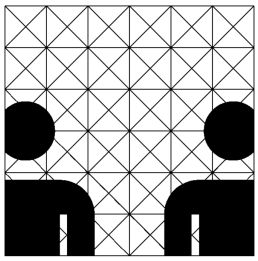


Universität Hamburg
Fakultät für Mathematik, Informatik
und Naturwissenschaften

Technische Aspekte Multimodaler Systeme (TAMS)

Diplomarbeit

3–dimensionale Rekonstruktion einer Tischszene aus monokularen Handkamera–Bildsequenzen im Kontext autonomer Serviceroboter

vorgelegt im
Juni 2006

Sascha Jockel

Julius-Leber-Straße 12
22765 Hamburg

9jockel@informatik.uni-hamburg.de
Matrikelnummer: 5200923

Erstbetreuung: Prof. Dr. Jianwei Zhang
Zweitbetreuung: Dr. Werner Hansmann

„Computer sind das bis heute genialste Werk menschlicher Faulheit.“
– *Slogan einer IBM Werbekampagne in den ’70iger Jahren*

Abstract

Image driven environment perception is one of the main research topics in the field of autonomous robot applications. This thesis will present an image based three dimensional reconstruction system for such robot applications in case of daily table scenarios.

Perception will be done at two spatial-temporal varying positions by a micro-head camera mounted on a six-degree-of-freedom robot-arm of our service-robot TASER. Via user interaction the epipolar geometry and fundamentalmatrix will be calculated by selecting 10 corresponding corners in both input images predicted by a Harris-corner-detector. The images then will be rectified by the calculated fundamentalmatrix to bring corresponding scanlines together on the same vertical image coordinates. Afterwards a stereocorrespondence is made by a fast Birchfield algorithm that provides a 2.5 dimensional depth map of the scene. Based on the depth map a three dimensional textured point-cloud is presented as interactive OpenGL scene model.

Zusammenfassung

In der Robotik stellt mittlerweile die bildbasierte Umgebungsmodellierung einen großen Forschungszweig dar. In dieser Diplomarbeit wird ein bildbasiertes dreidimensionales Rekonstruktionssystem für alltägliche Tischszenen entwickelt.

Die Umgebung wird mit einer Mikro-Kopf Kamera an der mobilen Hand eines Roboterarms mit sechs Freiheitsgraden akquiriert. Dabei werden die Bilder aus zwei verschiedenen Positionen zeitnah nacheinander aufgenommen. Durch nutzergesteuerte Selektion von 10 korrespondierenden Punkten beider Ansichten wird die Epipolargeometrie, mathematisch ausgedrückt durch die Fundamentalmatrix, berechnet. Die Korrespondenzpunkte können aus einer Menge von zuvor mit Harris-Kanten-Detektor ermittelten Ecken ausgewählt werden. Anschließend werden mit Hilfe der Fundmentalmatrix die Eingabebilder rektifiziert um korrespondierende Epipolarlinien in neuen virtuellen Ansichten auf die selbe vertikale Bildkoordinate abzubilden. Nachfolgend wird ein schneller Stereoalgorithmus nach Birchfield zur Disparitätsanalyse und der Erstellung einer zweieinhalbdimensionalen Tiefenkarte eingesetzt. Aus der Tiefenkarte wird ein dreidimensionales interaktives OpenGL Modell berechnet, das dem Nutzer als realistische, texturierte Punktwolke dargeboten wird.

Inhaltsverzeichnis

Abbildungsverzeichnis

Notation

In dieser Arbeit werden zweidimensionale Punkte sowie Geraden durch kleine, fettgedruckte Buchstaben gekennzeichnet. Matrizen seien durch große, fettgedruckte Buchstaben und dreidimensionale Punkte durch geneigte Großbuchstaben kenntlich gemacht.

Die erste Tabelle zeigt mathematische Notationsformen, während die zweite Tabelle die Abkürzungen dieser Arbeit zusammenfasst.

Ausdruck	*Bedeutung*
$\mathbf{A}_{n\times m}$	$n \times m$ Matrix $\mathbf{A}$
$[\ldots]^T$	Der/die transponierte Vektor/Matrix durch vertauschen der Zeilen- und Spaltenindizes
$[\ldots]^{-1}$	Inverse Matrix (wobei $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$)
$[\ldots]^{-T}$	Die transponierte Inverse einer Matrix, wobei die Reihenfolge unerheblich ist, da $[[...]^{-1}]^T = [[...]^T]^{-1}$
C_1, C_2	Optisches Zentrum der Kamera an Position 1 und Position 2 im dreidimensionalen euklidischen Raum
$\mathcal{I}_1, \mathcal{I}_2$	Die beiden Projektionsebenen der Kameras
I	Intensität eines Bildpunktes $\mathbf{m}$
M	3D-Vektor $(x,\ y,\ z)^T$. Raumpunkt im dreidimensionalen euklidischen Raum
$\mathbf{m}_i$	3D-Vektor $(u,\ v,\ w)^T$. Punkt in Pixelkoordinaten mit $w = 0$. Entspricht Projektion von M auf die Projektionsebene $\mathcal{I}_i$ mit $i \in \{1, 2\}$
$\tilde{\mathbf{m}}_i$	3D-Vektor $(x,\ y,\ z)^T$. Punkt in homogenen Koordinaten $(x,\ y,\ z)^T$ mit $z = 1$
$\mathbf{m'}_i$	3D-Vektor $(x,\ y,\ z)^T$. Punkt in Sensorkoordinaten. Entspricht Projektion von M auf die Fokalebene der i-ten Kamera mit $z = f$
x, y, z	x-,y-,z-Komponente eines dreidimensionalen Punktes
u, v, w	Spalten u und Zeilen v der von den Kameras aufgenommenen Bilder in Pixelkoordinaten

(*Fortsetzung nächste Seite*)

Ausdruck	*Bedeutung*
$\mathbf{I}_{n\times n}$	Einheitsmatrix mit $diag\{1,..,1\}$
B	Basislinie, Verbindungsgerade zwischen den optischen Zentren
e	Epipol
ℓ	Epipolarlinie
π	Epipolarebene, aufgespannt durch die drei euklidischen Raumpunkte M, C_1, C_2
$\mathbf{F}$	Fundamentalmatrix (3×3 Matrix)
$\mathbf{E}$	Essentialmatrix (3×3 Matrix)
$\mathbf{K}$	Kameramatrix (3×3 Matrix)
$\mathbf{R}$	Rotationsmatrix (3×3 mit $\mathbf{R} = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{pmatrix}$)
$det(X) = y$	Die Determinante der Matrix $\mathbf{X} = y$, wobei $y \in 0, 1, 2$. Eine Determinante ist genau dann ungleich 0, wenn die Menge der Spaltenvektoren linear unabhängig über Körper K ist
$\Re^n$	Euklidischer Vektorraum der Dimension n, im Allgemeinen $n \in \{1, 2\}$
$\langle \mathbf{t}, \mathbf{u} \rangle$	Skalarprodukt (inneres Produkt) zweier Vektoren $\mathbf{t}$, $\mathbf{u}$
$\mathbf{t} \times \mathbf{u}$	Kreuzprodukt (äußeres Produkt) zweier Vektoren $\mathbf{t}$, $\mathbf{u}$

Abkürzung	*Bedeutung*
2D	zweidimensional
3D	dreidimensional
Abb.	Abbildung
Abk.	Abkürzung
engl.	englisch
Gl.	Gleichung
LGS	lineares Gleichungssystem
NCC	Normierte Kreuzkorrelation (engl. normalized cross-correlation)
RAC	Radial alignment constraint
SAD	Mittlerer absoluter Fehler (engl. sum of absolute differences)
SSD	Mittlerer quadratischer Fehler (engl. sum of squared differences)
SVD	Singulärwertzerlegung (engl. singular value decomposition)
u.a.	unter anderen
vgl.	vergleiche

1 Einleitung

„We see because we move; we move because we see.“

– James J.Gibson, *The Perception of the Visual World*

Diese Worte des Verhaltensforschers J.J. Gibson bringen zum Ausdruck, dass Wahrnehmung im Allgemeinen ein Prozess darstellt, um mit der Umwelt in Verbindung zu bleiben und in ihr zu existieren. Seiner Ansicht nach wird die Wahrnehmung eines Objektes von Lebewesen nur auf dessen Interaktionsmöglichkeit reduziert.

Wahrnehmung und Objektinteraktion sind ebenfalls Forschungsschwerpunkte der Robotik. Um jedoch mit Objekten der Umwelt zu interagieren, wird zuvor eine Repräsentation der Umwelt benötigt. Findet die Wahrnehmung der Umwelt dabei mittels einer Kamera statt, spricht man von maschinellem Sehen (engl. *Computer-Vision*). Ziel dieses Forschungsbereiches ist es unter anderem, den bei der Abbildung von 3D→2D verlorenen Tiefeneindruck durch mehrere Kameras oder zeit- und örtlich divergierende Aufnahmen einer Kamera zurückzuerlangen, um ein dreidimensionales maßgetreues Modell der Umwelt zu erstellen. Die Anwendungsgebiete dreidimensionalen maschinellen Sehens sind äußerst vielfältig und reichen über den Bereich Robotik hinaus. Aufgabengebiete sind Fertigungstechniken, Inspektion, Qualitätskontrolle, Fahrassistenzsystemen und Planetenvisualisierung, um nur einige zu nennen.

Ferner bietet die dreidimensionale Modellierung der Umgebung eines mobilen Roboters gegenüber den üblichen zweidimensionalen Grundrißkarten eine Reihe von Vorteilen sowie zusätzliche Möglichkeiten bei der Gestaltung der Mensch-Maschine Schnittstelle, der Navigation bzw. Positionsbestimmung sowie der Visualisierung von Systemzuständen.

Der Bereich Computer-Vision hat in den letzten Jahren einige Verfahren hervorgebracht, die eine 3D Rekonstruktion mit immer weniger erforderlichem Vorwissen über die zu modellierende Szenerie und die Aufnahmeumstände bzw. -parameter gestatten. So kann anstelle einer kalibrierten Stereokamera aus monokular aufgenommenen Bildern die Anordnung der Aufnahmen zueinander extrahiert werden [Fau93, Zha96]. Die Abbildungseigenschaften der verwendeten Kamera lassen sich ebenfalls aus einer Bildsequenz ableiten, so dass auch eine Kalibrierung vorab nicht mehr zwingend

erforderlich ist. Pollefeys und Heyden haben entsprechende Verfahren in [PKVG98] und [HÅ96] vorgestellt.

1.1 Motivation und Ziel dieser Arbeit

Die hier vorliegende Arbeit hat ein multimodales dreidimensionales Rekonstruktionssystem zum Ziel. Dieses soll im Kontext der Servicerobotik ein 3D Modell einer alltäglichen Tischszene erstellen. Der hier vorgestellte Ansatz benutzt die Bilder der Handkamera eines mobilen Roboters zur 3D Rekonstruktion einer vorab unbekannten Einsatzumgebung und resultiert in einem texturierten virtuellen Umgebungsmodell, das für weitere Interaktion des Roboters genutzt werden kann. Die Repräsentation des Modells beruht dabei auf einem interaktiven OpenGL Modell. Die Multimodalität entsteht aus der Verarbeitung und Fusion verschiedenster Sensordaten, in diesem Fall der Kamerabilder und der Armstellung.

Der Wissenschaftszweig der Servicerobotik beschäftigt sich u.a. mit der nutzbringenden Einbindung von Robotern in alltägliche Aufgabengebiete. Beispielhaft seien hier autonome Helfer für körperlich beeinträchtigte Menschen oder für Menschen die durch ihre Behinderung bei alltäglichen Aufgaben ihrem Handicap unterliegen, genannt. Beispielszenarien, die in genau jene Richtung gehen und eine Entwicklung eines flexiblen und mobilen 3D Rekonstruktionssystems dringend erforderlich machen, sind im folgendem aufgeführt.

Alltägliche Tischszene Ein Roboter soll dem Nutzer etwas vom Tisch reichen und dabei Acht geben, dass er beim Greifen des gewünschten Objektes nichts umstößt. Für eine robuste Greifplanung würde ein dreidimensionales Modell der Objekte auf dem Tisch benötigt. Um eine Repräsentation einer Tischszene zu erlangen, könnte natürlich anhand einer herkömmlichen Stereokamera ein 3D Model erstellt werden. Verdeckungen von Objekten, insbesondere dem eigenen Arm des Roboters, werden hierbei nur begünstigt und stellen ein Problem dar. Da von einer Bewegung des Roboters um den Tisch aufgrund häufigen Platzmangels abgesehen werden soll, bietet sich der Einsatz einer monokularen Handkamera an. Mit dem Arm können in sicherer Höhe Aufnahmen aus verschiedenen Betrachtungswinkeln akquiriert werden, ohne die mobile Roboterplattform dabei unnötig zu bewegen. Eine Verdeckung durch andere Roboterkomponenten wird durch die Nutzung der Handkamera vermieden.

Objekte in einem Unterschrank Der Roboter soll dem Nutzer Objekte aus einem Unterschrank heraus geben. Auch hier ist für eine erfolgreiche Greifplanung

ein virtuelles 3D Modell äußerst hilfreich. Hier scheidet die Nutzung eines Stereokopfes, meist montiert an der höchsten Stelle eines Roboters, wegen der schlechten Einsicht in solch einen niedrigen Schrank von vornherein aus. Mit einer Roboterhand kann der Unterschrank geöffnet und die Bildakquise mit der Handkamera direkt am Einsatzort bewerkstelligt werden.

In dieser Arbeit soll der Fokus auf das erste beschrieben Szenario gelegt werden. Die dreidimensionale Rekonstruktion einer Tischszene ist Inhalt der kompletten vorliegenden Arbeit.

1.2 Gliederung der Arbeit

Im zweiten Kapitel werden die für das Verständnis dieser Arbeit notwendigen mathematisch-theoretischen Abbildungseigenschaften eines Kameramodells zur Abbildung dreidimensionaler Szenen auf zweidimensionale Bildebenen erörtert. Dabei wird auch auf die Verzeichnungseigenschaften und deren mathematische Formulierung sowie auf Kameralinsen eingegangen. Zudem wird ein verbreiteter Algorithmus zur Kalibration von Kameras vorgestellt.

Die Einführung in die Basis-Prinzipien wird im dritten Kapitel fortgeführt, indem das Modell um eine weitere Kamera ergänzt wird. Es werden wichtige Beziehungen zwischen zwei Ansichten einer Szene erläutert, die sich unter dem Begriff der Epipolargeometrie vereinen. Es werden Verfahren vorgestellt die mittels dieser Beziehungen neue, virtuelle 2D Ansichten einer Szene generieren können.

Das vierte Kapitel behandelt mit der Korrespondenzanalyse eines der klassischen und fundamentalen Themenbereiche der Bildanalyse. Es werden aktuelle Berechnungsverfahren zur Ermittlung korrespondierender Bildpunkte aufgeführt und verglichen. Des weiteren werden allgemeine Probleme der Stereoanalyse dargelegt.

Die Hard- und Software mit der die Bilddaten für diese Arbeit akquiriert werden, wird im fünften Kapitel beschrieben.

Das sechste Kapitel stellt das in dieser Arbeit entwickelte 3D Rekonstruktionssystem anhand der theoretischen Grundlagen der vorherigen Kapitel vor. Es werden einige Ergebnisse, sowie Analysen zur Qualität der erlangten Tiefenwerte vorgestellt und Probleme der einzelnen Abschnitte eines solchen Systems diskutiert. Zudem wird eine Übertragbarkeit des Rekonstruktionssystems, über das eigentliche Aufgabengebiet einer Tischszene hinaus, beschrieben.

Eine Reflektion dieser Arbeit wird in der Zusammenfassung im siebten Kapitel gegeben. Spezielle Erweiterungsmöglichkeiten des entwickelten 3D Rekonstruktionssystems werden ebenso gegeben, wie ein allgemeiner Ausblick möglicher weiterer wissenschaftlicher Arbeiten. Aufbauend auf den Ergebnissen dieser Arbeit und den bildbasierten Rekonstruktionssystemen allgemein werden diese Erweiterungsmöglichkeiten diskutiert.

Um die Lesbarkeit dieser Arbeit zu erleichtern, wurden einige Themen in den Anhang verlagert.

1.3 Hinweise

Ich werde in dieser Arbeit einige Begrifflichkeiten bei deren englischen Nomenklatur nennen, da sie sich einerseits im deutschen Fachkollegium schon seit Jahren mit ihrer englischen Bezeichnung etabliert haben, oder sie meiner Meinung nach nicht präzise übersetzbar sind. Des weiteren werden einige Grundlagen dieser Arbeit, die sich in annähernd allen Standardwerken zu dem Thema *3D Vision* überschneiden, nicht referenziert seien. Sie werden als bekannt vorausgesetzt und können in der Großzahl der angegebenen Publikationen nachgeschlagen werden, und werden somit nicht zwangsläufig auf einen einzigen Autor oder eine Liste von Autoren verweisen.

Einige Abschnitte dieser Arbeit sind als Exkurs gekennzeichnet und wurden aufgenommen, um einen vollständigeren Überblick über den Stand der Technik zu geben. Die Exkurse haben rein informativen Charakter und die beschriebenen Verfahren wurden in dieser Arbeit nicht realisiert. Sie wurden Zwecks Vollständigkeit aufgenommen.

2 Geometrie einer Ansicht

Bevor die Informationsgewinnung aus 3D-Szenen erörtert werden kann, ist es wichtig die Prozesse der Bildproduktion – genauer den Abbildungsprozess eines beliebigen dreidimensionalen Raumpunktes auf dessen zweidimensionale Bildkoordinaten – zu verstehen. Daher soll im folgendem zuerst das approximierende Modell einer Kamera ohne Linsenverzeichnung vorgestellt werden, gefolgt von der mathematischen Formulierung der Abbildungs- und Projektionseigenschaften. Das bekannteste und auch in dieser Arbeit verwendete Modell ist das Lochkamera-Modell[1]. Darauf aufbauend wird kurz auf mögliche Verzeichnungen eingegangen. Abschließend wird ein Kalibrierungsverfahren zur Ermittlung der Kameraparameter des Lochkameramodells vorgestellt.

2.1 Kameramodell

2.1.1 Lochkamera-Modell

Das Lochkamera-Modell beschreibt die perspektivische Abbildung des sichtbaren projektiven dreidimensionalen Raumes $\mathcal{P}^3$ über ein optisches Projektionszentrum C auf die zweidiemensionale projektive Ebene $\mathcal{P}^2$, die das Projektionszentrum nicht enthält [Fau95]. Das optische Zentrum der Kamera entspricht in der vorherigen Formulierung von O. Faugeras dem Projektionszentrum C und liegt in der Fokalebene, auch Brennebene genannt. Die zweidiemensionale projektive Ebene $\mathcal{P}^2$ ist die Bildebene $\mathcal{I}$ und stellt eine skalierte Punktspiegelung der (sichtbaren) betrachteten Szene am Brennpunkt C dar. Die optische Achse verläuft orthogonal durch das Projektionszentrum C und durchstößt die Bildebene im Kamerahauptpunkt c. Zur intuitiven grafischen Darstellung wird die Bildebene in der Literatur im allgemeinen vor der Fokalebene notiert[2]. Die Abbildung 2.1 stellt das Prinzip der Lochkamera mit der punktsymme-

[1] Die sog. Camera Obscura oder auch perspektivische Kamera ist lediglich ein approximierendes Modell einer realen Kamera.

[2] Dies führt in der folgenden mathematischen Beschreibung und Herleitung zu keinerlei Konsequenzen und hat nur eine Spiegelung der x- und y-Achse zur Folge. Es erleichtert jedoch die grafische Darstellung und wird in allen weiteren Abbildungen so verwendet. Die einzige Ausnahme stellt Abbildung 2.1 dar.

trischen, skalierten Abbildung eines Szenenobjektes des dreidimensionalen Raumes anschaulich dar.

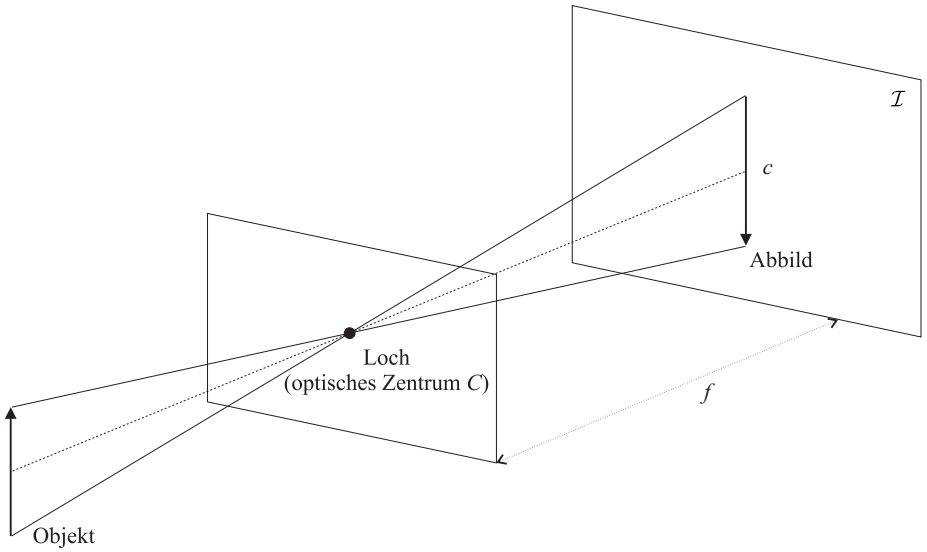


Abbildung 2.1: Projektion einer 3D-Szene auf eine Bildebene $\mathcal{I}$.

Der Prozess der optischen Abbildung ist unterteilbar in drei Schritte: Der externen, der perspektivischen und der internen Transformation. Wie zuvor erläutert befindet sich das Projektionszentrum der Kamera in der Fokalebene und bildet den Ursprung des Kamerakoordinatensystems. Die *externe Transformation* (Gl. 2.1) stellt die eindeutige Relation zwischen dem Kamerakoordinatensystem und einem frei wählbaren Weltkoordinatensystem über eine euklidische Transformation her.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_c = \mathbf{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix}_w + \mathbf{t} \tag{2.1}$$

In Gl. 2.1 beschreibt $\mathbf{R}$ eine 3×3 Rotationsmatrix mit drei Freiheitsgeraden entsprechend der Drehwinkel um die Koordinatenachsen und $\mathbf{t}$ verkörpert einen 3×1 Translationsvektor. $\mathbf{R}$ ist orthogonal und hat die Eigenschaft $det(\mathbf{R}) = 1$. Folglich beschreiben $\mathbf{R}$ und $\mathbf{t}$ die Orientierung und Position der Kamera im Raum relativ zum Weltkoordinatensystem.

Die *perspektivische Transformation* (Gl. 2.3) überführt durch eine Projektion die Punkte aus dem Kamerakoordinatensystem in die Sensorkoordinaten des CCD-Chip[3]

[3] CCD, *Charge-coupled Device*: Elektronisches Bauteil, aufgebaut aus einer Matrix von lichtempfindlichen Zellen zur ortsauflösenden Messung der Lichtstärke.

der Kamera. Eine Projektion eines 3-D Raumpunktes in 3D Kamerakoordinaten $(x, y, z)_c$ in einen Punkt (x, y) in 2D Sensorkoordinaten ist durch Gl. 2.3 in homogenen Koordinaten[4] bis auf den Skalierungsfaktor λ eindeutig bestimmt. Diese Projektion bezieht sich auf eine verzeichnisfreie, perspektivische Projektion ausgedrückt durch die Zentralprojektion (Gl. 2.2). Die Matrix **P'** in Gl. 2.4 wird als perspektivische Projektionsmatrix bezeichnet.

$$\frac{x}{x_c} = \frac{y}{y_c} = \frac{f}{z_c} \tag{2.2}$$

$$\lambda\tilde{\mathbf{m}}\text{'} = \begin{pmatrix} U \\ V \\ \lambda \end{pmatrix} = \mathbf{P}\text{'} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \mathbf{P}\text{'}\tilde{M}_c, \qquad \text{mit} \qquad x = \frac{U}{\lambda},\ y = \frac{V}{\lambda} \text{ für } \lambda \neq 0 \tag{2.3}$$

$$\text{wobei} \qquad \mathbf{P}\text{'} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.4}$$

Der Parameter f beschreibt den Abstand der Bildebene zum Ursprung des Kamerakoordinatensystems. Für $f = 1$ gelangt man zur Definition der normierten Kamera (Gl. 2.5). Durch $f = 1$ lässt sich die Gl. 2.2 umformen zu Gl. 2.6.

$$\text{wobei} \qquad \mathbf{P}\text{'} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.5}$$

$$x = \frac{x_c}{z_c} \qquad \text{und} \qquad y = \frac{y_c}{z_c} \tag{2.6}$$

Die *interne Transformation* schließt als letzte Transformation den Prozess der optischen Abbildung ab. Dabei werden die metrischen Daten eines Punktes **m'** aus den vorherigen Transformationen in diskrete Pixelwerte übertragen. Die Transformation von den Sensorkoordinaten in die Bildkoordinaten besteht aus einer horizontalen und

[4] Ein Vektor $\mathbf{v} = (v_1, \ ..., \ v_n)^T$ der Dimension n in homogenen Koordinaten $\tilde{\mathbf{v}} = (v_1, \ ..., \ v_n, \ 1)^T$ entspricht einer Dimensionserweiterung, wobei die zusätzliche Komponente zu 1 gesetzt den Skalierungsfaktor darstellt. Somit wird ein Punkt im projektiven Raum $\mathcal{P}^n$ der Dimension n durch einen Vektor **v** mit n+1 Komponenten beschrieben. Bei der homogenen Komponente handelt es sich um einen formalen Kunstgriff, der eine einheitliche Behandlung von Skalierung, Rotation und Translation erlaubt und somit die Zusammmenfassung zu einer Matrix ermöglicht.

vertikalen Skalierung (k_u, k_v). Da eine Bildmatrix in der Regel in einer Ecke des Bildes beginnt, erfolgt außerdem eine Verschiebung (u_0, v_0) des Kamerahauptpunktes (engl. principle point), dem Schnittpunkt $\mathbf{c}$ der optischen Achse mit der Bildebene, in den Ursprung des Bildkoordinatensystems. Der *Skew*-Parameter s beschreibt eine schiefsymmetrische Ausrichtung (Scherung) der Achsen des Bildsensors. Aufgrund der qualitativ hochwertigen Fertigung von CCD-Chips ist dieser Parameter jedoch im Allgemeinen vernachlässigbar. Daher wird im folgenden Umgang mit der Matrix $\mathbf{A}$ ihr Parameter s als 0 betrachtet.

Führt man gleichzeitig zur internen Transformation die perspektivische Transformation aus, und geht man von normierten Bildkoordinaten aus, können die Parameter in der sogenannten *intrinsischen Kamermatrix* $\mathbf{A}$ (Gl. 2.7) zusammengefasst werden.

$$\mathbf{A} = \begin{pmatrix} fk_u & s & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.7}$$

Die Transformation eines Punktes von Sensorkoordinaten in Bildkoordinaten lautet somit $\mathbf{m} = \mathbf{Am'}$. Die vollständige Transformation eines 3D Weltpunktes über die Kamera- und Sensorkoordinaten in seine resultierende Abbildung in der 2D Bildebene $\mathcal{I}$ kann zusammengefasst werden zu Gl. 2.8 und führt zu der allgemeinen Projektionsmatrix $\mathbf{P}$ der perspektivischen Projektion (Gl. 2.9):

$$\lambda\tilde{\mathbf{m}} = \mathbf{A}\ [\mathbf{R}\ \mathbf{t}]\ \tilde{M}_w = \mathbf{P}\ \tilde{M}_w \tag{2.8}$$

$$\text{und somit} \qquad \mathbf{P}_{3\times4} = \mathbf{A}\ [\mathbf{R}\ \mathbf{t}] \tag{2.9}$$

Bis auf den Skalierungsfaktor λ ist nun der Zusammenhang zwischen einem dreidimensionalen Weltpunkt und seiner zweidimensionalen Abbildung hergestellt. In Abbildung 2.2 sind die Relationen der verschiedenen Koordinatensysteme noch einmal veranschaulicht. Durch Eliminierung des Skalierungsfaktors (Division durch die dritte Komponente) erhält man die sogenannte *Kollinearitätsgleichung* (Gl. 2.10) für die Bildkoordinaten u und v des *verzeichnisfreien perspektivischen Kameramodells*, wobei a_{ij} die Koeffizienten in Zeile i und Spalte j der Projektionsmatrix $\mathbf{P}$ (Gl. 2.9)

darstellen.

$$u = \frac{a_{11}x_w + a_{12}y_w + a_{13}z_w + a_{14}}{a_{31}x_w + a_{32}y_w + a_{33}z_w + a_{34}},$$
$$v = \frac{a_{21}x_w + a_{22}y_w + a_{23}z_w + a_{24}}{a_{31}x_w + a_{32}y_w + a_{33}z_w + a_{34}} \tag{2.10}$$

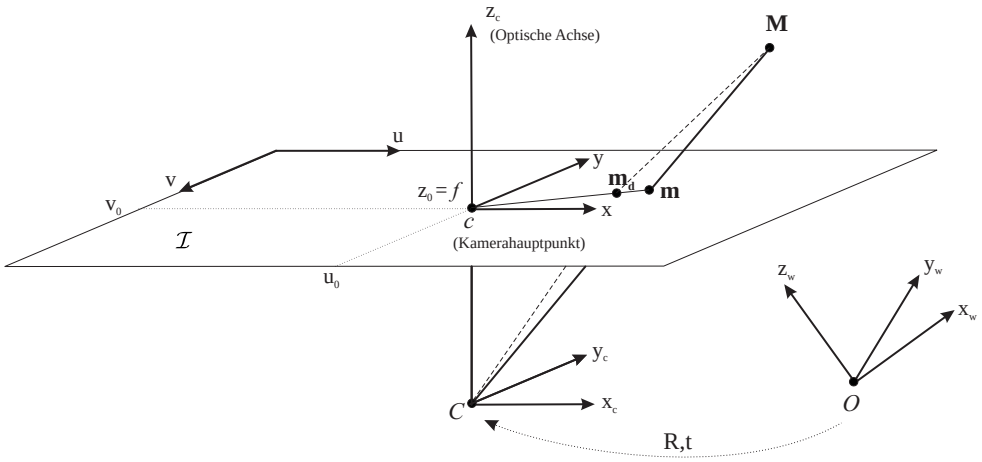


Abbildung 2.2: Die intrinsische und extrinsische Transformation.

2.1.2 Berücksichtigung von Linsenverzeichnung

Zu den in Abschnitt 2.1.1 erwähnten linearen Verzeichnungsparameter der Matrix **A** kommt es bei Kameras mit geringer Brennweite, insbesondere mit zunehmendem radialen Brennpunkt-Abstand, zu nichtlinearen Verzerrungen durch die Krümmung der Linse. Das einfachste und effektivste Modell für solch radiale Verzeichnung ist nach [MSKS04, TV98]:

$$\begin{aligned} u &= u_d(1 + \kappa_1 r^2 + \kappa_2 r^4), \\ v &= v_d(1 + \kappa_1 r^2 + \kappa_2 r^4) \end{aligned} \tag{2.11}$$

mit den Koordinaten (u_d, v_d) für die verzerrten Bildpunkte und $r^2 = u_d^2 + v_d^2$. Die Parameter κ_1, κ_2 geben den Grad der Verzerrung an. Untersuchungen haben gezeigt, dass bei geringer Brennweite vor allem der erste radiale Verzerrungskoeffizient zu

berücksichtigen ist [Sch05, Tsa87] und somit, wie auch in dieser Arbeit, $\kappa_2 = 0$ gesetzt werden kann. In Abbildung 2.3 sind verscheidene Formen der radialen Verzerrung schematisch dargestellt.

In der Literatur werden weitere Verzeichnungen betrachtet, die jedoch in dieser Arbeit nicht näher untersucht werden sollen. Weitere Kameramodelle für die hauptsächlichen Verzerrungsformen wie der radialen Linsenverzeichnung, dezentrierenden Verzeichnung und Verzeichnung des dünnen Prismas sind in [WCH92] aufgeführt. Wobei die beiden letzteren Modelle sowohl radiale als auch tangentiale Verzerrungsanteile aufweisen. Während das Lochkameramodell nur eine Approximation einer realen Kamera darstellt, wäre es mit diesen Erweiterungen möglich, die Projektion handelsüblicher Kameras zu modellieren.

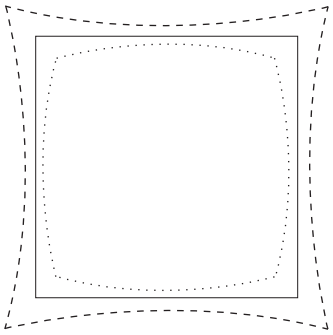


Abbildung 2.3: Schematische Darstellung des Einflusses von radialen Verzerrungen auf das Bild eines Quadrates. Positive oder Kissenverzerrung (gestrichelte Linie) und negative oder Tonnenverzerrung (gepunktete Linie).

2.2 Kamerakalibrierung

Das Lochkameramodell für den Abbildungsprozess von einer dreidimensionalen Szene auf eine zweidimensionale Bildebene enthält 10 unbekannte Parameter in der Projektionsmatrix $\mathbf{P}$ (Gl. 2.9), wenn der Scherungsparameter s, wie in Abschnitt 2.1.1 erwähnt, vernachlässigt wird. Die extrinsische Transformation hat drei Freiheitsgrade für die Translation und drei für die Rotation (die Eulerwinkel yaw Θ, pitch Φ, roll Ψ bestimmen die Rotationsmatrix $\mathbf{R}$ eindeutig). Die intrinsische Transformation trägt weitere zwei Freiheitsgrade für die horizontale und vertikale Skalierung und zwei Freiheitsgrade für die Verschiebung des Bildkoordinatensystems bei. Die Ermittlung dieser Unbekannten nennt sich Kalibrierung. Die Kamerakalibrierung im Kontext dreidimensionaler maschineller Bildverarbeitung ist also die Bestimmung der intrinsischen und/oder extrinsichen Kameraparameter.

Einige Ansätze zur Ermittlung der Parameter gehen von einer bekannten dreidimensionalen Struktur der Szene aus, wie etwa von einen planarem Kalibrationsaufbau [Tsa86, Zha98]. Andere Autoren versuchen durch einschränkendes Wissen über die Kamera, wie etwa konstante intrinsische Parameter [HÅ96], oder zeitlich variierende und unbekannte intrinsische Parameter [HÅ97, PKG98], eine Kalibrierung durchzuführen.

In dieser Arbeit wird die Kalibrierung nach [Tsa87] genutzt, um die intrinsischen Parameter der Projektionsmatrix vorab zu bestimmen. Für die Berechnung des Lochkameramodells nach Tsai kam ein Softwarepaket der Carnegie Mellon School of Computer Science zum Einsatz[5]. Tsai hat eine akkurate zweistufige Methode zur Berechnung der Kameraparameter entwickelt. Die Methode besteht aus einer linearen und einer nichtlinearen Berechnungsstufe. Über das Vorwissen der Lage des Kamerahauptpunktes in der Sensorebene hinaus, benötigt Tsai's Ansatz eine oder mehrere Ebenen mit bekannten Kontrollpunkten – einem sogenannten Kalibrationskörper. Der in dieser Arbeit genutzte Kalibrationskörper ist in der Abbildung 2.4 zu sehen. An den Rändern des Bildes ist der Einfluss der radialen Verzeichnung deutlich an den gekrümmten Kanten des Kalibrationsaufbaus ersichtlich.

Da Tsai, wie im vorherigen Abschnitt erwähnt, den tangentialen Anteil der Verzeichnung vernachlässigt, gelangt man zu der Betrachtungsweise der rein radialen Verzeichnung, wie schon in Gl. 2.11 erwähnt. Tsai berechnet in seinem Algorithmus nur den ersten radialen Verzeichnungskoeffizienten κ_1. In den zwei Stufen dieses Verfahrens werden folgende die Parameter wie folgt berechnet:

1. *Stufe:* Lineare Schätzung der extrinsischen Parameter Θ, Φ, Ψ der Rotationsmatrix $\mathbf{R}$, sowie t_x, t_y Komponente des Translationsvektors $\mathbf{t}$ und eine erste Schätzung der Fokallänge f.

2. *Stufe:* Iteratives Schema zur Näherung der Parameterwerte für t_x, κ_1 und der effektiven Fokallänge f.

Für den ersten Schritt macht Tsai sich das *radial alignment constraint* (RAC)[6] zu Nutzen. Dieses erhält man, wenn neben der radialen Verzeichnung keine anderen Verzerrungen auftreten. Das RAC bietet eine Entkopplung der Parameter und ermöglicht eine rein lineare Lösung für die Komponeten t_x, t_y und aller Komponenten

[5] Erhältlich unter: http://www.cs.cmu.edu/~rgw/TsaiCode.html (letzter Aufruf: 12.01.2006).

[6] Das RAC beschreibt die Tatsache, dass der Richtungsvektor von jedem Punkt der optischen Achse zum Objektpunkt M, dem Vektor vom Kamerahauptpunkt zum verzerrten Bildpunkt $\mathbf{m}_d$, radial ausgerichtet ist [TL87]. Vergleiche Abbildung 2.2.

der Rotationsmatrix **R** durch deren lineare Abhängigkeit. Diese Parameter weisen jedoch Unabhängigkeit gegenüber κ_1, f und t_z auf.

In [LT87] haben Lenz und Tsai diesen Kalibrations-Ansatz erneut untersucht und in der zweiten Berechnungsstufe eine Bestimmung der zwei Komponenten u_0 und v_0 des Kamerahauptpunktes integriert. Lenz optimierte in [Len87] nochmals die zweite Stufe mit einer nicht-iterativen Methode.

Nachteile des Tsai Algorithums sind: Bei geringer oder nicht vorhandener radialer Verzeichnung bietet der Algorithmus keine vollständige Lösung. Das RAC ist nur gültig, wenn der Abbildungsprozess deutliche radiale Verzerrungen aufweist. Des weiteren muss sich die Kamera bei der Aufnahme des Referenzbildes nahe dem Kalibrationskörper befinden und möglichst nicht genau senkrecht zu der Kalibrationsebene ausgerichtet sein [MMW04].

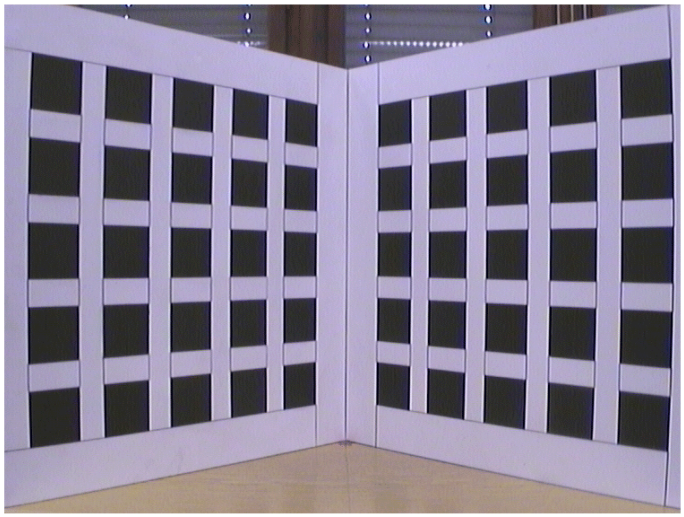


Abbildung 2.4: Kalibrationskörper mit deutlich sichtbaren radialen Verzerrungen (Tonnenverzerrung), insbesondere an den Bildrändern.

Der Algorithmus von Tsai zählt zu den klassischen Kalibrationsverfahren. In [Arm96, Pol99, Zha98] werden Selbstkalibrationsverfahren vorgestellt, die anhand von Bilddaten der realen Welt die intrinsischen und extrinsischen Kameraparameter bestimmen und somit nicht von „teurem Equipment“ [Zha98], insbesondere einem Kalibrationskörper, abhängig sind. Eine gute Übersicht einiger Verfahren der Selbstkalibration findet sich in [Fus00].

2.3 Zusammenfassung

In diesem Kapitel wurde das Lochkameramodell vorgestellt und dessen Abbildungseigenschaft erläutert. Der Abbildungsprozess einer 3D-Szene auf eine 2D-Bildebene wurde durch ein mathematisches Modell beschrieben. Basierend auf diesem Modell ist unter Berücksichtigung von Verzeichnungen eine Kalibrierung der Kamera ermöglicht worden. Durch das vorliegende Kameramodell können in den nächsten Kapiteln genauere Betrachtungen hinsichtlich der geometrischen Beziehung zwischen zwei Kameras angestellt werden.

3 Geometrie zweier Ansichten

In Kapitel 2 wurde ausführlich auf die Abbildungseigenschaften einer Kamera eingegangen. Es wurde deutlich, das Geraden des Raumes durch den Brennpunkt der Kamera auf Punkte in der Ebene abgebildet werden. Es ist umgekehrt nicht möglich, die Position eines Objektes aus den Pixeln eines Bildes zu bestimmen. Um räumliche Informationen aus digitalen Bildern zu gewinnen, werden deshalb mindestens zwei Kameras benötig. Dieses Kapitel beschäftigt sich mit der Erweiterung durch eine zusätzliche, zweite Kamera zu einem Stereoansatz. Nach der Aufnahme eines Stereobildpaares werden korrespondierende Punkte gesucht, d.h. Punkte bei denen im linken und rechten Bild der gleiche Weltpunkt abgebildet wird. Es müsste also für jeden Bildpunkt im linken Bild das komplette rechte Bild nach dem korrespondierenden Punkt durchsucht werden.

In diesem Kapitel wird folgende grundlegende Fragestellung genauer betrachtet. *Sei* $\boldsymbol{m}_1$ *ein Bildpunkt einer perspektivischen Abbildung des 3D-Raumpunktes M auf eine Bildebene* $\mathcal{I}_1$*. Können darauf aufbauend Einschränkungen der Positionen des korrespondierenden Bildpunktes* $\boldsymbol{m}_2$ *in einer weiteren Bildebene* $\mathcal{I}_2$ *gemacht werden?*

Es soll gezeigt werden, dass Einschränkungen aus der Kalibrierung und bekannter Lage zweier Kameras zueinander formuliert werden können. In Abschnitt 3.1 wird gezeigt, dass die Suche über den ganzen Bildraum auf eine Suche über eine Linie eingeschränkt werden kann. Durch eine Transformation der Bilder kann zudem erreicht werden, dass diese Linien mit den Zeilen der transformierten Bilder zusammenfallen und folglich eine horizontale Suche ermöglicht wird. Diese Transformation, auch Rektifikation genannt, wird in Abschnitt 3.3 erläutert.

3.1 Epipolargeometrie

Die Epipolargeometrie behandelt einen Teilbereich der projektiven Geometrie. Sie wird auch als allgemeine Stereogeometrie bezeichnet. Die Epipolargeometrie deckt im wesentlichen alle Projektionen und Operationen im zwei- bzw. dreidimensionalen euklidischen Raum ab. Sie beschreibt vollständig die geometrischen Informationen korrespondierender Punkte zwischen zwei perspektivischen Bildern zueinander und kann

durch die 3×3 Fundamentalmatrix $\mathbf{F}$, oder bei bekannten intrinsischen Kameraparametern durch die 3×3 Essentialmatrix $\mathbf{E}$, ausgedrückt werden [HZ03, Fau93, Zha96]. Die Fundamentalmatrix $\mathbf{F}$ beschreibt die geometrische Beziehung in Pixelkoordinaten, während die Essentialmatrix $\mathbf{E}$ die geometrische Beziehung in Kamerakoordinaten wiederspiegelt.

Seien nun zwei Kameras mit ihren optischen Zentren C_1, C_2 auf den gleichen Raumpunkt M ausgerichtet wie in Abbildung 3.1 dargestellt. Die Verbindungslinie zwischen den beiden optischen Zentren C_1, C_2 wird Basislinie B genannt. Aufgrund der leicht zueinander gedrehten Bildebenen $\mathcal{I}_1$ und $\mathcal{I}_2$ schneidet die Basislinie beide Bildebenen. Die Schnittpunkte der Basisline mit den Bildebenen werden *Epipole* genannt und mit e_1 und e_2 gekennzeichnet. Anders formuliert stellen die Epipole die Projektion der optischen Zentren in die jeweils andere Bildebene dar. Ihre Position in den Bildebenen hängt ausschließlich von der Anordnung der Kameras zueinander ab. Sie können, müssen sich aber nicht in den jeweils aufgenommenen Bildern befinden. Die beiden optischen Zentren C_1, C_2 spannen gemeinsam mit dem 3D-Raumpunkt M eine Ebene auf, die als *Epipolarebene* π bezeichnet wird[1]. Es ist offensichtlich, dass auch die Punkte $\mathbf{m}_1$ und $\mathbf{m}_2$ auf dieser Ebene π liegen, da die optischen Strahlen von M zu C_1 und M zu C_2 Geraden auf der Epipolarebene sind.

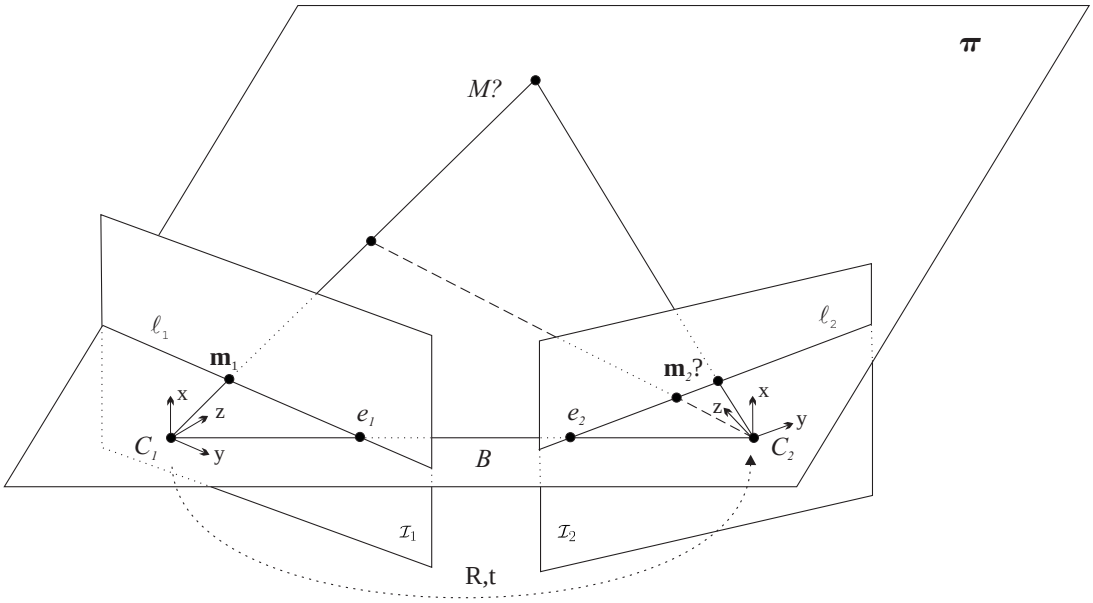


Abbildung 3.1: Schematische Darstellung der Epipolargeometrie.

Der 3D-Raumpunkt M liegt auf der Geraden durch das Projektionszentrum C und

[1] Die Epipolarebene kann ebenfalls durch die beiden Epipole e_1, e_2 und dem 3D-Raumpunkt M aufgespannt werden.

seinem Bildpunkt $\mathbf{m}$. Diese Gerade, in das andere Bild projeziert, kennzeichnet die Epipolarlinie ℓ_i und entspricht genau der Schnittgeraden der Bildebenen mit der Epipolarebene π. Da die Bildpunkte $\mathbf{m}_1, \mathbf{m}_2$ koplanar der Epipolarebene π sind, müssen Sie sich folglich auf den Epipolarlinien ℓ_1, ℓ_2 befinden. Alle Abbildungen $\mathbf{m}_{1\pi}, \mathbf{m}_{2\pi}$ der zu π koplanaren Raumpunkte M_π müssen auf den Epipolarlinien $\ell_{1\pi}, \ell_{2\pi}$ liegen.

Rotiert man die Epipolarebene um die Basislinie, so kann die Gesamtheit aller 3D-Punkte im Raum erfasst werden. Aus jeder neuen Epipolarebene resultieren jeweils neue Schnittgeraden mit den Bildebenen und somit neue Epipolarlinien. Gemein haben diese Epipolarlinien, dass sie sich in den jeweiligen Epipolen ihrer Bildebene schneiden. Die Gesamtheit der Epipolarlinien in einer Bildebene wird *Epipolarlinienbüschel* (engl. *pencil of epipolarlines*) genannt. Als wesentliche Eigenschaft erhält man somit folgende wichtige Beziehung. Wird ein 3D-Punkt M in einem Bild an einer bestimmten Bildposition abgebildet, so beschränkt sich die Suche des korrespondierenden Punktes $\mathbf{m}_2$ auf die Epipolarlinie ℓ_2, statt auf das gesamte Bild. Das 2D-Korrespondenzproblem wird vereinfacht zu einem 1D-Korrespondenzproblem entlang der Epipolarlinien ausgehend vom Epipol. Diese Vorschrift wird Epipolarbedingung genannt und im nächsten Unterabschnitt mathematisch durch die Epipolargleichung erläutert.

Zuvor sei noch die Bestimmung einer Linie ℓ im Bild $\mathcal{I}$ durch den Punkt $\mathbf{m} = (u, v)^T$ durch die Geradengleichung $au + bv + c = 0$ beschrieben. Mit $\ell = (a, b, c)^T$ kann die Geradegleichung umgeschrieben werden zu $\ell^T \tilde{\mathbf{m}} = 0$, beziehungsweise $\tilde{\mathbf{m}}^T \ell = 0$.

3.1.1 Eigenschaften der Essential- und Fundamentalmatrix

In der allgemeinen Stereogeometrie sind die beiden Kameras nicht nur verschoben, sondern auch zueinander verdreht. Diese Positions- und Orientierungsänderung kann durch eine starre Transformation ausgedrückt werden. Sie beschreibt eine Überführung des Raumpunktes M hinsichtlich des Kamerakoordinatensystems $(x,\ y,\ z)^T_{c_1}$ der ersten Kamera in das zweite Kamerakoordinatensystem $(x,\ y,\ z)^T_{c_2}$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_{c_2} = \mathbf{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix}_{c_1} + \mathbf{t} \tag{3.1}$$

Hierbei stellt $\mathbf{R}$ eine orthogonale Drehmatrix dar. Für den dreidimensionalen Verschiebungsvektor $\mathbf{t}$ der Kameras gilt $\mathbf{t} = C_1 - C_2$. Diese Transformation ist ähnlich der Gl. 2.1 in Abschnitt 2.1. In diesem Fall wird jedoch das Kamerakoordinatensystem

der zweiten Kamera in das Kamerakoordinatensystem der ersten Kamera verschoben und ausgerichtet.

In dieser Arbeit werden die Bilder nicht von zwei, sondern von einer Kamera zu verschiedenen Zeitpunkten mit unterschiedlicher Orientierung und Position akquiriert. Dieses Vorgehen nennt man allgemein Struktur aus Bewegung (engl. *structure-from-motion*). Im weiteren Verlauf dieser Arbeit seien C_1 und C_2 nicht als Projektionszentren zweier konvergenter Kameras anzusehen, sondern als Kamerakoordinatensystem derselben Kamera an der ersten und zweiten Betrachtungsposition zu verstehen. Wenn in im Folgenden von zwei Kameras gesprochen wird, sind die beiden verschiedenen Lagen der einen Kamera gemeint. Somit gilt im weiteren Verlauf dieser Arbeit $\mathbf{A} = \mathbf{A}_1 = \mathbf{A}_2$ (vgl. Gl. 2.7).

Setzt man nun in Gl. 3.1 die projezierten Punkte $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2$ in homogenen Koordinaten ein, gelangt man zur zentralen Gleichung der Epipolargeometrie, der Epipolargleichung (Gl. 3.2).

Epipolargleichung: Seien $\tilde{\mathbf{m}}'_1, \tilde{\mathbf{m}}'_2$ zwei Punkte des gleichen Raumpunktes M in jeweiligen Kamerakoordinaten mit relativem Versatz ($\mathbf{R}$,$\mathbf{t}$), dann erfüllen $\tilde{\mathbf{m}}'_1, \tilde{\mathbf{m}}'_2$ [MSKS04]:

$$\tilde{\mathbf{m}}'^{\top}_2 \mathbf{E} \tilde{\mathbf{m}}'_1 = 0 \qquad \text{mit} \qquad \mathbf{E} = (\mathbf{t})_\times \mathbf{R} \tag{3.2}$$

Durch $(\mathbf{t})_\times$ ist die antisymmetrische Kreuzproduktmatrix mit $(\mathbf{t})_\times \mathbf{x} = \mathbf{t} \times \mathbf{x}$ für alle dreidimensionalen Vektoren $\mathbf{x}$ gegeben (vgl. Gl. 3.3).

$$(\mathbf{t})_\times = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}_\times = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} = -(\mathbf{t})_\times^\top \tag{3.3}$$

Die 3×3 Matrix $\mathbf{E}$ wird *Essentialmatrix* genannt und wurde erstmals von Longout-Higgins [Lon81] vorgestellt. Die Essentialmatrix beschreibt die geometrische Beziehung zweier korrespondierender Punkte in den beiden Ansichten des Stereokamerasystems in Sensorkoordinaten.

Da die Determinante[2] des Vetrschiebungsvektors $det(\mathbf{t}) = 0$, ist aufgrund des Pro-

[2] Die Determinante ist eine spezielle Funktion, die einer quadratischen Matrix $\mathbf{A}_{n \times n}$ einen reellen Zahlenwert zuordnet. Die Determinante gibt Auskunft, ob ein lineares Gleichungssystem eindeutig lösbar ist. Bei einer $det(\mathbf{A}) = 0$ ist ein LGS nicht eindeutig lösbar und die Matrix $\mathbf{A}$ ist singulär. Der Rang der Matrix beträgt $Rg(\mathbf{A}) < n$. Ist $det(\mathbf{A}) \neq 0$, so ist $\mathbf{A}$ regulär und die n Spalten (n Zeilen) sind linear unabhängig. Der Rang der Matrix lautet $Rg(\mathbf{A}) = n$. Das LGS ist lösbar.

duktsatzes[3] für Determinanten $det(\mathbf{E}) = det(\mathbf{t}_\times)\ det(R) = 0$. Somit enthält die Essentialmatrix nur zwei linear unabhängige Zeilen- oder Spaltenvektoren und hat den Rang[4] $Rg(\mathbf{E}) = 2$.

Für den Epipol ergeben sich aus der Essentialmatrix folgende Eigenschaften. Die Epipole e_1, e_2 bezüglich der linken und rechten Kamera können aus dem links- und rechtsseitigen Nullraum der Essentialmatrix bestimmt werden [MSKS04].

$$\begin{array}{lllll} & \tilde{e'}_2^T\mathbf{E} = 0 & \text{und} & \mathbf{E}\tilde{e}'_1 = 0 \\ \text{da} & \tilde{e}'_1 = \mathbf{R}^T\mathbf{t} & \text{und} & \tilde{e}'_2 = \mathbf{t} \end{array} \tag{3.4}$$

Der Zusammenhang zwischen normierten Kamerakoordinten und Pixelkoordinaten wird durch die intrinsische Transformation (vgl. Abschnitt 2.1.1 und Gl. 2.7) mit der Matrix $\mathbf{A}$ hergestellt und lautet für beide Kameras des Stereosystems:

$$\tilde{\mathbf{m}}_1 = \mathbf{A}_1\tilde{\mathbf{m}}\text{'}_2 \qquad \text{und} \qquad \tilde{\mathbf{m}}_2 = \mathbf{A}_2\tilde{\mathbf{m}}\text{'}_2 \tag{3.5}$$

Setzt man nun Gl. 3.5 in Gl. 3.2 ein, so ergibt sich die geometrische Beziehung in Pixelkoordinaten durch die *Fundamentalmatrix* $\mathbf{F}$.

$$\tilde{\mathbf{m}}_2^\top \mathbf{A}_2^{-T}\mathbf{E}\mathbf{A}_1^{-1}\tilde{\mathbf{m}}_1 = \tilde{\mathbf{m}}_2^\top \mathbf{F}\tilde{\mathbf{m}}_1 = 0 \qquad \text{mit} \qquad \mathbf{F} = \mathbf{A}_2^{-T}\mathbf{E}\mathbf{A}_1^{-1} \tag{3.6}$$

Sie enthält die intrinsischen und extrinsischen Parameter der euklidischen Transformationen beider Kameras. Da die Determinante der Essentialmatrix $det(\mathbf{E}) = 0$ ist, gilt auch für die Fundamentalmatrix durch den Produktsatz $det(\mathbf{F}) = 0$. Somit hat die Fundamentalmatrix ebenfalls Rang $Rg(\mathbf{F}) = 2$. Analog zu der Betrachtung in Kamerakoordinaten gelten folgende Beziehungen für die Fundamentalmatrix und die Epipole in Pixelkoordinaten (siehe Gl. 3.7).

$$\tilde{e}_2^T\mathbf{F} = 0 \qquad \text{und} \qquad \mathbf{F}\tilde{e}_1 = 0 \tag{3.7}$$

Die Epipolarlinien können durch die Punkte in Pixelkoordinaten wie in Gl. 3.8 bis auf einen Skalierungsfaktor berechnet werden. Dabei drücken die Epipolarlinien die

[3] Der Produktsatz für Determinanten besagt: $det((\alpha_{ik}) \cdot (\beta_{ik})) = det(\alpha_{ik}.) \cdot det(\beta_{ik})$

[4] Der Rang einer Matrix $\mathbf{A}$ mit $\mathbf{A} \neq 0$ ist die maximale Anzahl linear unabhängiger Spaltenvektoren. Homogene lineare Gleichungssysteme $\mathbf{A}x = 0$ haben bei Übereinstimmung von Rang $Rg(\mathbf{A})$ und Anzahl der Variablen nur die triviale Lösung $x = 0$. Ist die Anzahl der der Variablen größer als der Rang $Rg(\mathbf{A})$, dann besteht lineare Abhängigkeit zwischen den Zeilenvektoren der Koeffizientenmatrix $\mathbf{A}$.

Spannvektoren der Epipolarebene π aus.

$$\begin{aligned} \ell_1 &= \mathbf{F}^T \tilde{\mathbf{m}}_2, \\ \ell_2 &= \mathbf{F}\tilde{\mathbf{m}}_1 \end{aligned} \tag{3.8}$$

Aus der Epipolargleichung ergibt sich unter Verwendung der Gleichung für die Epipolarlinien (Gl. 3.8), dass für jeden Punkt einer Ansicht der korrespondierende Punkt in der anderen Ansicht auf der entsprechenden Epipolarlinie liegt. Ebenso gilt das für die Epipole der jeweiligen Ansicht. Daraus folgt:

$$\ell_i^T \tilde{e}_i = 0 \qquad \text{und} \qquad \ell_i^T \tilde{\mathbf{m}}_i = 0 \qquad \text{für} \qquad i \in \{1, 2\} \tag{3.9}$$

Gl. 3.6 und Gl. 3.9 stellen die wohl wichtigste Eigenschaft der Fundamentalmatrix für den Abschnitt 3.3 dar.

3.1.2 Berechnung der Essential- und Fundamentalmatrix

In den Abschnitten zuvor wurde erläutert, dass korrespondierende Punkte zweier Bilder über die Epipolargleichung in Relation gesetzt werden können. Daher kann bei einer gegebenen Anzahl n von korrespondierenden Punkten über die Epipolargleichung die Kameraposition ermittelt werden. Relevante Punkte, deren Korrespondenz es zu beweisen gilt, können vorab beispielsweise durch Merkmalsdetektoren wie dem Harris-Ecken-Detektor ermittelt werden (vgl. Anhang B). Die in der Fachliteratur weit verbreiteten Verfahren zur Ermittlung der Essential- und Fundamentalmatrix durch lineare und nichtlineare Ansätze sollen als aktueller Stand der Technik erläutert werden.

Lineare Verfahren, wie in Abschnitt 3.1.3 angeführt, liefern zufriedenstellende Ergebnisse, wenn die Punktkorrespondenzen hinreichend genau sind. Es zeigt sich jedoch in der Literatur, dass durch Verwendung von klassischen nichtlinearen Verfahren (Abschnitt 3.1.4) aus der numerischen Mathematik bessere Ergebnisse zu erzielen sind. Nichtlineare Standardverfahren seien etwa das *Gauß-Newton-Verfahren* oder die *Levenberg-Marquardt-Optimierung* [Sch05].

3.1.3 Lineare Berechnungsverfahren

8-Punkt-Algorithmus In der Menge der Ansätze zur Berechnung der Essential- und Fundamentalmatrix ist der lineare 8-Punkt-Algorithmus bei weitem der einfachste.

Er wurde erstmals in [Lon81] zur Schätzung der Essentialmatrix vorgestellt. Dies ist historisch gesehen das erste Verfahren, das die Schätzung der Fundamentalmatrix beschreibt.

Die Idee des 8-Punkt-Algorithmus ist sehr einfach. Seien n Punktkorrespondenzen zwischen den Bildern gegeben. Jedes Paar korrespondierender Punkte $\tilde{\mathbf{m}}_{n1} \leftrightarrow \tilde{\mathbf{m}}_{n2}$, eingesetzt in Gl. 3.6, führt zu einer linearen Gleichung mit einem der neun Koeffizienten der Fundamentalmatrix. Die Gleichung für ein Punktepaar $\tilde{\mathbf{m}}_1 = (u, v, 1)^T$ und $\tilde{\mathbf{m}}_2 = (o, p, 1)^T$ in homogenen Pixelkoordinaten lautet:

$$ouf_{11} + ovf_{12} + of_{13} + puf_{21} + pvf_{22} + pf_{23} + uf_{31} + vf_{32} + f_{33} = 0 \tag{3.10}$$

Durch n korrespondierende Punktpaare entsteht ein lineares Gleichungssystem wie in Gl. 3.11. Dabei stellt $\mathbf{f}$ die Vektorform der neun Fundamentalmatrix-Koeffizienten dar.

$$\mathbf{A}_n\mathbf{f} = \begin{pmatrix} o_1u_1 & o_1v_1 & o_1 & p_1u_1 & p_1v_1 & p_1 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_nu_n & o_nv_n & o_n & p_nu_n & p_nv_n & p_n & u_n & v_n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ \vdots \\ f_{33} \end{pmatrix} = 0 \tag{3.11}$$

Seien nun mindestens $n \geq 8$ Punktkorrespondenzen gegeben und bilden die Punktepaare keine instabilen Anordnungen, lassen sich die neun Komponenten der Fundamentalmatrix durch Lösen des Gleichungssystems bestimmen. Die Aufgabe lässt sich dann als Minimierungsproblem der quadratischen Fehlerfunktion $Q(\mathbf{F})$ formulieren:

$$\min Q(\mathbf{F}) = \min \sum_n \| \tilde{\mathbf{m}}_{n2}^T \mathbf{F} \tilde{\mathbf{m}}_{n1} \|^2 \qquad \text{wobei} \qquad \tilde{\mathbf{m}}_{n2}^T \mathbf{F} \tilde{\mathbf{m}}_{n1} = \mathbf{a}_n\mathbf{f} \tag{3.12}$$

Das Gleichungssystem liefert mehrdeutige Ergebnisse, wenn der Rang $Rg(\mathbf{A}) < 8$ trotz $n > 8$ Punktpaaren. Dieser Fall tritt ein, wenn alle Punktpaare beispielsweise auf einer Ebene liegen. Handelt es sich bei der Matrix $\mathbf{A}$ in Gl. 3.11 um eine homogene Matrix mit Rang $Rg(\mathbf{A}) = 8$, so ist die Lösung bis auf einen Skalierungsfaktor eindeutig. Um die triviale Lösung $\mathbf{f} = 0$ auszuschließen, bietet sich als zusätzliche Bedingung $||\mathbf{f}|| = 1$, die Norm des Vektors $\mathbf{f}$, an.

Sollten $n > 8$ Punktkorrespondezen vorliegen und die Matrix $\mathbf{A}$ Rang $Rg(\mathbf{A}) > 8$ besitzen, so ist das Gleichungssystem überbestimmt. Ein Hilfsmittel zur Berechnung von überbestimmten Gleichungssystem ist die Singulärwertzerlegung (engl. *singular*

value decomposition, SVD)[5]. Sei $\mathbf{A}$ die Ausgangsmatrix, so kann sie durch die SVD zerlegt werden in $\mathbf{A} = \mathbf{UDV}^\top$. Die Spalte von $\mathbf{V}$, die mit dem einzigen Singulärwert 0 der Matrix $\mathbf{A}$ übereinstimmt, stellt die Lösung dar. Aufgrund von Rauschen der Bilddaten ist dieser Satz zu korrigieren, denn die Matrix $\mathbf{A}$ wird sicherlich nichtsingulär sein. Folglich ist die Lösung dann die Spalte von $\mathbf{V}$, die dem kleinsten Singulärwert von $\mathbf{A}$ entspricht.

Hartley widerlegt in [Har97a] die allgemeine Ansicht, dass der Algorithmus durch die angeblich starke Anfälligkeit gegenüber Rauschen für die meisten Anwendungen den nichtlinearen Verfahren weit unterlegen sei. Seine Optimierung besteht aus einer einfachen Normalisierung (Translation und Skalierung) der Bildpunktkorrespondenzen in einem unerheblichen Vorverarbeitungschritt des 8-Punkt-Algorithmus [Har97a].

Die geschilderte Mindestzahl von acht Punktkorrespondenzen trifft nur aus Bequemlichkeit zu. Nach Ma et al. [MSKS04] hat die Essentialmatrix $\mathbf{E}$, als Funktion von ($\mathbf{R}$, $\mathbf{t}$), eigentlich nur fünf Freiheitsgrade. Drei für die Rotation, und bis auf einen Skalierungsfaktor zwei für die Translation. Dies wurde schon 1913 von Kruppa festgestellt. Die Herleitung von Kruppa's Gleichungen aus der Fundamentalmatrix sind in [Har97b] zusammengefasst. Unter Nutzung einiger algebraischer Eigenschaften kann bei bekannter Determinante $det(\mathbf{E}) = 0$ statt Rang $Rg(A) = 8$ Rang $Rg(A) = 7$ angenommen werden. Werden noch weitere, komplexere Eigenschaften betrachtet, so kann die Essentialmatrix durch 6 Punktkorrespondenzen berechnet werden, oder gar durch 5 Punktpaare bis auf 10 komplexe Lösungen ermittelt werden. Desweiteren können ebene, oder symmetrische Bewegungen die Anzahl der benötigten Punktkorrespondenzen auf 4 beschränken (vgl. [MSKS04, NS04]).

3.1.4 Nichtlineare Berechnungsverfahren

Neben dem vorgestellten linearen Lösungsansatz existieren noch eine Reihe nichtlinearer Methoden, die meist eine deutlich höhere Anzahl an Punktpaaren fordern. Es handelt sich um iterative Verfahren, die durch Merkmalsextraktion Korrespondenzpunktepaare erstellen. Allerdings kommt es bei einer automatischen Korrespondenzpaar-Bestimmung zu einer neuen Klasse von Fehlern, den sogenannten Ausreißern (engl. *outlier*). Sie entstehen durch schlechte Lokalisierung oder falsche Punktkorrespondenzen. Die bisherigen Verfahren sind besonders gegenüber den falschen Korrespondenzen sehr empfindlich. Das folgende Verfahren kann falsche Punktpaare herausfiltern und stellt somit für eine Automatisierung einen hohen Nutzen dar.

Für die nichtlinearen Verfahren ist es im Allgemeinen unerlässlich, einen geeigneten

[5] Für weitere Details zur Singulärwertzerlegung sei der Leser auf Anhang A referenziert.

Startwert für die Parameter der Fundamentalmatrix zu wählen. Im Allgemeinen wird deshalb zuerst mit Hilfe eines linearen Verfahrens eine erste Schätzung der Fundamentalmatrix bestimmt, die dann mit den nichtlinearen Verfahren optimiert wird.

Iterative Verfahren liefern zwar gute bis sehr gute Ergebnisse, sind jedoch erheblich aufwendiger zu implementieren und berechnungsintensiver als der lineare Ansatz.

Betrachtet man das lineare Gütekriterium aus Gl. 3.12 geometrisch, so entspricht es dem Abstand des Messpunktes $\mathbf{m}_n$ zur Epipolarlinie ℓ_n. Möchte man dieses Abstandskriterium einfließen lassen, ist ein iteratives Vorgehen erforderlich. Da sich der Abstand des Messpunktes zur Epipolarlinie jedoch nur bei Kenntniss der Fundamentalmatrix ermitteln lässt, wird über Gl. 3.13 mit einer anfänglichen Gewichtung 1 für alle n Punktkorrespondenzen eine erste Fundamentalmatrix nach dem bereits beschriebenen 8-Punkt-Algorithmus berechnet. Die resultierende approximierte Fundamentalmatrix wird für eine Berechnung der Gewichtungsfaktoren c_n genutzt und mit dieser Gewichtung ist wiederum eine neue Berechnung der Fundamentalmatrix möglich. Bei den Gewichtungsfaktoren stellt l_{nij} die $i-te$ Komponenten des Epipolarlinienvektors im $j-ten$ Bild des $n-ten$ Bildpunktes dar. Dieser Prozess kann wiederholt werden, solange sich die Schätzung optimiert. Mit diesem Verfahren wird eine geringere Beeinflussung durch ungenaue Punktkorrespondenzen erzielt.

$$\min_{\mathbf{F}} \sum_n c_n^2 (\tilde{\mathbf{m}}_{n2}^T \mathbf{F}\ \mathbf{m}_{n1})^2, \qquad \text{mit} \qquad c_n = \left(\frac{1}{l_{n11}^2 + l_{n21}^2} + \frac{1}{l_{n12}^2 + l_{n22}^2}\right) \tag{3.13}$$

Least-Median-of-Squares (LMedS) Bei dieser Schätzmethode wird versucht den Median der Summe der Residuen r zu minimieren:

$$LMS = \min \underbrace{median}_{i} r_i^2 \tag{3.14}$$

Die LMedS Methode bestimmt die Parameter durch Gleichung 3.14, die ein nichtlineare Minimierungsfunktion darstellt. Als Residuum könnte etwa das Abstandsmaß der Punkte von den Epipolarlinien, $d^2(\tilde{\mathbf{m}}_2, \mathbf{F}\tilde{\mathbf{m}}_1) + d^2(\tilde{\mathbf{m}}_1, \mathbf{F}^T\tilde{\mathbf{m}}_2)$, genutzt werden.

Es handelt sich hierbei also um eine Schätzung in einem Suchraum der Größe aller mögliche Punktkorrespondenzen. Da dieser Suchraum zu groß ist, werden Untermengen ausgewählt. P. Rousseeuw und A. Leroy schlugen in [RL87] folgenden Algorithmus zur Berechnung der Fundamentalmatrix vor:

Seien n korrespondierende Punktpaare $\tilde{\mathbf{m}}_{n1} \leftrightarrow \tilde{\mathbf{m}}_{n2}$ mit $n = \{1, 2, \ldots, n\}$ gegeben:

1. Eine Monte-Carlo-Methode[6] wird benutzt, um m Tupel mit 8 Korrespondenzen auszuwählen[7].

2. Für jede Stichprobe J wird der in Abschnitt 3.1.3 beschriebene 8-Punkt-Algorithmus zur Berechnung der Fundamentalmatrix verwendet.

3. Für jedes $\mathbf{F}_J$ kann der Median der Residuen-Quadrate r_i^2, bezeichnet durch M_J , in Bezug auf den gesamten Datensatz bestimmt werden, d.h.:

$$M_J = \underbrace{median}_{i=1,\ldots,n}(d^2(\tilde{\mathbf{m}}_{n2}, \mathbf{F}_J\tilde{\mathbf{m}}_{n1}) + d^2(\tilde{\mathbf{m}}_{n1}, \mathbf{F}_J^T\tilde{\mathbf{m}}_{n2})) \tag{3.15}$$

4. Das Ergebnis der Fundmentalmatrix ist jenes $\mathbf{F}_J$, für das M_J minimal über die Menge m aller M_J.

Auf die richtige Wahl der Untermengen, mit „guten“ Punktkorrespondenzen wird hier nicht weiter eingegangen, der Leser sei diesbezüglich auf [Zha96] verwiesen. In [RL87] wird gezeigt, dass die Effizienz der LMedS Methode bei Gauß-Rauschen schlecht ist.

RANSAC Die Idee des RANSAC-Algorithmus (RANdom SAmpling Consensus) ist ebenfalls einfach. Aus einer großen verfügbaren Anzahl von Korrespondenzmessungen wird nur eine Teilmenge, also nur so viele Werte genommen, wie für die Berechnung der Parameter des Modells benötigt. Die restlichen Punktpaare werden dann mit dem erstellten Modell über einen Schwellwert bezüglich ihrer Güte kontrolliert. Die RANSAC Schätzung maximiert die Größe einer Teilmenge der vorhandenen Daten, welche mit dem geschätzten Ergebnis „konsistent“ ist [FB81]. Das zu erfüllende Modell ist auch hier für jedes Punktpaar n die Fundamentalmatrix $\tilde{\mathbf{m}}_{n2}^T\mathbf{F}\tilde{\mathbf{m}}_{n1}$ (vgl. Gl. 3.6). Eine initiale Schätzung der Fundamentalmatrix wird meist durch den linearen 8-Punkt-Algorithmus ermittelt.

Die RANSAC Methode ähnelt somit stark der LMedS Methode, in der Idee wie auch der Implementation. Unterschiede bestehen lediglich darin, dass der Nutzer beim RANSAC einen Schwellwert für den Konsistenztest angeben muss, während dieser bei LMedS selbstständig berechnet wird. Zudem berechnet die RANSAC Methode in Stufe 3 des zuvor beschriebenen LMedS Algorithmus nicht den Median der quadratischen Residuen, sondern die Anzahl der Punktkorrespondenzen, die konsistent zur geschätzten Fundamentalmatrix F_J sind.

[6] Die Monte-Carlo-Methode verwendet Prinzipien der Wahrscheinlichkeitsrechnung und Statistik, um komplexe Probleme zumindest näherungsweise zu lösen. Sie wird deshalb auch als Methode der statistischen Versuche bezeichnet.

[7] Es sei zu beachten, dass mindestens 7 Korrespondenzenpunkte benötigt werden, um die Fundamentalmatrix zu bestimmen.

Die Anzahl der Iterationen hängt von der Menge der Ausreißer ab. Die wesentliche Herausforderung besteht also in der geeigneten Wahl eines Schwellwertes, ab dem die geschätzte Fundamentalmatrix als optimal angesehen wird. Ein Algorithmus, der den Schwellwert adaptiv in jedem Iterationsschritt anpasst, ist in [HZ03] vorgeschlagen. Eine effiziente Kombination des RANSAC-Verfahrens mit einem 5-Punkt-Algorithmus von Nister ist in [Nis04] vorgestellt.

LMedS kann mit $> 50\%$ Ausreißern nicht umgehen, während RANSAC dazu in der Lage ist. RANSAC ist in der Berechnung etwas effizieter, da es die Stichprobenschleife verlassen kann, sobald ein konsistentes Ergebnis relativ zum gegeben Schwellwert gefunden wurde.

In dieser Arbeit kann sich der Nutzer aus dem Methodenschatz zwischen dem linearen 8-Punkt-Algorithmus, oder den nichtlinearen Verfahren RANSAC oder LMedS zur Berechnung der Fundamentalmatrix entscheiden (vgl. Abschnitt 6.4).

3.2 Achsparallele Stereogeometrie

Im Gegensatz zu der im Abschnitt 3.1 angesprochenen allgemeinen Stereogeometrie zeichnet sich ein achsparalleles Stereosystem durch zwei Karemas aus, deren Koordinatensysteme nur horizontal verschoben und nicht gegeneinander verdreht sind. In Abbildung 3.2 ist die Frontsicht auf ein achsparalleles Stereosystem dargestellt. Die beiden Bildebenen $\mathcal{I}_1$ und $\mathcal{I}_2$ sind parallel und die optischen Zentren C_1 und C_2 sind nur horizontal verschoben.

Disparität Das Wort Disparität stammt vom spätlateinischen „Disparitas“ ab. Im Allgemeinen bezeichnet es die Ungleichheit, bzw. Verschiedenheit. Aus der achsparallelen Stereogeometrie lässt sich der Begriff der *Disparität* herleiten.

> „Sei $\mathbf{m}_1$ mit den Koordinaten (u_1, v_1) in der ersten Bildebene $\mathcal{I}_1$ und $\mathbf{m}_2$ mit den Koordinaten (u_2, v_2) in der zweiten Bildebene $\mathcal{I}_1$ zwei korrespondierende Punkte, so ist die *Disparität* die Differenz $\delta = u_1 - u_2$.“ [Fau93, Kapitel 6.2.3.1, S. 175, Übersetzung des Authors]

In der Differenz δ betrachtet Faugeras nur die Disparität von Pixeln der gleichen Höhe in beiden Bilder. Verdeutlicht wird dies durch Abbildung 3.2. Er geht also von dem selben v-Wert der Pixel aus und bezieht sich somit auf rektifizierte Bilder, in denen diese Bedingung gegeben ist. Wie Bilder rektifiziert werden, soll in den nachfolgenden Abschnitten erläutert werden.

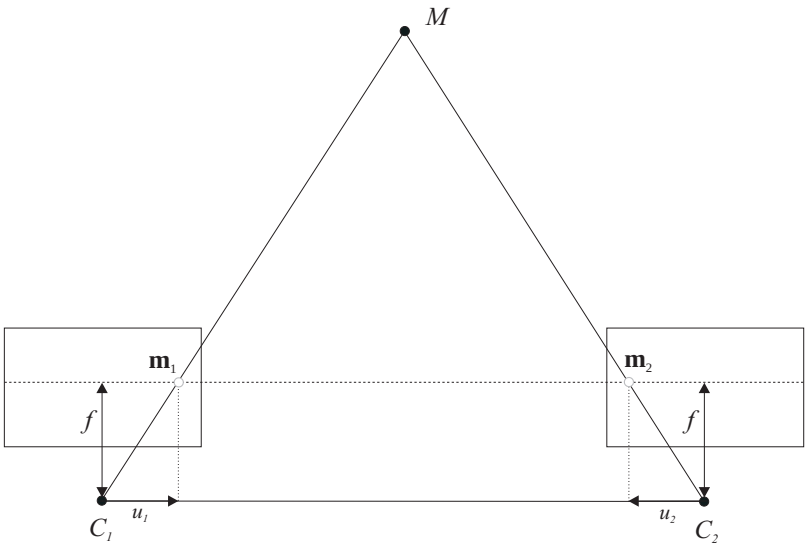


Abbildung 3.2: Frontansicht einer achsparallelen Stereogeometrie mit den jeweiligen Disparitäten u_1, u_2.

Eine Disparität von 0 impliziert, dass sich der abgebildete 3D-Raumpunkt in unendlicher Entfernung befindet. Die Disparität ist umgekehrt proportional zur Tiefe des Raumpunktes [TV98]. Disparitäten werden im Allgemeinen in Pixelkoordinaten berechnet und durch den *euklidischen Abstand* $d_k(m_1, m_2)$ für $k = 2$ angegeben[8].

$$d_k(m_1, m_2) = ((u1 - u2)^k + (v1 - v2)^k)^{\frac{1}{k}} \tag{3.16}$$

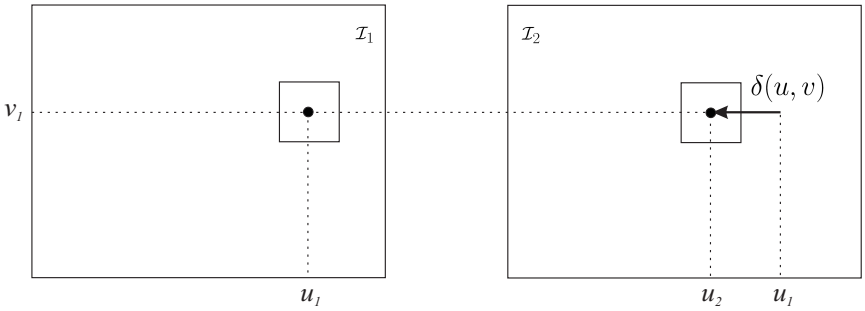


Abbildung 3.3: Fenster um Pixel (u_1, v) im ersten, und um Pixel (u_2, v) im zweiten Bild mit resultierender Disparität $\delta(u, v)$.

Es wird ersichtlich, dass die Tiefe eines Raumpunktes in einem achsparallelen Stereosystem durch simple *Triangulation* berechnet werden kann. Die Position und somit

[8]Die Formel für den euklidischen Abstand sei allgemeiner definiert, um den vertikalen Anteil zu berücksichtigen. Folglich sind auch Abstandsberechnungen für korrespondierende Punktpaare mit unterschiedlichen v-Werten möglich.

die Tiefe eines Raumpunktes sei durch den Schnittpunkt der Geraden, resultierend aus der Rückprojektion der optischen Zentren C_1, C_2 durch die korrespondierenden Punkten $\mathbf{m}_1, \mathbf{m}_2$, gegeben. Dies verdeutlicht zudem die benötigte Genauigkeit der Korrespondenzen. Abweichungen der ermittelten Positionen der Korrespondenzpunkte führen bei der Rückprojektion zu windschiefen Geraden. Auf die Ermittlung von Korrespondenzen und das Korrespondenzproblem wird genauer in Kapitel 4 eingegangen. Im Kapitel 3.3 werden zuvor verschiedene Verfahren vorgestellt, die eine allgemeine Stereogeometrie in eine achsparallele Stereogeometrie überführen. Dieser Prozess wird im Allgemeinen als Rektifikation bezeichnet.

3.3 Rektifikation

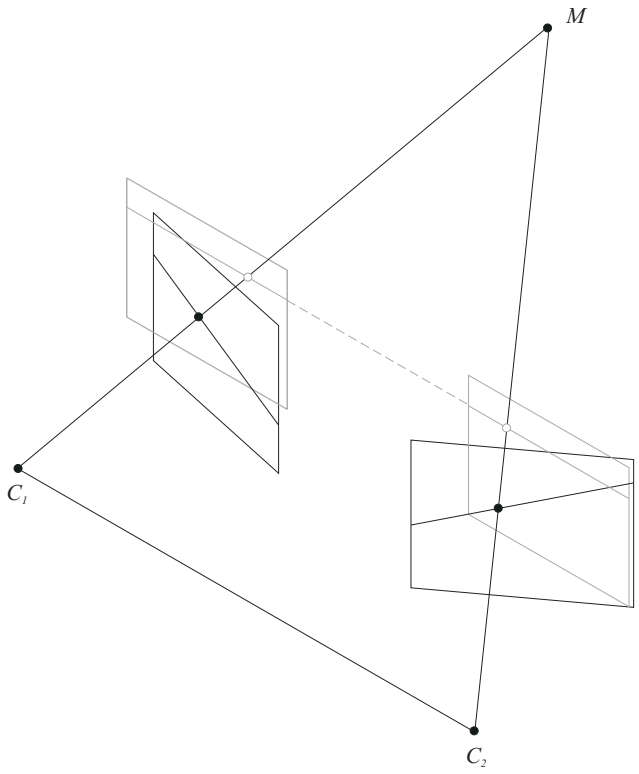


Abbildung 3.4: Rektifikation (schematisch) durch Reprojektion auf eine gemeinsame Ebene.

Wie bereits in der Einleitung des 3. Kapitels erwähnt, sollen die Bilder so ausgerichtet („rektifiziert“) werden, dass korrespondierende Epipolarlinien horizontal und auf gleicher Höhe verlaufen, um die nachfolgende Disparitätensuche effizient implementieren zu können. Die Epipolargeometrie aus Abschnitt 3.1 hat gezeigt, dass der Suchraum

auf eine Dimension entlang der Epipolarlinie reduziert werden kann. Jedoch stellen die schräg durch das Bild verlaufenden Linien eine ungünstige Struktur für eine Suche dar. Die Position der Pixel im Bild müsste jeweils noch berechnet werden und es dürfte kein Pixel doppelt ausgewertet oder vergessen werden. Daher soll die Rektifikation eine kolineare (horizontale) Anordnung der Epipolarlinien berechnen. Im Folgenden werden drei Verfahren zur Rektifikation aufgeführt.

3.3.1 Rektifikation mittels intrinsischer und extrinsischer Kameraparameter

In [AH88] und [FTV00, TV98] werden Rektifikationsverfahren vorgestellt, die auf bekannten intrinsischen und extrinsischen Parametern der Kameras basieren. Abbildung 3.4 verdeutlicht das folgende Prinzip der Rektifikation. Es handelt sich um eine Art Rückprojektion, wobei das konvergente System durch eine Rotation um die Projektionszentren der originalen Kameras in ein virtuelles achsparalleles Stereosystem überführt wird. Auf die rektifizierten Bildebenen werden die Bildpunkte des jeweiligen Originalbildes projeziert. Die neuen, künstlichen Bildebenen sind zueinander koplanar und parallel zur Basislinie. Die Basislinie muss zudem parallel zur neuen x-Achse der rektifizierten Bildebene sein. So ist sichergestellt, dass die Epipole ins Unendliche der horizonatelen Bildachse verschoben werden. In Abschnitt 3.1 wurde erläutert, dass sich alle Epipolarlinien im Epipol treffen. Zusammen mit der Tatsache, dass sich parallele Linien im Unendlichen treffen, stellen die Epipolarlinien somit parallele Linien im rektifizierten Bild dar. Um dieselbe y-Koordinate korrespondierender Epipolarlinien im Bild zu gewährleisten, müssen die beiden neuen Kameras dieselben intrinsischen Parameter haben.

Es entstehen also zwei neue virtuelle Kameraansichten (vgl. Abbildung 3.5), die sich nur noch durch ihren Verschiebungsterm $\mathbf{t}$ voneinander unterscheiden. Die Projektionszentren gleichen den alten und die Kameras sind so rotiert, dass sie dieselbe Orientierung haben. Dies kann durch die neuen perspektivischen Projektionsmatrizen formuliert werden zu:

$$\tilde{\mathbf{P}}_{old1} = \mathbf{A}_1[\mathbf{R}_1|\mathbf{c}_1] \quad \Longrightarrow \quad \tilde{\mathbf{P}}_{new1} = \mathbf{A}^*[\mathbf{R}^*| - \mathbf{R}^*\mathbf{c}_1] \tag{3.17}$$

$$\tilde{\mathbf{P}}_{old2} = \mathbf{A}_2[\mathbf{R}_2|\mathbf{c}_2] \quad \Longrightarrow \quad \tilde{\mathbf{P}}_{new2} = \mathbf{A}^*[\mathbf{R}^*| - \mathbf{R}^*\mathbf{c}_2] \tag{3.18}$$

Die intrinsische Kameramatrix $\mathbf{A}^*$ ist die gleiche für die beiden neuen Projektionsmatrizen. Die optischen Zentren $\mathbf{c}_1, \mathbf{c}_2$ sind gegeben durch die alten Projektionszentren.

Weiterhin sei:

$$\mathbf{A}^* = \frac{\mathbf{A}_1 + \mathbf{A}_2}{2}, \qquad \mathbf{R}^* = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{pmatrix} \tag{3.19}$$

$$\text{mit} \qquad \mathbf{r}_1 = \frac{\mathbf{c}_1 - \mathbf{c}_2}{\parallel \mathbf{c}_1 - \mathbf{c}_2 \parallel}, \qquad \mathbf{r}_2 = \frac{\mathbf{k} \times \mathbf{r}_1}{\parallel \mathbf{k} \times \mathbf{r}_1 \parallel}, \qquad \mathbf{r}_3 = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\parallel \mathbf{r}_1 \times \mathbf{r}_2 \parallel}, \tag{3.20}$$

wobei $\mathbf{k}$ in Gl. 3.20 ein beliebiger Vektor ist, der die neue y-Achse anzeigt (nach [FTV00]). Die Matrix $\mathbf{R}^*$ spannt das neue, normierte Koordinatensytem auf.

Da in dieser Arbeit dieselbe Kamera mit unveränderten intrinsischen Parametern genutzt wird, gilt für Gl. 3.19 $\mathbf{A}^* = \mathbf{A}_1 = \mathbf{A}_2 = \mathbf{A}$. Somit berechnen sich die rektifizierenden Transformationsmatrizen $\mathbf{T}_i$ wie folgt:

$$\mathbf{T}_1 = \tilde{\mathbf{P}}_{old1}\tilde{\mathbf{P}}_{new1}^{-1} \qquad \text{und} \qquad \mathbf{T}_2 = \tilde{\mathbf{P}}_{old2}\tilde{\mathbf{P}}_{new2}^{-1} \tag{3.21}$$

In dieser Arbeit wird dieses Verfahren verwendet, da durch die Kalibrierung nach Tsai [Tsa87] die intrinsischen Kameraparameter – wie für diesen Algorithmus gefordert – bekannt sind.

3.3.2 Exkurs: Weitere Rektifikationsmethoden

Rektifikation mittels Polarkoordinaten

Bei einem anderen Verfahren nach Pollefeys [Pol00] wird ein Epipol als Ursprung eines Polarkoordinatensystems genutzt. Die Rektifikation wandelt die kartesischen Pixelkoordinaten in Polarkoordinaten um. Einzige Voraussetzung ist die orientierte Fundamentalmatrix. Die Epipolarlinien entsprechen den verschiedenen Radien r_j des Winkels Θ_i. Unter der Bedingung, dass die Winkelschrittweite $\Delta\Theta_i$ und die radiale Schrittweite Δr_i die Größe eines Pixels nicht unterschreiten, kann das Bild zwischen den äußeren Bildrändern in die Polarkoordinatendarstellung (r, Θ) transformiert werden. Die Größe des rektifizierten Bildes ergibt sich durch die Höhe $\frac{\Theta_{max}-\Theta_{min}}{\Delta\Theta_i} \leq 2(w+h)$ und Breite $\frac{r_{max}-r_{min}}{\Delta r_j} \leq \sqrt{w^2+h^2}$ der Ränder des Ursprungsbildes, mit Breite w und Höhe h (vgl. Abbildung 3.6). Die rektifizierten Bilder werden anhand eines beliebigen Epipolarlinienpaares horizontal zueinander ausgerichtet, damit die Epipolarlinien die gleichen Abstände haben. Die Lage des Epipols[9] im, oder

[9] Es gibt neun Möglichkeiten für die Lage des Epipols. Acht davon ausserhalb des Bildes und eine im Bild. Sternförmige Epipolarlinien im Bild lassen auf eine Vorwärts- oder Rückwärtsbewegung

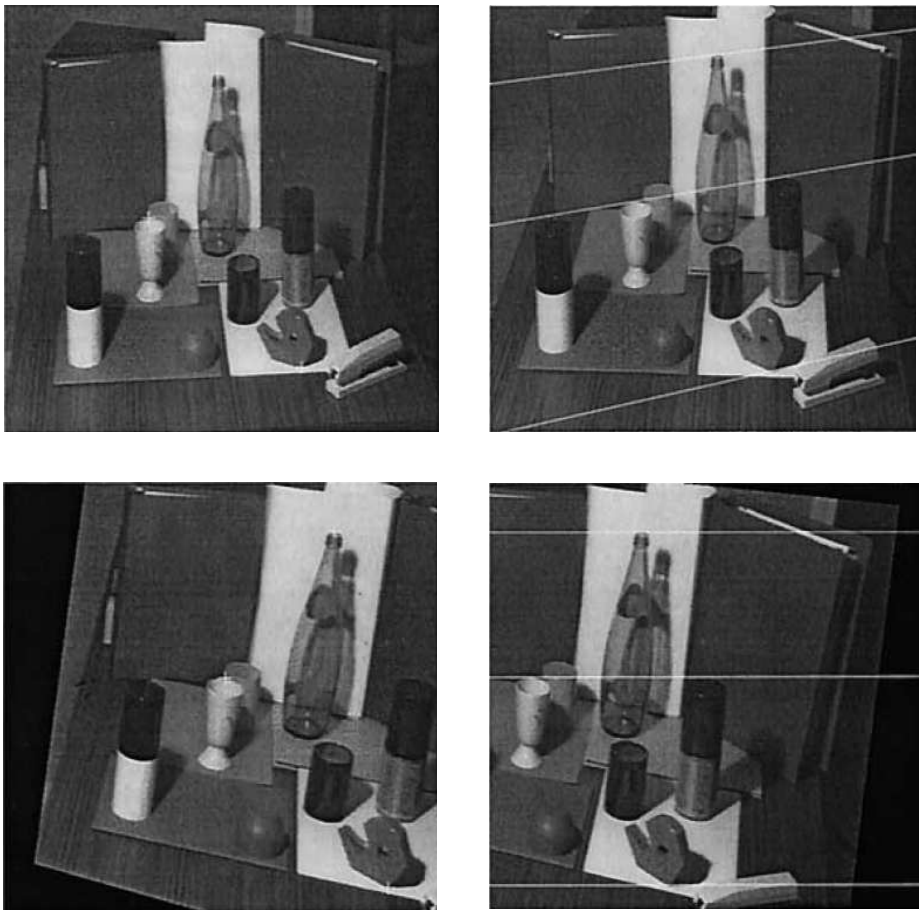


Abbildung 3.5: Linke und rechte Originalaufnahmen (oben) und rektifizierte Bilder mit horizontalen Epipolarlinien gleicher y-Koordinate (unten). Quelle: [FTV00].

um das Bild herum, hat keine Auswirkung auf die Transformation des Bildes. Jedoch sind die Bilder, bei denen der Epipol im Bild lag, in ihrer rektifizierten Form recht unübersichtlich (vgl. Abbildung 3.7).

Die Vorteile dieses Algorithmus liegt klar in seiner Unempfindlichkeit gegenüber im Bild liegender Epipole. Geraden, die nicht entlang der Epipolarlinien verlaufen, werden bei diesen Verfahren gekrümmt. Da die beiden rektifizierten Bilder des Stereobildpaares gleichermaßen gekrümmt sind, sollte es keine Auswirkungen auf die folgende Dispäritätensuche haben. Im Gegensatz zu dem vorher genannten Ansatz von Fusiello et al., wird in der Arbeit von Pollefey ausserdem auf die Minimierung der Bildgröße des rektifizierten Bildes geachtet. Fusiello et al. sind sich des Problems von Verzerrungen der Ausgabebilder bewusst, behaupten jedoch, dass diese erst bei einem schwach kalibrierten System auftreten.

der Kamera schließen.

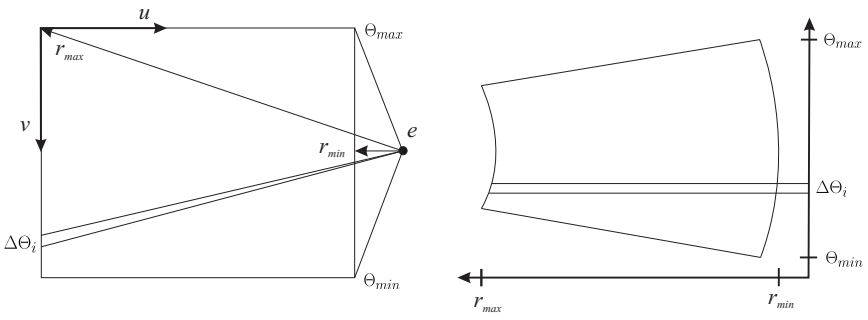


Abbildung 3.6: Rektifikationsprinzip. Umwandlung der kartesischen Koordinaten (u, v) in Polarkoordinaten (r, Θ). Die Θ-Achse ist nicht einheitlich und jede Epipolarline hat eine optimale Breite.

Rektifikation mittels Homographien

Ähnlich wie der vorhergehende Ansatz, benötigt Zhangs Verfahren [Zha98] zusätzlich zur Lage des Epipols die Kenntnis der Fundamentalmatrix. Beim Prinzip der Rektifikation mittels Homographien[10] wird eine projektive Abbildung zwischen zwei Ebenen verwendet. Die Zentralprojektion (siehe Gl. 2.2) selbst ist ebenfalls eine Homographie.

Normalerweise wird die Welt als euklidischer 3D-Raum wahrgenommen. In manchen Fällen jedoch (beispielsweise bei Bildern) ist es nicht anders möglich oder gar wünschenswert, nicht die volle euklidische Struktur zu betrachten. Dies führt zu einer Unterteilung in verschiedenen geometrischen Strata, die einfach übereinander gelegt werden können. Sie bestehen aus der projektiven Schicht, in der nur Kreuzverhältnisse und Incidencen gültig sind. Die projektive Schicht ist Teil der affinen Schicht. Die affine Schicht wiederum, in der zudem noch relative Verhältnisse und Parallelität vorherrschen, ist Teil der euklidischen Schicht, in der absolute Distanzen und Winkel gelten. In der Literatur sind viel weiter reichende Erläuterungen der geometrischen Strata zu finden, der interessierte Leser sei u.a. auf Pollefeys [Pol99], sowie Hartley [HZ03] verwiesen. Ein Vergleich der verschiedenen geometrischen Schichten, mit ihren Freiheitsgraden und Invarianten, ist in Tabelle 3.1 ersichtlich. Um die Bilder vergleichbar zu machen, müssen sie, wie schon mehrfach erwähnt, transformiert werden. Projektive Transformationen auf Bildern haben beim Rektifikationsprozess zur Folge, dass zuvor parallele Linien in den neuen Bildern verformt werden könnten[11].

[10] Eine Homographie (griech: *homos*, gleich; *gráphein*, schreiben, zeichnen.) ist nach Schreer [Sch05] eine perspektivische Projektion eines Punktes im projektiven Raum. Bei der Abbildung handelt es sich um eine projektive Abbildung. Zwischen zwei Ebenen erhält jedes Element der ersten Ebene auch ein Entsprechung in der zweiten Ebene.

[11] Schon der Abbildungsprozess über die Zentralprojektion ist eine projektive Transformation und hat Verzerrungen zur Folge hat.

Abbildung 3.7: Beispiel für rektifizierte Bilder durch Umwandlung in Polarkoordinaten (jeweils nur für ein Bild des Stereobildpaares). In der Abbildung links oben liegt der Epipol rechter Hand des Bildes, in der Abbildung rechts oben mitten im Bild auf dem Volleyball (weißer Punkt). Quelle: [Pol00, Kapitel 7, S. 69].

In diesem Ansatz wird von der weitaus einfacheren Struktur der projektiven Geometrie ausgegangen. Ziel ist es ebenfalls, den Epipol ins Unendliche zu verschieben, um zu einer parallelen Anordnung der Epipolarlinien zu gelangen. Durch eine Überführung der projektiven Geometrie in den affinen Raum soll über die Homographiematrizen eine Rektifikation der Bilder durchgeführt werden. Es soll hier zwar nicht auf die Ermittlung der einzelnen Matrizen im Detail eingegangen werden, jedoch anhand einer Kamera den grundlegenden Ansatz sehr kurz erläutert werden. Verfahren zur Bestimmung der Homograpiehen sind u.a. beschrieben in [Har99].

Zur Ermittlung der rektifizierenden Transformationsmatrix wird die Homographie $\mathbf{H}_i$ der Ansicht i in vier einfachere Matrizen zerlegt. Die erste Aufteilung ist die Zerlegung in eine projektive und eine affine Teilmatrix $\mathbf{H}_i = \mathbf{H}_{i,P}\mathbf{H}_{i,A}$. Die Matrix $\mathbf{H}_{i,P}$ beinhaltet sämtliche projektiven Transformationen um den Epipol e_i ins Unendliche

Bezeichnung	Transformationsmatrix	DOF	Invarianten
Projektiv	$\mathbf{T}_P = \begin{pmatrix} \mathbf{A}_{3\times3} & \mathbf{b}_{3\times1} \\ \mathbf{v}_{1\times3}^T & v_{1\times1} \end{pmatrix}$	15	Doppelverhältnisse
Affin	$\mathbf{T}_A = \begin{pmatrix} \mathbf{A}_{3\times3} & \mathbf{b}_{3\times1} \\ \mathbf{0}_{1\times3}^T & 1 \end{pmatrix}$	12	Relative Abstände entlang einer Richtung, Parallelität
Euklidisch	$\mathbf{T}_E = \begin{pmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ \mathbf{0}_{1\times3}^T & 1 \end{pmatrix}$	6	Absolute Distanzen

Tabelle 3.1: Freiheitsgrade (DOF) und Transformationsmatrizen der verschiedenen geometrischen Räume, mit ihren Invarianten.

zu verlagern und die Epipolarlinien ℓ_i parallel, jedoch nicht achsparallel der neuen u-Achse, anzuordnen. Die zweite Zerlegung trennt die affine Matrix $\mathbf{H}_{i,A} = \mathbf{H}_{i,R}\mathbf{H}_{i,s}$ in einen Rotations- und einen Scherungsanteil auf. Die Rotationsmatrix $\mathbf{H}_{i,R}$ richtet die Epipolarlinien parallel zur neuen u-Achse und jeweils auf dieselbe v-Koordinate aus. Die Scherungsmatrix $\mathbf{H}_{i,s}$ dient dazu, die durch $\mathbf{H}_{i,P}$ entstandene projektive Verzerrung bestmöglich auszugleichen. Die letzte Teilmatrix ist die uniforme Skalierung $\mathbf{H}_u$ und dient der Begrenzung der entstehenden Größe des neuen Bildes. Denn je näher der Epipol an dem Bild liegt, desto größer wird das rektifizierte Bild. Mit $\mathbf{H}_u$ kann durch eine passende Skalierung diesem Wachstum entgegengewirkt werden. Sie ist für beide Bilder gleich und trägt daher kein Index i.

Die Homographie kann somit ausgedrückt werden durch $\mathbf{H}_i = \mathbf{H}_u\mathbf{H}_{i,R}\mathbf{H}_{i,s}\mathbf{H}_{i,P}$, $i \in \{1, 2\}$. Die Rektifikation aller Bildpunkte $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2$ erfolgt demnach durch Gl. 3.22.

$$\tilde{\mathbf{m}}_{1,new} = \mathbf{H}_1\tilde{\mathbf{m}}_{1,old} \qquad \text{und} \qquad \tilde{\mathbf{m}}_{2,new} = \mathbf{H}_2\tilde{\mathbf{m}}_{2,old} \tag{3.22}$$

Diese recht aufwendige Art der Rektifikation zeichnet sich durch minimale Voraussetzungen an die Ursprungsbilder aus. Alle notwendigen Informationen zur Rektifikation werden aus der Fundamentalmatrix gewonnen. Ein besonderes Augenmerk liegt auf der verzerrungsarmen Rektifikation. Jedoch eignet sich dieses Verfahren nicht für solche Fundamentalmatrizen, bei denen der Epipol nah oder gar im Bild liegt. Diese Bilder würden bei der Rektifikation stark verzerrt oder gespalten, da ein Teil des Bildes ebenfalls ins Unendliche projeziert werden würde.

Durch die Kalibration der Kamera wird der projektiven Raum verlassen und man gelangt in den euklidischen Raum. Da in dieser Arbeit die Kamera kalibriert wird, bietet sich die Nutzung dieses Verfahrens somit nicht an.

3.4 Zusammenfassung

In diesem Kapitel wurde die Beziehung zwischen zwei Kameras mathematisch erläutert. Das wegweisende Resultat ist die Epipolargleichung, welche die Abbildungen beider Ansichten mit der Geometrie der Kameras verknüpft. Ihre wesentliche Aussage ist die Einschränkung, dass korrespondierende Bildpunkte in der anderen Ansicht auf den Epipolarlinien liegen müssen. Es wurden häufig zitierte und etablierte Verfahren vorgestellt, die unter Verwendung der Epipolarbedingung die Schätzung der Fundamentalmatrix allein aus Punktkorrespondenzen zweier Stereoansichten ermöglichen.

Die Korrespondenzanalyse konnte um eine Dimension reduziert werden. Durch die Rektifikation, die Generierung virtueller achsparalleler Stereosysteme, kann die Korrespondenzanalyse zudem wesentlich einfacher implementiert werden. Korrepondierende Bildpunkte liegen infolge der Rektifikation auf der gleichen Zeile beider Bilder.

4 Korrespondenzanalyse

Die Korrespondenzanalyse, auch Stereoanalyse genannt, bezeichnet die Analyse zweier Ansichten eines Stereokamerasystems. Die Zielsetzung besteht darin, korrespondierende Bildpunkte oder Bildmerkmale in zwei unterschiedlichen Bildern zu finden. Die Korrespondenzanalyse ist ein klassisches Aufgabenfeld der Bildanalyse. Dank der Entwicklung der Zufallspunktstereogramme von Bela Julesz (1959) konnte bewiesen werden, dass der Mensch das Korrespondenzproblem allein auf der Basis von Texturen zu lösen vermag. Ganz so einfach gestaltet es sich in der Computer-Vision leider nicht.

Die meisten Methoden zur Ermittlung von Korrespondenzen in Bildpaaren unterliegen zweier grundlegender Annahmen:

1. Die meisten Szenenpunkte sind in beiden Ansichten sichtbar.

2. Korrespondierende Bildregionen sind gleich.

Das mag für Stereosysteme gelten, deren Szenenobjekte viel weiter entfernt sind als die Länge der Basisline. Allerdings kann es sein, dass beide Annahmen keine Gültigkeit haben.

Das Korrespondenzproblem kann als ein Suchproblem aufgefasst werden: Sei ein Element der einen Ansicht gegeben, wird das dazu korrespondierende Element der anderen Ansicht gesucht [TV98]. Hierbei müssen zwei Entscheidungen getroffen werden: 1) Welches Bildelement soll verglichen werden und 2) Welches Ähnlichkeitsmaß kann dafür verwendet werden. Ma et al. drücken das Korrespondenzproblem natürlichsprachlich folgendermaßen aus:

> „Das *Korrespondenzproblem* besteht darin, herauszufinden welcher Punkt eines Bildes zu welchem Punkt eines anderen Bildes korrespondiert, unter der Annahme, dass es sich um Abbilder desselben Raumpunktes handelt.“ [MSKS04, Kapitel 4.1, S. 76, Übersetzung des Authors]

Die Algorithmen der Korrespondenzanalyse können grob in zwei Klassen unterteilt werden. Es handelt sich dabei um die pixelbasierten- und die merkmalsbasierten Verfahren.

Bei den *pixelbasierten Verfahren* werden Regionen fester Größe um Pixel herum miteinander verglichen und dessen Korrelation über ein Ähnlichkeitsmaß bestimmt. Das korrespondierende Bildelement findet sich an der Stelle, an der ein Maximum der Ähnlichkeitsfunktion herrscht. Die pixelbasierten Verfahren betrachten die Menge aller Pixel. Wegen ihres Bezugs auf einen Bildblock werden sie auch als *Block-Matching* bezeichnet.

Die *merkmalsbasierten Verfahren* beschränken den Suchraum auf eine Untermenge von Merkmalen im Bild. Merkmale können unter anderem Ecken- oder Liniensegmente sein. Im Gegensatz zu der Korrelationsmessung nutzen diese Verfahren die Abstände zwischen den Merkmalen als Ähnlichkeitskriterium. Als Vorverarbeitungsschritt ist somit eine Merkmalsextraktion notwendig (vgl. Anhang B).

Durch die erfolgreichen Einschränkungen der vorherigen Kapitel (vgl. Abschnitte 3.1 und 3.3), gestaltet sich das Korrespondenzproblem etwas einfacher. Beispielsweise durch die Bedingung, dass korrespondierende Punkte in den beiden Bildern auf einer Geraden, der Epipolarlinie, liegen müssen. Da sich zudem diese Epipolarlinien durch den Rektifikationsprozess zu achsparallelen, horizontalen Linien anordnen lassen, kann der Suchraum auf die gleiche Zeile des anderen Bildes beschränkt werden[1]. Können zudem vorab Annahmen über den Tiefenbereich der Szenen gemacht werden, so ist auch der Disparitätsbereich durch seine reziproke Beziehung zur Tiefe einschränkbar (vgl. Kapitel 3.2).

Zusammenfassend ist also die Zielsetzung der Korrespondenzanalyse eine robuste, fehlerfreie und eindeutige Zuordnung zwischen Bildmerkmalen zweier unterschiedlicher Bilder derselben Szene. Die Wahl des Verfahrens hängt von der Art der Anwendung, sowie von den Hard- und Softwareanforderungen ab.

4.1 Pixelbasierte Verfahren

Die pixelbasierten Verfahren sind leichter zu implementieren und berechnen dichte Disparitätskarten. Sie nutzen die Bildstruktur in der Umgebung eines Pixels, da diese aussagekräftiger ist als ein einzelner Bildpunkt. Daher wird bei der Korrespondenzanalyse ein Fenster, bzw. Block, um den Aufpunkt betrachtet. Die Block-Matching Verfahren sind translatorische Schätzverfahren. Es werden nur geradlinige

[1] Aufgrund von Ungenauigkeiten in dem Rektifikationsprozess kann es vorkommen, dass der Suchraum auf einige Zeilen um die Referenzzeile ausgeweitet werden muss.

Bewegungen der einzelnen betrachteten Blöcke berücksichtigt. Bewegungs- und Helligkeitsänderungen zweier Bilder (Rotation, Zoom, starke Helligkeitänderung) führen zu schwerwiegenden Problemen. Es gibt zwar Schätzverfahren, die genau auf solche Probleme abgestimmt sind, jedoch schwierig zu implementieren sind.

Beim Block-Matching wird für jede Position (u_1, v_1) der ersten Ansicht um das Pixel ein Referenzblock der Größe (m, n) gewählt, und mit entsprechenden Musterblöcken selber Größe in der zweiten Ansicht um $\delta(u, v)$ verschoben, verglichen. In Abschnitt 3.2 wurden Disparitäten schon eingehend betrachtet. Die Ähnlichkeit zweier Bildblöcke wird duch eine Bewertungsfunktion berechnet. Je kleiner das Fenster um den Bildpunkt gewählt wird, desto mehr Informationen lassen sich im Disparitätenbild wiederfinden. Je größer das Fenster gewählt wird, desto eher werden kleine Abweichungen der Bildsignale in dem Fenster von den Bewertungsfunktionen verschluckt. Wie aus [MSKS04, Sch05, TV98] ersichtlich, finden folgende parametrische Ähnlichkeitsmaße häufige Verwendung.

4.1.1 Mittlerer absoluter Fehler

Mit dem mittleren absoluten Fehler (engl. *sum of absolute differences*, SAD) berechnet man die absolute Differenz zweier Bildblöcke (siehe Gl. 4.1). Die Ähnlichkeit ist dort am größten, wo die Differenz minimal wird. Der Abstand der Bildblockzentren zueinander stellt die Disparität $\delta(u, v)$ dar. Seien $f_i(u, v), i \in \{1, 2\}$ die Intensitätsbilder der ersten und zweiten Ansicht der Szene, so lautet die Formel:

$$\delta_{SAD}(u,v) = \arg\min_{\delta(u,v)} \frac{1}{|\Lambda|} \sum_m \sum_n |f_1(u+m, v+n) - f_2(u+\delta(u,v)+m, v+n)| \quad (4.1)$$

wobei Λ die Umgebung, also die Anzahl der Pixel in dem betrachteten Fenster gekennzeichnet. Durch unterschiedliche Blenden der beiden Kameras können Differenzen der mittleren Helligkeit auftreten. Die resultierenden Mehrdeutigkeiten in der Korrespondenzanalyse können durch eine Mittelwertbefreiung des Muster- und Referenzblocks

vermieden werden. Eine mittelwertbefreite Form der SAD ist in Gl. 4.2 aufgeführt.

$$\begin{aligned}\delta_{\widehat{SAD}}(u,v) = \arg\min_{\delta(u,v)} \frac{1}{|\Lambda|} \sum_m \sum_n |(f_1(u+m,v+n) - \bar{f_1}) \\ - \sum_m \sum_n (f_2(u+\delta(u,v)+m,v+n) - \bar{f_2})| \\ \text{mit} \quad \bar{f_1} = \frac{1}{|\Lambda|} \sum_m \sum_n f_1(u+m,v+n) \\ \text{und} \quad \bar{f_2} = \frac{1}{|\Lambda|} \sum_m \sum_n f_2(u+\delta(u,v)+m,v+n)\end{aligned} \tag{4.2}$$

4.1.2 Mittlerer quadratischer Fehler

Häufiger als die SAD wird der mittlere quadratische Fehler (engl. *sum of squared differences*, SSD) als Abstandsmaß genutzt. Eine anschließende Quadrierung der Differenz (Gl. 4.3) führt zu einer stärkeren Gewichtung von größeren Fehlern. Somit ist die SSD sensibler als die SAD gegenüber Fehlern. Auch hier seien wieder $f_i(u,v), i \in \{1,2\}$ die Intensitätsbilder der ersten und zweiten Ansicht der Szene.

$$\delta_{SSD}(u,v) = \arg\min_{\delta(u,v)} \frac{1}{|\Lambda|} \sum_m \sum_n (f_1(u+m,v+n) - f_2(u+\delta(u,v)+m,v+n))^2 \tag{4.3}$$

Multipliziert man diese kompakte Form der SSD-Formel aus, gelangt man zu Gl. 4.4. Dabei stellen die ersten beiden Summanden die konstante Energie des Muster- und Referenzblockes dar, während der dritte Term die Korrelation der beiden Blöcke beschreibt. Der quadratische Fehler wird minimal, wo die Ähnlichkeit am größten ist.

$$\begin{aligned}\delta_{SSD}(u,v) = \arg\min_{\delta(u,v)} \frac{1}{|\Lambda|} \sum_m \sum_n (f_1(u+m,v+n))^2 \\ + \sum_m \sum_n (f_2(u+\delta(u,v)+m,v+n))^2 \\ -2 \sum_m \sum_n (f_1(u+m,v+n) f_2(u+\delta(u,v)+m,v+n))\end{aligned} \tag{4.4}$$

4.1.3 Normierte Kreuzkorrelation

Um eine Abhängigkeit von der Energie des Muster- und Referenzblockes, wie in Gl. 4.4, zu vermeiden, wird die normierte Kreuzkorrelation (engl. *normalized cross-correlation*, NCC) verwendet. Hierbei wird, wie in Gl. 4.5 ersichtlich, nur ein Vergleich der relativen Unterschiede zwischen den Bildblöcken ermittelt, da im Nenner auf die Energie der beiden Blöcke normiert wird. Die normierte Kreuzkorrelation liefert dort das Maximum, wo die Ähnlichkeit am größten ist. Auch die NCC weist ähnlich der SSD ein sensibles Verhalten gegenüber Ausreißern (engl. *outlier*) auf. Die Intensitätsbilder der ersten und zweiten Ansicht der Szene sind mit $f_i(u,v), i \in \{1,2\}$ gekennzeichnet.

$$\delta_{NCC}(u,v) = \arg\max_{\delta(u,v)} \frac{\sum_m \sum_n f_1(u+m,v+n) f_2(u+\delta(u,v)+m,v+n)}{\sqrt{\sum_m \sum_n (f_1(u+m,v+n))^2 \sum_m \sum_n (f_2(u+\delta(u,v)+m,v+n))^2}} \tag{4.5}$$

Neben den parametrischen Bewertungsfunktionen gibt es auch noch nicht-parametrische Ähnlichkeitsmaße. Bei der *Rank-Transformation* beispielsweise wird die Anzahl jener Bildpunkte berechnet, die einen geringeren Intensitätswert aufweisen, als der Pixel des Aufpunktes. Es wird ein neues Bild berechnet, bei dem diese Werte an der Pixelposition abgetragen werden. Die neuen Bilder werden schließlich mit einem parametrischen Ähnlichkeitsmaß (wie im Abschnitt erläutert) verglichen. Dieses Verfahren ist invariant gegenüber Rotation, Reflektion und monotonen Grauwerttransformationen.

Bei der *Census-Transformation* wird jedem Fenster eine Bitkette zugewiesen. Sie beschreibt die Relation der Intensitäten der Bildpunkte im Messfenster bezüglich der Intensität des Aufpunktes. Die länge der Bitkette entspricht daher der Anzahl der verglichenen Pixel im Messfenster. Ähnlich der parametrischen Ähnlichkeitsmaße wird die Ähnlichkeit aus der Summe von Hamming-Distanzen[2] berechnet.

Für weitere Informationen zur Rank- und Census-Transformation siehe [Sch05].

4.2 Exkurs: Merkmalsbasierte Verfahren

Wie in der Einleitung dieses Kapitels erwähnt, werden bei den merkmalsbasierten Verfahren Bildmerkmale hinsichtlich ihrer Korrespondenz untersucht. Die Merkmale

[2]Die Hamming-Distanz beschreibt die Anzahl der unterschiedlichen Bits in zwei Bitketten.

werden in einem Vorverarbeitungsschritt ermittelt (vgl. Anhang B). Die merkmalsbasierten Verfahren nutzen unterschiedliche Methodiken, je nachdem, ob sie sich auf die Korrespondenzanalyse von Punktmerkmalen oder Liniensegmenten stützen.

4.2.1 Korrespondenzanalyse von Punktmerkmalen

Als Punktmerkmale eines Bildes würden sich Ecken eignen, die sich in beiden Bildern wieder finden. Die Extraktion der Punktmerkmale lässt sich durch Standardverfahren wie dem *Moravec-Operator* oder dessen Erweiterung, dem *Harris-Ecken-Detektor*, realisieren (vgl. Anhang B). Ist die Vorverarbeitung abgeschlossen und liegen interessante Punkte für beide Bilder vor, so werden über die Epipolarbedingung $\tilde{\mathbf{m}}_2^\top \mathbf{F} \tilde{\mathbf{m}}_1 = \tilde{\mathbf{m}}_2^\top \tilde{\ell}_2 = 0$ (siehe Abschnitt 3.1.1) alle Punkte im linken Bild ausgewählt, die nur einen entsprehenden Punkt auf der Epipolarlinie im rechten Bild haben. Die Betrachtung der Bildstrukturen in einem Fenster um die Punkte herum kann die Zuverlässigkeit zusätzlich erhöhen.

Die verbleibenden Merkmalspunkte des linken Bildes haben folglich mehrere Kandidaten im rechten Bild. Auch hier kann die Berücksichtigung der Umgebung weitere Einschränkungen erbringen. Ausserdem können auch geometrische Kriterien, wie etwa die Lage mehrere Punkt zueinander, in den verschiedenen Bildern untersucht werden. Des weiteren kann durch die Glattheitsbedingung[3] die Güte von Korrespondenzen optimiert werden. Bei bekannter Epipolargeometrie und durch Nutzung der eben erwähnten Bedingungen lassen sich somit robust Korrespondenzen bestimmen.

4.2.2 Korrespondenzanalyse von Liniensegmenten

Eine komplexere Merkmalsstruktur als die Punktmerkmale stellen Liniensegmente dar. Sie kommen weitaus weniger im Bild vor als Punkte und vereinfachen die Korrespondenzanalyse. Durch die geringere Anzahl ist eine zuverlässige Zuordnung der Merkmale möglich.

Liniensegmente zeichnen sich durch starke Hell-Dunkel-Übergänge an 3D-Objektkanten im Bild aus und sind im $\Re^2$ durch vier Parameter eindeutig definiert. Die Beschreibung kann über die Anfangs- und Endpunkte des Segmentes erfolgen (u_a, v_a, u_e, v_e) oder über den Mittelpunkt (u_m, v_m) mit Länge l und Orientierung α. Die Eigenschaften Länge und Orientierung eignen sich wegen der Blickwinkelabhänbgigkeit

[3] Die Glattheitsbedingung besagt, dass örtlich benachbarte Bildpunkte nur geringe Disparitätsänderungen aufweisen.

nur eingeschränkt als Ähnlichkeitsmaß, da sie durch perspektivische Verzerrungen in beiden Ansichten sehr unterschiedliche Anordnungen und Formen aufweisen können.

Um ein Liniensegment zu erstellen müssen zuerst die Kantenpunkte im Bild gefunden werden und in einem 2. Schritt zu Linien zusammengefasst werden. Die Bestimmung von Kantenpunkte erfolgt meist über Ableitungsfilter 2. Ordnung. Sie geben einen Nulldurchgang an der Stelle an, wo die 1. Ableitung[4] maximal wird, bzw. sich im Originalsignal die Mitte des Hell-Dunkel-Übergangs befindet. Vorher wird meist eine Glättung über einen Gauß-Filter vorgenommen, um die Rauschstörungen zu vermindern. Ein weit verbreitetes Verfahren stellt der *Canny-Kantendetektor* dar [Can83]. Das resultierende Kantenpunktbild kann durch Hough-Transformation oder die Verfolgung von Kantenpunkten zu Liniensegmenten zusammengefasst werden.

Ein robustes Verfahren zur Korrespondenzanalyse von Liniensegmenten wurde von N. Ayache 1991 vorgestellt und beruht auf den drei Phasen Prädikation, Propagierung und Validierung. Eine gute Erläuterung des Verfahrens ist in [Sch05] gegeben. Bevor in der Prädikationsphase durch Schätzung potentielle Zuordnungen zu Liniensegmenten in beiden Bildern gemacht werden, wird die Nachbarschaft für Liniensegmente beider Ansichten bestimmt. In der rekursiven Propagierungsphase werden Ergebnisse der Prädikation unter Verwendung des Nachbarschaftsgraphen fortgepflanzt (propagiert). Diese Phase liefert für alle in der Prädikationsphase ermittelten Segmente des Bildes $\mathcal{I}_1$ mehrere Hypothesen von korrespondierenden Liniensegementen des Bildes $\mathcal{I}_2$, basierend auf Ähnlichkeitsbedingungen. Je länger der Nachbarschaftsgraph, desto mehr Segmente beinhaltet die Hypothese. In der Validierungsphase ist das erste Kriterium für die Auswahl der richtigen Korrespondenzen daher die Länge der Hypothese. Bei gleicher Hypothesenlänge kann das korrespondierende Segment über das Ähnlichkeitsmaß J_s zwischen Segment S^l des Bildes $\mathcal{I}_1$ und S^r des Bildes $\mathcal{I}_2$ ermittelt werden (siehe Gl. 4.6). Dafür können die verschiedensten geometrischen und Grauwerteigenschaften von Liniensegmenten genutzt werden.

$$J_s(S^l, S^r) = w_0 \frac{\Delta\alpha}{\Delta\alpha_{max}} + w_1 \frac{\Delta l}{\Delta l_{max}} + \frac{\Delta Grad}{\Delta Grad_{max}} + \frac{\Delta MGW}{\Delta MGW_{max}}$$

$$\text{mit} \quad 0 < w_{1,2,3} < 1, \quad \forall DIR^l = DIR^r, \quad \Delta\alpha = |\alpha^l - \alpha^r|, \quad \Delta l = |l^l - l^r|, \tag{4.6}$$

$$\Delta Grad = |Grad^l - Grad^r|, \quad \Delta MGW = |MGW^l - MGW^r|$$

Dabei ist w eine Gewichtung, $Grad$ der Gradient, DIR die Richtung der Intesitäts-

[4] Ableitungsfilter 1. Ordnung sind die sogenannten Gradientenfilter.

änderung und MGW der Mittlere Grauwert. Die Korrespondenz zweier Segmente geht aus dem besten Ähnlichkeitsmaß hervor.

4.3 Stereoalgorithmus von Birchfield und Tomasi

Von Birchfield und Tomasi wird in [BT98] ein Zweiphasen Stereoalgorithmus vorgestellt. In einem ersten Matchingschritt werden grobe Disparitäten zwischen den Epipolargeraden der beiden Eingabebilder gesucht. Dabei werden im Gegensatz zu den in Abschnitt 4.1 verglichenen Fensterblöcken lediglich die Grauwertintensitäten auf Pixelebene innerhalb einer Bildzeile verglichen und durch dynamische Programmierung eine lokale Kostenfunktion minimiert. In dem zweiten Matchingschritt wird versucht die groben Disparitäten durch Verknüpfung der einzelnen Bildzeilen zu verfeinern. In dieser Arbeit wird eine `OpenCV` [Int05] Implementation des Birchfield-Algorithmus genutzt.

Korrespondenzen beschreiben Birchfield und Tomasi als übereinstimmende Sequenzen (engl. *match sequence*), wobei jede Übereinstimmung ein Tupel (u_1, u_2) ergibt. Die zentrale Rolle spielt ihre Kostenfunktion $\gamma(S)$ aus Gl. 4.7. Für jede übereinstimmende Sequenz S wird die Kostenfunktion berechnet. Sie definiert die Wahrscheinlichkeit, dass S tatsächlich eine Beschreibung der wahren Korrespondenz ist.

$$\gamma(S) = N_{occ}\kappa_{occ} - N_s\kappa_r + \sum_{i=1}^{N_s} d(u_1, u_2) \tag{4.7}$$

wobei κ_{occ} einer konstanten Strafe für verdeckte Sequenzen s, κ_r einer konstanten Belohnung für korrekte Sequenzen und $d(u_1, u_2)$ der Unähnlichkeit der Pixel u_1 und u_2 aus Bild $\mathcal{I}_1$ und $\mathcal{I}_2$ entspricht. N_{occ} und N_s beschreiben die Anzahl von verdeckten und korrekten Sequenzen, nicht die Anzahl der verdeckten Pixel.

Die Abstandsfunktion $d(u_1, u_2)$ ist unempfindlich gegenüber Bildabtastfehlern, da sie nicht nur die Differenz der Intensitäten $I_1(u_1), I_2(u_2)$ einzelner Pixel betrachtet, sondern auch die Nachbarpixel in die Berechnung mit einbezieht. Die aktuell betrachteten Pixel einer Zeile v haben die Intensität $I_{u1} = I_1^0$ im linken Bild und $I_{u2} = I_2^0$ im rechten Bild. Die Nachbarn von I_2^0 in Zeile v seien I_{u2-1} und I_{u2+1}. Zwischen I_2^0 und den Nachbarn werden die Intensitäten interpoliert. Man erhält das Minimum I_2^- sowie das Maximum I_2^+ für I_2^0. Sie bilden ein Intervall mit den Grenzen $[I_{2min}, I_{2max}]$. Das Abstandsmaß $d_1(u_1, u_2) = 0$, wenn $I_{u1} = I_1^0$ innerhalb des Intervalls liegt, ansonsten entspricht es der Distanz zur nächsten Flächengrenze. Analog dazu wird $d_2(u_1, u_2)$ über eine Fläche im linken Bild berechnet. Als Abstandsmaß von Birchfield ergibt

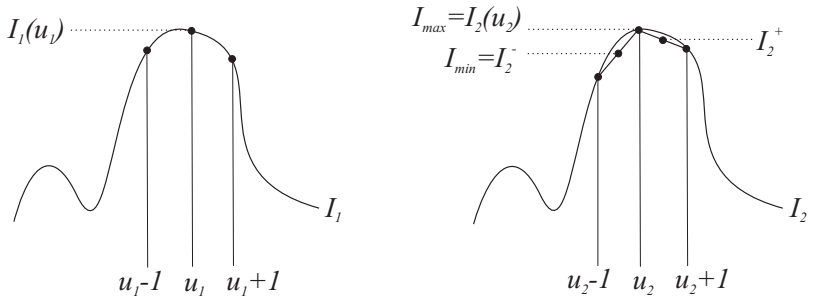


Abbildung 4.1: Definition und Berechnung von $\bar{d}(u_1, u_2, I_1, I_2)$. Dabei entspricht $I_1(u_1)$ der Intensität des Pixels u_1 im linken Bild. I_2^-, I_2^+ sind die interpolierten Intensitäten an Pixelposition $u_2 - \frac{1}{2}$ und $u_2 + \frac{1}{2}$. Zudem ist $I_{min} = \min(I_2^-, I_2^+, I_2(u_2))$ und $I_{max} = \max(I_2^-, I_2^+, I_2(u_2))$.

sich somit $d(u_1, u_2) = \min\{\bar{d}(u_1, u_2, I_1, I_2), \bar{d}(u_2, u_1, I_2, I_1)\}$, wobei $\bar{d}(u_1, u_2, I_1, I_2) = \max\{0, I_1^0 - I_{max}, I_{min} - I_1^0\}$ und $\bar{d}(u_2, u_1, I_2, I_1) = \max\{0, I_2^0 - I_{max}, I_{min} - I_2^0\}$. Abbildung 4.1 verdeutlicht dieses Berechnungsprinzip.

Laut Birchfield und Tomasi liegt der Vorteil ihres Algorithmus gegenüber klassischen Verfahren (vgl. 4.1) insbesondere in der Unanfälligkeit bei großen untexturierten Regionen[5], sowie der schnellen Berechnungszeit von etwa 1,5 ms pro Pixel auf einer Workstation [BT98]. Diese Gründe führen zur Verwendung des Algorithmus in der vorliegenden Arbeit.

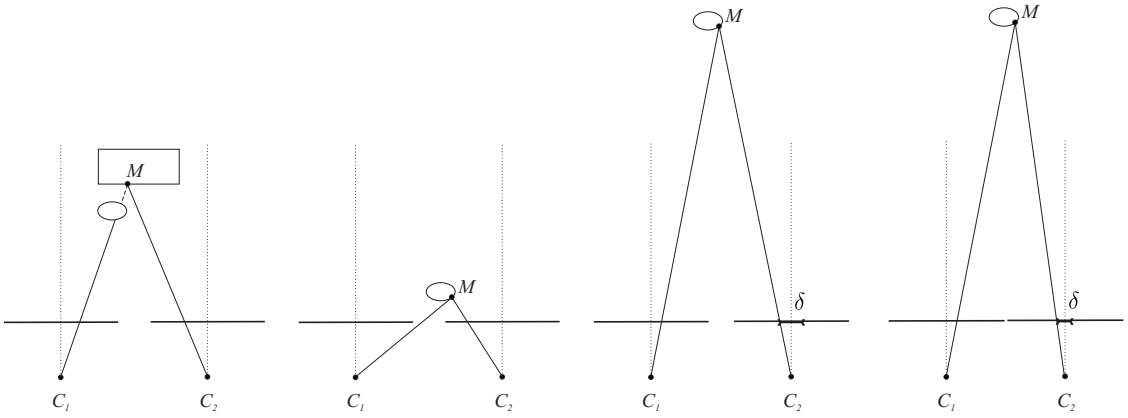
4.4 Probleme der Korrespondenzanalyse

Probleme der Korresponenzanalyse resultieren aus Verdeckungen oder dem eingeschränkten Blickfeld der Kameras. In Abbildung 4.2(a) ist der Objektpunkt M in der ersten Ansicht verdeckt und es ist unmöglich für M in der zweiten Ansicht einen Korrespondenzpunkt im anderen Bild zu finden. Abbildung 4.2(b) verdeutlicht den eingeschränkten Tiefenbereich von Objektpunkten. Hier ist ebenfalls M nur in einem Bild sichtbar.

Aus unterschiedlichen Abständen der Stereokameraanordnung zueinander können sich ebenfalls gewisse Probleme, wie auch Vorteile, ergeben. Die Stereokamerasysteme können abhängig von dem Verhältnis der Basislänge zur mittleren Tiefe der Szene unterteilt werden in Systeme mit kleiner Basisline (engl. *small baseline stereo*) und Systeme mit großer Basislinie (engl. *wide baseline stereo*)[Sch05, Pol00].

[5]Solange die Bedingung zutrifft, dass eine Intensitätsschwankung von mindestens 5 zwischen minimalem und maximalem Grauwert eine Tiefendiskontinuität beschreibt.

Die Vorteile bei Systemen mit *kleiner Basislinie* liegen in der einfacheren Korrespondenzanalyse, wenn es sich bei den beiden Bilder um Ansichten derselben Szene handelt. Es ist offensichtlich, dass bei geringem Abstand der Kameras der perspektivische Unterschied der Bilder kleiner und demzufolge die Ähnlichkeiten größer sind. Außerdem kommt es nur zu geringen, blickrichtungsabhängig verdeckten[6] Bildblöcken. Aus dem nur geringen perspektivischen Unterschied ergibt sich jedoch eine verringerte Disparitätsauflösung (vgl. Abbildungen 4.2(c) und 4.2(d)). Aus dem kleinem Triangulationswinkel resultiert eine große Ungenauigkeit bezüglich der Tiefe in der Szene.



(a) Verdeckung (b) Eingeschränkter Sichtbereich (c) System mit großer Basislinie (d) System mit kleiner Basislinie

Abbildung 4.2: Verschiedene Probleme der Korrespondenzanalyse (2D) für einen Oberflächenpunkt M eines Objektes (Elipse und Rechteck).

Wird hingegen der Abstand der Kameras vergrößert, so folgt aus dem erhöhten perspektivischen Unterschied der Bilder eine aufwendigere Korrespondenzanalyse, da die Bilder nur bedingt gleich sind und zudem größere verdeckte Bereiche aufweisen. Durch den erhöhten Disparitätsabstand und größeren Triangulationswinkel lässt sich jedoch die Tiefe der Szene präziser bestimmen.

Weitere Probleme bei der Korrespondenzanalyse treten in der Regel durch periodische Strukturen und homogene, schwach texturierte Bildbereiche auf. Diese Bildmuster führen zu mehrdeutigen Ergebnissen der Ähnlichkeitsfunktion aus Abschnitt 4.1.

Durch falsch detektierte Korrespondenzen wird eine nicht zu vernachlässigende Zahl von Ausreißern produziert. In diesem Fall liefert ein klassisches Ähnlichkeitsmaß keine zuverlässigen Ergebnisse. Bessere Ergebnisse lassen sich mit dem Verfahren von

[6] Eine blickrichtungsabhängige Verdeckung sei zu betrachten als eine Region, die in dem Bild der einen Kamera sichtbar ist, während sie im anderen Bild nicht abgebildet wird. Ursachen hierfür seien sich selbst verdeckende Objekte, oder Informationen die zwar in dem Sichtbereich der einen, jedoch nicht in dem der anderen Kamera liegen.

Birchfield und Tomasi erzielen (vgl. Abschnitt 4.3), das etwas unanfälliger gegenüber homogenen Regionen ist.

4.5 Zusammenfassung

In diesem Kapitel wurde auf Methoden zur Ermittlung von Korrespondenz-Punktpaaren eingegangen und die wesentlichen Herausforderungen erläutert. Durch die Stereo- / Korrespondenzanalyse kann die Disparität für korrespondierende Bildpunkte ermittelt werden, die Abbildungen des gleichen 3D Punktes sind. Im Bereich der Bildsynthese, zu dem sich auch diese Diplomarbeit zuordnen lässt, werden im Allgemeinen Disparitätskarten von hoher örtlicher Auflösung benötigt. Deshalb werden meist rechenintensive pixelbasierte Verfahren verwendet. Mit dem Birchfield-Stereoalgorithmus wurde eine schnelle Alternative zu den pixelbasierten Verfahren vorgestellt. Wird eine schnelle und robuste Bestimmung von wenigen Punktkorrespondenzen benötigt, sind die merkmalsbasierten Verfahren sinnvoller. Dies bietet sich beispielsweise für die Navigation von mobilen Robotern an.

5 Hardware und Software

Diese Arbeit wurde am Arbeitsbereich TAMS (Technische Aspekte Multimodaler Systeme) der Universität Hamburg in der Fakultät für Mathematik, Informatik und Naturwissenschaften erstellt. Es soll im Folgenden kurz darauf eingegangen werden, welche Hard- und Software zur Ausarbeitung dieser Diplomarbeit zur Verfügung stand.

5.1 Hardware

Der Service-Roboter TASER (siehe Abbildung 5.2) ist aus verschiedenen Standardkomponenten zusammengesetzt. Das Robotersystem besteht aus einer mobilen Plattform, einem Roboterarm, einer Dreifingerhand und diversen Kamerasystemen. Als aktive Sichtsysteme stehen ein Stereo-Sicht-System, eine omnidirektionale Kamera, sowie eine an dem Manipulator installierte Mikro-Kopf Kamera zur Verfügung. Informationen über das Robotersystem können aus dem Paper [WSZ06] entnommen werden.

Die *mobile Plattform* ist eine modifizierte MP-L-655 der Firma NEOBOTIX und ausgestattet mit einem Differentialantrieb, Rad-Encodern, zwei SICK-Lasermesssysteme und einem Gyroskop. Durch die Modifikation kann die Plattform mit zwei Roboterarmen ausgestattet werden. Der *Manipulator* besteht aus einem Roboterarm PA10-6C der Firma Mitsubishi Heavy Industries, hat einen Operationsradius von etwa einem Meter und bietet sechs Freiheitsgrade[1]. Eine Übersicht aller Achsen und Koordinatensysteme des Roboterarms ist in Abbildung 5.1 zu sehen. Als *Greifwerkzeug* ist am Manipulatorende eine Dreifingerhand von Barrett Technologies® Inc. mit vier Freiheitsgraden montiert. Ausgestattet mit TorqueSwitch™ Mechanismen[2], ist die Beweglichkeit der Finger ähnlich der menschlichen mit ihren Sehnen. Zwei der drei Finger sind über ein gemeinsames Winkelgelenk miteinander verbunden.

An der Handfläche der BarrettHand ist eine *Mikro-Kopf Farbkamera* montiert. In dieser Arbeit wird die Bildakquisition über diese Kamera bewerkstelligt, daher wird

[1] Der menschliche Arm hat im Vergleich fünf Freiheitsgrade.
[2] Der TorqueSwitch™ Mechanismus verbindet zwei Gelenke je Finger miteinander.

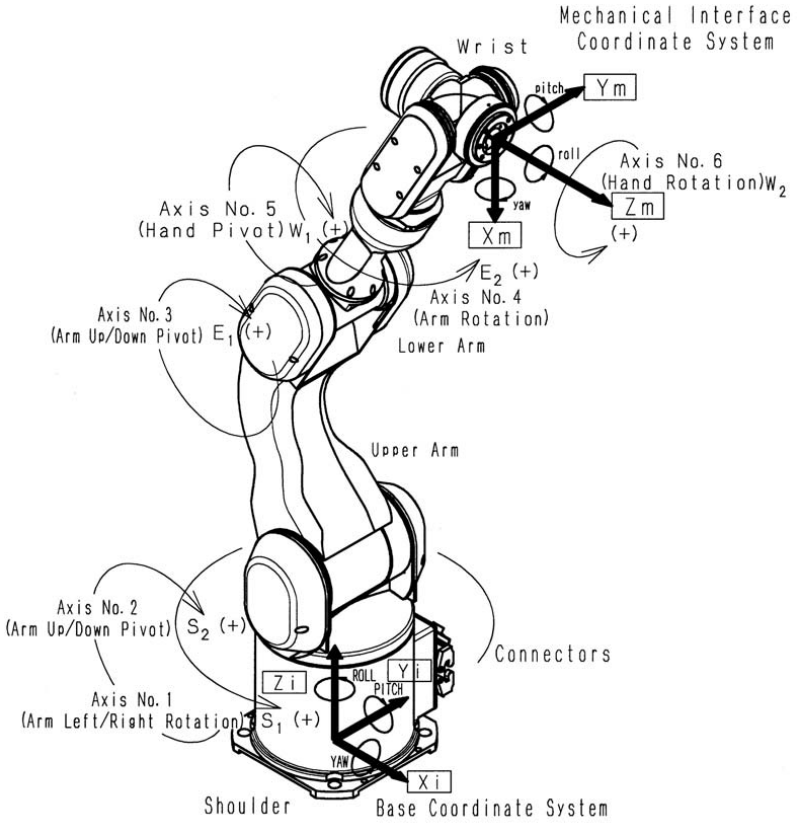


Abbildung 5.1: Übersicht aller sechs Achsen und Koordinatensysteme des PA10-6C Roboterarms der Firma Mitsubishi Heavy Industries. Quelle: [MHI].

im Folgenden nur die Spezifikation dieses Sichtsystem genauer erläutert. Einige spezifische Merkmale der Mikro-Kopf Kamera der Firma jAi[3] und der verwendeten Linse JK-L7.5M von Toshiba[4] sind in der Tabelle 5.1 dargestellt. Weitere Sichtsysteme bilden der *Stereokopf* aus zwei Sony DFW-VL500 Firewire-Digitalkameras mit 12x Zoom auf einer Pan-Tilt-Einheit von DirectPerception sowie ein *Omnivisionsystem* mit Sony DFW-SX900 Firewire-Digitalkamera und hyperbolischem Spiegel für Panoramaaufnahmen.

Die Handkamera ist mit einem Aluprofil an der Basis der Hand justiert (vergleiche Abbildung 5.3). Die Genauigkeit und Stabilität dieser Konstruktion hat einen nicht unerheblichen Einfluss auf die Qualität des entwickelten Rekonstruktionssystems.

[3] http://www.jai.com
[4] http://www.toshiba.ch/ics/

Kamera:	jAi CV-M2250
CCD-Chip:	1/2“, Farbe
Aufnahmesystem:	PAL: 625 Linien bei 25 Bilder/Sekunde
Größe Sensorelement:	8,6 μm × 8,3 μm
Effektive Pixel:	752 horizontal × 582 vertikal
Größe des Kamerakopfes:	17 mm × 99 mm
Gewicht des Kamerakopfes:	5 g-7 g
Linse:	Toshiba JK-L7,5M
Fokallänge:	7,5 mm
Winkel:	48° horizontal und 36° vertikal
Blende:	F = 1.6

Tabelle 5.1: Spezifikation der Kamera CV-M2250 von jAi.

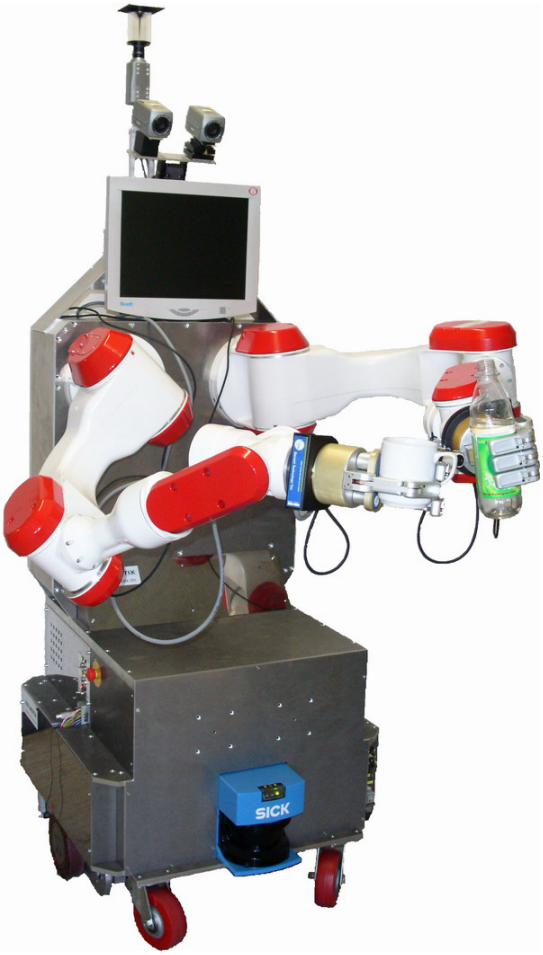


Abbildung 5.2: Der Roboter TASER des Arbeitsbereich TAMS in der angestrebten finalen Ausbaustufe mit zwei Armen.

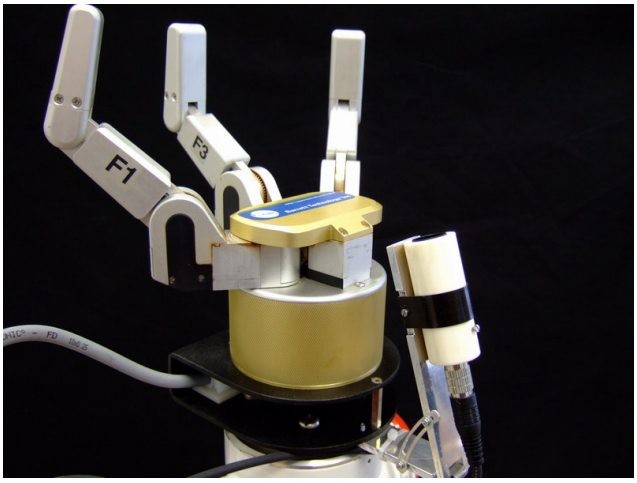


Abbildung 5.3: TASER's BarrettHand mit Konstruktionsaufbau der Mikro-Kopf Kamera montiert an der Manipulatorbasis.

5.2 Software

Zur Steuerung des Manipulators dient die *Robot Control C Library* (kurz RCCL) entwickelt von V. Hayward und J. Lloyd (1986). Ende der Achtziger wurde die Bibliothek erweitert, um mehrere Roboter steuern und Multiprozessormaschinen nutzen zu können. Dieses C Paket bietet die Möglichkeit, zielgerichtete Roboteranwendungen unter UNIX zu implementieren, ohne sich um die Trajektorie des Armes kümmern zu müssen[LH92].

Für die Verknüpfung des Armes und der Hand wurde eigens eine Bibliothek[5] am Arbeitsbereich TAMS entwickelt. Sie bietet leistungfähige Funktionalitäten für das komplette Manipulationssystem. Diese Funktionalitäten umfassen kartesische Bewegungen in unterschiedlichsten Koordinatensystemen, Steuerung der Bildakquise der Handkamera sowie verschiedenste Griffe mit der BarrettHand.

Über die Software ist es möglich, das Kamerakoordinatensystem der Handkamera im Weltkoordinatensystem des Roboters zu verschieben. Anders ausgedrückt kann eine Anweisung verfasst werden, die besagt, dass sich das Kamerazentrum zu einem gewünschten Weltpunkt bewegen soll.

[5]Die ZRobot-Bibliothek, entwickelt von Tim Baier.

6 Experimentelle Ergebnisse

Ziel dieser Arbeit sollte es sein, die Tiefeninformation aus einer alltäglichen Tischszene durch ein Stereobildpaar zurück zu gewinnen. Aufgenommen wurde das Bildpaar aus verschiedenen Blickrichtungen mittels der Mikro-Kopf Handkamera des Serviceroboter TASER (vgl. Kapitel 5.1). Parallel zu den Bildern wird die Aufnahmeposition und Orientierung des Kamerazentrums[1] in Weltkoordinaten des Roboters gespeichert um in späteren Verarbeitungschritten diese Information in Betracht ziehen zu können.

Abbildung 6.1 zeigt das Flussdiagramm des entwickelten Rekonstruktionssystems. Das in die Entzerrung einfließende Kameramodell wird zuvor offline berechnet mittels Tsai-Algorithmus (vgl. Abschnitt 2.2). Die Bilder werden parallel entzerrt und die erforderlichen Merkmale berechnet. Erst nach der Selektion von 10 korrespondierenden Punktpaaren werden die Informationen beider Eingabebilder zusammengefügt. Dieses Kapitel beschreibt alle Prozesse des Flussdiagramms anhand experimenteller Ergebnisse.

Die folgenden Abschnitte erläutern genauer die einzelnen Arbeitsstufen des entwickelten Rekonstruktionssystems und die experimentellen Ergebnisse dieser Arbeit. In Abschnitt 6.1 werden die Kalibrationsergebnisse vorgestellt. Der Basisdatensatz, die zurgrundeliegende Tischszene, wird in Abschnitt 6.2 vorgestellt. Abschnitt 6.3 erläutert die Merkmalsextraktion und die darauf aufbauende Berechnung der Fundamentalmatrix in Abschnitt 6.4. Die Rektifikation der Bilder mittels der Fundamentalmatrix wird in Abschnitt 6.5 beschrieben. Nach der Präsentation der Ergebnisse des genutzten Stereoalgorithmus in Abschnitt 6.6 wird in Abschnitt 6.7 eine ausführliche Analyse des Rekonstruktionssystems vorgestellt. Eine mögliche Übertragbarkeit des entwickelten Rekonstruktionssystems auf andere Aufgabenbereiche wird abschließend in Abschnitt 6.8 thematisiert.

[1] Die Lage des Kamerazentrums wird dabei durch eine homogene Transformation, ausgedrückt durch eine 3×4 Matrix, beschrieben. Wobei die Orientierung $\mathbf{R}$ durch die ersten drei Zeilen und drei Spalten und die Translation $\mathbf{T}$ durch die vierte Spalte beschrieben wird. Die TTransform wird von der RCCL-Steuerung bereit gestellt.

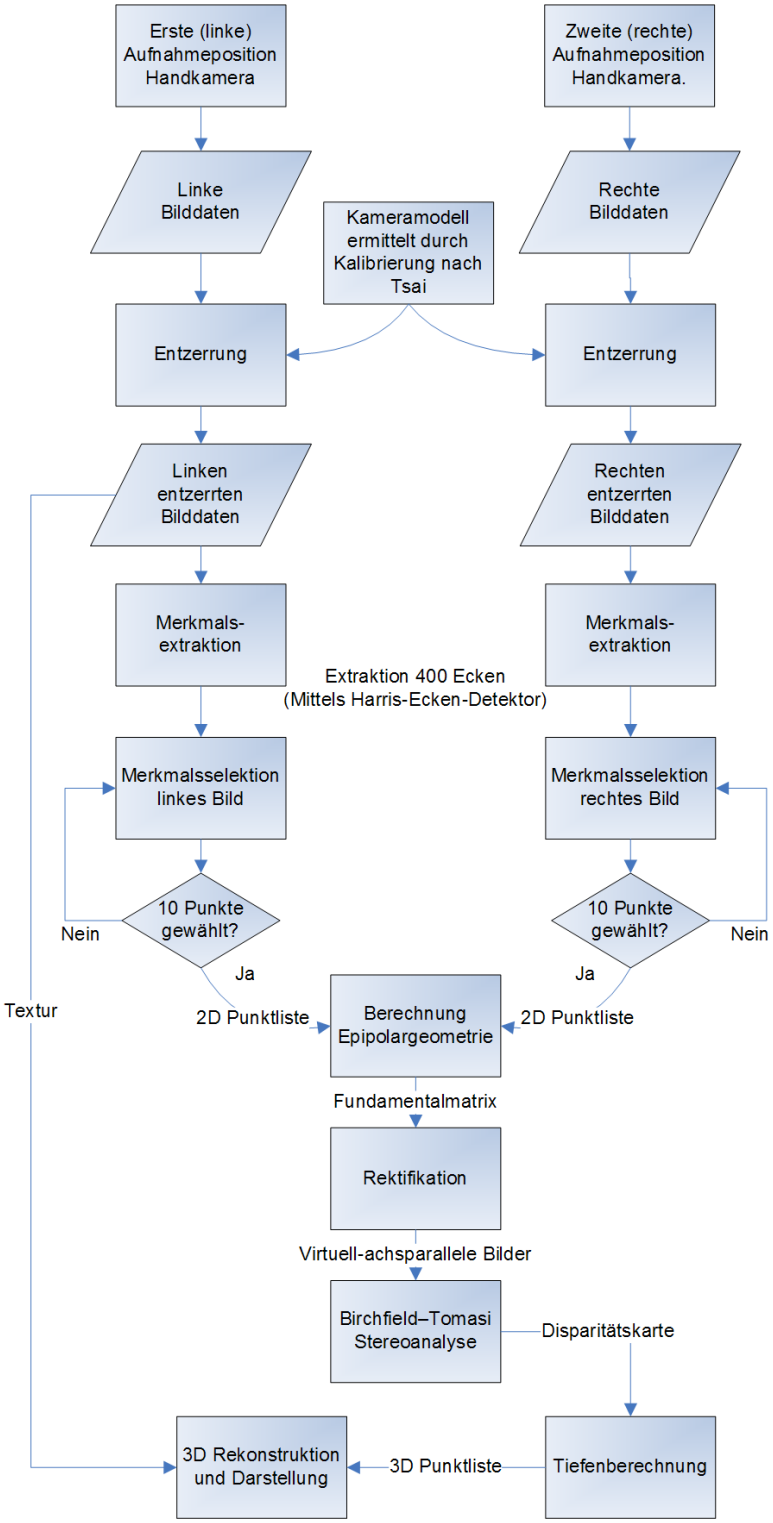


Abbildung 6.1: Flussdiagramm des Rekonstruktionssystems.

6.1 Kalibrierung

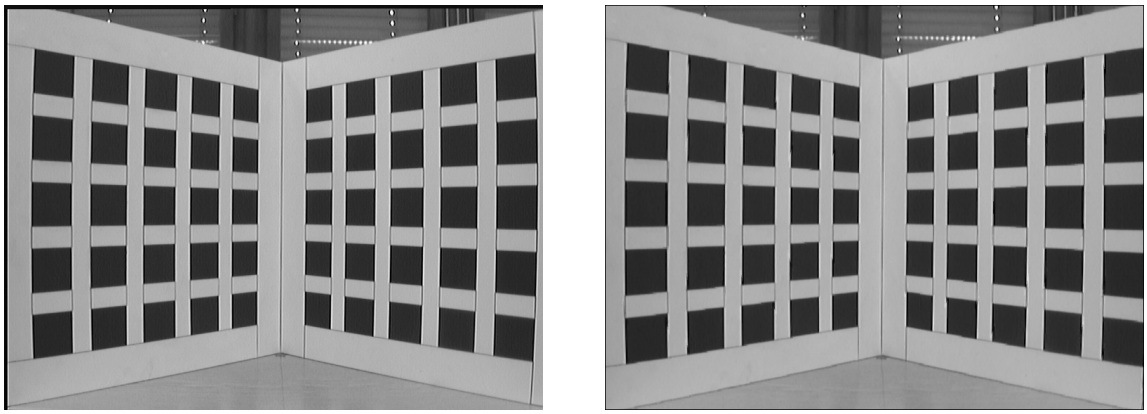
Vorab wurde die Mikro-Kopf Handkamera fest an dem Manipulator des Serviceroboters TASER installiert und mit dem in Abschnitt 2.2 vorgestellten Tsai-Algorithmus kalibriert. Dabei ergaben sich die in Tabelle 6.1 aufgeführten Parameter für die Handkamera.

f:	8,673491 mm	u_0:	384,00 $pixel$
κ:	-0,000776 $\frac{1}{mm^2}$	v_0:	288,00 $pixel$
s_x:	1,026295	normlisierter Fehler:	12,068445

Tabelle 6.1: Relevante Kalibrationsparamter nach Tsai's Algorithmus [Tsa87], berechnet für die jAi Kamera.

Die Nutzung des Kalibrationsverfahrens von Tsai hat gegenüber dem Kalibrationsverfahren von Pollefeys [Pol99] den Vorteil höherer Robustheit, wenngleich die Verwendung des Kamera-Zooms ausscheidet. Da sich jedoch diese Arbeit auf Innenaufnahmen beschränkt und die Kamera zudem fest in einer Gehäusehalterung montiert ist, stellt die einmalige Kalibrierung keine wesentliche Einschränkung dar.

Abbildung 6.2(a) stellt die Originalaufnahme des Kalibrationskörpers dar. Im Randbereich ist die radiale Verzerrung ersichtlich, die eigentlichen Geraden des Kalibrationskörpers unterliegen einer konvexen Deformation. Abbildung 6.2(b) zeigt das resultierende entzerrte Bild. Das verzerrte Bild wird dabei mittels der berechneten Parametern aus dem Tsai Algorithmus über die Gl. 2.11 aus Abschnitt 2.1.2 entzerrt, wobei der zweite Term $\kappa_2 r^4$ vernachlässigt werden kann [Tsa86] (vgl. Abschnitt 2.2).



(a) Verzerrte Ansicht (b) Entzerrte Ansicht

Abbildung 6.2: Kalibrationskörper vor und nach der Bereinigung von Linsenverzeichnung.

Das Ergebnis der Entzerrung ist hinreichend genau, wenngleich die Entzerrung in den Bildecken immer noch leicht zu erkennen ist. Wie aber schon in Kapitel 2.2 thematisiert, ist das RAC nur gültig wenn deutliche radiale Verzerrungen vorliegen. Der Grad der Verzerrung ist bei dieser Kamera jedoch noch recht gering und wird beschrieben durch den Parameter κ aus Tabelle 6.1.

6.2 Original Szene

Wie schon in der Einleitung dieses Kapitels erwähnt, werden die Bilddaten über die in Abschnitt 5.1 beschriebenen Mikro-Kopf Kamera am Arm des Service Roboters TASER aquiriert. Abbildung 6.3 zeigt die noch verzerrten Eingangsdaten einer alltäglichen Tischszene.

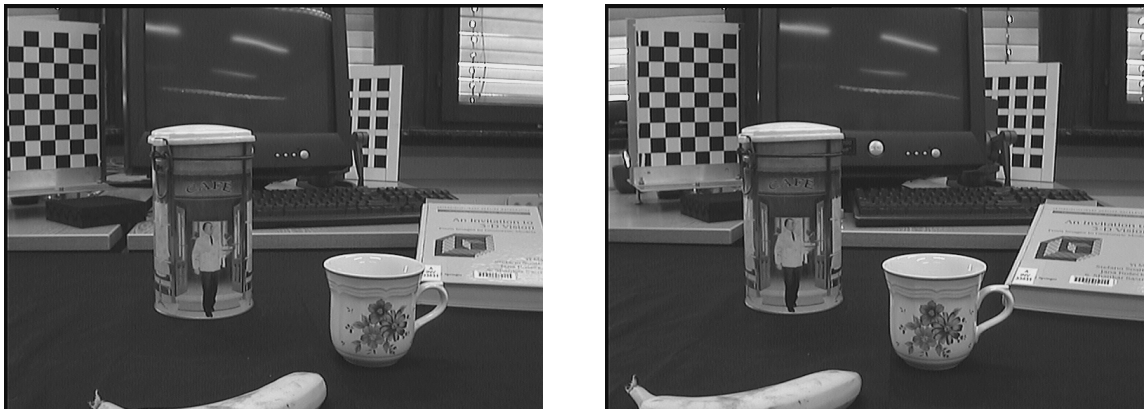


Abbildung 6.3: Linkes und rechtes Eingabebild in Rohfassung (verzerrt), aquiriert durch die Mikro-Kopf Kamera am Roboter Arm.

Dabei ist die Kamera mittels des Roboter Armes auf einer Höhe von 90cm zwischen der ersten und zweiten Aufnahme um 10cm horizontal transliert worden. Für beide Aufnahmepositionen hat die Kamera einen Pitch-Winkel von 8° und einen Yaw-Winkel von 8° zueinander. Für die Kamera gilt ein rechtshändiges Koordinatensystem mit Ursprung im optischen Zentrum. Die z-Achse verläuft entlang des optischen Strahls, die x-Achse ist die vertikale Achse und folglich entspricht die y-Achse der horizontalen Achse. Um Bildrauschen zu minimieren werden an beiden Aufnahmeposition jeweils 10 Bilder aufgenommen und ein Mittel aus den 10 Aufnahmen gebildet.

Die Eingangsbilder werden nach der Aufnahme mit den in Abschnitt 6.1 aufgeführten Parametern entzerrt und stehen für die weiteren Bearbeitungsschritte nahezu verzeichnisfrei zur Verfügung.

6.3 Merkmalsextraktion

Der nächste Verarbeitungsschritt besteht in der Ermittlung von Merkmalen in den Eingabebildern. In dieser Arbeit werden Ecken als Merkmale genutzt und können auf zweierlei Arten berechnet werden. Die Funktion `cvGoodFeaturesToTrack` der OpenCV Bibliothek von Intel® (vgl. Anhang C) findet Ecken die sich besonders gut verfolgen lassen. Das sind bespielsweise Ecken mit besonders großen Eigenwerten. Dabei berechnet die Funktion für jedes Pixel die Kovarianzmatrix[2] $\mathbf{M}$ und speichert in einem separaten Bild lediglich den minimalen Eigenwert von $\mathbf{M}$ für das jeweilige Pixel. Darauf wird eine non-maxima Unterdrückung angewendet, so dass nur die Maxima in einer 3×3 Pixel Umgebung übrig bleiben. Nachfolgend werden die Ecken mit kleineren Eigenwerten als ein bestimmter Schwellwert entfernt. Des weiteren kann ein minimaler Merkmalsabstand gewählt werden. Dadurch werden alle Merkmale entfernt, die in dem angegebenen Umkreis eines Merkmals mit größerem Eigenwert liegen [ST94].

Diese Funktion kann auch mit dem Harris-Operator [HS88] aus Gl. B.3, Kapitel B.2, genutzt werden. Bei beiden Varianten kann vorab gewählt werden, wieviele Ecken durch die Kantendetektion erkannt werden sollen. Abbildung 6.4 zeigt 400 mittels Harris-Operator detektierte Ecken in den Eingabebildern. Tests ergaben, dass die Anzahl zu entdeckender Eckpunkte den empirischen Wert 400 nicht unterschreiten sollte, damit in beiden Eingabebildern genug reale korrespondierende Merkmalspunkte während der nächsten Verarbeitungschritte des Rekonstruktionssystems gewählt werden können.

Abbildung 6.4: Ergebnis der Merkmalsextraktion. Die Ecken, detektiert mittels Harris-Operator, liegen teilweise etwas neben den tatsächlichen Ecken.

[2]Entsprechend der Kovarianzmatrix aus Gl. B.2, Kapitel B.2.

6.4 Merkmalsselektion und Ermittlung der Fundamentalmatrix

Der Nutzer muss nun interaktiv 10 korrespondierende Eckpunkte in beiden Bildern mit der Maus selektieren (vgl. Abbildung 6.5). Die Punkte können wechselseitig oder in einer kompletten Sequenz selektiert werden. Wichtig ist jedoch, darauf zu achten, dass die Punkte wirklich korrespondierende Punkte darstellen und in gleicher Reihenfolge von 1 bis 10 vom Nutzer selektiert werden. Dies ist notwendig, weil die nachfolgende Berechnung der Fundamentalmatrix eine geordnete Korrespondenzpunktmenge benötigt.

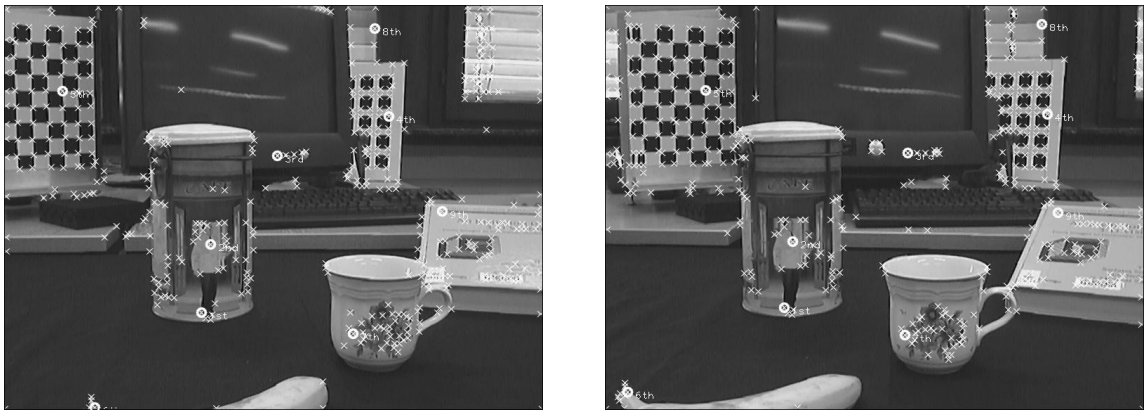


Abbildung 6.5: Durch den Nutzer interaktiv markierte Ecken (durch Kreise gekennzeichnet) für die Berechnung der Fundamentalmatrix. Dabei sei zu beachten, dass möglichst von jeder Tiefenebene Punkte markiert werden. Ansonsten kommt es überwiegend zu einer fehlerhaften Fundamentalmatrix.

Während der experimentellen Erprobung stellte sich heraus, dass die 10 Merkmalspunkte möglichst gleichmäßig über die verschiedenen Tiefenebenen, über das gesamte Bild verteilt, gewählt werden sollten. Andernfalls kann die Berechnung der Fundamentalmatrix fehl schlagen. In Abschnitt 6.3 wurde schon erwähnt, dass mindestens 400 Ecken detektiert werden sollten, andernfalls werden auf einigen Tiefenebenen keine Eckpunkte dargeboten und eine möglichst gleichverteilte Auswahl von Korrespondenzpunkten über alle Tiefenebenen ist folglich nicht möglich.

Die Fundamentalmatrix wird über die Funktion `cvFindFundamentalMat` berechnet. Hierbei stehen dem Nutzer 3 Varianten zur Berechnung der Fundamentalmatrix zur Verfügung. Es kann gewählt werden zwischen dem linearen Berechnungsverfahren, dem 8-Punkt-Algorithmus, oder zwischen den nichtlinearen Verfahren RANSAC oder LMedS (vgl. Abschnitte 3.1.3, 3.1.4). Dabei benötigen alle Verfahren acht oder mehr

Korrespondenzen. Der 8-Punkt-Algorithmus stellte sich bei der Wahl von 10 Korrespondenzpunkten als zuverlässigste Berechnungsvariante heraus[3]. Die Verfahren RANSAC und LMedS berechnen erst ab 22 vorgegebenen Korrespondenzen überhaupt eine Fundamentalmatrix. Tests, bei denen ungeordnete Punktmengen von bis zu 1200 Ecken den Methoden RANSAC und LMedS zur Verfügung gestellt wurden, ergaben trotzdem keine Fundamentalmatrix, die der Qualität der Fundamentalmatrix aus dem 8-Punkt-Algorithmus entprachen. Dies stellt den Grund dar, warum weiterhin eine Nutzer-Interaktion zur Selektion von 10 Korrespondenzpunkten gefordert wird.

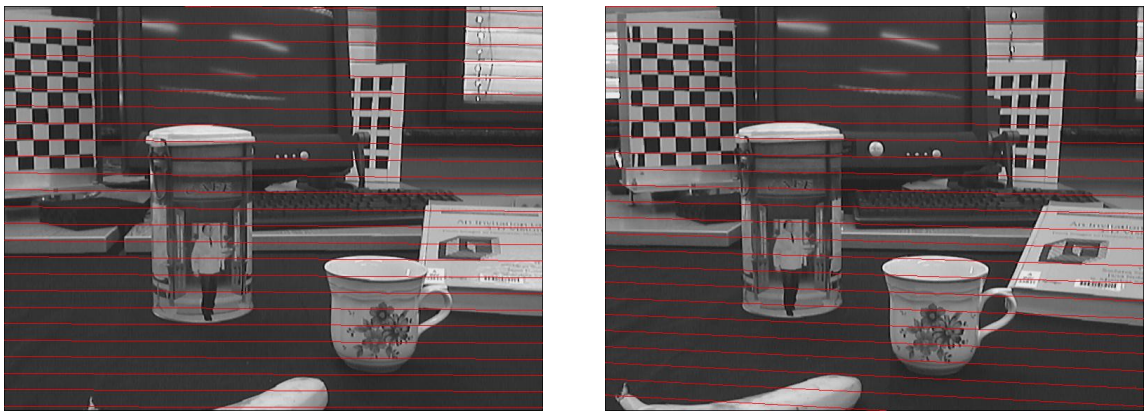


Abbildung 6.6: Visualisierung der Epipolarlinien.

Die Abbildung 6.6 zeigt die Epipolarlinien in beiden Bildern, die durch Gl. 3.8 aus Abschnitt 3.1.1 berechnet werden können. Bei genauer Betrachtung wird ersichtlich, dass einige Epipolarlinen nicht genau die gleiche Gerade in beiden Bildern abdecken, sondern einen leichten Versatz aufweisen. Die ermittelte Fundamentalmatrix lautet:

$$\mathbf{F} = \begin{pmatrix} -0.00000030 & 0.00002216 & -0.00119938 \\ 0.00000437 & 0.00000774 & -0.22008468 \\ -0.00309049 & 0.20801485 & 1.00000000 \end{pmatrix} \tag{6.1}$$

6.5 Rektifikation

Die Rektifikation (vgl. Abschnitt 3.3) stellt einen wesentlichen Schritt der Tiefeninformationsgewinnung stereogeometrischer Bildpaare dar. Um die Eingabebilder in eine

[3]Bei genau acht Korrespondenzpunkten ist es auch bei diesem Verfahren schwer eine passende Fundamentalmatrix zu ermitteln.

achsparallele Geometrie (vgl. Abschnitt 3.2) überführen zu können, bildet die Fundamentalmatrix aus dem vorherigen Abschnitt 6.4 den elementaren Grundbaustein. Anhand der Fundamentalmatrix können die Transformationsmatrizen (vgl. Abschnitt 3.3.1) berechnet werden, mit deren Hilfe die Eingabebilder so verzerrt werden, dass sie einem zeilenweise arbeitenden Disparitätenalgorithmus übergeben werden können.

Die Epipolarlinien werden dabei horizontal so ausgerichtet, dass sie im linken wie auch im rechten rektifizierten Bild auf der selben vertikalen Bildkoordinate verlaufen. Dies ist für die Disparitätenalgorithmen zwingend erforderlich.

Abbildung 6.7: Die neu ermittelten virtuellen Kameraansichten aus der Szene nach der Rektifikation. Korrespondierende Bildzeilen befinden sich nun auf der selben Höhe.

Abbildung 6.7 stellt die neu gewonnen virtuellen achsparallelen Kameraansichten dar. Je nach Qualität der Fundamentalmatrix ist die Rektifikation mehr oder weniger genau. Tests ergaben, dass bei einer Rektifikation die korrespondierenden Bildzeilen meist nicht weiter als 1-4 Pixel auseinander lagen.

6.6 Stereoanalyse und Rekonstruktion

Für die Stereoanalyse nach Birchfield und Tomasi [BT98] wird die OpenCV Funktion `cvFindStereoCorrespondance` genutzt. Als Eingabe erhält der Algorithmus das linke und rechte entzerrte, rektifizierte Bild[4]. Zudem können die Belohnungs- und Straffaktoren der Kostenfunktion (vgl. Abschnitt 4.3, Gl. 4.7) vom Nutzer gewählt

[4]Das in dieser Arbeit entwickelte Rekonstruktionssystem ist noch zweigeteilt. Die Stereoanalyse und 3D Rekonstruktion wurde auf achsparallele Aufnahmen angewendet und nicht mit den rektifizierten Bildern verknüpft. Der Grund dafür ist, dass je nach Fundamentalmatrix die rektifizierten Bilder gespiegelt und/oder rotiert werden. Dies erforderte eine Nutzerinteraktion um

Parameter	Wert
Konstanter Verdeckungs-Malus:	25
Bonus korrekter Korrespondenzen:	5
Malus stark vertrauenswürdiger Regionen im Intervall:	12
Malus moderat vertrauenswürdiger Regionen im Intervall:	15
Malus schwach vertrauenswürdiger Regionen im Intervall:	25

Tabelle 6.2: Parameter der Kostenfunktion für den Birchfield Algorithmus

werden. In dieser Arbeit stellten sich die in Tabelle 6.2 beschriebenen Werte für die Parameter des Birchfield-Algorithmus als sinnvoll heraus. Diese Werte entsprechen denen aus [BT96]. Ferner kann die maximal auftretende Disparität im Intervall $[0, 255]$ angegeben werden.

Das Ergebnis des Algorithmus ist in Abbildung 6.8(a) als Grauwertbild zu sehen. An den Objektkonturen ist ein deutliches Verwischen mit dem Hintergrund zu erkennen. Das liegt daran, dass der Algorithmus wirklich nur Zeilen der selben vertikalen Bildkoordinate miteinander vergleicht und erst in späteren Schritten die umliegenden Zeilen berücksichtigt. Um diese Artefakte etwas zu verringern und kontinuierlichere Grauwertverläufe zu erhalten, sowie kleine Löcher zu füllen, wird das Grauwertbild anschließend mit einem kantenerhaltenden Medianfilter überarbeitet. Für den Medianfilter hatte sich eine Fenstergröße von 5×5 Pixeln in dieser Arbeit bewährt. Das Ergebnis ist in Abbildung 6.8(b) dargestellt.

In der Disparitätskarte wird die Disparität zweier korrespondierender Punkte durch unterschiedliche Grauwerte beschrieben. Ein Pixel mit einem Grauwert von 40 beschreibt eine Disparität des korrespondierenden Pixels von 40 Pixeln bezüglich der anderen Ansicht. Objekte, die sich näher an der Kamera befinden, haben eine große Disparität und somit einen höheren Grauwert, während entferntere Objekte wegen ihrer geringeren Disparität einen niedrigeren Grauwert erhalten (siehe Abschnitt 3.2). Je heller also das Objekt im Grauwertbild, desto näher ist es der Kamera.

Im Vergleich zu ersten Tests, das Korrespondenzproblem mit Hilfe der in Kapitel 4.1 erwähnten Ähnlichkeitsmaße zu lösen, ist die Schnelligkeit des Birchfield-Algorithmus bei weitem nicht zu erreichen. Der jeweilige Vergleich von Fenstern um das betrachtete Pixel herum kostet bei den Methoden SAD, SSD und NCC viel Zeit und viele Ressourcen.

Ziel ist es, aus dem Disparitätsbild eine räumliche, dreidimensionale Rekonstruktion der abgebildeten Szene zu erlangen. Der Zusammenhang zwischen 3D Kamerakoor-

die Bilder wieder in die richtige Lage zu spiegeln und zu rotieren. Die Komponenten können jedoch theoretisch miteinander verknüpft werden.

(a) Ergebnis des Birchfield-Algorithmus. Das Grauwertbild enthält deutlich fälschliche Ausfransungen der Objektkonturen.

(b) Um ein kontinuierlicheres Disparitätsbild zu erhalten, wird die eigentliche Tiefenkarte mit einem Median der Größe 5×5 Pixel geglättet.

Abbildung 6.8: Mittels Birchfield-Tomasis-Stereoalgorithmus [BT98] ermittelte Disparitätskarten.

dinaten und Bildkoordinaten wurde schon in Abschnitt 2.1.1 ausführlich erläutert. Nimmt man sich nun erneut die Zentralprojektions-Gleichung 2.2 und formt diese für die jeweilige Kamera um, erhält man die perspektivischen Abbildungen der linken (Gl. 6.2) und rechten (Gl. 6.3) Kamera.

$$\begin{pmatrix} u_1 \\ v_1 \end{pmatrix} = \frac{f}{z_c} \begin{pmatrix} x_c \\ y_c \end{pmatrix} \tag{6.2}$$

$$\begin{pmatrix} u_2 \\ v_2 \end{pmatrix} = \frac{f}{z_c} \begin{pmatrix} x_c + b \\ y_c \end{pmatrix} \tag{6.3}$$

Wegen der Betrachtung der selben vertikalen Koordinate v ergibt sich für die Disparität d (vgl. Abschnitt 3.2) folgende Beziehung:

$$d = u_2 - u_1 = (\frac{fx_c}{z_c} + \frac{fb}{z_c}) - \frac{fx_c}{z_c} = \frac{fb}{z_c} \tag{6.4}$$

Formt man nun Gl. 6.2, Gl. 6.3 und Gl. 6.4 um, lässt sich aus der resultierenden Gl. 6.5 die 3D Kamerakoordinate $\mathbf{m'}_1$ für $\mathbf{m}_1$ rekonstruieren.

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \frac{b}{d} \begin{pmatrix} u_1 \\ v_1 \\ f \end{pmatrix} \tag{6.5}$$

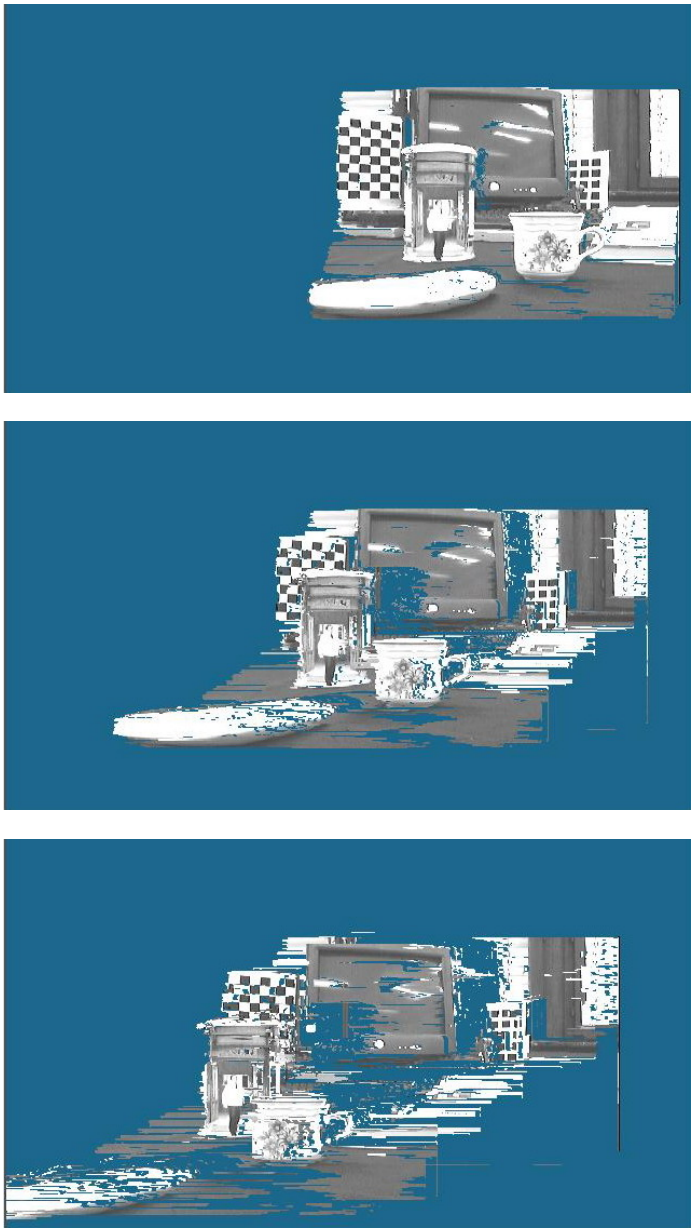


Abbildung 6.9: Aus der Disparitätskarte rekonstruiertes 3D Modell der Szene aus verschiedenen virtuellen Ansichten. Dargestellt wird die Szene als texturierte OpenGL Punktwolke.

Mittels Gl. 6.5 wird anschließend für jedes Pixel der Disparitätskarte ein Raumvektor ermittelt. Visualisiert werden all diese 3D Raumpunkte durch OpenGL, wobei jeder Raumpunkt als Textur den jeweiligen Grauwert des linken, entzerrten Originalbildes erhält. Das Resultat ist in Abbildung 6.9 ersichtlich. Der Nutzer kann die 3D Szene durch interaktive Maussteuerung aus beliebigen neuen Blickwinkeln betrachten. Die größeren untexturierten Bereiche in Abbildung 6.9(b) und 6.9(c), etwa in der lin-

ken unteren Ecke des Monitors oder der Fensterfront, entstehen durch Verdeckungen anderer Objekte.

6.7 Analyse

Allgemein gilt für bildbasierte dreidimensionale Rekonstruktionssysteme, dass sich bei einem aufwendigen Prozess wie der Tiefeninformationsgewinnung die Fehler aus jedem einzelnen Verarbeitungsschritt immer weiter aufsummieren. Das lässt sich leicht erklären. Es beginnt schon bei der Entzerrung der Bilder. Werden Bilder nicht richtig entzerrt, so ergibt sich eine nicht wirklich euklidische Dimensionalität. In dieser Arbeit wurden mit nur einer Kamera beide Bilder spatial und temporal divergierend aufgenommen und unterlagen somit den selben Abbildungseigenschaften (vgl. Abschnitt 2.1.1, Gl. 2.7). Nicht korrekt detektierte Merkmalspunkte können zu einer falschen Fundamentalmatrix führen. Die Fundmentalmatrix ist zwar für die Menge der detektierten Merkmale korrekt, aber beschreibt die Realität nicht exakt. Eine falsche Fundamentalmatrix führt zudem zu einem Rektifikationsergebnis, bei dem die korrespondierenden Bildzeilen nicht auf die selbe vertikale Bildkoordinate abgebildet werden und leicht divergieren. Das führt zu Problemen bei der zeilenweisen Korrespondenzanalyse und verursacht unter anderem die in den Disparitätskarten ersichtlichen zerlaufenden Objektkonturen. Durch diese fälschlichen Disparitätswerte werden folglich unkorrekte Tiefenwerte berechnet.

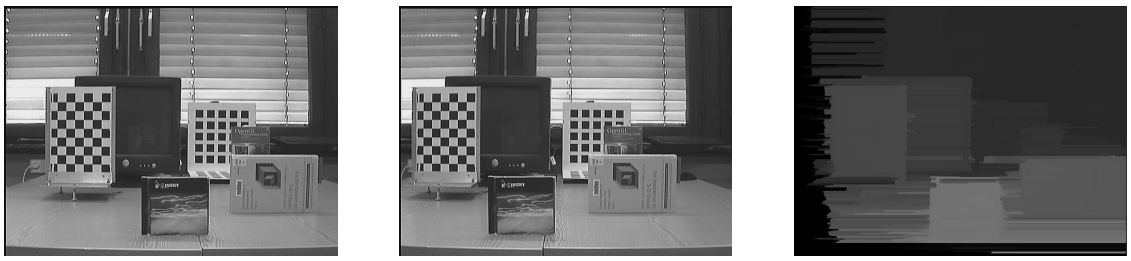
Im Rahmen dieser Arbeit wurde eine analytische Auswertung gemacht, um eine angemessene Basislinie für eine Tischszene und eine qualitative Bewertung der jeweils berechneten Disparitätsbilder des Stereoalgorithmus zu erlangen.

Versuchsaufbau Für die Analyse wurden auf einem Tisch sechs Objekte in gleichmäßigen Abständen orthogonal zur optischen Achse platziert. Die Objekte wurde so angeordnet, dass sie in beiden Kamerabildern zumindest partiell sichtbar sind. Der Abstand der Kamera zum vordersten Objekt beträgt 80cm. Der Abstand zwischen den Objekten jeweils 25cm. Das entfernteste Objekt ist somit 205cm von der Kamera entfernt. Das Intervall $[80cm, 205cm]$ stellt einen repräsentativen Tiefenbereich für die Vielfältigkeit von Tischszenarien dar. Die Kamera wurde vor der Szene mit verschiedenen Basislängen versetzt und das Disparitätsbild der Stereoanalyse gespeichert. Die Ergebnisse für die Basislinien 2.5cm, 5cm, 7.5cm und 10cm sind in Abbildung 6.10 zu sehen, während Abbildung 6.11 die Versuche für die Basislinien 15cm, 20cm, 25cm und 30cm zeigt. Die Kamera wurde bei dieser Versuchsreihe lediglich achsparallel verschoben, um das Verhalten des Stereoalgorithmus genauer zu untersuchen.

Eingabebilder mit 2.5cm Basislinie. Resultierende Tiefenkarte mit GVM 30 und DA 17.

Eingabebilder mit 5cm Basislinie. Resultierende Tiefenkarte mit GVM 57 und DA 34.

Eingabebilder mit 7.5cm Basislinie. Resultierende Tiefenkarte mit GVM 85 und DA 51.



Eingabebilder mit 10cm Basislinie. Resultierende Tiefenkarte mit GVM 113 und DA 68.

Abbildung 6.10: Vergleich der verschiedenen Basisabstände der Kamerazentren und der resultierenden Tiefenkarte mit angegebener Disparitätsauflösung (DA) zwischen vorderstem und hinterstem Objekt, sowie dem Grauwert des vordersten Objektes (GVM). Basislinien sind 2.5cm, 5cm, 7.5cm, 10cm.

Ergebnis Betrachtet man die Disparitätsbilder aller Basislinien aus Abbildung 6.10 und 6.11, so lässt sich auf den ersten Blick ein enormer Qualitätsunterschied der Grauwertbilder feststellen. Bei einer Basisline von mehr als 25cm hat das vordere Objekt – die CD – mit einer realen Disparität von 275 Pixeln den maximal darstellbaren

Eingabebilder mit 15cm Basislinie. Resultierende Tiefenkarte mit GVM 169 und DA 101.

Eingabebilder mit 20cm Basislinie. Resultierende Tiefenkarte mit GVM 190 und DA 100.

Eingabebilder mit 25cm Basislinie. Resultierende Tiefenkarte mit GVM 255 und DA 142.

Eingabebilder mit 30cm Basislinie. Resultierende Tiefenkarte mit GVM 255 und DA 88.

Abbildung 6.11: Vergleich der verschiedenen Basisabstände der Kamerazentren und der resultierenden Tiefenkarte mit angegebener Disparitästauflösung (DA) zwischen vorderstem und hinterstem Objekt, sowie dem Grauwert des vordersten Objektes (GVM). Basislinien sind 15cm, 20cm, 25cm, 30cm.

Disparitätsraum überstiegen. Der genutzte Birchfield-Algorithmus erstellt ein Disparitätsbild mit einem Wertebereich von einem Byte pro Pixel. Folglich kann es maximal eine Disparität von 256 Pixeln abbilden. Die maximalen Grauwerte (Abk. GVM) sind in Abbildung 6.10 und 6.11 für jeden Einzeltest der Versuchsreihe aufgeführt. Schon

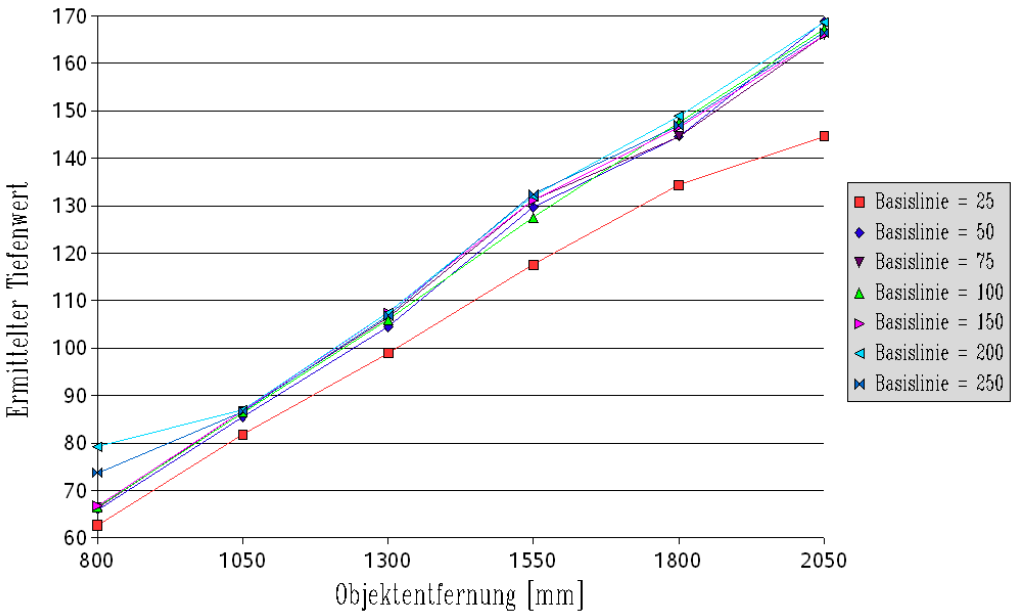


Abbildung 6.12: Relation der ermittelten Tiefenwerte zur Objektentfernung.

bei einer Basislinie von 20cm treten Fehlklassifikationen des Birchfield-Algorithmus auf, die zu unstetigen Grauwertverläufen der jeweiligen Objekte führen. Je größer die maximale Disparität der Eingabebilder des genutzten OpenCV Algorithmus zur Stereoanlyse, desto schlechter wird die Qualität der ermittelten Disparitätskarten. Somit ist für eine Rekonstruktion einer Szene in einer Entfernung von 80-205cm eine Basislinie von mehr als 15 cm wegen ihrer schlechten Ergebnisse ungeeignet. Wird die Funktion mit einer maximalen Disparität bis zu 170 Pixeln verwendet, liefert sie im allgemeinen gute und stetige Ergebnisse.

Betrachtet man die Disparitätsbilder mit niedrigen Basislinien, wie beispielsweise 2.5cm in Abbildung 6.10, sind die verschiedenen Objekte anhand der geringen Disparitätsauflösung (Abk. DA) von 17 Pixeln zwischen vordersten und hintersten Szenenobjekt kaum zu unterscheiden. Die Eingabebilder haben lediglich einen sehr geringen Versatz und führen allgemein zu geringen Disparitäten von maximal 30 Pixeln.

Die Tabelle 6.12 zeigt die Relation der wahren Objektentfernung zu den nach Gl. 6.5 ermittelten Tiefenwerten des Disparitätsbildes. Die meisten Basislinien weisen ein lineares Verhältnis zwischen Objektentfernung und ermittelter Tiefe auf und es kann die Annahme gemacht werden, dass sie sich bei größeren Objektentfernungen vorerst linear fortpflanzt. Die Basislinie mit 2.5cm weicht allgemein sehr von den anderen Basislinien ab und insbesondere bei zunehmender Tiefe wird die Abweichung immer größer. Wegen der geringen Disparitätsauflösung sind tiefere Objekte schlecht rekon-

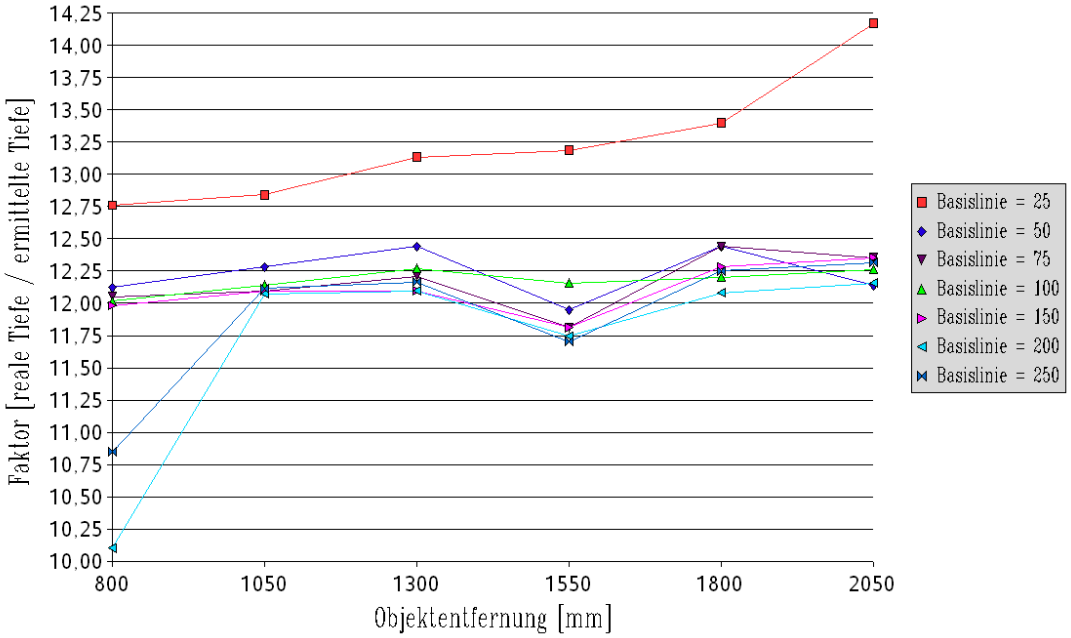


Abbildung 6.13: Multiplikationsfaktoren, um die ermittelten Objekttiefen in reale Objekttiefen zu überführen.

struierbar. Nach Kapitel 4.4 war dies jedoch zu erwarten. Bei einer großen Basisline über 15cm lässt sich ebenfalls in der Tabelle 6.12 ersehen, dass die ermittelten Tiefenwerte nicht ihrer linearen Eigenschaft folgen. Für nahe Objekte lassen sich aufgrund der unzulänglichen Relationen des Stereoalgorithmus die exakten Tiefenwerte nicht genau bestimmen. Die Basislinien von 5-15cm weisen alle ein sehr lineares Verhältnis zwischen realer Objekttiefe und errechnetem Tiefenwert auf, wobei die Basisline mit 10cm die deutlich linearste darstellt[5].

Wegen des linearen Verhältnisses zwischen realer Objekttiefe und berechneter Tiefe lässt sich ein Faktor bestimmen, der – multipliziert mit den berechneten Tiefenwerten – die reale Objekttiefe annähernd genau beschreibt. Die Division der realen Objekttiefe durch die ermittelte Tiefe ist in Tabelle 6.13 dargestellt. Auch hier zeigt sich, dass die 10cm Basislinie mit einem Streckungsfaktor zwischen 12.02 bis 12.27 und dem Mittelwert 12.17 den konstantesten Verlauf über alle untersuchten Tiefenbereiche aufweist. Auch in dieser Tabelle ist erkennbar, dass die Basislinie mit 5cm stark von der Masse divergiert, sowie Basislinien ab 20cm im Nahbereich stark abweichende Multiplikationsfaktoren für die Ermittlung der realen Objekttiefe ergeben. Der berechnete Streckungsfaktor ähnelt stark dem normalisierten Fehler des Tsai-Algorithmus aus

[5]Die Ausreißer fast aller Basislinien bei einer Objekttiefe von 155cm in Tabelle 6.12 und Tabelle 6.13 entstehen wahrscheinlich durch Spiegelungen der Mattscheide des Monitors.

Tabelle 6.1.

Für die Rekonstruktion einer Tischszene in einer ungefähren Entfernung von 80-205cm ist eine Basislinie von 10cm, bis hinauf auf 15cm, folglich der optimale Abstand zwischen den Kamerazentren. Bei einer Multiplikation jedes rekonstruierten Raumvektors mit dem Mittel des Streckungsfaktors für die Basislinie von 10cm ist die reale Objektentfernung mit nur geringer Abweichung bestimmbar[6].

6.8 Weitere Ergebnisse

Das Ziel dieser Arbeit, eine dreidimensionale Rekonstruktion einer alltäglichen Tischszene zu berechnen, ist somit erfolgreich absolviert worden. Es stellt sich jedoch die Frage, ob sich dieses System auch auf weitere Aufgabengebiete, wie beispielsweise Gesichtsrekonstruktion, Rekonstruktion von Objekten größerer Tiefe oder gar die Rekonstruktion ganzer Räume, übertragen lässt. Zwei dieser Aufgabengebiete wurden über den eigentlichen Rahmen dieser Arbeit hinaus etwas genauer betrachtet. Die Ergebnisse einer Gesichtsrekonstruktion sind in Abbildung 6.14 und die einer Rekonstruktion eines ganzen Raumes in Abbildung 6.15 zu sehen.

Die Rekonstruktion eines Gesichtes aus Abbildung 6.14 mit einer Basislinie von 7.5cm bei einem Mindestobjektabstand von etwa 65cm liefert erstaunlich gute Ergebnisse. Die runden Formen von Gesicht und Körper werden dabei gut auf verschiedene Tiefenwerte abgebildet, wie an den Grauwerten in den Disparitätsbildern gut erkennbar ist. Die Augenbrauen sehen im Disparitätsbild jedoch viel heller und somit der Kamera näher liegend aus, als sie tatsächlich sind. Hier müsste eine Anpassung der Parameter des Birchfield-Algorithmus aus Tabelle 6.2 vorgenommen werden, um bessere Ergebnisse zu erziehlen. Die Malus Faktoren für stark, moderat und schwach vertrauenswürdige Regionen müssten dabei anders aufeinander abgestimmt werden. Auch der Notaus-Knopf in einer Entfernung von etwa 140cm in der unteren linken Ecke des Bildes wird sehr gut dargestellt.

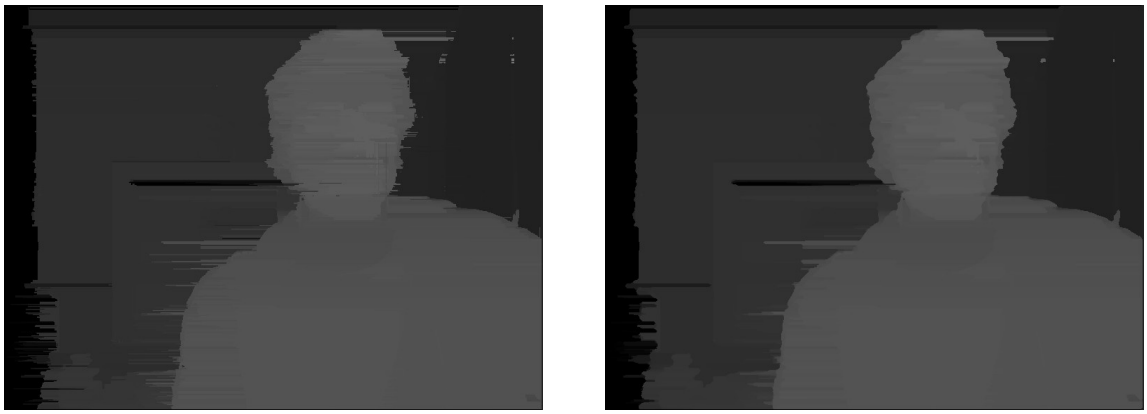
Tiefere Szenen verlangen eine größere Basislinie, um eine bessere Tiefenauflösung zu gewährleisten (vgl. Abschnitt 4.4). Während bei der Gesichtsrekonstruktion weiterhin mit einer Basislinie von 7.5cm sehr gute Ergebnisse erzielt wurden, reicht auch eine Basislinie von 10cm für die Rekonstruktion eines großen Raumes wie in Abbilgung 6.15 nicht aus. Die Basislinie muss etwa 20cm betragen, um ein angemessenes und gutes 3D Modell zu bekommen. Der Grund für die größere Basislinie liegt in der

[6]Bei einer Multiplikation des Raumvektors mit einem Skalar verändert sich natürlich auch die x- und y-Dimension der jeweiligen Objekte. Eine Analyse jenes Verhältnisses wurde in dieser Arbeit jedoch nicht betrachtet (vgl. Kapitel 7).

Tiefe des Raumes. Das nächste Objekt der Raumszene zur Kamera liegt etwa 400cm entfernt und der Raum hat eine weitaus größere Gesamttiefe als die untersuchten Szenen. Eine deutliche Rekonstruktion beispielsweise der Regale und Schränke an der hinteren Wand ist jedoch trotz allem nicht sehr gut möglich. Bei einem solch großen Raum sieht das berechnete 3D Modell dann doch etwas „zerstückelt“ und ungenau aus.

Diese Ergebnisse zeigen jedoch – wenn auch erst in geringem Maße für große Räume – die Übertragbarkeit des in dieser Arbeit entwickelten dreidimensionalen Rekonstruktionssystems auf andere Szenarien. Eine Analyse dieser Ergebnisse stellt unter anderem einen hervorragenden Ansatzpunkt für weitere wissenschaftliche Studien an diesem Rekonstruktionssystem dar.

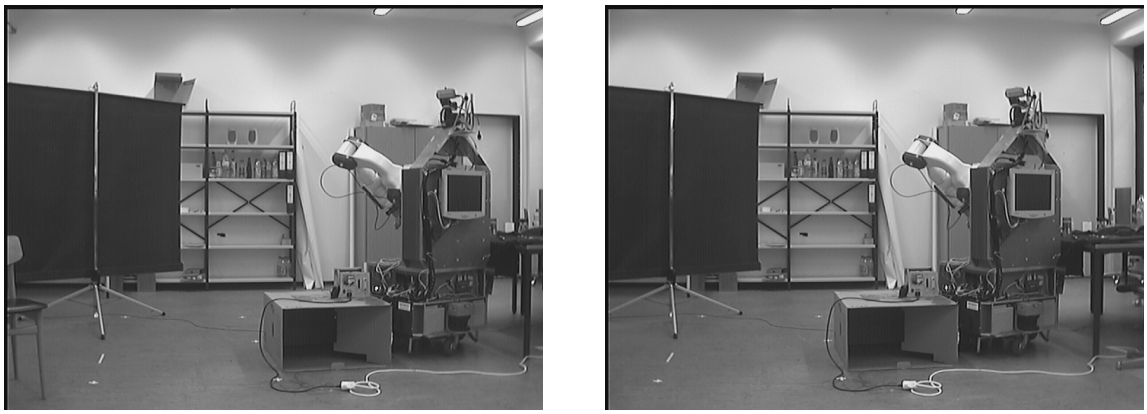
Eingabebilder mit Basislinie 7.5cm.



Originale (links) und geglättete (rechts) Disparitätskarte.

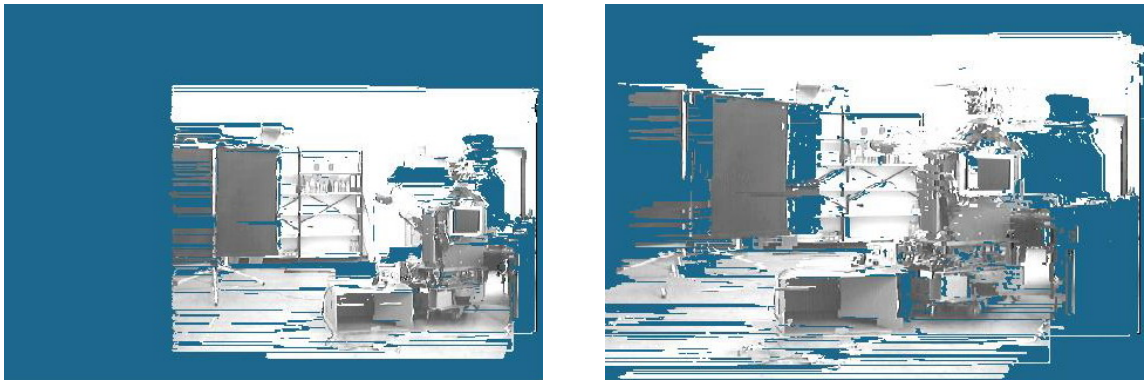
Verschiedene Ansichten des 3D Modells.

Abbildung 6.14: Versuch der Übertragbarkeit des entwickelten Rekonstruktionssystems auf Gesichtsrekonstruktion. Die Basisline bei diesem Versuch betrug 7.5cm.



Eingabebilder mit Basislinie 20cm.

Originale (links) und geglättete (rechts) Disparitätskarte.



Verschiedene Ansichten des 3D Modells.

Abbildung 6.15: Versuch der Übertragbarkeit des entwickelten Rekonstruktionssystems auf große Räume. Die Basisline bei diesem Versuch betrug 20cm.

7 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde ein multimodales, dreidimensionales Rekonstruktionssystem für Serviceroboter entwickelt. Augenmerk der Szenarien galt alltäglichen Tischszenen, um dem Serviceroboter eine Repräsentation potentieller Interaktionsmöglichkeiten, beispielsweise für Greifaktionen in näherer Arbeitsumgebung, dreidimensional zu modellieren. Das Modell basierte auf einem Stereobildpaar, aquiriert mittels Mikro-Kopf Kamera, montiert am Manipulator des Roboterarms. Unter Berücksichtigung der theoretischen Grundlagen des Forschungsgebietes *Struktur aus Bewegung* wurde eine Übertragung auf den Serviceroboter TASER realisiert. Dabei wurde das Gesamtproblem spezifisch in Aufgabenbereiche unterteilt und der theoretische Hintergrund ausführlich in den Kapiteln 2, 3, 4 vermittelt.

Über Grundlagen von Abbildungseigenschaften und Relationen zwischen mehreren Ansichten einer Szene wurde mittels epipolarer Geometrie die Stereogeometrie in eine achsparallele Geometrie rektifiziert. Auf diesen virtuellen achsparallelen Ansichten wurde das Korrespondenzproblem mit einem schnellen Stereoalgorithmus gelöst und die Disparitäten der einzelnen Pixel in eine Tiefenkarte übertragen. Daraufhin wurde ein dreidimensionales, texturiertes OpenGL Modell der abgebildeten Szene rekonstruiert, das neue virtuelle Ansichten der Szene ermöglicht.

7.1 Bewertung der Ergebnisse

Nachdem in dem vorangegangenen Kapitel 6 die Implementation eines dreidimensionalen Rekonstruktionssystems vorgestellt wurde, gilt es nun, diese Implementierung zu reflektieren. Die theoretischen Grundlagen aus den Kapiteln 2 bis 4 beschrieben alle fundamentalen Grundlagen zur Umsetzung eines dreidimensionalen, bildbasierten Rekonstruktionssystems.

Die Nutzung der Hand-Kamera hat im Gegensatz zum Stereokopf den Vorteil, dass der Großteil des Bildes nicht durch Roboterkomponenten, wie dem Arm, verdeckt würden und dadurch die Szenerie in ihrer Ganzheit abgebildet werden kann. Zudem kann eine Szene vor dem Roboter wegen der Mobilität der Kamera am Roboterarm

auch von der Seite aufgenommen werden ohne den gesamten Roboter zu bewegen und positionieren.

Die Ermittlung der Fundamentalmatrix, der Lagebeziehung der Kameras zueinander, stellt durch die derzeit notwendige Nutzerinteraktion der Korrespondenzpunkt-Selektion für den 8-Punkt-Algorithmus eine enorme Einschränkung und Fehlerquelle dar. Eine Automatisierung mittels RANSAC oder LMedS wäre eine wünschenswerte Ergänzung, doch dafür müssten die aufgetretenen fehlerhaften Berechnungen mittels vorgenannter Methoden erst genauer Untersucht werden. Leider passte eine diesbezügliche ausführliche Fehleranalyse nicht mehr in den Rahmen dieser Arbeit.

Durch die Rektifikation der Stereobilder mittels Epipolargeometrie ist eine freizügige Positionierung der Kameras möglich ohne auf die gängigen und schnellen Methoden der Korrespondenzanalyse für achsparallele Bilddaten verzichten zu müssen. Ansonsten müssten rechenintensive Algorithmen entwickelt werden um auf stereoskopisch ausgerichteten Bildern Disparitäten ermitteln zu können. Eine genauere Untersuchung und Anpassung der Rektifikationsergebnisse seien über diese Arbeit hinaus noch erforderlich, um auf den rektifizierten Bildern direkt die Stereoalgorithmen anwenden zu können. Derzeit wird durch eine willkürlich erscheinende Rotation und/oder Spiegelung der rektifizierten Bildpaare eine Nutzerinteraktion erzwungen, bei der die Bilder für den nachfolgenden Stereoalgorithmus manuell ausgerichtet werden müssen[1].

Die Verwendung einiger Funktionalitäten der OpenCV Bibliothek von Intel® erhöht die gesamte Performance des Systems wegen ihrer speziellen Abstimmung auf die Intel-Chipsätze. Insbesondere der Birchfield-Stereoalgorithmus stellt eine sehr schnelle Alternative zu den klassischen Korrespondenzanalysen dar. Andernfalls müsste für die Akquise der Bilddaten bestimmte Voraussetzungen eingehalten werden, die das Einsatzgebiet wieder veringern würden. Doch auch diese günstigen Rahmenbedingungen wie beispielsweise eine achsparallel Anordnung der Kameras ist mit dem entwickelten Rekonstruktionssystem leicht zu bewerkstelligen.

Durch die Kombination aus Hand-Kamera und Roboterarm mit 6 Freiheitsgraden bietet das in dieser Arbeit entwickelte System eine enorme Flexibilität für verschiedenste Anwendungsgebiete der bildbasierten Rekonstruktionssysteme in der Servicerobotik.

[1] Der Birchfield-Algorithmus untersucht Zeilenweise das linke Eingabebild von rechts nach links um korrespondierende Pixel im rechten Eingabebild zu ermitteln.

7.2 Ausblick

Die Ergebnisse und die Bewertung dieser Arbeit haben gezeigt, dass ein dreidimensionales Rekontruktionssystem anhand von Bildsequenzen einer Handkamera – allgemeiner einer Kamera an einem mobilen Roboterarm – im Kontext autonomer Serviceroboter umgesetzt werden kann und ein überaus nützliches Hilfsmittel für Umgebungsrepräsentationen darstellt. In diesem Kapitel sollen noch einige Erweiterungs- und Ergänzungsmöglichkeiten für das vorgestellte Rekonstruktionssystem aufgeführt werden um einen Ausblick für weitere Forschung zu geben. Sie gliedern sich in Detailergänzungen und konzeptionelle Erweiterungen.

Detailergänzungen Detailergänzungen beinhalten Änderungen der Implementation, die jedoch keinen Einfluss auf den Gesamtansatz haben.

- Eine genauere Analyse der resultierenden Fundamentalmatrix und des benutzten Rektifikationsalgorithmus könnten zu einer Vorhersage der Rotation und Spiegelung der rektifizierten Bilder führen. Würden die Bilder nicht rotiert und/oder gespiegelt wie es bei dem derzeitigen Algorithmus der Fall ist, wäre eine Kopplung der Rektifikationsergebnisse mit dem Stereoalgorithmus ein Schritt in Richtung eines automatisches Rekonstruktionssystems.
- Die bisher unzuverlässigen Ergebnisse des RANSAC und LMedS Verfahrens der OpenCV Funktion `cvFindFundamentalMat` zur Ermittlung der Fundamentalmatrix müssten genauer untersucht werden. Eine funktionstüchtige Fassung könnte die Berechnung mittels 8-Punkt-Algorithmus und somit die benötigte Auswahl von 10 Korrespondenzpunkten durch den Nutzer ersetzen. Folglich könnte der Parameter für die Anzahl der maximal zu detektierenden Ecken nach oben korrigiert werden und die Berechnung der Fundamentalmatrix über alle detektierten Eckpunkte durch den RANSAC oder LMedS Ansatz automatisiert werden.
- Eine strukturierte grafische Bedienoberfläche zur Präsentation der Einzelergebnisse und Auswahl der Parameter einzelner Funktionen würde die Akzeptanz und Nutzbarkeit des entwickelten Systems erhöhen.

Konzeptionelle Erweiterungen Konzeptionelle Erweiterungen haben Einfluss auf den Gesamtansatz und erweitern diesen um bisher nicht eingeflossene Aspekte.

- Die Ergänzung des verwendeten Stereoansatzes um weitere Kameraansichten würde zu realistischeren 3D Modellen führen, die aus jeder erdenklichen Lage

neu betrachtet werden könnten. Diese Art der Erweiterung fällt in den Bereich *multi-view geometry* und wurde ausführlich u.a. von Hartley, Pollefeys und Beardsley in [HZ03], [Pol00] und [BTZ96] untersucht. Dabei werden nicht nur zwei Kameraaufnahmen, sondern beliebig viele Ansichten aus unterschiedlichsten Betrachtungswinkeln für eine 3D Rekonstruktion genutzt. Nachbarschaften einzelner 3D Punkte könnten dadurch ermittelt werden und unter Verwendung der Nachbarschaftsinformation ein trianguliertes Oberflächennetz als 3D Modell präsentiert werden.

- Unter Betrachtung der Forschungsergebnisse von Chen aus [CL05, CL04] könnte eine Vorhersage für die nächste Aufnahmeposition, die möglichst viele Informationen für das zu berechnende 3D Modell beitragen kann, implementiert werden. Dies wird im Allgemeinen als das *next-best-view* Problem bezeichnet. In seinem Verfahren werden bei jeder neuen Aufnahme *sensing* Parameter [CL05] ermittelt um das 3D Modell sukzessiv zu erstellen. Dabei wird der Oberflächentrend eines Objektes berechnet um eine globale Vorhersage des Oberflächenverlaufs für bisher unbekannte Objektregionen zu machen.

- Analyseverfahren verschiedenster Algorithmen des Bereichs *Struktur aus Bewegung* nach Oliensis in [Oli00] offenbaren für Weiterentwicklungen einen grundlegenden Theoriebaustein. Oliensis schlägt Systeme vor, in denen verschiedene Struktur aus Bewegung Konzepte über eine Zwischenschicht miteinander zu mächtigen Systemen verknüpft werden.

- Es wäre zu untersuchen, ob beispielsweise ein 3D Rekonstruktionssystem einer omnidirektionalen Kamera, wie durch Fleck *et al.* [FBB+05] und Bunschoten *et al.* [BK03] beschrieben, mit dem hier entwickelten Rekonstruktionssystem verknüpft werden kann. Somit könnten einzelne Bereiche des groben 3D Modells einer omnidirektionalen Kamera mit detailierteren 3D Modellen eines Rekonstrukionssystems einer Handkamera zu einem Gesamtmodell verknüpft werden.

Als wesentliche konzeptionelle Erweiterung stellt sich die *Ergänzung um weitere Kameraansichten* heraus. Mit ihr wären realistischere 3D Modelle möglich. Aufbauend auf diesem Modell könnten dann Aktionsplanungen für Greifaktionen oder etwa eine dreidimensionale Kollisionserkennung zwischen Objekten der Szene und dem Roboterarm während einer Greifaktion berechnet werden.

Weitere Analysen wären jedoch vorerst der Ausgangspunkt für die Fortführung praxisbezogener Forschung an dem entwickelten Rekonstruktionssystem. Dabei wäre zu der in Kapitel 6.7 geführten Analyse der Qualität der Tiefe sicherlich eine Untersuchung zur Genauigkeit der Breite und Höhe der rekonstruierten Objekte ein erster

Ansatzpunkt um reale Objektgrößen zu berechnen. Diese Analyse lag leider außerhalb des Rahmens dieser Diplomarbeit.

Nach einer Montage des zweiten Roboterarms würde eine Modifizierung des entwickelten Rekonstruktionssystems zu einem Ansatz mit zwei physischen Handkameras ebenfalls Potential für weitere Forschung offenbaren.

Literaturverzeichnis

[AH88] Ayache, Nicolas ; Hansen, Charles: Rectification of images for binocular and trinocular stereovision / INRIA – Institut Nationale de Recherche en Informatique et en Automatique. Version: 1988. `http://www.inria.fr/rrrt/rr-0860.html` (860). – Rapport de recherche. – Online–Ressource. Letzter Aufruf 10. Feb. 06

[Arm96] Armstrong, Martin: *Self-Calibration from Image Sequences*, Department of Engineering Science, University of Oxford, Ph.D. thesis, 1996

[BK03] Bunschoten, Roland ; Kröse, Ben: Robust Scene Reconstruction from an Omnidirectional Vision System. In: *IEEE Transaction on Robotics and Automation* 19 (2003), April, Nr. 2, S. 351–357

[BT96] Birchfield, Stan ; Tomasi, Carlo: Depth Discontinuities by Pixel-to-Pixel Stereo / Stanford University. Version: July 1996. `http://www.ces.clemson.edu/~stb/publications/CS-TR-96-1573.pdf` (STAN-CS-TR-96-1573). – Technical report. – Online–Ressource. Letzter Aufruf 28. Dec. 05

[BT98] Birchfield, Stan ; Tomasi, Carlo: Depth Discontinuities by Pixel-to-Pixel Stereo. In: *Proceedings of the Sixth International Conference on Computer Vision*. Bombay, India, January 1998, S. 1073–1080

[BTZ96] Beardsley, Paul A. ; Torr, Philip H. S. ; Zisserman, Andrew: 3D Model Acquisition from Extended Image Sequences. In: *European Conference on Computer Vision (ECCV)* 2 (1996), 15. - 18. April, S. 683–695

[Can83] Canny, John: *Finding Edges and Lines in Images*. Cambridge, MA, USA, Massachusetts Institute of Technology, Master thesis, June 1983

[CL04] Chen, S. Y. ; Li, Y. F.: Automatic Sensor Placement for Model-Based Robot Vision. In: *IEEE Transactions on Systems, Man and Cyberne-*

tics, Part B: Cybernetics Bd. 34, IEEE Systems, Man, and Cybernetics Society, February 2004, S. 393–408

[CL05] Chen, S. Y. ; Li, Y. F.: Vision Sensor Planning for 3D Model Acquisition. In: *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* Bd. 35, IEEE Systems, Man, and Cybernetics Society, August 2005, S. 1–12

[Fau93] Faugeras, Olivier: *Three-dimensional computer vision: a geometric viewpoint*. Cambridge, MA, USA : MIT Press, 1993. – ISBN 0–262–06158–9

[Fau95] Faugeras, Olivier: Stratification of 3-Dimensional Vision: Projective, Affine, and Metric Representations. In: *Journal of the Optical Society of America (JOSA-A)* 12 (1995), March, Nr. 3, 465–484. `http://www-sop.inria.fr/odyssee/team/Olivier.Faugeras/publications.html`. – Letzter Aufruf 17. Jan. 06

[FB81] Fischler, Martin A. ; Bolles, Robert C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Commun. Assoc. Comp. Mach.* 24 (1981), Nr. 6, S. 381–395. – ISSN 0001–0782

[FBB+05] Fleck, Sven ; Busch, Florian ; Biber, Peter ; Strasser, Wolfgang ; Andreasson, Henrik: Omnidirektional 3D Modeling on a Mobile Robot using Graph Cuts. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Barcelona, Spain : IEEE Computer Society Press, April 2005, S. 1760–1766

[FTV00] Fusiello, Andrea ; Trucco, Emanuele ; Verri, Alessandro: A Compact Algorithm for Rectification of Stereo Pairs. In: *Machine Vision and Applications* 12 (2000), Nr. 1, 16–22. `http://www.sci.univr.it/~fusiello/papers/00120016.pdf`. – Letzter Aufruf 26. Dec. 05

[Fus00] Fusiello, Andrea: Uncalibrated Euclidean reconstruction: A review. In: *Image and Vision Computing* 18 (2000), May, Nr. 67, 555–563. `citeseer.ist.psu.edu/fusiello00uncalibrated.html`. – Letzter Aufruf 3. Oct. 05

[HÅ96] Heyden, Anders ; Åström, Kalle: Euclidean Reconstruction from Constant Intrinsic Parameters. In: *Proc. 13th International Conference of Pattern Recognition, Vienna, Austria*, IEEE Computer Society Press, 1996, S. 339–343

[HÅ97] Heyden, Anders ; Åström, Kalle: Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In: *International Conference on Computer Vision and Pattern Recognition*. Puerto Rico, 1997, 438–443

[Har97a] Hartley, Richard I.: In Defense of the Eight-Point Algorithm. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (1997), 6, Nr. 6, S. 580–593

[Har97b] Hartley, Richard I.: Kruppa's Equations Derived from the Fundamental Matrix. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* Bd. 19. Washington, DC, USA : IEEE Computer Society, 1997. – ISSN 0162–8828, S. 133–135

[Har99] Hartley, Richard I.: Theory and Practice of Projective Rectification. Hingham, MA, USA : Kluwer Academic Publishers, 1999, S. 115–127. – ISSN 0920–5691

[HS88] Harris, Chris ; Stephens, Mike: A combined corner and edge detector. In: *Fourth Alvey Vision Conference*, 1988, S. 147–151

[HZ03] Hartley, Richard I. ; Zisserman, Andrew: *Multiple View Geometry in Computer Vision*. Camebridge University Press, 2003

[Int05] Intel, Corp.: *Open CV 0.9.7*. `http://www.intel.com/research/mrl/research/opencv/`. Version: July 2005

[Len87] Lenz, Reimar K.: Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit. In: Paulus, Erwin (Hrsg.): *9. DAGM-Symposium*. Braunschweig : Springer, 1987 (Informatik-Fachberichte 149 Mustererkennung), S. 212–216. – Signatur 228 Kel 8424

[LH92] Lloyd, John ; Hayward, Vincant: *Multi-RCCL User's Guide*. 4.0. Montréal, Québec, Canada: McGill University, April 1992

[Lon81] Longuet-Higgins, H.-C.: A computer algorithm for reconstructing a scene from two projections. In: *Nature* 293 (1981), September, S. 133–135

[LT87] Lenz, Reimar K. ; Tsai, Roger Y.: Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. In: *International Conference on Robotics and Automation* Bd. 4, IEEE Society Press, March 1987, S. 68–75

[MHI] MITSUBISHI HEAVY INDUSTRIES, LTD.: *General Purpose Robot PA10 series PA10-6CE Instruction Manual for Installation, Maintenance & Safety.* – Letzter Aufruf 14. Jun. 06. `http://www.mhi.co.jp/kobe/mhikobe/products/mechatronic/download/new/loadfile/e_6c_m.pdf`

[MMW04] Meagher, T. ; Maire, F. ; Wong, O.: A Robust Multi-Camera Approach to Object Tracking and Position Determination using a Self-Organising Map Initialised through Cross-Ratio Generated Virtual Points. In: *International Conference on Computational Intelligence for Modelling Control and Automation*. Sheraton Mirage Hotel, Gold Coast, Australia, 12 - 14 July 2004

[MSKS04] Ma, Yi ; Soatto, Stefano ; Kosecka, Jana ; Sastry, S. S.: *An Invitation to 3-D Vision: From Images to Geometric Models*. Berlin, Heidelberg : Springer, 2004. – ISBN 0387008934

[Nis04] Nistér, David: An Efficient Solution to the Five-Point Relative Pose Problem. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* Bd. 26, 756–777

[NS04] Nistér, David ; Schaffalitzky, Frederik: What do four points in two calibrated images tell us about the epipoles? In: *Lecture Notes in Computer Science* 3022 (2004), S. 41–57

[Oli00] Oliensis, John: A Critique of Structure-from-Motion Algorithms. In: *Computer Vision and Image Understanding (CVIU)* 80 (2000), Nr. 2, S. 172–214

[PKG98] Pollefeys, Marc ; Koch, Reinhard ; Gool, Luc J. V.: Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters. In: *International Conference on Computer Vision (ICCV)* (1998), S. 90–95

[PKVG98] Pollefeys, Marc ; Koch, Reinhard ; Vergauwen, Maarten ; Gool, Luc J. V.: Flexible 3D Acquisition with a Monocular Camera. In: *IEEE International Conference on Robotics and Automation (ICRA)*, 1998, S. 2771–2776

[Pol99] Pollefeys, Marc: *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, ESAT-PSI, Katholieke Universiteit Leuven, Ph.D. thesis, 1999. – Scientific Prize BARCO 1999

[Pol00] Pollefeys, Marc: 3D Modeling from Images / Katholieke Universiteit Leuven. Dublin, Ireland, In conjunction with ECCV 2000, 26. June 2000. – Lecture notes

[RL87] Rousseeuw, Peter J. ; Leroy, Annick M.: *Robust regression and outlier detection*. New York : Wiley, 1987. – Signatur M ROU 32035. – ISBN 0–471–85233–3

[Sch05] Schreer, Oliver: *Stereoanalyse und Bildsynthese*. Berlin, Heidelberg : Springer, 2005. – ISBN 3–540–23439–X

[ST94] Shi, Jianbo ; Tomasi, Carlo: Good Features to Track. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, June 1994

[TL87] Tsai, Roger Y. ; Lenz, Reimar K.: Review of the two-stage camera calibration technique plus some new implementation tips and new techniques for center and scale calibration. In: *2nd Topical Meeting on Machine Vision, Optical Society of America* (1987), 18 - 20 March. – Also IBM RC 12301

[Tsa86] Tsai, Roger Y.: An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In: *International Conference on Computer Vision and Pattern Recognition*. Miami Beach, Fla. : IEEE Computer Society Press, June 1986, S. 364–374. – Signatur 228 Kel 8029

[Tsa87] Tsai, Roger Y.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using off-the-Shelf TV Cameras and Lenses. In: *IEEE Journal of Robotics and Automation* RA-3 (1987), Nr. 4, S. 323–344. ISBN 0–86720–294–7

[TV98] Trucco, E. ; Verri, A.: *Introductory Techniques for 3–D Computer Vision*. New York : Prentice Hall, 1998

[WCH92] Weng, Juyang ; Cohen, Paul ; Herniou, Marc: Camera Calibration with Distortion Models and Accuracy Evaluation. In: *IEEE Transaction on Pattern Analysis and Machine Intelligence* Bd. 14. Washington, DC, USA : IEEE Computer Society, 1992. – ISSN 0162–8828, S. 965–980

[WSZ06] Westhoff, Daniel ; Stanek, Hagen ; Zhang, Jianwei: Distributed Applications for Robotic Systems using Roblet-Technology. In: *Proceeding of the ISR/Robotik 2006 Joint Conference on Robotics*. Munich, Germany, May 2006

[Zha96] Zhang, Zhengyou: Determining the Epipolar Geometry and its Uncertainty: A Review / INRIA – Institut Nationale de Recherche en Informatique et en Automatique. Sophia-Antipolis Cedex, France, 1996 (2927). – Rapport de recherche

[Zha98] Zhang, Zhengyou: A Flexible New Technique for Camera Calibration / Microsoft Research. 1998 (MSR-TR-98-71). – Technical report

Singulärwertzerlegung

A

Mittels der Singulärwertzerlegung (engl. *singular value decomposition*, SVD) werden vorzugsweise komplexe Matritzen[1] in der numerischen Mathematik gelöst. Kapitel D.2.1 in [Sch05] und Kapitel A.7 in [MSKS04] geben weiteren Aufschluss zur Singulärwertzerlegung. Kapitel A4.4 in [HZ03] gibt darüber hinaus weitere Informationen über die Berechnungskomplexität, die abhängig von der Menge der Rückgabeinfomation ist.

Bei der Singulärwertzerlegung wird die Ausgangsmatrix Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ in eine Diagonalmatrix $\mathbf{D} = diag(d_1, \ldots, d_n) \in \mathbb{R}^{n \times n}$ und zwei orthogonale Matrizen $\mathbf{U} \in \mathbb{R}^{m \times n}$ und $\mathbf{V} \in \mathbb{R}^{n \times n}$ zerlegt, die auf folgende Weise faktorisiert ist:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top \tag{A.1}$$

Unter der Verwendung der Orthogonalitätseigenschaften der Spaltenvektoren für $\mathbf{U}$ und $\mathbf{V}$ folgt $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}$. Die Spalten von $\mathbf{U}$ und $\mathbf{V}$ heißen Links- bzw. Rechtssingulär. Die Matrix $\mathbf{D}$ ist eine Diagonalmatrix mit nicht-negativen Elementen, den Singulärwerten der Matrix $\mathbf{A}$. Die Zerlegung kann so permutiert werden, dass ohne Beschränkung der Allgemeinheit gilt:

$$d_1 \geq d_2 \geq \ldots \geq d_n \geq 0$$

Folglich kann die Lösung entsprechend dem kleinsten quadratischen Fehler mittels SVD gefunden werden.

[1] Die Matrizen müssen nicht notwendigerweise quadratische Form haben

B Merkmalsextraktion

Merkmale sind lokale, aussagekräftige, feststellbare Teile eines Bildes. Typische Merkmale sind beispielsweise starke Veränderung der Intensität der Farbewerte oder des Kontrastes, ausgelöst durch Objektkonturen – also Kanten, deren Anfangs- und Endpunkte, sowie Ecken homogener Flächen. An uniformen Flächen und homogenen Bildregionen des gleichen Intensitätswertes I, bei denen kaum Schwankungen der Intensität vorherrschen, lassen sich nur schwer aussagekräftige Merkmale berechnen. Bei einer Merkmalsextraktion werden die wichtigen strukturellen Eigenschaften eines Bildes hervorgehoben, während eine signifikante Menge nutzloser Information aus dem Bild herausgefiltert wird. Dieser Vorgang führt zu einer enormen Datenreduktion. Im folgenden sollen nur einige Kanten-Detektoren betrachtet werden, insbesondere der Moravec-Interest Operator und der auf ihn aufbauende Harris-Kanten-Detektor.

Verfahren zur Bestimmung von Ecken, spüren diese an Bildpositionen auf, an denen große Gradienten in alle Richtungen vorliegen. Drei Kriterien bei der Eckenextraktion eines optimalen Detektors sollten bestmöglich erfüllt werden.

1. Die wohl wichtigste Eigenschaft ist die Minimierung der Erkennung von falschen Ecken, die durch Rauschen entstehen (*false positives*). Ebenso wichtig ist die zuverlässige Bestimmung echter Ecken (*true positives*).
2. Eine detektierte Ecke sollten einer guten Lagebestimmung unterliegen, d.h. der Abstand zwischen dem berechneten Eckpixel und der echten Ecke sollte minimiert werden.
3. Für einen Eckpunkt soll nur ein Punkt zurück gegeben werden. Die Anzahl der lokalen Maxima, ausgelöst durch Rauschen um den wirklichen Punkt herum, ist zu minimieren.

B.1 Moravec-Interest Operator

Der Moravec Operator dient zur Detektion isolierter Punkte oder Ecken. Er analysiert die mittlere Änderung der Bildintensität um einen Bildpunkt in 45° Schritten.

Der Operator liefert den größten Wert dort, wo die Grauwertänderung in mehrere Richtungen besonders groß ist (vgl. Abbilung B.1). Dies ist an Ecken der Fall. Der Moravec Operator klassifiziert Merkmale als Ecken, die im Auffälligkeitsbild lokale Maxima darstellen. Sei $f(u, v)$ das Intensitätsbild, so berechnet der Moravec-Interest Operator $MO(u, v)$ nach Gl. B.1.

0	0	0	0	0
0	0	0	0	0
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1

(a) Original: Ecke

X	X	X	X	X
X	1	2	3	X
X	2	5	3	X
X	3	3	0	X
X	X	X	X	X

(b) Moravec Ergebnis

0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1

(c) Original: Kante

X	X	X	X	X
X	3	3	0	X
X	3	3	0	X
X	3	3	0	X
X	X	X	X	X

(d) Moravec Ergebnis

Abbildung B.1: Anwendung des Moravec-Operators auf eine Ecke und eine Kante. Die Ergebnisse des Operators zeigt in diesem Fall die Anzahl der Richtungen in denen ein Intensitätswechsel um ein Pixel stattgefunden hat.

$$MO(u, v) = \frac{1}{8} \sum_{m=-1}^{1} \sum_{n=-1}^{1} |f(u + m, v + n) - f(u, v)| \tag{B.1}$$

Die Nachteile des Moravec-Interest Operators liegen in seiner Rotationsinvarianz und Problemen bei symmetrischen Grauwertverteilungen. Das rechteckige Betrachtungsfenster um ein Pixel herum führt zu Rauschen. Des weiteren liefert der Moravec Operator ein Maximum nicht direkt über der Ecke, sondern leicht versetzt.

B.2 Harris-Ecken-Detektor

Der im Jahre 1988 von Harris und Stephens vorgestellte Ecken-Detektor [HS88] stellt eine Erweiterung des Moravec-Interest Operators dar und löst ihn damit weitestgehend ab. Er ist wie der Moravec-Operator ein sogenannter intensitätsbasierter (engl. *intensity based detector*) Detektor. Die Erweiterung basiert auf der Entwicklung einer Taylorreihe. Es werden zuerst die Gradienten in horizontaler und vertikaler Richtung berechnet. Durch eine Faltung des Bildes mit einer Gewichtsfunktion werden die Quadrate der örtlichen Ableitung einer Tiefpassfilterung unterzogen. Für jeden Bildpunkt lässt sich somit eine Kovarianzmatrix $\mathbf{M}$ (Gl. B.2) aufstellen. Störungen durch Bildrauschen werden dadurch unterdrückt.

Im Gegensatz zur diskreten Anzahl der betrachteten 45° Drehungen des Moravec-Operators werden alle möglichen Drehungen durch die Gradienten der Intensitäten

berechnet. Dadurch wird der Harris-Ecken-Detector invariant gegenüber Rotationen. Moravecs Indikation für eine Ecke wird von Harris und Stephens erweitert indem auch die Richtung und Variation einer Intensitätsänderung mit in die Berechnung einfließt. Die Richtungsänderungen lassen sich mit Hilfe des Spektralsatzes[1] und der Kovarianzmatrix aus Gl. B.2 berechnen.

$$\mathbf{M} = \begin{pmatrix} I_u^2 & I_{uv} \\ I_{uv} & I_v^2 \end{pmatrix} \tag{B.2}$$

Wobei I_u und I_v die Gradienten der Intensität in horizontaler und vertikaler Bildrichtung darstellen. Die Eigenwerte der Kovarianzmatrix sind groß, genau dann wenn eine Ecke vorliegt. Ist jeweils nur ein Wert der Kovarianzmatrix groß, liegt eine Kante vor. Die Eigenwerte sind folgliche ein Maß für die „Kantigkeit“ zweier orthogonaler Richtungen.

Für einen ausgezeichneten Eckpunkt definieren Harris und Stephen das in Gl. B.3 dargestellte Auswahlkriterium. Es sei $det(\mathbf{M}) = I_u^2 I_v^2 - I_{uv}^2$ und $trace(\mathbf{M}) = I_u^2 + I_v^2$. Der Faktor k ist ein Gewichtungsfaktor und wurde von Harris empirisch zu 0.04 gewählt.

$$\mathbf{K} = det\mathbf{M} - k(trace\mathbf{M})^2 \tag{B.3}$$

Mit Hilfe des Spektralsatzes kann die Kovarianzmatrix in beliebige Richtungen rotiert werden. Die Orientierung der Gradienten wird dabei ebenfalls rotiert und die Matrix ermöglicht eine überprüfen, ob eine Ecke oder Kante in jene Richtung zu finden ist.

Der Harris-Ecken-Detektor führt bei geringen Variationen des Bildes und eng benachbarten Kanten zu verlässlichen Ergebnissen. Um zwei Bilder unterschiedlicher Größe miteinander zu vergleichen, besteht der Nachteil des Harris-Ecken-Detektor jedoch in seiner hohen Sensibilität gegenüber Skalierungen.

[1] Allgemein besagt der Spektralsatz für selbstadjungierte bzw. normale lineare Operatoren $\mathbf{A}$ über einem komplexen Hilbertraum die Existenz einer Spektralzerlegung, also eines Projektionsoperator-wertigen Wahrscheinlichkeitsmaßes λ über den reellen Zahlen bzw. komplexen Zahlen, so dass $\mathbf{A}$ sich als Spektralintegral $\mathbf{A} = \int_{\mathbb{R}} x\, \lambda(dx)$ schreiben lässt. Quelle: http://de.wikipedia.org/wiki/Spektralsatz.

Open Computen Vision Library

C

Die *Open Computer Vision Library*[1] (Abk. OpenCV) ist eine Open-Source Bibliothek. Sie wurde von Intel® im Jahre 2000 entwickelt und speziell für die Architektur der Intel Chipsätze optimiert. Sie läuft unter den Betriebssystem Linux ebenso wie unter Windows. Die Bibliothek ist besonders auf Echtzeit Computer Vision Anwendungen ausgelegt und in ANSI C und C++ geschrieben. Die OpenCV Bibliothek umfasst unter anderem optimierte Filter wie etwa den Harris-Ecken-Detektor, 3D-Funktionalitäten, Tracking-Algorithmen, sowie Gesichts- und Gesten-Erkennung und viele andere Funktionen aus dem Anwendungsgebiet der Bewegungs- und Bildanalyse. Zudem bietet sie einige Algorithmen im experimentellen Stadium aus aktuellen Forschungsergebnissen in dem Bereich Struktur aus Bewegung, die in dieser Arbeit zum Einsatz kommen. Die hier verwendete Version von OpenCV[Int05] ist die Veröffentlichung 0.9.7 beta 5.

Den Kern von OpenCV bildet die Image Processing Library (Abk. IPL), die von Intel® um einige komplexe Funktionalitäten erweitert wurde. Intel entwickelte die IPL 1997. OpenCV kann aufgrund der Open-Source Bestimmungen frei von Kosten genutzt, verändert und weitergegeben werden, und unterliegt offensichtlich einer BSD-Lizenz[2].

Die Dokumentation von Intel selbst hält sich leider sehr in Grenzen und kann als schlecht bezeichnet werden. Dies führt zu einem sehr zeitintensiven Einbindungsprozess. Jedoch gibt es ein sehr gutes, von Intel empfohlenes Diskussionsforum bei Yahoo[3] rund um das Thema OpenCV, in dem sich viele Fragen zu Handhabung und Problemen lösen lassen.

Zusammenfassend seien nun kurz die genutzten Funktionalitäten der Open Coputer Vision Library im Rahmen dieser Arbeit dargestellt. Mit `cvGoodFeaturesToTrack`

[1] Offizielle Informationen über OpenCV und Verknüpfungen zu Quelltext, Internet Gruppen und Dokumentationen unter: http://www.intel.com/technology/computing/opencv/, letzter Aufruf: 25.04.2006.

[2] http://www.intel.com/technology/computing/opencv/license.htm, letzter Aufruf: 25.04.2006.

[3] OpenCV Diskussionsgruppe unter http://www.yahoogroups.com/group/OpenCV, letzter Aufruf: 28.04.2006.

wurden durch Nutzung eines Harris-Ecken-Detektors (vgl. Anhang B.2) die Eckpunkte in den Eingabebildern bestimmt und nach dem stärksten Eigenwert für die jeweiligen Ecke geordnet. Über die Funktion `cvFindFundamentalMat` wird mit dem klassischen 8-Punkt-Algorithmus von Longuet-Higgins [Lon81] die Fundamentalmatrix berechnet (vgl. Abschnitt 3.1.3). Durch die Verknüpfung verschiedener Funktionalitäten wie `cvMakeScanline` und `PreWarpImage` werden die entzerrten Bilder rektifiziert (vgl. Kapitel 3.3). Abschließend wird durch `cvFindStereoCorrespondance` mit dem Stereoalgorithmus von Birchfield und Tomasi [BT98] eine Disparitätenkarte der Szene erstellt (siehe Abschnitt 4.3).

Danksagung

Die Bearbeitung dieser Diplomarbeit wäre ohne die Unterstützung vieler Menschen nicht möglich gewesen. An dieser Stelle möchte ich die Gelegenheit nutzen, mich bei jenen Menschen zu bedanken, die mir dieses Projekt „Diplomarbeit“ ermöglichten.

In erster Linie bedanke ich mich bei Prof. Dr. Jianwei Zhang für die Möglichkeit an diesem überaus interessanten Thema wissenschaftlich zu forschen und für seine Bemühungen meiner beruflichen Zukunft. Zudem geht mein Dank an Dr. Werner Hansmann, der während meines gesamten Studiums mein Interesse für den Bereich Computer-Vision und Computer-Grafik geweckt hat.

Außerdem bedanke ich mich bei allen Mitarbeitern des Arbeitsbereich TAMS für die förderlichen, fachlichen Diskussionen und Anregungen. Natürlich aber auch für die gute Stimmung während der ganzen Bearbeitungszeit, ohne die mir diese Diplomarbeit wesentlich schwerer gefallen wäre. Stellvertretend seien für den Arbeitsbereich hier insbesondere Tim Baier, Markus Hüser und Daniel Westhoff genannt, die mir nicht nur bei fachlichen Problemen mit Rat und Tat zur Seite standen.

Ferner bedanke ich mich bei meiner ganzen Familie für ihren Zuspruch, ihre unendliche Geduld sowie aufbauende Art. Insbesondere danke ich meiner Schwester Manuela Jockel für die finanzielle Unterstützung während der Dauer dieses Projektes.

Für den großartigen seelischen Beistand gerade in den Anfängen der Diplomarbeit und die kreative Ablenkung in der geringen verbleibenden Freizeit danke ich allen Mitgliedern meiner Band Bozinsky.

Schließlich danke ich noch Dirk Bade und Martin Weser, die es bestens zu verstehen wussten, meine grammatikalischen Ausflüge und Experimente zu zügeln.

Eidesstattliche Erklärung

Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Die vorliegende Diplomarbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und ist noch nicht veröffentlicht.

Ich bin mit einer Einstellung meiner Diplomarbeit in den Bestand der Bibliothek des Fachbereiches einverstanden.

Hamburg, den _____ Unterschrift: _____