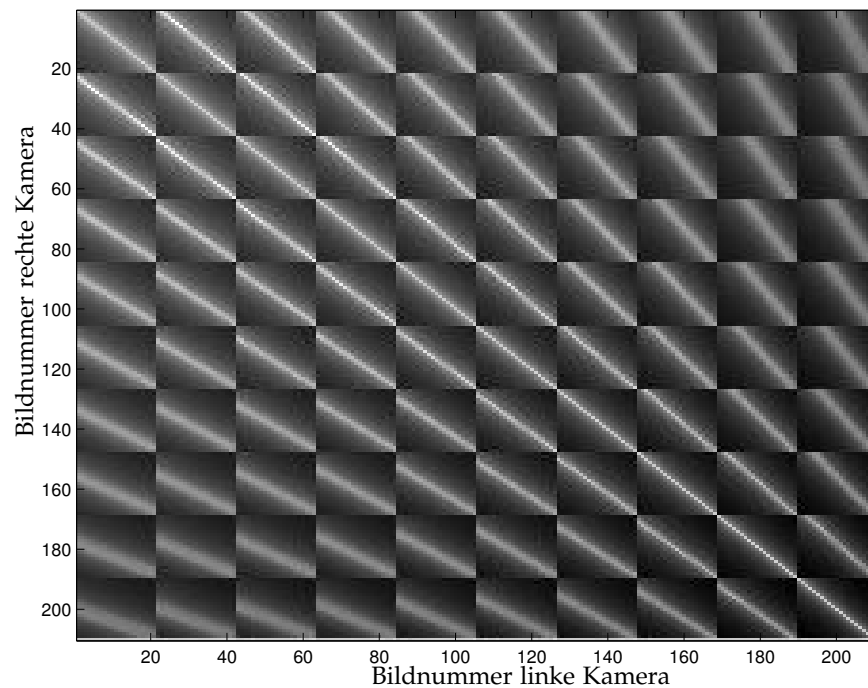


Bild D.3 Rekonstruktionsfehler e des Objekts in Abhängigkeit der Translation, wobei ein heller Grauwert einen großen Fehler bedeutet.

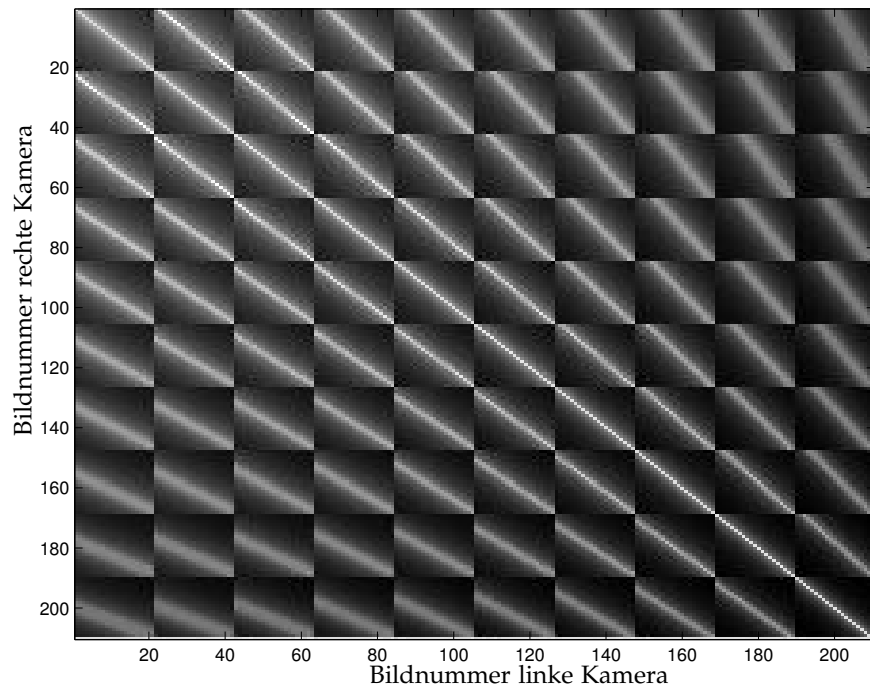
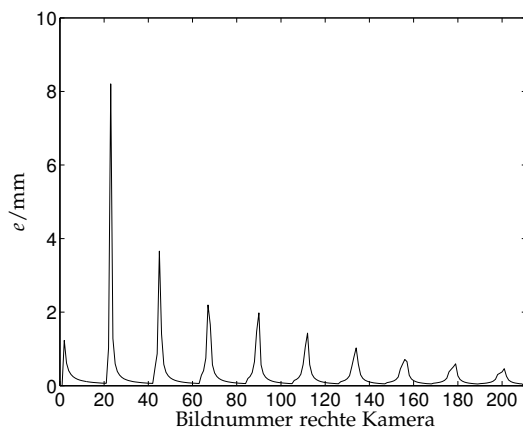
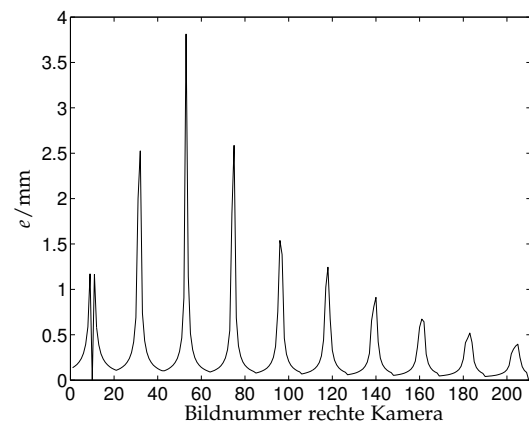


Bild D.4 Rekonstruktionsfehler e_p des Objekts in Abhängigkeit der Translation, wobei ein heller Grauwert einen großen Fehler bedeutet.



(a)



(b)

Bild D.5 Rekonstruktionsfehler e in mm in Abhängigkeit der Bildnummer der rechten Kamera für die linke Kamera mit Bildnummer (a) 1 bzw. (b) 10. An der Stelle, wo die Bildnummer der linken und rechten Kamera identisch ist, wurde der Fehlerwert auf Null gesetzt.

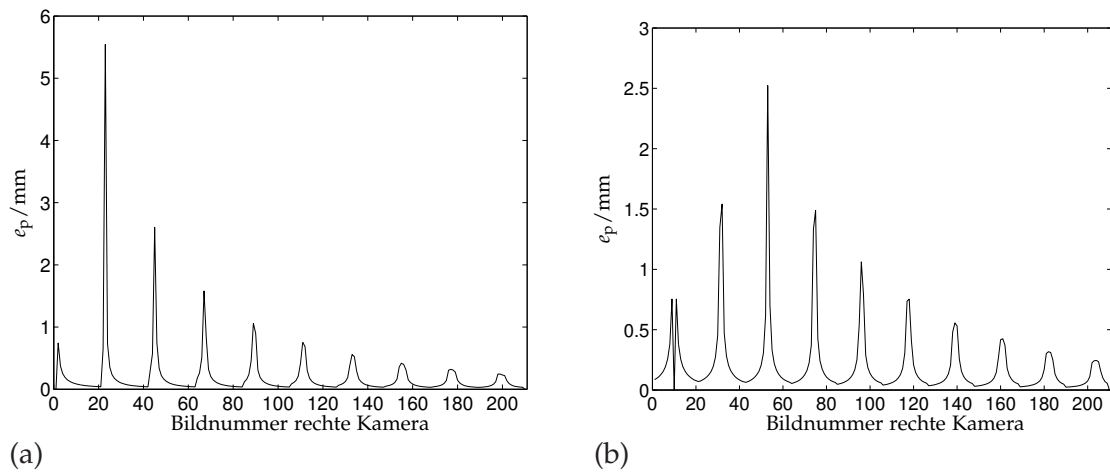


Bild D.6 Rekonstruktionsfehler e_p gem. Abschnitt 6.1.3 in mm in Abhängigkeit der Bildnummer der rechten Kamera für die linke Kamera mit Bildnummer (a) 1 bzw. (b) 10. An der Stelle, wo die Bildnummer der linken und rechten Kamera identisch ist, wurde der Fehlerwert auf Null gesetzt.

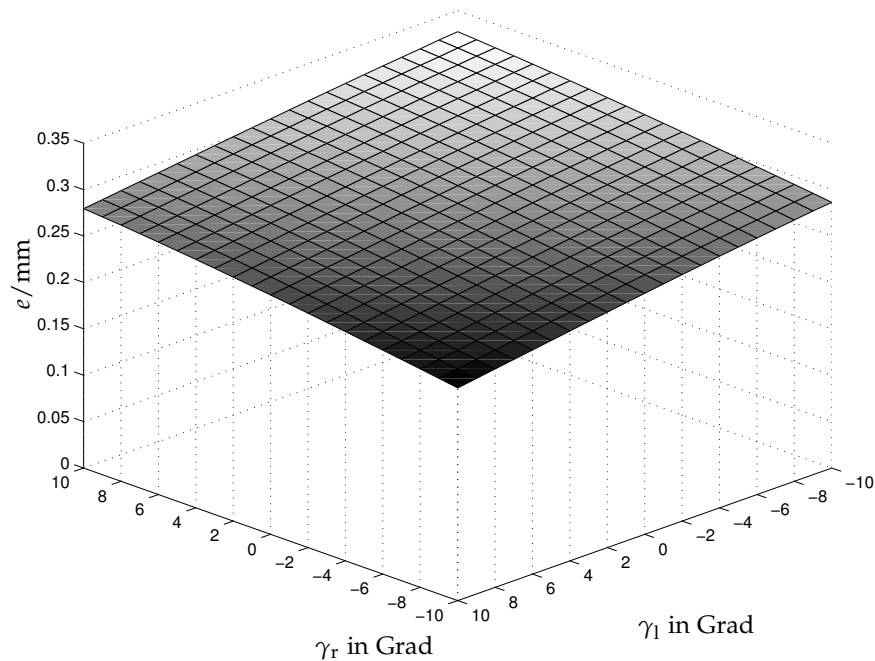


Bild D.7 Darstellung des Rekonstruktionsfehlers e in mm bei der Rekonstruktion des Objekts in Abhängigkeit des Rotationswinkels von linker und rechter Kamera. Die linke Kamera ist bei $+10^\circ$ vom Objekt weggedreht, die rechte bei -10° .

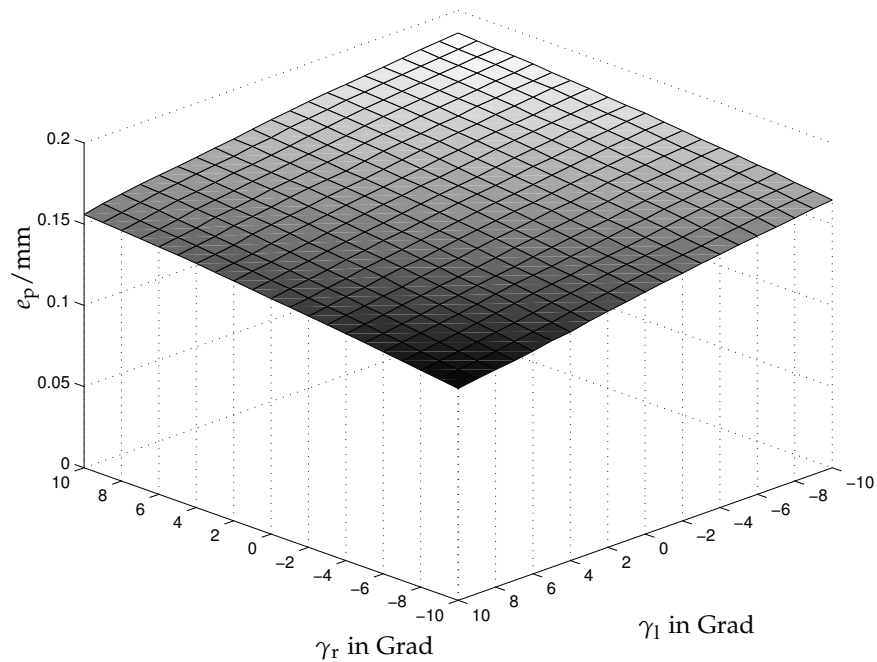
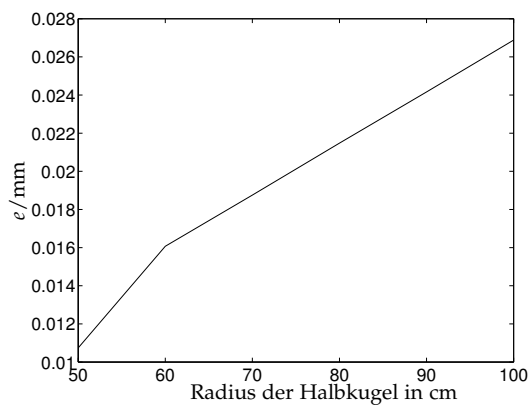
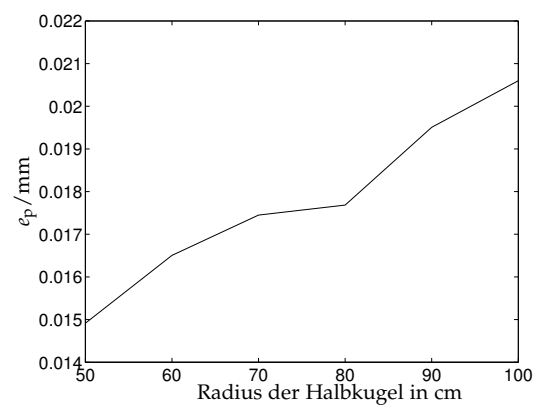


Bild D.8 Darstellung des Rekonstruktionsfehlers e_p in mm bei der Rekonstruktion des Objekts in Abhängigkeit des Rotationswinkels von linker und rechter Kamera. Die linke Kamera ist bei $+10^\circ$ vom Objekt weggedreht, die rechte bei -10° .



(a)



(b)

Bild D.9 Fehler der Stereorekonstruktion des L-förmigen konvexen Objekts in Abhängigkeit des Radius bei Kameraanordnung auf einer Halbkugel mit dem direkt berechneten 3-D-Fehler e in Bild (a) bzw. e_p in Bild (b)

$\gamma_r =$	-10°	-9°	-8°	-7°	-6°	-5°	-4°	-3°	-2°	-1°	0°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
$\gamma_l = -10^\circ$	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.33	0.33
$\gamma_l = -9^\circ$	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.33	0.33
$\gamma_l = -8^\circ$	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32
$\gamma_l = -7^\circ$	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.32
$\gamma_l = -6^\circ$	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.32	0.32
$\gamma_l = -5^\circ$	0.27	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.32	0.32
$\gamma_l = -4^\circ$	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31	0.31
$\gamma_l = -3^\circ$	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31	0.31
$\gamma_l = -2^\circ$	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31
$\gamma_l = -1^\circ$	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31	0.31
$\gamma_l = 0^\circ$	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.31	0.31
$\gamma_l = 1^\circ$	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30
$\gamma_l = 2^\circ$	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.30	0.30	0.30
$\gamma_l = 3^\circ$	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.30	0.30	0.30
$\gamma_l = 4^\circ$	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29	0.29	0.29
$\gamma_l = 5^\circ$	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29	0.29
$\gamma_l = 6^\circ$	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29	0.29
$\gamma_l = 7^\circ$	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.29	0.29
$\gamma_l = 8^\circ$	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28
$\gamma_l = 9^\circ$	0.23	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.27	0.28	0.28	0.28	0.28
$\gamma_l = 10^\circ$	0.23	0.24	0.24	0.24	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.27	0.28	0.28	0.28

Tabelle D.4 Der 3-D-Fehler e in Abhängigkeit der Rotation der linken und rechten Kamera.

$\gamma_r =$	-10°	-9°	-8°	-7°	-6°	-5°	-4°	-3°	-2°	-1°	0°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
$\gamma_l = -10^\circ$	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19	0.19
$\gamma_l = -9^\circ$	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19
$\gamma_l = -8^\circ$	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = -7^\circ$	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = -6^\circ$	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = -5^\circ$	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18
$\gamma_l = -4^\circ$	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.18
$\gamma_l = -3^\circ$	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.18
$\gamma_l = -2^\circ$	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18
$\gamma_l = -1^\circ$	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 0^\circ$	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 1^\circ$	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 2^\circ$	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.17	0.17
$\gamma_l = 3^\circ$	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17
$\gamma_l = 4^\circ$	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.17
$\gamma_l = 5^\circ$	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.17
$\gamma_l = 6^\circ$	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
$\gamma_l = 7^\circ$	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16
$\gamma_l = 8^\circ$	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16
$\gamma_l = 9^\circ$	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16
$\gamma_l = 10^\circ$	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.16

Tabelle D.5 Der Rekonstruktionsfehler e_p in Abhängigkeit der Rotation der linken und rechten Kamera.

Literaturverzeichnis

- [Alo90] Aloimonos, J. Y.: *Perspective Approximations*, *Image and Vision Computing*, Bd. 8, Nr. 3, 1990, S. 177–192.
- [Ana89] Anandan, P.: *A Computational Framework and an Algorithm for the Measurement of Visual Motion*, *International Journal of Computer Vision*, Bd. 2, Nr. 3, 1989, S. 283–310.
- [Arb99] Arbel, T.; Ferrie, F.: *Viewpoint selection by navigation through entropy maps*, in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, Bd. 1, 1999, S. 248–254.
- [Arb01] Arbel, T.; Ferrie, F.: *Entropy-based gaze planning*, *Image and Vision Computing*, Bd. 19, Nr. 11, 2001, S. 779–786.
- [Ban00] Banta, J. E.; Wong, L. R.; Dumont, C.; Abidi, M. A.: *A Next-Best-View System for Autonomous 3-D Object Reconstruction*, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Bd. 30, Nr. 5, 2000, S. 589–598.
- [Bay06] Bay, H.; Tuytelaars, T.; Gool, L. V.: *SURF: Speeded Up Robust Features*, in *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, Bd. 1, 2006, S. 404–417.
- [Beß96] Beß, R.; Paulus, D.; Niemann, H.: *3D Recovery Using Calibrated Active Camera*, in *International Conference on Image Processing*, Bd. 2, IEEE Computer Society Press, Lausanne, 1996, S. 855–858.
- [Beu04] Beutelspacher, A.; Rosenbaum, U.: *Projektive Geometrie*, Vieweg Studium, Wiesbaden, 2. erw. Ausg., 2004.
- [Bla47] Blaschke, W.: *Projektive Geometrie*, Wolfenbütteler Verlagsanstalt GmbH, Hannover, 1947.
- [Bli95] Blickle, T.; Thiele, L.: *A Mathematical Analysis of Tournament Selection*, in Eshelman, L. (Hrsg.): *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Francisco, CA, 1995, S. 9–16.
- [BS88] Bar-Shalom, Y.; Fortmann, T.: *Tracking and Data Association*, Academic Press, Boston, San Diego, New York, 1988.
- [BS95] Bar-Shalom, Y.; Li, X.-R.: *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, Connecticut, USA, 3. Ausg., 1995.
- [Bur82] Burt, P. J.; Yen, C.; Xu, X.: *Local Correlation Measures for Motion Analysis: A Comparative Study*, in *IEEE Conference on Pattern Recognition and Image Processing*, IEEE Computer Society Press, Las Vegas, 1982, S. 269–274.

- [Bär01] Bär, C.: *Elementare Differentialgeometrie*, Walter de Gruyter GmbH & Co. KG, Berlin, New York, 2001.
- [Cap71] Capelle, W.: *Die griechische Philosophie II*, Walter de Gruyter & Co., Berlin, 3. Ausg., 1971.
- [Cha94] Chaumette, F.; Boukir, S.; Bothemy, P.; Juvin, D.: *Optimal estimation of 3D structures using visual servoing*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Seattle, USA, 1994, S. 347–354.
- [Cha96] Chaumette, F.; Boukir, S.; Bouthemy, P.; Juvin, D.: *Structure from controlled motion*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 18, Nr. 5, 1996, S. 492–504.
- [Che91] Chen, H.: *A Screw Motion Approach to Uniqueness Analysis of Head-Eye Geometry*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Maui, Hawaii, 1991, S. 145–151.
- [Che05] Cheung, K.; Baker, S.; Kanade, T.: *Shape-From-Silhouette Across Time Part I: Theory and Algorithms*, *International Journal of Computer Vision*, Bd. 62, Nr. 3, 2005, S. 221–247.
- [Cho91] Chou, J. C. K.; Kamel, M.: *Finding the Position and Orientation of a Sensor on a Robot Manipulator Using Quaternions*, *International Journal of Robotics Research*, Bd. 10, Nr. 3, 1991, S. 240–254.
- [Chu91] Chui, C. K.; Chen, G.: *Kalman Filtering*, Springer, Heidelberg, 1991.
- [Con85] Connolly, C. I.: *The determination of next best views*, in *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE Computer Society Press, St. Louis, Missouri, USA, 1985, S. 432–435.
- [Cov91] Cover, T.; Thomas, J.: *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley and Sons, New York, 1991.
- [Cra04] Craig, J. J.: *Introduction to Robotics: Mechanics and Control*, Prentice Hall, Upper Saddle River, USA, 3. Ausg., 2004.
- [Dan99] Daniilidis, K.: *Hand-Eye Calibration Using Dual Quaternions*, *International Journal of Robotics Research*, Bd. 18, 1999, S. 286–298.
- [Dan01] Daniilidis, K.: *Using the Algebra of Dual Quaternions for Motion Alignment*, in Sommer, G. (Hrsg.): *Geometric Computing with Clifford Algebras*, Kap. 20, Springer-Verlag, 2001, S. 489–500.
- [Dei03] Deinzer, F.; Denzler, J.; Niemann, H.: *Viewpoint Selection – Planning Optimal Sequences of Views for Object Recognition*, in *Computer Analysis of Images and Patterns*, Springer, Groningen, Netherlands, 2003, S. 65–73.
- [Dei05] Deinzer, F.: *Optimale Ansichtenauswahl in der aktiven Objekterkennung*, Logos Verlag, Berlin, 2005.

- [Dei06a] Deinzer, F.; Denzler, J.; Derichs, C.; Niemann, H.: *Integrated Viewpoint Fusion and Viewpoint Selection for Optimal Object Recognition*, in Chanteler, M.; Trucco, E.; Fisher, R. (Hrsg.): *British Machine Vision Conference 2006*, Bd. 1, Malvern Worcs, UK, 2006, S. 287–296.
- [Dei06b] Deinzer, F.; Denzler, J.; Derichs, C.; Niemann, H.: *Aspects of Optimal Viewpoint Selection and Viewpoint Fusion*, in *Proceedings of the Asian Conference on Computer Vision*, Bd. 3852, Hyderabad, India, 2006, S. 902–912.
- [Den55] Denavit, J.; Hartenberg, R. S.: *A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices*, *ASME Journal for Applied Mechanics*, Bd. 22, Nr. 2, 1955, S. 215–221.
- [Den94] Denzler, J.; Beß, R.; Hornegger, J.; Niemann, H.; Paulus, D.: *Learning, Tracking and Recognition of 3D-Objects*, in *International Conference on Intelligent Robots and Systems*, München, 1994, S. 89–96.
- [Den97] Denzler, J.: *Aktives Sehen zur Echtzeitobjektverfolgung*, Infix, St. Augustin, 1997.
- [Den02] Denzler, J.; Brown, C.: *An Information Theoretic Approach to Optimal Sensor Data Selection for State Estimation*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 24, Nr. 2, 2002, S. 145–157.
- [Den03a] Denzler, J.: *Probabilistische Zustandsschätzung und Aktionsauswahl im Rechnersehen*, Logos Verlag, Berlin, 2003.
- [Den03b] Denzler, J.; Zobel, M.; Niemann, H.: *Information Theoretic Focal Length Selection for Real-Time Active 3-D Object Tracking*, in *International Conference on Computer Vision*, IEEE Computer Society Press, Nizza, Frankreich, 2003, S. 400–407.
- [Den04] Dennerlein, S.: *Modellierung der Kinematik von Objekten im 3D durch geometrische Algebren*, Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Mustererkennung, 2004.
- [Der06] Derichs, C.; Niemann, H.: *Handling Camera Movement Constraints in Reinforcement Learning Based Active Object Recognition*, in Franke, K.; Müller, K.-R.; Nickolay, B.; Schäfer, R. (Hrsg.): *Pattern Recognition - 28th DAGM Symposium*, Lecture Notes in Computer Science, Berlin, 2006, S. 637–646.
- [Der08] Derichs, C.; Deutsch, B.; Wenhardt, S.; Niemann, H.; Denzler, J.: *Information Theoretic Approaches for Next Best View Planning in Active Computer Vision*, in Chanda, B.; Murthy, C. A. (Hrsg.): *Advances in Intelligent Information Processing*, Statistical Science and Interdisciplinary Research, World Scientific, New Jersey, 2008, S. 55–83.
- [Deu05] Deutsch, B.; Niemann, H.; Denzler, J.: *Multi-Step Active Object Tracking with Entropy Based Optimal Actions Using the Sequential Kalman Filter*, in IEEE, (Hrsg.): *International Conference on Image Processing - ICIP'05*, Bd. 3, 2005, S. 105–108.

- [Deu06] Deutsch, B.; Wenhardt, S.; Niemann, H.: *Multi-Step Multi-Camera View Planning for Real-Time Visual Object Tracking*, in Franke, K.; Müller, K.; Nickolay, B.; Schäfer, R. (Hrsg.): *Pattern Recognition - 28th DAGM Symposium*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, S. 536–545.
- [Dev04] Deventer, R.: *Modeling and Control of Static and Dynamic Systems with Bayesian Networks*, Logos, Berlin, 2004.
- [Dou01] Doucet, A.; Freitas, N.; Gordon, N. (Hrsg.): *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, London, 2001.
- [Esp92] Espiau, B.; Chaumette, F.; Rives, P.: *A New Approach to Visual Servoing in Robotics*, *IEEE Transactions on Robotics and Automation*, Bd. 8, Nr. 3, 1992, S. 313–326.
- [Euk97] Euklid, : *Die Elemente*, Verlag Harri Deutsch, Thun und Frankfurt, 3. Ausg., 1997, Herausgeber: Clemens Thaer.
- [Fau93] Faugeras, O.: *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.
- [Fit98] Fitzgibbon, A. W.; Zisserman, A.: *Automatic Camera Recovery for Closed or Open Image Sequences*, in *Proceedings European Conference on Computer Vision (ECCV)*, Springer-Verlag, Freiburg, Germany, 1998, S. 311–326.
- [Fof01] Fofi, D.; Salvi, J.; Mouaddib, E.: *Uncalibrated Vision based on Structured Light*, in *Proc. of IEEE Int. Conf. on Robotics and Automation*, Bd. 4, IEEE Computer Society Press, Seoul, Korea, 2001, S. 3548–3553.
- [För91] Förstner, W.; Pertl, A.: *Photogrammetric Standard Methods and Digital Image Matching Techniques for High Precision Surface Measurements*, Elsevier Science Publications, 1991.
- [For97] Fortune, S.: *Voronoi diagrams and Delaunay triangulations*, in Goodman, J.; O'Rourke, J. (Hrsg.): *Handbook of discrete and computational geometry*, The CRC Press series on discrete mathematics and its applications, CRC Press, Boca Raton, 1997, S. 377–388.
- [Fua99] Fua, P.: *Using Model-Driven Bundle-Adjustment to Model Heads from Raw Video Sequences*, in *Proceedings of the 7th International Conference on Computer Vision (ICCV)*, Bd. 1, IEEE Computer Society Press, Corfu, 1999, S. 46–53.
- [Goh47] Gohlke, P.: *Aristoteles, Über die Seele*, Ferdinand Schöningh, Paderborn, 1947.
- [Gol89] Goldberg, D. E.: *Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley, Bonn, 1989.
- [Gol96] Golub, G. H.; van Loan, C. F.: *Matrix Computations*, The Johns Hopkins University Press, Baltimore, USA, 3. Ausg., 1996.

- [Gor96] Gortler, S.; Grzeszczuk, R.; Szeliski, R.; Cohen, M. F.: *The Lumigraph*, in *Computer Graphics (Proceedings of Siggraph 96)*, Addison-Wesley, New Orleans, Louisiana, 1996, S. 43–54.
- [Gov09] Govender, N.: *Evaluation of feature detection algorithms for structure from motion*, in *3rd Robotics and Mechatronics Symposium (ROBMECH 2009)*, Pretoria, South Afrika, 2009.
- [Gre97] Grethlein, M.: *Modellierung und Vermessung linearer Gelenkbewegungen bei Industrierobotern*, Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik Berlin, Druckhaus Berlin-Mitte GmbH, Berlin, 1997.
- [Ham44] Hamilton, W. R.: *On a New Species of Imaginary Quantities Connected with a Theory of Quaternions*, *Proceedings of the Royal Irish Academy*, Bd. 2, 1844, S. 424–434.
- [Ham47] Hamilton, W. R.: *On Quaternions*, *Proceedings of the Royal Irish Academy*, Bd. 3, 1847, S. 1–16.
- [Ham48] Hamilton, W. R.: *Researches Respecting Quaternions: First Series*, *Transactions of the Royal Irish Academy*, Bd. 21, 1848, S. 199–296.
- [Han77] Han, S. P.: *A Globally Convergent Method for Nonlinear Programming*, *Journal of Optimization Theory and Applications*, Bd. 22, Nr. 3, 1977, S. 297–309.
- [Har94] Hartley, R.: *Euclidean Reconstruction from Uncalibrated Views*, in *Applications of Invariance in Computer Vision*, Bd. 825 von *Lecture Notes in Computer Science*, Springer-Verlag, 1994, S. 237–256.
- [Har03] Hartley, R.; Zisserman, A.: *Multiple View Geometry in computer vision*, Cambridge University Press, Cambridge, U. K., 2. Ausg., 2003.
- [Hee92] Heeger, D. J.; Jepson, A. D.: *Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation*, *International Journal of Computer Vision*, Bd. 7, Nr. 2, 1992, S. 95–117.
- [Hei04] Heigl, B.: *Plenoptic Scene Modeling from Uncalibrated Image Sequences*, ibidem-Verlag Stuttgart, 2004.
- [Her96] Herrmann, J.: *dtv-Atlas zur Astronomie*, Deutscher Taschenbuch Verlag, München, 12. Ausg., 1996.
- [Heu04] Heuser, H.: *Lehrbuch der Analysis Teil 2*, B. G. Teubner, Wiesbaden, 13. Ausg., 2004.
- [Hey97] Heyden, A.; Åström, K.: *Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Puerto Rico, 1997, S. 438–443.
- [Hof05] Hoffmann, A.; Marx, B.; Vogt, W.: *Mathematik für Ingenieure 1*, Pearson Studium, München, 2005.

- [Hor87] Horn, B. K. P.: *Robot Vision*, MIT Press, Cambridge, Massachusetts, 1987.
- [Hun99] Hung, Y.; Ho, H.: *A Kalman Filter Approach to Direct Depth Estimation Incorporating Surface Structure*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 6, 1999, S. 570–575.
- [Irl05] Irle, A.: *Wahrscheinlichkeitstheorie und Statistik*, B. G. Teubner Verlag, Wiesbaden, 2005.
- [Iso92] Isobe, T.; Nagasaka, K.; Yamamoto, S.: *A new approach for kinematic control of simple manipulators*, *IEEE Transactions on Systems, Man, and Cybernetics*, Bd. 22, Nr. 5, 1992, S. 1116–1124.
- [Jac64] Jacobson, N.: *Lectures in Abstract Algebra*, Bd. III: Theory of Fields and Galois Theory, D. van Nostrand Company, Princeton, USA, 1964.
- [Jäh05] Jähne, B.: *Digitale Bildverarbeitung*, Springer-Verlag, Berlin, 6. erw. Ausg., 2005.
- [Kal60] Kalman, R.: *A new Approach to Linear Filtering and Prediction Problems*, *Journal of Basis Engineering*, Bd. 82, 1960, S. 35–44.
- [Kan96] Kanatani, K.: *Statistical Optimization for Geometric Computation: Theory and Practice*, Bd. 18 von *Machine Intelligence and Pattern Recognition*, Elsevier Science, Amsterdam, 1996.
- [Kay93] Kay, S. M.: *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall PTR, Upper Saddle River, New Jersey, USA, 1993.
- [Kin94] Kinnear, K. E.: *Advances in Genetic Programming*, The MIT Press, Cambridge, USA, 1994.
- [Kle00a] Klein, K.; Sequeira, V.: *The view-cube: An efficient method of view planning for 3d modelling from range data*, in *Proceedings of the 5th IEEE Workshop on Applications of Computer Vision*, IEEE Computer Society Press, Palm Springs, Kalifornien, USA, 2000, S. 186–191.
- [Kle00b] Klein, K.; Sequeira, V.: *View Planning for the 3D Modelling of Real World Scenes*, in *Proceedings of the International Conference on Intelligent Robots and Systems*, Bd. 2, IEEE Computer Society Press, Takamatsu, Japan, 2000, S. 943–948.
- [Kli07] Klippenstein, J.; Zang, H.: *Quantitative Evaluation of Feature Extractors for Visual SLAM*, in *Proceedings of the Fourth Canadian Conference on Computer and Robot Vision*, 2007, S. 157–164.
- [Koc97] Koch, R.: *Automatische Oberflächenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten*, Nr. 499 in *VDI Fortschritt-Berichte, Informatik/Kommunikationstechnik*, VDI Verlag GmbH, Düsseldorf, 1997.
- [Koz92] Koza, J. R.: *Genetic Programming, On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, USA, 1992.

- [Kuh06] Kuhl, A.; Wöhler, C.; Krüger, L.; d'Angelo, P.; Groß, H.-M.: *Monocular 3D Scene Reconstruction at Absolute Scales by Combination of Geometric and Real-Aperture Methods*, in *Pattern Recognition, 28th DAGM Symposium*, Berlin, 2006, S. 607–616.
- [Kut94a] Kutulakos, K. N.; Dyer, C. R.: *Global Surface Reconstruction by Purposive Control of Observer Motion*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Seattle, Washington, USA, 1994, S. 331–338.
- [Kut94b] Kutulakos, K. N.; Dyer, C. R.: *Recovering shape by purposive viewpoint adjustment*, *International Journal of Computer Vision*, Bd. 12, Nr. 2, 1994, S. 113–136.
- [Kut94c] Kutulakos, K. N.; Dyer, C. R.; Lumelsky, V. J.: *Provable Strategies for Vision-Guided Exploration in Three Dimensions*, in *Proc. 1994 IEEE Int. Conf. Robotics and Automation*, IEEE Computer Society Press, San Diego, Kalifornien, USA, 1994, S. 1365–1372.
- [Kut94d] Kutulakos, K. N.; Seales, W. B.; Dyer, C. R.: *Building Global Object Models by Purposive Viewpoint Control*, in *Proc. 2nd CAD-Based Vision Workshop*, Champion, Pennsylvania, USA, 1994, S. 169–182.
- [Kut95] Kutulakos, K. N.; Dyer, C. R.: *Global surface reconstruction by purposive control of observer motion*, *Artificial Intelligence*, Bd. 78, 1995, S. 147–177.
- [Lev96] Levoy, M.; Hanrahan, P.: *Light Field Rendering*, in *Computer Graphics (Proceedings of Siggraph 96)*, Addison-Wesley, New Orleans, Louisiana, 1996, S. 31–42.
- [Li03] Li, Y. F.; Chen, S. Y.: *Automatic Recalibration of an Active Structured Light Vision System*, *IEEE Transactions on Robotics and Automation*, Bd. 19, Nr. 2, 2003, S. 259–268.
- [Low04] Lowe, D. G.: *Distinctive Image Features from Scale-Invariant Keypoints*, *International Journal of Computer Vision*, Bd. 60, 2004, S. 91–110.
- [Lun06] Lunze, J.: *Regelungstechnik 2*, Springer-Verlag, Berlin, 4. Ausg., 2006.
- [Luo97] Luong, Q.-T.; Faugeras, O.: *Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices*, *International Journal of Computer Vision*, Bd. 22, Nr. 3, 1997, S. 261–289.
- [Luó10] Luó, X.; Feuerstein, M.; Reichl, T.; Kitasaka, T.; Mori, K.: *An Application Driven Comparison of Several Feature Extraction Algorithms in Bronchoscope Tracking During Navigated Bronchoscopy*, in *Medical Imaging and Augmented Reality - 5th International Workshop*, 2010, S. 475–484.
- [Mar82] Marr, D.: *Vision: A Computational Investigation into the Human representation and Processing of Visual Information*, W. H. Freeman and Company, San Francisco, 1982.

- [Mar94] Marchand, E.; Chaumette, F.: *Real Time Estimation of 3D Environment with an Active Vision System*, in *Int. Workshop on Intelligent Robotic Systems*, Grenoble, France, 1994, S. 311–318.
- [Mar96] Marchand, E.; Chaumette, F.: *Controlled camera motions for scene reconstruction and exploration*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, San Francisco, USA, 1996, S. 169–176.
- [Mar99a] Marchand, E.; Chaumette, F.: *Active Vision for Complete Scene Reconstruction and Exploration*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 1, 1999, S. 65–72.
- [Mar99b] Marchand, E.; Chaumette, F.: *An Autonomous Active Vision System for Complete and Accurate 3D Scene Reconstruction*, *International Journal of Computer Vision*, Bd. 32, Nr. 3, 1999, S. 171–194.
- [Mas98] Massios, N. A.; Fisher, R. B.: *A Best Next View Selection Algorithm incorporating a Quality Criterion*, in P. H. Lewis, M. S. N. (Hrsg.): *9th British Machine Vision Conference*, British Machine Vision Association, Southampton, England, 1998, S. 780–789.
- [Mat89] Matthies, L.; Kanade, T.; Szeliski, R.: *Kalman Filter-based Algorithms for Estimating Depth from Image Sequences*, *International Journal of Computer Vision*, Bd. 3, Nr. 3, 1989, S. 209–236.
- [Mav93] Maver, J.; Bajcsy, R.: *Occlusion as a guide for planning the next view*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 17, Nr. 5, 1993, S. 417–433.
- [May79] Maybeck, P.: *Stochastic Models, Estimation and Control*, Bd. I, Academic Press, New York, 1979.
- [McC89] McCloy, D.; Harris, M.: *Robotertechnik*, VCH Verlagsgesellschaft mbH, New York, 1989.
- [McL99] McLauchlan, P.: *Gauge Invariance in Projective 3D Reconstruction*, in *Proceedings of the IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes*, IEEE Computer Society Press, Fort Collins, Colorado, 1999, S. 183–199.
- [Mey01] Meyberg, K.; Vachenauer, P.: *Höhere Mathematik 1*, Springer-Verlag, Berlin, 6. korrig. Ausg., 2001.
- [Mun06] Munkelt, C.; Kühmstedt, P.; Denzler, J.: *Incorporation of a-priori information in planning the next best view*, in Kobbelt, L.; Kuhlen, T.; Aach, T.; Westermann, T. (Hrsg.): *Proceedings Vision, Modelling and Visualization*, Akademische Verlagsgesellschaft Aka GmbH, Berlin, Aachen, 2006, S. 261–268.
- [Mun07] Munkelt, C.; Trummer, M.; Wenhardt, S.; Denzler, J.: *Benchmarking 3D Reconstructions from Next Best View Planning*, in *Proc. of IAPR Conference on Machine Vision Applications (MVA 2007)*, Tokyo, Japan, 2007, S. 552–555.

- [Mun11] Munkelt, C.: *Aktive daten- und modellbasierte Sensorpositionierung zur 3-D Vermessung*, Dissertation, Chair for Computer Vision, Friedrich-Schiller University of Jena, Jena, Germany, 2011.
- [Nie96] Niem, W.; Steinmetz, M.: *Camera Viewpoint Control for the Automatic Reconstruction of 3D Objects*, in *Proceedings of the International Conference on Image Processing*, Bd. 3, IEEE Signal Processing Society, Los Alamitos, Calif., 1996, S. 655 – 658.
- [Nie99] Niem, W.: *Automatische Rekonstruktion starrer dreidimensionaler Objekte aus Kamerabildern*, VDI Verlag GmbH, 1999.
- [Nie03] Niemann, H.: *Klassifikation von Mustern*, Internetausgabe, 2. verbesserte. Ausg., 2003, <http://www5.informatik.uni-erlangen.de/Personen/niemann/klassifikation-von-mustern/m00links.html>.
- [Nip02] Nipp, K.; Stoffer, D.: *Lineare Algebra*, vdf Hochschulverlag AG an der ETH Zürich, 5. durchgesehene. Ausg., 2002.
- [Oht81] Ohta, Y.; Maenobu, K.; Sakai, T.: *Obtaining Surface Orientation from Textels Under Perspective Projection*, in *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981, S. 746–751.
- [Oht99] Ohta, Y.; Tamura, H.: *Mixed Reality*, Ohmsha, Tokyo, 1999.
- [Ola97] Olague, G.; Mohr, R.: *Optimal 3D Sensor Placement to Obtain Accurate 3D Point Positions*, in *Primer Encuentro de Computacion ENC 97: Vision Robotica*, Queretaro, Mexico, 1997, S. 116–123.
- [Ola98] Olague, G.; Mohr, R.: *Optimal Camera Placement to Obtain Accurate 3D Point Positions*, in *Proceedings of 14th International Conference on Pattern Recognition*, Bd. 1, IEEE Computer Society Press, 1998, S. 8–10.
- [Ola02] Olague, G.; Mohr, R.: *Optimal Camera Placement for Accurate Reconstruction*, *Pattern Recognition Journal*, Bd. 35, 2002, S. 927–944.
- [Oli94] Oliensis, J.: *A Linear Solution for Multiframe Structure from Motion*, in *Image Understanding Workshop*, Monterey, California, 1994, S. 1225–1231.
- [Oli01] Oliensis, J.; Genc, Y.: *Fast and Accurate Algorithms for Projective Multi-Image Structure from Motion*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 23, Nr. 6, 2001, S. 546–559.
- [Pap02] Papoulis, A.: *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Inc, Singapore, 4. Ausg., 2002.
- [Pea88] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo, Kalifornien, USA, 1988.
- [Pit99] Pito, R.: *A Solution to the Next Best View Problem for Automated Surface Acquisition*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 10, 1999, S. 1016–1030.

- [Poe97] Poelman, C. J.; Kanade, T.: *A Paraperspective Factorization Method for Shape and Motion Recovery*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 19, Nr. 3, 1997, S. 206–218.
- [Pol99] Pollefeys, M.: *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, Dissertation, Katholieke Universiteit Leuven – Faculteit Toegepaste Wetenschappen, 1999.
- [Pow78a] Powell, M.: *A Fast Algorithm for Nonlinearly Constrained Optimization Calculations*, *Numerical Analysis*, Bd. 630, 1978, S. 144–175.
- [Pow78b] Powell, M.: *The Convergence of Variable metric Methods For Nonlinearly Constrained Optimization Calculations*, in O.L. Mangasarian, R. M.; Robinson, S. (Hrsg.): *Nonlinear Programming*, Bd. 3, Academic press, New York, USA, 1978, S. 27–63.
- [Pre02] Press, W.; Teukolsky, S.; Vetterling, W.; Flannery, B.: *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, 2. Ausg., 2002.
- [Puk93] Pukelsheim, F.: *Optimal Design of Experiments*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapur, 1993.
- [Rin03] Rinne, H.: *Taschenbuch der Statistik*, Verlag Harri Deutsch, Frankfurt a. M., 2003.
- [Ris99] Ristow, R.: *Modeling Real World Objects Combining Photogrammetric Reconstruction and Semantic Description*, in *Proceedings of Vision, Modeling, and Visualization*, Erlangen, Germany, 1999, S. 101–108.
- [Sal98] Salvi, J.; Batlle, J.; Mouaddib, E.: *A robust-coded pattern projection for dynamic 3D scene measurement*, *Pattern Recognition Letters*, Bd. 19, 1998, S. 1055–1065.
- [Sch01] Schmidt, J.; Niemann, H.: *Using Quaternions for Parametrizing 3-D Rotations in Unconstrained Nonlinear Optimization*, in Ertl, T.; Girod, B.; Greiner, G.; Niemann, H.; Seidel, H.-P. (Hrsg.): *Vision, Modeling, and Visualization 2001*, AKA/IOS Press, Berlin, Amsterdam, Stuttgart, Germany, 2001, S. 399–406.
- [Sch02] Schmidt, J.; Niemann, H.; Vogt, S.: *Dense Disparity Maps in Real-Time with an Application to Augmented Reality*, in *Proceedings Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002)*, IEEE Computer Society Press, Orlando, FL USA, 2002, S. 225–230.
- [Sch03] Schmidt, J.; Vogt, F.; Niemann, H.: *Robust Hand-Eye Calibration of an Endoscopic Surgery Robot Using Dual Quaternions*, in Michaelis, B.; Krell, G. (Hrsg.): *Pattern Recognition, 25th DAGM Symposium*, Bd. 2781 von *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, Heidelberg, New York, 2003, S. 548–556.

- [Sch06a] Schmidt, J.: *3-D Reconstruction and Stereo Self-Calibration for Augmented Reality*, Logos Verlag, Berlin, 2006.
- [Sch06b] Schwarz, H. R.; Köckler, N.: *Numerische Mathematik*, B. G. Teubner Verlag, Stuttgart, 6. überarb. Ausg., 2006.
- [Sch08] Scholz, I.: *Reconstruction and Modeling of Static and Dynamic Light Fields, Reihe: Studien zur Mustererkennung*, Dissertation, Universität Erlangen-Nürnberg, Berlin, 2008.
- [Sco01] Scott, W. R.; Roth, G.; Rivest, J.-F.: *View Planning with a Registration Component*, in *Proceedings of the 3D Imaging and Medling Conference*, National Research Council of Canada, Quebec, 2001, S. 127–134.
- [Sco03] Scott, W.; Roth, G.; Rivest, J.-F.: *View Planning for Automated Three-Dimensional Object Reconstruction and Inspection*, *ACM Computing Surveys*, Bd. 35, Nr. 1, 2003, S. 64–96.
- [Shi89] Shiu, Y. C.; Ahmad, S.: *Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX = XB$* , *IEEE Transactions on Robotics and Automation*, Bd. 5, Nr. 3, 1989, S. 16–29.
- [Shi94] Shi, J.; Tomasi, C.: *Good Features to Track*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Seattle, WA, 1994, S. 593–600.
- [Shu99] Shum, H.-Y.; Ke, Q.; Zhang, Z.: *Efficient Bundle Adjustment with Virtual Key Frames: A Hierarchical Approach to Multi-Frame Structure from Motion*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Bd. 2, Fort Collins, Colorado, 1999, S. 538–543.
- [Sie96] Siegert, H.-J.; Bocionek, S.: *Robotik: Programmierung intelligenter Roboter*, Springer Verlag, Berlin, 1996.
- [Sla80] Slama, C.: *Manual of Photogrammetry*, American Society of Photogrammetry, Falls Church, VA, 4. Ausg., 1980.
- [Sla02] Slater, M.; Steed, A.; Chrysanthou, Y.: *Computer Graphics and Virtual Environments. From Realism to Real-Time*, Addison-Wesley Publishing Co., Reading, MA, 2002.
- [Sta94] Stark, H.; Woods, J. W.: *Probability, Random Processes and Estimation Theory For Engineers*, Prentice Hall, Englewood Cliffs, New Jersey, 2. Ausg., 1994.
- [Sze98] Szeliski, R.; Torr, P.: *Geometrically Constrained Structure from Motion: Points on Planes*, in *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE)*, Freiburg, Germany, 1998, S. 171–186.
- [Tar95a] Tarabanis, K.; Allen, P.; Tsai, R.: *A survey of sensor planning in computer vision*, *IEEE Transactions on Robotics and Automation*, Bd. 11, Nr. 1, 1995, S. 86–104.

- [Tar95b] Tarbox, G.; Gottschlich, S.: *Planning for complete sensor coverage in inspection*, *Computer Vision and Image Understanding*, Bd. 61, Nr. 1, 1995, S. 84–111.
- [Tom91] Tomasi, C.; Kanade, T.: *Detection and Tracking of Point Features*, Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991.
- [Tre97] Trefethen, L.; Bau III, D.: *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1997.
- [Tri00] Triggs, B.; McLauchlan, P. F.; Hartley, R. I.; Fitzgibbon, A. W.: *Bundle Adjustment — A Modern Synthesis*, in Triggs, B.; Zisserman, A.; Szeliski, R. (Hrsg.): *Vision Algorithms: Theory and Practice*, Springer-Verlag, 2000, S. 298–372.
- [Tru98] Trucco, E.; Verri, A.: *Introductory Techniques for 3-D Computer Vision*, Addison-Wesley, Massachusetts, 1998.
- [Tru09a] Trummer, M.; Munkelt, C.; Denzler, J.: *Combined GKL Feature Tracking and Reconstruction for Next Best View Planning*, in Denzler, J.; Notni, G.; Suesse, H. (Hrsg.): *DAGM 2009*, Bd. 5748 von *Lecture Notes in Computer Science*, Springer, 2009, S. 161–170.
- [Tru09b] Trummer, M.; Munkelt, C.; Denzler, J.: *Extending GKL Tracking – Feature Tracking for Controlled Environments with Integrated Uncertainty Estimation*, in Salberg, A.-B.; Hardeberg, J.; Jenssen, R. (Hrsg.): *Proceedings of IAPR Scandinavian Conference on Image Analysis (SCIA)*, Bd. 5575 von *Lecture Notes in Computer Science (LNCS)*, Springer Berlin Heidelberg, 2009, S. 460–469.
- [Tru10] Trummer, M.; Munkelt, C.; Denzler, J.: *Online Next-Best-View Planning for Accuracy Optimization Using an Extended E-Criterion*, in International Association for Pattern Recognition, (Hrsg.): *Proceedings of 20th International Conference on Pattern Recognition (ICPR)*, IEEE Computer Society Press, Istanbul, 2010, S. 1642–1645.
- [Tru11] Trummer, M.: *Purposive Three-dimensional Reconstruction by Means of a Controlled Environment*, Dissertation, Chair for Computer Vision, Friedrich-Schiller University of Jena, Jena, Germany, 2011.
- [Tsa85] Tsai, L. W.; Morgan, A. P.: *Solving the kinematics of the most general six and five-degree-of-freedom manipulators by continuation methods*, *Transactions of the ASME, Journal of Mechanisms, Transmissions and Automation in Design*, Bd. 107, 1985, S. 189–200.
- [Tsa87] Tsai, R. Y.: *A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses*, *IEEE Journal of Robotics and Automation*, Bd. RA-3, Nr. 4, 1987, S. 323–344.
- [Tsa89] Tsai, R. Y.; Lenz, R. K.: *A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration*, *IEEE Transactions on Robotics and Automation*, Bd. 5, Nr. 3, 1989, S. 345–358.

- [Vog01] Vogt, S.: *Anwendung von Stereoverfahren zur Verdeckungsrechnung im Bereich erweiterte Realität*, Diplomarbeit, Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, 2001, Supervisors: J. Schmidt, B. Heigl.
- [Vog06] Vogt, F.: *Augmented Light Field Visualization and Real-Time Image Enhancement for Computer Assisted Endoscopic Surgery*, Der Andere Verlag, Tönnig, 2006.
- [Wal02] Walter, W.: *Analysis 2*, Springer-Verlag, Berlin, 5., erw. Ausg., 2002.
- [Wan92] Wang, C.: *Extrinsic Calibration of a Vision Sensor Mounted on a Robot*, *IEEE Transactions on Robotics and Automation*, Bd. 8, Nr. 2, 1992, S. 161–175.
- [Wel06] Welch, G.; Bishop, G.: *An Introduction to the Kalman Filter*, University of North Carolina at Chapel Hill, Department of Computer Science, 2006, http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf; zuletzt überprüft: 15.06.2007.
- [Wen04] Wenhardt, S.; Denzler, J.; Niemann, H.: *On Minimizing Errors in 3-D-Reconstruction for Stereo Camera Systems*, in Geppener, V.; Gurevich, I.; Ivanova, S.; Nemirko, A.; Niemann, H.; Puzankov, D.; Trusova, Y.; Zhuravlev, Y. (Hrsg.): *7th International Conference on Pattern Recognition and Image Analysis 2004: New Information Technologies*, SPbETU, St. Petersburg, 2004, S. 562–565.
- [Wen06] Wenhardt, S.; Deutsch, B.; Hornegger, J.; Niemann, H.; Denzler, J.: *An Information Theoretic Approach for Next Best View Planning in 3-D Reconstruction*, in Tang, Y.; Wang, S.; Lorette, G.; Yeung, D.; Yan, H. (Hrsg.): *The 18th International Conference on Pattern Recognition*, Bd. 1 von *IEEE Computer Society Order Number P2521*, IEEE Computer Society Press, Hong Kong, 2006, S. 103–106.
- [Wen07a] Wenhardt, S.; Denzler, J.; Niemann, H.: *On Minimizing Errors in 3D Reconstruction for Stereo Camera Systems*, *Pattern Recognition and Image Analysis*, Bd. 17, Nr. 2, 2007, S. 337–348.
- [Wen07b] Wenhardt, S.; Deutsch, B.; Angelopolou, E.; Niemann, H.: *Active Visual Object Reconstruction using D-, E-, and T-Optimal Next Best Views*, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, IEEE Computer Society Press, Los Alamitos, California, Washington, Tokyo, 2007, S. 1–7.
- [Won98] Wong, L. M.; Dumont, C.; Abidi, M. A.: *A next best view algorithm for object reconstruction*, in *Proc. of SPIE Conference on Sensor Fusion and Decentralized Control in Robotics Systems*, Bd. 3523, Boston, Massachusetts, 1998, S. 191–200.
- [Woo83] Wood, G. A.: *Realities of Automatic Correlation Problem*, *Photogrammetric Engineering and Remote Sensing*, Bd. 49, 1983, S. 537–538.
- [Yu05] Yu, Y.-K.; Wong, K.-H.; Chang, M. M.-Y.: *Recursive 3D Model Reconstruction Based on Kalman Filtering*, *IEEE Transactions on Systems, Man and Cybernetics - Part B*, Bd. 35, Nr. 3, 2005, S. 587–592.

- [Zha96] Zhang, Z.: *On the Epipolar Geometry Between Two Images With Lens Distortion*, in *Proceedings International Conference Pattern Recognition (ICPR)*, Wien, 1996, S. 407–411.
- [Zha00] Zhang, Z.: *A Flexible New Technique for Camera Calibration*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 22, Nr. 11, 2000, S. 1330–1334.
- [Zin02] Zinßer, T.: *Robuste Schätzung der Korrespondenzen zwischen 3-D Punktemengen*, Diplomarbeit, Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, 2002, Supervisors: J. Schmidt, F. Vogt.
- [Zin04] Zinßer, T.; Gräßl, C.; Niemann, H.: *Efficient Feature Tracking for Long Video Sequences*, in Rasmussen, C. E.; Bülthoff, H. H.; Giese, M. A.; Schölkopf, B. (Hrsg.): *Pattern Recognition, 26th DAGM Symposium*, Bd. 3175 von *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 2004, S. 326–333.
- [Zin05] Zinßer, T.; Schmidt, J.; Niemann, H.: *Point Set Registration with Integrated Scale Estimation*, in Sadykhov, R.; Ablameiko, S.; Doudkin, A.; Podenok, L. (Hrsg.): *Proceedings of the Eighth International Conference on Pattern Recognition and Image Processing (PRIP 2005)*, Minsk, Republic of Belarus, 2005, S. 116–119.
- [Zin11] Zinßer, T.: *Efficient and Robust Algorithms for Sparse 3-D Reconstruction from Image Sequences*, Dissertation, Friedrich-Alexander Universität Erlangen-Nürnberg, Technische Fakultät, Uelvesbüll, 2011.
- [Zob04] Zobel, M.: *Optimale Brennweitenwahl für die multiokulare Objektverfolgung*, Bd. 11, Logos Verlag, Berlin, 2004.

Index

- Abbruchkriterium, 102, 104–106
- Absolute Konik, 24, 26
- Arbeitsbereich (Roboter), 34, 35, 99–101, 129, 157, 162
- Arbeitsbereichs (Roboter), 102
- Auffrischung, *siehe* Kalman-Filter
- Bündelausgleich, 15, 36, 40, 41
- Bayes-
 - Ansatz, 72, 84
 - Formel, 73, 75
 - Netze, 9
- Beobachtung, 72–76, 79–83, 85–87, 91, 92, 94, 97, 102, 104, 145–147, 158, 161, 162
- Beobachtungs-
 - funktion, 72, 74, 75, 77, 78, 80, 81, 85, 97, 105
 - matrix, 75, 77, 81, 85
 - rauschen, 72, 74–77, 79, 80, 83–85, 93, 156, 158
 - vektor, 80, 81
- Bewegungsmodell, 99, 100, 102, 105, 117, 118, 133
- Bildkoordinatensystem, *siehe* Koordinatensystem
- Blockdiagonalstruktur, 81–83, 87
- Brennweite, 16, 48, 78, 85, 86, 110, 120, 152, 153, 158, 160, 162
 - effektive, 17, 27, 110, 120
- Bundle Adjustment, *siehe* Bündelausgleich
- Cardan-Winkel, 61, 85
- Cauchy-Verteilung, 78
- Delauny-Triangulation, 95
- Denavit-Hartenberg-
 - Matrix, 34–36, 100, 101, 105, 162
 - Parameter, 34, 100, 102
 - Regeln, 32
- Dreiecksnetz, 95, 96, 103, 104, 161
- Eigenvektor, 30
- Eigenwert, 30, 37
- Eigenwertkriterium, 89
- Endeffektor, 32, 34, 35
- Entropie, 85–87, 154, 162
 - bedingte, 86
- Entzerrung, 17
- Fischer-Ansatz, 72
- Gauß
 - Dichte, 94
 - Fehlerfunktion, 94
 - Rauschen, 24, 58
 - System, 75, 76
- Gauß-Newton-Verfahren, 41
- Gedächtnislosigkeit, 74
- Gelenk, 27, 31
 - ideales, 31
- Gelenkkoordinatensystem, *siehe* Koordinatensystem
- Hand-Auge-Kalibrierung, 28, 29, 31, 36, 100, 105
- Hand-Auge-Transformation, 15, 27–29, 31, 162
- Hauptpunkt, 16, 17, 27, 48
- Homographie, 26
- ICP, 109, 152
- Kalibriermuster, 15, 22, 27, 105, 108, 110–112, 114, 115, 117, 118, 121, 122, 131–133, 137, 152, 159, 162, 163
- Kalman-Filter, 5, 12, 13, 18, 71, 74–81, 83, 84, 86, 87, 92, 94, 95, 97, 102–105, 145–147, 157, 159, 161
 - Auffrischung, 76, 77, 79–83, 91, 92, 103

- erweitertes, 77
- Kalman-Verstärkungsmatrix, 76, 80, 82
- Kamerakalibrierung, 17, 22, 29, 36, 105
- Kamerakoordinatensystem, *siehe* Koordinatensystem
- Kameramatrix, 17
- Kameraparameter, 10, 15, 22, 36, 40, 41, 48, 50, 62, 69, 74, 85, 86, 103, 105, 110, 117, 155–159, 162, 163
 - extrinsische, 15, 17, 22, 26, 27, 31, 34, 36, 38
 - intrinsische, 16, 22, 24, 36, 38
- Kinematik, 15
 - direkte, 31, 34–36, 156, 159, 162
 - inverse, 31, 34, 35, 99, 100, 157
 - kinematische Kette, 32, 34
- kinematische Kette, 32, 34
- Koordinatensystem, 16, 32
 - Basis-, 15, 32
 - Bild-, 16
 - Endeffektor-, 15, 32
 - Gelenk-, 32
 - Kamera-, 16, 38, 85, 103, 157
 - Welt-, 16, 23, 28, 38, 85, 108, 109, 157, 159
- Korrespondenzproblem, 108
- MAP-Schätzer, 73, 75
- Maximum-Likelihood-Schätzung, 24
- ML-Schätzer, 73
- MMSE-Schätzer, 73–76
- Monte-Carlo-Integration, 94, 95, 98
- Monte-Carlo-Simulation, 58, 59, 95
- Objektverfolgung, 71
- Optimierungskriterium, 72, 84, 87, 91–93, 97, 98, 104, 148
- Projektionsmatrix, 15, 23, 41
- Projektionsmodell, 15, 18
 - affines, 18
 - paraperspektivisches, 16, 19, 94
 - perspektivisches, 16, 18
- Punkt detektion, 36, 37, 150
- Punkt detektor, 146, 159
- Punktkorrespondenzen, 36, 79, 91, 103, 104, 110, 145, 148, 154, 155
- Punktverfolger, 79, 91, 103, 104, 110, 133, 138, 140, 146, 147, 153, 155–157
- Punktverfolgung, 36, 37, 40, 150
- Quaternionen, 29, 85
 - duale, 29
- Rückprojektionsfehler, 40, 41
- Rotation, 48, 54
 - Achse/Winkel, 29–31
- Rotationsmatrix, 18, 26, 27, 30, 31, 33
- Rotationswinkel, 50
- Selbstkalibrierung, 22
- Selbstverdeckung, 71, 91, 93, 95, 103, 107, 110, 131, 138, 142, 144, 152–154, 161, 163
- Sichtbarkeit, 80, 84, 93, 97, 98, 104, 120, 161
- Skew, 17, 48
- Stereosystem, 107, 110, 152, 163
- Streifenprojektion, 3, 109, 142, 143
- Strukturmatrix, 36
- Taylor-Reihe, 21, 22, 37, 66, 77, 78, 102, 119, 159–161
- Transformationssatz für Dichten, 94, 96
- Translation, 48–50
- Translationsvektor, *siehe* Kameraparameter, extrinsische
- Triangulation, 15, 36, 38, 41, 82, 103, 104, 112, 117, 120, 159
- Verzerrungen
 - radiale, 17
 - tangentiale, 17
- Verzerrungskoeffizienten, 26
- Verzerrungsparameter, 16, 48
- Vorhersageschritt, 76, 77, 79
- Weltkoordinatensystem, *siehe* Koordinatensystem
- Wiederholgenauigkeit, 131, 132, 144, 157
- Zustand, 72–76, 80–82, 84, 87, 90, 93, 97, 103, 104, 161
- Zustandsübergang, 74, 75, 79, 92
- Zustandsschätzung, 71–75, 79, 85, 90, 103, 158, 161, 163