Aktive Stereoskopie,

Neue Verfahren zur dreidimensionalen Vermessung von Objekten

von

Helmut Hamfeld

Im Fachbereich Informatik
der Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
eingereichte Dissertation

September 2001

Vorwort

Die vorliegende Dissertation ist während meiner mehrjährigen Forschungsarbeit als wissenschaftlicher Mitarbeiter der Arbeitsgruppe "Robotik und Prozessrechentechnik" im Fachbereich Informatik der Universität Kaiserslautern entstanden und dort unter dem Titel *Aktive Stereoskopie - Neue Verfahren zur dreidimensionalen Vermessung von Objekten* erschienen.

Forschungsgebiet der Arbeitsgruppe ist der Bereich der Autonomen Mobilen Roboter, mit den Schwerpunkten Sensoren, Hardwarearchitektur und Kontrollsoftware. Grundlage dieser Forschungsarbeit sind die Ergebnisse der beiden Forschungsprojekte TOPAS[1] und SMARAGD[2] auf dem Gebiet der Sensorik, insbesondere der 3D-Vermessung von Objekten.

Beide Forschungsprojekte wurden von der Stiftung Rheinland-Pfalz für Innovation gefördert, die nicht zuletzt durch die unproblematische Finanzierung meiner Arbeiten die hier dokumentierten Ergebnisse ermöglicht hat. Dem Arbeitsgruppenleiter Prof. Dr. E. v. Puttkamer danke ich für seine vielfältige Unterszützung, sowie die stets fruchtbaren Anregungen und das daraus resultierende hervorragende Arbeitsklima.

Ebenso hat das gute Verhältnis zu meinen (ehemaligen) Kollegen Markus Berg, Michael Kasper, Dr. Klaus-Werner Jörg, Guido Hansen und Dr. Rainer Trieb zum Gelingen meiner Arbeit beigetragen. Ganz besonderer Dank gilt dabei meinem Kollegen Joachim Weber, sowie meinem Bruder Heinz Hamfeld für die Mühe des Korrekturlesens dieser Arbeit.

Weiterer Dank gilt natürlich den Studenten, die im Rahmen von Projekt- und Diplomarbeiten, sowie im Rahmen einer Anstellung als wissenschaftliche Hilfskraft einen wesentlichen Anteil an der Forschungsarbeit haben. Namentlich danke ich an dieser Stelle Boris Heiduk†, Ewald Dieterich, Arne Hoffmann, Oliver Mezger, Klaus Hofmann, Stefan Hülbrock, René Joachimi, Michael Kapp, Martin Hein, Kushalappa Paleyanda und Frank Peters, denen ich viel Erfolg für ihr weiteres Studium, bzw. ihren beruflichen Werdegang wünsche.

Meinem jetzigen Arbeitgeber, der Firma TecMath AG, insbesondere Dr. Andreas Seidl gilt mein Dank für die Unterstützung bereits während der beiden Forschungsprojekte. Bei Prof. G. Häusler bedanke ich mich für die Übernahme des Korreferats.

Kaiserslautern, im Sepember 2001 Helmut Hamfeld

1. Touchless Perception and Anthropometric Adaption System
2. Smart Acquisition of Geometric Data

Zusammenfassung

Die dreidimensionale Erfassung von Objekten ist seit Mitte der 80er Jahre Inhalt von Forschungsprojekten und hat seitdem zu einer Vielzahl unterschiedlicher Produkte für die verschiedensten Anwendungsgebiete geführt. Erste Ansätze computergestützter Datenverarbeitung zur Erfassung von Objektoberflächen sind bereits in den 70er Jahren zu finden, diese wurden aber nur begrenzt weiterentwickelt. Die Ursache lag vor allem darin, dass die damals noch teure Rechenzeit der vorhandenen Computersysteme keine effiziente Realisierung zuließ. Die grundsätzlichen Ideen vieler heute bekannter Verfahren wurden aber schon viel früher entwickelt. So bilden Stereogramme die Vorläufer der Stereoskopie: mit Computern wird versucht, den 3D-Effekt, der problemlos vom menschlichen Sehsystem vermittelt wird, nachzubilden. Die Technologien, die zur Kartografierung und Landvermessung (Geodätik) notwendig sind, bilden die Basis der heute verfügbaren Triangulationssysteme. Die Messung der Laufzeit eines Lichtimpulses von einem Sensor zu einem Objekt ist ebenfalls ein heute verfügbares Prinzip zur Erfassung von Objektoberflächen. Bereits Galileo Galilei hat Überlegungen zur Messung der Lichtgeschwindigkeit angestellt, die erste Berechnung der Lichtgeschwindigkeit wurde im 17ten Jahrhundert von Römer durchgeführt.

Aktuelle Forschungsprojekte decken die Vermessung von Strukturen im atomaren Bereich bis hin zur Oberflächenvermessung ganzer Planeten ab. Inhalt dieser Arbeit ist vor allem die Vermessung von Objekten, die eine Größe von wenigen Zentimetern bis hin zu einigen Metern haben. Während der bisherige Stand der Technik für jedes Anwendungsgebiet und jede Objektgröße ein spezialisiertes Sensorsystem vorsieht, ist der Mittelpunkt dieser Arbeit die Entwicklung von Sensorsystemen, mit einem besonders großen Anwendungsspektrum.

Nach einer groben Vorstellung des heutigen Stands der Technik zur dreidimensionalen Objekterfassung und der dafür gebräuchlichsten Verfahren werden als weitere Grundlage die wichtigsten Methoden zur Kalibrierung von 3D-Sensoren vorgestellt, die die Hauptvoraussetzung einer genauen Vermessung von Objekten bilden.

Im zweiten Teil dieser Arbeit wird die im Rahmen der Forschungsarbeit entwickelte Technologie vorgestellt. Das nach dem Prinzip der aktiven Stereoskopie arbeitende Basissystem *HandyScan3D* bildet dabei die Grundlage der darauf aufbauenden Sensorsysteme *HandyScanTT* und *HandyScanCLM*. Schwerpunkte dieser Arbeit sind dabei Aspekte der Selbstkalibrierung von 3D-Sensoren, der Echtzeitbildverarbeitung und der Echtzeitvisualisierung. Diese Technologien zusammengefasst ergeben eine Sensorsystemkonzeption, die sich durch einfache Handhabung, kostengünstige Realisierung und hohe Flexibilität auszeichnet, was nicht zuletzt durch die realisierten, anwendungsnahen Systeme *HandyScanTT* und *HandyScanCLM* belegt werden soll.

Inhaltsverzeichnis

Vorwort **1**

Zusammenfassung **3**

Zielsetzung und Beitrag der Arbeit **7**

1 Dreidimensionale Objekterfassung **9**

1.1 Begriffserklärung **10**

1.2 Klassifizierung von 3D-Sensoren **12**

1.2.1 Aktive optische 3D-Sensoren 13
1.2.1.1 Eindimensional messende 3D-Sensoren 13
1.2.1.2 Zweidimensional messende 3D-Sensoren 17
1.2.1.3 Räumlich messende 3D-Sensoren 19
1.2.1.4 Aktive Stereoskopie 23

1.2.2 Passive optische 3D-Sensoren 24
1.2.2.1 Monoskopie 24
1.2.2.2 Volumenschnitt 25
1.2.2.3 Passive Stereoskopie 27

1.2.3 Andere Sensortypen 30
1.2.3.1 Ultraschall 31
1.2.3.2 Taktile Systeme 31

1.3 Beispiele für 3D-Sensorsysteme **32**

2 Kalibrierung von 3D-Sensoren **43**

2.1 Kamerakalibrierung **43**

2.1.1 Elektronische Kameras 43
2.1.1.1 CCD-Kameras 44
2.1.1.2 CMOS-Kameras 44
2.1.1.3 Bildübertragung in den Computer 45

2.1.2 Abbildungseigenschaften von Kameraobjektiven 46

2.1.3 Kameramodelle 47

2.1.4 Kalibrierung nach Tsai 49

2.2 Modellbasierte Kalibrierung **53**

2.3 Modellfreie Kalibrierung **54**

2.3.1 Interpolierende Kalibrierung 55

2.3.2 Approximierende Kalibrierung 56

2.4 Stereokalibrierung 57
2.4.1 Berechnung der Epipolargeometrie 57
2.4.2 Nichtlineare Berechnung der Stereokalibrierung 61

3 Aktive Stereoskopie 63

3.1 HandyScan3D 63
3.1.1 HandyScan3D im Überblick 63
3.1.2 Verwendete Hardware 65
3.1.3 Bildverarbeitung 67
3.1.3.1 Extraktion eines Punktes 68
3.1.3.2 Extraktion einer Linie 70
3.1.3.3 Extraktion des Gray-Codes 73
3.1.4 Stereokalibrierung 77
3.1.5 3D-Rekonstruktion und Visualisierung 82
3.1.5.1 3D-Rekonstruktion 82
3.1.5.2 Visualisierung 85
3.1.5.3 Erstellung von Mehrfachansichten 88
3.1.6 Beispiele 89
3.2 HandyScanTT 97
3.2.1 Systemkonzeption 97
3.2.2 Kalibrierung und 3D-Rekonstruktion 100
3.2.3 Verschmelzen mehrerer Ansichten 103
3.2.3.1 Genauigkeitssteigerungen 103
3.2.3.2 Korrektur der Textur 106
3.2.4 Beispiele 107
3.3 HandyScanCLM 110
3.3.1 Systemkonzeption 111
3.3.2 Kalibrierung und Registrierung 113
3.3.2.1 Kalibrierung der Ortungskameras 115
3.3.2.2 Registrierung von Einzelansichten 117
3.3.3 Beispiele 119

4 Resumee und Perspektiven 129

A Abbildungsverzeichnis 131

B Literaturverzeichnis 133

C Index 139

Zielsetzung und Beitrag der Arbeit

Am Anfang der Planung dieser Forschungsarbeit stand die Überlegung, einen handgeführten 3D Laserscanner zur dreidimensionalen Rekonstruktion von Objekten im Rechner zu entwikkeln. Das auf aktiver Stereoskopie basierende Konzept stellt ein Höchstmaß an Flexibilität und einfacher Handhabung in Aussicht, erforderte aber auch die Entwicklung von echtzeitfähigen Verfahren zur Bildverarbeitung, Selbstkalibrierung und zur Datenvisualisierung. Denn erst durch die Echtzeitfähigkeit des Gesamtsystem wird die höchstmögliche Interaktion erreicht, die Voraussetzung für die Einsetzbarkeit eines Handscanners ist. Weitere Überlegungen zielen auf die Erweiterung der Grundkonzeption zum Nachweis der praktischen Anwendbarkeit und in Richtung integrierter Sensorsysteme zur hochgenauen Vermessung großer Objekte.

Im Rahmen dieser allgemeinen Zielsetzung sind vor allem die folgenden Fragestellungen und vertiefenden Überlegungen zu berücksichtigen:

- Wie können bestehende Beschränkungen verfügbarer 3D-Systeme, z.B. ein begrenztes Messvolumen, überwunden werden?
- Wie können bekannte Verfahren aus der passiven Stereoskopie und der Lasertriangulation kombiniert werden bei gleichzeitiger Unterdrückung der Nachteile und Verbindung der Vorteile beider Verfahren?
- Unter welchen Voraussetzungen kann die robuste Selbstkalibrierung eines 3D-Systems auf Basis aktiver Stereoskopie durchgeführt werden?
- Welche Vereinfachungen können getroffen werden, um ein echtzeitfähiges 3D-Sensorsystem zu realisieren?
- Welche Konzepte und Vereinfachungen führen zu einem echtzeitfähigen Visualisierungssystem für 3D Scandaten?
- Welche Anforderungen können an die praxisnahe Konzeption eines 3D Sensorsystems gestellt werden?

Die Teilprobleme, die zur Lösung dieser Probleme im Rahmen dieser Arbeit gelöst wurden beinhalten die

- Entwicklung von echtzeitfähigen Bildverarbeitungssystemen zur Extraktion eines Laserpunktes oder einer Laserlinie aus Stereobildern,
- Entwicklung eines robusten Verfahrens zur Online-Stereokalibrierung auf Basis sequentiell aufgenommener Stereokorrespondenzen,
- Entwicklung eines Verfahrens zur Mehrkamerakalibrierung und zur Registrierung von 3D-Scans auf Basis optisch projizierter Lichtmarken,
- Entwicklung von echtzeitfähigen Verfahren zur Rekonstruktion und Visualisierung der Scandaten.

Der Aufbau dieser Arbeit trägt diesen Fragestellungen und Problemen Rechnung, wobei in den einzelnen Kapiteln die nachfolgend aufgeführten Themenbereiche behandelt werden:

Kapitel 1 ist eine Übersicht und Klassifizierung bestehender Technologien zur dreidimensionalen Erfassung von Objekten auf Basis optischer Sensorsysteme. Das Kapitel soll den Einstieg in die Methoden der dreidimensionalen Objekterfassung erleichtern und erklärt die wichtigsten Prinzipien heute verfügbarer Systeme. Den Abschluss bildet die kurze Vorstellung einer Auswahl kommerziell verfügbarer Sensorsysteme.

Kapitel 2 beschreibt verschiedene Methoden zur Kalibrierung von 3D-Sensoren. Vor dem Hintergrund, dass die Kalibrierung oft ein vernachlässigter Aspekt von 3D-Sensorsystemen ist, wird dargelegt, dass erst eine genaue Kalibrierung die tatsächliche Qualität eines 3D-Sensors ausmacht.

Kapitel 3 enthält die eigentlichen Systeme, die im Rahmen dieser Arbeit entwickelt wurden. Das Basissystem bildet *HandyScan3D*, auf dem die beiden Weiterentwicklungen *HandyScanTT* und *HandyScanCLM* aufbauen. Schwerpunkte sind dabei die Verfahren der Bildverarbeitung, die im speziellen verwendeten und entwickelten Kalibrierverfahren und die Methoden zur Visualisierung der Scandaten.

Die Arbeit enthält somit die folgenden, neuartigen Beiträge:

- *HandyScan3D* als Basiskonzept eines flexiblen, selbstkalibrierenden 3D-Messsystems. Das System zeichnet sich durch seine echtzeitfähige Bildverarbeitung und seine robuste, während der Messung sich stetig verbessernde Kalibrierung und Genauigkeit aus. Zudem vereinigt *HandyScan3D* erstmalig verschiedene Messverfahren auf Basis eines Laserpunktes, einer Laserlinie und des kodierten Lichtansatzes unter einem gemeinsamen Konzept.
- Ein System zur Visualisierung von Scandaten, das die Triangulation der Scandaten, eine Multiresolution-Methode zur Visualisierung unterschiedlicher Auflösungen und ein Verfahren zur Lückenfüllung bei unvollständigen oder spärlichen Datensätzen beinhaltet. Alle diese Komponenten tragen der Forderung der Echtzeitvisualisierung Rechnung.
- *HandyScanTT* als anwendungsnahes Sensorsystem zur vollständigen 3D-Erfassung von Objekten, ausgelegt vor allem auf eine kostengünstige Realisierung und einfache Handhabung.
- *HandyScanCLM* als produktorientiertes Gesamtkonzept zur hochgenauen Vermessung von Objekten verschiedenster Größe, das sich vor allem durch die erstmalige Integration von photogrammetrischen und topometrischen Sensorsystemen bei gleichzeitiger hoher Flexibilität und einfacher Handhabung auszeichnet.

1 Dreidimensionale Objekterfassung

Die stetig fortschreitende Entwicklung des Computers seit den 80er Jahren hat den Ablauf moderner Fertigungsprozesse entscheidend geprägt. Die zunehmende Leistungsfähigkeit von Computern hat dazu geführt, dass immer mehr Teilprozesse in der Produktfertigung vom Computer unterstützt oder gänzlich vom Computer durchgeführt werden konnten. Waren beispielsweise früher für die Planung und Entwicklung eines Autos noch viele Iterationszyklen nötig, bis die Fließbandproduktion eines neuen Modells anlaufen konnte, so wird heutzutage fast der gesamte Fertigungsprozess mit dem Computer geplant, bevor das erste Serienfahrzeug hergestellt wird. Designstudien werden zwar nach wie vor von Hand erstellt, aber bereits in einer sehr frühen Phase werden Computermodelle der neuen Typen erstellt, um Simulationen bezüglich des Aussehens der Karosserie, des Designs des Fahrzeuginnenraums oder der Fahrzeugergonomie durchführen zu können. Der Entwurf der einzelnen Fahrzeugkomponenten geschieht mit Hilfe von CAD-Programmen. Bevor das erste Teil des Fahrzeugs gefertigt wird, existiert bereits ein vollständiges Modell im Computer. Ein realer visueller Eindruck wird nach wie vor mit verkleinerten Modellen gewonnen, die entweder von Hand vom Designer entworfen oder aufgrund der Computerdaten aus Kunststoff gefräst werden. Ein erstes Problem bei der Fahrzeugentwicklung mit CAD-Programmen war die Übertragung von Änderungen, die an den realen Modellen durchgeführt wurden, zurück in den Computer.

Die erste Lösung zur dreidimensionalen Digitalisierung solcher Objekte ist der Einsatz von Koordinatenmessmaschinen. Ein Tastkopf auf einer Positioniereinheit wird an das Objekt herangefahren. Jede Berührung des Tastkopfes mit dem Objekt wird registriert und die gemessene Koordinate gespeichert. Aufwändige Mechaniken erlauben die Messung von Oberflächenkoordinaten mit Genauigkeiten von Bruchteilen eines Millimeters, so dass diese Systeme noch heute als Referenzmaß für moderne optische 3D-Sensorsysteme dienen. Sowohl die unerwünschte Berührung des Tastkopfes mit dem Objekt, als auch die prinzipbedingte geringe Datenrate hat die Entwicklung neuer Verfahren erzwungen. Der Austausch des Messkopfes durch einem optischen Entfernungsmesser war der erste Fortschritt. Die Koordinatenmessmaschinen mussten nur noch bis auf einige Zentimeter an das Objekt herangeführt werden. Zudem erlaubte die Datenrate der optischen Entfernungsmesser oft ein scannendes Abtasten der Oberflächengeometrie des Objektes, so dass die Vermessung des Objektes nicht mehr nur auf markante Strukturen auf der Objektoberfläche eingeschränkt war, sondern eine flächendekkende Digitalisierung möglich wurde. Bei kompletten Fahrzeugen oder großen Modellen war der Zeitaufwand jedoch so hoch, dass nach wie vor noch Kompromisse bezüglich der Genauigkeit, Datendichte oder Auflösung eingegangen werden mussten. Ende der achtziger Jahre wurden die ersten 3D-Sensoren nach dem sogenannten Lichtschnittverfahren und dem Coded-Light Ansatz entwickelt. Erst mit diesen Technologien wurde es möglich, mit sehr hohen Datenraten und in ausreichender Genauigkeit eine dichte, flächendeckende Abtastung von Objekten durchzuführen. Die Computer waren jedoch kaum in der Lage, die anfallenden Datenmengen zu verwalten oder weiter zu verarbeiten. Erst die Leistungsfähigkeit aktuell verfügbarer Computer sorgte für einen Boom in der dreidimensionalen Objektvermessung und für

die Entstehung immer neuer Anwendungsgebiete. Durch moderne Technologie haben 3D-Sensoren vor allem hinsichtlich ihrer Genauigkeit und Effizienz deutlich an Wert gewonnen. Bei der Entwicklung neuer 3D-Sensoren waren dabei besonders zwei Trends zu beobachten. Zum einen wurde versucht, die Auflösung der Sensoren weiter zu steigern, wodurch einige optische 3D-Sensoren mittlerweile Auflösungen von unter einem Mikrometer erreichen. Der andere Trend stellte sich in einer zunehmenden Spezialisierung von 3D-Sensoren hinsichtlich spezifischer Anwendungen dar. Für Rapid-Prototyping Anwendungen wurden Systeme entwickelt, die weitestgehend automatisch zumeist Objekte von einigen Zentimetern Größe in hoher Auflösung digitalisieren konnten [LaserDesign]. Dementgegen wurden für multimediale Anwendungen Systeme entwickelt, die kostengünstig sehr schnell ästhetisch ansprechende Computer-Modelle erstellen können, dafür aber keine Anforderungen an besondere Genauigkeiten erfüllen müssen [Busch91]. Andere Anwendungsbeispiele finden sich im Bereich der Bekleidungsindustrie zur Erfassung von Maßgrößen des menschlichen Körpers [Contour], in der Robotik zur automatischen Kartografierung von Gebäudeumgebungen [Weber01] und zur Hindernisvermeidung bei autonomen mobilen Robotern [Jörg94] oder aber bei der Dokumentation und Konservierung von Denkmälern [Levoy99][Neugebauer98].

Wenig Beachtung hat bisher die dreidimensionale Erfassung von großen Objekten gefunden. Mit immensem technischen Aufwand werden ganze Planeten dreidimensional erfasst und kartografiert. Die Vermessung von Autos oder Häusern ist aber ebenso oft nur mit hohem Zeitaufwand oder starken Einschränkungen möglich. Vor allem hier gilt nach wie vor, dass für jede neue Aufgabe neue Systeme entwickelt werden müssen, was einer großflächigen Verbreitung solcher Systeme im industriellen Umfeld entgegensteht. Im Rahmen dieser Arbeit werden Konzepte und neue Sensorsysteme vorgestellt, die in diesem Umfeld neue Lösungen bieten, da sie sich vor allem durch ihre hohe Flexibilität und kostengünstige Realisierung auszeichnen. Im Folgenden soll daher zunächst der aktuelle Stand der Technik in Form der zur Zeit verbreitetsten 3D-Sensoren aufgezeigt werden. Den Abschluss dieses Kapitels bildet die Vorstellung einer Auswahl von verfügbaren Sensorsystemen und der Anwendungsgebiete in deren Umfeld diese Systeme entstanden sind.

1.1 Begriffserklärung

Ein Problem einschlägiger Literatur über die dreidimensionale Objekterfassung ist die unterschiedliche Verwendung von Begriffen und Definitionen. Vor allem in Verkaufs- und Werbeprospekten wird oft der Begriff der Genauigkeit mit der sensoriellen Auflösung gleichgesetzt. Da im weiteren keine einheitliche Spezifikation für die Begriffe 3D-Sensor, 3D-Scanner und 3D-Sensorsysteme existieren, soll hier kurz auf die im Rahmen dieser Arbeit verwendete Bedeutung der wesentlichen Begriffe der dreidimensionalen Objekterfassung eingegangen werden.

3D-Sensor: Das minimal nötige Instrumentarium zur Entfernungsmessung wird als 3D-Sensor bezeichnet. Bei aktiven optischen 3D-Sensoren bezeichnet das im Allgemeinen die Einheit aus Lichtquelle, Empfangssystem und gegebenenfalls vorhandener mechanischer Kopplung zwischen beiden, nicht jedoch mechanische Ablenkeinheiten.

Das Messfeld des 3D-Sensors kann sich dabei über eine Dimension erstrecken, wenn alle erfassbaren Raumpunkte auf einer Geraden liegen, oder bei komplexeren Systemen über zwei bis drei Dimensionen, wenn die erfassbaren Raumpunkte eine Ebene bzw. ein Volumen aufspannen.

3D-Scanner: Die zusätzliche Ausstattung eines 3D-Sensors mit mechanischen Ablenkeinheiten zur Vergrößerung des Messfeldes des Sensors führt zum 3D-Scanner. Die Ablenkeinheit kann den 3D-Sensor als Ganzes bewegen oder nur Teilkomponenten des 3D-Sensors, wie z.B. Lichtquelle oder Kamera bei optischen Sensoren. Alternativ kann die Ablenkung auch indirekt über Spiegel erfolgen, die im Sichtfeld der Kamera liegen oder den Strahlengang einer Lichtquelle ändern.

3D-Sensorsystem: Die anwendungsspezifische Realisierung eines 3D-Sensors oder eines 3D-Scanners in Kombination mit der notwendigen Rechnerhardware und Software wird als 3D-Sensorsystem bezeichnet. Merkmale von 3D-Sensorsystemen bestehen oft in der Kopplung von mehreren 3D-Sensoren oder -Scannern zu einem Gesamtsystem, das auf eine Anwendung hin optimiert ist. Ein 3D-Sensorsystem umfasst aber auch die vorbereitende Präparation der zu vermessenden Objekte zur Unterstützung der dreidimensionalen Objekterfassung, z.B das Aufbringen von optischen Marken auf dem Objekt.

Auflösung: Ein wesentliches Merkmal eines 3D-Sensors besteht in der erreichbaren Auflösung. Damit wird der geometrische Abstand beschrieben, den zwei Punkte im Messraum haben dürfen, um vom 3D-Sensor noch unterschieden werden zu können. Im weiteren existiert der Begriff der *Tiefenauflösung*, also der Auflösung des Sensors entlang der Hauptmessrichtung. Die *laterale Auflösung* beschreibt die Auflösung orthogonal zur Hauptmessrichtung. Da die Auflösung nur in seltenen Fällen im gesamten Messraum gleich ist, sondern in vielen Fällen in der Nähe des 3D-Sensors hoch ist und mit größerem Abstand schlechter wird, wird der Begriff der *mittleren Auflösung* eingeführt; der durchschnittlichen Auflösung über den gesamten Messraum.

Kennzahl: Die maximale Zahl der Punkte, die über den gesamten Tiefenbereich eines 3D-Sensors aufgelöst werden können, wird als Kennzahl des Sensors beschrieben. Die Kennzahl kann im Allgemeinen aus der Sensorkonfiguration und der eingesetzten Hardware abgeleitet werden. Der Quotient aus bekanntem Tiefenmessbereich und Kennzahl ergibt die mittlere Tiefenauflösung. Bei kamerabasierten Sensoren ist die Pixelzahl der Kamera in vielen Fällen ein Maß für die Kennzahl des Sensors. Bei einer Detektionsgenauigkeit von einem halben Pixel und einer Bildabmessung von 500 Pixeln erreicht ein einfacher Triangulationssensor eine Kennzahl von 1000.

Genauigkeit: Die qualitative Umsetzung von realen Messpunkten über Messwerte des Sensorsystems in geometrische Koordinaten als Repräsentation im Rechner wird als (Mess-) Genauigkeit bezeichnet. Die Maßzahl der *mittleren Genauigkeit* beschreibt den mittleren Abstand, den ein realer Punkt im Rechner innerhalb eines

Referenzkoordinatensystems zur tatsächlichen Lage im Raum im gleichen Koordinatensystem erhält. Die Genauigkeit ist in jedem Fall durch die halbe Auflösung begrenzt. Bei einer mittleren Auflösung von beispielsweise einem Millimeter kann die Genauigkeit bestenfalls einen halben Millimeter betragen.

Kalibrierung: Das algorithmische Verfahren zur Berechnung der Abbildungsfunktion von Sensormesswerten in geometrische Koordinaten, wird als Kalibrierung bezeichnet. Dazu werden in den meisten Fällen mit dem 3D-Sensor Testmessungen an Referenzkörpern bekannter Ausprägung durchgeführt. Da mit einer begrenzten Zahl solcher Testmessungen normalerweise nicht alle Sensormesswerte erreicht werden können, wird über ein mathematisches Modell eine Repräsentation des 3D-Sensors im Computer gebildet. Die Parameter dieses Modells werden im Rahmen der Kalibrierung berechnet. Die Güte des Modells ist ein unmittelbares Maß für die erreichbare Genauigkeit des 3D-Sensors. Im Rahmen dieser Arbeit wird die Kalibrierung in zwei Zusammenhängen verwendet. Die *Kamerakalibrierung* beschreibt die Bestimmung der internen Parameter von elektronischen Kameras, wie zum Beispiel Brennweite oder Maßzahl der so genannten Kissenverzerrung. Unter *Kalibrierung* oder *Stereokalibrierung* wird die Bestimmung der externen Parameter, also der Lage der Kameras bezüglich eines Referenzkoordinatensystems verstanden.

1.2 Klassifizierung von 3D-Sensoren

Zum weiteren Verständnis dieser Arbeit wird im Folgenden, ein Überblick über den aktuellen Stand der Technik in Bezug auf verfügbare 3D-Sensoren erstellt. Die Auflistung erhebt keinen Anspruch auf Vollständigkeit, sondern dient lediglich dazu, die wichtigsten 3D-Sensoren vorzustellen und den Bereich ihrer Einsatzgebiete, sowie die Grenzen der jeweiligen Verfahren vorzustellen.

Zur Differenzierung der hier vorgestellten 3D-Sensoren wird eine Klassifizierung nach unterschiedlichen Kriterien durchgeführt. Das Hauptmerkmal von 3D-Sensoren ist das zugrundeliegende technische Prinzip. Die meistverwendeten Prinzipien basieren auf der Ausnutzung der physikalischen Eigenschaften des Lichts, aber auch Sensoren auf Basis von Ultraschall oder taktile Systeme sind weit verbreitet.
Der Schwerpunkt bei der Klassifizierung liegt entsprechend dem in dieser Arbeit verwendeten Grundprinzip auf optischen Sensoren. Ein mögliche Unterscheidung dieser Sensoren ist die Klassifizierung nach aktiven oder passiven Systemen. Aktive optische Sensoren zeichnen sich dadurch aus, dass der Sensor optische Signale in die zu erfassende Szene einbringt. Die Reflektion dieser optischen Signale auf dem Objekt kann wiederum im Sensor ausgewertet werden und ermöglicht die Entfernungsberechnung. Passive Sensoren dagegen werten ausschließlich die optische Information aus, die von der Szene reflektiert wird. Zur Kontrastverstärkung wird die Szene üblicherweise diffus beleuchtet, oft reicht allerdings auch das vorhandene Tageslicht aus.

Die weitergehende Unterteilung unterscheidet die aktiven Sensoren nach der Dimensionalität, mit der gemessen wird. Eindimensional messende Sensoren zeichnen sich dadurch aus, dass

sich alle erfassbaren Punkte im Raum als Funktion in Abhängigkeit eines Parameters beschreiben lassen. Üblicherweise beschreibt diese Funktion eine Gerade mit dem Ursprung im 3D-Sensor. Alle Punkte, die sich von einem zweidimensional messenden Sensor erfassen lassen, können dementsprechend als Funktion in Abhängigkeit von zwei Parametern beschrieben werden. Analog dazu lassen sich räumlich messende Sensoren als Funktion in Abhängigkeit von drei Parametern beschreiben.

Die passiven Sensoren werden detaillierter nach ihrer Funktionsweise vorgestellt, da prinzipbedingt oft Einschränkungen bezüglich der Anwendungsgebiete in Kauf genommen werden müssen oder Kontextwissen über die zugrundeliegende Aufgabenstellung notwendig ist. Das Funktionsprinzip des Sensors ist oft sehr stark an eine spezielle Aufgabenstellung gebunden.

Die folgende Grafik gibt einen Überblick über die vorgestellten 3D-Sensoren. Neben den aktiven und passiven optischen Sensoren werden vergleichend noch zwei alternative Prinzipien vorgestellt: Ultraschallsensoren und taktile Sensoren.

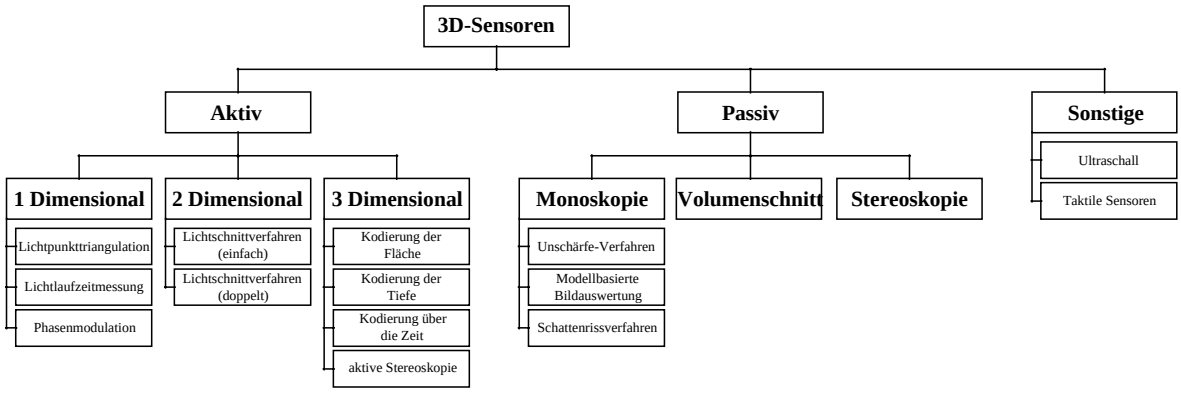
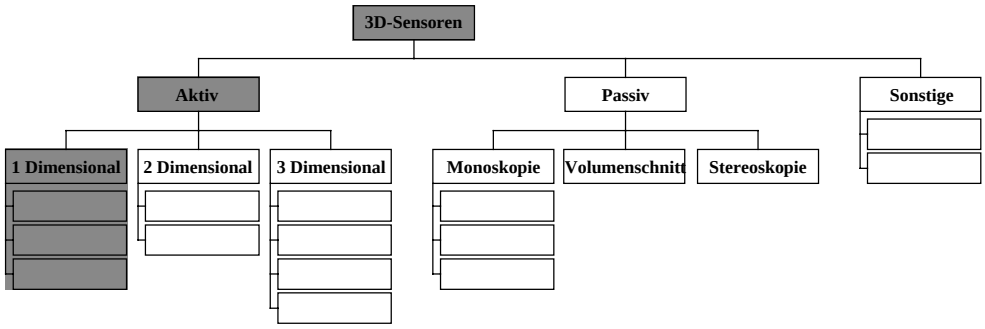


Abbildung 1.1: Klassifizierung von 3D-Sensoren

1.2.1 Aktive optische 3D-Sensoren

1.2.1.1 Eindimensional messende 3D-Sensoren



Die Entfernungsmessung in nur einer Dimension scheint auf den ersten Blick wenig sinnvoll, da nur mit zusätzlicher mechanischer Ablenkeinheit ein Messbild mit räumlicher Ausdehnung erzeugt werden kann. Andererseits werden besonders diese Sensoren oft spezialisiert, bei-

spielsweise auf eine kostengünstige Realisierung, eine hohe Genauigkeit oder einen großen Messbereich.

Lichtpunkttriangulation
Das einfachste und älteste optische Messprinzip ist die Lichtpunkttriangulation. Ein Sensor besteht aus einer photosensitiven Zeile und einer Lichtquelle. Bei der photosensitiven Zeile werden normalerweise CCD-Zeilen eingesetzt. Alternativ können auch PSD[1]-Sensoren verwendet werden. Als Lichtquelle werden idealerweise Laser eingesetzt, mit denen ein sehr eng gebündelter Strahl erzeugt werden kann. Zudem erlaubt das monochromatische Licht von Lasern die Verwendung von angepassten Farbfiltern vor der photosensitiven Zeile, die nur das Licht des Lasers durchlassen.
Die Messrichtung des Sensors wird durch den Laser vorgegeben, die Reflektion des Lasers auf einem Messobjekt wird von der photosensitiven Zeile erfasst. Durch den Abstand der photosensitiven Zeile vom Laser in einer definierten Entfernung (Basisabstand) entsteht ein funktionaler Zusammenhang zwischen der Entfernung des Lichtpunktes auf dem Objekt und der Lage dieser Reflexion auf der photosensitiven Zeile (Abbildung 1.2).

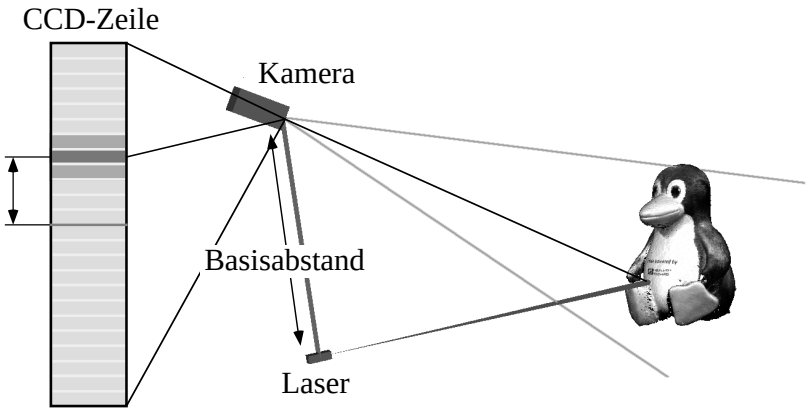


Abbildung 1.2: Lichtpunkttriangulation

Da die photosensitive Zeile eine begrenzte Größe hat, wird durch den Aufbau des Sensors auch der Messraum begrenzt. Die Auflösung wird direkt durch die Eigenschaften der photosensitiven Zeile bestimmt. So wird beispielsweise beim Einsatz einer CCD-Zeile mit 500 Photozellen und einer doppelt pixelgenauen Erfassung des Lichtpunktes eine mittlere Kennzahl von 1000 erreicht. Die Auflösung ist aber systembedingt nicht über den ganzen Messraum linear. In der Nähe des Sensors wird eine hohe Auflösung erreicht, die mit zunehmender Entfernung abnimmt. Ein weiteres Problem besteht in potentiellen Abschattungen, die auftreten, wenn der Laserpunkt gegenüber der photosensitiven Zeile verdeckt ist. Tiefe Löcher auf einem Objekt können so beispielsweise nicht erfasst werden. Im Weiteren muss häufig ein Kompromiss zwischen der Lichtstärke des Lasers und der Empfindlichkeit des Sensors geschlossen werden. Ein starker Laser erlaubt die Vermessung von dunklen Oberflächen, führt aber bei hellen Flächen schnell zu einer Überlastung der photosensitiven Zeile. Mit einem zu schwachen Laser können andererseits dunkle Flächen nicht mehr vermessen werden.

1. Photo Sensitive Device

Die einzelnen Komponenten dieses Sensortyps sind einfach und billig, so dass Sensoren dieses Typs sehr günstig hergestellt werden können. Sensoren mit hochauflösenden CCD-Zeilen erreichen auch auf Messbereichen bis zu einem Meter hohe Genauigkeiten von einem zehntel Millimeter und darunter, was einer Kennzahl von 10.000 entspricht.

Lichtlaufzeitmessung
Ein naheliegendes Prinzip, das technisch gesehen allerdings nur sehr aufwändig zu realisieren ist, ist die Messung der Lichtlaufzeit. Von einer Lichtquelle, üblicherweise einem Laser, wird ein kurzzeitiger starker Lichtimpuls ausgesandt. Gleichzeitig wird eine Zeitmessung begonnen. Sobald die Reflexion des Lichtimpulses in einem Detektor empfangen wird, wird die Zeitmessung gestoppt. Die verstrichene Zeit ist linear proportional zur doppelten Objektentfernung (Abbildung 1.3). Bei einer Lichtgeschwindigkeit von $2.9979*10^8$m/s ist es für eine Ortsauflösung von einem Zentimeter notwendig, ein Zeitintervall von 66 Pikosekunden aufzulösen. Beim Einsatz eines digitalen Zählers zur Zeitmessung muss dafür eine Zählfrequenz von etwa 15 Gigahertz erreicht werden. Bekannte Systeme erreichen Genauigkeiten von wenigen Zentimetern. Der auftretende Messfehler ist normalerweise stochastisch verteilt ist, daher kann die Genauigkeit durch Mittelung mehrerer Messwerte auf weniger als einen Zentimeter gesteigert werden.

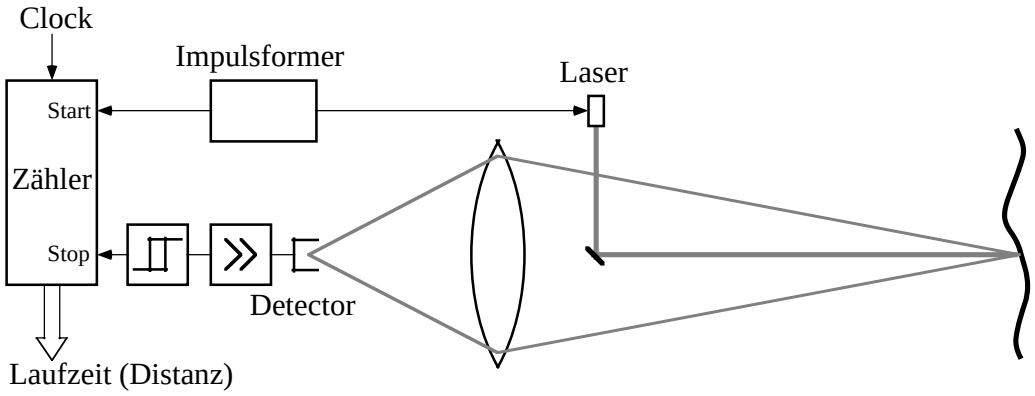


Abbildung 1.3: Lichtlaufzeit

Da Sender und Empfänger auf einer Achse liegen, entfällt das Problem der Abschattung. Die Aussendung von sehr kurzen aber starken Lichtimpulsen ermöglicht einen Messbereich von wenigen Metern bis zu mehreren Kilometern für spezielle Anwendungen. Über den gesamten Messbereich wird eine gleichmäßige Auflösung erreicht, die ausschließlich durch das Zeitmesssystem bestimmt wird. Zudem wird durch die starken Lichtimpulse auch die Vermessung von dunklen oder transparenten Oberflächen möglich.

Phasenmodulation
Ein weiteres Prinzip, bei dem Lichtquelle und Detektor auf einer Achse angebracht sind, beruht auf dem Prinzip der Phasenmodulation. Auf das Lichtsignal, üblicherweise einem kontinuierlich strahlenden Laser, wird ein periodisches Intensitätsprofil aufmoduliert. Die Frequenz des Intensitätsprofils wird dabei so eingestellt, dass eine Phasenlänge von mehreren Metern entsteht. Das Lichtsignal, das von einem Objekt reflektiert wird, kehrt normalerweise phasenverschoben in den Empfänger zurück (Abbildung 1.4).

Über einen Phasenvergleicher kann dieser Phasenversatz gemessen werden. Innerhalb des Bereichs einer Phasenlänge besteht ein linearer Zusammenhang zwischen Objektentfernung und Phasenversatz, so dass die Entfernung zum Objekt gemessen werden kann. Die erreichbare Genauigkeit liegt bei diesem Verfahren bei wenigen Millimetern, allerdings wird der Messbereich durch die Phasenlänge begrenzt. Zur sicheren Messung des Phasenversatzes muss ein kontinuierlich strahlender Laser eingesetzt werden. Aus Sicherheitsgründen kann keine hohe Lichtintensität eingesetzt werden, da der Laser ab einer gewissen Grenze nicht mehr als augensicher gilt. Das Phasenmodulationsverfahren ist aus diesem Grund deutlich schlechter bei dunklen oder spiegelnden Oberflächen. Der wesentliche Vorteil gegenüber dem Laufzeitverfahren besteht aber in der höheren Genauigkeit, außerdem wird gegenüber der Lichtpunkttriangulation das Problem der Abschattung vermieden.

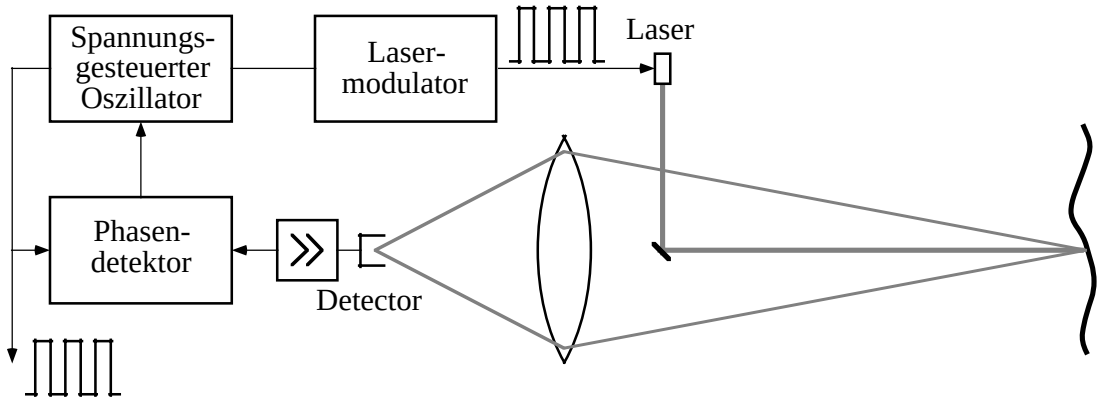
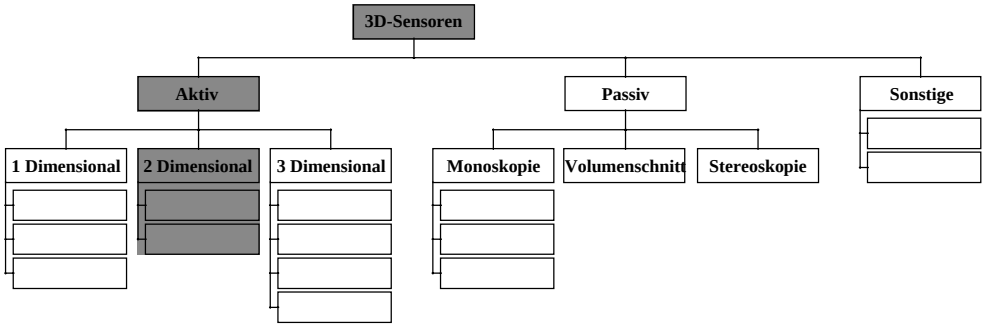


Abbildung 1.4: Phasenmodulationverfahren

Robuster ist das beim AccuRange3000LV [Acuity] eingesetzte Verfahren. Die aufmodulierte Frequenz wird so gesteuert, dass die Objektentfernung ein vielfaches der aufmodulierten Phasenlänge beträgt, die gemessene Phase also immer der ausgesendeten Phase entspricht. Der Sensor wird unempfindlicher gegenüber der Oberflächenbeschaffenheit des Objektes, da kein echter Phasenversatz mehr gemessen werden muss, sondern nur noch die Frequenz so eingestellt wird, dass die Summe aus empfangenem und ausgesendetem Signal maximal ist. Die Objektentfernung ist damit proportional zur Modulationsfrequenz.

Neben diesen beispielhaft erwähnten Verfahren gibt es eine Vielzahl weiterer eindimensional messender optischer 3D-Sensoren. Die meisten dieser Verfahren sind Varianten oder Kombinationen der vorgestellten drei Prinzipien. Alle Sensoren haben spezifische Stärken und Schwächen und qualifizieren sich dadurch für spezielle Aufgaben. Bei der Messung von großen Entfernungen wird häufig das Laufzeitverfahren eingesetzt (militärische Zielsysteme). Für sehr genaue Distanzen auf kurze Entfernungen kommen Lichtpunkttriangulationssysteme zum Einsatz (Koordinatenmessung). Phasenmodulationsverfahren bilden oft einen Kompromiss zur Messung von Entfernungen im Bereich von wenigen Metern.
Zur Vermessung von Oberflächen oder Objekten mit räumlicher Ausdehnung werden die Sensoren oft mit mechanischen Ablenkeinheiten versehen. Der Sick Sensor [Sick] ist ein Beispiel für die Kombination eines eindimensional messenden 3D-Sensoren mit einem Rotationsspiegel.

1.2.1.2 Zweidimensional messende 3D-Sensoren



Lichtschnittverfahren (einfache Triangulation)
Die bekannteste und am weitesten verbreitete Methode zur dreidimensionalen Vermessung ist das Lichtschnittverfahren. Das Funktionsprinzip ist der Lichtpunkttriangulation sehr ähnlich. Anstatt eines Laserpunktes wird ein zur Linie aufgeweiteter Laser verwendet, die photosensitive Zeile wird durch eine Matrixkamera ersetzt. Wenn die Kamera in einem definierten Abstand zur Laserebene steht (Basisabstand), dann bildet sich die Reflexion des Lasers auf dem Objekt als Profillinie im Kamerabild ab (Abbildung 1.5).

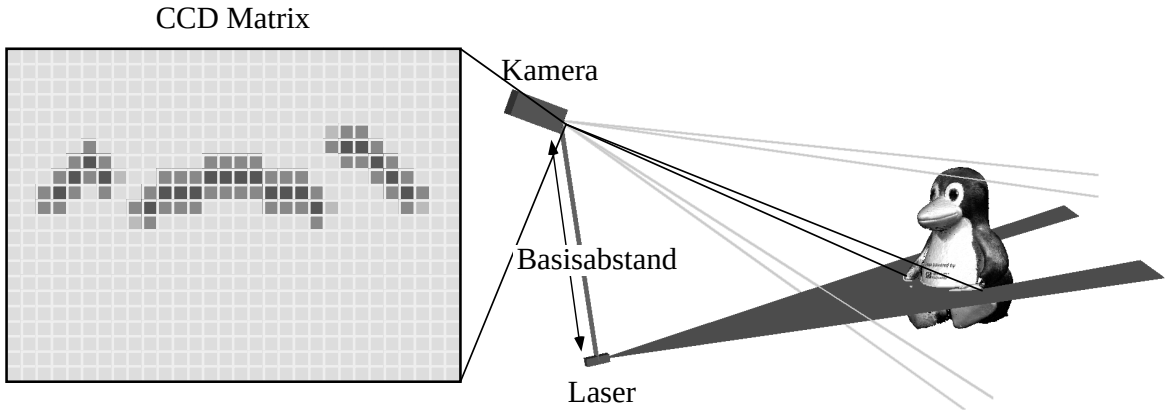


Abbildung 1.5: Lichtschnittverfahren (einfache Triangulation)

Das Lichtschnittverfahren entspricht damit einem vielfach parallelen Lichtpunkttriangulationsverfahren. Die Berechnung zur Entfernungsbestimmung, die bei der Lichtpunkttriangulation nötig ist, kann beim Lichtschnittverfahren für jede Spalte der Matrixkamera durchgeführt werden. Bei heute üblichen Auflösungen elektronischer Kameras können dadurch mit einer Bildaufnahme mehrere hundert Entfernungswerte berechnet werden. Allerdings ist die Bildwiederholfrequenz von elektronischen Kameras in vielen Fällen an Videonormen gebunden, so dass in der Regel nicht mehr als dreißig Bilder pro Sekunde ausgewertet werden. Im Einsatz sind allerdings auch Anwendungen mit Hochgeschwindigkeitskameras [Ranger], die mehr als tausend Aufnahmen pro Sekunde zulassen. Bei einer Auflösung von beispielsweise 256 Pixeln im Quadrat wird so eine Datenrate von 256.000 3D-Punkten pro Sekunde erreicht. Diese Geschwindigkeit wird allerdings durch hohe Systemkosten erkauft, die vor allem durch den Preis der Kamera erzeugt werden. Zusätzlich erfordern die kurzen Belichtungszeiten starke Laser und sehr definierte Beleuchtungsbedingungen.

Lichtschnittverfahren (Doppeltriangulation)
Eine Erweiterung, die das Problem der Abschattung von einfachen Triangulationsköpfen zum Teil löst, besteht in der zusätzlichen Verwendung einer zweiten Kamera, die symmetrisch zur ersten Kamera auf der anderen Seite der Lichtebene angebracht ist (Abbildung 1.6).

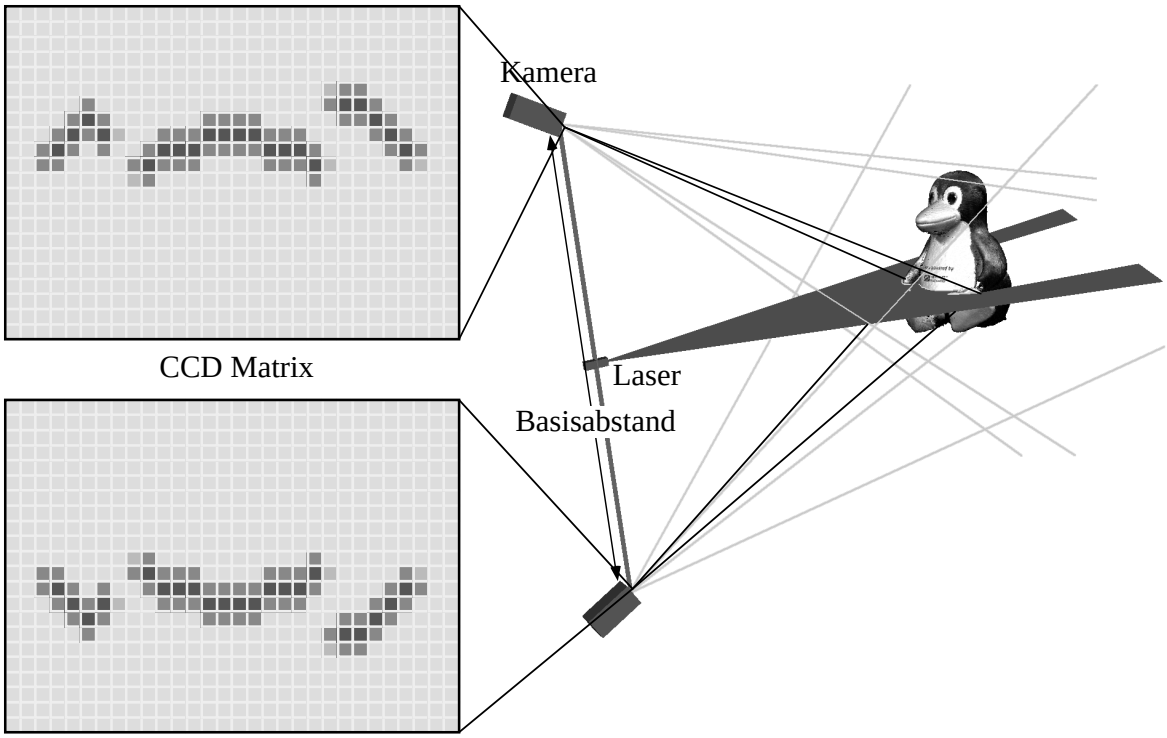


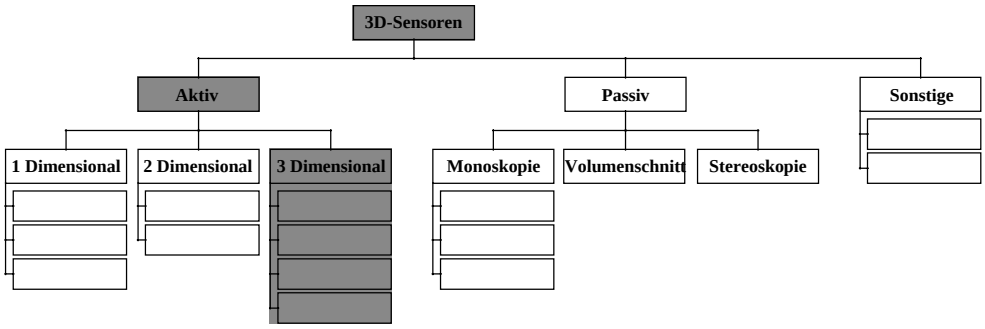
Abbildung 1.6: Lichtschnittverfahren mit zwei Kameras (Doppeltriangulation)

Bei synchroner Aufnahme der Bilder beider Kameras wird jeder Punkt redundant in beiden Kameras erfasst. Diese Redundanz kann zur Erhöhung der Genauigkeit ausgenutzt werden. Punkte, die gegenüber einer Kamera abgeschattet werden, werden aber nur in einfacher Genauigkeit erfasst. Tiefe Löcher im Objekt können zu Abschattungen in beiden Kameras führen, so dass keine 3D-Punkte berechnet werden können. Wenn die Kameras abwechselnd Bilder aufnehmen, kann die Datenrate des Systems verdoppelt werden.

Die Auflösung, die mit dem Lichtschnittverfahren erreicht werden kann, liegt im Allgemeinen deutlich unter der Auflösung, die mit der Lichtpunkttriangulation erreicht wird. Das ist vor allem darauf zurückzuführen, dass die Auflösung einer Bildspalte einer Matrixkamera deutlich geringer ist, als die Pixelzahl, die mit Zeilensensoren erreicht werden. Typische Kennzahlen von 3D-Sensoren nach dem Lichtschnittverfahren liegen bei Werten zwischen 1.000 und 2.000.
Zur Erweiterung des Erfassungsraums des Sensors um eine Dimension wird der 3D-Sensor häufig mit Linearführungen kombiniert. Beispiele für derartige Sensoren sind der TOPAS-Scanner, der Vitus Scanner und der Cyberware Scanner WB4, die in Kapitel 1.3 (Beispiele für 3D-Sensorsysteme) vorgestellt werden. Beim Cyberware Desktop Scanner wird das Objekt auf einem Drehteller gedreht, während der Sensor selber fest steht.

Aufgrund des weit verbreiteten Prinzips des Lichtschnittverfahrens existieren eine Vielzahl von weiteren Varianten von Sensorköpfen und Kombinationen der Sensorköpfe mit Positioniersystemen. Die hier und in Kapitel 1.3 (Beispiele für 3D-Sensorsysteme) dargestellten Beispiele sollen nur eine Vorstellung über die Vielfältigkeit der möglichen Variationen und Einsatzgebiete dieses Verfahrens vermitteln.
Eine Fortsetzung der Idee der Lichtpunkttriangulation und des Lichtschnittverfahrens stellt der folgende Abschnitt dar: die Erfassung von Objekten mit räumlich messenden 3D-Sensoren.

1.2.1.3 Räumlich messende 3D-Sensoren



Eine naheliegende Erweiterung zu den beschriebenen Lichtschnittsensoren mit dem Zweck, mehr 3D-Information pro Bild aufzunehmen, liegt in der Projektion von vielen Linien gleichzeitig. Da die Linien allerdings identifizierbar sein müssen, um eine eindeutige Berechnung der Tiefenwerte zu gewährleisten, können bereits kleine Sprünge oder Stufen auf der Objektoberfläche zu Problemen führen. Die projizierten Linien sind dann ebenfalls im Bild unterbrochen und die eindeutige Zuordnung einer Liniennummer ist dann nicht mehr gewährleistet (Abbildung 1.7).

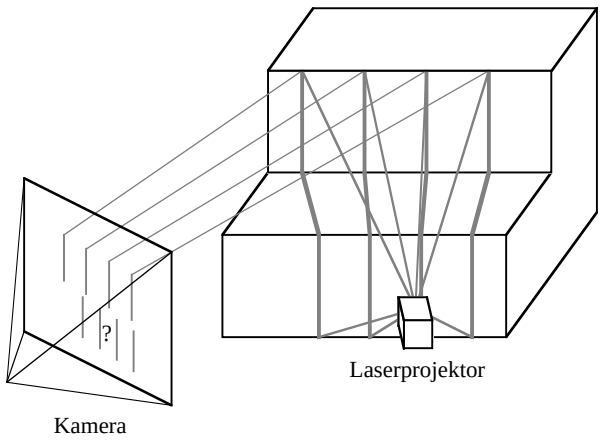


Abbildung 1.7: Linienzuordnung beim Mehrlinienverfahren

Vor allem aufgrund dieser Schwäche haben sich die einfachen Mehrlinienverfahren nicht durchgesetzt. Es wurden allerdings Erweiterungen dieses Verfahren entwickelt, die die Unterscheidung der Linien erlauben. Zur Zuordnung einer Liniennummer werden die Linien mit

einer Kodierung versehen, die im Rahmen dieser Arbeit nach drei verschiedenen Varianten unterschieden werden:

- Kodierung der Fläche (Strukturlichtverfahren),
- Kodierung der Tiefe (Phasen-Shift Verfahren, Moirée-Verfahren),
- Kodierung über die Zeit (Gray-Code Verfahren).

Obwohl auch hier nur ein grober Überblick über die möglichen Varianten des kodierten Lichtansatzes gegeben werden kann, stellen die im Folgenden vorgestellten Verfahren die wesentlichen Grundprinzipien dar.

Kodierung der Fläche
Bei der Kodierung der Fläche wird ein strukturiertes Muster auf das zu vermessende Objekt projiziert, das in den Kamerabildern wiedererkannt werden kann. Aus der Forschung sind zwei Beispiele für dieses Verfahren bekannt. In [Vuylsteke90] wird ein Schachbrettmuster verwendet. Die Kodierung erfolgt über die Ecken der Felder, die entweder hell oder dunkel abgebildet werden. Bei vier Ecken pro Feld können sechzehn Zustände kodiert werden. Die eindeutige Identifizierung der gesamten Fläche ist damit zwar noch nicht gegeben, der eindeutige Tiefenbereich vergrößert sich allerdings beträchtlich (Abbildung 1.8). Die Analyse des Codes eines Feldes und der Vergleich mit benachbarten Felder sichert aber im Allgemeinen die Identifizierung des Musters.

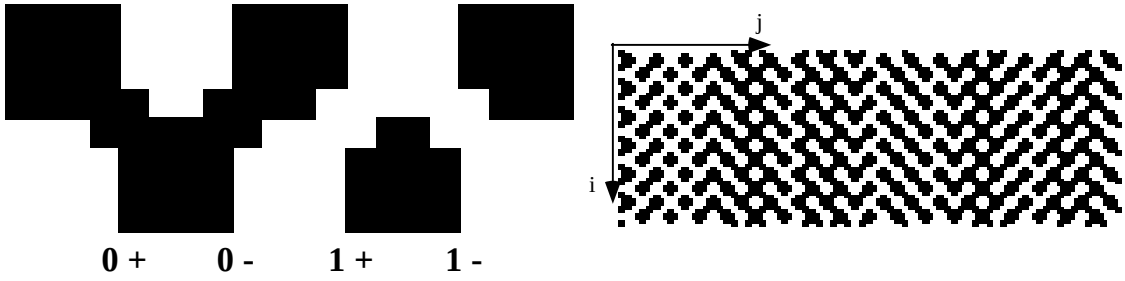


Abbildung 1.8: Flächig kodiertes Muster

In [Sasse94] wird Farbinformation zur Unterscheidung von Linien verwendet. Die Analyse der Farbinformation in den Kamerabildern ermöglicht die korrekte Zuordnung der Liniennummer zur projizierten Linie. Der Schwachpunkt dieses Verfahrens liegt in der Anfälligkeit gegenüber der Farbe des erfassten Objektes. Das Farbsignal, das vom Kamerasystem aufgenommen wird, setzt sich aus der Farbe des projizierten Musters und der Farbe des Objektes zusammen. Wenn das Objekt selbst bunt ist, wird das Farbsignal des Projektors verfälscht und es kann zu Fehlmessungen kommen. Der große Vorteil der Kodierung der Fläche liegt darin, dass für die Erfassung eines Objektes genau ein Bild aufgenommen werden muss. Die Vermessung lässt sich also innerhalb eines Bruchteils einer Sekunde durchführen. Zugunsten einer robusteren und genaueren Vermessung wird dieser Vorteil bei der Kodierung der Tiefe und der Kodierung über die Zeit aufgegeben. Diese Varianten eignen sich dafür allerdings nicht mehr für bewegte Objekte.

Kodierung der Tiefe
Die bekanntesten Verfahren zur Kodierung der Tiefe sind das Phasen-Shift Verfahren und das Moirée Verfahren. Beim Phasen-Shift Verfahren werden wenige Linien projiziert, die ein sinusförmiges Intensitätsprofil aufweisen. Zur Berechnung der Tiefe wird die Liniennummer und der Phasenwinkel bestimmt. Da vor allem die Berechnung des Phasenwinkels sehr anfällig gegenüber der Oberflächenstruktur des Objektes ist, werden mehrere Aufnahmen des Objektes gemacht, bei denen die Linienordnung des projizierten Musters um den Bruchteil einer Phase verschoben wird. Bei der verbreitetsten Variante werden vier Linienmuster, deren Phase um jeweils 90 Grad verschoben ist, verwendet. Die genaue Phasenlage kann dann aus den vier Intensitätswerten berechnet werden, die den Pixeln des Kamerabildes entnommen werden. Der Tiefenmessbereich ist allerdings nur innerhalb einer Phase eindeutig, dafür kann der Phasenwinkel sehr genau berechnet werden. Bekannte Realisierungen erreichen eine Genauigkeit von wenigen Grad.

Durch die Verkleinerung des Schiebewinkels bei entsprechender Erhöhung der Aufnahmezahl kann die Genauigkeit fast beliebig gesteigert werden. Die Erhöhung der Phasenbreite vergrößert den Tiefenmessbereich. Die Auslegung des Verfahrens wird im Allgemeinen von den Anforderungen an die Messdatenqualität abhängig gemacht. Es ist jeweils ein Kompromiss zwischen Tiefenmessbereich, Genauigkeit und Aufnahmedauer zu finden (Abbildung 1.9 a).

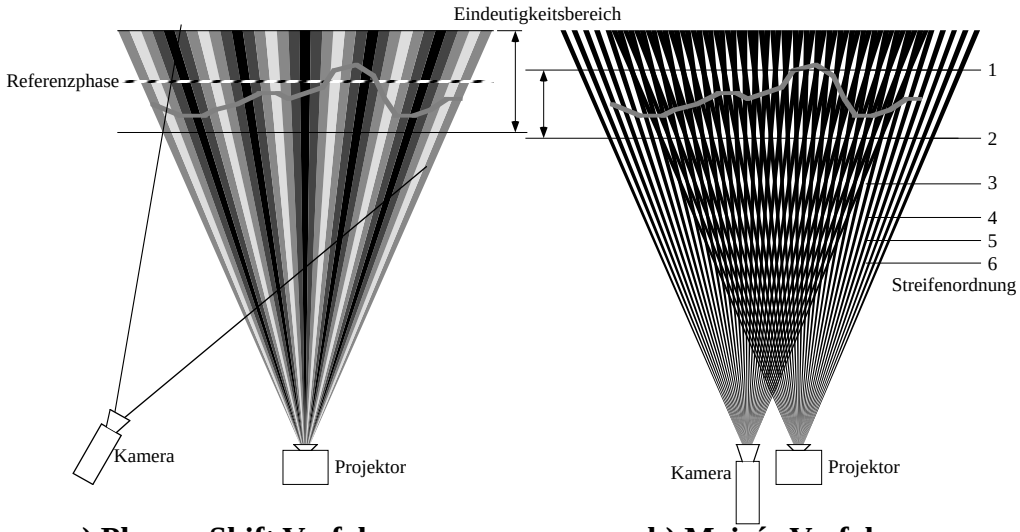


Abbildung 1.9: Kodierung der Tiefe

Beim Moirée-Verfahren wird die Interferenz zwischen zwei regelmäßigen Linienstrukturen ausgenutzt. Je nach Anforderung und Applikation sind aus Forschung und Praxis eine Vielzahl von Varianten des Moirée-Verfahrens bekannt, die sich vor allem durch den Aufbau unterscheiden, mit dem die Linienstrukturen überlagert werden. Hier soll nur eine einfache Variante vorgestellt werden, bei dem ein feines Linienmuster mit Hilfe eines Projektors auf das Messobjekt projiziert wird. Die Pixelmatrix der Aufnahmekamera stellt die zweite Linienstruktur dar, über die die Interferenz gebildet wird (Abbildung 1.9 b). Wenn die Bildebene der Kamera koplanar zur Projektionsebene des Projektors ist, entsteht ein direkter Zusammenhang zwischen der Streifenordnung der Interferenzlinien und der Tiefe des Objektes - alle Punkte mit der selben Streifenordnung haben die gleiche Tiefe. Solange nur eine Aufnahme gemacht

wird, kann im Allgemeinen nur die relative Tiefe des Objektes gemessen werden. Zur absoluten Vermessung der Tiefe innerhalb eines definierten Bereichs wird die Streifenordnung innerhalb mehrerer Aufnahmen variiert. Das kann in einfacher Weise durch die Veränderung des Abstandes zwischen Projektor und Kamera geschehen. Je größer der Abstand ist, um so feiner werden die Interferenzlinien; ein kleinerer Abstand führt zu gröberen Streifenordnungen. Wird der Abstand zwischen Projektor und Kamera so variiert, dass sich die Streifenordnung mit jeder Aufnahme verdoppelt, dann können die Interferenzlinien der einzelnen Aufnahmen direkt in einen Binär-Code umgesetzt werden, der die Abzählung der Interferenzlinien erlaubt und damit die absolute Vermessung der Tiefe des Objektes ermöglicht. Es sei allerdings erwähnt, dass dieses Moirée-Verfahren sehr präzise Kameras, Projektoren und Justierungen verlangt, um die beschriebenen Eigenschaften zu bieten. In der Praxis wird daher das Moirée-Verfahren hauptsächlich zur qualitativen Vermessung von Objekten eingesetzt, beispielsweise zur Auffindung von Beulen in Karosserieteilen oder zum Vergleich von Bauteilen mit Referenzobjekten zur Qualitätskontrolle. Vor allem die Eigenschaft des Verfahrens, dass mit sehr feiner Streifenordnung sehr geringe Tiefenunterschiede erkannt werden können, ohne dass die absolute Tiefe bekannt sein muss, kommt dabei zum Tragen.

Kodierung über die Zeit
Beim sogenannten Gray-Code Verfahren werden in einer zeitlichen Folge mehrere Muster mit Linien unterschiedlicher Breite auf das Objekt projiziert. Dabei halbiert sich die Linienbreite mit jedem neuen Muster. Die Linien überlagern sich derart, dass die zeitliche Folge der Hell-Dunkel Information jedes Pixels einen binären Gray-Code ergeben. Alle Pixel mit der gleichen Kodierung beschreiben eine Projektionslinie (Abbildung 1.10).

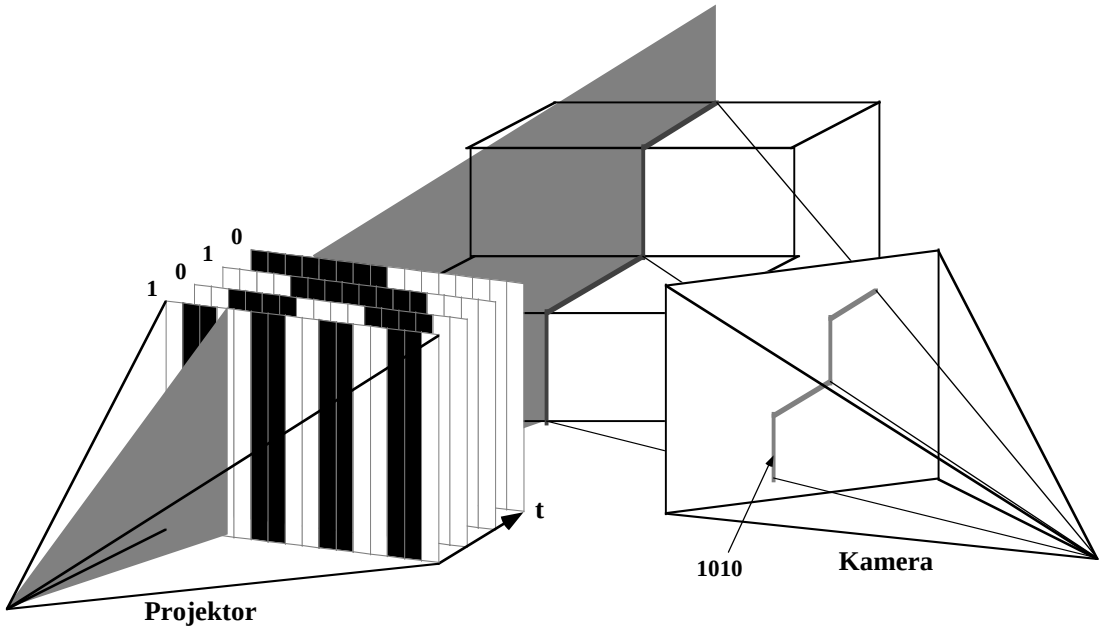


Abbildung 1.10: Gray-Code Verfahren

Mit jeder Verfeinerung der Linienstruktur in einem neuen Projektionsbild verdoppelt sich die Zahl der kodierten Linien. So können mit acht Projektionsmustern bereits 256 verschiedene Linien-Codes erzeugt werden. Die Kodierung im Gray-Code System garantiert, dass sich benachbarte Linien nur in einem Bit unterscheiden. Bitfehler aufgrund von Kamerarauschen

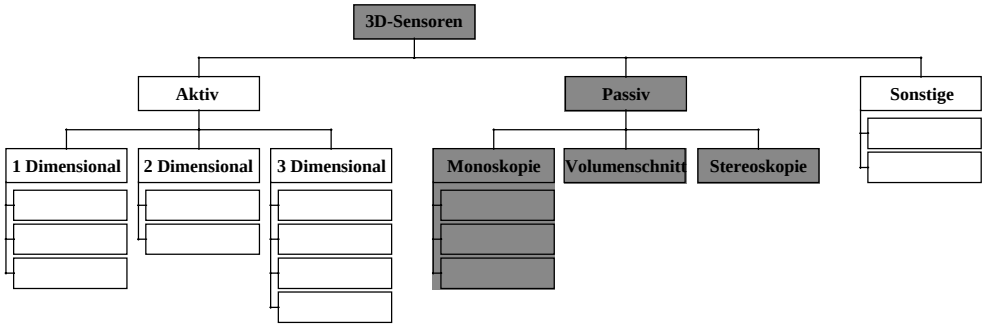
erzeugen also maximal einen Fehler von einer Linienbreite, sofern die Linien noch ausreichend vom Kamerasystem aufgelöst werden.
Das Gray-Code Verfahren ist sehr robust gegenüber Oberflächenstrukturen der Messobjekte, da nur binäre Information ausgenutzt wird. Jede Linie ist eindeutig identifizierbar, somit ist auch eine absolute Vermessung der Tiefe problemlos möglich. Ein Nachteil des Gray-Code Verfahrens ist die geringe Tiefenauflösung, die durch die Zahl der kodierten Linien begrenzt wird. Die Projektion von beispielsweise 256 verschiedenen Linien-Codes in einer Sequenz von acht Bildern erlaubt auch nur die Unterscheidung von 256 Tiefenstufen. Bei einer Tiefe von einem Meter entspricht das einer Auflösung von vier Millimetern. Aus diesem Grund wird das Gray-Code Verfahren häufig mit dem Phasen-Shift Verfahren kombiniert. Die Linien-Codes werden dabei zur eindeutigen Zuordnung einer Phasennummer verwendet, über den Phasenwinkel wird eine hohe Auflösung erzielt. Die Kombination beider Verfahren erlaubt bereits mit normal auflösenden Kameras eine Steigerung der Tiefenauflösung um einen Faktor von fünf bis zehn gegenüber dem einfachen Gray-Code Verfahren. Im industriellen Bereich hat das Gray-Code Verfahren in Kombination mit dem Phasen-Shift Verfahren die weiteste Verbreitung erfahren, da es sich neben seiner hohen Auflösung ebenso durch eine hohe Robustheit auszeichnet. Einsatzbereiche finden sich besonders im Reverse-Engineering, bei dem Produkte hochgenau dreidimensional erfasst werden müssen, um dann im Computer wieder in CAD-Daten umgesetzt zu werden.

1.2.1.4 Aktive Stereoskopie

Die aktive Stereoskopie sei hier zunächst nur der Vollständigkeit halber erwähnt. Dieses Verfahren beschreibt alle stereoskopischen Sensorsysteme, die um eine strukturierte Lichtquelle erweitert werden. Die Berechnung von Raumkoordinaten geschieht über Triangulation zwischen Korrespondenzen in den Bildern der eingesetzten Kameras. Ein zentrales Problem der rein passiv arbeitenden Stereo-Verfahren, die Suche von korrespondierenden Punkten, wird durch die zusätzliche Lichtquelle gelöst. Wenn im einfachsten Fall ein Punktlaser eingesetzt wird, besteht die Bestimmung von korrespondierenden Punkten in den Kamerabildern im Auffinden des Laserpunktes. Erweiterte Verfahren sehen den Einsatz von farbigen Mustern oder Sequenzen von strukturierten Bildern als aktive Lichtquelle vor.

Da die aktive Stereoskopie das zentrale Thema dieser Arbeit darstellt, soll hier nicht weiter auf die verschiedenen Varianten dieses Verfahrens eingegangen werden. Die zugrundeliegenden Ideen, Verfahren und mathematischen Grundlagen sind in Kapitel 3 (Aktive Stereoskopie) dargestellt.

1.2.2 Passive optische 3D-Sensoren



1.2.2.1 Monoskopie

Ziel der monoskopischen Verfahren ist die dreidimensionale Erfassung von Objekten mit nur einer Kamera. Dabei werden hier alle Verfahren als monoskopisch bezeichnet, bei denen die Daten entweder aufgrund einer einzigen Aufnahme gewonnen werden oder bei denen mehrere Aufnahmen von einer Szene gemacht werden, die aber nicht direkt miteinander in Beziehung gebracht werden.
Ohne zusätzliches Wissen über die Szene kann im Allgemeinen allerdings mit nur einer Kamera und einer Aufnahme keine dreidimensionale Information gewonnen werden, daher werden verschiedene Voraussetzungen an die zu vermessenden Objekte gestellt, die die im Folgenden vorgestellten Verfahren charakterisieren.

Fokus-Defokus Analyse (FDA)
Bei strukturierten und kontrastreichen Objekten kann das sogenannte FDA-Verfahren eingesetzt werden. Die Kamera nimmt das Messobjekt auf, und im Bild wird das Maß der Bildschärfe berechnet. Die gewonnene Kennzahl pro Pixel ist proportional zum Abstand des Objektes von der Kamera und kann bei bekannter Lage und Parametern der Kamera in eine dreidimensionale Koordinate umgesetzt werden. Da die Bildschärfe üblicherweise über eine Gradientenanalyse, also dem Vergleich der Helligkeit von benachbarten Pixeln, bestimmt wird, ist die Qualität der Daten von mehreren Faktoren abhängig. Je kleiner der Tiefenschärfebereich der Kamera ist, desto genauer kann einem scharfen Bereich im Bild eine Objekttiefe im Bild zugeordnet werden. Allerdings verkleinert sich der auswertbare Tiefenbereich, da oberhalb und unterhalb der fokussierten Tiefe sehr schnell Bereiche großer Unschärfe erreicht werden, die nicht mehr ausgewertet werden können. Im Weiteren kann nur bei einer kontrastreichen Oberfläche eine Bildschärfe berechnet werden. Bildbereiche mit geringem Kontrast führen zu Messfehlern. Aufgrund dieser schwerwiegenden Mängel wird in erweiterten Verfahren die Fokussierung der Kamera in mehreren Schritten variiert [Wei94], [Liu98]. Für jede Einstellung der Fokussierung wird die Bildschärfe jedes Pixels berechnet. Zur Berechnung eines Tiefenwertes wird für jedes Pixel das Bild mit der jeweils größten Schärfe verwendet. Auf Kosten der Messzeit wird dadurch eine deutliche Genauigkeitssteigerung erreicht, Objekte ohne kontrastreiche Oberfläche können jedoch nach wie vor nicht vermessen werden.

Modellbasierte Bildauswertung
Sofern bereits detailliertes Vorwissen über die aufgenommene Szene vorliegt und im Wesentlichen nur noch Lage, Größe oder Form der aufgenommenen Objektenbestimmt werden müssen, wird die sogenannte modellbasierte Bildauswertung betrieben. Das aufgenommene Bild wird nach Strukturen und Mustern durchsucht, um bekannte Objekte wiederzufinden. Gefundene Objekte werden analysiert und ausgewertet. In einfachen Fällen kann beispielsweise ein Ball im Bild gefunden werden, wenn nach kreisförmigen Objekten gesucht wird. Bei bekannter Größe des Balles kann über den Radius des Kreises im Bild auf die Entfernung des Balles geschlossen werden. Bei bekannter Entfernung kann alternativ die Größe des Balles berechnet werden. Bei Quadern oder Pyramiden kann zusätzlich die Lage und Orientierung des Objektes im Raum bestimmt werden. Bei komplexeren Objekten wird aber vor allem die automatische Auswertung zunehmend schwieriger, da die Zahl der beschreibenden Parameter des Objektes so groß wird, dass sie sich nicht mehr in einen Algorithmus zur Bildanalyse umsetzen lassen. Vor allem in diesen Fällen wird daher oft auf neuronale Algorithmen zurückgegriffen, die mit Beispielszenen trainiert werden. Eine begrenzte Menge von Objekten kann so wiedererkannt oder klassifiziert werden. In [Wichert95] ist eine Kamera auf einem AMR[1] montiert, der für die Arbeit in Bürogebäuden vorgesehen ist. Während der Fahrt des AMR werden Aufnahmen von der Umgebung gemacht, die mittels neuronaler Netze mit einer Datenbank verglichen werden. Bereits erfasste Flure oder Räume können wiedererkannt werden, so dass sich der AMR innerhalb des Gebäudes lokalisieren[2] kann.

Schattenrissverfahren
Das Tecmath-Contour System [Contour][Heiduk96] geht in der Spezialisierung noch einen Schritt weiter. Das eingesetzte Schattenrissverfahren dient ausschließlich zur Vermessung von Menschen. Die Szene, die von einer Kamera aufgenommen wird, besteht aus einer beleuchteten, weißen Wand, vor der eine Person in definierter Haltung steht. Durch den hohen Kontrast zwischen Person und beleuchtetem Hintergrund kann die Kontur der Person als Schattenriss extrahiert werden. Da die Haltung der Person bekannt ist, können über die Analyse der Kontur markante Bereiche, wie z.B. Hals, Schritt, Achseln oder Taille gefunden werden. Die Berechnung von Längen kann direkt über diese Bildinformation erfolgen. Zur Berechnung von Umfängen und zur Validierung von berechneten Größen wird die Person zusätzlich von der Seite und in einer weiteren Haltung aufgenommen. Über eine Datenbank mit Größentabellen werden den berechneten Werten Konfektionsgrößen der Bekleidungsindustrie zugeordnet. Zusätzlich werden noch individuelle Merkmale, wie die Rückenkontur oder die Ärmellänge festgehalten, so dass die Daten ausreichen, um passgenaue Maßkonfektion zu fertigen. Bildbeispiele für dieses Verfahren finden sich in Kapitel 1.3 (Beispiele für 3D-Sensorsysteme).

1.2.2.2 Volumenschnitt

Das Volumenschnittverfahren setzt die Separation des Messobjektes vom Hintergrund voraus. Das Messobjekt wird von mehreren Seiten aufgenommen, und die Objektdaten werden in ein Volumenmodell übertragen. Dazu werden über die Lage der Kamera und die bekannten Abbildungsparameter alle Bereiche im Bild, die dem Hintergrund zugeordnet werden können, im Volumenmodell als Freiraum markiert. Wenn das Objekt von mehreren Seiten aufgenommen

1. autonomer mobiler Roboter
2. die Lage des Gebäudes relativ zum Aufnahmesystem (AMR) ist dazu mathematisch äquivalent

wird und die Hintergrundbereiche in den entsprechenden Ansichten im Volumenmodell als Freiraum gekennzeichnet werden, bleibt letztendlich nur noch das Objekt im Volumenmodell übrig (Abbildung 1.11).

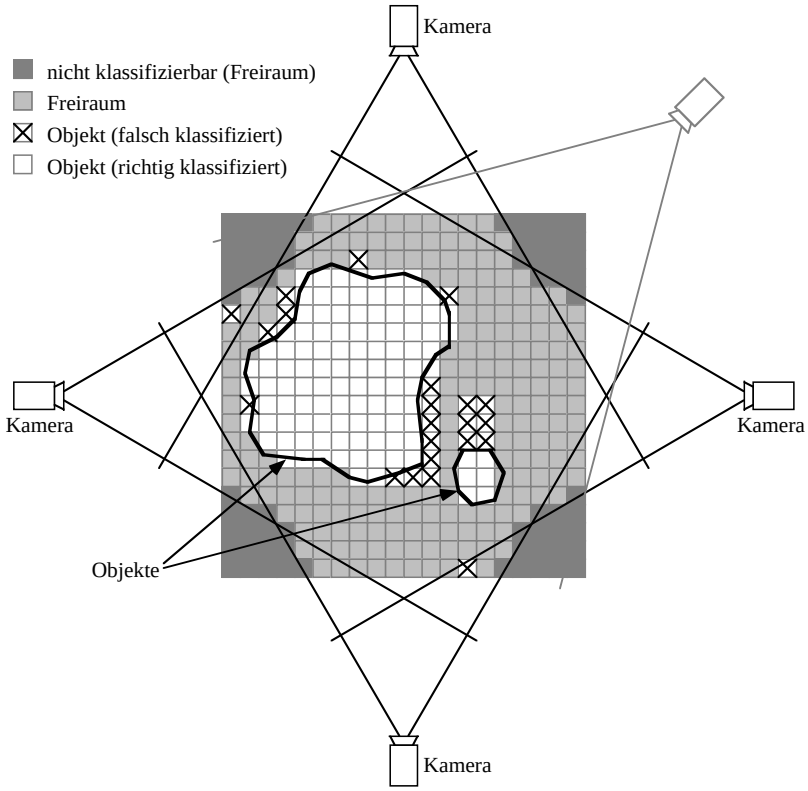


Abbildung 1.11: Volumenschnittverfahren (in zwei Dimensionen, vier Ansichten)

Die Zahl und Lage der Ansichten bestimmt dabei entscheidend die Qualität der Rekonstruktion. Der Freiraum zwischen den beiden Objekten im Beispiel wird zwar erkannt, das kleinere Objekt wird aber deutlich zu groß rekonstruiert. Bereits eine weitere Ansicht aus der richtigen Richtung (von oben rechts) könnte das Ergebnis deutlich verbessern.

Das Volumenmodell, das im Allgemeinen als Voxelraum realisiert wird, kann im Anschluss an die Objektvermessung in Flächenmodelle umgesetzt werden. Zur einfachen Durchführung von mehreren Aufnahmen aus unterschiedlichen Richtungen werden häufig Drehteller eingesetzt [Busch91]. Die Kamera wird fest positioniert und die Lage der Kamera zum Drehteller wird mit einem geeigneten Kalibrierkörper bestimmt. Nach jeder Aufnahme wird der Drehteller weiter gedreht, der aktuelle Winkel wird über Enkoder in den Rechner übertragen.

Die technischen Voraussetzungen für das Volumenschnittverfahren sind sehr einfach, allerdings können Objekte nur in Grenzen vollständig und genau erfasst werden. Konkave Bereiche, die aus keiner Bildansicht den Rand des Objektes bilden, können nicht vermessen werden. So ist beispielsweise der Innenraum einer Tasse nicht zu erfassen. Die Genauigkeit des Volumenschnittverfahrens hängt von der Auflösung der Kamera und des Volumenmodells ab. Meistens begrenzt das Volumenmodell das System. Falls ein Voxelraum eingesetzt wird, kann bei einer Voxelgröße von einem Millimeter auch nur eine Genauigkeit von einem Millimeter erreicht werden. Allerdings ist zu beachten, dass bei einer Objektgröße von 20 Zentimetern

und einer Auflösung von einem Millimeter ein Voxelraum mit 200 Voxeln im Kubik, also bereits acht Millionen Voxeln verwaltet werden müssen.
Volumenschnittverfahren werden daher weniger in messtechnischen Aufnahmen, als vielmehr im Bereich des Multimedia eingesetzt, da optisch ansprechende 3D-Modelle auch bei geringer Auflösung durch die Darstellung mit hochauflösender Textur erreicht werden. Das Scanbook der Firma 3D-Dimension ist eine Realisierung des Volumenschnittverfahrens. In Kapitel 1.3 (Beispiele für 3D-Sensorsysteme) finden sich Bildbeispiele für Objektmodelle, die mit dem Scanbook erstellt wurden.

1.2.2.3 Passive Stereoskopie

Die dreidimensionale Vermessung von Objekten mit passiven 3D-Sensoren wird nach heutigem Stand der Technik hauptsächlich mit stereoskopischen Sensorsystemen durchgeführt. Die zu erfassende Szene wird entweder mit einer Kamera in mehreren Bildern aus unterschiedlichen Ansichten erfasst, oder mehrere Kameras erfassen gleichzeitig die gleiche Szene aus verschiedenen Ansichten. Zur dreidimensionalen Rekonstruktion der Szene müssen in den verschiedenen Kamerabildern Korrespondenzen gefunden werden, also Punkte in den Bildern, die dem gleichen Ort auf dem Objekt entsprechen. Aus einer Korrespondenz, die in wenigstens zwei Kamerabildern enthalten ist, kann ein Raumpunkt berechnet werden. Bei ausreichend vielen Korrespondenzen, die über die ganze Szene verteilt sind, kann über die berechneten Raumpunkte die Szene rekonstruiert werden (Abbildung 1.12).

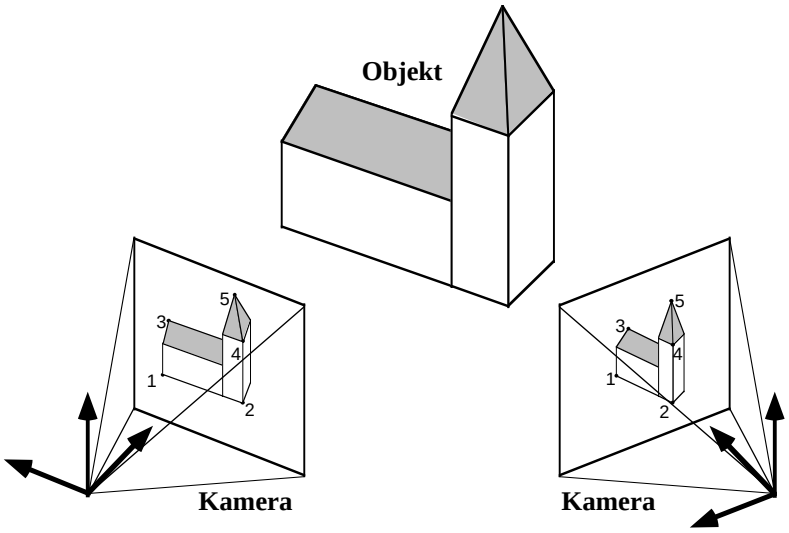


Abbildung 1.12: Passive Stereoskopie

Als Grundlage zur Rekonstruktion der Szene müssen zwei wesentliche Probleme gelöst werden. Zum Einen müssen die Positionen im Raum bekannt sein, an denen die Aufnahme gemacht wurden. Bei bekannten Abbildungseigenschaften der Kameras sichern sie die korrekte Berechnung von Raumpunkten aus korrespondierenden Punkten. Die Verfahren zur Bestimmung der Aufnahmepositionen und der Abbildungseigenschaften werden im Allgemeinen als Kalibrierung bezeichnet. Zum Anderen müssen die aufgenommenen Bilder der Szene nach gleichen oder ähnlichen Bereichen durchsucht werden, aus denen dann Korrespondenzen

abgeleitet werden können. Erst die Sicherstellung beider Funktionen ermöglicht die dreidimensionale Erfassung eines Objektes.

Kalibrierung
Für die Bestimmung der Aufnahmepositionen der Bilder sind zwei Verfahren üblich. Sehr robust und genau ist die Kalibrierung mit Hilfe von Kalibrierkörpern. Ein Referenzkörper mit leicht erkennbaren Marken, die bei der Bildauswertung des Stereosystems als Korrespondenzen dienen können, wird in das Sichtfeld der Kameras gestellt. Aus bekannter Lage der Marken können mit bekannten Verfahren [Tsai87] Lage und Orientierung sowie Abbildungseigenschaften der Kameras berechnet werden. Solange die Lage der Kameras nicht verändert wird, können mit dieser Information aus Korrespondenzen Raumpunkte berechnet werden. Weitere Details und Information zu dieser modellbasierten Kalibrierung sind in Kapitel 2.1 (Kamerakalibrierung) und Kapitel 2.2 (Modellbasierte Kalibrierung) zu finden.
Die sogenannte Stereokalibrierung kommt ohne Kalibrierkörper aus. Die Lagebestimmung der Kameras wird ausschließlich mit Hilfe von gefundenen Korrespondenzen durchgeführt. Die Beziehung zwischen gleichen Punkten in zwei verschiedenen Kamerabildern kann durch die sogenannte Epipolargeometrie beschrieben werden (Abbildung 1.13). Zu jedem Punkt in einem Kamerabild kann eine Epipolarlinie im Bild der anderen Kamera berechnet werden, auf der der korrespondierende Punkt liegen muss. Diese sogenannte Epipolareinschränkung kann mathematisch berechnet werden, wenn eine ausreichende Zahl von Korrespondenzen vorliegt. Im Weiteren kann aus dem Ergebnis dieser Rechnung die relative Lage der beiden Kameras zueinander (R, T) abgeleitet werden [Huang86].

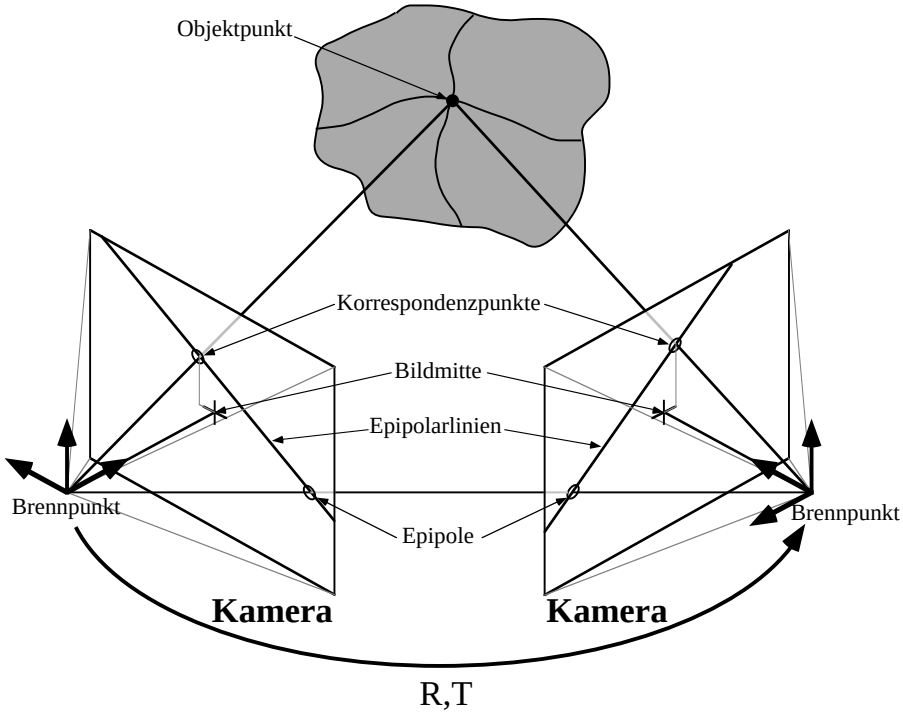


Abbildung 1.13: Epipolargeometrie

Die Qualität der berechneten Epipolargeometrie hängt sehr stark von der Qualität der Korrespondenzen ab. Bereits ein Fehler von wenigen Pixeln in der Positionsbestimmung der Bildpunkte einer Korrespondenz kann zu Fehlern in der Lageinformation der Kameras führen, die

keine vernünftige Rekonstruktion mehr zulässt. Zusätzlich fehlt dem Gesamtsystem die Skalierungsinformation der Szene. Die Geometrie der Szene kann im Idealfall zwar rekonstruiert werden, die Größe der erfassten Objekte ist aber nicht berechenbar. Eine genaue Beschreibung des Verfahrens der Stereokalibrierung und der mathematischen Grundlagen ist in Kapitel 2.4 (Stereokalibrierung) zu finden.

Korrespondenzpunktsuche
Das eigentliche Problem der Stereoskopie, vor allem bei automatischen Verfahren, liegt in der Suche von korrespondierenden Punkten in zwei Kamerabildern. Häufig eingesetzt werden dabei Korrelationsverfahren, die zu einem Bildausschnitt eines Bildes den Bereich mit der höchsten Übereinstimmung im anderen Bild suchen. Um Mehrdeutigkeiten zu vermeiden und die Berechnungsgeschwindigkeit zu steigern, wird dabei der Suchbereich im zweiten Bild deutlich eingeschränkt. Das setzt aber voraus, dass die Stereokameras in definierter Konfiguration relativ dicht zueinander stehen, damit sich die Bildpaare ausreichend ähnlich sehen. Sofern das System vorab kalibriert ist, kann zur weiteren Vereinfachung der Berechnung der Korrelation die Epipolareinschränkung verwendet werden. Probleme entstehen, wenn die Bilder kontrastarm sind, da die Korrelation dann falsche oder ungenaue Ergebnisse liefert [Casott99]. Im ungünstigsten Fall liefern alle getesteten Bereiche den gleichen Korrelationswert, so dass keine eindeutige Korrespondenz berechnet werden kann.
In [Digiclops] werden drei Kameras miteinander verbunden, von denen eine neben und eine über der zentralen Kamera angeordnet ist. Die optischen Achsen der Kameras verlaufen parallel, so dass die Bildebenen koplanar zueinander liegen. Die Epipolarlinien verlaufen damit exakt zeilenweise durch die Pixel bei den nebeneinanderliegenden Kameras, beziehungsweise spaltenweise bei den übereinanderliegenden Kameras, so dass die Korrelation in Echtzeit berechnet werden kann. Zudem können die Ergebnisse aufgrund der redundanten Auslegung des Stereosystems über drei Kameras validiert werden. Mit dem Stereosystem können reale Szenen aufgenommen werden, die Genauigkeit ist aber prinzipbedingt sehr begrenzt. Die Messtiefe reicht von 0,5 Metern bis unendlich. Bei einer Entfernung von 1,5 Metern beträgt die Auflösung etwa fünf Millimetern, nimmt aber mit größerer Entfernung schnell ab.

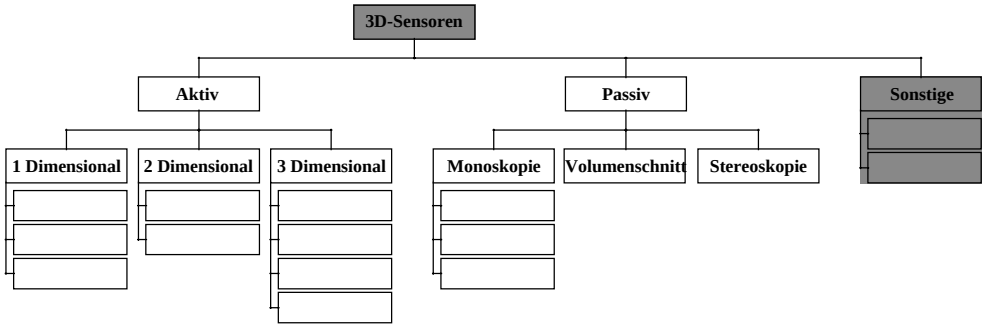
Flexibler, aber auch rechenaufwändiger sind Verfahren, die auf der Extraktion von Merkmalen in den Bildern basieren. Die Zuordnung von gleichen Merkmalen führt zu den gesuchten Korrespondenzen. Zu der Klasse dieser Verfahren zählen unter anderem

- *feature based visibility*: Die Methode der merkmalsbasierten Sichtbarkeit basiert auf dem Ansatz der Approximation von Sichtbarkeitsberechnungen und liefert Informationen für die dreidimensionale Rekonstruktion von Objekten [Lerios95].
- *feature extraction*: In den Stereobilder werden aus geometrischen Erscheinungsformen (z.B. elektronische Bauteile) bestimmte Muster extrahiert. Der Vergleich der gefundenen Muster in beiden Stereobildern führt zu Korrespondenzen, mit denen das Bauteil rekonstruiert werden kann [Haberäcker95].
- *pattern matching*: In einem Bild werden markante Muster gesucht, die im anderen Bild wiedergefunden werden können. Das Verfahren ist der Korrelation sehr ähnlich, nur dass der Auswertung eine Bildvorverarbeitung vorausgeht, die die interessanten Bereiche extrahiert. Die Korrelation mit homogenen Bereichen, bei denen keine vernünftigen Ergebnisse erwartet werden können, wird damit vermieden [Ottmann90].

Allerdings setzen auch diese Verfahren das Vorhandensein von Merkmalen voraus, die sicher in beiden Bildern wiedererkannt werden. Die erreichbare Genauigkeit ist im Allgemeinen höher als bei Korrelationsverfahren, da in gewissem Maß Modellwissen in die Auswertung eingebracht wird. Eine flächige Vermessung der Szene ist aber im Allgemeinen nicht möglich. Vielmehr sind die Verfahren darauf spezialisiert, Lage und Form von Objekten im Raum zu bestimmen.

Vor allem um die Unsicherheit bei der Auswertung der Bilder in Bezug auf die Korrespondenzpunktsuche zu minimieren, wurden zahlreiche Verfahren entwickelt, die die manuell interaktive Unterstützung des Benutzers erfordern. Das Programm präsentiert die aufgenommenen Bilder der Szene und durch Anklicken mit der Maus werden vom Benutzer korrespondierende Punkte benannt. Die Software verbessert die Genauigkeit der Punkte und berechnet die Geometrie. Zudem kann der Benutzer Beziehungen zwischen Punkten definieren, um so geometrische Primitive, wie Linien oder Flächen zu bilden. Diese Primitive wiederum werden zu größeren Objekten der Szene zusammengesetzt [Ristow98].
Die Programme Canoma und Facade [Debevec96] erfordert vom Benutzer die Einpassung von einfachen geometrischen Körpern, wie Quader und Prismen in die Szene. Die Ecken der Körper werden auf vergleichbare Merkmale im Bild eingepasst. So kann beispielsweise ein Haus durch einen Quader, auf dem ein Dreiecksprisma angeordnet ist, modelliert werden. Über die Verschiebung und Verzerrung, die notwendig ist, um die geometrischen Körper mit der Szene in Deckung zu bringen, kann die Lage der Kamera berechnet werden, mit der die Szene aufgenommen wurde. Prinzipiell reicht für diese Art der Rekonstruktion bereits ein Bild. Um aber die Abbildungseigenschaften der Kamera genau bestimmen zu können und eine vollständige Rekonstruktion der Szene zu erreichen, werden mehrere in zeitlicher Folge und aus unterschiedlichen Richtungen aufgenommenen Bilder verwendet. Neben der Einpassung der geometrischen Körper muss der Benutzer dann noch die Zuordnung von gleichen Körpern in den verschiedenen Bildern leisten.
Beispiele für die manuell interaktive Stereoskopie, die beispielsweise mit Hilfe des Software Systems Facade erstellt wurden, finden sich in Kapitel 1.3 (Beispiele für 3D-Sensorsysteme).

1.2.3 Andere Sensortypen



Die bisher beschriebenen Verfahren zur dreidimensionalen Erfassung von Objekten beruhen auf optischen Verfahren. Es ist aber festzustellen, dass sich auch andere technische Verfahren dazu eignen, Entfernungen und Positionen im Raum zu messen. Exemplarisch werden hier ultraschallbasierte und taktile Verfahren erklärt. Im praktischen Einsatz sind aber auch Radar- und Sonarverfahren, oder die Positionsbestimmung über die Laufzeitmessung von Radiowel-

len, als Beispiel sei hier das GPS[1] genannt. Die genauere Beschreibung aller Verfahren würde allerdings den Rahmen der Arbeit sprengen.

1.2.3.1 Ultraschall

Ein sehr bekanntes und verbreitetes Verfahren zur Entfernungsmessung basiert auf Ultraschall. Ein Sender erzeugt einen Ultraschallimpuls, die Laufzeit des Signals zum Objekt und wieder zurück zum Empfänger wird gemessen. Bei einer Schallgeschwindigkeit von etwa 300 Metern pro Sekunde ist bereits mit sehr geringem technischen Aufwand eine Auflösung von wenigen Millimetern problemlos realisierbar. Ultraschallentfernungsmesser werden dadurch sehr kostengünstig und robust realisierbar. Eine bekannte Realisierung solcher Sensoren findet sich in den Objektivsteuerungen von Polaroid-Kameras. Die Entfernungsinformation wird zur Scharfstellung des erfassten Objektes auf der Bildebene verwendet. Eine neuere Anwendung sind Einparkhilfen bei Fahrzeugen. Die Ultraschallsensoren werden in der Stoßstange angebracht und messen die Entfernung zu Hindernissen. Bei zu kurzen Abständen wird Alarm ausgelöst, um eine mögliche Kollision zu vermeiden.

Ein Besonderheit von Ultraschallsensoren besteht in der räumlichen Ausbreitung von Schallwellen. Die Entfernung eines Objektes ist auf wenige Millimeter genau messbar, die Lage des Objektes zur Schallquelle kann aber nur mit einer Genauigkeit von 20 bis 30 Grad bestimmt werden. Im Weiteren kann es an schallharten Oberflächen zu Totalreflexionen des Ultraschallimpulses kommen, so dass keine oder zu große Entfernungen gemessen werden. Fremdschall kann zu Phantomhindernissen führen, die wegen der Störinformation detektiert werden. Bei autonomen mobilen Robotern werden daher mehrere Ultraschallsensoren rund um den Roboter angebracht. Sobald der Roboter in Bewegung ist, werden vorhandene Hindernisse mehrfach aus unterschiedlichen Richtungen gemessen. Über stochastische Verfahren kann dann gültige von ungültiger Information getrennt werden [Jörg94].

In aktuellen Forschungsprojekten werden die Schallimpulse mit Signalen moduliert. Die Analyse der Schallsignale im Empfänger erlaubt das Wiedererkennen der ausgesendeten Signale und damit die Vermeidung von Phantomhindernissen. Zusätzlich kann so mit mehreren Ultraschallsensoren, die unterschiedliche Signalformen verwenden, gleichzeitig gearbeitet werden. Über Korrelationsverfahren können in einem Empfänger die Signale von unterschiedlichen Sendern wiedererkannt werden. Damit ist es möglich, durch die Kombination verschiedener Sender und Empfänger die Datendichte und Datensicherheit deutlich zu steigern. Wenn sich die Empfänger an unterschiedlichen Positionen befinden, wird im Weiteren die Ortsauflösung verbessert, da ein Objekt in einer Messung aus unterschiedlichen Richtungen gemessen wird [Berg99].

1.2.3.2 Taktile Systeme

Als Referenzsysteme für optische 3D-Messsysteme werden heutzutage noch taktile Koordinatenmesssysteme eingesetzt. Eine Sonde wird durch mechanische Positioniersysteme über eine zu vermessende Fläche verfahren. Mit jeder Berührung der Sonde auf dem Objekt wird die aktuelle Lageinformation gespeichert und die Sonde neu positioniert. Alternativ dazu kann ein

1. Global Positioning System

Bediener bestimmte Positionen auf dem Objekt ferngesteuert anfahren und vermessen. Die Genauigkeiten solcher Systeme liegen bei wenigen Mikrometern, allerdings ist der notwendige technische Aufwand sehr hoch. Da zusätzlich die Messzeit zur Erfassung eines Punktes sehr hoch ist, wird mit Koordinatenmesssystemen oft nur ein Ausschnitt des Objektes erfasst. Der Messvorgang selbst muss also sehr genau geplant werden, damit nur relevante Information gemessen wird. Weiche Objekte können im Allgemeinen nicht vermessen werden, da die Berührung der Sonde mit der Objektoberfläche nicht mehr hinreichend sicher detektiert werden kann und zudem die Gefahr der Verformung der Objektoberfläche durch die Sonde besteht. Vor allem aus diesen Gründen werden die Messköpfe daher mit optischen 3D-Sensoren versehen, die Vermessung kann dann berührungslos und mit einer höheren Datenrate geschehen.

Der Faro-Arm [Faro] kann manuell geführt werden. Über Encoder, die in den Gelenken des Arms angebracht sind, wird die Position und Lage der Sonde berechnet. Die Reichweite ist abhängig vom Modell auf wenige Dezimeter beschränkt. Die Genauigkeit liegt laut Herstellerangaben je nach Modell bei 10 bis 100 Mikrometern. Vor allem auf Grund seiner vielen Freiheitsgrade ist das System sehr flexibel einsetzbar. Der 3D-Systems Scanner kombiniert den Faro-Arm mit einem Lichtschnittsensor. Die Datenauswertung erfolgt in Echtzeit, so dass es in einfacher Weise möglich ist, auch komplexe Objekte dreidimensional zu vermessen. Bildbeispiele für beide Sensoren sind im folgenden Kapitel zu finden.

1.3 Beispiele für 3D-Sensorsysteme

Obwohl mittlerweile eine Vielzahl unterschiedlicher Anwendungen im Bereich der dreidimensionalen Objekterfassung bekannt sind, können hier nur einige wenige verfügbare Sensorsysteme vorgestellt werden. Ein Bereich, in dem sehr spezialisierte 3D-Sensorsysteme entwickelt wurden, ist die Vermessung von Menschen. Ziel dabei ist die Erfassung einer Vielzahl von Körpermaßen des Probanden in möglichst kurzer Zeit, z.B. zur Berechnung von Kleidergrößen oder zur Durchführung von anthropometrischen Reihenuntersuchungen. Eine maximale Körpergröße von 2,10m stellt dabei die erste Herausforderung in der Systemkonzeption des 3D-Scanners dar. Zudem darf die Aufnahmezeit zur Erfassung des Menschen nur wenige Sekunden betragen, da andernfalls natürliche Körperbewegungen des aufgenommenen Menschen zu ungenauen Daten führen können. Dass das Sensorsystem nicht gesundheitsgefährdend sein darf, versteht sich von selbst.

Im Rahmen des *TOPAS*-Projektes [Hamfeld98] wurde ein sehr kompaktes 3D-Sensorsystem entwickelt, das in nur 1,5 Sekunden einen Menschen erfassen kann. An zwei gegenüberliegenden Linearführungen ist jeweils ein einfacher Lichtschnittsensor angebracht, der während des Messvorgangs vertikal entlang des Menschen verfahren wird. Die Linearführungen sind miteinander gekoppelt, so dass sich die beiden 3D-Sensoren immer auf gleicher Höhe befinden. Die Laser der beiden 3D-Sensoren spannen eine Lichtebene auf, über die die Lichtschnitttriangulation durchgeführt wird. Besonders bemerkenswert beim *TOPAS*-Scanner ist die kurze Messzeit, die Bewegungsartefakte fast vollständig ausschließt. Erreicht wird die dafür notwendige Datenrate durch Spezialkameras des Typs *RANGER2200*, bei denen die Bildauswertung bereits auf einer auf dem CCD-Chip der Kamera integrierten Logik durchgeführt wird. Die Kameras werden digital angesteuert und sind nicht an Videonormen gebunden, so dass meh-

rere hundert Bilder pro Sekunde ausgewertet werden können. Um aber trotz dieser relativ teuren Kameras einen insgesamt kostengünstigen Scanner realisieren zu können, wurde auf redundante Auslegung der 3D-Sensoren verzichtet. Die dadurch entstehenden Abschattungen und Lücken in den Scandaten (vgl. Abbildung 1.14) werden durch eine intelligente Sensordatenauswertung ausgeglichen.

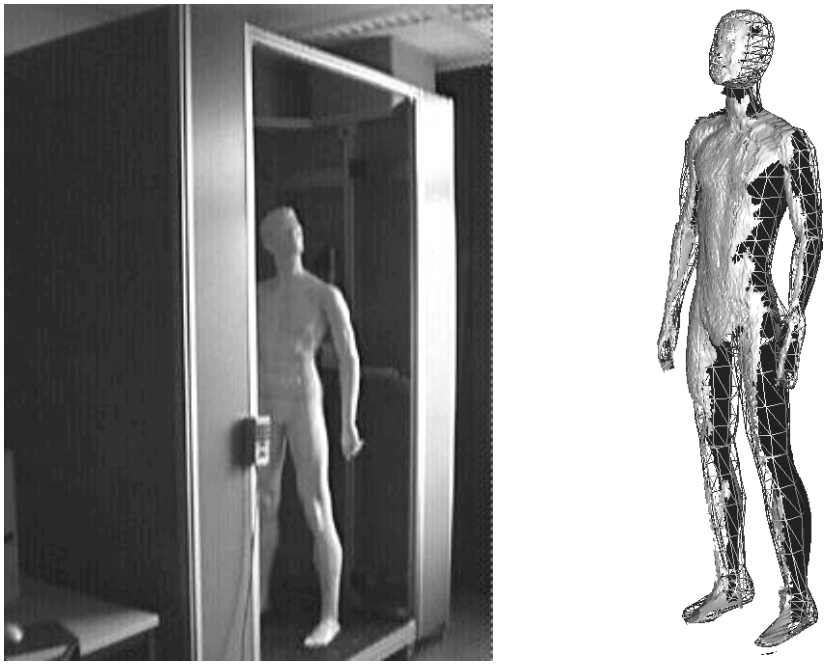


Abbildung 1.14: TOPAS-Scanner

Da das direkte Messen vor allem von komplexen Maßgrößen, direkt auf den Scandaten kaum möglich ist, wird das CAD-Menschmodell RAMSIS an die Daten angepasst. Durch einen evolutionären Algorithmus [Hansen96] werden die Parameter des Modells solange optimiert, bis das Modell die höchstmögliche Übereinstimmung mit der erfassten Person hat. Neben einer extremen Datenkompression hat dieses Verfahren den besonderen Vorteil, dass die individuellen Körpermaßdaten als CAD-Modell vorliegen. Spezifische Daten müssen nicht mehr aufwändig aus den Sensordaten extrahiert werden, sondern können direkt dem CAD-Modell entnommen werden. Zusätzlich kann das Modell als Ganzes, z.B. im Rahmen von Ergonomieuntersuchungen, verwendet werden.

Bei der Entwicklung der *VITUS* 3D-Scanner wurde mehr Wert auf Datenvollständigkeit, als auf hohe Messgeschwindigkeit gelegt. Der *VITUS Pro* Scanner kombiniert acht Lichtschnittsensoren (Doppeltriangulation), die in vier Gondeln an zwei Linearführungen aufgehängt sind. Die 3D-Sensoren haben zusammen vier Laser, die wie beim *TOPAS*-Scanner eine Laserebene aufspannen. Da die eingesetzten Kameras nach der europäischen CCIR-Norm arbeiten, werden pro Sekunde 25 Lichtschnitte aufgenommen und verarbeitet. Bei einer Gesamtmesszeit von 15 Sekunden entstehen so 375 Schnitte mit insgesamt etwa 750.000 Messpunkten. Durch die Vielzahl von Kameras wird eine fast vollständige Erfassung eines Menschen gewährleistet. Probleme entstehen nur noch im Kopfbereich bei dunklen Haaren, die optisch schwer zu erfassen sind, oder in Abschattungsbereichen unter den Achseln oder im Schritt des Menschen, die von den Lasern nicht erreicht werden. Durch vier zusätzliche Farbkameras kann der *VITUS*

Pro Scanner zusätzlich die Farbtextur des menschlichen Körpers erfassen, so dass auch optisch sehr realistische 3D-Scans erzeugt werden können.

VITUS Pro **VITUS Smart**

Abbildung 1.15: VITUS 3D-Scanner

Eine applikationsspezifische Weiterentwicklung des *VITUS Pro* Scanners ist der *VITUS Smart* Scanner. Auf unnötige Hardware wurde verzichtet, so dass der Scanner kostengünstig und modular realisiert werden kann. Üblicherweise vier voneinander unabhängige arbeitende Scan-Säulen werden um ein Podest platziert, auf dem der zu vermessende Mensch steht. Jede Säule ist mit einem Lichtschnittsensor (Doppeltriangulation) auf einer Linearführung ausgestattet. Damit der Laser eines Lichtschnittsensors nicht die Aufnahme eines anderen Sensors stört, wird jeder Laser synchron zu den zugehörigen Kameras gepulst. Jede Kamera belichtet also nur dann, wenn der entsprechende Laser im gleichen Lichtschnittsensor angeschaltet ist. Dieses zunächst aufwändig erscheinende Verfahren hat den Vorteil, dass eine mechanische Kopplung der Linearführungen entfallen kann. Die Forderung, dass alle Laser zusammen eine einzige Laserebene aufspannen, muss nicht mehr beachtet werden.

Bereits Anfang der 90er Jahre wurde in den USA der *Cyberware WB4* Scanner entwickelt. Auch bei diesem Scanner werden mehrere Lichtschnittsensoren an Linearführungen angebracht und vertikal am Menschen entlang verfahren. Das Messprinzip ist dem *VITUS Pro* Scanner sehr ähnlich. Im Gegensatz zum *VITUS Pro* Scanner wird beim *Cyberware WB4* Scanner pro Lichtschnittsensor nur eine Kamera benötigt. Durch Spiegelsysteme vor den Kameras wird erreicht, dass die obere Bildhälfte einer Kamera als virtuelle obere Kamera und die untere Bildhälfte als virtuelle untere Kamera eines Doppeltriangulationssensors verwendet werden kann. Da dadurch nur noch die halbe Kameraauflösung für einen Scanpunkt zur Verfügung steht, werden gleiche Punkte in oberer und unterer Bildhälfte zu Punkten doppelter Genauigkeit kombiniert. Die Kameras laufen entsprechend der EIA-Norm mit 30 Hertz, sodass der *Cyberware WB4* Scanner etwas schneller messen kann als der *VITUS Pro* Scanner. Allerdings werden in einer Messung nur etwa 150.000 Scanpunkte aufgenommen. Die Ergebnisse

beider Scannersysteme sind hinsichtlich ihrer Vollständigkeit und Qualität dennoch vergleichbar, zumal auch der *Cyberware WB4* Scanner Farbtextur erfassen kann.

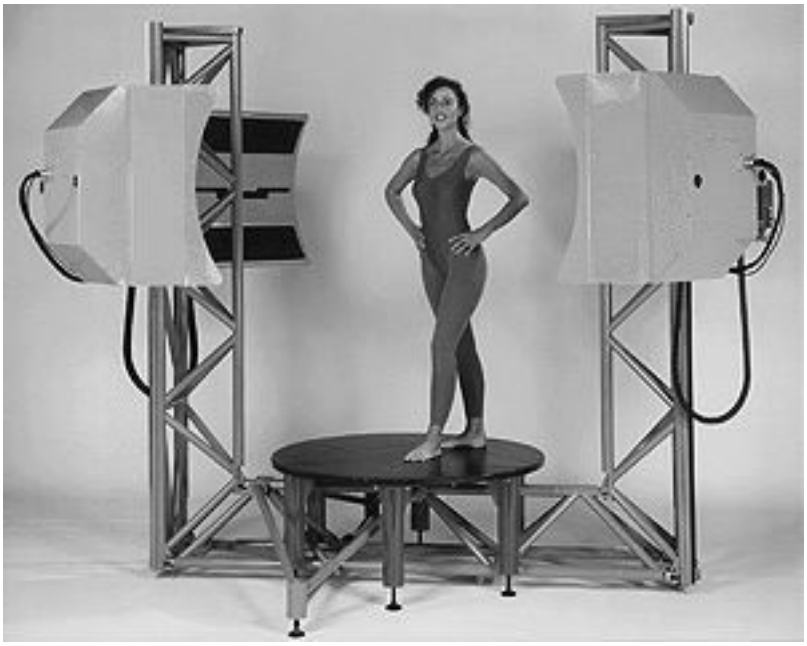


Abbildung 1.16: Cyberware WB4

Einen völlig anderen Weg im Bereich der Menschvermessung geht das *Tecmath Contour*-System [Contour]. Ein Person wird, vor einem beleuchteten Hintergrund stehend, in mehreren Körperhaltungen von einer Kamera aufgenommen.

Abbildung 1.17: Tecmath Contour

Durch den hohen Kontrast zwischen Person und Hintergrund kann in den Kamerabildern ein sogenannter Schattenriss extrahiert werden, der den Umriss der Person beschreibt (Abbildung 1.17). Über die direkte Analyse einzelner Schattenrisse können Grundkörpermaße ermittelt werden. Die Kombination von Daten aus mehreren Schattenrissen einer Person in unterschiedlichen Körperhaltungen führt zu komplexen Körpermaßen, sowie zur Steigerung der Genauigkeit und zur Validierung der Grundkörpermaße. Vor allem aufgrund der kostengünstigen Realisierung und hohen Zuverlässigkeit wird das *Tecmath Contour*-System im Bereich der Bekleidungsindustrie zur Erstellung von maßkonfektionierten Herrenanzügen eingesetzt. Das Messsystem wird dazu im Bekleidungsgeschäft installiert. Der Kunde kann sich innerhalb weniger Sekunden vermessen lassen und äußert anschließend seine individuellen Wünsche bezüglich Schnitt, Applikationen und Stoffs seines Anzuges. Diese Daten werden per EDV an den Hersteller versandt, bei dem der Wunschanzug dann innerhalb weniger Tage produziert wird. Ein weiteres Einsatzgebiet ist die Einkleidung von Rekruten bei der Armee. Durch die mögliche Reihenvermessung vieler Rekruten und die Verwaltung der Körpermaße in Verbindung mit Datenbanksystemen kann so die Lagerhaltung von Uniformen deutlich optimiert und effizienter gestaltet werden.

Ein weiteres Anwendungsgebiet für 3D-Scanner Systeme sind Tischmodelle zur einfachen Erfassung von kleinen Objekten. Ein klassischer Vertreter ist der *Cyberware Desktop Scanner Modell 15*. Das Objekt wird auf einem Drehteller platziert und von einem Lichtschnittsensor innerhalb weniger Sekunden vermessen. Der Drehteller ermöglicht die Rundumvermessung von Objekten, der verwendete 3D-Sensor erfasst neben der Geometrie noch die Farbtextur des Objektes. Die gesamte Objekterfassung wird vom Rechner gesteuert und verläuft somit automatisch. Der Drehteller garantiert zwar, dass das Objekt von allen Seiten vermessen werden kann. Eine vollständige Erfassung ist aber nur bei einfachen Objekten möglich, bei denen keine Abschattungen auftreten. Eine Nachbearbeitung der Daten durch ein geeignetes Programm kann daher oft nicht vermieden werden. Zudem fällt der hohe Preis des Scanners von etwa 27.900$ auf.

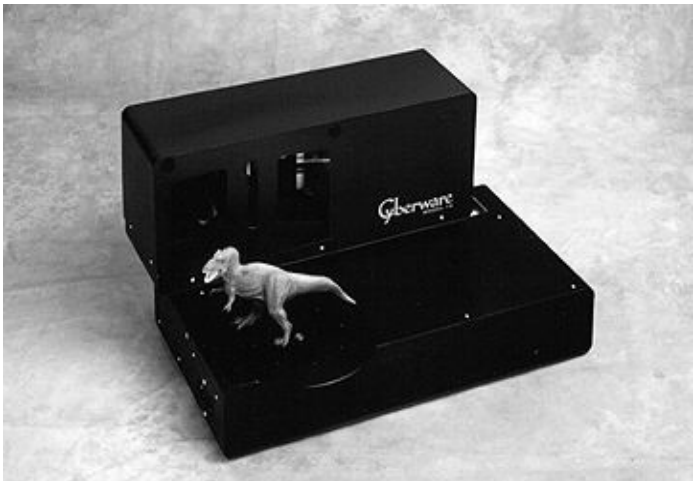


Abbildung 1.18: Cyberware Desktop Scanner Modell 15

Günstiger und einfacher realisiert ist das *3D-Dimension Scanbook*. Auch hier wird das Objekt auf einem Drehteller platziert. Das verwendete Messprinzip ist aber das Volumenschnittverfahren. Die Position der Aufnahmekamera relativ zum Drehteller wird zu Beginn einer Messung mit Hilfe eines Kalibrierkörpers bestimmt. Durch den einfarbigen Hintergrund des

Scanbooks und die zusätzliche diffuse Beleuchtung kann das Objekt in den Kamerabildern vom Hintergrund separiert werden. Der Drehteller erlaubt die Einstellung von fast beliebig vielen Ansichten, so dass nur der eingesetzte Voxelraum die Auflösung des Modells begrenzt. Mit der hochauflösenden Digitalkamera werden zusätzlich hochwertige Farbtexturen erzeugt, so dass fast fotorealistische 3D-Bilder erzeugt werden können. Da die Qualität der 3D-Daten aufgrund des verwendeten Volumenschnittverfahrens sehr begrenzt wird, finden sich die meisten Anwendungsgebiete des *Scanbooks* im Multimediabereich. Eine typische Applikation ist das Internetshopping, Produkte sind als dreidimensionale Bilder aus dem Internet abrufbar und damit für den Kunden sehr viel plastischer darstellbar als einfache Fotografien.

Abbildung 1.19: 3D-Dimension Scanbook

Das direkte Gegenteil des *Cyberware Desktop Scanners* und des *Scanbooks* sind Systeme zur Vermessung von besonders großen Objekten, wie zum Beispiel Gebäuden.

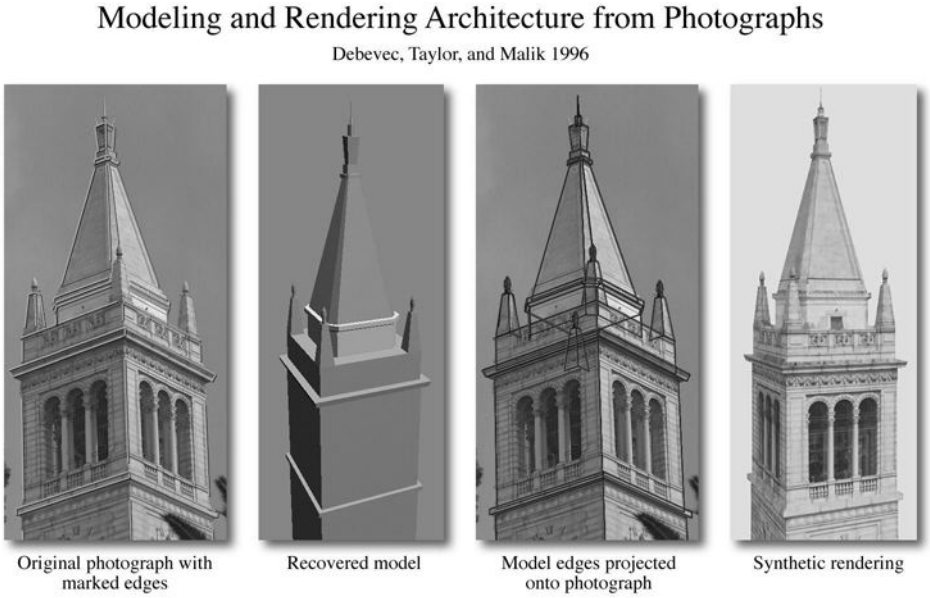


Abbildung 1.20: Objektrekonstruktion mit Facade

Mit der Software *Facade* kann aus einfachen Fotografien eines Gebäudes interaktiv eine 3D-Rekonstruktion berechnet werden. In jedes Bild des zu rekonstruierenden Gebäudes werden vom Benutzer geometrische Körper, wie zum Beispiel Quader, Prismen oder Pyramiden eingepasst. Aber auch komplexere Körper, wie zum Beispiel Torbögen, Säulen oder Erker sind möglich (Abbildung 1.20). Durch die Zuordnung von gleichen Körpern in unterschiedlichen Bildern, werden die einzelnen Ansichten miteinander verbunden und die Rekonstruktion vervollständigt. Da die Modellierung jedes einzelnen Details des Gebäudes zuviel Aufwand bedeutet, werden sehr starke Vereinfachungen bei der Rekonstruktion getroffen. Ein Haus besteht im einfachsten Fall aus einem Quader als Basis und einem Prisma als Dach. Da die Bilder des Hauses aber zusätzlich zur Texturierung der Rekonstruktion verwendet werden können, enstehen dennoch fotorealistische Darstellungen.

Der *Riegl* 3D-Scanner dient hauptsächlich zur Erfassung von Industrieanlagen, Gebäuden oder Geländen. Ein Lichtlaufzeit messender Entfernungssensor wird über Spiegel abgelenkt und in Zeilen und Spalten über das Objekt geschwenkt. Die Reichweite des Sensors beträgt bis zu siebenhundert Metern, bei einer Auflösung von 2,5 Zentimetern. Innerhalb weniger Minuten wird so ein Panoramascan mit einer Auflösung von mehreren tausend Zeilen und Spalten erstellt. Die Daten werden dann interaktiv im Rechner weiterverarbeitet. Der Anwender rekonstruiert beispielsweise eine erfasste Industrieanlage, indem er Bauteile, wie zum Beispiel Rohrleitungen, vom Computer unterstützt in die Scandaten einpasst. Neben der grundsätzlichen Generierung von Computermodellen bestehender Industrieanlagen, können auch Änderungen, die nachträglich in einer Anlage durchgeführt wurden leichter in ein Computermodell übertragen werden.

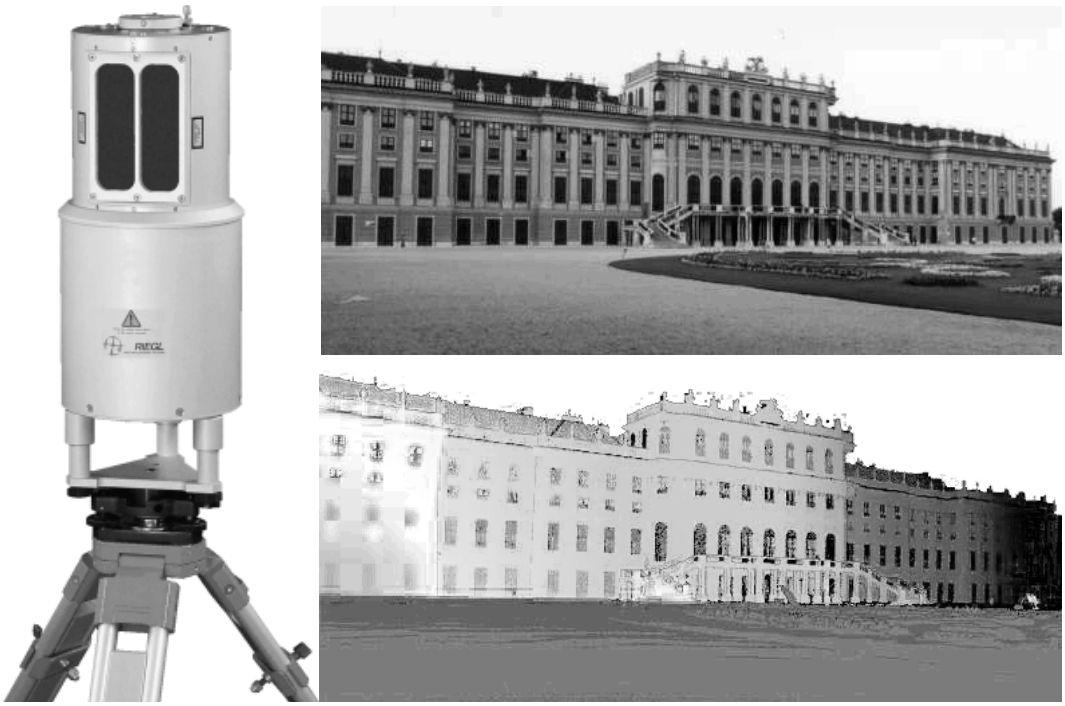


Abbildung 1.21: Riegl Scanner (l.) mit Gebäude (r.o.) und erzeugtem Tiefenbild (r.u.)

Im Bereich der Computer unterstützten Produktion (CAM) und der Computer unterstützten Qualitätskontrolle (CAQ) werden von 3D-Scanner Systemen Genauigkeiten und Flexibilitäten gefordert, wie sie von keinem der bisher vorgestellten Systeme geleistet werden können. Bei

der Qualitätskontrolle im Rahmen der Fertigung von Automobilen müssen Beulen in der Karosserie von Fahrzeugen gefunden werden. Dazu muss eine Fläche von vielen Quadratmetern auf Fehlstellen mit einer Tiefen von etwa 50 Mikrometern durchsucht werden. Für das sogenannte Reverse Engineering müssen Design Modelle von Fahrzeugen vollständig und lükkenlos mit einer ähnlich hohen Auflösung erfasst werden. Die beiden Hauptanforderungen bestehen also zum einen in einer hohen Auflösung und Genauigkeit und zum anderen in der flexiblen Rundumerfassung von Objekten.

Vor allem auf die zweite Anforderung zugeschnitten ist der *ModelMaker* der Firma 3D-Scanners. Mit einem Lichtschnittsensor, der auf einem Positionsarm montiert ist, können vor allem mittelgroße und kleine Objekte mit einer Größe von einigen Kubikdezimetern erfasst werden. Der Lichtschnittsensor wird per Hand geführt und der Laser des Sensors über das Objekt geschwenkt. Über die Position des Arms können die gemessenen Entfernungswerte des Lichtschnittsensors in ein gemeinsames globales Koordinatensystem umgerechnet werden.

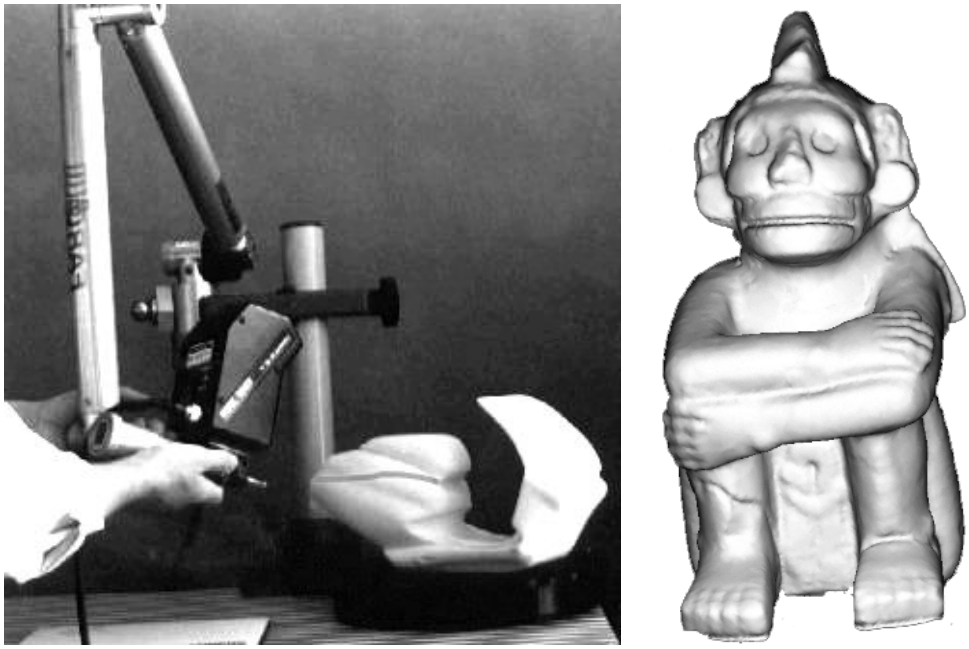


Abbildung 1.22: 3D-Scanners ModelMaker

Die manuelle Führung des Sensorkopfes garantiert die einfache Handhabung des Sensors, die auch einem Laien die Vermessung komplexer Objekte ermöglicht. Die Kopplung des Sensorkopfes an einen Positionsarm schränkt aber den Arbeitsraum des Sensors auf die Reichweite des Arms ein. Große Objekte können daher nur teilweise erfasst werden.

Mit dem *ATOS* Sensor in Kombination mit dem *TRITOP* System der Firma GOMmbh können nahezu beliebig große Objekte in hoher Auflösung vermessen werden. Der *ATOS* Sensor ist ein Sensor, der nach dem kodierten Lichtansatz arbeitet. Über zwei Kameras wird stereoskopisch gemessen, so dass der Sensor unabhängig von möglichen Temperaturdriftfehlern des Projektors ist. Kalibriert wird der Sensor über Eichkörper, mit denen über eine Bündelausgleichsrechnung die internen und externen Parameter der Kameras bestimmt werden. Vom Hersteller wird eine Genauigkeit von 50 Mikrometern angegeben, bei einer Diagonalen durch das Messvolumen von etwa 40 Zentimetern [Bergmann97],[Winter97]. Um große Objekte, wie zum

Beispiel ganze Fahrzeuge zu vermessen, ist eine vorherige Präparation des Objektes mit retroreflektierenden Marken notwendig. Mit dem hochauflösenden photogrammetrischen Messsystem *TRITOP* werden die globalen Koordinaten der Marken bestimmt. Sofern in einer Messung, die vom freipositionierbaren *ATOS* Sensor aufgenommen wurde, wenigstens drei dieser Marken wiedergefunden werden, dann kann eine Transformation der Messung aus dem Sensorkoordinatensystem in das globale Koordinatensystem berechnet werden. Die Vermessung eines ganzen Fahrzeugs kann so innerhalb von acht Stunden mit einer Genauigkeit von etwa 0,1mm durchgeführt werden. Die dafür erforderlichen 150 Einzelmessungen werden bereits während des Messvorgangs visualisiert, so dass der Anwender korrigierend in die Messung eingreifen kann um gegebenenfalls zusätzliche Aufnahmen durchzuführen.

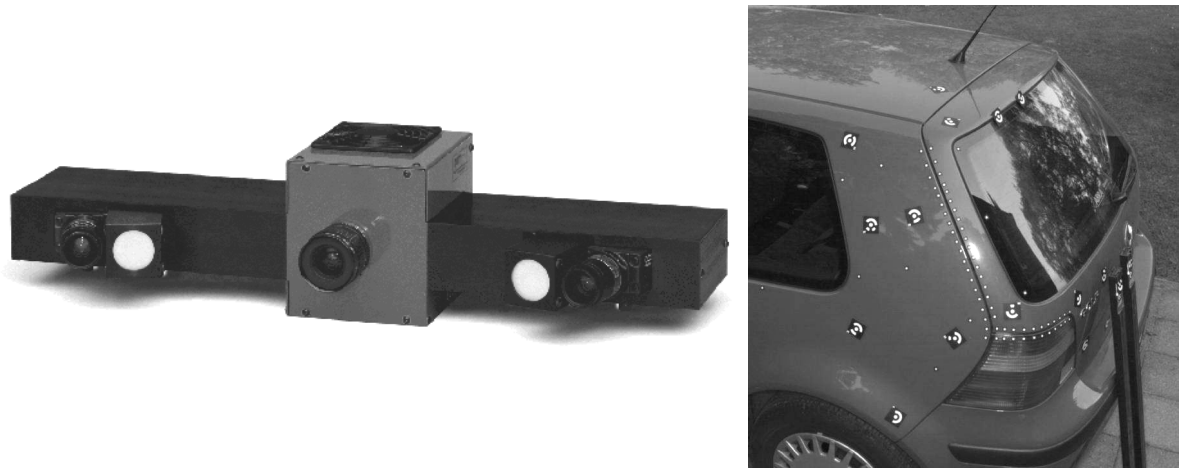


Abbildung 1.23: ATOS Sensor

Wenn das Aufbringen von Marken auf dem Objekt nicht möglich ist, kann für Objekte mit einer Größe bis zu einem halben Meter ein Drehteller eingesetzt werden, bei dem alternativ das Objekt oder ein System von Referenzmarken vermessen werden kann. Die genaue Drehmechanik garantiert dabei, dass über die Referenzmarken die genaue Position des Sensors relativ zum Objekt bestimmt werden kann, indem alternativ die Marken oder das Objekt in das Sichtfeld des Sensors geschwenkt wird. An Grenzen stößt das System aber, wenn auch große Objekte ohne Marken vollständig vermessen werden sollen. Zusätzlich tauchen die Marken als Fehlstellen in den Messdaten auf, so dass bestenfalls unwichtige Oberflächenbereiche des Objektes mit Marken versehen werden dürfen.

Das Softwaresystem *SLIM* ermöglicht grundätzlich die Zusammenführung und Registrierung von unabhängig aufgenommenen Einzelansichten eines Objektes. Das Aufkleben von Referenzmarken auf das Objekt kann vollständig entfallen. Einzelansichten, die beispielsweise mit einem ABW Projektor aufgenommen wurden, werden in die Software eingelesen. Der Benutzer muss nun wenigstens drei gleiche Punkte in jeweils zwei einander überlappenden Scans bestimmen, damit die Software eine Groborientierung der Datensätze bestimmen kann. Danach wird automatisch mit einem ICP[1] Verfahren eine Feinregistrierung der Einzelansichten vorgenommen, in dem die Abstände in den Überlappungsbereichen der Einzelansichten minimiert werden (Abbildung 1.24). In einem Nachbearbeitungsschritt werden die Dreiecksnetze

1. Iterative Closest Point

der Einzelansichten zu einem Netz verschmolzen. Bestehende Restungenauigkeiten, die sich als Rauschen im resultierenden Dreiecksnetz bemerkbar machen, werden durch einen problemangepassten Filter geglättet [Karbacher97]. Das Ergebnis ist ein „wasserdichtes“ Dreiecksnetz, das zu CAD-Flächen weiterverarbeitet werden kann, oder mit dem direkt Rapid Prototyping Maschinen programmiert werden können.

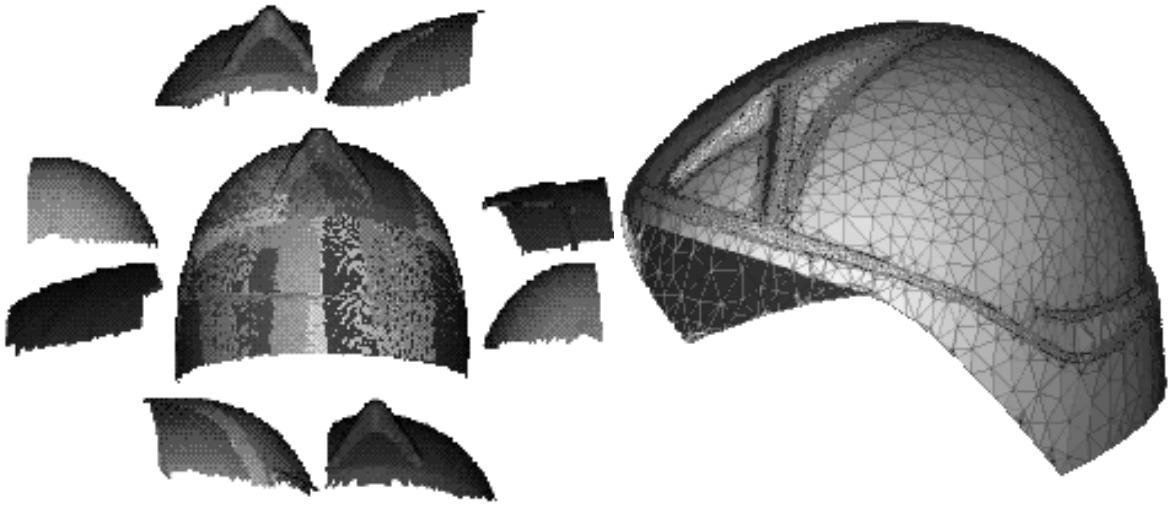


Abbildung 1.24: Softwaresystem SLIM

Voraussetzungen für das Funktionieren der Registrierung sind große Überlappungsbereiche zwischen den Einzelansichten. Zusätzlich müssen ausreichend charakteristische Merkmale, wie zum Beispiel Auswölbungen oder Kanten, in den Datensätzen vorhanden sein. Während der automatischen Registrierung sorgen vor allem diese markanten Bereiche für eine gute Konvergenz des Verfahrens und ein eindeutiges Ergebnis. Sind keine derartigen Merkmale vorhanden, dann scheitert das Verfahren. So können viele Einzelansichten einer Kugel nicht korrekt registriert werden, da es beliebig viele Lösungen für die relative Lage der Einzelansichten zueinander gibt. Im Weiteren hängt die Komplexität des Verfahrens von der Zahl der Einzelansichten ab, die zueinander in Deckung gebracht werden sollen. Bei Objekten, die über zu viele Einzelansichten erfasst wurden, wird einerseits das Datenaufkommen zu hoch, andererseits leidet die Konvergenz der automatischen Registrierung, so dass Rechenzeiten von mehreren Tagen in Kauf genommen werden müssen, bis eine zufriedenstellende Qualität erreicht wird.

2 Kalibrierung von 3D-Sensoren

Während die Auflösung eines 3D-Sensors direkt über den Aufbau und die Eigenschaften der eingesetzten Sensorkomponenten bestimmt werden kann, wird eine hohe Genauigkeit erst durch eine Kalibrierung des Sensors erreicht. Das wesentliche Ziel der Kalibrierung ist dabei, die Genauigkeit möglichst nah an die erreichbare Auflösung zu bringen. Ein Sensor kann als optimal kalibriert angesehen werden, wenn die Genauigkeit mit der doppelten Auflösung übereinstimmt.

Bei optischen Systemen werden im Allgemeinen elektronische CCD-Kameras eingesetzt. Wichtige Voraussetzung für das Verständnis zur Kalibrierung von 3D-Sensoren ist daher das Wissen über die Funktionsweise einer CCD-Kamera und die Abbildungseigenschaften der verwendeten Kameraoptiken, sowie die Bestimmung der Abbildungsparameter. Die Kalibrierung von 3D-Sensoren kann grundsätzlich in modellfreier Kalibrierung oder modellbasierter Kalibrierung erfolgen. Beide Verfahren werden am Beispiel optischer Triangulationssensoren erläutert. Im Weiteren wird das Prinzip der Stereokalibrierung vorgestellt, das die wesentliche Grundlage der vorliegenden Arbeit darstellt.

2.1 Kamerakalibrierung

Elektronische Kameras sind die wesentliche Voraussetzung optischer 3D-Sensoren. Die Bestimmung der Abbildungsparameter einer Kamera ist zwar nicht für jedes Kalibrierverfahren notwendig, das Wissen über die relevanten Parameter der Abbildungsfunktion einer Kamera ist aber wichtig für die Entwicklung von Kalibrierverfahren und vor allem für die Entscheidung, welche Ausprägung der möglichen Kalibrierverfahren für die gegebene Aufgabenstellung am sinnvollsten ist.

Für die mathematische Repräsentation im Rechner bildet die Kamera mit dem Objektiv eine Einheit. Zunächst werden daher die wesentlichen technischen Verfahren zur Bildaufnahme mit elektronischen Kameras beschrieben. Anschließend werden die Eigenschaften von Kameraobjektiven erläutert, die einen bestimmenden Einfluss auf die Abbildung von realen Szenen auf die Bildebene der Kamera haben.

2.1.1 Elektronische Kameras

Mit der Entwicklung von integrierten Schaltungen und der damit möglichen Miniaturisierung haben sich zwei Typen von Kameras durchgesetzt: die CCD-Kameras[1] und die CMOS[2]-Kameras. Das Bildfeld dieser Kameras besteht aus vielen einzelnen Bildelementen die in einem

1. Charge Coupled Device
2. Complementary Metal Oxide Semiconductor

Raster angeordnet sind und separat ausgewertet werden. Die Größe solcher Bildfelder liegt bei üblichen Kameras bei etwa 6 bis 40 Millimetern im Quadrat, auf denen zwischen hunderttausend und mehreren Millionen Bildelementen integriert sein können.

2.1.1.1 CCD-Kameras

Ein Pixel einer CCD-Kamera wandelt Licht im sichtbaren Wellenlängenbereich in elektrische Ladung um. Die nach einer definierten Integrationszeit gesammelte Ladung wird durch die Kameraelektronik ausgelesen und entweder in ein analoges Videosignal gewandelt und zum Computer übertragen oder direkt in der Kamera digitalisiert. Die Größe der Ladung ist proportional zur Menge des eingefallenen Lichtes und entspricht damit der Helligkeit des Lichtes am Ort des Pixels. Die Integrationszeit wird üblicherweise auf 1/50s eingestellt[1]. Um schnelle Bewegungen scharf abzubilden, wird ein elektronischer Shutter verwendet, der die Verkürzung der Integrationszeit bei Standardkameras auf 1/10.000s zulässt. Allerdings verringert sich dabei auch die Bildhelligkeit entsprechend.

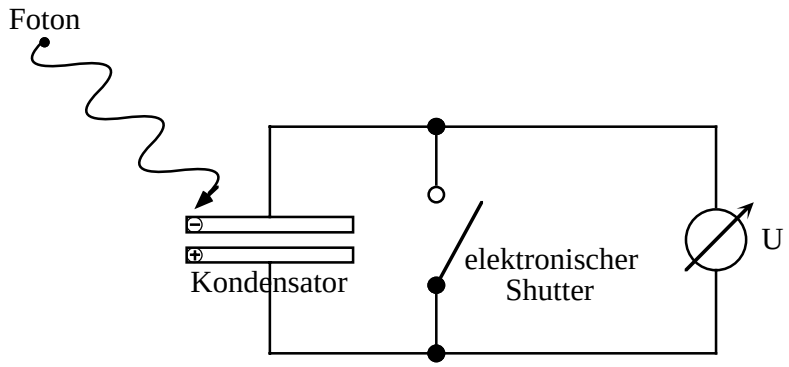


Abbildung 2.1: CCD-Sensor

Zu helles Licht oder zu lange Belichtung führt zum unkontrollierten Abfluß von sogenannten Störladungen. Der Kondensator wird überladen und die Ladung eines Pixel fließt zum Teil in die Nachbarpixel ab. Die Bilder weisen helle Streifen, ausgehend von der überbelichteten Stelle auf. Besonders bei Szenen mit starken Kontrasten oder lokaler Überstrahlung wird daher die Wahl der richtigen Belichtungszeit schwierig. Der Vorteil von CCD-Kameras liegt vor allem in der hohen Linearität, der hohen Empfindlichkeit, sowie dem großen Widerstand gegen mechanische und magnetische Störungen.

2.1.1.2 CMOS-Kameras

CMOS-Sensoren basieren auf fotoempfindlichen, jeweils mit einem Widerstand in Serie geschalteten Dioden. Im Gegensatz zu CCD-Sensoren kann dadurch eine kontinuierliche Wandlung des Fotonenstroms in eine Spannung erreicht werden. Der logarithmische Zusammenhang zwischen Lichtintensität und Spannung ermöglicht zudem die Aufnahme von sehr kontrastreichen Szenen. Im Weiteren kann jedes Pixel zu jeder Zeit ausgelesen werden, daher wird bei CMOS-Kameras oft ein wahlfreier Zugriff auf die Pixel realisiert.

1. europäische Norm, 1/60s bei amerikanischer Norm

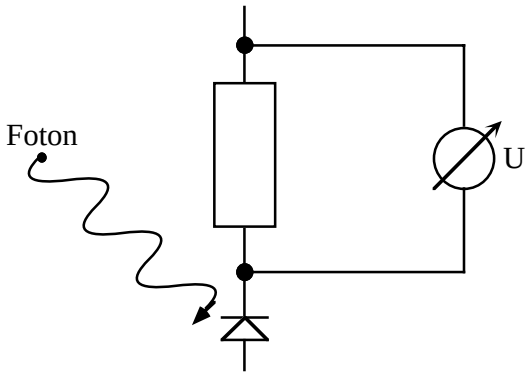


Abbildung 2.2: CMOS-Sensor

Die Realisierung eines CMOS-Sensors entsprechend Abbildung 2.2 würde allerdings in einer hervorragenden Rauschquelle resultieren. Daher wird zu jedem Pixel Funktionalität integriert, die unter anderm der Rauschminderung dienen kann. Die dafür verbrauchte Fläche geht allerdings der fotosensitiven Fläche verloren. Die Empfindlichkeit von CMOS-Sensoren ist daher üblicherweise eine Dekade geringer als die der CCD-Sensoren.

2.1.1.3 Bildübertragung in den Computer

Der Transfer der Bilddaten von der Kamera in den Computer geschieht in den meisten Fällen über Framegrabber, Einsteckkarten die im Computer die Digitalisierung und Bildwandlung der analogen Videosignale der Kameras übernehmen. Um eine möglichst einfache und flexible Ankopplung des Framegrabbers an die unterschiedlichsten Kameratypen zu gewährleisten, werden aus der Fernsehtechnik bekannte, standardisierte Videosignale verwendet. Die europäische CCIR[1]-Norm erlaubt die Übertragung von 25 Bildern mit 575 Zeilen pro Sekunde. Pro Zeile werden üblicherweise 767 Pixel mit einer Frequenz von 14,75 MHz übertragen. Die amerikanische EIA[2]-Norm überträgt 30 Bilder mit 485 Zeilen und üblicherweise 647 Pixeln pro Zeile in einer Sekunde. Die Pixelfrequenz beträgt dann 12,3 MHz. Zusätzlich sind beide Videonormen noch mit Signalen, wie Zeilenrücklauf und Bildrücklauf ausgestattet, die die Synchronisation des Framegrabbers mit der Kamera ermöglichen.

Kameras, die nicht diesen Standards entsprechen, da sie hohe Auflösungen oder schnelle Bildwiederholraten bieten, benötigen normalerweise spezielle Framegrabber. Da derartige Lösungen schnell sehr teuer werden, haben sich mehrere digitale Übertragungsformate durchgesetzt. Die wichtigsten sind die IEEE1394[3]-Standard und der CameraLink. Aus dem Multimediabereich ist zudem der USB-Standard bekannt, der aufgrund seiner geringen Übertragungsrate allerdings nur eingeschränkt für die Bildübertragung geeignet ist. Genau wie der USB-Standard ist auch der IEEE1394-Standard als Bussystem ausgelegt. Das heißt, dass beim IEEE1394 System bis zu 63 Geräte an einer Interfacekarte gleichzeitig betrieben werden können. Die Bildübertragung wird allerdings durch die Übertragungsrate von etwa 400 MBit/s (32 MByte/s effektiv) begrenzt, die unter allen Kameras aufgeteilt werden muss. Eine hochauflösende Echtzeitbildverarbeitung ist daher nur mit maximal zwei Kameras möglich. Neuere Spe-

1. Comité Consultativ International des Radiocommunications
2. Electronic Industries Association
3. auch FireWire oder ILink

zifikationen sehen allerdings Übertragungsraten von bis zu 1.6 GBit/s vor. Der CameraLink lässt nur den Anschluss einer Kamera zu, erreicht allerdings Übertragungsraten bis zu 2.38 GBit/s. Er ist daher besonders für hochauflösende Kameras geeignet.
Der wesentliche Vorteil der digitalen Bildübertragung ist die Unabhängigkeit der Interfacekarte im Rechner von den Bildformaten der Kameras. An Bussystemen können gleichzeitig Kameras unterschiedlicher Formate betrieben werden. Zudem kann die Elektronik zur Digitalisierung der Bilddaten in der Kamera genau auf die Eigenschaften der Kamera abgestimmt werden. Aliasingeffekte, die bei Framegrabberlösungen aufgrund schlechter Synchronisation entstehen können, treten bei digitaler Übertragung nicht mehr auf. Allerdings sind digitale Kameras aufgrund der notwendigen Elektronik im Allgemeinen teurer als Kameras mit analoger Datenübertragung. Ein Nachteil, der nur selten über den geringeren Preis der Interfacekarte gegenüber dem Framegrabber ausgeglichen werden kann.

2.1.2 Abbildungseigenschaften von Kameraobjektiven

Die Aufnahme von Bildern mit einer Kamera kann nur durch die Verwendung von Objektiven erfolgen, die für eine scharfe Abbildung der Szene auf die Bildfläche der Kamera sorgen. Ideale Objektive bilden eine Lochblende mit unendlich kleinem Loch nach und bilden jeden Punkt der Szene spiegelverkehrt auf der Bildfläche scharf ab. Reale Objektive bestehen allerdings aus mehreren hintereinander angeordneten Linsen und können den idealen Zustand daher nur annähern, so dass zum einen nur ein begrenzter Bereich vor der Kamera scharf abgebildet wird und zum anderen Verzeichnungsfehler entstehen. Einige wesentliche Verzeichnungsfehler nach [Schröder80] sind

- *Sphärische Aberration*: Unschärfe im Bild, da die durch die Randzone einer sphärischen Linse gehenden Strahlen stärker gebrochen werden als die Strahlen in der Mitte.
- *Koma*: Punkte fern der optischen Achse erscheinen in radialer Richtung verwischt; Punkte in der Szene werden bei der Abbildung kometenförmig verzerrt.
- *Astigmatismus*: Entsteht, wenn die Linse im waagerechten und senkrechten Durchmesser unterschiedlich stark gekrümmt ist. Ein in den waagerechten Durchmesser fallendes Lichtband ergibt eine andere Brennweite als senkrechte Bündel, z.B. bei schräg einfallendem Licht.
- *Bildfeldkrümmungen*: Die Ränder der Bilder werden unscharf, weil ein scharfes Bild auf einer gewölbten Fläche statt auf einer Ebene erzeugt wird.
- *Verzerrung*: Die Abbildung eines rechtwinkligen Gitternetzes ist kissen- oder tonnenförmig, d.h., vor allem an den Rändern werden die Geraden nach innen oder außen gekrümmt.

Für die Messtechnik spielt vor allem die Verzerrung eine große Rolle, da besonders bei Objektiven mit kurzen Brennweiten Bildfehler von mehreren Pixeln entstehen können. Die anderen Verzeichnungsfehler sind dagegen oft vernachlässigbar klein oder haben keine Auswirkung auf die Genauigkeit der Bildauswertung. So sorgt die sphärische Aberration zwar dafür, dass Kontraste verschwimmen; Schwerpunktberechnungen zur genauen Bestimmung von Bildpunkten werden aber nicht beeinflusst.

Die mathematische Beschreibung von Kameramodellen muss den jeweils für die Anwendung relevanten Verzeichnisfehlern Rechnung tragen. Die im Rahmen dieser Arbeit verwendeten

Kameramodelle, sowie die Verfahren zur Berechnung der Parameter dieser Modelle werden im Folgenden vorgestellt.

2.1.3 Kameramodelle

Das einfachste Modell zur Beschreibung von Abbildungseigenschaften von Kameras ist das Lochkameramodell. Es berücksichtigt keine Verzerrungen, sondern beschreibt lediglich die Abbildung von Objektpunkten über eine Zentralprojektion auf die virtuelle Bildebene der Kamera. Wenn von einem quadratischen Bildfeld ausgegangen wird, für das o.B.d.A die Abmessung 1 angenommen wird, kann das gesamte Modell hinreichend durch die Brennweite (f), sowie die externe Lageinformation, bestehend aus einer Translation ($\vec{T}$) und einer Rotation (R), in Form dreier Rotationswinkel (Θ,Ψ,Φ), beschrieben werden. Zur besseren Repräsentation realer Kameras wird das Modell zusätzlich durch die Parameter der horizontalen und vertikalen Bildausdehnung erweitert (N_X, N_Y). Die optische Achse läuft in jedem Fall durch den Bildmittelpunkt, der Brennpunkt bildet den Ursprung des Kamerakoordinatensystems.

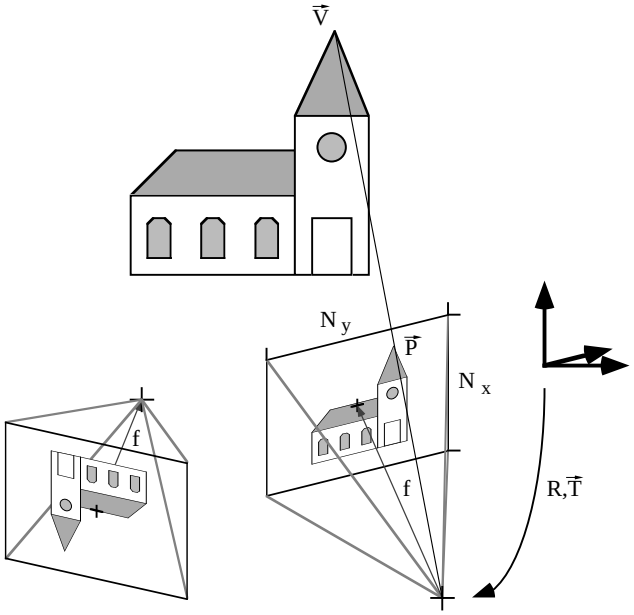


Abbildung 2.3: Lochkamera- und Mattscheibenmodell

Wie bei realen Kameras bildet das Lochkameramodell die Szene spiegelverkehrt auf die Bildebene ab. Die mathematisch äquivalente Anordnung der Bildebene vor dem Brennpunkt und der daraus folgenden seitenrichtigen Abbildung der Szene auf die Bildebene führt zum Mattscheibenmodell, das zum überwiegenden Teil aufgrund der einfacheren Handhabung in Algorithmen Anwendung findet.

Zur Berechnung der Bildkoordinate $\vec{P}$ aus einem Raumpunkt $\vec{V}$, wird der Raumpunkt zunächst in das Kamerakoordinatensystem transformiert:

$$\vec{V}' = (\vec{V} - \vec{T}) \bullet R^{-1}, \tag{2.1}$$

die Berechnung der Bildkoordinate geschieht dann mit

$$\begin{pmatrix} P_X \\ P_Y \end{pmatrix} = \begin{pmatrix} N_X \bullet V'_X / V'_Z \\ N_Y \bullet V'_Y / V'_Z \end{pmatrix} \bullet f + \begin{pmatrix} N_X / 2 \\ N_Y / 2 \end{pmatrix}. \tag{2.2}$$

Das Modell besitzt sieben Unbekannte, den Translationsvektor mit drei Unbekannten, drei Rotationswinkel zur Beschreibung der Rotation und die Brennweite. Für die Berechnung dieser Parameter im Rahmen einer Kalibrierung muss daher die Abbildung für wenigstens sieben Raumpunkte bekannt sein.

Zur Beschreibung realer Systeme ist dieses Modell meistens unzureichend, da die beschriebenen Verzeichnungsfehler einen beträchtlichen Bildfehler verursachen können. Das Mattscheibenmodell wird daher um zusätzliche Parameter erweitert, mit denen die Auswirkungen der Verzeichnungsfehler kompensiert werden können. Ein Satz von Parametern für jede Klasse von Verzeichnungsfehlern ist allerdings nicht sinnvoll, da es zu Mehrdeutigkeiten zwischen den verschiedenen Parametern kommen kann, die sich mathematisch nur schwer bestimmen lassen. Zudem werden die Parameter oft in Bezug auf ihre mathematische Berechenbarkeit und nicht in Bezug auf eine reale Entsprechung gewählt.

Ein weit verbreitetes, erweitertes Modell wurde in [Tsai87] eingeführt. Zusätzlich zu den sieben beschriebenen Parametern werden weitere fünf Parameter verwendet. Die kissen- oder tonnenförmige Verzerrung wird durch einen Parameter der radialen Linsenverzerrung erster Ordnung (κ) beschrieben. Die optische Achse der Kamera wird in den seltensten Fällen durch die Bildmitte verlaufen. Die ungenaue Montage des Bildchips im Kameragehäuse oder Ungenauigkeiten im Objektiv führen meistens zu einer Verschiebung der Bildmitte. In [Tsai87] wird die Bildmitte als Zentrum der radialen Linsenverzerrung beschrieben und geht mit zwei weiteren Parametern (C_X, C_Y) in das Kameramodell ein. Diese Definition hat zwar keine direkte reale Entsprechung, unterstützt aber die Parameterbestimmung im Rahmen einer Kamerakalibrierung. Da der Einzug der Bilddaten in den Computer meistens mit einem Framegrabber geschieht, wird der Skalierungsfaktor s eingeführt. Er soll die Streckung oder Stauchung der Bilddaten entlang der Bildzeilen beschreiben, die entstehen kann, wenn ein unzureichend synchronisierter Framegrabber die Zeilen zu schnell oder zu langsam digitalisiert. Ein Wert von eins entspricht einer korrekten Abtastung, ein größerer Wert entspricht einer zu schnellen Digitalisierung, die das Bild zu breit erscheinen lässt. Alternativ kann dieser Wert auch zur Kompensierung nicht quadratischer Pixel dienen. Der Skalierungsfaktor entspricht dann dem Verhältnis zwischen Breite und Höhe eines Pixels. Aus der technischen Dokumentation der Kamera können in vielen Fällen die Abmessungen eines Pixels oder die Größe des CCD-Chips entnommen werden. Diese Größen sind kein wesentlicher Bestandteil der Kalibrierung, können aber dazu verwendet werden, die Parameter des Modells mit realen Größen zu vergleichen. Das ist vor allem hilfreich, um die Plausibilität der berechneten Brennweite mit der auf dem Kameraobjektiv angegebenen Größe zu überprüfen.

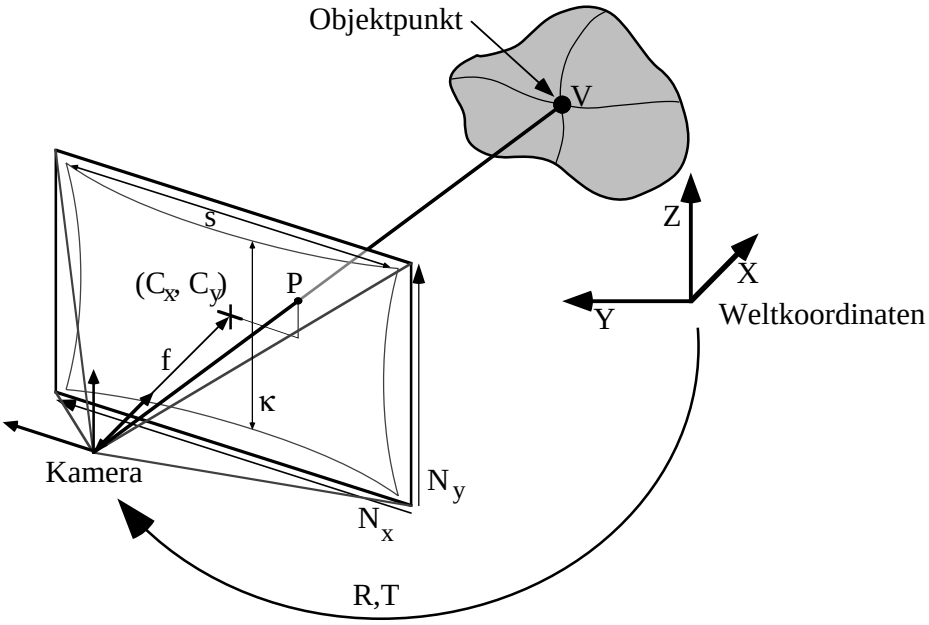


Abbildung 2.4: Kameramodell nach Tsai

Für die Bestimmung der Parameter im Rahmen einer Kalibrierung muss wenigstens die Abbildung für elf Raumpunkte bekannt sein. Das im Rahmen dieser Arbeit realisierte Verfahren zur Bestimmung der elf Parameter[1] basiert auf dem in [Tsai87] vorgeschlagenen Verfahren und ist Inhalt des folgenden Abschnitts.

2.1.4 Kalibrierung nach Tsai

Zur Bestimmung von Kameraparametern im Rahmen einer Kalibrierung werden Kalibrierkörper verwendet, die in ihrem Aussehen und ihrer Form die Bestimmung von punktförmigen Raummarken zulassen. Der Parameterbestimmung zuträglich ist dabei die gleichmäßige Verteilung der Marken über das Bildfeld der Kamera und die Verteilung über eine gewisse Tiefe. Als Faustregel kann dabei gelten, dass die Breiten- und Tiefenausdehnung aller Marken in etwa gleich sein sollte. Im Weiteren sollte die Erkennung der Kalibriermarken im Kamerabild mit hoher Genauigkeit möglich sein. Bei dem hier vorgestellten Verfahren kann eine ungenaue Detektion allerdings auch durch eine hohe Zahl an Marken kompensiert werden, sofern der Fehler, der bei der Markendetektion gemacht wird, stochastisch verteilt ist. Schließlich müssen die Marken im Kamerabild gegenüber Verzeichnisfehlern und perspektivischen Verzerrungen invariant sein. So werden große Marken vor allem im Randbereich eines Bildes stark verzerrt. Eine Lokalisierung einer Marke, die diese Verzerrung nicht berücksichtigt, führt zu einem systematischen Fehler, der eine ungenaue Berechnung der Kameraparameter zur Folge hat.

Bei dem im Rahmen dieser Arbeit realisierten Verfahren werden ringförmige Marken verwendet, die in einem gleichmäßigen Raster auf einer ebenen Kalibrierwand angeordnet sind. Um eine Verteilung der Marken über einen Tiefenbereich zu erreichen, ist die Kalibrierwand auf

1. $\Theta,\Psi,\Phi,t_x,t_y,t_z,f,\kappa_1,s,C_x,C_y$

einer Linearführung angebracht, mit der die genaue Positionierung der Kalibrierwand vor der Kamera möglich ist.

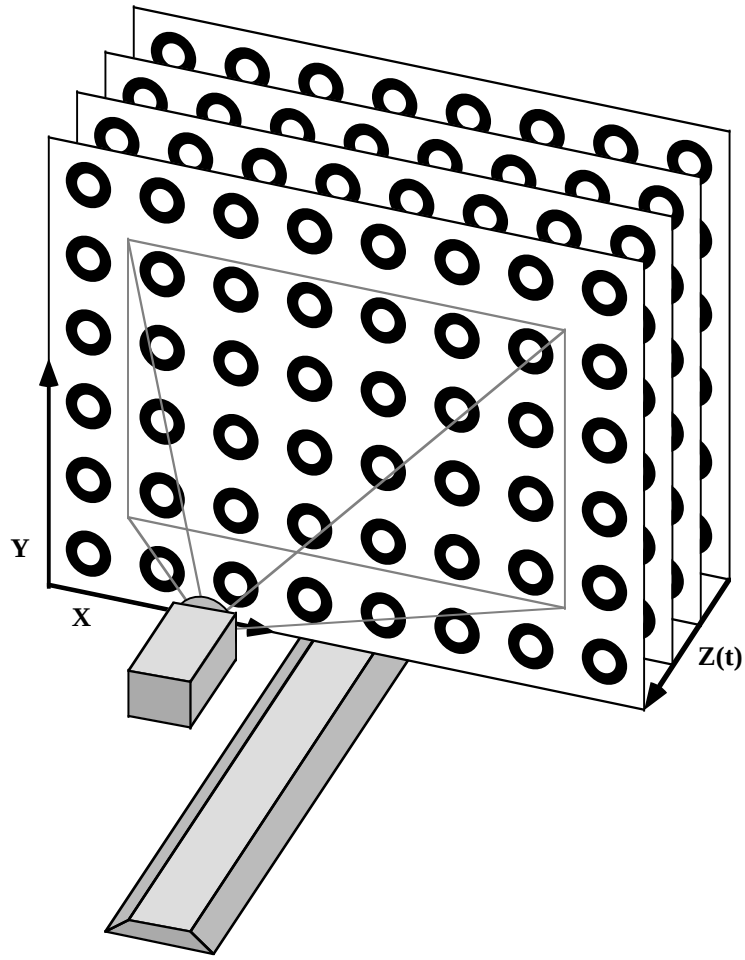


Abbildung 2.5: Kalibriersystem

Zur Kalibrierung wird die Kalibrierwand auf eine Startposition gefahren und die erste Aufnahme der Marken gemacht. Anschließend wird die Wand um eine definierte Strecke auf die Kamera zubewegt und eine weitere Aufnahme gemacht. Das Verfahren wird so lange fortgesetzt, bis eine ausreichende Tiefe mit Kalibriermarken durchsetzt ist. In jedem Kamerabild werden die Ringmarken detektiert. Der schwarze Ring erleichtert dabei das Auffinden der Marken. Die weiße Mitte dient der genauen Bestimmung der Markenmitte. Alle gefundenen Marken werden in jedem Bild horizontal und vertikal abgezählt, so dass jeder Marke eine Raumposition zugeordnet werden kann. Die Marke aus dem ersten Bild, die der Bildmitte am nächsten ist, definiert den Koordinatenursprung und damit die Raumlage aller weiteren Marken. Als Eingabewerte für den Algorithmus [TsaiCode] zur Bestimmung der Kameraparameter dienen dann Tupel, die jeweils aus zwei Bildkoordinaten und drei Raumkoordinaten bestehen.

Die Parameterbestimmung läuft in drei Schritten ab. Zunächst werden aus den Kalibrierdaten die Parameter des Mattscheibenmodells berechnet. Da sich dieses Modell als lineares Gleichungssystem beschreiben lässt, kann die Berechnung der Parameter mit einem Least-Square Verfahren erfolgen. Das Ergebnis wird als Initialschätzung für den zweiten Schritt verwendet,

bei dem eine nichtlineare Optimierung zum Einsatz kommt. Zusätzlich zur Rotation, Translation und Brennweite wird im zweiten Schritt noch die Radialverzerrung bestimmt. Der dritte Schritt optimiert wieder mit einem nichtlinearen, iterativen Verfahren das gesamte Kameramodell. Eine genaue Beschreibung der mathematischen Verfahren würde den Rahmen dieser Arbeit sprengen. Deshalb sei an dieser Stelle auf [Tsai87],[TsaiCode] und [MinPack] verwiesen. Zusätzlich werden im Folgenden beispielhaft einige Ergebnisse dokumentiert, die mit unterschiedlichen Kameras und unterschiedlichen Objektiven durchgeführt wurden.

Die Kamerakalibrierung wurde mit drei unterschiedlichen Kameratypen durchgeführt. Die Schwarz-Weiß Kamera SONY XC 55 hat eine Auflösung von 628x492 Pixeln und war an einen Framegrabber des Typs PCEye4 der Firma Eltec angeschlossen. Bei den beiden weiteren Farbkameras handelt es sich um eine JAI CVM 70 mit einer Auflösung von 620x490 Pixeln, die an einen Framegrabber des Typs PCEye2 angeschlossen ist. Bei der SONY DFW V500 handelt es sich um eine Digitalkamera mit einer Auflösung von 640x480 Pixeln, bei der die Bildübertragung über einen IEEE1394 Bus erfolgt. Im Testszenario wurden Brennweiten von 4,8 mm bis 12 mm verwendet.

Der Vollständigkeit halber sei noch erwähnt, dass bei der SONY DFW V500 das Format des CCD-Chips nur geschätzt ist. Daher sind die berechneten Brennweiten nur bedingt aussagekräftig.

Kameratyp	**Objektiv**	**C_X [Pixel]**	**C_Y[Pixel]**	**κ**	**f [mm]**	**Genauigkeit [Pixel]**
SONY XC 55	4,8 mm	331,8	253,0	10724,4	4,847	0,374
Bildzentrum 314/296	6 mm	329,6	245,8	5111,7	6,155	0,294
	12 mm	321,3	252,3	2766,4	12,144	0,315
JAI CV M70	4,8 mm	305,5	240,5	10514,8	4,823	0,351
Bildzentrum 310/245	6 mm	298,0	236,9	4541,5	6,151	0,358
	12 mm	306,8	242,9	2763,6	12,188	0,339
SONY DFW V500	4,8 mm	314,9	239,9	10535,7	4,846	0,323
Bildzentrum 320/240	6 mm	311,4	234,6	5128,6	6,159	0,305
	12 mm	296,1	255,1	2834,0	12,158	0,326

Tabelle 2.1: Ergebnisse der Kamerakalibrierung

Es ist auffällig, dass die berechnete Bildmitte vom Bildzentrum abweicht. Unabhängig vom verwendeten Objektiv variiert die berechnete Bildmitte aber jeweils nur über einem Bereich von fünf bis zehn Pixeln. Die einzige Ausnahme bildet das 12mm Objektiv auf der SONY DFW V500. Daraus kann geschlossen werden, dass die Bildmitte weniger vom Objektiv als

viel mehr von der Lage des CCD-Chips in der Kamera abhängt. Der Parameter κ als Maß der Kissenverzerrung ist erwartungsgemäß umgekehrt proportional zur Brennweite des verwendeten Objektivs. So erzeugt die Kissenverzerrung im Bildrandbereich bei der Verwendung eines 4,8mm Objektiv einen Fehler von 35 Pixeln. Ein 12mm Objektiv führt dagegen nur zu einem Fehler von 10 Pixeln. Die Brennweite der Objektive wird bei allen Kameras relativ gut berechnet. Die Abweichungen zwischen den verschiedenen Kameras liegen bei gleichen Objektiven unter einem halben Prozent. Die leichte Überschätzung der tatsächlichen Werte ist darauf zurück zu führen, dass die mathematische Definition der Brennweite nicht exakt mit der optischen Definition übereinstimmt.

Die angegebene Genauigkeit wurde dadurch bestimmt, dass die Kalibriermarken durch das berechnete Kameramodell zurück in die Bildebene projiziert werden. Der mittlere Abstand zwischen den berechneten und den tatsächlich gemessenen Koordinaten entspricht dieser Genauigkeit. Das Ergebnis liegt unabhängig von verwendeter Kamera und verwendetem Objektiv bei etwa 0,3 bis 0,35 Pixeln. Das ist besonders dadurch bemerkenswert, dass die SONY DFW V500 aufgrund des verwendeten Datenformats (YUV411[1]) eine geringere Auflösung hat, als die anderen beiden Kameratypen. Die Ursache wird allerdings darin zu suchen sein, dass die Genauigkeit das Maß wiederspiegelt, mit dem die Marken detektiert wurden. Praktische Anwendungen der Kalibrierdaten im Rahmen von 3D-Vermessungen zeigen zudem, dass die tatsächliche Genauigkeit der Kamerakalibrierung um den Faktor zwei bis drei genauer ist, als der hier dargestellte Wert. Mit der SONY XC 55 wird eine tatsächliche Genauigkeit von etwa 0,15 Pixeln erreicht. Bei der SONY DFW V500 liegt die tatsächliche Genauigkeit aufgrund der niedrigeren Auflösung bei etwa 0,25 Pixeln. Beispiele für 3D-Vermessungen, mit denen diese tatsächlichen Größen berechnet wurden sind im Kaptitel 3 (Aktive Stereoskopie) zu finden.

Abschließend lässt sich sagen, dass das verwendete Kameramodell sehr gut die verwendeten Kameras beschreibt. Die Kalibrierung liefert unabhängig von der verwendeten Kamera und dem verwendeten Objektiv immer gleich gute Ergebnisse. Über den Vergleich dieser Ergebnisse mit den Genauigkeiten, die experimentell im Rahmen von 3D-Vermessungen erreicht wurden, sind aber noch systematische Fehler feststellbar. Die höchste Genauigkeit einer 3D-Vermessung, die durch die Kombination von modellbasierter und modellfreier Kalibrierung erreicht wurde, lag in Versuchen bei etwa 0,08 Pixeln. Allerdings ist diese Größe mit Vorsicht zu genießen, da besonders bei einer modellfreien Kalibrierung die Gefahr besteht, Ungenauigkeiten im Kalibrierkörper zu modellieren[2]. Die Methoden zur Kalibrierung von 3D-Sensoren, mit denen unter anderen die Ergebnisse der Kamerakalibrierung verifiziert wurden, sind Inhalt den folgenden beiden Kapitel zur modellfreien und modellbasierten Kalibrierung.

1. Vier Grauwerte (Y) und eine Farbe (U,V) für vier nebeneinanderliegende Pixel
2. Die Genauigkeit wurde berechnet, indem der verwendete Kalibrierkörper als Referenzkörper vermessen wurde.

2.2 Modellbasierte Kalibrierung

Bei der modellbasierten Kalibrierung wird der gesamte Triangulationssensor über ein mathematisches Modell beschrieben. Das Ziel dabei ist eine möglichst getreue Nachbildung aller relevanten Komponenten des Sensors mit Hilfe mathematischer Beschreibungen. Besonders bei optischen Triangulationssensoren spielt daher das Modell der eingesetzten Kamera eine wesentliche Rolle. Das gesamte Modell setzt sich aus den Beschreibungen der einzelnen Komponenten des Sensors zusammen und umfasst zusätzlich die Verbindung der Komponenten untereinander. Vereinfachungen des Modells werden dabei in soweit getroffen, wie sie der Modellbeschreibung nicht schaden oder zur Vermeidung von Grenzen der Mathematik dienen. Am Beispiel eines Lichtschnittsensors sollen hier die wesentlichen Aspekte der modellbasierten Kalibrierung erläutert werden.

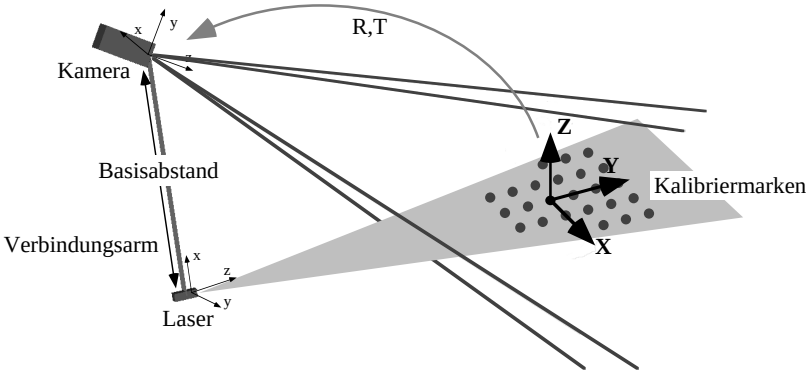


Abbildung 2.6: Kalibrierung eines Lichtschnittsensors

Aus Abbildung 2.6 lassen sich Kamera, Laser und Verbindungsarm als elementare Komponenten des Sensors identifizieren. Für den zur Ebene aufgeweiteten Laser wird angenommen, dass er eine ideale, völlig planare Halbebene aufspannt. Das Querschnittsprofil des Lasers wird als Gauß-förmig angenommen. Daher kann bei der Bildauswertung über eine einfache Schwerpunktbildung subpixelgenau die Linienmitte bestimmt werden. Für die eigentliche Kalibrierung ist diese Annahme allerdings irrelevant. Krümmungen in der vom Laser projizierten Ebene werden ebenfalls nicht berücksichtigt, da sie bei realen Systemen vernachlässigbar gering sind. Die Beschreibung der Kamera entspricht dem von Tsai vorgestellten Modell. Der Verbindungsarm als mechanische Kopplung zwischen Kamera und Laser definiert die Lage des Koordinatensystems, in dem sich die Kamera und der Laser befinden. Ohne Einschränkung wird die XY-Ebene des Koordinatensystems mit der vom Laser aufgespannten Ebene gleichgesetzt. Die genaue Position des Koordinatensystems innerhalb der Ebene wird während der Kalibrierung durch eine definierte Kalibriermarke festgelegt. Die Höhe der Kamera über der Ebene wird als Basisabstand bezeichnet.

Für die eigentliche Kalibrierung wird eine Vielzahl von Marken bekannter Positionen ausgemessen. Dafür kann beispielsweise ein Stab mit Hilfe eines Positioniertisches durch die Laserebene verfahren werden. Die Bildauswertung liefert dann über den reflektierten Laser auf dem Stab für jede angefahrende Position Bildkoordinaten. Mit einer Kamerakalibrierung nach [Tsai87] können dann die Parameter der Kamera bestimmt werden. Die externen Parameter entsprechen dabei der Lage der Kamera im Sensorkoordinatensystem. Besonders bei kleinem

Basisabstand ist die Kalibrierung allerdings schlecht konditioniert. Die Kamera sieht in einem sehr flachen Winkel auf die Kalibriermarken und besonders die Bestimmung der internen Kameraparameter wird dadurch instabil. Eine zweischrittige Kalibrierung, die im ersten Schritt mit einer geeigneten Kalibrierung nur die internen Kameraparameter bestimmt und im zweiten Schritt mit dem beschriebenen Verfahren die weiteren Parameter des Sensors berechnet, ist daher stabiler und genauer.

Die Parameter der Kamerakalibrierung erlauben die Berechnung eines Sehstrahls für jedes Pixel der Kamera. Wird eine Laserlinie im Bild detektiert, so kann die Linie in einzelne Bildpunkte zerlegt werden. Der Schnitt jedes Sehstrahls mit der XY-Ebene des Kalibrierkoordinatensystems ergibt die Raumpunkte des erfassten Objektes und erlaubt so mit jedem Bild die Rekonstruktion eines Schnitts durch das Objekt.

Die modellbasierte Kalibrierung hat den Vorteil, dass die Realität im Computer nachgebildet wird. Das Ergebnis einer Kalibrierung lässt sich einfach validieren, da die berechneten Modellparameter mit den real bekannten oder geschätzten Parametern verglichen werden können. Im Weiteren sind die Anforderungen an das Kalibriersystem als eher gering anzusetzen. Sofern die Fehler, die bei der Detektion der Kalibriermarken gemacht werden, stochastisch verteilt sind, ist das Ergebnis der Kalibrierung bei einer ausreichenden Zahl von Marken sehr viel genauer als das Maß, mit dem die Marken gemessen wurden. Einzelne Ausreißer bei der Markenerfassung haben eine sehr geringe Wirkung auf das Gesamtergebnis. Reale Modellparameter, die nicht bekannt sind oder bei der Kalibrierung nicht berücksichtigt wurden, gehen nicht in die Modellberechnung ein und begrenzen die Genauigkeit des Systems. Eine zu große Zahl von Modellparametern führt zudem oft zu Mehrdeutigkeiten in der Modellberechnung und damit zu mathematischen Instabilitäten. Daher ist bei hohen Genauigkeitsanforderungen der Einsatz von hochwertigen Objektiven mit geringen Verzeichnungen sinnvoll, da die Modellparameter dadurch einfach gehalten werden können.

2.3 Modellfreie Kalibrierung

Vor allem, wenn wenig über den Aufbau oder die Funktion eines 3D-Sensors bekannt ist, bietet sich eine modellfreie Kalibrierung an. Die Abbildungseigenschaften werden dabei als Abbildung vom Bildraum in ein globales Koordinatensystem beschrieben. Bei einem Lichtschnittsensor spannt das Kamerabild das zweidimensionale Bildkoordinatensystem auf. Das globale Koordinatensystem wird in einer Ebene von der Laserlinie definiert. Die Kalibrierung muss also die Abbildung zwischen zwei zweidimensionalen Räumen bestimmen. Die technische Durchführung der Kalibrierung geschieht wie bei der modellbasierten Kalibrierung durch die Vermessung einer Vielzahl von Kalibriermarken. Die Umsetzung dieser Stützpunkte in eine funktionale Abbildung kann dann mit interpolierenden oder approximierenden Verfahren geschehen, die im Folgenden exemplarisch am Beispiel des Lichtschnittsensors vorgestellt werden.

2.3.1 Interpolierende Kalibrierung

Wenn die Detektion der Kalibriermarken mit hoher Genauigkeit und Sicherheit erfolgt und zudem eine hohe Zahl von Marken durch den gesamten Messraum erfasst werden kann, bietet sich eine interpolierende Kalibrierung an. Die Kalibrierung an sich speichert die globalen Koordinaten der erfassten Marken zusammen mit den zugehörigen Bildkoordinaten in einer Tabelle ab. Bei einer Objektvermessung werden die gemessenen Bildkoordinaten mit der Tabelle verglichen. Sofern eine Bildkoordinate in der Tabelle vorliegt, kann direkt die gespeicherte globale Koordinate als Ergebnis der Vermessung verwendet werden. Falls kein Treffer existiert, werden die drei oder vier, den gesuchten Bildkoordinaten am nächsten liegenden Einträge ermittelt. Entsprechend der Bildlage der gesuchten Koordinate wird zwischen den globalen Koordinaten der gefundenen Einträge interpoliert und so ein Ergebnis berechnet.

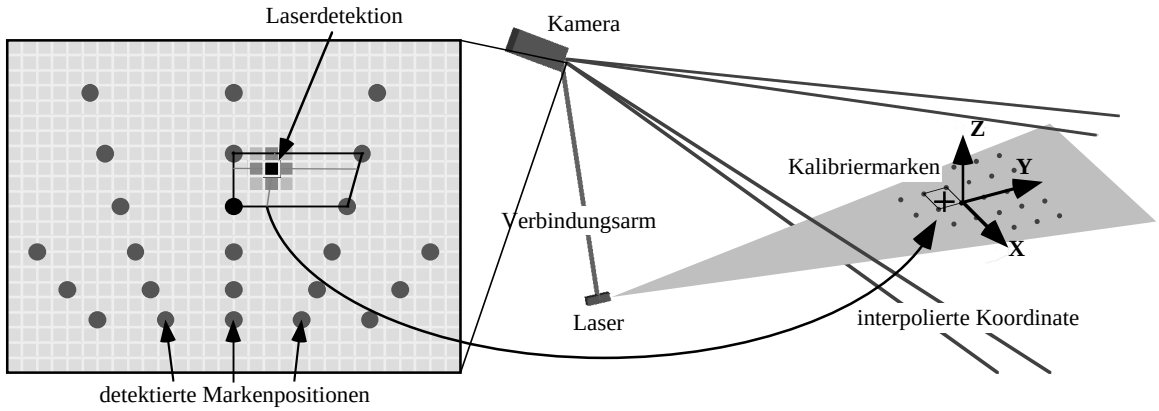


Abbildung 2.7: Interpolierende Kalibrierung

Bei Bildräumen begrenzter Größe kann die Berechnung der Abbildung von Bildkoordinaten in Raumkoordinaten für jedes Pixel einmal vorab bestimmt werden. Der Speicheraufwand der Kalibrierung erhöht sich dadurch zwar beträchtlich, die Umrechnung von Bildkoordinaten in globale Koordinaten kann dann aber durch einen einfachen Tabellenzugriff erfolgen. Da die Interpolation im Bildraum durchgeführt wird, bietet es sich im Weiteren an, die Kalibriermarken so anzuordnen, dass die berechneten Stützpunkte der Kalibrierung im Bildkoordinatensystem eine möglichst gleichmäßige Verteilung ergeben. So kann gewährleistet werden, dass eine gleichmäßige Genauigkeit über den gesamten Bildraum erzielt wird. Es ist leicht einzusehen, dass die Gesamtgenauigkeit des Systems stark von der Zahl der Kalibriermarken abhängt. Ein absolut genauer Messwert kann nur berechnet werden, wenn die Bildkoordinaten des Messwertes mit den Koordinaten einer bereits gespeicherten Kalibriermarke übereinstimmen. Je weiter die Bildkoordinate von den Koordinaten der Kalibriermarken entfernt ist, um so ungenauer wird die Messung. Ein hohe Dichte von Kalibriermarken sichert also eine hohe Messgenauigkeit. Alternativ dazu kann die Genauigkeit dadurch verbessert werden, dass nicht nur die dem Messpunkt nächsten Bildkoordinaten der Kalibrierung zur Interpolation verwendet werden, sondern weitere Bildkoordinaten aus der größeren Umgebung verwendet werden. Da dann nicht mehr nur linear interpoliert werden kann, sondern Interpolationsverfahren höherer Dimension verwendet werden müssen, vergrößert sich natürlich auch der Rechenaufwand entsprechend.

Die Gesamtgenauigkeit der Kalibrierung wird allerdings auch durch die Genauigkeit der am schlechtesten gemessenen Kalibriermarke begrenzt. Da jede Kalibriermarke als sichere Referenzinformation vorausgesetzt wird, führt eine schlecht oder ungenau gemessene Marke zu einer Krümmung in der Abbildung zwischen Bild und globalem Koordinatensystem. Die Messung eines Objektes kann in keinem Fall genauer geschehen, als die Genauigkeit, mit der die Kalibriermarken erfasst wurden. Für Bildkoordinaten, die außerhalb des kalibrierten Bildbereichs erfasst wurden, muss zudem eine Extrapolation berechnet werden, so dass besonders im Randbereich des Bildkoordinatensystems die Qualität der Daten sehr stark abnimmt.

2.3.2 Approximierende Kalibrierung

Einen Kompromiss zwischen der modellbasierten und der interpolierenden Kalibrierung stellt die approximierende Kalibrierung dar. Es wird auf ein Modell des Sensors verzichtet, die Abbildung zwischen Bildkoordinaten und Weltkoordinaten wird zur Fehlerkompensation approximierend über hochgradig, mehrdimensionale Polynome berechnet. Die Aufgabe der Kalibrierung besteht in der Berechnung der Koeffizienten dieser Polynome. Ein allgemeine Formulierung der Polynome für einen Lichtschnittsensor ist

$$g_X(X,Y) = \sum_i c_{Xi} f_{Xi}(X,Y), \tag{2.3}$$

$$g_Y(X,Y) = \sum_j c_{Yj} f_{Yj}(X,Y). \tag{2.4}$$

Jede globale Koordinate wird durch ein eigenes Polynom berechnet (g_X,g_Y). Dabei werden die Bildkoordinaten durch X und Y beschrieben. Die Funktionen (f_{Xi},f_{Yj}) beschreiben die Summanden der Polynome, die mit den Koeffizienten (c_{Xi},c_{Yj}) gewichtet werden.

Mit jeder Kalibriermarke und der zugehörigen Bildkoordinate lässt sich eine Gleichung aufstellen, deren Unbekannte die Koeffizienten des Polynoms sind. Wenn wenigstens soviele Kalibriermarken wie Koeffizienten vorliegen, kann das Gleichungssystem gelöst und die Koeffizienten berechnet werden. Es bietet sich allerdings an, zur Bestimmung der Koeffizienten deutlich mehr Kalibriermarken als Koeffizienten zu verwenden. Die Berechnung der Koeffizienten kann dann mit einem Least-Square Ansatz geschehen. Aus dem Ergebnis lässt sich direkt die Qualität der berechneten Kalibrierung ermitteln. Dazu wird mit dem berechneten Polynom aus den Bildkoordinaten der Kalibriermarken die zugehörige globale Position berechnet und mit den tatsächlichen Positionen verglichen.

Die Ausgleichsrechnung zur Bestimmung der Polynomkeoffizienten hat den Vorteil, dass vereinzelte Fehler während der Markenbestimmung bei der Kalibrierung ausgeglichen werden. Die Gesamtgenauigkeit der Kalibrierung kann über den Grad der verwendeten Polynome und den verbleibenden Approximationsfehler kontrolliert werden. Solange die Zahl der Kalibriermarken deutlich höher als die Zahl der Koeffizienten der Polynome ist, kann der Grad der Polynome erhöht werden. Wenn keine Qualitätsverbesserung mehr erreicht wird, ist die bestmögliche Kalibrierung bestimmt. Es ist allerdings anzumerken, dass diese Kalibrierung nicht optimal sein muss, da vor allem Bildbereiche, die außerhalb der mit Kalibriermarken durchsetzten Gebiete liegen, sehr schnell große Fehler produzieren. Zudem können Bildbereiche mit besonders hoher Dichte von Kalibriermarken andere Bildbereiche dominieren, so dass lokal

zwar eine gute Kalibrierung vorliegt, das Gesamtsystem allerdings noch recht hohe Fehler produziert. Auch bei der approximierenden Kalibrierung ist also eine sorgfältige Durchführung der Kalibrierung zu planen. Sie stellt aber einen guten Kompromiss einerseits zwischen der modellbasierten Kalibrierung dar, wenn die Modellparameter nicht hinreichend gut bekannt sind und andererseits der interpolierenden Kalibrierung, wenn die Markenbestimmung nicht ausreichend genau erfolgen kann.

2.4 Stereokalibrierung

Inhaltlich ist die Stereokalibrierung der modellbasierten Kalibrierung zuzuordnen. Der zu kalibrierende Stereosensor besteht aus zwei Kameras, die auf das zu vermessende Objekt ausgerichtet werden. Die Lage der Kameras zueinander ist a priori unbekannt und wird im Rahmen der Stereokalibrierung berechnet. Der Unterschied zu den bisher beschriebenen Kalibrierverfahren ist das Fehlen eines Kalibrierkörpers. Die Kalibrierung wird aus den Messdaten berechnet. Der Sensor wird daher als selbstkalibrierend bezeichnet.

Das Sensormodell der Stereokalibrierung setzt ideale Kameramodelle voraus. Die inneren Parameter der eingesetzten Kameras müssen daher vorab durch eine geeignete Kamerakalibrierung einmalig bestimmt werden. Jeder Bildpunkt, der von einer Kamera aufgenommen wird, wird dann zunächst über das innere Kameramodell in eine unverzerrte Bildkoordinate umgerechnet, bevor er in die Stereokalibrierung eingeht. Die zweite Annahme zur Berechnung der Stereokalibrierung besteht in der Festlegung der Brennweiten beider Kameras auf einen Wert von Eins. Diese Festlegung stellt keine Einschränkung dar, da jede Kamera mit bekannter Brennweite durch Division der Kamerakoordinaten durch die Brennweite in eine Kamera mit der Brennweite Eins transformiert werden kann. Zudem vereinfacht diese Vereinheitlichung die Berechnung der Stereokalibrierung beim Einsatz von Kameras mit unterschiedlichen Bildformaten oder Brennweiten.

2.4.1 Berechnung der Epipolargeometrie

Die Grundlagen der im Folgenden erläuterten Stereokalibrierung sind aus dem Forschungsgebiet *Structure from Motion* abgeleitet. Inhalt dieses Forschungsgebietes ist die Berechnung der Bewegung eines Objektes vor einer Kamera durch die Analyse der Kamerabilder und der Verfolgung von homologen Punkten auf dem Objekt. Aus der zweidimensionalen Bewegungsbahn mehrerer homologen Punkte in den Kamerabildern wird dann die dreidimensionale Bewegung des Objektes berechnet. Auf die Stereokalibrierung bezogen kann das Problem umgekehrt betrachtet werden. Das Objekt steht still und eine Kamera wird um das Objekt herum bewegt. Mathematisch gesehen sind beide Definitionen äquivalent und können durch eine einfache Transformation ineinander überführt werden. Für die Stereokalibrierung wird im Weiteren angenommen, dass das Objekt nur aus zwei Ansichten aufgenommen wurde. Dabei ist es unerheblich, ob die Bilder durch zwei Aufnahmen einer Kamera oder durch jeweils eine Aufnahme von zwei Kameras aufgenommen wurden. Schließlich wird ohne Beschränkung der Allgemeinheit durch die Aufnahmeposition des ersten Bildes das Koordinatensystem definiert, in dem die Stereokalibrierung berechnet wird. Die Aufnahmeposition des zweiten Bildes definiert sich dann im Koordinatensystem des ersten Bildes. Inklusive der im Folgenden verwendeten Bezeichnungen wird das Modell des Stereosensors entsprechend Abbildung 2.8 dargestellt.

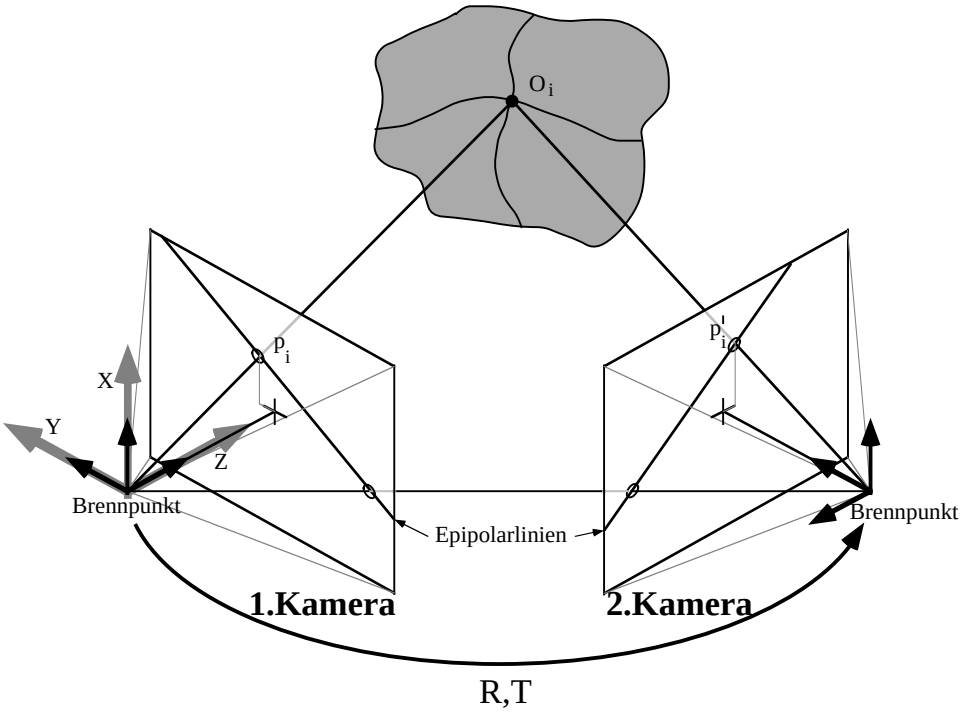


Abbildung 2.8: Modell des Stereosensors

Die Eingangswerte für die Stereokalibrierung sind Paare von Bildpunkten (p_i,p_i') in homogenen Koordinaten, die in den beiden Kamerabildern der Abbildung der gleichen Objektpunkte (O_i) entsprechen. Der erste Schritt bei der Berechnung der Stereokalibrierung besteht in der Berechnung der Epipolargeometrie, die durch die Epipolarmatrix (M) repräsentiert wird. Im zweiten Schritt wird die Epipolarmatrix in die Objekttranslation $(\hat{T})$ und -rotation $(\hat{R})$ zerlegt, aus denen dann die relative Lage der Kameras zueinander berechnet werden kann (Abbildung 2.8).

$$M = \begin{pmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{pmatrix} \tag{2.5}$$

Die Epipolargeometrie beschreibt die Lage der korrespondierenden Punkte beider Kameras zueinander. Die Gleichung, die für zwei Punkte gelten muss, lautet dabei

$$p_i^T \bullet M \bullet p_i' = 0. \tag{2.6}$$

Die Ausmultiplikation des ersten Produktes führt zu

$$(a\ b\ c)\, p_i' = 0 \tag{2.7}$$

oder

$$(\,a\;b\,)\begin{pmatrix} p_{Xi}' \\ p_{Yi}' \end{pmatrix} + c = 0, \tag{2.8}$$

was der Hess'schen Geradengleichung entspricht, auf der die Korrespondenz zu einem Punkt im zweiten Bild liegen muss. Diese sogenannte Epipolarlinie ist anschaulich der Schnitt der Ebene, die durch die beiden Brennpunkte der Kameras und den Raumpunkt O_i definiert wird, und der Bildebene der zweiten Kamera. Analog dazu kann zu jedem Punkt der zweiten Kamera die Epipolarlinie im ersten Bild berechnet werden. Die Epipolarmatrix (M'), die diese Beziehung beschreibt, berechnet sich als

$$M' = M^T. \tag{2.9}$$

Der Schnittpunkt aller Epipolarlinien eines Bildes wird als Epipol bezeichnet. Er ist anschaulich der Punkt im Bild, an dem die andere Kamera zu sehen ist. Wenn die Bildfelder beider Kameras koplanar sind, liegen die Epipole im Unendlichen.

Aus (2.6) ergibt sich mit allen Korrespondenzen ein homogenes Gleichungssystem mit neun Unbekannten:

$$\begin{pmatrix} p_{X1}p_{X1}' & p_{X1}p_{Y1}' & p_{X1} & p_{Y1}p_{X1}' & p_{Y1}p_{Y1}' & p_{Y1} & p_{X1}' & p_{Y1}' & 1 \\ p_{X2}p_{X2}' & p_{X2}p_{Y2}' & p_{X2} & p_{Y2}p_{X2}' & p_{Y2}p_{Y2}' & p_{Y2} & p_{X2}' & p_{Y2}' & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \end{pmatrix} = \vec{0}. \tag{2.10}$$

Da das Stereosystem bis auf einen Skalierungsfaktor eindeutig ist, kann eine Komponente der Epipolarmatrix beliebig gewählt werden, so dass sich ein lösbares Gleichungssystem mit acht Unbekannten ergibt. Für den Fall, dass e_9 auf den Wert Eins gesetzt wird ergibt sich das folgende Gleichungssystem:

$$\begin{pmatrix} p_{X1}p_{X1}' & p_{X1}p_{Y1}' & p_{X1} & p_{Y1}p_{X1}' & p_{Y1}p_{Y1}' & p_{Y1} & p_{X1}' & p_{Y1}' \\ p_{X2}p_{X2}' & p_{X2}p_{Y2}' & p_{X2} & p_{Y2}p_{X2}' & p_{Y2}p_{Y2}' & p_{Y2} & p_{X2}' & p_{Y2}' \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{pmatrix} = \vec{-1}. \tag{2.11}$$

Wenn acht Korrespondenzen vorliegen, ist das Gleichungssystem bestimmt und eindeutig lösbar. Um Ungenauigkeiten in der Korrespondenzpunktbestimmung zu kompensieren, bietet es sich allerdings an, eine Vielzahl von Korrespondenzen in das Gleichungssystem einzubringen und die Lösung mit einem Least-Square Verfahren [Press86] zu berechnen.

Nach [Huang86] setzt sich die Epipolarmatrix im Weiteren aus der Objekttranslation und der Objektrotation zusammen:

$$M = \hat{R} \bullet A, \tag{2.12}$$

mit

$$A = \begin{pmatrix} 0 & \hat{t}_X & -\hat{t}_Y \\ -\hat{t}_X & 0 & \hat{t}_Z \\ \hat{t}_Y & -\hat{t}_Z & 0 \end{pmatrix}, \hat{T} = \begin{pmatrix} \hat{t}_X \\ \hat{t}_Y \\ \hat{t}_Z \end{pmatrix}. \tag{2.13}$$

Die eindeutige Zerlegung der Epipolarmatrix wird im Detail in [Huang86] beschrieben. Die relative Lage der Kameras zueinander berechnet sich dann aus

$$\begin{aligned} R &= \hat{R}^{-1} \\ T &= \hat{R}^{-1} \bullet -\hat{T} \end{aligned}. \tag{2.14}$$

Der große Vorteil der Stereokalibrierung besteht in der Unabhängigkeit von einem Kalibrierkörper. Die Messdaten dienen zur Bestimmung der relativen Lage der Kameras zueinander. Die erreichbare Genauigkeit von Stereosensoren liegt abgesehen von der fehlenden Skalierung, in der Größenordnung, die auch mit vorkalibrierten optischen Systemen erzielt werden kann. Allerdings ist die Kondition des Gleichungssystem aus (2.11) sehr schlecht, so dass die Sensordaten zur Berechnung der Epipolarmatrix transformiert werden müssen, um numerische Instabilitäten zu vermeiden [Hartley95]. Praktische Untersuchungen im Rahmen dieser Arbeit

haben ergeben, dass eine Genauigkeit der Sensordaten von wenigstens einem Hundertstel der Bilddiagonale notwendig ist, um brauchbare Ergebnisse zu erzielen. Optimale Ergebnisse werden mit bis zu hundert Korrespondenzen erreicht, danach führen weitere Korrespondenzen nicht mehr zu Verbesserungen. Daher ist es sinnvoll, eine sorgfältige Auswahl der Korrespondenzen zu treffen, so dass zum einen eine gleichmäßige Verteilung über beide Bildbereiche gewährleistet ist, und zum anderen sicher detektierte Bildpunkte unsicheren Bildpunkten vorgezogen werden.

Die bekannten inneren Kameraparameter garantieren die Rekonstruktion von Stereokorrespondenzen in Raumpunkte. Durch die berechnete Rotation und Translation zwischen beiden Kameras können beiden Bildpunkten Projektionsstrahlen zugeordnet werden, deren Schnittpunkt im Raum dem aufgenommenen Objektpunkt entspricht. Im normalen Fall, in dem die Projektionsstrahlen windschief aneinander vorbeilaufen, wird der Raumpunkt verwendet, der beiden Geraden am nächsten liegt. Der Abstand der beiden Geraden zueinander kann zusätzlich als Kriterium zur Bewertung der Qualität oder internen Genauigkeit des rekonstruierten Punktes verwendet werden. Die Summe der Qualitäten aller Punkte entspricht der Gesamtqualität der berechneten Stereokalibrierung.

2.4.2 Nichtlineare Berechnung der Stereokalibrierung

Die im letzten Kapitel berechnete Epipolarmatrix enthält neun Unbekannte. Wenn die Skalierung des Gesamtsystems unberücksichtigt bleibt, müssen noch acht Parameter berechnet werden. Tatsächlich enthält das hier beschriebene Stereosystem allerdings nur drei translatorische und drei rotatorische Freiheitsgrade. Aufgrund der fehlenden Skalierung würde sich das System sogar nur mit fünf Parametern beschreiben lassen. Bei schlecht konditionierten Stereosystemen, also Fällen, in denen beide Kameras sehr dicht beieinander stehen oder sehr lange Brennweiten verwendet werden, führen die überzähligen Parameter in der Epipolarmatrix zu ungenauen Ergebnissen. Die Rekonstruktion der Daten ist zwar mit einer hohen internen Genauigkeit möglich, die externe Genauigkeit wird aber sehr schnell sehr schlecht. Das liegt vor allem daran, dass die aus der Epipolarmatrix berechnete Matrix $(\hat{R})$ nicht orthogonal ist, also keiner echten Rotation entspricht. Die berechnete Geometrie wird im Raum gekrümmt, so dass besonders bei schlecht konditionierten Systemen große globale Fehler entstehen.

Die direkte Berechnung der relativen Lage der Kameras zueinander vermeidet dieses Problem. Ein System, das aus der Translation und den drei Rotationswinkeln zwischen den beiden Kameras besteht, kann allerdings nicht mehr als lineares System beschrieben werden. Die Berechnung der Parameter erfolgt daher iterativ mit einem nichtlinearen Optimierungsverfahren [MinPack]. Im Weiteren ist zu beachten, dass die Berechnung der Stereokalibrierung mehrere Lösungen hat. Die vier trivialen Lösungen (Abbildung 2.9) unterscheiden sich im Vorzeichen des Translationsvektors oder in der Rotation der zweiten Kamera um 180 Grad entlang einer Bildachse.

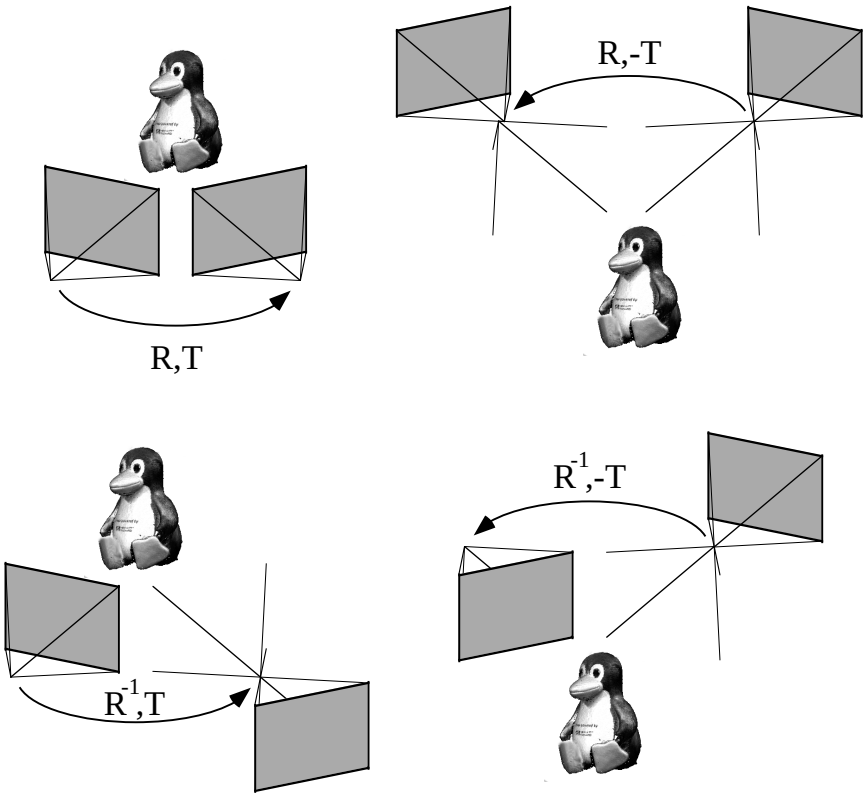


Abbildung 2.9: Lösungen der Stereokalibrierung

Die Objektgeometrie wird zwar in allen vier Fällen korrekt berechnet, aber nur in einem Fall werden alle Raumpunkte vor beiden Kameras rekonstruiert. In den drei anderen Fällen findet die Rekonstruktion hinter wenigstens einer der beiden Kameras statt. Die drei ungültigen Fälle können daher in den korrekten Fall transformiert werden und sind unproblematisch. Allerdings enthält das Stereosystem je nach Konfiguration mehrere lokale Minima, in denen sich ein nichtlinearer Optimierer verfangen kann. Der typische singuläre Fall ist ein Translationsvektor der Länge Null, der dazu führt, dass die Rekonstruktion in einem einzigen Punkt stattfindet, also wertlos ist, obwohl eine optimale Qualität erreicht wird. Es bietet sich daher an, mit dem linearen Verfahren einen Startwert zu berechnen, der dann optimiert wird. Die Größe, die optimiert wird, ist der windschiefe Abstand zwischen den Projektionsstrahlen korrespondierender Punkte. Dabei wird die Länge des Translationsvektors auf Eins festgelegt, um den singulären Fall zu vermeiden.

Die nichtlineare Optimierung zur Berechnung der Stereokalibrierung ist um ein Vielfaches aufwändiger als das lineare Verfahren. Die im Rahmen dieser Arbeit erreichte interne Genauigkeit der rekonstruierten Daten entspricht der des linearen Verfahrens. Beide Verfahren sind also in dieser Hinsicht als gleichwertig anzusehen. Da die nichtlineare Optimierung aber direkt die Rotation zwischen den beiden Kameras berechnet, erreicht sie zudem eine hohe externe Genauigkeit bei der Objektrekonstruktion. Globale Krümmungen sind auch bei schlecht konditionierten Stereosystemen kaum noch nachzuweisen.

3 Aktive Stereoskopie

Das Hauptproblem der passiven Stereoskopie, die Suche nach korrespondierenden Punkten in den Bildern zweier Kameras, wird bei der aktiven Stereoskopie durch den Einsatz von Lichtquellen gelöst. Im einfachsten Fall wird ein Laserstrahl verwendet, um einen einzelnen Punkt auf die Objektoberfläche zu projizieren. Die Korrespondenzpunktsuche beschränkt sich dann auf die Suche dieses Punktes in den Kamerabildern. Nachteil dieses Verfahrens ist die niedrige Datenrate des Sensors. Mit jedem Bildpaar kann nur ein 3D-Punkt aufgenommen werden. Damit wird beim Einsatz von Standardkamerasystemen eine maximale Datenrate von 30 Punkten pro Sekunde erreicht.

Höhere Datenraten werden durch den Einsatz komplexerer Muster erreicht. Der Einsatz eines zur Linie aufgeweiteten Lasers ermöglicht bereits die Berechnung von mehreren hundert Korrespondenzen pro Bild. Die Projektion eines sequentiell kodierten Gray-Code Musters schließlich führt zu mehreren hunderttausend Korrespondenzen in einer Sequenz von zwanzig aufeinanderfolgenden Bildpaaren. Obwohl eine Vielzahl weiterer Muster denkbar sind, wurden diese drei Varianten im Rahmen der Entwicklung des 3D-Sensors *HandyScan3D* untersucht und werden im Folgenden genauer erläutert.

Die beiden Sensorsysteme *HandyScanTT* und *HandyScanCLM*, die auf Basis von *HandyScan3D* entwickelt wurden und zur Aufnahme komplexer Objektgeometrien dienen, bilden den Abschluss dieses Kapitels.

3.1 HandyScan3D

3.1.1 HandyScan3D im Überblick

Die Basisausstattung von *HandyScan3D* besteht im einfachsten Fall aus zwei CCD-Kameras und einem handelsüblichen Klasse-III-Laserstrahler, sowie der nötigen Hardware, um die Bilddaten der Kameras in den Rechner zu übertragen (Abbildung 3.1). Das im Rahmen dieses Projektes entwickelte Softwarepaket übernimmt die Bildverarbeitung, Kalibrierung und 3D-Rekonstruktion der Daten.

Als Vorbereitung für eine 3D-Erfassung werden die beiden Kameras auf das zu vermessende Objekt ausgerichtet, so dass das Objekt möglichst bildfüllend aufgenommen wird und beide Kameras in etwa den gleichen Bereich des Objektes erfassen. Zur eigentlichen 3D-Erfassung wird das Objekt mit der Lichtquelle angestrahlt. Die Software extrahiert die Lichtreflexion auf dem Objekt aus den Bildern der beiden Kameras und ordnet identische Punkte einander zu. Die Korrespondenzen werden einerseits zur Stereokalibrierung des Sensors verwendet, andererseits kann bei bekannter Kalibrierung gleichzeitig eine 3D-Rekonstruktion durchgeführt werden. Durch eine geschickte Auswahl der Korrespondenzen und mit zunehmender Dauer

der 3D-Erfassung kann die Genauigkeit und Auflösung der Vermessung sukzessive gesteigert werden.

Abbildung 3.1: Hardware von HandyScan3D und Messobjekt (Tasse)

Wenn die Lichtquelle manuell geführt ist, können sehr gezielt markante Punkte des Objektes angestrahlt werden. Die 3D-Erfassung kann dadurch der Topologie des Objektes angepasst werden. Unwichtige Bereiche werden grob erfasst, während Details hochauflösend abgetastet werden können. Für die bessere Kontrolle der aktuellen Messergebnisse findet eine Online-Visualisierung der Daten statt. Die Kamerabilder können zur Texturierung der Darstellung verwendet werden und sorgen dadurch für fotorealistische Visualisierung des Objektes.

a) Punkte b) trianguliertes Dreiecksnetz c) texturierte Darstellung

Abbildung 3.2: Rekonstruiertes Objekt (Tasse)

Die Detektion der Lichtquelle in den Kamerabildern wurde in drei Varianten realisiert. Bei der einfachsten Variante, dem Einsatz eines Laserstrahlers wird in jedem Bild nach einem einzelnen Lichtpunkt gesucht. Da der Laserstrahler eindeutige Korrespondenzen in beiden Bildern markiert, können diese Daten gleichzeitig zur Kalibrierung des Stereokamerasystems und zur 3D-Rekonstruktion des Objektes verwendet werden.

Die zweite Variante besteht in der Detektion eines zur Ebene aufgeweiteten Laserstrahls in beiden Bildern. Da durch die Linie keine eindeutigen Korrespondenzen mehr markiert werden, muss das System vorab kalibriert worden sein. Über die Epipolareinschränkung kann dann zu

den detektierten Bildpunkten eines Bildes der korrespondierende Punkt im anderen Bild gefunden werden. Praktischerweise wird der 3D-Sensor zunächst mit einem einfachen Laserstrahler kalibriert, danach wird der Betriebsmodus in die Liniendetektion umgeschaltet und im Weiteren ein Linienlaser zur Beleuchtung des Objektes verwendet.

Die aufwändigste Variante ist der Einsatz eines Gray-Code Projektors. Der Gray-Code Projektor wird mit Hilfe eines Statives fest auf das Objekt ausgerichtet. Im Gegensatz zur herkömmlichen 3D-Vermessung mit Gray-Code Projektoren werden zwei Gray-Code Sequenzen verwendet. Die zweite Sequenz entspricht der ersten in soweit, dass sie lediglich um 90 Grad verdreht ist. Dadurch werden nicht nur Linien kodiert, sondern auch einzelne Punkte, und es ist wiederum möglich, eindeutige Korrespondenzen in beiden Stereobildern zu finden und somit die Kalibrierung und die 3D-Rekonstruktion gleichzeitig zu berechnen.

Inhalt der folgenden Kapitel sind die Details dieser Verfahren, angefangen mit der im Rahmen des Projektes eingesetzten Bildverarbeitungshardware. Im Kapitel Bildverarbeitung werden die Ansätze zur Detektion des Laserpunktes, der Laserlinie und des Gray-Codes erläutert. Stereokorrespondenzen sind die Ergebnisse der Bildverarbeitung. Sie dienen zum einen zur Stereokalibrierung und zum anderen zur 3D-Rekonstruktion und Visualisierung des Objektes, dem Inhalt der letzten beiden Kapitel.

3.1.2 Verwendete Hardware

Eine Idee bei der Entwicklung von *HandyScan3D* war der Einsatz von preisgünstigen Standardkomponenten. Dennoch muss systemspezifischen Voraussetzungen Rechnung getragen werden. Für den Einsatz des Laserpunktes und der Laserlinie muss die Bilderfassung in beiden Kameras synchron erfolgen. Da der Laserpointer manuell geführt wird, kann er im Einzelfall sehr schnell bewegt werden, so dass nur bei gleichem Aufnahmezeitpunkt beider Stereobilder garantiert werden kann, dass auch die gefundenen Korrespondenzen zueinander passen. Im normalen Betriebsmodus belichtet eine Standardkamera etwa eine 30tel Sekunde. Der projizierte Punkt eines schnell bewegten Laserstrahlers stellt sich in diesem Zeitraum nicht mehr als Punkt sondern als Linie im Bild dar. Um eine hohe zeitliche Bewegungsauflösung zu erreichen, ist es daher sinnvoll, eine Kamera mit elektronischem Shutter einzusetzen und damit die Belichtungszeit zu verkürzen. Ein zusätzlicher Nutzen besteht in der Abdunkelung des Bildes bei kurzen Belichtungszeiten, mit der Überblendeffekte unterdrückt werden können. Allerdings wird dadurch auch die Helligkeit des Laserpunktes verringert.

Beim Einsatz eines Gray-Code Projektors ist weder die Synchronität, noch ein steuerbarer Shutter notwendig, da der Projektor fest auf das Objekt ausgerichtet wird und während der Aufnahme der Bildsequenz weder Objekt, noch Projektor bewegt werden dürfen. Bei identischen Code-Bildern kann damit auch die Aufnahme der Stereobildpaare zeitlich entkoppelt durchgeführt werden.

Die Entwicklung von *HandyScan3D* wurde auf Basis eines Stereokamerasystems entwickelt, das aus zwei SONY XC55 Monochromkameras und einem Framegrabber des Typs PCEye4 besteht. Die Kameras erzeugen 30 Vollbilder pro Sekunde und haben eine Auflösung von 684*492 Pixeln. Über einen speziellen Triggermodus werden die Kameras vom Framegrabber synchronisiert und der elektronische Shutter der Kameras gesteuert. Dadurch kann über den Rechner die Belichtungszeit der Kameras kontrolliert werden. Die Bildverarbeitung wurde mit

einem Windows NT Rechner mit einem 450 MHz Intel Pentium III Prozessor durchgeführt. Als Alternativsystem wurde die Einsetzbarkeit von zwei IEEE1394 (FireWire) Farbkameras des Typs SONY DFW V500 untersucht. Der Vorteil bei dieser Lösung liegt in der Digitalisierung der Bilddaten bereits in der Kamera und der digitalen Übertragung zum Rechner, der Framegrabber wird durch eine digitale Interface-Karte ersetzt. Alle Kameraparameter, wie z.B. Shutter-Zeit, Farbtemperatur und Verstärkung können über das digitale Interface gesteuert werden. Die Kameras haben eine Bildauflösung von 640*480 Pixeln. Durch die begrenzte Datenrate des FireWire-Busses kann bei dieser Auflösung eine Datenrate von 15 Bildern pro Sekunde und Kamera erreicht werden. Die Farbdaten werden im YUV411 Farbformat [YuvColor] übertragen, die Farbauflösung ist daher um den Faktor vier geringer als die Bildauflösung der Kamera.

In beiden Fällen wird die Bildfrequenz von den Kameras vorgegeben. Alle aufgenommenen Bilder werden im Hauptspeicher abgelegt und als reserviert gekennzeichnet, sofern sie noch nicht von der nachfolgenden Bildverarbeitung ausgewertet wurden. Die Zahl der Bildspeicher im Hauptspeicher ist sinnvollerweise begrenzt. Sobald alle Bildspeicher reserviert sind, wird die Bildaufnahme gestoppt und zu einem späteren Zeitpunkt wieder fortgesetzt. Solange die nachfolgende Bildverarbeitung und Datenauswertung echtzeitfähig ist, werden immer genug Bildspeicher zur Verfügung stehen und die überzähligen Bildspeicher können zur Überbrükkung von kurzfristigen Spitzen der Rechnerlast genutzt werden.

Als Lichtquellen kommen zur Projektion des Punktes und der Linie handelsübliche Laser zum Einsatz. Zur Projektion des Punktes reicht ein preiswerter batteriebetriebener 1mW Laserpointer. Zur Projektion der Linie wird ein 5mW Laser der Firma Lasiris verwendet. Die Linie wird durch Aufweitung des Laserstrahls mit Hilfe einer Diffraktionsoptik erreicht, die ein gleichmäßiges Intensitätsprofil über die gesamte Linie garantiert.

Für das Gray-Code Verfahren wird ein Halogenprojektor des Typs ABW 320 mit den Optionen CROSS und SWITCH eingesetzt. Der Projektor verfügt in der Basisausstattung über einen programmierbaren Linien-LCD mit 320 programmierbaren Linien und eine 150 Watt Halogenlampe. Über einen Microcontroller mit Speicher können bis zu hundert Linienmuster programmiert und abgerufen werden. Die Option CROSS erlaubt über einen zweiten LCD die Projektion von weiteren Linienmustern, die um 90 Grad gedreht sind.

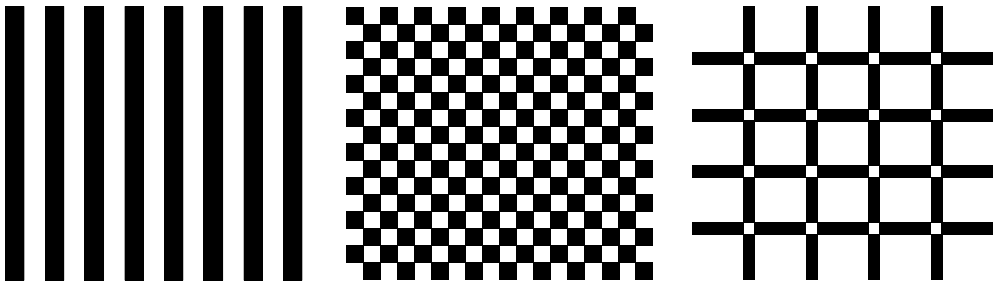


Abbildung 3.3: Projektionsmuster des Linienprojektors ABW 320

Die Option SWITCH beinhaltet eine stärkere 250 Watt Halogenlampe und die Möglichkeit, die Lampe vom Steuerrechner aus ein- oder auszuschalten sowie die Regelung der Lampenhellig-

keit. Bei der Verwendung eines einfachen Gray-Codes können bis zu 320 Linien mit 9-Bit Codes versehen werden. Alternativ kann der Gray-Code in 8 Bit ausgeführt werden. Die feinste Liniensequenz wird dann zur Projektion von vier Phasenbildern zur subpixelgenauen Bildkodierung verwendet. Besonders durch die Option CROSS können aber auch andere Projektionsmuster, wie Schachbrett oder Gitternetz erzeugt werden.

Zusätzlich zum Halogenprojektor wurde die Einsetzbarkeit eines handelsüblichen Videobeamers des Typs Philips Hopper XG20 untersucht. Der Beamer hat eine Pixelauflösung von 1024*768 Farbpixeln und wird wie ein zweiter Monitor angesteuert, so dass der Rechner mit einer entsprechenden Dual-Head Grafikkarte ausgestattet sein muss. Zum Einsatz kam dafür die Grafikkarte SYNERGY III der Firma Elsa. Der Beamer ist lichtstärker und flexibler als der Halogenprojektor, da beliebige Muster und Strukturen im Rahmen der gegebenen Auflösung projiziert werden können. Das System aus Grafikkarte und Beamer ist allerdings nicht für den technischen Einsatz zur dreidimensionalen Objektvermessung ausgelegt. So existiert keine Kontrolle über die Lichtleistung der Lampe des Beamers, der Beamer kann nicht zwischen zwei Messungen abgeschaltet werden. Bei einer dadurch drohenden Überhitzung schaltet sich der Beamer automatisch ab, was die 3D-Messung zerstört. Im Weiteren kann auch das eingesetzte Betriebssystem (Windows NT) nicht garantieren, dass während der Messung kein Fenster auf dem zweiten Monitor geöffnet wird, was ebenfalls zu Fehlern in der Messung führen kann. Im Einzelnen ist also zu überlegen, welches Projektionssystem sich am besten eignet.

3.1.3 Bildverarbeitung

Ein wesentliches Ziel von *HandyScan3D* war die hohe Interaktion des Benutzers mit dem Rechner. Der Scan-Vorgang und die 3D-Visualisierung sollen in Videoechtzeit durchgeführt werden. Das heißt, dass alle aufgenommenen Daten direkt dreidimensional und trianguliert visualisiert werden, um den Benutzer ein direktes Feedback über die aktuell aufgenommenen Daten zu geben. Dafür ist aber für alle Einzelkomponenten der Verarbeitungssoftware eine weitgehende algorithmische Optimierung notwendig.

Die Aufgabe der Bildverarbeitung besteht dabei in der Reduktion der Bilddaten auf die wesentliche Information, der Extraktion der aufprojizierten optischen Signale aus den Kamerabildern. Im einfachsten Fall, bei der Verwendung eines Laserstrahlers, besteht diese Aufgabe in der Extraktion eines einzelnen Bildpunktes. Bei der Verwendung der Laserlinie ist das die Extraktion der Lichtlinie. Beim Einsatz des Gray-Code Systems wird über die Bildverarbeitung jedem Bildpunkt ein Gray-Code zugeordnet.

Obwohl die Zuordnung von korrespondierenden Punkten in den Stereobildpaaren sinngemäß nicht Bestandteil der Bildverarbeitung ist, wird sie dennoch in diesem Kapitel behandelt, da sie algorithmisch sehr stark an die Bildverarbeitung gekoppelt ist.
Im Folgenden werden die einzelnen Verfahren der Bildverarbeitung für den Fall des Punktes, der Linie und des Gray-Codes genauer erläutert. Dabei finden insbesondere die Optimierungen zur Erlangung der Echtzeitfähigkeit Beachtung.

3.1.3.1 Extraktion eines Punktes

Die Extraktion eines einzelnen Punktes aus einem Kamerabild ist dann besonders einfach, wenn sich der gesuchte Punkt als hellstes Pixel im Bild darstellt. Soll das System unter normalen Umgebungslichtbedingungen arbeiten, ist diese Forderung aber im Allgemeinen nicht zu erfüllen.

Für die Entwicklung von *HandyScan3D* kann ausgenutzt werden, dass die Szene während der Vermessung nicht verändert werden darf. Die einzige veränderliche Größe während des Scan-Vorgangs ist der reflektierte Laserpunkt. Die Aufnahme eines Referenzbildes ohne Laserpunkt vor dem eigentlichen Scan-Vorgang ermöglicht die Bildung von Differenzbildern während der Messung, die dann nur noch den Laserpunkt als einziges markantes Signal enthalten

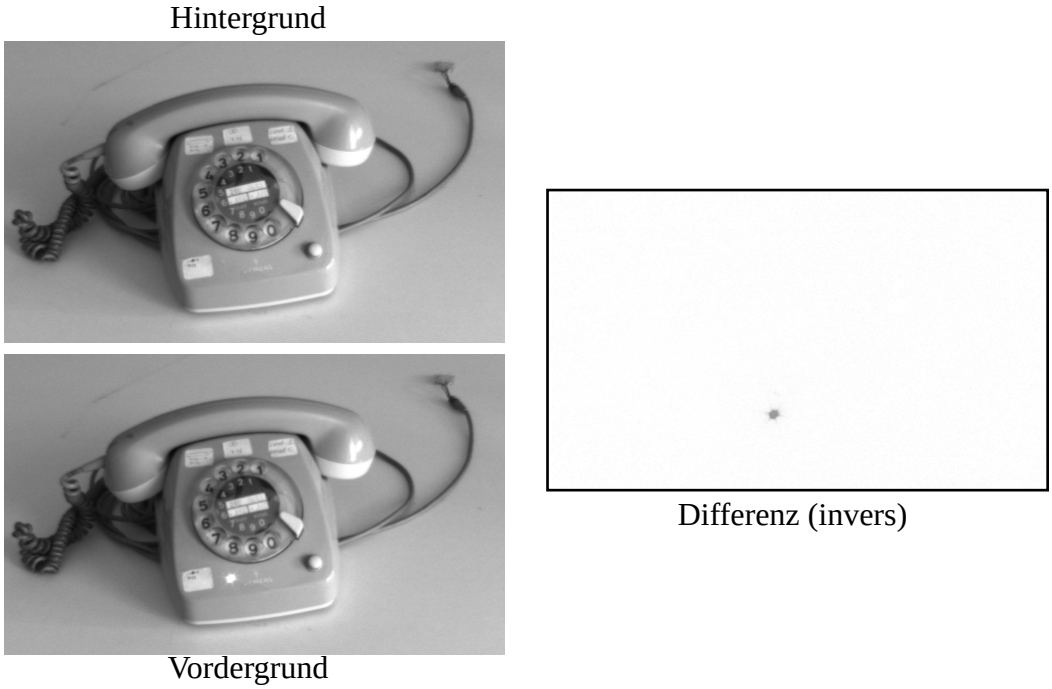


Abbildung 3.4: Differenzbildverfahren

Mit einem einfachen Schwellwertverfahren kann das Differenzbild nach dem Laserpunkt durchsucht werden. Die Schwelle wird dabei sinnvollerweise so gewählt, dass das Grundrauschen der Kamera ausgeblendet wird. Das hellste Pixel markiert den gefundenen Laserpunkt.

Zur Berechnung einer subpixelgenauen Bildkoordinate des Laserpunktes wird die Umgebung des hellsten Pixels nach weiteren Pixeln durchsucht, die ebenfalls zum Laserpunkt gehören. Dazu wird ein zweiter Schwellwert unterhalb der Intensität des hellsten Pixels festgelegt, der den Suchraum begrenzt. Über alle gefundenen Pixel wird der Schwerpunkt gebildet.

Die komplette Berechnung eines Differenzbildes, sowie das vollständige Absuchen des Differenzbildes ist zu rechenaufwändig. Daher werden die folgenden zwei Annahmen getroffen:

- Der Laserpunkt überdeckt im Bild wenigstens zwei Pixel.
- In zwei zeitlich aufeinander folgenden Bildern finden sich die Laserpunkte etwa an der gleichen Stelle im Bild.

Die erste Annahme stellt zwar eine Einschränkung dar, da Laserpunkte, die in großen Entfernungen oder auf dunklem Untergrund aufgenommen werden, sich tatsächlich nur in einem Pixel darstellen können. Derartige Bildpunkte sind aber wenig vertrauenswürdig und eignen sich auch nicht zur Schwerpunktberechnung. Aufgrund dieser Annahme ist es aber möglich, die Suche im Bild signifikant zu beschleunigen, da es reicht, nur noch jedes zweite Pixel in jeder zweiten Bildzeile zu durchsuchen.

Die zweite Annahme rührt aus der Vorgehensweise her, wie der Scan-Vorgang durchgeführt wird. Ein Laserstrahler wird per Hand über das Messobjekt geschwenkt. Bei einer ausreichend hohen Abtastrate werden nicht mehr unabhängige Laserpunkte aufgenommen, sondern Punkte entlang einer Bewegungsbahn. Durch diese Lokalitätsannahme kann der Suchbereich deutlich eingeschränkt werden, wenn im vorangegangenen Bild ein Laserpunkt gefunden wurde. Wenn die lokale Suche nicht zum Erfolg geführt hat, wird beim folgenden Bild wieder eine Suche über das gesamte Bild durchgeführt.

Die Abbildung 3.5 verdeutlicht noch einmal diese Vorgehensweise.

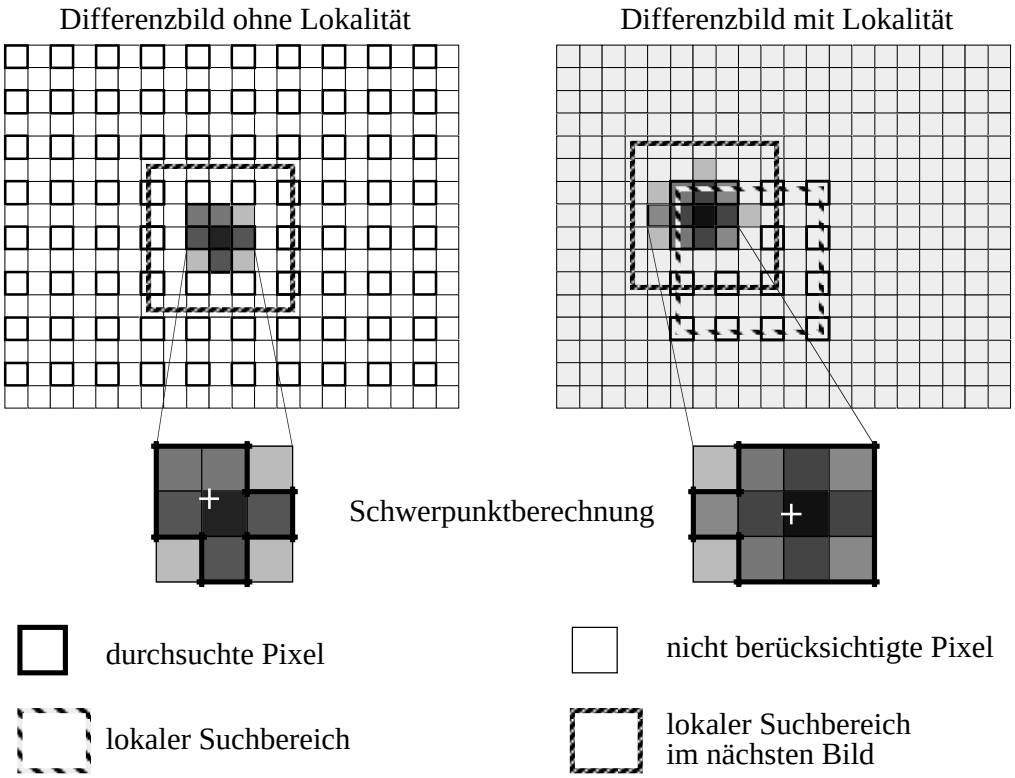


Abbildung 3.5: Laserpunktsuche

Daraus ergeben sich im zusammengefasst die folgenden vier Schritte:

1. Solange kein Laserpunkt gefunden wurde, wird jedes vierte Pixel des Differenzbildes abgesucht.
2. Wenn ein Laserpunkt gefunden wurde, wird die lokale Umgebung nach weiteren Punkten durchsucht.

3. Über alle gefunden Punkte wird der Schwerpunkt gebildet.
4. Der berechnete Schwerpunkt ist die Mitte des lokalen Suchbereichs für das nächste Bild.

Die Bilder beider Stereokameras werden in dieser Art bearbeitet. Da jedes Bild mit einem Zeitstempel versehen ist, können gefundene Laserpunkte mit gleichem Zeitstempel zu Korrespondenzen gekoppelt werden und dienen damit im Weiteren als Eingabe für die Stereokalibrierung und die 3D-Rekonstruktion.

3.1.3.2 Extraktion einer Linie

Im Gegensatz zum Laserpunkt können mit einem zur Ebene aufgeweiteten Laserstrahl keine eindeutigen Punkte mehr markiert werden. Auch wenn die Laserlinie in den Stereobildern sicher erkannt wird, können ohne zusätzliches Wissen keine Korrespondenzen gebildet werden. Daher muss vor dem Einsatz der Linie das Stereosystem kalibriert werden. Das kann beispielsweise durch den Laserpunkt erfolgen. Über die Epipolargeometrie ist es dann möglich, für jeden Punkt der Laserlinie in einem Bild eine Epipolarlinie im zweiten Bild zu berechnen. Der Schnittpunkt dieser Linie mit der Laserlinie im zweiten Bild markiert dann den gesuchten korrespondierenden Punkt.

Es ist unmittelbar einzusehen, dass dieses Verfahren nur funktioniert, wenn die Epipolarlinie und die Laserlinie im zweiten Bild in erster Näherung nicht parallel sind. Anderenfalls besteht die Gefahr, dass die gesamte Laserlinie im zweiten Bild als Korrespondenz für jeden Punkt der Linie im ersten Bild erkannt wird. Dieses Problem wirkt sich zum einen auf die Art aus, wie das Objekt mit der Laserlinie abgescannt werden kann, zum anderen hat es Einfluß auf die Algorithmen zur Detektion der Laserlinie im Bild. Zum besseren Verständnis dieser Problematik wird daher zunächst der Begriff der Stereoorientierung eingeführt:

Stereoorientierung eines Bildes = mittlere Orientierung aller Epipolarlinien im Bild.

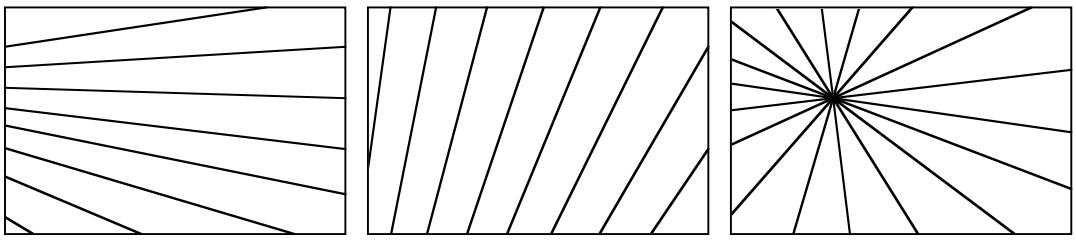


Abbildung 3.6: Stereoorientierungen

Die Stereoorientierung beschreibt die grobe Lage einer Stereokamera relativ zur zweiten Kamera. Steht die zweite Stereokamera neben der ersten Kamera, so verlaufen die Epipolarlinien im Mittel horizontal. Steht die zweite Kamera über oder unter der ersten Kamera, so verlaufen die Epipolarlinien im Mittel vertikal. Der Fall, dass die beiden Kameras hintereinander stehen, führt zu keiner eindeutigen Stereoorientierung[1] (Abbildung 3.6).

1. Dieser Fall bedeutet aber auch, dass sich eine Kamera im Bildfeld der anderen Kamera befindet. Da diese Konfiguration für HandyScan3D nicht sinnvoll ist, wird sie nicht weiter betrachtet.

Bei der Durchführung der Objektvermessung mit der Laserlinie ist die Stereoorientierung in soweit zu beachten, als dass ein möglichst großer Winkel zwischen der Stereoorientierung und den Laserlinien eingehalten werden muss. Das heißt, dass die Linie senkrecht projiziert werden muss, wenn die Kameras nebeneinander stehen, bzw. horizontal, wenn die Kameras übereinander stehen.

Das Verfahren zur Korrespondenzbestimmung beim Einsatz einer Laserlinie unterscheidet bei den Stereobildern zwischen einer Master- und einer Slave-Kamera. In den Bildern der Master-Kamera wird als erstes nach der Laserlinie gesucht. Für alle gefundenen Punkte der Laserline werden dann die Korrespondenzen im Bild der Slave-Kamera gesucht. Dabei unterscheiden sich die Algorithmen im Master und im Slave-Bild. Wie bei der Suche nach dem Laserpunkt kommt aber wieder die Differenzbildung zum Einsatz.

Das Bild der Master-Kamera wird zunächst entsprechend der Stereoorientierung durchsucht. Bei horizontaler Stereoorientierung wird zeilenweise vorgegangen, bei vertikaler Orientierung spaltenweise. Dabei gelten vereinfachend alle Orientierungen mit einem Winkel von weniger als 45 Grad als horizontal, alle Orientierungen mit einem Winkel größer 45 Grad als vertikal. In jeder durchsuchten Pixelreihe wird das hellste Pixel oberhalb eines definierten Schwellwertes bestimmt. Danach werden die benachbarten Pixel, die ebenfalls zur Laserlinie gehören, entsprechend eines zweiten Schwellwertes unterhalb der Intensität des hellsten Pixels bestimmt. Der subpixelgenaue Schnittpunkt der Laserlinie mit der Pixelreihe wird dann mittels Schwerpunktbildung berechnet.

Für jeden gefundenen Schnittpunkt zwischen Laserlinie und Pixelreihe wird eine Korrespondenz im Slave-Bild gesucht. Da diese Operationen nur mit Hilfe der Epipolargeometrie einfach gelöst werden kann, wird zunächst für jeden Schnittpunkt die Epipolarlinie im Slave-Bild berechnet. Entlang der Epipolarlinie werden dann die Pixel im Slave-Bild nach Pixeln durchsucht, die zur Laserlinie gehören. Das Verfahren funktioniert prinzipiell genau so, wie im Master-Bild. Zunächst wird das hellste Pixel oberhalb eines Schwellwertes bestimmt, dann wird die lokale Umgebung durchsucht, und eine Schwerpunktbildung führt zum gewünschten Korrespondenzpunkt. Es ist allerdings zu beachten, dass die Epipolargeometrie im normierten Bildkoordinatensystem definiert ist, während die Bildoperationen im Pixelkoordinatensystem durchgeführt werden. Punkte der Laserlinie im Master-Bild müssen daher zunächst ins normierte Bildkoordinatensystem transformiert werden, damit dann die Epipolarlinie berechnet werden kann. Die Transformation der Epipolarlinie zurück ins Pixelkoordinatensystem der Slave-Kamera stellt sich allerdings als nicht mehr trivial dar, da mögliche Kissenverzerrungen in der Optik der Slave-Kamera zu Krümmungen der Epipolarlinie im Pixelkoordinatensystem führen. Eine pixelweise Transformation der Epipolarlinie ins Pixelkoordinatensystem der Slave-Kamera ist rechnerisch sehr aufwändig. Daher wird eine Approximation berechnet, bei der die Epipolarlinie in Teilstücke aufgeteilt wird. Für die Endpunkte der Teilstücke wird jeweils die Pixelkoordinate berechnet. Die Teilstücke werden dann mit dem Bresenham-Algorithmus [Foley92] im Bild der Slave-Kamera rekonstruiert.

Zur weiteren Optimierung der Bildauswertung werden wieder Lokalitätsannahmen gemacht, die im Gegensatz zur Detektion des Laserpunktes nicht mehr auf Lokalität in Bildsequenzen beruhen, sondern die Linieneigenschaft des Lasers selber nutzen. Wenn im Bild der Master-Kamera ein Punkt der Laserlinie in einer Pixelreihe gefunden wurde, dann ist in der nächsten

Pixelreihe an gleicher Stelle ebenfalls ein Punkt der Laserlinie zu erwarten. Solange die Laserlinie im Bild stetig verläuft, kann der Suchbereich auf wenige Pixel eingeschränkt werden. Erst wenn die Linie im Bild unterbrochen ist, wird der Suchbereich wieder auf die gesamte Pixelreihe ausgeweitet. Im Bild der Slave-Kamera gilt grundsätzlich das gleiche, nur dass als erstes in dem Linienstück der Epipolarlinie gesucht wird, das auch bei der letzten Korrespondenz den Treffer geliefert hat.

Genau wie bei der Suche nach dem Laserpunkt wird eine weitere Geschwindigkeitsoptimierung durch die Unterabtastung des Bildes erreicht. Das heißt, sofern kein Vorwissen über die Laserlinie im aktuellen Kamerabild vorliegt, wird nur jedes zweite Pixel in jeder zweiten Reihe abgesucht. Bei einer angenommenen Dicke der Linie von wenigstens zwei Pixeln im Bild kann so eine deutliche Beschleunigung des Algorithmus erreicht werden, ohne dass Daten verloren gehen.

In Abbildung 3.7 wird noch einmal die Bildverarbeitung des Linienalgorithmus aufgezeigt. Die Auswertung des Differenzbildes der Slave-Kamera wird exemplarisch für nur einen Schwerpunkt der Laserlinie im Differenzbild der Master-Kamera und damit für nur eine Epipolarlinie aufgezeigt.

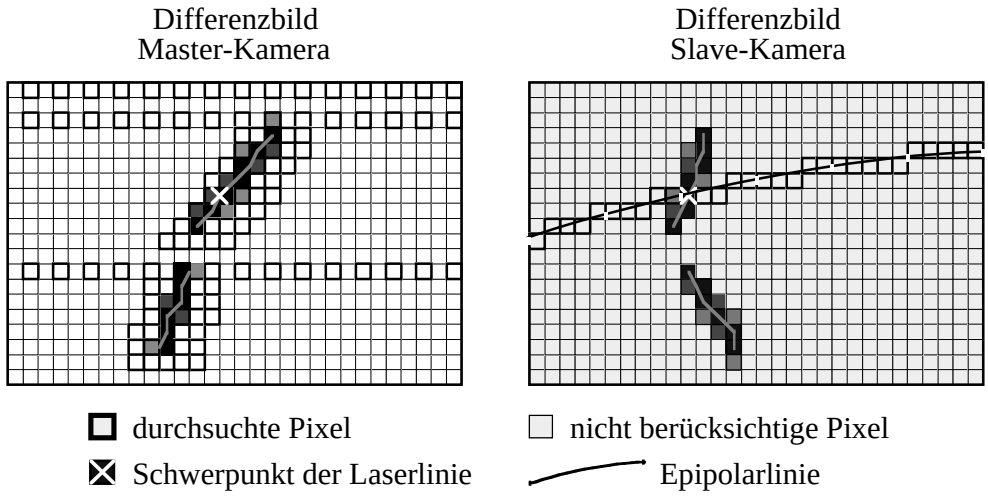


Abbildung 3.7: Korrespondenzsuche beim Einsatz eines Linienlasers

Die einzelnen Schritte die dafür notwendig sind, sind im Einzelnen:

1. Entsprechend der Stereoorientierung wird das Bild der Master-Kamera reihenweise durchsucht. Wenn in der vorherigen Reihe ein Punkt der Laserlinie gefunden wurde, wird lokal gesucht, sonst in der ganzen Reihe.
2. Über alle Pixel der Laserlinie in einer Pixelreihe wird der Schwerpunkt berechnet und dann ins normierte Koordinatensystem transformiert.
3. Für jeden gefundenen Schwerpunkt wird die Epipolarlinie im Bild der Slave-Kamera berechnet.

4. Die Epipolarlinie wird in Teilstücke gleicher Länge aufgeteilt. Jedes Teilstück wird nach Bresenham in das Pixelkoordinatensystem der Slave-Kamera transformiert. Wurde auf der letzten Epipolarlinie ein Punkt der Laserlinie gefunden, so wird nur das entsprechende Teilstück durchsucht, sonst werden alle Teilstücke durchsucht. Über alle gefunden Pixel der Laserlinie wird wieder der Schwerpunkt berechnet.

Die gefundenen Korrespondenzen dienen als Eingabe der 3D-Rekonstruktion. Es sei hier aber noch angemerkt, dass die Extraktion der Laserlinie zwei potentielle Schwächen hat. Wenn der Verbindungsvektor zwischen den beiden Kameras parallel zur Ebene verläuft, die vom Linienlaser aufgespannt wird, degeneriert die Auswertung der Epipolarlinie im Bild der Slave-Kamera. Zum einen werden viele Punkte der Laserlinie im Bild der Master-Kamera auf die selbe Epipolarlinie abgebildet. Zum anderen verlaufen die Epipolarlinien im Bild der Slave-Kamera parallel zur Laserlinie, so dass entweder keine Korrespondenz gefunden wird, oder der Schnitt der Epipolarlinie verläuft in Längsrichtung durch die Laserlinie und der berechnete Schwerpunkt ist nicht mehr zu gebrauchen. Idealerweise sollte die Laserlinie also vorzugsweise orthogonal zur Stereoorientierung auf das Objekt projiziert werden. Die zweite potentielle Schwäche zeigt sich, wenn sich die Richtung, von der aus die Laserlinie auf das Objekt projiziert wird, deutlich von den Richtungen unterscheidet, in der die Kameras das Objekt aufnehmen. Das ist beispielsweise dann der Fall, wenn die Kameras von vorne auf das Objekt gerichtet sind und die Laserlinie von oben auf das Objekt projiziert wird. Dann kann es bei Sprüngen oder Hinterschneidungen auf dem Objekt dazu kommen, dass die Epipolarlinie mehr als einen Schnittpunkt mit der Laserlinie im Bild der Slave-Kamera hat. Welcher Schnittpunkt korrekt ist, kann im Allgemeinen nicht entschieden werden, so dass es schnell zu Fehlmessungen kommen kann.

In der Praxis stellen beide potentielle Schwächen kaum ein Problem dar, da sie durch ausreichende Sorgfalt beim Führen der Linie vermieden werden können. Zusätzlich sind beide Fälle detektierbar. Im ersten Fall ist die Breite des Querschnitts durch die Laserlinie ein Maß für die Orientierung des Lasers. Ist der Schnitt der Epipolarlinie durch die Laserlinie zu breit, kann die Korrespondenz verworfen werden. Die Zahl der Schnittpunkte zwischen Epipolarlinie und Laserlinie ist ein einfaches Kriterium zur Vermeidung des zweiten Problems.

3.1.3.3 Extraktion des Gray-Codes

Anders als beim Punktlaser- und Linienlaserverfahren, bei denen das Objekt sukzessive mit dem Laser überstrichen werden muss, um eine 3D-Ansicht zu erzeugen, bietet das Gray-Code Verfahren die Möglichkeit, mit einer kurzen Sequenz von Projektionsmustern innerhalb weniger Sekunden eine 3D-Ansicht zu generieren. Zudem werden durch den Einsatz von Halogenprojektoren die unerwünschten Speckles[1] von Laserlicht vermieden.

Mit einer einfachen Gray-Code Sequenz werden lediglich Linien codiert. Zur stereoskopischen Auswertung der Gray-Code Sequenz muss in diesem Fall, genau wie beim Einsatz der Laserlinie, die Epipolargeometrie bekannt sein. Durch die Projektion von zwei Gray-Code Sequenzen, die um 90 Grad zueinander verdreht sind, können Punkte codiert werden. Die Zuordnung von Pixeln gleichen Codes in beiden Bildern führt zu eindeutigen Korrespondenzen, so dass

1. Lichtstreuungen bei Laserlicht aufgrund von Interferenzen

Kalibrierung und 3D-Rekonstruktion aus ein und demselben Datensatz berechnet werden können.

Der eingesetzte Projektor des Typs ABW 320 erlaubt die Codierung von 320 horizontalen und vertikalen Linien. Damit lassen sich bis zu 102.400 Punkte codieren. Durch den zusätzlichen Einsatz des Phase-Shift Verfahrens und der interpolierenden Zuordnung von Punkten ist die Zahl der möglichen Korrespondenzen nur noch durch die Kameraauflösung begrenzt.
In beiden Fällen werden zunächst zwei Referenzbilder erzeugt. Beim ersten Bild sind alle Linien undurchsichtig geschaltet, beim zweiten Bild sind alle Linien durchsichtig. Die mittlere Helligkeit der gleichen Pixel in beiden Referenzbildern dient als Schwellwert zur Unterscheidung von Hell- und Dunkelcodes in der nachfolgenden Gray-Code Sequenz. Sofern der Unterschied zwischen Hellreferenz und Dunkelreferenz nicht ausreichend hoch ist, werden die Pixel bei der nachfolgenden Gray-Code Auswertung ausmaskiert.

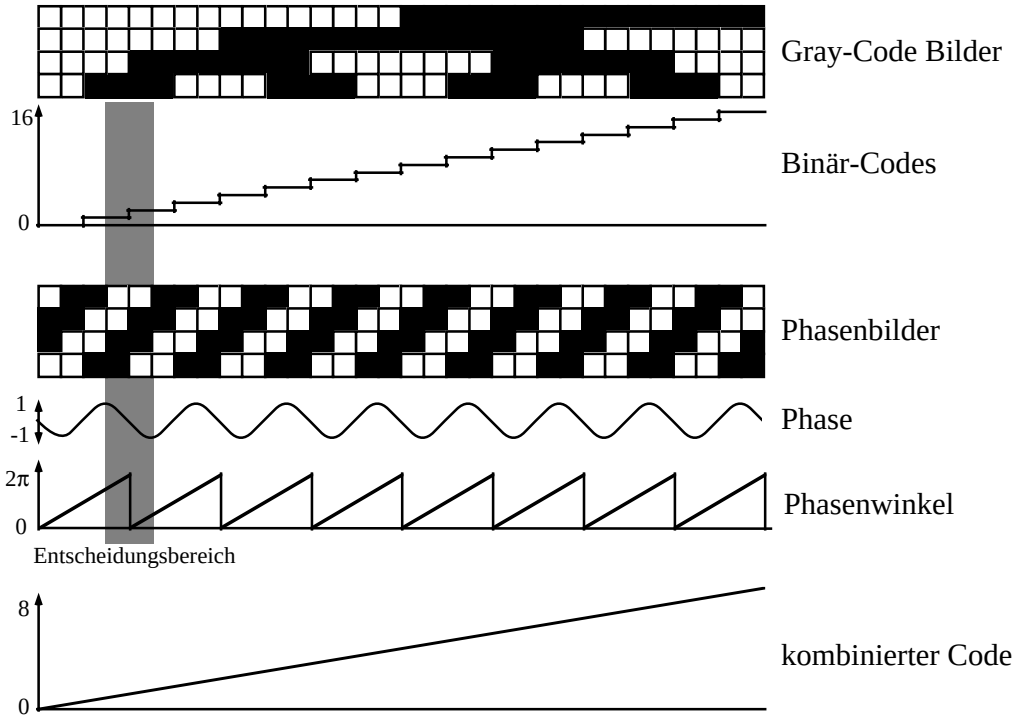


Abbildung 3.8: Auswertung von Gray-Code und Phase

Der einfache GrayCode besteht aus einer Sequenz von acht Linienbildern, mit denen 256 binär codierte Linien unterschieden werden können. Zur deutlichen Genauigkeitssteigerung erfolgt anschließend die Projektion von jeweils vier feinen Linienmustern, die abwechselnd aus zwei hellen und zwei dunklen Linien bestehen und um jeweils eine Zeile zueinander verschoben sind. Idealerweise müssten sinusförmige Profile projiziert werden, da sich die Linien des Projektors jedoch nur binär steuern lassen, wird das Objektiv des Projektors leicht unscharf gestellt, um einen gleichartigen Effekt zu erzielen. Bei der Auswertung der Phasenbildern wird der Grauwert jedes Kamerapixels ausgewertet. Im Verhältnis zur Hell- und Dunkelreferenz kann dann eine relativer Phasenwert mit einem Wertebereich von -1 bis +1 berechnet werden. Ein Phasenbild alleine wäre zu rauschanfällig, als dass die Auswertung der Phasenwerte zu brauchbaren Ergebnissen führen würde. Im realisierten Verfahren werden vier Phasenbilder

projiziert, die um jeweils 90 Grad zueinander verschoben sind. Aus den vier Intensitäten wird mit [Osten91]

$$\varphi = \operatorname{atan}\left(\frac{I_2 - I_4}{I_3 - I_1}\right) \tag{3.15}$$

ein Phasenwinkel berechnet. Bei der Vereinigung der Binär-Codes der Linien mit den Phasenwinkeln ist zu beachten, dass eine komplette Phase zwei Schritten beim Binär-Code entspricht. An den Unstetigkeitsstellen der Phasenwinkel muss im Einzelnen entschieden werden, welcher Binär-Code mit dem Phasenwinkel verrechnet werden muss. Ungerade Linien-Codes können nur auf Phasenwinkel im Bereich von 0 bis π Fallen. Gerade Linien-Codes fallen dementsprechend nur auf Phasenwinkel im Bereich von π bis 2π. Ist diese Regel verletzt, dann wird von einer Fehlzuordnung ausgegangen. Der genauere Phasenwinkel gibt im Entscheidungsbereich vor, wie der Linien-Code korrigiert wird (Abbildung 3.8) um die Stetigkeit wieder herzustellen. So wird beispielsweise aus einem Binär-Code von 2 und einem Phasenwinkel von $1{,}8\pi$ ein kombinierter Code mit dem Wert 0,9. Der erreichbare Wertebereich der kombinierten Codes spannt zwar nur noch einen Bereich von 0 bis 127 auf, durch die Phasenauswertung werden aber Genauigkeiten erreicht, die dem 10- bis 20-fachen dessen entsprechen, was nur mit den Gray-Codes erreichbar ist.

Mit dem bisher beschriebenen Gray-Code Verfahren lassen sich in einem Bild nur Linien codieren, die Stereokalibrierung erfordert aber korrespondierende Bildpunkte in beiden Bildern. Durch die Projektion einer zweiten, um 90 Grad verdrehten Linien-Sequenz, kann jedem Pixel ein zweistelliger Wert, bestehend aus horizontalem und vertikalem Code, zugeordnet werden. Die Bildung von Korrespondenzen mit gleichen Codes in beiden Bilder ist aber nach wie vor nicht trivial, da im Allgemeinen keine zwei Pixel den selben Wert erhalten. Das gilt sowohl in einem Bild, als auch im Vergleich von zwei Stereobildern. Eine einfache Suche nach Pixeln gleicher Werte in Stereobildpaaren wird also kaum zu Treffern führen. Daher werden die Werte der Pixel im Bild der Master-Kamera skaliert in ganzzahlige Codes umgewandelt. Die Skalierung wird so gewählt, dass 10 bis 20 benachbarte Pixel den gleichen Code erhalten. Der neue, „ungenaue“ Pixel-Code wird zusammen mit der Bildkoordinate, die den Schwerpunkt aller Pixel gleichen Codes bildet in einer Hash-Tabelle abgelegt. Bei der Auswertung des Bildes der Slave-Kamera wird nun für jedes Pixel in gleicher Weise ein „ungenauer“ Code berechnet. Über die Hash-Tabelle kann dann eine „grobe“ Korrespondenz gebildet werden. Eine lokale Suche im Bild der Master-Kamera führt anschließend zu den vier Pixeln, deren genauer Wert am besten mit dem Wert des Pixels der Slave-Kamera übereinstimmt. Zwischen diesen vier Pixeln wird schließlich mit linearer Interpolation eine exakte Korrespondenz bestimmt (Abbildung 3.9).

Reale Versuche haben gezeigt, dass so für über 90 Prozent aller Pixel, für die eine Korrespondenz existiert, auch eine Korrespondenz gefunden wird. Bei Pixeln, für die keine Korrespondenz gefunden wurde, lag das Problem zumeist darin, dass keine plausible Interpolation berechnet werden konnte. Das ist dann der Fall, wenn der gesuchte Wert nicht in dem Bereich liegt, der von den vier gefundenen Pixeln aufgespannt wird, also bestenfalls eine Extrapolation berechnet werden kann. Das Weglassen dieser Pixel fällt im Allgemeinen nicht auf, während hingegen eine Extrapolation leicht eine Störung in der Rekonstruktion verursachen kann, die das Ergebnis in Hinsicht auf seine Genauigkeit verschlechtert. Aufwändigere Verfahren zur

Korrespondenzpunktberechnung wurden nicht in Betracht gezogen, da der kaum vorhandene Nutzen den nötigen Aufwand nicht mehr rechtfertigt.

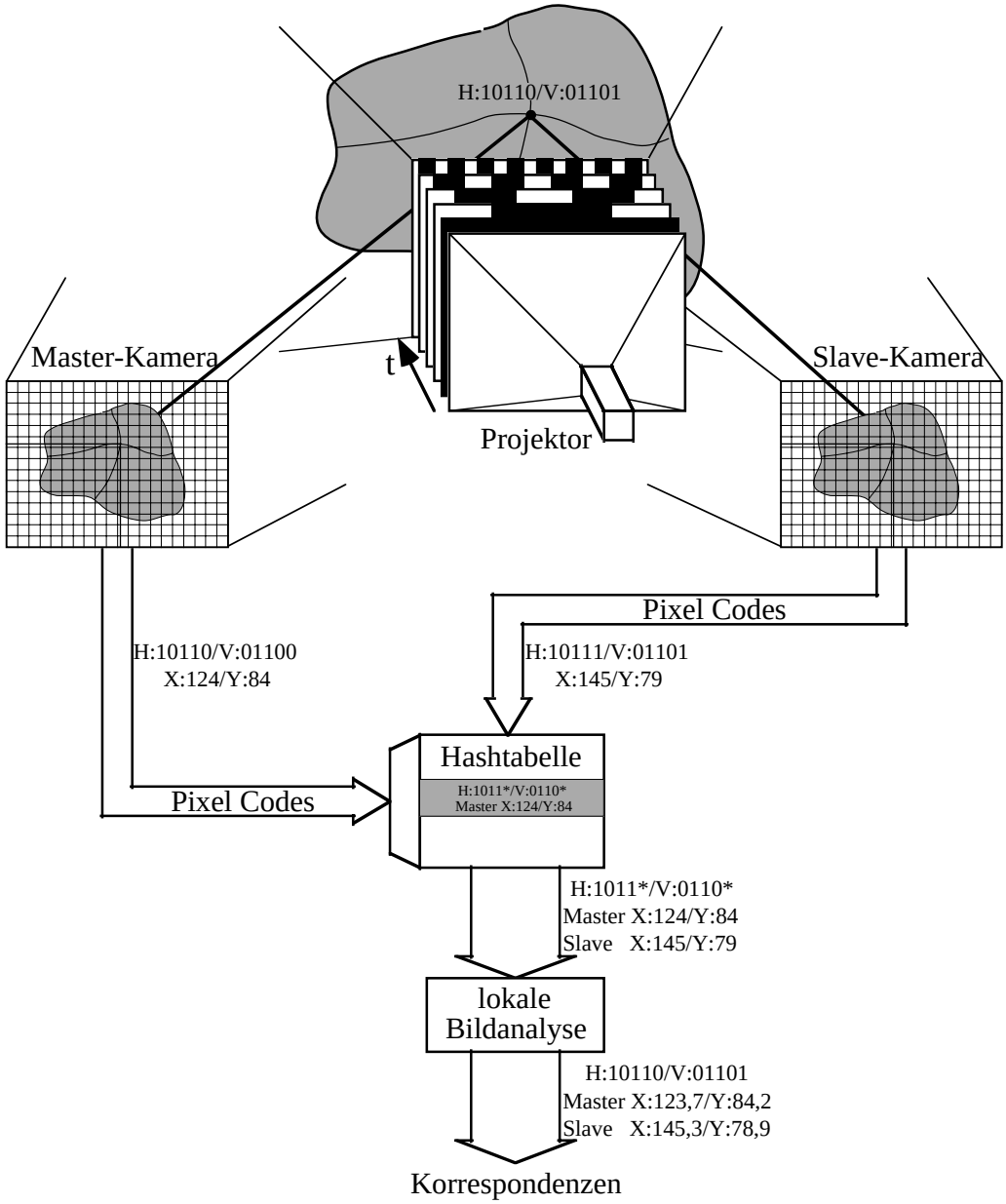


Abbildung 3.9: Bildverarbeitung beim Gray-Code Verfahren

In Kürze besteht das kombinierte Verfahren aus den folgenden Schritten:

1. Projektion und Aufnahme der Hell- und Dunkelreferenz.
2. Projektion der horizontalen Gray-Code Sequenz und der vier horizontalen Phasenbildern. Auswertung des Gray-Codes und Berechnung der Phasenwinkel aus den Phasenbildern.
3. Wie 2., für die vertikale Sequenz.
4. Berechnung der zweistelligen Werte für jedes Pixel in beiden Stereobildern.

5. Für das erste Bild werden „ungenaue“ Codes zusammen mit den zugehörigen Pixelkoordinaten in eine Hash-Tabelle eingetragen.
6. Für jedes Pixel des zweiten Bildes wird über die Hash-Tabelle ein Pixel ähnlichen Codes im ersten Bild gesucht.
7. Durch lokale Bildanalyse und Interpolation werden subpixelgenaue Koordinaten im Bild der Master-Kamera berechnet, die zusammen mit den Pixeln des zweiten Bildes die gesuchten Korrespondenzen ergeben.

Im Gegensatz zur Detektion der Laserlinie können die Korrespondenzen aus der Laserpunktdetektion und aus dem Gray-Code Verfahren zur Stereokalibrierung verwendet werden. Die Verfahren zur Bestimmung einer optimalen Kalibrierung aus diesen Korrespondenzen sind Inhalt des folgenden Kapitels.

3.1.4 Stereokalibrierung

Die Stereokalibrierung dient zur Berechnung der relativen Lage zweier Kameras zueinander. Das eingesetzte Verfahren ist das aus der passiven Stereoskopie bekannte erweiterte Acht-Punkte Verfahren, dessen primäres Ergebnis die Epipolarmatrix der beiden Kameras ist. Zur späteren dreidimensionalen Rekonstruktion der Daten können aus dieser Matrix ein Translationsvektor und eine Rotationsmatrix berechnet werden, die die Lage einer Kamera im Koordinatensystem der anderen Kamera beschreiben. Voraussetzung für die Berechnung der Epipolarmatrix nach dem Acht-Punkte Verfahren sind wenigstens acht korrespondierende Bildpunktepaare. Bei mehr als acht Korrespondenzen ist das System überbestimmt und kann mit einer Least-Square Berechnung gelöst werden. In jedem Fall ist aber eine sorgfältige Auswahl von „guten“ Korrespondenzen sehr wichtig für die Berechnung einer genauen Epipolargeometrie.

Unabhängig von der Quelle der Korrespondenzen, dem Punktverfahren oder dem Gray-Code Verfahren muss davon ausgegangen werden, dass alle Korrespondenzen mit mehr oder weniger großen stochastischen und systematischen Fehlern behaftet sind. Primäre Ursache für diese Fehler ist das Rauschen der CCD-Pixel der verwendeten Kameras. Bei den laserbasierten Verfahren kommen noch Speckle-Effekte, also seitliche Lichtstreuungen als Fehlerquelle hinzu. Sie verfälschen die Schwerpunktbildung bei der Detektion des Laserpunktes im Kamerabild. Aber auch die Objektoberfläche des gescannten Objektes kann zu signifikanten Fehlern bei der Berechnung der Korrespondenzen führen. Im einfachen Fall führt eine starke Texturierung der Objektoberfläche zu einem geringfügigen Fehler bei der Schwerpunktberechnung der laserbasierten Verfahren, da der Laserpunkt nicht mehr gleichmäßig vom Objekt reflektiert wird. Beim kombinierten Gray-Code Verfahren führt die Texturierung zu einer Verschiebung in der Phase. Aber auch dieser Fehler ist als relativ geringfügig zu betrachten. Fatale Auswirkungen können reflektierende Objektoberflächen haben, bei denen das projizierte Licht auf der Oberfläche gespiegelt wird und mehrere Reflexionen ein und desselben Lichtsignals im Kamerabild gemessen werden können.

Bei den laserbasierten Verfahren lassen sich drei Fälle von Korrespondenzen unterscheiden:

1. In beiden Kameras wird die primäre Reflektion detektiert.

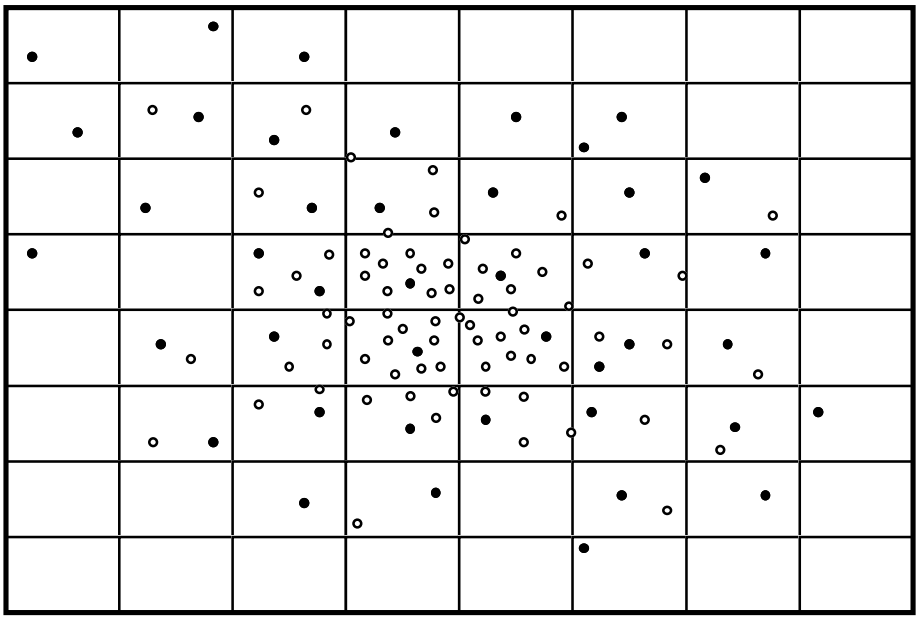
2. In einer Kamera wird die primäre Reflektion, in der anderen Kamera wird die Spiegelung detektiert.
3. In beiden Kameras wird die Spiegelung detektiert.

Die Fälle 1 und 3 sind unproblematisch, da sie zu korrekten Korrespondenzen führen. Der Fall 2 führt im Allgemeinen zu einer völlig falschen Korrespondenz. Gerade bei den laserbasierten Verfahren ist das Problem der Reflektion zwar bereits während der Bildverarbeitung durch einfache Heuristiken zu lösen, allerdings gehen dann auch die Lokalitätsannahmen verloren, da im Zweifelsfall das gesamte Bild nach Laserpunkten durchsucht werden muss. Andererseits sollten Reflektionen auf dem Objekt eher als Ausnahmen angesehen werden, so dass der Mehraufwand an Rechenleistung durch den Nutzen nicht mehr gerechtfertigt ist. Daher wird bei der Auswahl von Korrespondenzen zur Berechnung der Epipolargeometrie besondere Rücksicht auf potentiell falsche Korrespondenzen genommen.

Um vor allem das Problem der fehlerhaften Korrespondenzen zu lösen, werden mehrere Strategien zur Berechnung der Epipolargeometrie gemischt. Die Hauptstrategie beruht auf der Annahme, dass die Zahl der falschen Korrespondenzen gering ist gegen die Zahl der genauen Korrespondenzen und dass weitaus mehr Korrespondenzen vorhanden sind, als für die Kalibrierung nötig sind. Wenn eine Vielzahl von Epipolargeometrien auf Basis zufälliger Teilmengen der vorhandenen Korrespondenzen berechnet werden, so ist zu erwarten, dass auch die meisten Lösungen relativ genau sind. Ein Vergleich und eine Bewertung der berechneten Epipolarmatrizen ermöglicht dann die selektive Auswahl von „guten“ Ergebnissen. Es gibt allerdings Fälle, in denen die berechnete Epipolargeometrie sehr gute Ergebnisse liefert, aber wenig repräsentativ für die Konfiguration der Stereokameras ist. Das passiert vor allem dann, wenn alle Korrespondenzen nur eine kleine Region in beiden Bildern abdecken.

Die zweite Strategie dient zur zufälligen Auswahl einer Menge von Korrespondenzen, die möglichst repräsentativ für die Konfiguration der Stereokameras sind. Dazu werden alle Korrespondenzen in Bezug auf eine Kamera in einem Raster sortiert. Zur Suche einer Korrespondenz wird dann zunächst zufällig ein Raster bestimmt, aus dessen Inhalt dann eine Korrespondenz gewählt wird. Bei der Wahl vieler Korrespondenzen wird so gewährleistet, dass die Korrespondenzen unabhängig von lokalen Häufungen über das gesamte Bildfeld einer Kamera verteilt sind (Abbildung 3.10).

Die aus solchen Mengen berechneten Epipolargeometrien entsprechen mit sehr hoher Sicherheit der Aufnahmekonfiguration des Stereosystems. Um bei einer Vielzahl von Berechnungen den Anteil schlechter Epipolargeometrien gering zu halten, ist es weiter sinnvoll, die Qualität jeder Korrespondenz zu bestimmen, um sukzessive schlechte Korrespondenzen zu eliminieren und durch gute zu ersetzen. Wenn die Zuverlässigkeit der Epipolargeometrie bereits relativ hoch ist, können schlechte Korrespondenzen bereits vorab aussortiert werden. Korrespondenzen, deren Qualität ausreichend hoch ist, werden in dem oben beschriebenen Raster gesammelt. Sobald die Zahl der Einträge in einem Rasterfeld zu hoch ist, wird die Qualität aller Korrespondenzen dieses Rasterfeldes bezüglich der aktuellen Epipolargeometrie bewertet. Die schlechtesten Korrespondenzen werden anschließend gelöscht.



● ausgewählte Korrespondenzen ○ nicht berücksichtigte Korrespondenzen

Abbildung 3.10: Unterteilung des Kamerabildes zur Auswahl von Korrespondenzen

Durch die beschriebenen Ansätze ist immer eine bestmögliche Verteilung der Korrespondenzen im Bildbereich der Kameras gewährleistet. Zudem wird durch die Aussonderung schlechter Korrespondenzen eine zunehmende Verbesserung der Epipolargeometrie ermöglicht. Die Zahl der Korrespondenzen, die zur Berechnung einer Epipolargeometrie verwendet werden, wurde empirisch ermittelt. Dabei hat sich eine Zahl von etwa 100 Korrespondenzen als optimal erwiesen. Weniger Korrespondenzen sind nicht mehr repräsentativ für die Aufnahmesituation. Bei zu vielen Korrespondenzen wird das System numerisch zunehmend instabiler und es besteht die Gefahr, dass zuviele schlechte Korrespondenzen in die Berechnung eingehen und die Ergebnisse verfälschen. Zudem steigt der nötige Rechenaufwand quadratisch mit der Zahl der Korrespondenzen. Zur Berechnung der Qualität einer Epipolargeometrie wird der mittlere Abstand der Korrespondenzpunkte eines Bildes zur Epipolarlinie des zugehörigen Punktes im anderen Bild herangezogen. Obwohl dieser Wert keine Aussage über die absolute Genauigkeit der später rekonstruierten Geometrie zuläßt, sondern lediglich als Maßzahl dienen kann, eignet er sich sehr gut für den qualitativen Vergleich zweier Epipolargeometrien. Diese Maßzahl wird im Folgenden als interne Qualität bezeichnet, sie wird in der Größenordnung von Pixeln gemessen. Die geometrische Genauigkeit der Rekonstruktion wird als externe Qualität bezeichnet. Sie beschreibt als Kennzahl das Verhältnis der Genauigkeit zum durchmessenen Volumen. Ein Wert von 1:1000 bedeutet dabei, dass bei einer durchmessenen Tiefe von einem Meter eine Genauigkeit von einem Millimeter vorliegt. Die externe Qualität kann für eine einzelne Rekonstruktion im Allgemeinen nicht trivial berechnet werden. Eine tatsächliche Genauigkeit kann normalerweise nur exemplarisch anhand der Vermessung eines Normkörpers bestimmt werden.

Bei der Verwendung des Laserpunktes zur Kalibrierung des Stereosystems fallen maximal 30 Korrespondenzen pro Sekunde an. Nach 20 Korrespondenzen wird die erste Kalibrierung berechnet. Dazu wird mehrfach eine zufällige Auswahl von maximal 100 Korrespondenzen aus der vorhandenen Menge zur Berechnung gewählt. Die Epipolargeometrie mit der besten

Güte gilt als erste Kalibrierung. Alle weiteren 20 Korrespondenzen wird der Vorgang wiederholt. Die neuen Ergebnisse nach jedem Schritt werden jeweils mit der bisher besten Epipolargeometrie verglichen und gegebenenfalls übernommen. Somit steht bereits nach wenigen Sekunden eine erste Näherung der Konfiguration des Stereosystems zur Verfügung. In praktischen Tests hat sich gezeigt, dass bereits nach wenigen Schritten sehr gute Ergebnisse erreicht werden, die im Weiteren nur noch geringfügig verbessert werden können. Interne Genauigkeiten von einem zehntel Pixel lassen sich relativ schnell erreichen, bei längerem Scannen wurden Ergebnisse mit bis zu zwanzigfacher Pixelgenauigkeit erreicht.

Zur Berechnung der externen Genauigkeit wurde für das System zunächst eine Kalibrierung durchgeführt. Danach wurde eine Ebene gescannt, die 3D-Rekonstruktion der Daten wird geometrisch mit einer Ebene approximiert und für jeden Punkt der 3D-Rekonstruktion ist der Abstand zur Ebene der Bewertungsmaßstab der externen Qualität. Die externe Qualität ist sehr stark vom Winkel zwischen den Hauptachsen beider Kameras abhängig. Bei geringem Winkel ist zwar eine sehr gute interne Qualität erreichbar, die externe Qualität wird aber vom Rauschen der Daten dominiert. Ein Winkel von 90 Grad ist optimal, da sich die Sehstrahlen durch die Korrespondenzpunkte beider Kameras annähernd im rechten Winkel schneiden, allerdings verringert sich das durchmessbare Volumen sehr stark und die Gefahr von Abschattungen ist sehr hoch. Zudem ist die Genauigkeit der internen Kameraparameter maßgeblich für die externe Genauigkeit der Stereokalibrierung verantwortlich. Im Folgenden werden die exemplarisch bestimmten externen Genauigkeiten für unterschiedliche Kameraanordnungen kurz dargestellt. Die interne Genauigkeit hat für alle Konfigurationen wenigstens zehnfache Pixelgenauigkeit betragen.

Das Testszenario wurde jeweils mit zwei SONY XC55 Kameras durchgeführt, die mit 12,5mm Objektiven ausgestattet waren. Mit der Laserlinie wurde jeweils eine ebene Tischplatte gescannt. Der Abstand der Kameras von der Platte hat jeweils etwa 0,65 Meter betragen, die erfassbare Fläche hatte damit einen Durchmesser von etwa 0,34 Meter. Die Scandaten werden dann mit einer Ebene approximiert und der Abstand jedes Scanpunktes zu dieser Ebene wird gemessen. Um vergleichbare Größen zu erhalten, wird dieser Abstand ins Verhältnis zum durchmessenen Volumen gesetzt (Kennzahl). Zum Vergleich wurde eine weitere Messung mit dem ABW-Projektor durchgeführt.

1. Winkel zwischen den Kameras: 9,6 Grad

   Basisabstand: 11 cm
   Kennzahl: 679
   Mittlere Genauigkeit: 0,49 mm
   Amplitude des Rauschen: 0,44 mm

2. Winkel zwischen den Kameras: 36,4 Grad

   Basisabstand: 44 cm
   Kennzahl: 2835
   Mittlere Genauigkeit: 0,12 mm
   Amplitude des Rauschen: 0,11 mm

3. Winkel zwischen den Kameras: 64,1 Grad

   Basisabstand: 75 cm
   Kennzahl: 4830
   Mittlere Genauigkeit: 0,075 mm
   Amplitude des Rauschen: 0,066 mm

4. Winkel zwischen den Kameras: 47,7 Grad (Messung mit Gray-Code Projektor)

   Basisabstand: 56 cm
   Kennzahl: 10677
   Mittlere Genauigkeit: 0,03 mm
   Amplitude des Rauschen: 0,018 mm

Es ist deutlich zu erkennen, wie die Kennzahl und die erreichbare Genauigkeit vom Winkel zwischen den Kameras, bzw. vom Basisabstand abhängt. Bei den Messungen mit der Laserlinie war die Amplitude des Rauschens nur wenig geringer als die Genauigkeit. Die Qualität der Gesamtmessung wurde also im Wesentlichen vom Sensorrauschen dominiert. Die Stereokalibrierung ist andererseits unabhängig von dieser Größe jeweils optimal abgelaufen. Beispielhaft sind für die zweite und vierte Messung die berechneten Fehlerbilder in Abbildung 3.11 dargestellt.

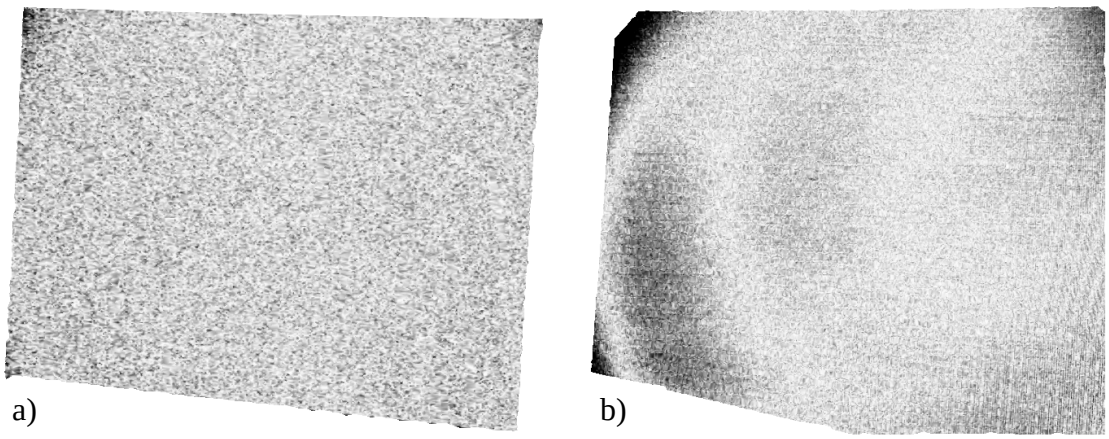


Abbildung 3.11: Rauschverhalten: a) Laserlinie b) Gray-Code Projektor

Die Vermessung mit der Laserlinie zeigt im Fehlerbild eine gleichmäßige Verteilung von Pixeln unterschiedlichster Qualität. Lediglich links und rechts oben in der Ecke des Bildes zeigen sich geringe systematische Fehler. Bei der Verwendung des Projektors wurde eine vergleichsweise deutlich höhere Genauigkeit erreicht. Im Fehlerbild dominieren systematische Fehler, die auf Restungenauigkeiten des verwendeten Kameramodells zurückzuführen sind. Der maximale Fehler oben links im Fehlerbild beträgt dabei etwa 0,1 Millimeter. Während bei der Verwendung der Laserlinie das Sensordatenrauschen die höchste Amplitude unter den verwendeten Verfahren hat, ist es beim Gray-Code Verfahren in Kombination mit dem Phase-Shift Verfahren am geringsten. Die Datenqualität bei der Verwendung des Laserpunktes findet sich zwischen diesen beiden Verfahren.

Die Verwendung des Gray-Code Verfahrens zur Kalibrierung des Stereosystems stellt in einem Schritt mehrere hunderttausend Korrespondenzen zur Verfügung. Die Berechnung der Epipo-

largeometrie erfolgt daher iterativ. Während jedem Iterationsschritt wird eine Auswahl von Korrespondenzen bestimmt, eine Epipolargeometrie berechnet und schlechte Korrespondenzen eliminiert. Der Vorgang wird solange wiederholt, bis sich keine nennenswerten Verbesserungen mehr ergeben.
Die Qualität der Korrespondenzen aus dem Gray-Code Verfahren ist deutlich höher als die Qualität der Laserkorrespondenzen, da Fehler auf Grund der Schwerpunktberechnung beim Laserverfahren oder auf Grund von Speckles nicht auftreten. Dennoch ist die interne Qualität der Kalibrierung bei beiden Verfahren vergleichbar, da nach ausreichend vielen Schritten alle schlechten Korrespondenzen eliminiert sind und die Berechnung nur noch auf Grund sehr genauer Korrespondenzen erfolgt. Die externe Qualität war bei gleicher Konfiguration deutlich besser. Das Gray-Code Verfahren zeichnet sich dabei vor allem durch eine hohe laterale Ortsauflösung aus, da durch die fehlende Schwerpunktbildung im Gegensatz zum laserbasierten Verfahren kein Tiefpass über mehrere Pixel hinweg gerechnet werden muss.
Beim Vergleich von Laserlinie und Laserpunkt ist jedoch die Laserlinie vorzuziehen. Die Genauigkeit der Tiefe ist zwar geringer als beim Laserpunkt, im Gegensatz zum Laserpunkt entspricht die laterale Auflösung aber in etwa der Auflösung der Tiefe, so dass das Ergebnis bei flächigen Vermessungen insgesamt homogener ist. Sollen nur einzelne Objektpunkte ausgemessen werden, gewinnt wieder der Laserpunkt.

Vergleichende Rekonstruktionsergebnisse aller vorgestellten Scan-Verfahren zur Illustration der beschriebenen Sachverhalte sind in Kapitel 3.1.6 (Beispiele) zu finden.

3.1.5 3D-Rekonstruktion und Visualisierung

Ein wesentliches Konzept von *HandyScan3D* ist die echtzeitfähige Datenverarbeitung. Besonders bei der Verwendung des Punktlasers und der Laserlinie besteht die Zielsetzung, die Datenverarbeitung in Videoechtzeit durchzuführen. Neben den beschriebenen Optimierungen im Bereich der Bildverarbeitung stellt dieses Ziel sehr hohe Anforderungen an die 3D-Rekonstruktion und die Visualisierung.
Die Verwendung von Gray-Codiertem Licht unterliegt keinen Echtzeitanforderungen, da die Kamerabilder verfahrensbedingt auch zeitversetzt aufgenommen werden können, sofern mit den korrespondierenden Bildern die gleichen Linienmuster erfasst werden. Die 3D-Rekonstruktion und Visualisierung wird aber dennoch analog zu den laserbasierten Verfahren durchgeführt.

3.1.5.1 3D-Rekonstruktion

Das Ziel der 3D-Rekonstruktion ist zum einen die Berechnung von Raumkoordinaten in einem Bezugskoordinatensystem aus den vorhandenen Stereokorrespondenzen. Als Bezugssystem wird üblicherweise das Koordinatensystem einer Kamera gewählt. Applikationsspezifisch kann aber auch jedes andere Bezugssystem verwendet werden. Das zweite Ziel ist die Triangulation der Raumpunkte zur Bildung eines Dreiecksnetzes, mit dem eine Visualisierung durchgeführt werden kann.

Die Berechnung von 3D-Koordinaten aus korrespondierenden Bildpunkten geschieht geometrisch. Über die bekannten Kameramodelle, sowie der aus der Stereokalibrierung berechneten Lagen der Stereokameras zueinander, kann zu jedem Bildpunkt ein Strahl berechnet werden,

der alle Raumkoordinaten beschreibt, die auf diesen Bildpunkt im Kamerabild abgebildet werden. Jeder dieser Sehstrahlen $\vec{L}$ wird in Punkt-Vektor Form beschrieben:

$$\vec{L} = \vec{P} + \lambda \vec{V} . \tag{3.16}$$

Der Punkt $\vec{P}$ entspricht dem Brennpunkt der Kamera und kann im Sinne der Stereokalibrierung als berechnete relative Lage der Kamera beschrieben werden. Der Richtungsvektor $\vec{V}$ wird aus der Orientierung der Kamera, den internen Kameraparametern und der Pixelposition des zugehörigen Bildpunktes berechnet. Er wird normiert, so dass der freie Parameter λ die Entfernung eines gesuchten Raumpunktes zum Brennpunkt der Kamera beschreibt.

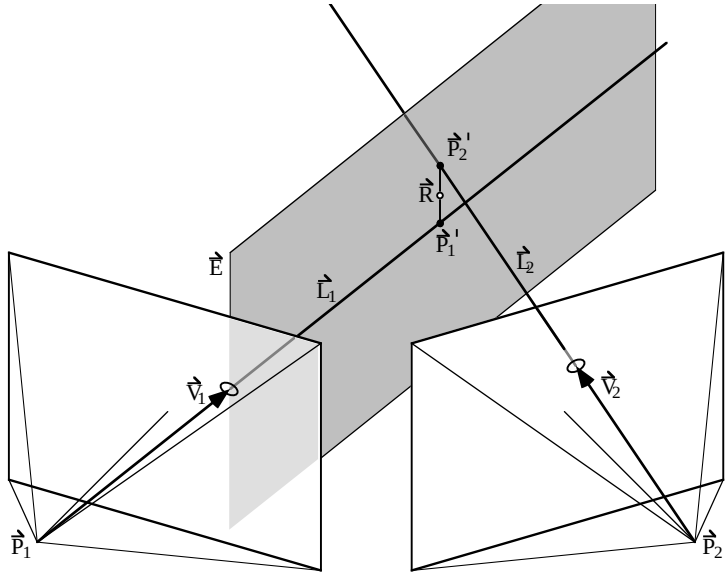


Abbildung 3.12: Berechnung von Raumkoordinaten

Ein Paar von Korrespondenzpunkten führt zu zwei Sehstrahlen, die im Allgemeinen windschief aneinander vorbeilaufen. Der Raumpunkt, der beiden Sehstrahlen am nächsten liegt, wird im Weiteren als Rekonstruktion des vermessenen Objektpunktes verwendet. Zur Berechnung dieses Punktes werden entsprechend Gleichung (3.16) die Sehstrahlen $\vec{L}_1$ und $\vec{L}_2$ der beiden Bildpunkte bestimmt. Dann wird über den Sehstrahl der ersten Kamera, sowie dem Normalenvektor auf beide Richtungsvektoren eine Ebene $\vec{E}$ gebildet (Abbildung 3.12):

$$\vec{E} = \vec{P}_1 + \lambda \vec{V}_1 + \mu (\vec{V}_1 \times \vec{V}_2) . \tag{3.17}$$

Der Schnittpunkt des zweiten Sehstrahls durch diese Ebene ergibt den Raumpunkt $\vec{P}_2'$ auf dem zweiten Sehstrahl, der dem ersten Sehstrahl am nächsten liegt. Über den dabei berechneten Parameter $\hat{\mu}$ und den Normalenvektor kann auf den entsprechenden Punkt $\vec{P}_1'$ des ersten Sehstrahls geschlossen werden:

$$\vec{P}_1' = \vec{P}_2' + \hat{\mu} (\vec{V}_1 \times \vec{V}_2) . \tag{3.18}$$

Die Mitte zwischen diesen beiden Punkten ist der rekonstruierte Raumpunkt $\vec{R}$:

$$\vec{R} = \frac{\vec{P}_1' + \vec{P}_2'}{2}. \tag{3.19}$$

Die Verwendung des Laserpunktes führt bei der Verwendung der beschriebenen Standard CCD-Kameras zu maximal 30 rekonstruierten Raumpunkten pro Sekunde. Die Verwendung der Laserlinie führt zu vielen hundert Raumpunkten pro Sekunde. Da der Laser in beiden Fällen unsystematisch über das Objekt bewegt wird, können keine Annahmen über die Anordnung der Raumpunkte gemacht werden. Die Berechnung eines Dreiecksnetzes auf Basis einer ungeordneten Punktewolke ist andererseits ein komplexes Problem, das normalerweise mit $O(n^3)$ skaliert, da potentiell viele Nachbarschaftsbeziehungen zwischen Punkten analysiert und Mehrdeutigkeiten aufgelöst werden müssen. Die zugrundeliegende Sensorkonfiguration und das erzeugende Verfahren sind aber bekannt, so dass zwei Vereinfachungen gemacht werden können, die die Komplexität des Problems bis auf eine lineare Komplexität O(n) verbessern.

1. Für jede Triangulation im dreidimensionalen Raum existiert eine äquivalente Triangulation im zweidimensionalen Raum.

   Da alle Punkte sich aus den Schnittpunkten der Sehstrahlen zweier Kameras berechnen, lassen sich alle Punkte in das zweidimensionale Bildfeld dieser Kameras projizieren. Das erfasste Objekt bleibt während der Vermessung ortsfest. Punkte, die in Bezug auf das Sichtfeld einer Kamera verdeckt sind, können nicht vermessen werden. Daher ist die Projektion bijektiv und eindeutig.
   Welche der Kameras sich besser für die Projektion eignet kann im Allgemeinen allerdings nicht entschieden werden.

2. Die laterale Auflösung des Sensors ist begrenzt

   Zwar kann ein Laserpunkt beliebig fein über das Objekt bewegt werden, die Position im Bildfeld ist also prinzipiell nur durch numerische Genauigkeiten begrenzt. Allerdings wird die Lage des Laserpunktes im Bild durch Schwerpunktbildung über mehrere Pixel berechnet. Dadurch können Details auf dem Objekt, die kleiner als ein bis zwei Pixel sind, nicht mehr aufgelöst werden. Für die Laserlinie gilt das Gleiche, da auch hier eine Schwerpunktbildung durchgeführt wird.
   Beim Gray-Code Verfahren wird für jedes Pixel ein Tiefenwert berechnet. Die laterale Auflösung beträgt damit genau ein Pixel.

Aufgrund dieser beiden Annahmen werden alle berechneten Raumpunkte in ein zweidimensionales Raster fester Größe abgebildet. Für die Zuordnung wird eine Kamera bestimmt, deren Pixelkoordinaten skaliert und diskretisiert unmittelbar den Positionen im zweidimensionalen Raster entsprechen. Die Abbildung eines Raumpunktes in dieses Raster wird damit trivial.

Jedes Element des Rasters nimmt genau einen Raumpunkt auf. Falls ein Rasterelement bereits belegt wird, wird der vorhandene Punkt verdrängt, sofern der neue Punkt „besser“ ist. Die Qualität eines Punktes wird von zwei Größen bestimmt. Zum einen kann für jeden Punkt eine Genauigkeit (Q) berechnet werden, die sich direkt aus der Epipolargeometrie ergibt. Zum anderen besteht ein Ziel darin, eine möglichst gleichmäßige Verteilung der Punkte im Raster zu erhalten. Punkte, die sich in der Mitte eines Rasterelements befinden, sind Punkten am Rand

des gleichen Rasterelements zu bevorzugen. Daher wird für jeden Punkt der Abstand[1] (D) zur Rastermitte berechnet. Die Entscheidung, ob ein Punkt (P_n) in ein Rasterelement ($R_{i,j}$) eingetragen wird, berechnet sich dann nach folgender Regel:

$$R_{i,j} = \begin{cases} P_n & \text{falls } R_{i,j} \text{ leer ist} \\ P_n & 0,5 \geq D(R_{i,j}) > 0,25 > D(P_n) \\ P_n & Q(P_n) < Q(R_{i,j}) \\ R_{i,j} & \text{sonst} \end{cases} \tag{3.20}$$

Der erste Fall ist trivial. Der zweite Fall sorgt dafür, dass ein Punkt nahe der Mitte eines Rasterelements einen Punkt verdrängt, der am Rand des Elements liegt. Falls die ersten beiden Fälle nicht zutreffen, entscheidet die höhere Genauigkeit über das Eintragen eines neuen Punktes. Insgesamt kann es zwar vorkommen, dass ein Punkt mit schlechter Genauigkeit einen Punkt mit hoher Genauigkeit verdrängt, häufiges Anstrahlen kritischer Bereiche auf dem Objekt sorgt aber dafür, dass zunehmend Punkte mit geringer Genauigkeit ausgesondert werden und eine gleichmäßige Belegung des Rasters entsteht.

Die eigentliche Rekonstruktion der Punkte und das Eintragen der Punkte in das Raster geschieht mit konstanter Komplexität O(1). Die Erzeugung eines Dreiecksnetzes auf Basis der Punkte im Rasters wird mit linearer Komplexität O(n) durchgeführt. Da die Größe des Rasters begrenzt ist, kann die maximale Laufzeit der Rekonstruktion in einem Visualisierungsschritt berechnet werden. Für eine Online-Visualisierung der aktuellen Scan-Ergebnisse wird mit jedem neuen Punkt ein neues Dreiecksnetz berechnet, daher sind die zugrundeliegenden Verfahren Bestandteil der Visualisierung und werden im folgenden Abschnitt beschrieben.

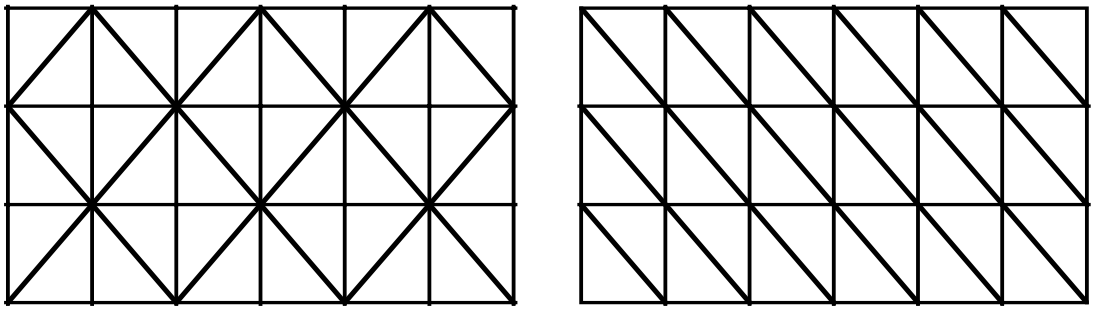
3.1.5.2 Visualisierung

Die Berechnung eines Dreiecksnetzes auf Basis eines gleichmäßigen zweidimensionalen Rasters ist grundsätzlich gesehen relativ trivial. Da aus dem Raster benachbarte Punkte durch Indexberechnungen abgeleitet werden können, können einfache Regeln aufgestellt werden, um Dreiecke zu bilden. Es sind allerdings nahezu beliebig viele Topologien denkbar, wie ein Dreiecknetz ein Raster abdeckt. Von symmetrischen Fällen abgesehen, werden hier nur die beiden regelmäßigen Topologien vorgestellt[2] (Abbildung 3.13).

Algorithmisch ist der Aufwand zur Berechnung beider Dreiecksnetze etwa gleich hoch. Die Entscheidung für die eine oder andere Typologie ist von dem zugrundeliegenden Grafiksubsystem abhängig. In der vorliegenden Implementierung wurde die Bildschirmdarstellung mit OpenGL durchgeführt. Die Darstellung eines einzelnen Dreiecks geschieht durch die Aufzählung von drei Raumpunkten. Die OpenGL-Engine führt dann die Darstellungstransformation durch und stellt das Dreieck als geschlossene Fläche in einem Fenster dar. Durch die zusätzliche Angabe von Flächen- oder Punktnormalen, Farben oder Texturkoordinaten kann die Dar-

1. Der Abstand wird aufgrund der einfachen Rechnung mit der Maximumsnorm bestimmt
2. Alle unregelmäßigen Topologien sind nur schwer zu berechnen und bringen keine Vorteile gegenüber den beiden regelmäßigen Topologien.

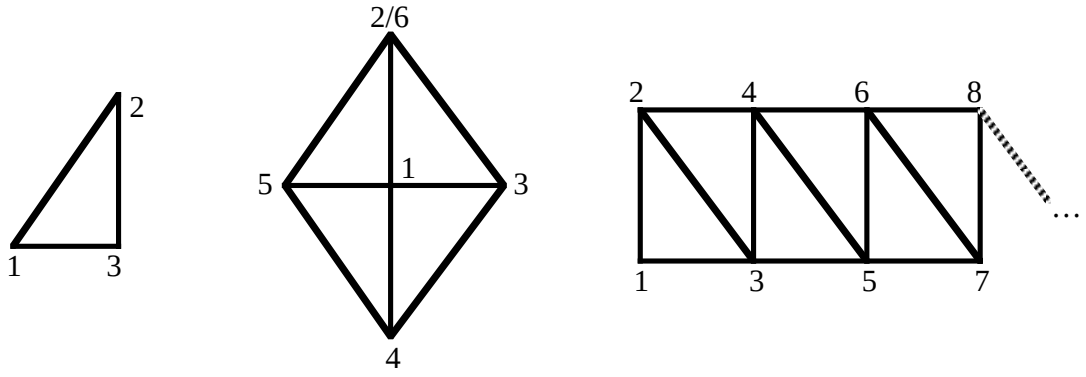
stellung verbessert werden. Zusätzlich können noch einfache Beleuchtungsmodelle berechnet werden.



Typ A Typ B

Abbildung 3.13: Die beiden regelmäßigen Netztopologien

Neben der Darstellung eines Dreiecks durch die Aufzählung von drei Punkten gibt es weitere optimierte Aufzählungen zur Darstellung von Dreiecksnetzen. Exemplarisch werden hier die beiden Varianten vorgestellt, die bei der realisierten Visualisierung Anwendung finden.



Triangle Triangle Fan Triangle Strip

Abbildung 3.14: Aufzählungsregeln zur Visualisierung von Dreiecksnetzen

Der *TriangleFan* erzeugt mit der Aufzählung von sechs Raumpunkten vier Dreiecke, dabei wird ein Punkt zweifach aufgezählt. Diese Aufzählung findet in *Typ A* Anwendung. Der *Typ B* wird mit dem *TriangleStrip* visualisiert. Hier werden für die Darstellung von n Dreiecken (n+2) Punkte aufgezählt, ohne dass Punkte innerhalb eines *TriangleStrips* wiederholt werden müssen. Der *TriangleStrip* ist also besonders dann sinnvoll, wenn lange Ketten gebildet werden können. In Abbildung 3.15 ist beispielhaft dargestellt, wie die Punkte in einem vollständig belegten Raster aufgezählt werden. Alle überzähligen Dreiecke, die nicht in das vorgegebene Schema passen, werden durch die Aufzählung von drei Punkten (*Triangle*) gebildet. Durch die Verwendung von *Typ A* zur Triangulation des dargestellten Rasters werden über die Aufzählung von 78 Raumpunkten 36 Dreiecke gebildet. Das Raster selbst besteht aus 28 Punkten. Einzelne Punkte tauchen also bis zu sechs mal in der Aufzählung auf. Insgesamt werden fünf *TriangleFan* gebildet, 16 überzählige Dreiecke werden über die Aufzählung von drei Punkten gebildet.

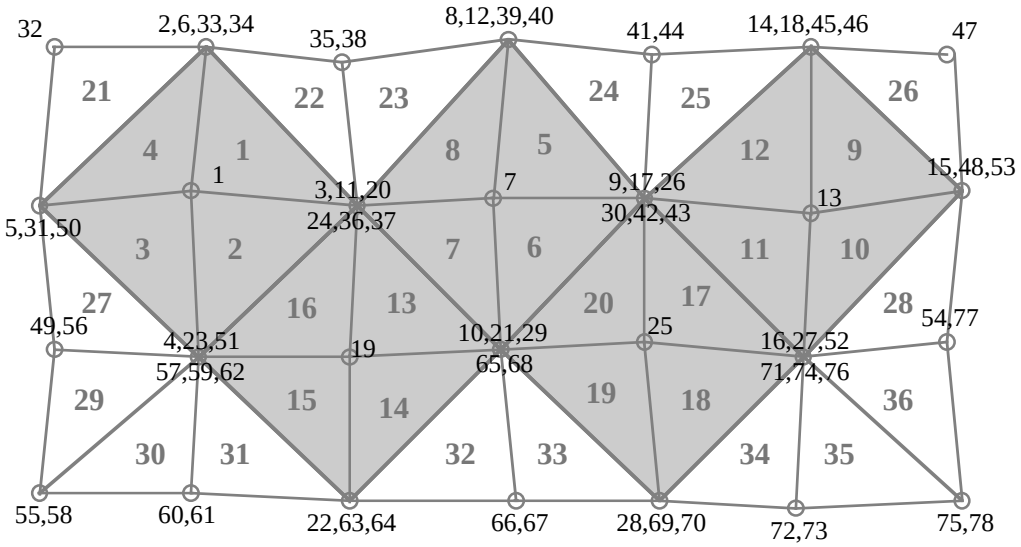


Abbildung 3.15: Triangulation des Rasters (*Typ A*)

Nicht immer ist es notwendig, alle Punkte zur Triangulation zu verwenden. Besonders bei großen Rastern oder Computern mit zu geringer Rechenleistung führt nur eine ausgedünnte Darstellung zu einer echtzeitfähigen Visualisierung. Daher wird die realisierte Implementierung mit zwei Parametern versehen, die eine lineare Ausdünnung in Spalten und Zeilenrichtung beschreiben. Ein Wert von zwei bedeutet dabei, dass nur jeder zweite Punkt in jeder Zeile, bzw. jeder Spalte des Rasters verwendet wird. Analog zu den Parametern kann eine proportionale Beschleunigung der Triangulation auf Kosten der Detailauflösung der Visualisierung erreicht werden. Allerdings ist auch bei recht grober Darstellung die Form des gescannten Objektes oft noch gut zu erkennen. Die Steuerung der Parameter wird jeweils dem Anwender überlassen.

Ein größeres Problem sind unvollständige Raster, also Rasterelemente ohne Einträge, die die Bildung von Dreiecken verhindern. Ursache dafür sind vor allem unvollständige Scans und fehlende Daten aufgrund von Abschattungen. Um vor allem am Anfang des Scan-Vorgangs früh eine Darstellung der Ergebnisse zu erhalten, wurde ein Verfahren entwickelt, mit dem Lücken bis zu einer begrenzten Größe überbrückt werden können. Die Aufzählung der Punkte geschieht wie beschrieben, bis es zu einer Unterbrechung aufgrund von fehlenden Punkten kommt. In der direkten Umgebung wird dann nach einem Ersatzpunkt gesucht, dabei wird von innen nach außen vorgegangen. Sobald ein Punkt gefunden wird, wird die Aufzählung weitergeführt. Die Suche wird spätestens bei den Nachbarn vierter Ordnung, also den Punkten, die maximal vier Rasterelemente vom Startraster entfernt liegen, beendet. Wenn kein Punkt gefunden wird, wird die Aufzählung abgebrochen und mit dem nächsten belegten Rasterpunkt wieder begonnen. Es ist leicht einzusehen, dass dieses Verfahren schnell zu entarteten Dreiecken führt, also Dreiecken, bei denen wenigstens zwei Eckpunkte identisch sind. Da das Verfahren aber ausschließlich zur Visualisierung dient, wirken sich diese Dreiecke nicht weiter störend aus, da sie vom Grafiksubsystem ohnehin nicht dargestellt werden können. Wenn die Triangulation ohnehin ausgedünnt dargestellt wird, kann mit dem beschriebenen Verfahren eine deutliche Aufwertung der Darstellung erzielt werden. Bei einer Begrenzung der Suchtiefe für

Ersatzpunkte auf maximal die Hälfte des Ausdünnungsfaktors entfallen zusätzlich die entarteten Dreiecke. In Abbildung 3.16 ist beispielhaft die Triangulation eines spärlich besetzten Rasters dargestellt.

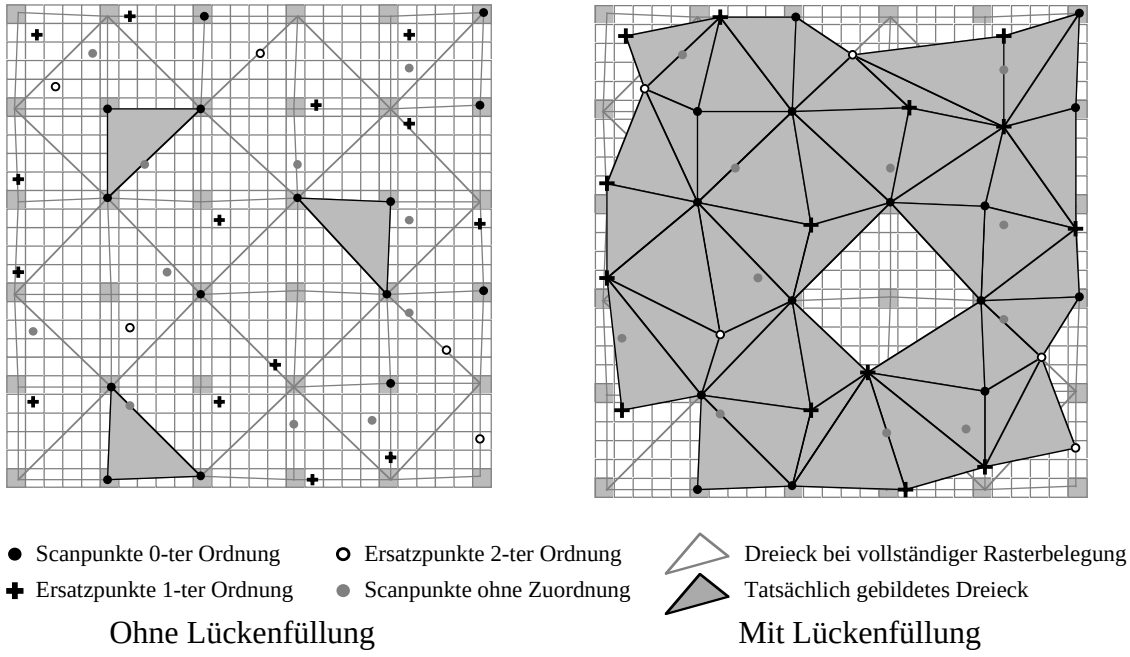


Abbildung 3.16: Triangulation eines spärlich besetzten Rasters

Für das Beispielraster wurden 45 willkürlich verteilte Punkte aufgenommen. Der Ausdünnungsfaktor ist für die Spalten und Zeilen auf den Wert fünf eingestellt. Ohne Lückenfüllung wird nur etwa jeder fünfundzwanzigste Punkt für die Triangulation herangezogen, im dargestellten Beispiel sind es immerhin neun Punkte. Es sind zwar weitere Dreiecke möglich, aufgrund der verwendeten Typologie (*Typ A*) werden die entsprechenden Beziehungen aber nicht gefunden. Mit Lückenfüllung werden bei einer Suchtiefe bis zu Punkten zweiter Ordnung im dargestellten Beispiel bereits 34 Punkte für die Triangulation verwendet, nur elf Punkte können nicht berücksichtigt werden. Die Zahl der Dreiecke erhöht sich von drei auf 44. Es treten keine entarteten Dreiecke auf und die Darstellung ist fast vollständig.

3.1.5.3 Erstellung von Mehrfachansichten

Das Basiskonzept von *HandyScan3D* sieht nur die Aufnahme einer Ansicht des Objektes vor. Eine einfache Erweiterung ermöglicht aber die Aufnahme von Mehrfachansichten durch Umstellen von Kameras. Dazu wird zunächst eine erste Ansicht aufgenommen. Wenn das Objekt in dieser ersten Ansicht ausreichend erfasst wurde, wird die Aufnahme unterbrochen und eine Kamera umgestellt. Der Scan-Vorgang wird dann mit einer neuen Rekonstruktion und einer erneuten Kalibrierung wieder aufgenommen. Da die neue Ansicht ebenfalls wieder unskaliert ist, muss nur noch die relative Skalierung zwischen beiden Ansichten berechnet werden. Die ortsfest gebliebene Kamera bildet das Referenzsystem, über das beide Ansichten zusammengeführt werden. Die neue relative Lage der umgestellten Kamera entspricht der Hintereinanderausführung der im Rahmen der beiden Stereokalibrierungen berechneten Lagetransformationen[1]. Zur Berechnung der Skalierung des neuen Stereosystems relativ zum alten

Stereosystem werden die Bildkoordinaten einzelner Scanpunkte in der ortsfesten Kamera mit den bereits in der ersten Ansicht aufgenommenen Scanpunkten verglichen. Ist an gleicher Position im Bild der ortsfesten Kamera bereits ein Scanpunkt aus der ersten Ansicht vorhanden, so ergibt der Quotient aus den Abständen beider Punkte zum Brennpunkt der ortsfesten Kamera die gesuchte Skalierung (Abbildung 3.17). Oder anders ausgedrückt ist das Verhältnis der Strecken e_1 und e_2 zueinander gleich dem Verhältnis der Strecken v_1 und v_2 zueinander.

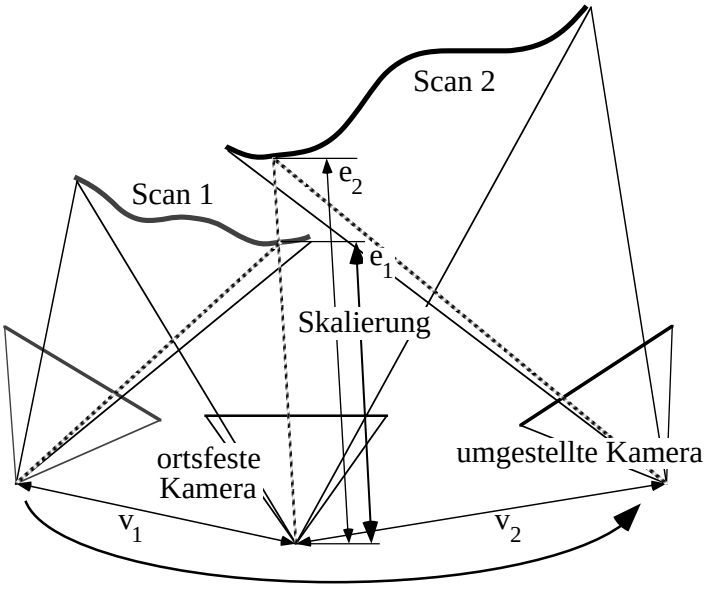


Abbildung 3.17: Skalierung von Mehrfachansichten

Weitere Ansichten werden in gleicher Weise erstellt, so dass das Objekt zunehmend vollständiger erfasst wird. Wenn die Kameras alternierend umgestellt werden, kann eine schnelle Vervollständigung des gescannten Objektes erreicht werden. Allerdings besteht dann auch die Gefahr, dass sich Ungenauigkeiten in den Stereokalibrierungen aufsummieren, da sich die Lage der zuletzt umgestellten Kamera aus der Hintereinanderausführung aller bisher berechneten Kalibrierungen zusammensetzt. Ein weitere Gefahr besteht darin, dass beim Umstellen einer Kamera versehentlich die ortsfeste Kamera berührt wird. Ein Verwackeln der ortsfesten Kamera hat zur Folge, dass keine zuverlässige Integration der neuen Ansicht in die vorhandenen Ansichten mehr möglich ist. Die gesamte Vermessung muss dann abgebrochen und gegebenfalls von vorne begonnen werden. Andererseits stellt dieses Verfahren eine einfache Möglichkeit dar, ohne zusätzlichen Aufwand von Rechner- oder Scannerhardware eine vollständige Rekonstruktion eines Objektes im Rechner zu erstellen.

3.1.6 Beispiele

Die beschriebenen Funktionen von *HandyScan3D* erlauben eine Vielzahl von Kombinationen in Hinsicht auf das Scan-Verfahren, die 3D-Rekonstruktion und die Visualisierung, die in ihrer Gänze hier sicher nicht bildlich zu dokumentieren sind. Vor allem ein wesentlicher Aspekt, das

1. Unter der Voraussetzung, dass die Master-Kamera umgestellt wurde

echtzeitfähige Scannen, Rekonstruieren und Visualisieren von Objekten lässt sich im Rahmen dieser Arbeit überhaupt nicht darstellen. Dennoch sollen die folgenden Beispiele unterschiedliche Kernaspekte von *HandyScan3D* beleuchten und vergleichen.

Die ersten Beispiele stellen wesentliche sensorische Eigenschaften und Unterschiede des Laserpunktes, der Laserlinie und des Gray-Code Verfahrens dar. Zunächst wurde mit jedem der drei Verfahren eine Kartonverpackung eingescannt. Die Anordnung der Kameras wurde dabei jeweils unverändert beibehalten. Die Verpackung ist bunt bedruckt, hat aber ansonsten eine gleichmäßig glatte Oberfläche. Dargestellt ist jeweils die rekonstruierte Geometrie im Vergleich zu den texturierten Daten.

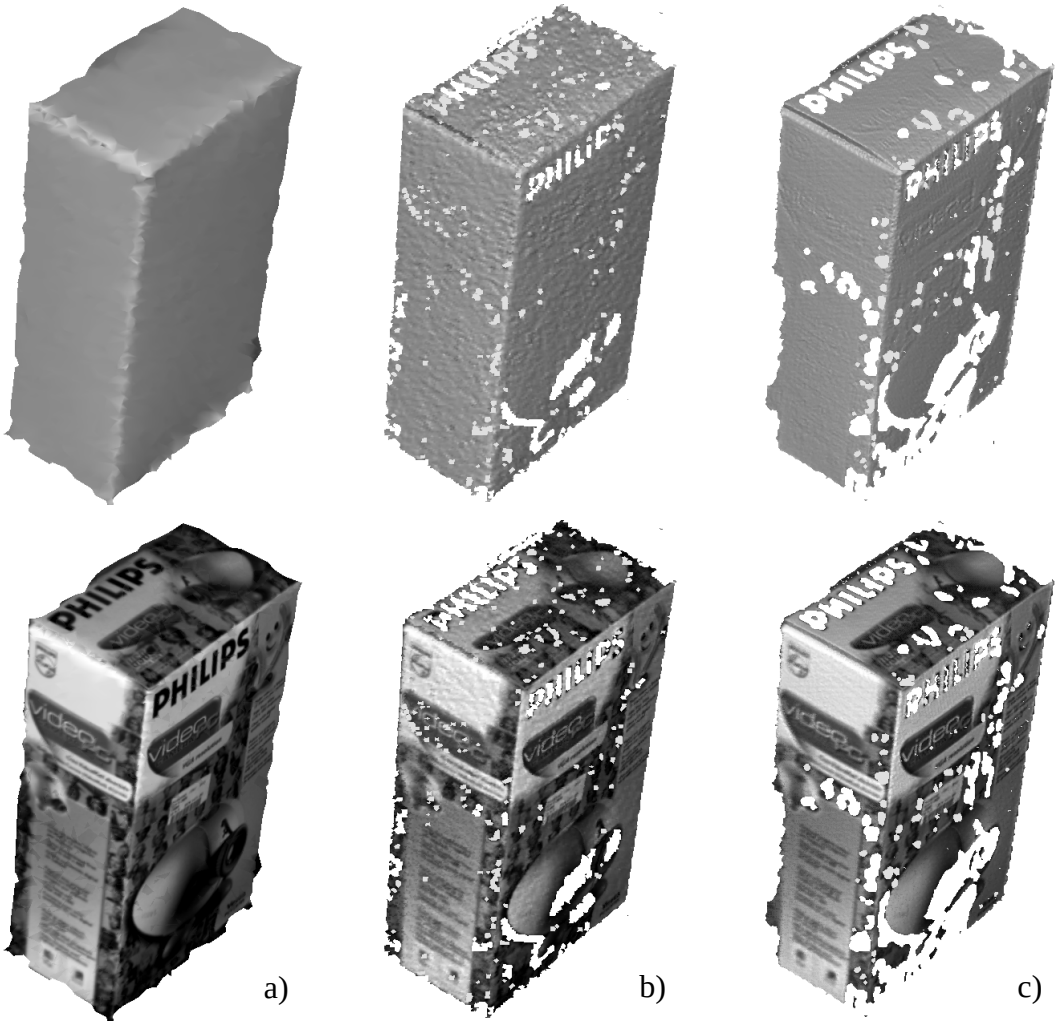


Abbildung 3.18: Rekonstruktion eines Kartons: a) Punktlaser, b) Linienlaser, c) Gray-Code Verfahren

Die Verwendung des Laserpunktes war am problemlosesten. Der helle Laserpunkt hat ausgereicht, auch die schwarz bedruckten Bereiche des Kartons zu erfassen. Zudem konnte die einfache Geometrie mit dem Laser sehr schnell ausreichend dicht erfasst werden, um die abgebildete Qualität zu erhalten. Bei der Verwendung der Laserlinie wurden in der gleichen Zeit zwar sehr viel mehr Daten aufgenommen, aufgrund der fehlenden Details des Kartons konnte die Rekonstruktion daraus aber keinen Nutzen ziehen. Zudem war die Lichtintensität der Laserlinie nicht hoch genug, um auch die dunkelsten Bereiche des Kartons zu erfassen. Bei

der Verwendung des Gray-Code Verfahrens verstärkt sich dieser Effekt noch. Die Daten sind zwar am genauesten, können aber nicht nennenswert zur Qualität des Ergebnisses beitragen. Zudem sorgen starke Farbkontraste auf der Kartonoberfläche für geringe systematische Messfehler, die auf eine fehlerhafte Phasenberechnung zurückzuführen sind. Die Auflösung der Textur ist bei allen drei Varianten identisch. Aufgrund seiner Vollständigkeit hinterlässt daher auch das Ergebnis des Laserpunktverfahrens den besten Eindruck.

Anders sieht dieser Vergleich aus, wenn Objekte mit strukturierten Oberflächen vermessen werden. Als Beispiel dient im Folgenden eine einfache Computertastatur. Wieder wird für alle Tests die Anordnung der Kameras unverändert gelassen. Die Bilder zeigen wieder die berechnete Geometrie und die texturierte Darstellung der 3D-Rekonstruktion.

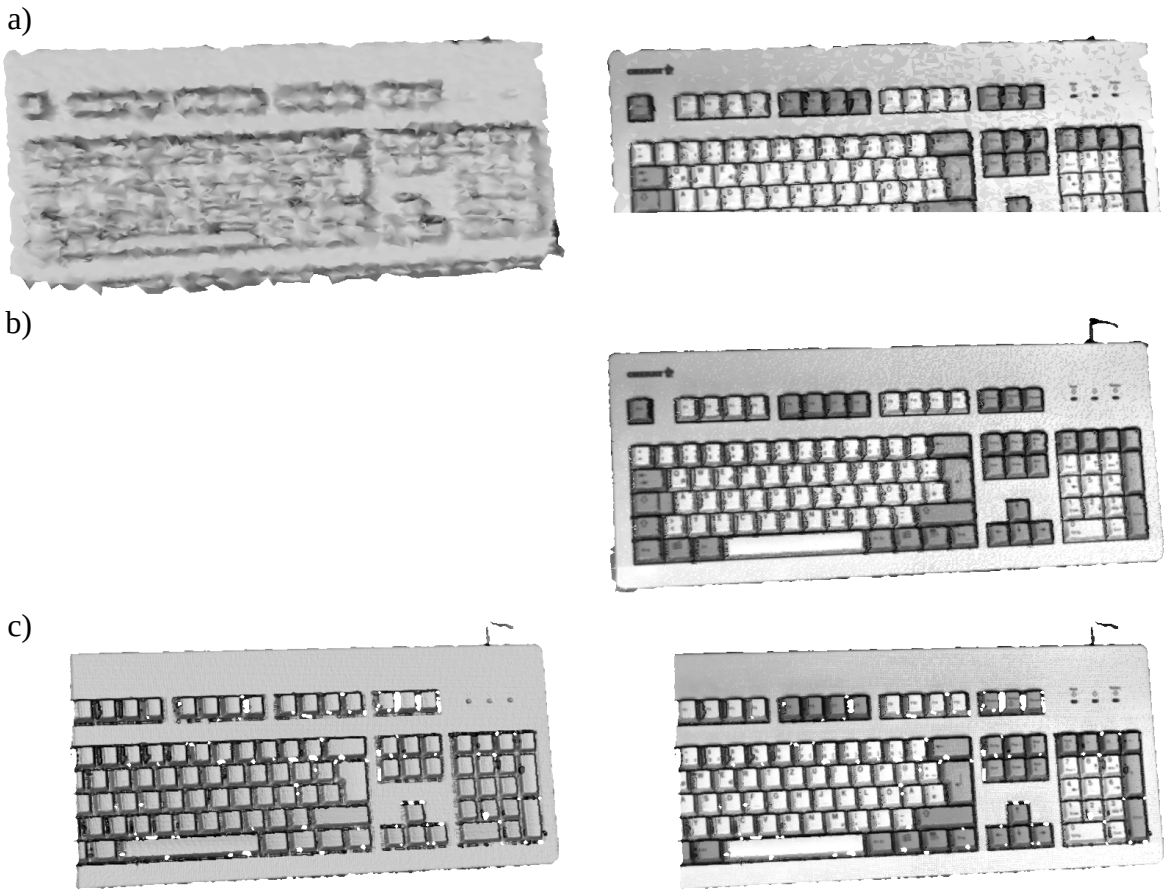


Abbildung 3.19: Rekonstruktion einer Tastatur: a) Punktlaser, b) Linienlaser, c) Gray-Code Verfahren[1]

Mit dem Laserpunkt kann das Tastenfeld der Tastatur kaum aufgelöst werden. Lediglich die Bereiche der einzelnen Tastenfelder können voneinander abgegrenzt werden. Im Vergleich dazu liefert das Gray-Code Verfahren ein exzellentes Ergebnis, bei dem sogar noch die Geometrie der Kontrollampen auf der Tastatur aufgelöst wird. Auch die Laserlinie liefert ein zufriedenstellendes Ergebnis, bei dem die Tasten noch gut zu erkennen sind. Im visuellen Vergleich der texturierten Darstellungen sind das Ergebnis der Laserlinie und des Gray-Code Ver-

1. Die Tastatur ist unvollständig, da sie nicht komplett vom Projektor ausgeleuchtet wurde

fahrens vergleichbar. Nur beim Laserpunkt kann die hochauflösende Textur die schlechte Qualität der Rekonstruktion nicht verbergen.

Der Vergleich von interner und externer Qualität ist Inhalt des nächsten Beispiels. Vor allem zur Bewertung der externen Qualität wurde wieder eine Holzplatte vermessen. Die Genauigkeit der Daten berechnet sich dann aus dem Abstand jedes Punktes zu einer Ebene, die durch die Daten approximiert wurde. Um nach Möglichkeit Sensorrauschen zu vermeiden, wurde das Gray-Code Verfahren zur Vermessung verwendet.

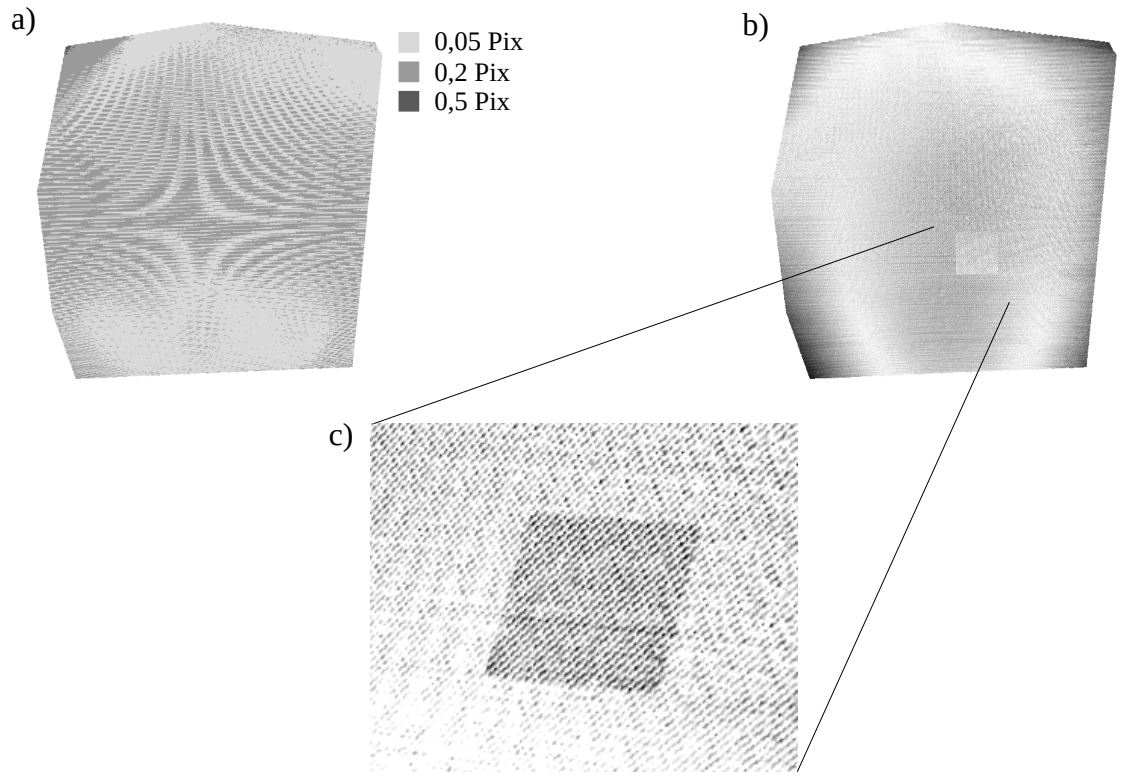


Abbildung 3.20: Rekonstruktionsqualität: a) interne Genauigkeit, b) externe Genauigkeit, c) Auflösung

In Abbildung 3.20 a) ist das Fehlerbild der internen Qualität dargestellt. Die Farbkodierung gibt dabei den Abstand jedes Korrespondenzpunktes zur Epipolarebene seines Korrespondenzpunktes an. Die mittlere Qualität der Punkte lag bei 0,1 Pixeln, lediglich im oberen rechten Bildbereich sind Punkte zu erkennen, die mit einer internen Genauigkeit von etwa 0,5 Pixeln rekonstruiert wurden. Das Fehlerbild aus Abbildung 3.20 b) zeigt die externe Genauigkeit. Weiße Bereiche haben einen Fehler von 0,0 Millimetern, bei den dunkelsten Bereichen liegt der Fehler bei maximal 0,5 Millimetern. Die mittlere Genauigkeit lag bei 0,18 Millimetern, was einer Kennzahl von 5600 entspricht. Das Fehlerbild zeigt eine deutliche Systematik, bei der der Fehler zum Rand hinzunimmt. Da die interne Genauigkeit ein sehr homogenes Bild liefert, deutet dieser Fehler auf Ungenauigkeiten des verwendeten Kameramodells hin. Obwohl natürlich eine Verbesserung des Kameramodells in Betracht gezogen werden kann, ist zu beachten, das eine Kennzahl von 5600 durchaus vergleichbar ist mit den Ergebnissen fest vorkalibrierter Gray-Code Systeme. Zur Dokumentation der Auflösung des Gray-Code Systems wurde ein Haftetikett auf die Holzplatte geklebt. Der entsprechende Ausschnitt aus dem Fehlerbild der externen Genauigkeit ist vergrößert in Abbildung 3.20 c) dargestellt. Das Hafteti-

kett hat eine Höhe von etwa 0,06 Millimetern und lässt sich noch deutlich in den Scandaten wiederfinden. Die maximale Auflösung der verwendeten Konfiguration dürfte daher bei wenigstens 0,03 Millimetern liegen.

Wie bereits in Kapitel 3.1.4 (Stereokalibrierung) erwähnt, hängt Auflösung und Genauigkeit der Daten sehr stark von der Anordnung der Kameras ab. Bei dem dokumentierten Beispielen wurde ein Basisabstand von 0,9 Metern und ein Winkel zwischen den Kameras von 41,2 Grad verwendet. Der Durchmesser der erfassten Fläche lag bei etwa einem Meter. Genau wie geringere Winkel zwischen den Kameras zu schlechterer externer Genauigkeit und Auflösung führen, könnten in dem dokumentierten Beispiel mit einem größeren Winkel bessere Ergebnisse erzielt werden. Die interne Genauigkeit ist aber jeweils unabhängig von dieser Größe und dementsprechend nur bedingt aussagekräftig für die externe Qualität der Daten.

Die weiteren Beispiele sollen unterschiedliche Aspekte der Visualisierung verdeutlichen. Wie in Kapitel 3.1.5.2 (Visualisierung) beschrieben, wird die Vollständigkeit und Qualität der Visualisierung über den Ausdünnungsfaktor und den Grad der Lückenfüllung gesteuert.

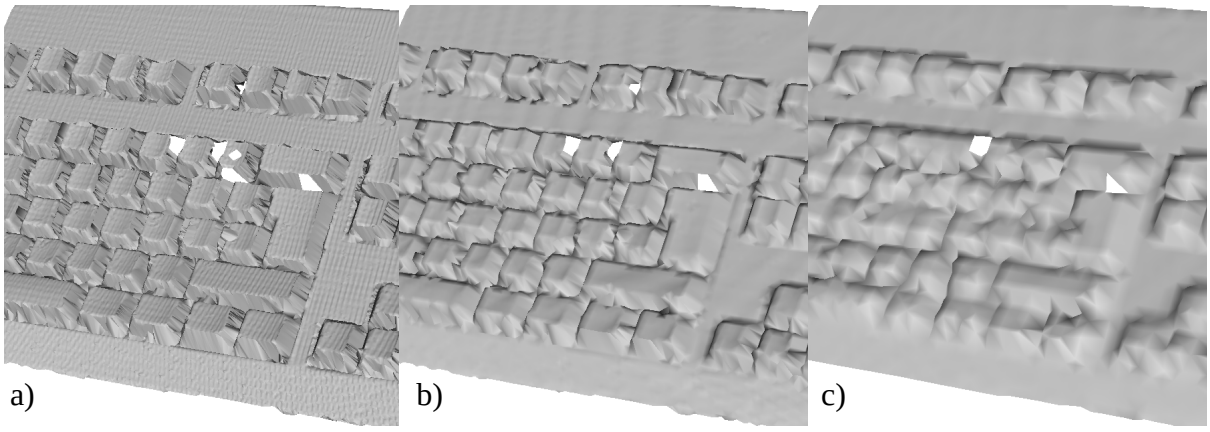


Abbildung 3.21: Unterschiedliche Ausdünnungen: a) 1 fach, b) 3 fach, c) 6 fach

Ein Ausschnitt aus der bereits bekannten Computertastatur ist in Abbildung 3.21 dargestellt. Ohne Ausdünnung wird der höchste Detaillierungsgrad der Rekonstruktion erreicht. In Abbildung 3.21 b) wurde ein horizontaler und vertikaler Ausdünnungsfaktor mit dem Wert drei eingestellt, so dass nur noch ein Neuntel der Daten zur Visualisierung verwendet wird. Die Abbildung 3.21 c) spiegelt das Ergebnis wieder, bei dem nur noch ein 36tel der Daten visualisiert wird. Die Detailauflösung nimmt zwar mit steigendem Ausdünnungsfaktor stetig ab, der Hauptzweck der Ausdünnung ist allerdings die Garantie der echtzeitfähigen Visualisierung der Scandaten auf Rechnern mit begrenzter Rechenleistung. Um einen ersten Eindruck über die bereits gescannten Flächen zu erhalten reicht es meistens vollkommen aus, den Scan-Vorgang mit einem Ausdünnungsfaktor von zwei oder drei durchzuführen. Zur genauen Beurteilung des Ergebnisses kann dann der Scan-Vorgang unterbrochen und die Ausdünnung auf einen Wert von eins gestellt werden.

Um schnell einen Eindruck über das Aussehen des rekonstruierten Objektes zu erhalten, kann der Grad der Lückenfüllung variiert werden. Beim Beginn des Startvorgangs empfiehlt sich eine Lückenfüllung vom höchsten Grad, damit sehr viele Dreiecke aus den Scandaten gebildet werden. Mit zunehmender Zahl von Scanpunkten kann der Grad der Lückenfüllung sukzessive

zurück genommen werden um Rechenzeit zu sparen. Im Beispiel aus Abbildung 3.22 wurde die Holzplatte mit 3428 Punkte aufgenommen. Der Ausdünnungsfaktor der Visualisierung lag in horizontaler und vertikaler Richtung bei einem Wert von zwei. Ohne Lückenfüllung wurden 79 Dreiecke gebildet (Abbildung 3.22 a). Lückenfüllung vom Grad 1 hat zu 3615 Dreiecken geführt, Grad 2 ergab 4963 Dreiecke (Abbildung 3.22 b). Durch Level 3 konnten 5431 Dreiecke gebildet werden. Mit Level 4 war die Darstellung lückelos und hat 5464 Dreiecke ergeben (Abbildung 3.22 c). Level drei und vier unterscheiden sich bei dem bereits fortgeschrittenen Scan-Vorgang kaum noch. Für eine weitere feinere Vermessung würde sich daher die Einstellung des Levels zwei anbieten. Anhand der Visualisierung mit Level 0 ist zu erkennen, dass es ohne Lückenfüllung sehr lange dauern kann, überhaupt ein wiedererkennbares Ergebnis zu erzielen. Bei der Verwendung von Level vier haben trotz der Verwendung einer vierfachen Ausdünnung der Daten bereits 80 Prozent der Scanpunkte an der Visualisierung teilgenommen.

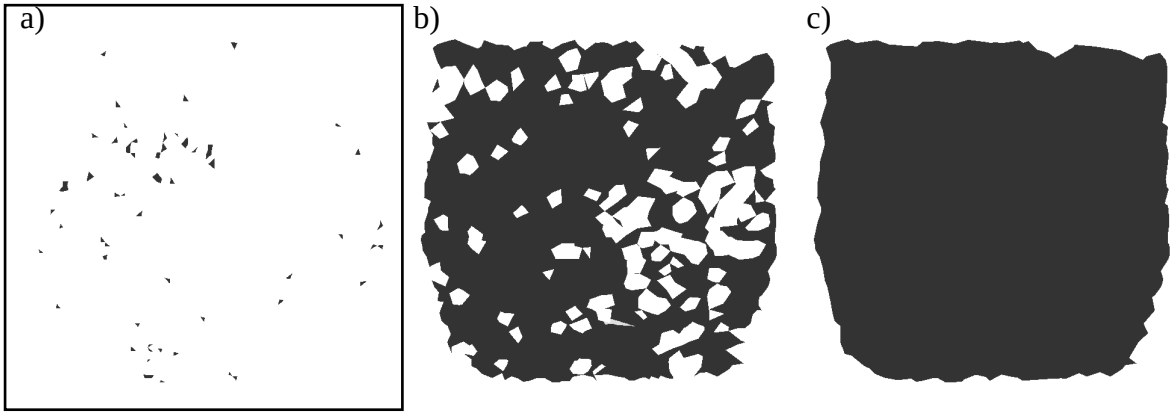


Abbildung 3.22: Unterschiedliche Grade von Lückenfüllern: a) Level 0, b) Level 2, c) Level 4

Zur Hervorhebung von Eigenschaften, zur plastischen Visualisierung oder zur realistischen Darstellung der 3D-Rekonstruktion können unterschiedliche Darstellungsattribute verwendet werden. Eine einfache schattierte Darstellung ist in Abbildung 3.23 a) dargestellt. Sie vermittelt einen guten Eindruck über die Geometrie und Plastizität des Objektes und ist die beste Variante für die Visualisierung der Rekonstruktion während des Scan-Vorgangs.

Über das Attribut Farbe können verschieden Eigenschaften des rekonstruierten Objektes hervorgehoben werden. In Abbildung 3.23 b) werden unterschiedliche Höhen in unterschiedlichen Helligkeiten eingefärbt. Während eines Scanvorgangs mit *HandyScan3D* wird Farbe hauptsächlich zur Darstellung der internen Genauigkeiten verwendet, um dem Anwender Hilfen zur Verfügung zu stellen, wo noch einmal gescannt werden sollte. Um schließlich eine photorealistische Visualisierung zu erlangen, in der auch Objekteigenschaften dargestellt werden, die nicht mit dem Scanner zu erfassen sind, wie z.B. Schriftzüge oder bunte Oberflächen, können die Bilder beider Kameras zur Texturierung der Darstellung verwendet werden. Für die Tastatur ist das Ergebnis noch einmal in Abbildung 3.23 c) zu sehen. Durch die Verwendung der Textur sind jetzt helle von dunklen Tasten zu unterscheiden und die Buchstaben auf den Tasten können erkannt werden.

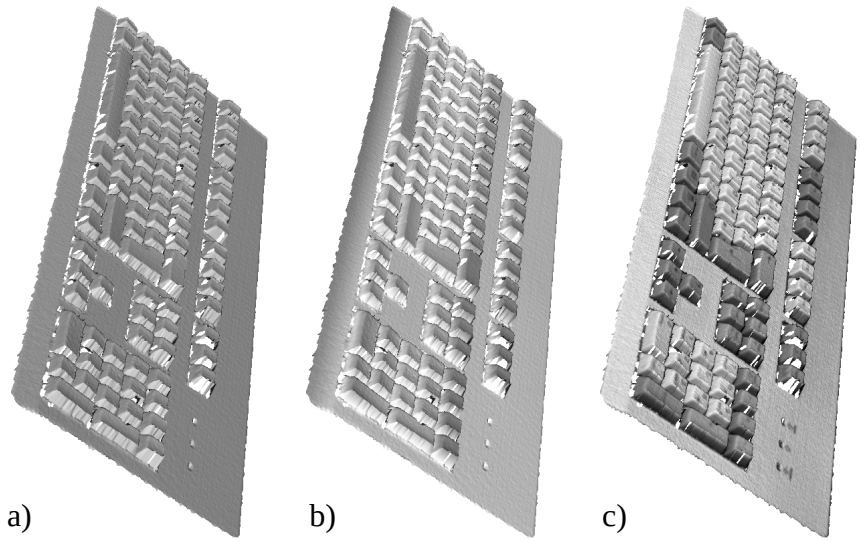


Abbildung 3.23: Unterschiedliche Darstellungsattribute: a) schattiert, b) Farbe, c) Textur

Den Abschluss von Kapitel 3.1.5 (3D-Rekonstruktion und Visualisierung) hat die Beschreibung eines Verfahrens zur Erzeugung von Mehrfachansichten durch Umstellen von Kameras gebildet. Auch wenn sich dieses Verfahren im praktischen Einsatz nur als begrenzt tauglich dargestellt hat, soll hier noch ein Beispiel für ein typisches Ergebnis dokumentiert werden. Das Scan-Objekt war eine kleine Pappschachtel mit aufgeklapptem Deckel. Der Scan-Vorgang wurde ausschließlich mit dem Laserpunkt durchgeführt.

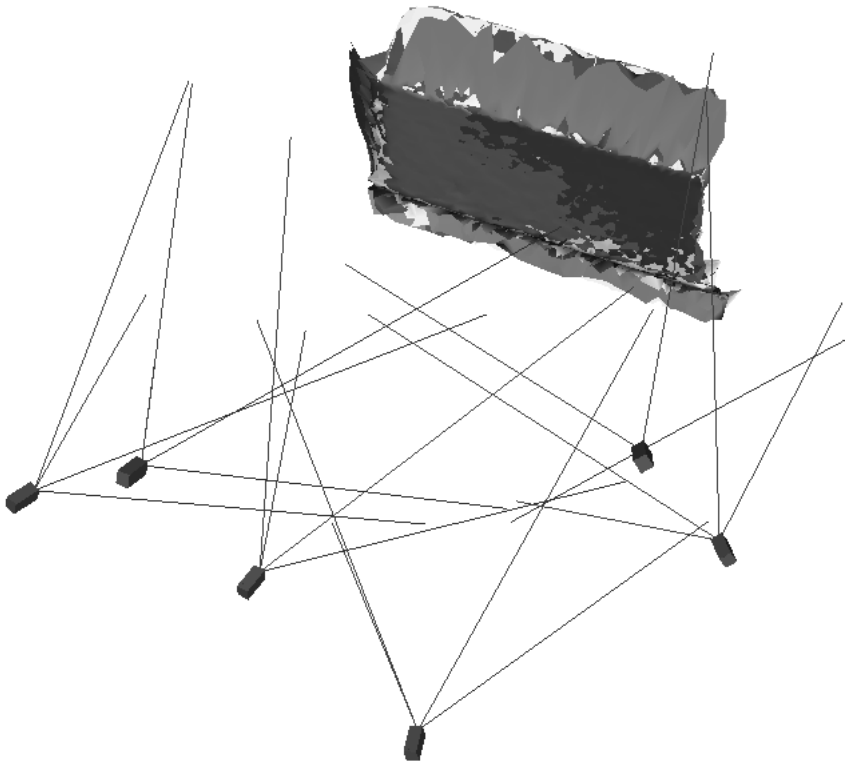


Abbildung 3.24: Erweiterte Ansichten durch Umstellen von Kameras

Neben dem Rekonstruktionsergebnis sind in Abbildung 3.24 zusätzlich die berechneten Kamerapositionen abgebildet. Da sich erst nach wechselseitigem Umstellen beider Kameras eine Erweiterung des Sichtfeldes ergibt, war es im dargestellten Beispiel erst nach dem vierten Umstellen einer Kamera möglich, die Rekonstruktion auf den Deckel und eine vierte Seite der Pappschachtel auszudehnen. Um auch noch die Rückseite der Pappschachtel zu scannen, währen wahrscheinlich weitere zehn bis fünfzehn Umstellvorgänge nötig. Um die Lage der letzten Kameraposition zu berechnen, muss dafür eine Rechnung über alle Zwischenpositionen durchgeführt werden. Für die vollständige Erfassung der Pappschachtel würden sich also die Fehler von fünfzehn Stereokalibrierungen bis zur letzten Ansicht aufsummieren.

Es wurde bereits angedeutet, dass im Rahmen des Forschungsprojektes neben der SONY XC55 Kamera noch zwei Farbkameras eingesetzt wurden. Zusätzlich wurde als Alternative zum ABW-Projektor der Einsatz eines Videobeamers untersucht. Die Ergebnisse, die sich mit diesen Systemen erzielen lassen, sind allerdings nur schwer im Rahmen dieser Dokumentation zu beschreiben. In aller Kürze besteht der große Vorteil von Farbkameras in der Farbtextur, mit der sich die Rekonstruktion visualisieren lässt. Im Vergleich zur SONY XC55 Kamera lässt sich damit noch einmal eine deutliche Aufwertung der texturierten Darstellung erzielen. Eine negative Eigenschaft von kostengünstigen Farbkameras[1] ist allerdings die niedrige Empfindlichkeit im roten Farbbereich. Die Aufteilung des CCD-Chips in 25 Prozent blaue, 25 Prozent rote und 50 Prozent grüne CCD-Elemente bei den eingesetzten Farbkameras sorgt zudem dafür, dass die Scandaten beim Einsatz eines roten Lasers sehr stark rauschen. Ein teurer grüner Laser schafft zwar Abhilfe, die Amplitude des Rauschens liegt aber dennoch um den Faktor zwei bis drei höher, als bei einem System, das aus SONY XC55 Kameras und rotem Laser besteht. Ähnlich ist das Ergebnis beim Gray-Code Verfahren, allerdings fällt der Fehler bei weitem nicht so stark ins Gewicht, da das Gray-Code Verfahren ohnehin nur wenig rauschanfällig ist. Der Videobeamer ist dem ABW-Projektor überlegen. Neben der höheren Auflösung spielt vor allem die Fähigkeit des Beamers eine wesentliche Rolle, die es ermöglicht, jedes Pixel in einer unterschiedlichen Farbe abzubilden. Für das Phasen-Shift Verfahren können sehr gute Sinusprofile erzeugt werden, auch ohne dass der Beamer unscharf gestellt werden muss. Die Daten sind daher frei von Moirée-Effekten und haben einen noch geringeren Rauschpegel als die Daten, die mit dem Projektor erzeugt werden. Die beiden wesentlichen Nachteile lagen in der Notwendigkeit einer Dual-Head Grafikkarte und der damit einhergehenden Abhängigkeit des Verfahrens von der Betriebssystemunterstützung für derartige Karten, sowie in der fehlenden Montagemöglichkeit des Beamers auf Stativen, um einen flexiblen Einsatz zu gewährleisten.

1. Kameras mit einem Farb CCD-Chip

3.2 HandyScanTT[1]

Das in diesem Kapitel eingeführte System *HandyScan3D* ist ein sehr flexibler und genauer 3D-Scanner zur Erfassung von Objekten. Der bislang noch nicht zufriedenstellend behandelte Aspekt zur Rekonstruktion von Objektgeometrien umfasst die vollständige Erfassung auch von komplexen Geometrien. Mit *HandyScan3D* ist es nur begrenzt möglich, mehrere Ansichten eines Objektes zu erfassen. Sofern erweiterte Ansichten, wie z.B. Rundumansichten, von Objekten erstellt werden müssen, sind erweiterte Systemkonzeptionen notwendig. *HandyScanTT* stellt eine mögliche Erweiterung dar, die aufbauend auf den Technologien von *HandyScan3D* die Kopplung des aktiven Stereosensors mit einem Drehteller realisiert.

3.2.1 Systemkonzeption

Zur Erfassung von Rundumansichten eines Objektes auf Basis von *HandyScan3D* muss das Objekt aus unterschiedlichen Ansichten erfasst werden. Dazu muss im Laufe der sensoriellen Erfassung die relative Lage zwischen Objekt und 3D-Sensor geändert werden. Technisch gesehen ist es dabei unerheblich, ob das Objekt oder der 3D-Sensor bewegt werden. Wichtig ist in jedem Fall, dass die relative Lageänderung zwischen Objekt und Sensor bekannt oder berechenbar ist, damit die einzelnen Aufnahmen im Rechner konsistent zusammengeführt werden können.

Die Konzeption von *HandyScanTT* sieht die Verwendung eines Drehtellers vor, mit dem das Objekt vor dem Sensor gedreht werden kann. Die Kameras des 3D-Sensors bleiben während des gesamten Scan-Vorgangs ortsfest, so dass eine einmalig berechnete Stereokalibrierung gültig bleibt. Ein weiterer konzeptioneller Schwerpunkt von *HandyScanTT* ist eine möglichst einfache Realisierung des Drehtellers. Auf eine Anbindung des Drehtellers an einen Computer wird gänzlich verzichtet, auch die Bestimmung der Lage des Drehtellers im Sensorkoordinatensystem geschieht mit einfachsten Mitteln. Zur weiteren Konzeption des Drehtellers werden noch einmal kurz die wesentlichen Merkmale von *HandyScan3D* rekapituliert, die die Spezifikation des Drehtellers beeinflussen.

Die Visualisierung und Rekonstruktion der Scandaten erfolgt in *HandyScan3D* wie beschrieben auf Basis von Rastern, in denen die aufgenommenen Scanpunkte sortiert werden. Der Aufbau eines solchen Rasters ist solange gültig, wie sich die relative Lage zwischen Objekt und Sensor nicht ändert. Es ist daher für *HandyScanTT* sinnvoll, zunächst eine vollständige Ansicht des Objektes aufzunehmen und erst dann die Lage des Objektes über den Drehteller zu verändern, anschließend kann eine weitere Ansicht des Objektes aufgenommen werden. Um eine Lageänderung des Drehtellers während der Aufnahme einer Ansicht zu vermeiden, werden Arretierungen am Drehteller vorgesehen, die die Justierung einer Drehtellerlage erlauben. Die Arretierungen befinden sich in festen Winkelinkrementen von 30 Grad am Drehteller. Die aktuelle Arretierposition muss vom Benutzer über einen entsprechenden Dialog dem Computer mitgeteilt werden. Das feste Winkelinkrement von 30 Grad stellt einen guten Kompromiss zwischen zu kleinen und zu großen Lageänderungen des Objektes dar. Zum einen wird dadurch vermieden, dass die gleiche Stelle des Objektes zu oft gescannt wird und damit ein unnötig hohes Datenaufkommen verwaltet werden muss. Zum anderen ist bei diesem Winkel-

1. HandyScan **T**urn **T**able

inkrement noch ausreichend gesichert, dass auch Hinterschneidungen auf dem Objekt erfasst werden können, vor allem vor dem Hintergrund, dass auch der optimale Winkel zwischen den Kameras des Stereosystems etwa 30 bis 45 Grad betragen sollte.

Im Weiteren muss zur konsistenten Zusammenführung der einzelnen Ansichten des Objektes bekannt sein, um welche Achse das Objekt um die definierten Winkelinkremente gedreht wurde. Zu diesem Zweck wird ein einfacher Kalibrierstab vorgesehen, der in die Drehachse des Drehtellers gesteckt werden kann. Auf dem Kalibrierstab sind wenigstens zwei Marken vorgesehen, die in den Kamerabildern des Stereosystems wiedererkannt werden können. Über die Stereokalibrierung des Sensors können dann aus den Markenpositionen in den Kamerabildern zwei Raumpunkte berechnet werden, die die Drehachse des Drehtellers im Sensorkoordinatensystem definieren. Der entwickelte Drehteller und der Kalibrierstab ist in folgender Abbildung dargestellt.

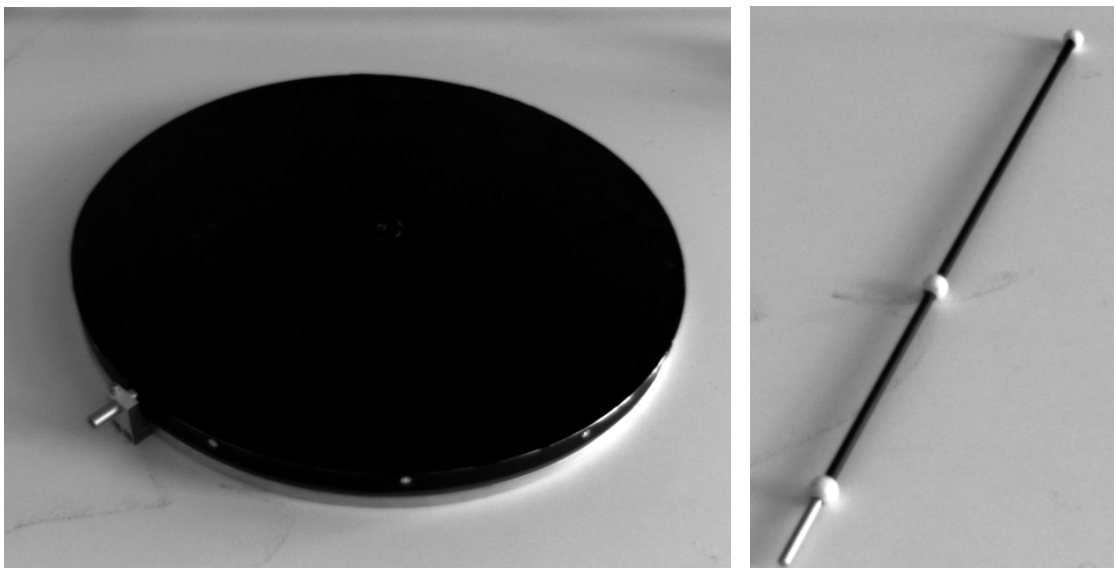


Abbildung 3.25: Drehteller und Kalibrierstab

Der Drehteller ist schwarz eingekleidet, um störende Reflexionen der Lichtquelle während der Vermessung zu vermeiden, ebenso wird dadurch erreicht, dass der Drehteller selbst nicht vom Sensor erfasst werden kann. Die Drehachse wird über ein Kugellager gehalten, das gleichzeitig zur Aufnahme des Kalibrierstabes dient. Zusätzlich ist die bewegliche obere Platte des Drehtellers über drei Räder auf der unteren Platte gelagert. Dadurch ist sichergestellt, dass auch größere oder schwere Objekte auf dem Drehteller positioniert werden können, ohne dass die obere Platte kippelt. Die Arretierungen zur Einstellung der festen Winkelinkremente sind auf der Seite der oberen Platte angebracht. Alle 30 Grad findet sich eine kleine Vertiefung in der Platte. Eine Justierung, die aus einer Kugel besteht, die von einer Feder an die obere Platte gedrückt wird, ist an der unteren Platte befestigt. Die Feder kann so eingestellt werden, dass die Arretierung sowohl einen ausreichenden Widerstand bietet, wenn die Kugel in einer Vertiefung einrastet, als auch der Drehteller aus einer Arretierung gelöst werden kann, ohne dass das Objekt auf dem Drehteller zu verrutschen droht. Der Kalibrierstab ist mit drei Marken bestückt, so dass auch bei großem Abstand der Kameras vom Drehteller der Kalibrierstab bildfüllend abgebildet wird. Die redundante Information der zusätzlichen Marke kann zur Genauigkeitssteige-

rung genutzt werden. Bei kleinen Objekten oder kurzem Abstand zwischen Kameras und Drehteller wird die Drehachse ausschließlich über die unteren Marken bestimmt.

Der Ablauf einer kompletten Rundumerfassung eines Objektes geschieht in mehreren Schritten. Zunächst werden die beiden Kameras auf den Drehteller ausgerichtet. Die Mitte des Drehtellers muss zu sehen sein, damit die spätere Drehachsbestimmung funktioniert. Im Weiteren sollte das Objekt in allen Drehtellerpositionen vollständig zu sehen sein. Zur Bestimmung der Drehachse werden von jeder Kamera zwei Aufnahmen des Drehtellers gemacht, eine mit und eine ohne eingesetztem Kalibrierstab. Über die Differenzberechnung zwischen diesen Vordergrund- und Hintergrundaufnahmen der Bilder können die Kalibriermarken im Bild wiedergefunden werden (Abbildung 3.26).

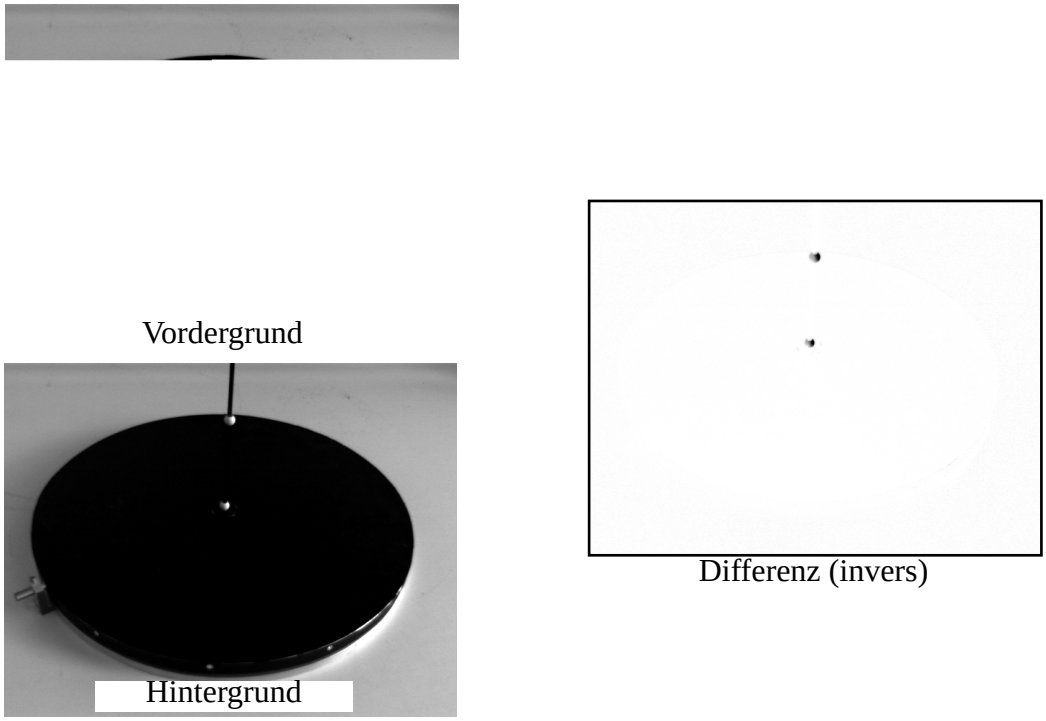


Abbildung 3.26: Markendetektion

Zur eigentlichen Erfassung kann das Objekt auf dem Drehteller plaziert und die Vermessung gestartet werden. Sobald die erste Ansicht ausreichend erfasst ist, wird die Vermessung unterbrochen und der Drehteller um eine Position weitergedreht. Bei einfachen Objekten kann der Drehteller um größere Winkelinkremente, z.B. 60 Grad, gedreht werden. Die Vermessung wird wieder gestartet, um die zweite Ansicht zu erfassen. Sobald alle Ansichten erfasst sind, werden die Daten noch einer automatischen Nachbearbeitung unterzogen, die zur Genauigkeitssteigerung und zur visuellen Aufbereitung der Daten dient.

3.2.2 Kalibrierung und 3D-Rekonstruktion

Neben der bereits beschriebenen Stereokalibrierung muss bei *HandyScanTT* noch die Lage und Orientierung der Drehachse berechnet werden. Da die Berechnung der Drehachse auf Basis der Stereokalibrierung durchgeführt wird, muss bei laufend aktualisierter Stereokalibrierung auch die Drehachse ständig korrigiert werden. Wenn eine neue verbesserte Stereokalibrierung berechnet wird, müssen also auch alle bisher aufgenommen Scandaten neu berechnet werden. Zur Erläuterung der dazu notwendigen Berechnungsvorgänge sollen zunächst die Koordinatensysteme eingeführt werden, mit denen gerechnet wird.

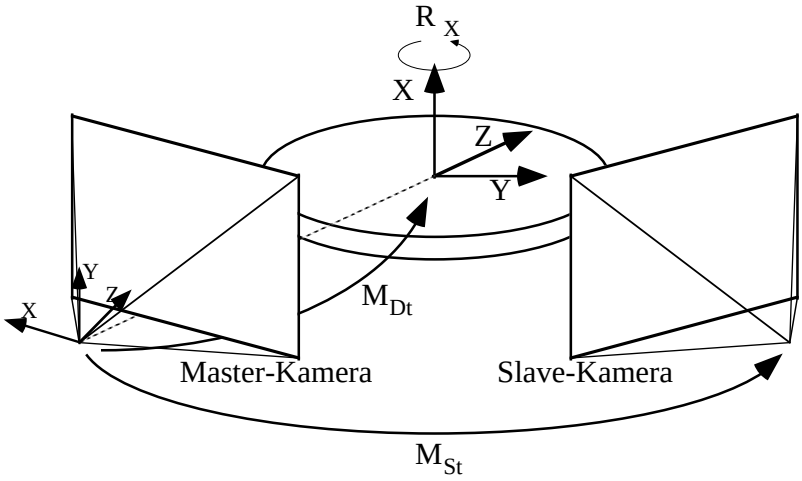


Abbildung 3.27: Koordinatensysteme von HandyScanTT

Das Stereokoordinatensystem M_{St} wird von *HandyScan3D* geliefert. Es beschreibt die relative Lage der beiden Stereokameras zueinander. Zur einfacheren Umsetzung der Drehlage des Drehtellers wird das globale Koordinatensystem in das Drehtellerzentrum gelegt, die X-Achse entspricht dabei der Drehachse des Tellers. Die Transformation M_{Dt} dient zur Abbildung des Stereokoordinatensystems in das globale Drehtellerkoordinatensystem. Die Lage einer rekonstruierten Ansicht kann dann durch eine einfache Rotation (R_X) um die globale X-Achse beschrieben werden.

Grundlage der Drehtellerkalibrierung ist die Berechnung der Drehachse. Da die Lage dieser Achse mit jeder neuen Stereokalibrierung neu berechnet werden muss, wird dafür von der einzigen fixen Information ausgegangen, den Bildkoordinaten aller Marken in den Kamerabildern, die zu Beginn der Rundumerfassung aufgenommen wurden. Mit jeder neuen Stereokalibrierung kann für jedes korrespondierende Bildpunktepaar die Lage einer Marke im Stereokoordinatensystem berechnet werden. Sofern zwei Marken erkannt wurden, bildet die Verbindungsgerade zwischen den beiden Raumkoordinaten der Marken die Drehachse. Bei mehr als zwei Marken ist die Berechnung der Drehachse überbestimmt, so dass eine Ausgleichsrechnung verwendet werden kann. Zur Berechnung der Transformation M_{Dt} muss die Lage des Drehtellers im Stereokoordinatensystem definiert werden. Die X-Achse des Drehtellers entspricht wie beschrieben der Drehachse des Tellers. Der Drehteller selbst sei koplanar zur YZ-Ebene des Drehtellerkoordinatensystems, die unterste Kalibriermarke bildet den Koordinatenursprung. Über die zusätzliche Festlegung, dass der Ursprung des Stereokoordinatensystems in der XZ-Ebene des Drehtellerkoordinatensystems liegt, ist das

Drehtellerkoordinatensystem vollständig definiert. Die Y-Achse kann über das Kreuzprodukt zwischen X-Achse des Drehtellers und dem Ortsvektor zur unteren Marke ($\vec{p}_1$) bestimmt werden, die obere Marke wird mit $\vec{p}_2$ bezeichnet:

$$\vec{X} = \vec{p}_2 - \vec{p}_1 \tag{3.21}$$

$$\vec{Y} = \vec{p}_1 \times \vec{X}. \tag{3.22}$$

Die Z-Achse wird schließlich aus dem Kreuzprodukt zwischen X-Achse und Y-Achse gebildet

$$\vec{Z} = \vec{X} \times \vec{Y}. \tag{3.23}$$

Aus den normierten Koordinatenachsen kann die Matrix berechnet werden, die die Rotationslage des Drehtellerkoordinatensystems im Stereosystem beschreibt (R_{Dt}). Die inverse Matrix führt zusammen mit dem Ortsvektor zur Marke $\vec{p}_1$ das Stereokoordinatensystem in das globale Drehtellerkoordinatensystem über und bildet die Transformation M_{Dt}

$$M_{Dt} = \begin{pmatrix} R_{Dt}^{-1} & -\vec{p}_1 \\ 0 & 1 \end{pmatrix}. \tag{3.24}$$

Jeder Scanpunkt, der im Stereokoordinatensystem aufgenommen wurde, muss über M_{Dt} in das globale Koordinatensystem überführt werden. Anschließend wird der globale Punkt über R_X der aktuellen Drehtellerposition zugeordnet und zur 3D-Rekonstruktion weitergeleitet.

Die Rekonstruktion der Daten geschieht analog zum *HandyScan3D* Verfahren. Jedem Scanpunkt werden Rasterkoordinaten zugeordnet, so dass eine Sortierung innerhalb zweidimensionaler Tabellen möglich ist. Auch im Fall des Drehtellers entstehen die Rasterkoordinaten über die Diskretisierung der Bildkoordinaten der aufgenommenen Scanpunkte in Bezug auf eine ausgezeichnete Kamera. Da sich aber im Laufe der Messung das Drehtellerkoordinatensystem relativ zum Stereokoordinatensystem bewegt, kann die Rekonstruktion nicht mehr ausschließlich über eine Tabelle geschehen. Vielmehr wird für jede neue Drehtellerposition eine neue Tabelle angelegt, in die dann die zugehörigen Scanpunkte eingetragen werden. So führt beispielsweise die Erfassung eines Objektes mit Winkelinkrementen von 60 Grad zu sechs Einzelansichten und damit zu sechs Tabellen.

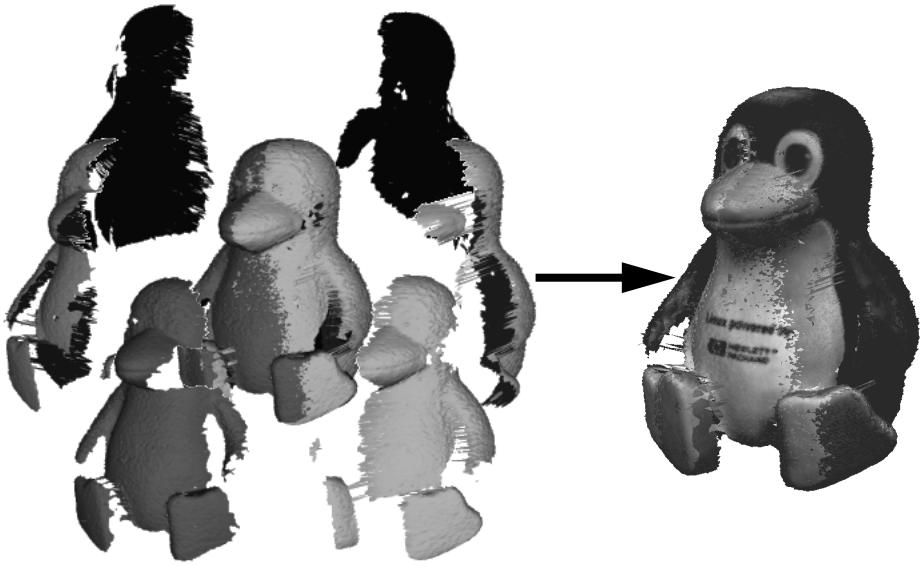


Abbildung 3.28: Rekonstruiertes Objekt

In Abbildung 3.28 ist die Rekonstruktion eines Spielzeugpinguins dargestellt. Neben den sechs aufgenommenen Einzelansichten ist der vollständig rekonstruierte Pinguin abgebildet. Abbildung 3.29 zeigt die virtuellen Kamerapaare, die während der Vermessung gebildet wurden. Im dargestellten Beispiel wurden zwei Kameras eingesetzt. Durch die Rotation des Drehtellers wird ein System simuliert, das aber zwölf Kameras enthält. Zur realitätsgetreuen Darstellung der Daten wird ein Kamerabild jeder Kamera in jeder Ansicht abgespeichert und zur Texturierung verwendet. Der einfach aufgebaute Drehteller erlaubt also die Bildung von sehr komplexen Kamerasystemen zur Rundumerfassung von Objekten.

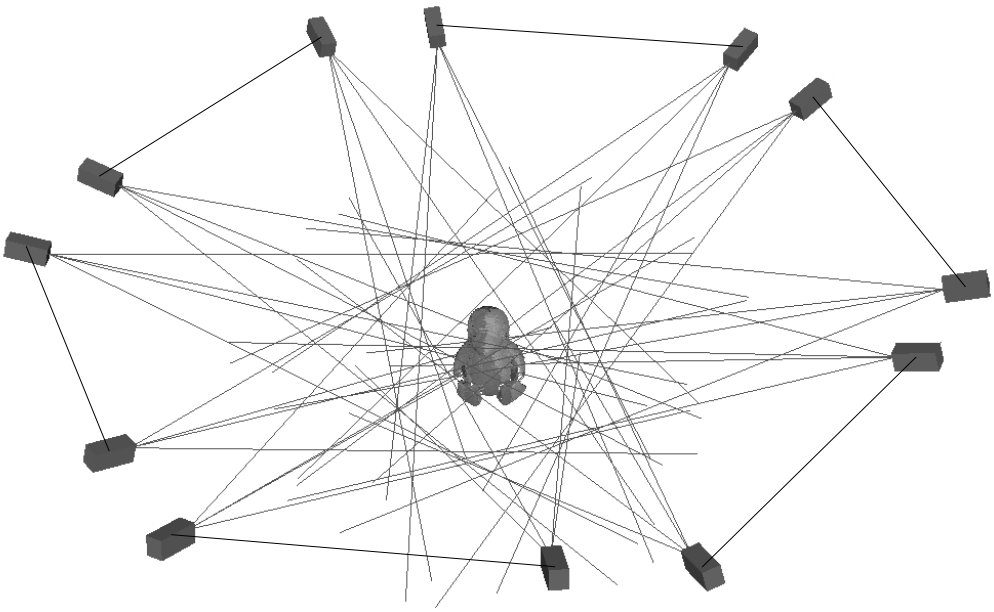


Abbildung 3.29: Virtuelle Kamerapaare

3.2.3 Verschmelzen mehrerer Ansichten

3D-Scansysteme können im Allgemeinen nur bis zu einer gewissen Grenze kalibriert werden. Je mehr Einzelkomponenten kalibriert werden müssen, um so höher ist die Unsicherheit, mit der das Scansystem kalibriert wird, da sich die Fehler der Einzelkalibrierungen potentiell summieren. In jedem Fall bestimmt aber die Qualität der schlechtesten Einzelkalibrierung die Qualität des Gesamtsystems. Bei *HandyScanTT* kommt zur Stereokalibrierung noch die Kalibrierung der Drehachse. Die Stereokalibrierung an sich kann als sehr genau angesehen werden, da durch die Vielzahl der aufgenommenen Stereokorrespondenzen und der damit erreichten hohen Redundanz eine hochwertige Kalibrierung berechnet werden kann. Die Kalibrierung der Drehachse wiederum ist von der Stereokalibrierung abhängig, zudem hängt die Genauigkeit der berechneten Drehachse von der Zahl und Qualität der Marken auf dem Kalibrierstab ab. Die Qualität der Kalibrierung von *HandyScanTT* wird also im Wesentlichen von der Drehachsbestimmung abhängen.

Andererseits ist aber besonders bei kleinen Winkelinkrementen zwischen den Einzelansichten einer Rundumvermessung zu erwarten, dass viele Bereiche des gescannten Objektes redundant in unterschiedlichen Einzelansichten erfasst werden. Die Überlappungsbereiche zwischen Einzelansichten werden bei einer optimalen Kalibrierung hundertprozentig in Deckung liegen. In der Realität klafft aber eine Lücke zwischen gemeinsamen Bereichen unterschiedlicher Ansichten. Der Abstand zwischen zwei Einzelansichten kann zur Bewertung der erreichten Kalibrierung verwendet werden, weiterhin kann die redundante Information aber auch zur Genauigkeitssteigerung und zur Verschmelzung der Einzelansichten verwendet werden.

3.2.3.1 Genauigkeitssteigerungen

Die Grundlage zur Anpassung zweier Einzelansichten aneinander sind sogenannte Korrekturvektoren, die die Klaffung zwischen zwei Ansichten beschreiben. Idealerweise wird für jeden Punkte einer Ansicht der korrespondierende Punkt der anderen Ansicht gesucht. Der Vektor zwischen diesen beiden Punkten ist der beschriebene Korrekturvektor. Normalerweise wird nur für einen Teil der Punkte ein Korrekturvektor berechenbar sein. Ziel ist also die Bestimmung einer Korrekturfunktion, mit der für jeden Scanpunkt ein Korrekturvektor berechnet werden kann.

Für die Korrektur einer Einzelansichten wird dazu eine Transformation in Form einer mehrdimensionalen polynomialen Funktion berechnet. Eingabewerte dieser Funktion sind die zweidimensionalen Bildkoordinaten eines Scanpunktes, Ausgabe ist ein Korrekturwert, der die Entfernung des Scanpunktes vom Brennpunkt der Kamera korrigiert, so dass die Bildposition des Scanpunktes unverändert bleibt. Die Berechnung der Koeffizienten der polynomialen Funktion erfolgt mittels einer Least-Square Approximation. Mit der im Folgenden beschriebenen Näherung werden zunächst für eine Vielzahl von Scanpunkten Korrekturvektoren berechnet, über die dann im Weiteren die Approximation berechnet werden kann. Sobald für jede Einzelansicht eine Korrekturfunktion vorliegt, wird die Korrektur durchgeführt.

Auch mit der Korrekturfunktion erfolgt eine genaue Anpassung der Einzelansichten erst in mehreren Iterationsschritten, aufgrund der massiven Ausnutzung der im Rechner modellierten Kameramodellen ist ein Iterationsschritt aber sehr einfach zu berechnen. Die Iteration wird solange fortgeführt, bis die Güte der Anpassung ein ausreichendes Maß erreicht hat. Üblicher-

weise terminiert das Verfahren, wenn die Klaffung zwischen den Einzelansichten der Genauigkeit des Stereosystems entspricht.

Zur Suche von korrespondierenden Punkten und zur Berechnung von Korrekturwerten werden in einer Einzelansicht eine Menge von Scanpunkten ausgewählt. Jeder Scanpunkt wird über das Koordinatensystem des Drehtellers und die Stereokoordinatensysteme in die Bildebenen der Kameras aller weiteren Ansichten abgebildet (Abbildung 3.30).

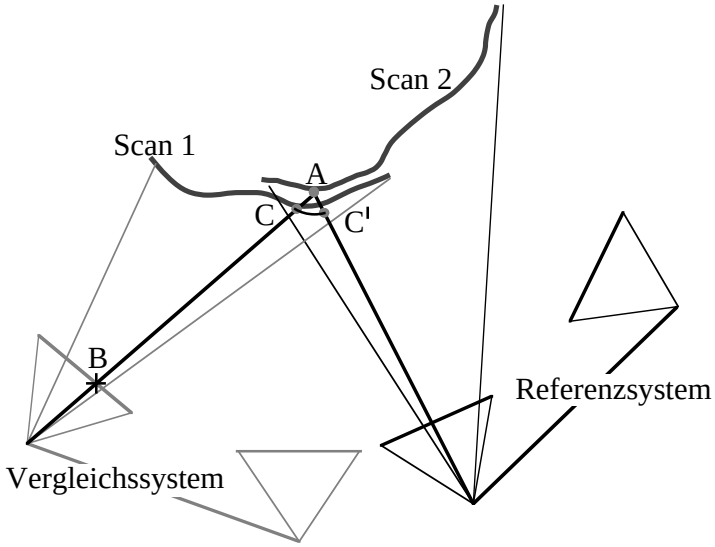


Abbildung 3.30: Berechnung von Korrekturvektoren

Ein Scanpunkt A wird also in die zweidimensionalen Bildkoordinaten B eines Vergleichssystem abgebildet. Wenn sich in der Nähe dieses Punktes Scanpunkte der Vergleichsansicht befinden, wird der interpolierte Korrespondenzpunkt C berechnet. Der Abstand zwischen A und C entspricht dem Korrekturwert, der zur Berechnung der Korrekturfunktion berechnet wird. Er beschreibt den Abstand, um den A bezüglich des zugehörigen Stereokoordinatensystems verschoben werden muss. Es ist leicht einzusehen, dass der tatsächlich verwendete Korrespondenzpunkt C' im Allgemeinen nicht mit der gescannten Fläche der Vergleichsansicht übereinstimmt. Er stellt aber eine gute Näherung dar, die genauer wird, je besser zwei Ansichten aufeinander passen. Falls mehrere Korrekturwerte in mehren Vergleichssystemen gefunden werden, wird der Mittelwert aller Korrekturwerte verwendet.

Zur Berechnung der Korrekturfunktion können die gefunden Korrekturwerte in einem zweidimensionalen Koordinatensystem angeordnet werden, das den Bildkoordinaten (x,y) der zugehörigen Scanpunkte entspricht.

Das Ziel ist im Weiteren die Berechnung einer Funktion, die eine bestmögliche Approximation der gefundenen Stützwerte beschreibt. Dabei muss der Grad der Funktion so gewählt werden, dass Krümmungen in der gesuchten Korrekturfunktion ausreichend modelliert werden können und vorhandenes Rauschen in den Stützwerten unterdrückt wird. Angesichts der Tatsache, dass

die Stützwerte besonders in den frühen Iterationsschritten der Genauigkeitssteigerung relativ ungenau sind, wurde eine Funktion mit niedrigem Grad der Form

$$f_k(x, y) = c_1x^2 + c_2y^2 + c_3xy + c_4x + c_5y + c_6 \tag{3.25}$$

gewählt. Bei einer Funktion höheren Grades besteht die Gefahr, dass Rauschen oder fehlerhafte Korrekturwerte zu starke Auswirkungen haben, zum anderen garantieren die quadratischen Terme, dass nach mehreren Iterationen auch Krümmungen hohen Grades modelliert werden können. Die Berechnung der Koeffizienten c_i erfordert das Vorhandensein von wenigstens sechs Stützpunkten. Um aber eine gute Approximation mit quadratischer Fehlerminimierung zu erhalten, empfiehlt es sich, eine sehr viel größere Zahl von Stützwerten zu suchen. In Abbildung 3.31 sind exemplarisch die berechneten Korrekturfunktion abgebildet, die aufgrund von Korrekturwerten einer Einzelansicht im ersten (a) und zweiten (b) Iterationsschritt berechnet wurde. Der dunkle Bereich zeigt die Gebiete, in dem Korrekturwerte gefunden wurden. Außerhalb dieses Bereichs werden die Korrekturwerte durch die berechnete Funktion extrapoliert. Zur besseren Darstellung der Ergebnisse ist zusätzlich die XY-Ebene dargestellt, so dass negative Korrekturwerte in der Darstellung verdeckt werden. Es ist deutlich zu erkennen, dass die Korrekturfunktion bereits nach einer Iteration deutlich abflacht. Nach einer weiteren Iteration wäre die Korrekturfunktion in der Darstellung nicht mehr von der XY-Ebene unterscheidbar.

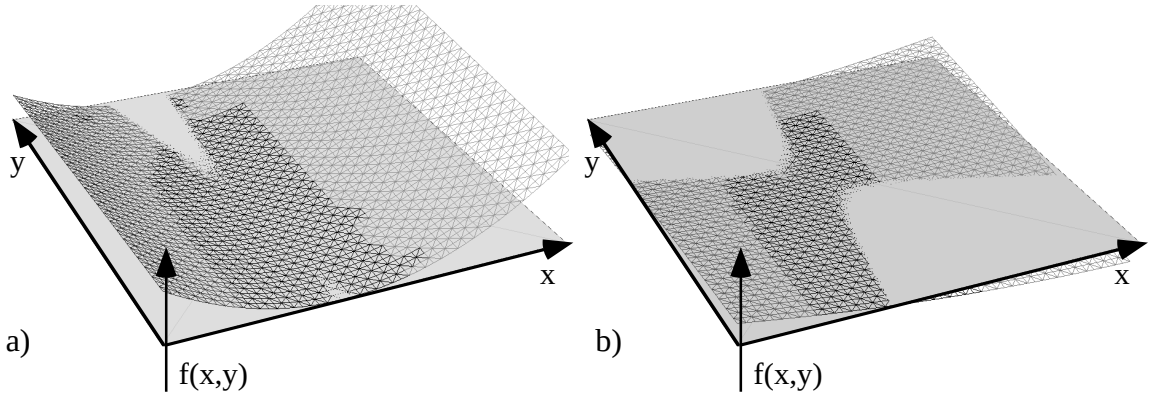


Abbildung 3.31: Berechnung der Korrekturfunktion

Sobald für jede Einzelansicht eine Korrekturfunktion berechnet wurde, müssen die Scanpunkte der Einzelansichten korrigiert werden. Es ist allerdings zu beachten, dass die direkte Anwendung des Funktionsergebnisses auf einen Scanpunkt nicht in jedem Fall sinnvoll ist. Unter der Annahme, dass genau zwei Einzelansichten in Deckung gebracht werden sollen und die Approximation ein exaktes Ergebnis darstellt, würde die Anwendung der Korrekturfunktion dazu führen, dass die Einzelansichten gerade ihre Lage tauschen und der Fehler der gleiche bleibt. Daher wird zusätzlich der Gewichtungsfaktor λ eingeführt, mit dem der Erfüllungsgrad der Korrekturfunktion gesteuert werden kann. Ein korrigierter Scanpunkt wird dann durch

$$\vec{P}' = \vec{P} + \lambda f(x, y)\,\vec{V}. \tag{3.26}$$

berechnet. Dabei bezeichnet $\vec{V}$ den normalisierten Ortsvektor des Scanpunktes $\vec{P}$ im zugehörigen Stereokoordinatensystem. Im beschriebenen Fall mit nur zwei Einzelansichten ist

$\lambda = 0{,}5$ eine optimale Wahl. Bei sehr vielen Einzelansichten hingegen führt $\lambda = 1{,}0$ zu den besten Ergebnissen. Als guter Kompromiss für alle Fälle hat sich $\lambda = 0{,}7$ erwiesen.

Mit diesem Verfahren ist es nicht möglich, lokale systematische Fehler zu korrigieren, da das Verfahren global auf den gesamten Datensatz wirkt. Praktische Versuche haben aber ergeben, dass bereits nach drei bis vier Iterationen mit dem beschriebenen Verfahren eine globale Annäherung zwischen den Einzelansichten erreicht wird, die im Bereich der Genauigkeit des Sensorsystems liegt, also nicht mehr verbessert werden kann.

3.2.3.2 Korrektur der Textur

Zur visuellen Aufbereitung der Scandaten werden neben der Oberflächengeometrie des gescannten Objektes zu jeder Einzelansicht noch Texturbilder abgespeichert. Da das Objekt für jede Einzelansicht auf dem Drehteller gedreht wird, ändert sich mit jeder Ansicht die Beleuchtung des Objektes. Bei einer insgesamt gleichmäßigen diffusen Beleuchtung ist der Effekt vernachlässigbar. Werden jedoch handelsübliche Lampen verwendet, die das Objekt einseitig beleuchten, sind deutliche Helligkeitssprünge in der texturierten Darstellung des rekonstruierten Objektes zu erkennen. Ähnliche Probleme entstehen, wenn die Blendeneinstellung beider Kameras unterschiedlich ist. In Abbildung 3.32 a) ist eine gescannte Kaffeedose abgebildet, die mit dem Gray-Code System erfasst wurde und bei der beide Probleme vorhanden sind. Der Aufdruck auf der Dose ist kaum zu erkennen, da sich unterschiedlich beleuchtete Bereiche aus verschiedenen Texturen in der Darstellung überlagern. Die Lücken in der Rekonstruktion entstehen aufgrund von Totalreflexionen der Linienmuster auf der glänzenden Dosenoberfläche.

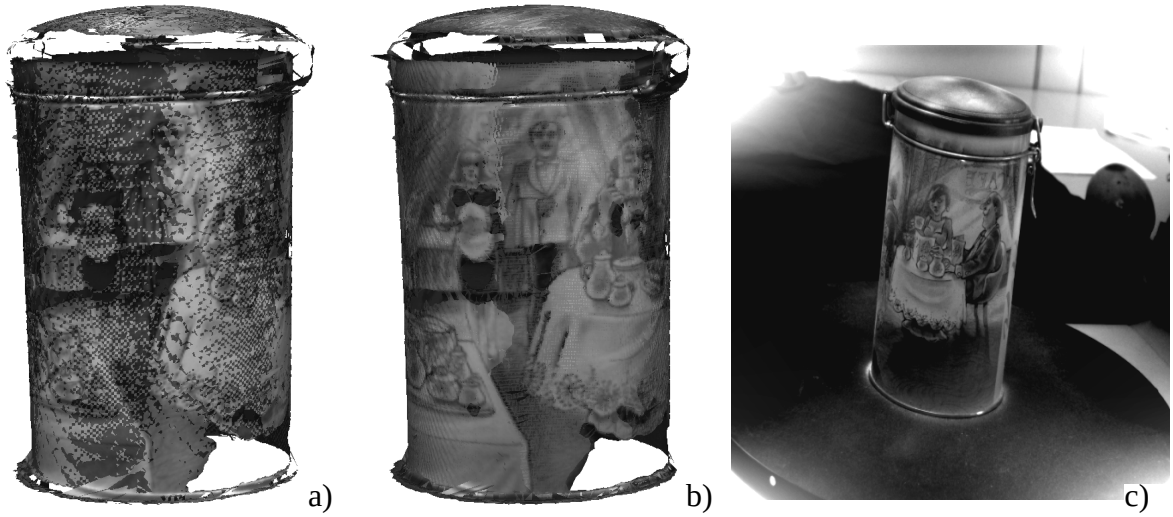


Abbildung 3.32: Beleuchtungsartefakte in texturierter Darstellung (a) und korrigierte Textur (b,c)

Die Korrektur dieser Beleuchtungsartefakte funktioniert analog zur beschriebenen Genauigkeitssteigerung. Zu einem Scanpunkt einer Einzelansicht werden alle korrespondierenden Punkte in den anderen Einzelansichten gesucht. Die Farbwerte der zugehörigen Pixel der Texturen werden gemittelt. Die Differenz dieses Farbwertes zum Farbwert des untersuchten Scanpunktes ergibt den Korrekturwert, der als Stützwert der später berechneten Korrekturfunktion dient. Sind für alle Einzelansichten ausreichend Stützwerte gefunden, werden die Koeffizienten der Korrekturfunktionen berechnet. Im Gegensatz zur Genauigkeitssteigerung muss aber

für jeden Farbkanal eine eigene Funktion berechnet werden. Zudem ist der Aufwand zur Korrektur der Texturen sehr hoch, da im Allgemeinen jedes Pixel jeder Textur auf die Korrekturfunktionen angewendet werden muss. In Abbildung 3.32 b) ist das Ergebnis einer solchen Textur dargestellt.

Das Objekt wurde in sechs Ansichten erfasst, so dass zwölf Texturen aufgenommen und abschließend korrigiert wurden. Zu sehen ist, dass die Korrektur der Texturen die Flecken beseitigt hat, die aufgrund starker Helligkeitsunterschiede in Abbildung 3.32 a) noch zu sehen waren. Der Aufdruck auf der Dose ist klar zu erkennen. Lokale Fehler, die aufgrund von Spiegelungen oder Schattenwurf entstehen, können mit dem beschriebenen Verfahren allerdings nicht behoben werden. Im abgebildeten Texturbild (Abbildung 3.32 c) sind starke Aufhellungen und Abdunkelungen besonders im Randbereich des Bildes zu erkennen, die aufgrund von extrapolierten Korrekturwerten entstehen. Da im Randbereich der Kamerabilder aber keine Scanpunkte aufgenommen wurden, wird auch keine Geometrie mit den extrapolierten Texturbereichen gebildet, diese Bereiche tauchen in der 3D-Visualisierung nicht auf und sind unproblematisch.

3.2.4 Beispiele

Die folgenden Beispiele zeigen exemplarisch die Rekonstruktionsergebnisse, die mit unterschiedlichsten Objekten erzielt wurden. Neben der bildlichen Darstellung wird auch noch einmal ein Vergleich der beiden Scan-Verfahren, Laser und Gray-Code, durchgeführt.

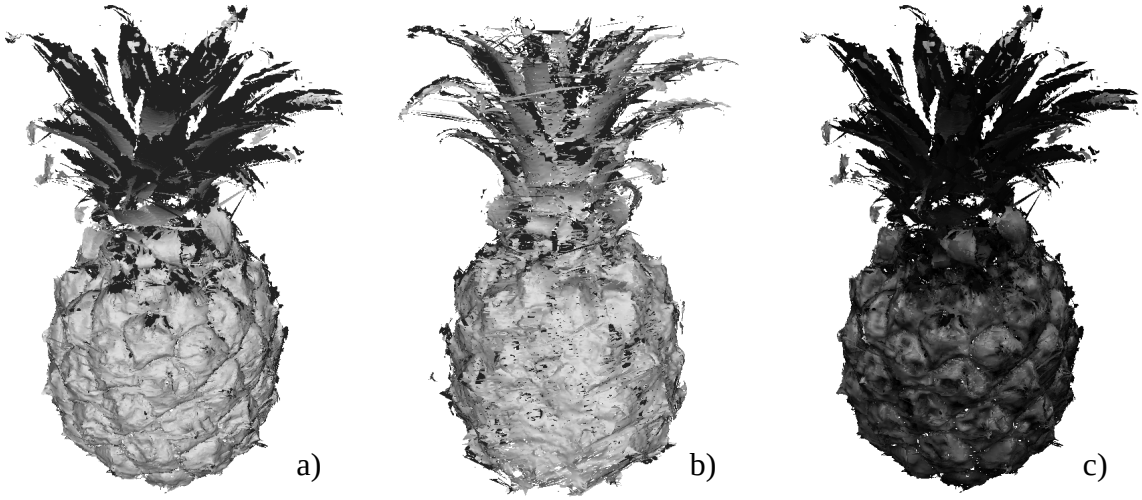


Abbildung 3.33: Ananas

Die in Abbildung 3.33 dargestellte Ananas ist das aufwändigste der hier dargestellten Beispiele. Besonders das Blattwerk der Frucht stellt eine besondere Herausforderung für ein 3D-Sensorsystem dar. Die Ananas wurde mit dem Gray-Code Verfahren in zwölf Ansichten (Abbildung 3.33 a,c) und mit der Laserlinie in sechs Ansichten (Abbildung 3.33 b) erfasst. Aufgrund der höheren Rekonstruktionsauflösung wird mit dem Gray-Code Verfahren ein deutlich höherer Detailgrad erreicht, so dass auch feine Details auf dem Fruchtkörper erkennbar sind. Bei Verwendung der Laserlinie entstehen vor allem im Blattwerk große Probleme, die zu vielen Stördaten und Fehlmessungen führen. Die meisten Fehlmessungen lagen allerdings weit

außerhalb der eigentlichen Geometrie und konnten problemlos manuell entfernt werden. Alle weiteren Darstellungsfehler sind im Wesentlichen auf das Visualisierungsverfahren zurückzuführen, mit dem im Einzelfall nur sehr schwer zu entscheiden ist, ob ein dargestelltes Dreieck sinnvoll ist. Daher tauchen vor allem wieder im Blattwerk viele Dreiecke auf, die der Darstellung abträglich sind. Die texturierte Darstellung (Abbildung 3.33 c) wurde mit dem Gray-Code Verfahren erzeugt, sie unterscheidet sich aber kaum von der entsprechenden Darstellung des Laserlinien-Verfahrens. Die erreichte Überlappung der Einzelansichten lag bei beiden Verfahren bei etwa 0,3mm, sie konnte durch das beschriebene Optimierungsverfahren nicht mehr wesentlich verbessert werden. Der Grund dafür ist wahrscheinlich darin zu suchen, dass die zerklüftete Oberfläche der Ananas die Berechnung geeigneter Korrekturvektoren erschwert.

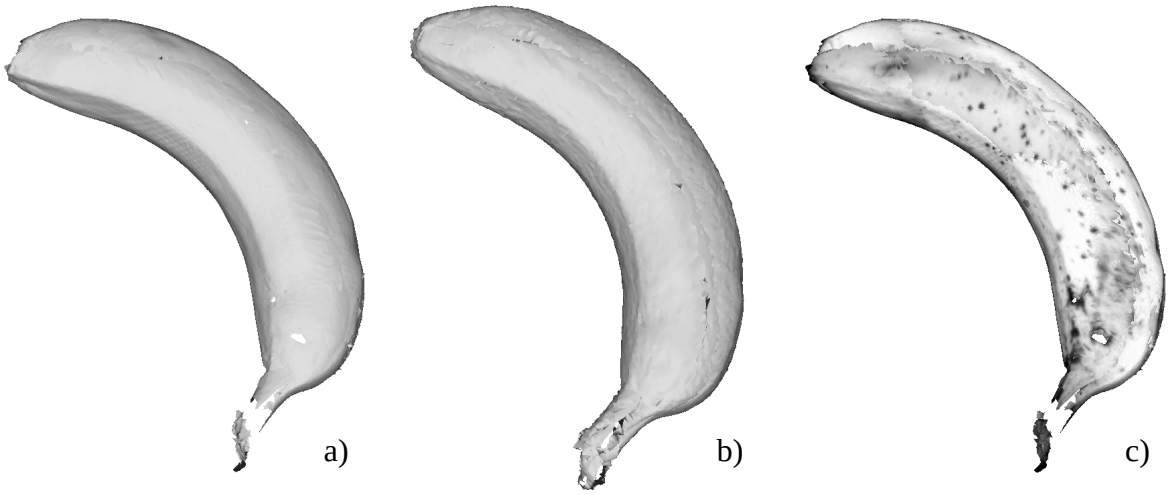


Abbildung 3.34: Banane

Im Gegensatz zur Ananas ist die in Abbildung 3.34 dargestellte Banane ein besonders einfaches Objekt für ein 3D-Sensorsystem. Sie hat eine weitgehend helle, glatte Oberfläche und eine insgesamt einfache Form. So ist es auch nicht verwunderlich, dass die Ergebnisse des Gray-Code Verfahrens (Abbildung 3.34 a,c) und der Laserlinie (Abbildung 3.34 b) fast identisch sind. Bei beiden Verfahren wurden sechs Ansichten aufgenommen. Der Laser hat gegenüber dem Gray-Code Verfahren den Vorteil, dass er unempfindlicher auf unterschiedliche Oberflächen reagiert. Der Stiel der Banane konnte mit dem Gray-Code Verfahren nicht mehr erfasst werden, war aber, genau wie ein etwas größerer Fleck auf der Bananenschale, kein Problem für den Laser. Die initiale Überlappung der Einzelansichten lag bei etwa 0,3mm. Mit dem Optimierungsverfahren wurde bereits nach zwei Iterationen eine Genauigkeit von weniger als 0,1mm erreicht, was in etwa der Sensorgenauigkeit entspricht. Die beiden folgenden Beispiele erreichen hinsichtlich der Genauigkeit wieder ähnlich gute Werte, sie dienen im Weiteren als Beispiele für die Auflösung und Vollständigkeit der Daten.

Der Pinguin in Abbildung 3.35 wurde wieder vergleichend mit dem Gray-Code Verfahren und der Laserlinie in sechs Ansichten erfasst. Bezüglich der Detailauflösung zeigt sich wieder die Überlegenheit des Gray-Code Verfahrens. Besonders im Bereich der Augen sind noch Details zu erkennen, die nur den Bruchteil eines Millimeters betragen (Abbildung 3.35 a). Derartig feine Strukturen gehen bei der Verwendung der Laserlinie im Sensorrauschen unter (Abbildung 3.35 b).

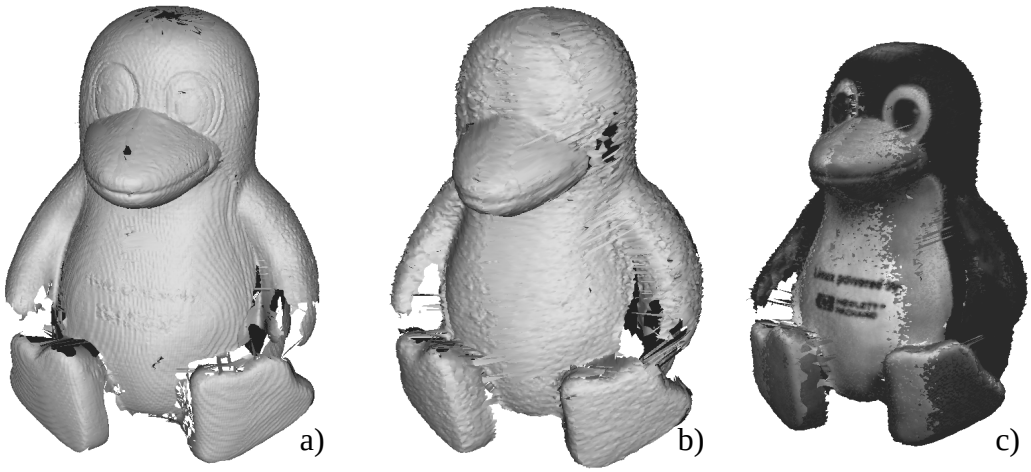


Abbildung 3.35: Pinguin

Die Daten, die mit dem Laser aufgenommen wurden, sind zwar geringfügig vollständiger, aufgrund der spiegelnden Füße des Pinguins führte die Auswertung der Laserlinie aber zu einigen falschen Korrespondenzen (im Bild nicht zu erkennen). Eine Störung, die beim Gray-Code Verfahren zu entdecken war, ist ein leichter Moirée-Effekt auf dem Bauch des Pinguins, der auf ein schlecht eingestelltes Phasen-Shift Verfahren zurückzuführen ist. Im Weiteren fällt auf, das sich auch der Aufdruck auf dem Bauch des Pinguins in den Daten des Gray-Code Verfahrens abhebt. Der starke Kontrast des Aufdrucks auf dem weißen Bauch des Pinguins führt zu einer fehlerhaften Berechnung der Phasenlage des Phasen-Shift Verfahrens und somit zu Ungenauigkeiten. Sowohl der Moirée-Effekt, als auch die Schrift auf dem Bauch führen aber zu Fehlern, die weitaus geringer als 0,1mm sind und fallen damit nicht sonderlich ins Gewicht. Die texturierte Darstellung des Pinguins auf Basis des Laser-Scans und des Gray-Code Verfahrens unterscheiden sich wiederum kaum voneinander, da alle beschriebenen Artefakte durch die Textur überdeckt werden. Kleinere geometrische Fehler fallen schlichtweg nicht mehr auf.

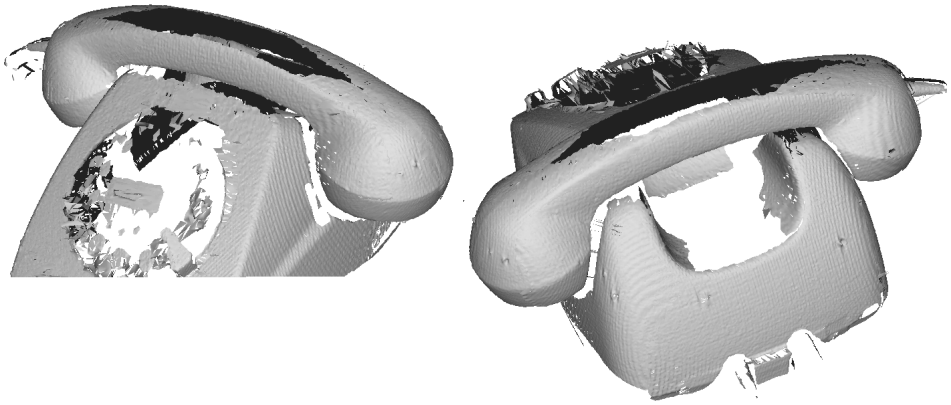


Abbildung 3.36: Telefon

Das in Abbildung 3.36 dargestellte Telefon wurde ausschließlich mit dem Gray-Code Verfahren erfasst. Sowohl die glänzende Oberfläche, als auch die transparente Wählscheibe führen beim Einsatz der Laserlinie zu massiven Störungen, die auch manuell kaum zu entfernen sind. Auf eine möglich Abhilfe, dem kompletten Einweißen des Telefons, zum Beispiel mit Trok-

kenshampoo, wurde verzichtet. Für das Gray-Code Verfahren stellt die transparente Wählscheibe ebenfalls ein Problem dar. Teilweise kommt es vor, dass von einer Kamera die Oberseite der Wählscheibe erfasst wird und von der anderen Kamera wird durch die Wählscheibe hindurch gemessen. In den meisten dieser Fälle werden keine Korrespondenzen gebildet, so dass Löcher entstehen, oft kommt es allerdings auch zu Fehlmessungen. Im Weiteren fällt auf, dass das Telefon im Bereich des Tragegriffs und des Hörers noch große Löcher aufweist. Der Sichtwinkel des Sensors auf das Telefon war zu flach, als dass die Oberfläche noch zu erkennen gewesen wäre. Ein steilerer Sichtwinkel würde zwar an diesen Stellen für Abhilfe sorgen, dafür würde es aber an anderen Stellen neue Probleme geben. So konnte die tiefe Aushöhlung des Tragegriffs nicht mehr erfasst werden. Ebenso besteht bei einem ungünstigen Winkel die Gefahr, dass das Licht des Projektors von der Front des Telefons direkt in eine Kamera reflektiert wird, so dass die Überstrahlung wieder zu Fehlmessungen oder Löchern führen kann. Die insgesamt sehr komplexe Geometrie des Telefons mit tiefen Aushöhlungen lässt sich nur sehr begrenzt mit Hilfe des Drehtellers erfassen. Ohne, dass wenigstens einmal die Ausrichtung des Sensors geändert wird, ist eine vollständige Erfassung praktisch nicht möglich. Dann stellt sich allerdings das Problem, wie die dann voneinander unabhängigen Messungen wieder zusammengerechnet werden können.

HandyScanTT ist ein insgesamt kostengünstiges Verfahren, das sehr einfach die Vermessung relativ kleiner Objekte mit geringer Komplexität zulässt. Aufwändige Objekte können nur mit Einschränkungen erfasst werden. Eine mögliche Lösung für die Vermessung großer oder komplexer Geometrien ist die Vermessung mit *HandyScanCLM*, das im nächsten Kapitel beschrieben wird.

3.3 HandyScanCLM[1]

Die bisher vorgestellten 3D-Sensoren und Sensorsysteme können durch eine Kennzahl beschrieben werden, also einer Zahl, die das Verhältnis zwischen erfassbarer Tiefe und Sensorgenauigkeit beschreibt. Kleine Objekte können mit einer hohen absoluten Genauigkeit erfasst werden, große Objekte werden mit einer entsprechend schlechteren absoluten Genauigkeit aufgenommen. Um besonders große Objekte, wie zum Beispiel Autos, mit einer Genauigkeit besser als einem zehntel Millimeter zu erfassen, wäre ein Sensor mit einer Kennzahl von 1:40.000 notwendig, vorausgesetzt die Sensorkonfiguration ist ähnlich der von *HandyScanTT*. Zur Zeit erreichen nur photogrammetrische Koordinatenmesssysteme auf Basis hochauflösender Digitalkameras eine derart hohe Genauigkeit. Allerdings ist mit diesen Systemen keine flächendekkende Vermessung möglich, da im Allgemeinen nur eine Koordinate pro Bildaufnahme gemessen werden kann und nicht mehr als ein bis zwei Koordinaten pro Sekunde erfasst werden können. Die Basiskonfiguration von *HandyScan3D* würde sich für eine derartige Aufgabe im Wesentlichen aufgrund der zu geringen Datenrate nicht eignen. Bei einer angenommenen Karosserieoberfläche von etwa 15m^2 und einer geforderten Abtastdichte von einem Punkt pro Quadratmillimeter müssten 15 Millionen Punkte erfasst werden. Für *HandyScan3D* würde das eine Scandauer von etwa sechs Tagen erfordern. Mit Coded-Light Systemen werden zwar aus-

1. HandyScan **C**oded **L**ight **M**arker

reichend hohe Datenraten erreicht, allerdings ist die Genauigkeit dieser Systeme mit einer Kennzahl von üblicherweise 1:5.000 nicht hoch genug. Zudem ist es schwierig, Flächen auszuleuchten, die größer als ein Quadratmeter sind.

Die Vermessung von großen Objekten erfordert also kombinierte Verfahren, die unterschiedliche Konzepte miteinander verbinden, um sowohl eine hohe Datenrate als auch eine hohe absolute Genauigkeit zu erreichen. Ein bekanntes Verfahren [Bergmann97] erfordert die Präparierung des Messobjektes mit kodierten Marken. Die Marken werden über photogrammmetrische Koordinatenmesssysteme innerhalb eines globalen Koordinatensystems lokalisiert. Anschließend wird mit einem Coded-Light System die stückweise Vermessung der Objektoberfläche durchgeführt. Jede Einzelaufnahme, die wenigstens drei kodierte Marken enthält, kann innerhalb des globalen Koordinatensystems registriert werden. Um eine ausreichend hohe Genauigkeit zu erreichen, wird mit einer Einzelaufnahme ein Bereich mit einer Breite und Höhe von maximal einem halben Meter aufgenommen. Für die Registrierung von etwa 180 Einzelaufnahmen müssen daher bei der Vermessung eines Autos 375 kodierte Marken auf der Karosserie angebracht und lokalisiert werden [Hightech00]. Ein weiteres Problem dieses Verfahrens besteht darin, dass die Marken aus einem matten, schwarzen Material bestehen. Die Wiedererkennung der Marken in einer Einzelaufnahme geschieht dann in der Analyse der Fehlstellen des Scans, da die Marken selbst nicht vom Sensor erfasst werden. Die Scandaten werden also an den Stellen der Karosserie, auf denen Marken angebracht waren, Lücken aufweisen. Sofern aus irgendwelchen Gründen die Oberfläche des Objektes nicht mit Marken präpariert werden darf, kann dieses Verfahren grundsätzlich nicht angewendet werden.

3.3.1 Systemkonzeption

Die Konzeption von *HandyScanCLM* sieht ein Verfahren vor, das dem beschriebenen Verfahren mit kodierten Marken grundsätzlich ähnelt, aber die beschriebenen Nachteile vermeidet und vor allem darauf ausgelegt ist, dass es einfach und ohne Expertenwissen angewendet werden kann. Auch *HandyScanCLM* profitiert dabei besonders von den Technologien des *HandyScan3D*.

Das Hauptkonzept von *HandyScanCLM* besteht in der Verwendung von Marken zur Registrierung von Einzelansichten innerhalb eines globalen Koordinatensystems. Im Gegensatz zu bekannten Verfahren muss das Messobjekt nicht präpariert werden. Die Marken bestehen aus Lichtmustern, die auf das Messobjekt projiziert werden. Die Erfassung einer Einzelansicht geschieht mit einem 3D-Sensor nach dem Codierten-Licht Ansatz, so dass der eingesetzte Projektor auch zur Projektion der Marken verwendet werden kann. Eine Einzelansicht kann dann innerhalb eines globalen Koordinatensystems registriert werden, wenn wenigstens von drei Marken die Lage sowohl innerhalb der Einzelmessung als auch im globalen System bekannt ist. Bei mehr als drei Marken kann die Redundanz zur Genauigkeitssteigerung verwendet werden.

Die Bestimmung der Lage einer Marke innerhalb einer Einzelansicht ist trivial. Insbesondere dann, wenn die Marken direkt mit dem verwendeten Projektionssystem erzeugt wurden, reicht der Vergleich der Markenpositionen in den Bildern der eingesetzten Kameras mit der an der gleichen Stelle erfassten Objektgeometrie. Für die Bestimmung der globalen Lage einer Marke werden feststehende Ortungskameras eingesetzt, deren Position innerhalb des globalen Koordinatensystems entweder vorab bekannt ist oder während der Messung über die in

HandyScan3D beschriebenden Verfahren bestimmt wird. Diese Ortungskameras haben ein größeres Sichtfeld als der 3D-Sensor und sie dienen ausschließlich dazu, die Ortsbestimmung der Marken jeder Einzelmessung durchzuführen. Über die Ortungskameras wird zusätzlich der verfügbare Messraum konfiguriert. Durch eine hohe Zahl von Ortungskameras kann eine hohe Genauigkeit der Markenortung erreicht werden. Werden die Ortungskameras weit auseinander gestellt, kann ein großes Volumen durchmessen werden.

Die Genauigkeit von *HandyScanCLM* wird von beiden Größen bestimmt, der Genauigkeit des 3D-Sensors und der Genauigkeit, mit der die Registrierung einer Einzelmessung erfolgt. Die erste Größe wird ausschließlich durch die Konfiguration des eingesetzten 3D-Sensors bestimmt. Die Registrierung einer Einzelmessung wird von mehreren Größen dominiert. Als erstes ist dabei die Zahl der Ortungskameras zu nennen, mit der eine Marke geortet wird. Für die eindeutige Lagebestimmung einer Marke sind wenigstens zwei Ortungskameras notwendig. Je mehr Ortungskameras eine Marke erfassen, um so höher ist die Redundanz und folglich die Genauigkeit, mit der diese Marke bestimmt wird. Als zweites müssen für die Berechnung der Lagetransformation einer Einzelmessung in das globale Koordinatensystem wenigstens drei Marken geortet werden. Auch hier gilt, dass mehr Marken für Redundanz sorgen, mit der die Genauigkeit der Registrierung gesteigert werden kann. Im Weiteren ist es natürlich sinnvoll, die Ortungskameras gleichmäßig zu verteilen, so dass alle Bereiche des Messraumes in der gleichen Qualität erfasst werden können. Genauso sollte die Ausprägung der projizierten Marken der Registrierung zuträglich sein. Sowohl die Lagebestimmung innerhalb einer Einzelmessung als auch die globale Ortung muss zuverlässig gewährleistet sein. Der Vollständigkeit halber sei hier erwähnt, dass die Registrierung einer Einzelmessung grundsätzlich auch mit nur einer Ortungskamera durchgeführt werden kann. Allerdings bestimmt dann die Genauigkeit der Ortungskamera die Genauigkeit der Registrierung, die aufgrund des größeren Bildfeldes im Allgemeinen deutlich schlechter sein wird als die Genauigkeit der Einzelmessung.

Die für *HandyScanCLM* realisierten Konfigurationen bestehen aus einem *HandyScan3D*-Sensor, der sich aus zwei Kameras und einem Projektor zusammensetzt. Für die Markenortung wurden im Weiteren Konfigurationen mit zwei bis vier CCD-Kameras getestet, aber auch Konfigurationen mit mehr als vier Kameras sind denkbar. Neben der normalen Mustersequenz, bestehend aus Gray-Code und Phasensequenzen, werden mit dem Projektor zusätzlich sechs horizontale und sechs vertikale Linien projiziert. Die 36 Kreuzungspunkte der Linien bilden die beschriebenen Marken. Die vorherige Analyse des Messraumes bezüglich Gradienten und Abschattungen erlaubt die Projektion der Markenlinien vorzugsweise in die Gebiete, in denen die Marken mit hoher Sicherheit detektiert werden können. Sollten dennoch zu wenig Marken geortet werden, wird die gesamte Einzelmessung verworfen.

Der Ablauf einer Messung mit *HandyScanCLM* sieht dann folgendermaßen aus:

1. Ausrichtung der Ortungskameras auf das Messobjekt, so dass die meisten Oberflächenpunkte des Objektes von wenigstens zwei Ortungskameras eingesehen werden.
2. Durchführung einer Einzelmessung mit dem 3D-Sensor, bestehend aus Gray-Code und Phasensequenzen.
3. Analyse der Einzelmessung bezüglich Oberflächengradienten und Fehlstellen mit anschließender angepasster Projektion der Markenlinien mit dem 3D-Sensor.

4. Suche der Kreuzungspunkte der Markenlinien in der Einzelmessung und in den Bildern der Ortungskameras.
5. Sofern die Lage der Ortungskameras noch nicht bekannt ist, werden die Marken, die von mehr als einer Ortungskamera erfasst werden, zur Stereokalibrierung der Ortungskameras verwendet.
6. Registrierung der Einzelmessung über die Berechnung der Lagetransformation zwischen den Markenpositionen der Einzelmessung und den zugehörigen globalen Markenpositionen, die über die Kalibrierinformation der Ortungskameras berechnet werden.
7. Wenn noch weitere Bereiche des Objektes erfasst werden sollen, fortfahren mit 2., ansonsten Ende.

Die Algorithmen zur Berechnung der Oberflächengeometrie innerhalb einer Einzelmessung wurden bereits in Kapitel 3.1 (HandyScan3D) beschrieben. Auch die Bildverarbeitung zur Extraktion der Bildpositionen der Marken entspricht im Wesentlichen den bereits beschriebenen Verfahren und soll hier nicht weiter erläutert werden. Die wesentliche Neuerung von *HandyScanCLM* im Rahmen dieser Arbeit stellt die Kalibrierung der Ortungskameras und die Registrierung der Einzelmessungen dar und ist Inhalt des nächsten Abschnitts.

3.3.2 Kalibrierung und Registrierung

Während die Stereokalibrierung von *HandyScan3D* mit nur zwei Kameras durchgeführt wird, muss im Rahmen der Kalibrierung von *HandyScanCLM* die Lagebestimmung für potentiell beliebig viele Kameras gelöst werden. Dabei wird zur Durchführung der Kalibrierung des Gesamtsystems in folgende vier Varianten unterschieden:

1. Die Lage der Ortungskameras wird mit Kalibrierkörpern bestimmt und der 3D-Sensor ist fest vorkalibriert.

   Diese Variante bietet die höchste Genauigkeit, ist andererseits aber auch am inflexibelsten. Als Kalibrierkörper kann beispielsweise ein Gestell mit Marken bekannter Position verwendet werden, das den gesamten Messraum durchsetzt. Die Lagebestimmung der Ortungskameras kann dann zum Beispiel nach dem in Kapitel 2.1 (Kamerakalibrierung) beschriebenen Verfahren durchgeführt werden. Auch für den 3D-Sensor kann eines der allgemein bekannten Kalibrierverfahren verwendet werden. Nachteilig ist die fehlende Flexibilität dieser Konzeption. Es müssen Kalibrierkörper eingesetzt werden, die etwa so groß sein sollten, wie das zu vermessende Objekt. Die Ortungskameras müssen fest montiert werden, da jede Lageänderung zwischen der Kalibrierung und der eigentlichen Vermessung die Genauigkeit des Gesamtsystems zerstört. Folglich muss auch die Konfiguration des Gesamtsystems sehr genau geplant werden, da eine nachträgliche Änderung während der Vermessung kaum noch möglich ist. Der Einsatz von Kalibrierkörpern erlaubt aber andererseits eine sehr genaue und robuste Kalibrierung. Die Gefahr von Fehlerfortpflanzungen über mehrere Teilsysteme hinweg besteht nicht, da jede Kamera unabhängig von allen anderen Kameras kalibriert werden kann.

2. Die Lage der Ortungskameras wird mit Kalibrierkörpern bestimmt und der 3D-Sensor arbeitet selbstkalibrierend

   Die meisten Argumente zur ersten Variante gelten auch hier. Der Vorteil, besteht in der zusätzlichen Flexibilität des 3D-Sensors. Mit jeder neuen Einzelmessung kann der 3D-Sensor individuell auf die neue Messsituation eingestellt werden. Soll beispielsweise eine Region mit tiefen Einhöhlungen vermessen werden, so können die Kameras des 3D-Sensors dicht am Projektor plaziert werden, um Abschattungen zu vermeiden. Soll eine ebene Oberfläche mit feinen Details vermessen werden, so können die Kameras mit einem großen Basisabstand zum Projektor plaziert werden, um eine hohe Auflösung zu erreichen. Bei wenig markanten Oberflächen kann der 3D-Sensor auf ein großes Bildfeld eingestellt werden, sofern eine hohe Auflösung nicht wichtig ist.

3. Die Lage der Ortungskameras wird selbstkalibrierend bestimmt und der 3D-Sensor ist vorkalibriert.

   Während der Aufnahme von Einzelansichten werden die projizierten Marken, die von wenigstens zwei Ortungskameras erfasst werden, zur stereoskopischen Kalibrierung dieser Ortungskameras verwendet. Die Skalierung des Systems der Ortungskameras wird aus den Einzelmessungen abgeleitet. Insgesamt kann diese Konfiguration als guter Kompromiss angesehen werden. Der vorkalibrierte 3D-Sensor garantiert eine hohe lokale Genauigkeit der Einzelansichten, die Ortungskameras können flexibel um das Messobjekt plaziert werden. Wird eine Lücke im System der Ortungskameras entdeckt, so kann auch noch während der Objektvermessung problemlos eine weitere Ortungskamera eingemessen werden.

4. Das Gesamtsystem, bestehend aus Ortungskameras und 3D-Sensor arbeitet selbstkalibrierend.

   Mathematisch gesehen existiert bezüglich der Kalibrierung grundsätzlich kein echter Unterschied mehr zwischen den Kameras des 3D-Sensors und den Ortungskameras. Die Ortungskameras werden um das Messobjekt plaziert, die Kameras des 3D-Sensors und der Projektor werden für jede neue Einzelmessung neu ausgerichtet. Für die Kalibrierung des Gesamtsystems könnten nun auch die Kameras des 3D-Sensors wie Ortungskameras behandelt werden. Mit jeder Einzelmessung würden dann aber zwei neue Kameras in das Gesamtsystem aufgenommen werden. Besonders bei vielen Einzelmessungen steigt dadurch schnell die Komplexität der Berechnungen zur Kalibrierung des Gesamtsystems. Daher wird der 3D-Sensor nach wie vor als Einheit betrachtet. Da aber sowohl dem 3D-Sensor und den Einzelansichten, als auch dem System der Ortungskameras der Skalierungsfaktor fehlt, muss dann für die Registrierung einer Einzelansicht neben der Lagetransformation noch eine Relativskalierung für die individuelle Einzelansicht berechnet werden. Neben der hohen Flexibilität zeichnet sich diese Variante auch durch eine besonders einfache Handhabung aus. Die Kameras des 3D-Sensors und der Projektor werden für jede Einzelansicht unabhängig voneinander auf das Objekt ausgerichtet. Der Messraum kann durch Hinzunahme neuer Kameras oder Umstellen einzelner Kameras jederzeit vergrößert oder neu konfiguriert werden. Großes Manko ist aber die fehlende Gesamtskalierung des Systems.

Schwerpunkt der folgenden Ausführungen ist die vierte Variante, da sie den allgemeinsten Ansatz darstellt und auch die wesentlichen Konzepte der ersten drei Varianten beinhaltet. Die Varianten 1 bis 3 stellen gegenüber der vierten Variante jeweils einen unterschiedlichem Grad der Spezialisierung dar, bei denen jeweils eine selbstkalibrierende Komponente durch eine vorkalibrierte Komponente ersetzt wird. Die kalibrierkörpergestützte Kalibrierung eines 3D-

Sensors oder Sensorsystems ist aber allgemeiner Stand der Technik und soll an dieser Stelle nicht weiter vertieft werden.

Im Folgenden wird also vorausgesetzt, dass unskalierte Einzelansichten vorliegen. Innerhalb jeder Einzelansicht sind die lokalen Koordinaten von mehreren Marken bekannt. Zu jeder dieser Marken existieren die Bildkoordinaten in jeweils wenigstens zwei Ortungskameras. A priori werden wie bei *HandyScan3D* nur die internen Parameter der Kameras als bekannt vorausgesetzt. Das erste zu lösende Problem besteht in der Lagebestimmung der Ortungskameras. Ist die Lage der Ortungskameras bekannt, so muss die Lagetransformation für die bisher nicht registrierten Einzelansichten, bestehend aus drei Translations-, drei Rotationsparametern und einem Skalierungsfaktor berechnet werden.

3.3.2.1 Kalibrierung der Ortungskameras

Die Lageberechnung der Ortungskameras geschieht in mehreren Schritten, die angefangen mit einem Startsystem, dass aus zwei Kameras besteht, zunehmend die Lage weiterer Kameras innerhalb des Gesamtsystems berechnen. Um eine Visualisierung der Ergebnisse nach jeder Einzelaufnahme zu ermöglichen, wird nach jeder Aufnahme bestimmt, welche Kameras kalibriert werden können. Kameras, die aufgrund einer vorhergehenden Einzelaufnahme als kalibrierbar gekennzeichnet waren, sind auch beim nächsten Schritt kalibrierbar. Daher brauchen zunächst nur die Kameras betrachtet werden, die bisher nicht kalibrierbar waren.

Zur Erzeugung eines Startsystems wird aus der Gesamtmenge der gefundenen Kameras eine Kamera ausgezeichnet, die als Referenzsystem für alle weiteren Kameras gilt. Grundsätzlich gibt es mehrere Möglichkeiten für die Wahl dieser Kamera. Die Gesamtzahl der Stereokorrespondenzen der Referenzkameras mit allen anderen Kameras ist dabei eine eher schlechte Heuristik, da sie nicht die Kalibrierbarkeit der anderen Kameras sichert. Ein Fall, der beispielsweise dann auftritt, wenn die Referenzkameras mit keiner anderen Kamera die zur Stereokalibrierung notwendigen Zahl von Korrespondenzen gemeinsam hat. Die hier verwendete Heuristik betrachtet nur die maximale Zahl von Korrespondenzen, die zwei Kameras gemeinsam haben. Die beiden Kameras mit den meisten gemeinsamen Korrespondenzen bilden das Startsystem, die Kamera dieses Startsystems die die meisten Korrespondenzen mit allen anderen Kameras hat, definiert im Speziellen die Referenzkamera, auf die sich alle weiteren Kameras beziehen.

Ausgehend von der Referenzkamera wird ein System von kalibrierten Kameras aufgebaut, das sukzessive um weitere Kameras erweitert wird, bis keine weitere kalibrierbare Kamera mehr bestimmt werden kann. Dabei wird jeweils die Kamera als nächste in das bestehende System aufgenommen, die die meisten Korrespondenzen mit allen kalibrierten Kameras gemein hat. Für die Berechnung der relativen Lage zweier Kameras zueinander werden die in Kapitel 2.4 (Stereokalibrierung) beschriebenen Verfahren verwendet. Insbesondere wird auf diese Weise die Lage der beiden Kameras des Startsystems berechnet. Da das Gesamtsystem bis auf eine Skalierung bestimmt ist, wird der Abstand zwischen den beiden Kameras des Startsystems auf den Wert Eins festgelegt. Die Berechnung der relativen Lage einer neuen Kamera (j) zum bestehenden System geschieht in zwei Schritten. Im ersten Schritt wird für jede kalibrierte Kamera (i), mit der die neue Kamera ausreichend Korrespondenzen gemein hat, eine Stereokalibrierung berechnet, die die relative Lage in Form von Rotation und Translation $(R, t)_{i-j}$ zwischen beiden Kameras beschreibt. Sofern nur eine kalibrierte Kamera gefunden wurde,

kann keine Initialschätzung für die Länge des unskalierten Translationsvektor berechnet werden. Die Translation ist demnach nur als grobe Schätzung zu verstehen. Wenn mehr als ein Stereokalibrierung zwischen der neuen und allen kalibrierten Kameras berechnet werden konnte, kann über alle Rotationen und Translationen eine Kreuzpeilung durchgeführt werden, die zu einer genauen Schätzung der Lage der neuen Kamera führt.

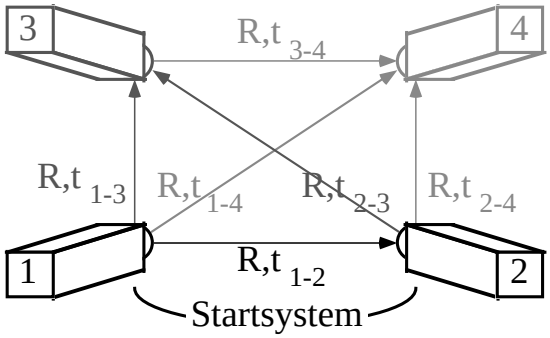


Abbildung 3.37: Startwertberechnung bei der Mehrkamerakalibrierung

Im zweiten Schritt wird die genaue Lage der neuen Kamera mittels nichtlinearer Optimierung bestimmt. Das Optimierungskriterium ist dabei der windschiefe Abstand zwischen den Projektionsstrahlen der verwendeten Korrespondenzen. Eine Korrespondenz, die in mehr als zwei Kameras vorhanden ist, wird dabei proportional stärker gewichtet. Um dem Gesamtsystem Rechnung zu tragen, werden bei der ersten Optimierung die Lageparameter der neuen Kamera berechnet. Danach wird für jede bereits kalibrierte Kamera eine erneute Lageberechnung durchgeführt, die die Korrespondenzen mit der neuen Kamera einschließt.

Grundsätzlich kann diese Iteration über die Kameras des bestehenden Systems beliebig oft durchgeführt werden. Praktische Versuche zeigen aber, dass bereits nach einer Iteration kein nenneswerter Fortschritt mehr zu erzielen ist. Im Weiteren werden mit jeder neuen Einzelaufnahme des 3D-Sensors neue Korrespondenzen der Kalibrierung der Ortungskameras zugeführt. Mit jedem Satz neuer Korrespondenzen wird die Kalibrierung der Ortungskameras entsprechend dem oben beschriebenen mehrschrittigen Verfahren erneut durchgeführt. Die bisherige Kalibrierung wird dabei als Startwert der neuen Kalibrierung verwendet, so dass eine schnelle Konvergenz der Optimierungsverfahren gewährleistet ist. Zudem wird die Qualität der Kalibrierung mit zunehmender Zahl von Korrespondenzen ständig verbessert. Kameras, die erst aufgrund der neuen Korrespondenzen kalibrierbar sind, werden automatisch in das Gesamtsystem aufgenommen.

Im Einzelnen sehen die Schritte zusammengefasst wie folgt aus:

1. Bestimmung des Startsystems und Auszeichnung der Referenzkamera,
2. Berechnung der Lage der Kameras des Startsystems mittels Stereokalibrierung,
3. Suche nach einer neuen, noch nicht kalibrierten Kamera, die ausreichend viele Korrespondenzen mit allen kalibrierten Kameras gemein hat,
4. Bestimmung einer ersten Lageschätzung der neuen Kamera durch paarweise Stereokalibrierung mit den kalibrierten Kameras und Kreuzpeilung,

5. nichtlineare Optimierung der Lage der neuen Kamera im aktuellen Gesamtsystem,
6. nichtlineare Optimierung der Lage jeder Kamera im aktuellen Gesamtsystem,
7. Wiederholung der Schritte 3bis 6, bis keine weitere Kamera mehr kalibriert werden kann.

Erwähnt sei, dass auch die gleichzeitige Berechnung der Lageparameter aller Kameras mit einem nichtlinearen Optimierungsverfahren denkbar ist. Praktische Versuche zeigen aber, dass auch bei sehr guten Startwerten nur selten eine gute Konvergenz des Verfahrens erreicht wird. Zusätzlich ist die Komplexität des verwendeten nichtlinearen Optimierers quadratisch von der Zahl der Parameter abhängig, so dass besonders bei vielen Kameras der Aufwand zur Berechnung der Kalibrierung enorm ansteigt. Messbar bessere Ergebnisse lassen sich aber dennoch nicht erreichen.

3.3.2.2 Registrierung von Einzelansichten

Die Kalibrierung der Ortungskameras ermöglicht die Berechnung von globalen Koordinaten für jede Marke. Dazu wird über alle Kameras, von der die Marke erfasst wurde, ein Projektionsstrahl berechnet. Gesucht wird nun der Punkt, dessen Summe aller quadratischen Abstände zu den Strahlen am geringsten ist.

Sei $\vec{T}_i$ die Position einer Kamera, sowie R_i die Orientierung dieser Kamera, dann lässt sich der Projektionsstrahl p_{ij} einer Marke m_j mit den normierten und unverzerrten Bildkoordinaten $\vec{B}_j$ darstellen in der Form:

$$p_{ij}(\lambda) = \vec{T}_i + \lambda \vec{v}_{ij}, \tag{3.27}$$

$$\text{mit } \vec{v}_{ij} = R_i \vec{B}_j \tag{3.28}$$

Sofern die Marke m_j auf dem Projektionsstrahl liegt, dann existiert genau ein λ, für das die Gleichung erfüllt ist. Sofern mehr als ein Projektionsstrahl vorhanden ist, entsteht für die Berechnung der globalen Markenposition ein System von Gleichungen der Form:

$$\begin{pmatrix} M_{jx} \\ M_{jy} \\ M_{jz} \end{pmatrix} = \begin{pmatrix} T_{ix} \\ T_{iy} \\ T_{iz} \end{pmatrix} + \lambda_{ij} \begin{pmatrix} v_{ijx} \\ v_{ijy} \\ v_{ijz} \end{pmatrix} \tag{3.29}$$

oder

$$\begin{pmatrix} M_{jx} \\ M_{jy} \\ M_{jz} \end{pmatrix} - \lambda_{ij} \begin{pmatrix} v_{ijx} \\ v_{ijy} \\ v_{ijz} \end{pmatrix} = \begin{pmatrix} T_{ix} \\ T_{iy} \\ T_{iz} \end{pmatrix}. \tag{3.30}$$

Neben den drei Koordinaten der Markenpositionen geht in das Gleichungssystem noch die Entfernung λ_{ij} der Marke vom Ursprung jeder Kamera ein. Sofern eine Marke in wenigstens

zwei Kameras liegt, ist das Gleichungssystem mit fünf Unbekannten und sechs Gleichungen bereits überbestimmt. Mit jeder weiteren Kamera wird das Gleichungssystem um eine Unbekannte und drei Gleichungen erweitert, so dass sich ein System mit n Kameras zur Berechnung der Marke m_j darstellen lässt als

$$\begin{pmatrix} -v_{1jx} & 0 & \dots & 0 & 1 & 0 & 0 \\ -v_{1jy} & 0 & \dots & 0 & 0 & 1 & 0 \\ -v_{1jz} & 0 & \dots & 0 & 0 & 0 & 1 \\ 0 & -v_{2jx} & \dots & 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -v_{njx} & 1 & 0 & 0 \\ 0 & 0 & \dots & -v_{njy} & 0 & 1 & 0 \\ 0 & 0 & \dots & -v_{njz} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \dots \\ \lambda_{nj} \\ M_{jx} \\ M_{jy} \\ M_{jz} \end{pmatrix} = \begin{pmatrix} T_{1x} \\ T_{1y} \\ T_{1z} \\ T_{2x} \\ \dots \\ T_{nx} \\ T_{ny} \\ T_{nz} \end{pmatrix}. \tag{3.31}$$

Die Lösung dieses Systems lässt sich mit Hilfe bekannter Methoden, wie zum Beispiel Least-Square Verfahren berechnen. Über die λ_{ij} ist es im Weiteren einfach möglich, den nächsten Punkt zur berechneten Marke auf dem Projektionsstrahl zu finden, so dass über die Summe der Abstände dieser Punkte zu der Markenposition einfach eine Markenqualität berechnet werden kann.

Der zweite Schritt zur Registrierung der Einzelansichten besteht in der Berechnung der Transformation (M_k), die die Einzelansicht k aus dem Sensorkoordinatensystem in das globale Koordinatensystem überführt. In anderen Worten die Berechnung der Transformation, die alle sensorlokalen Markenkoordinaten $\hat{m}_j$ einer Einzelansicht in die zugehörigen globalen Markenkoordinaten m_j überführt:

$$m_j = M_k \bullet \hat{m}_j . \tag{3.32}$$

Sofern die Skalierung des 3D-Sensors und des Systems der Ortungskameras bekannt ist, hat die Matrix M_k die Form

$$M_k = \begin{bmatrix} R & \vec{T} \\ 0 & 1 \end{bmatrix}. \tag{3.33}$$

Wenn entweder vom System der Ortungskameras oder vom 3D-Sensor die Skalierung nicht bekannt ist, dann wird zusätzlich die Skalierung als Unbekannte in die Matrix aufgenommen:

$$M_k = \begin{bmatrix} R & \vec{T} \\ 0 & s \end{bmatrix}. \tag{3.34}$$

Aufgrund des Skalierungsfaktors lässt sich das Gleichungssystem (3.32) nur selten stabil mit einem linearen Verfahren lösen. Daher werden zunächst Startwerte berechnet, die anschließend mit einem nichtlinearen Optimierungsverfahren verbessert werden. Der Vergleich zwischen den relativen Abständen gleicher Marken im lokalen und im globalen Koordinatensystem führt direkt zum Skalierungsfaktor. Die Schwerpunkte der skalierten Marken werden auf den gleichen Koordinatenursprung transformiert, so dass eine reine Rotation berechnet werden kann. Die ermittelten sieben Parameter, bestehend aus einem dreidimensionalen Translationsvektor, drei Rotationswinkeln und einem Skalierungsvektor bilden dann den Startwert der nichtlinearen Optimierung, die auf Basis dieser Parameter den Abstand zwischen transformierten lokalen Marken und den zugehörigen globalen Marken minimiert.

Bei der Aufnahme mehrerer unskalierter Einzelansichten fällt für jede Einzelansicht ein Skalierungsfaktor an. Sofern die Konfiguration des 3D-Sensors während den Aufnahmen nicht geändert wurde, muss der Skalierungsfaktor s für alle Einzelansichten gleich sein. Die berechneten Skalierungsfaktoren der Einzelansichten werden dann gemittelt. Anderenfalls sind die Skalierungsfaktoren voneinander unabhängig, und für jede Einzelansicht wird eine neue Skalierung berechnet. In welchem Zusammenhang die Skalierungsfaktoren der Einzelansichten zueinander stehen und im Rahmen des Gesamtsystems zu interpretieren sind, hängt im Einzelfall von der verwendeten Auslegung des Gesamtsystems ab. Bezogen auf die beschriebenen vier Varianten zur Kalibrierung des Gesamtsystem heißt das im einzelnen:

1. Wenn Ortungskameras und 3D-Sensor vorkalibriert sind, muss die Skalierung jeweils den Wert eins ergeben. Ein Skalierungsfaktor ungleich eins deutet auf eine Instabilität bei der Registrierung oder auf eine falsch detektierte Marke hin.
2. Wenn nur der 3D-Sensor selbstkalibrierend arbeitet und die Ortungskameras vorkalibriert sind, definiert der Skalierungsfaktor direkt den Abbildungsmaßstab des 3D-Sensors.
3. Wenn nur die Ortungskameras selbstkalibrierend arbeiten und der 3D-Sensor vorkalibriert ist, wird mit dem Skalierungsfaktor der Maßstab der Ortungskameras bestimmt.
4. Wenn Ortungskameras und 3D-Sensor selbstkalibrierend arbeiten, kann grundsätzlich keine Aussage über die Skalierung des Gesamtsystems gemacht werden. Der Vergleich der Skalierungsfaktoren für jede Einzelmessungen lässt aber Aussagen über Instabilitäten bei der Registrierung oder falsch detektierte Marken zu.

3.3.3 Beispiele

Im Folgenden werden einige Ergebnisse dargestellt, die mit *HandyScanCLM* erzielt wurden. Der eingesetzte 3D-Sensor besteht aus zwei Kameras des Typs SONY XC55, die an einen Framegrabber vom Typ Eltec PCEye4 angeschlossen sind. Als Projektionssystem wird wahlweise ein Projektor vom Typ ABW320 oder ein Videobeamer vom Typ Philips Hopper XG20 eingesetzt. Der 3D-Sensor arbeitet in allen Beispielen selbstkalibrierend. Die Ortungskameras sind ebenfalls vom Typ SONY XC55. Das System der Ortungskameras ist in allen Fällen ebenfalls initial unkalibriert.

Neben der visuellen Darstellung wird zusätzlich auf die erreichte Genauigkeit der Daten eingegangen. Dabei wird zwischen den folgenden Kenngrößen unterschieden:

- Die *Genauigkeit der Einzelmessung* gibt die Genauigkeit des 3D-Sensors an. Zur Ermittlung dieser Zahl wird mit dem 3D-Sensor eine möglichst quadratische, ebene Fläche gescannt. Durch die gewonnenen Daten wird dann eine Ebene approximiert. Der mittlere Abstand der Scanpunkte zur Ebene im Verhältnis zur Ausdehnung der Scandaten ergibt die Kennzahl des 3D-Sensors. Zusätzlich wird noch der mittlere Abstand zur Ebene als absolute Genauigkeit angegeben.
- Die *globale Markengenauigkeit* beschreibt den mittleren Abstand einer Marke zu den Projektionsstrahlen, mit denen die Marke berechnet wurde. Die Einheit dieser Größe wird bei skalierten Systemen in Millimetern angegeben. Bei unskalierten Systemen wird eine Schätzung der Skalierung durchgeführt, die ebenfalls in Millimeter angegebenen Werte können daher von den tatsächlichen Größen abweichen.
- Die *Genauigkeit der Registrierung* beschreibt den mittleren Abstand, den eine lokale Marke nach der Registrierungstransformation von der zugehörigen globalen Marke hat. Die Maßeinheit wird wiederum in Millimetern angegeben. Bei unskalierten Systemen wird wieder eine Schätzung durchgeführt.

Die *Genauigkeit der Einzelmessung* ist die einzige Größe, die eine externe Genauigkeit darstellt. Die beiden anderen Größen werden ausschließlich durch interne Größen bestimmt. Sie sind nicht direkt auf absolute Zahlen abbildbar und können daher nur eine Vorstellung der erreichten Qualität geben. Die *globale Genauigkeit der Marken* ist im Wesentlichen ein Maß für die Qualität, mit der die Ortungskameras kalibriert wurden. Über die *Genauigkeit der Registrierung* schließlich werden die beiden ersten Kenngrößen zueinander in Beziehung gesetzt. Über diese Größe kann festgestellt werden, ob sich die Fehler der einzelnen Kalibrierungen addiert oder ausgelöscht haben. Eine weitere interessante Größe als Maß für die globale Genauigkeit, ist der Abstand, den zwei Einzelansichten in ihren Überlappungsbereichen haben. Im allgemeinen Fall ist dieser Abstand allerdings nur aufwändig zu bestimmen, besonders, wenn zerklüftete oder feinstrukturierte Oberflächen vorliegen. Daher wurde nur in den ersten beiden Beispielen eine Berechnung durchgeführt, in dem eine ebene Fläche vermessen wurde. Die Berechnung der Genauigkeit geschieht dann durch die Approximation einer Ebene durch die Scandaten und die Messung des Abstandes jedes Scanpunktes zur Ebene.

Vermessung einer ebenen Fläche mit zwei Ortungskameras
Die einfachste Variante zur Vermessung mit *HandyScanCLM* besteht aus zwei Ortungskameras und zwei Messkameras. Im nachfolgenden Beispiel wurden fünf überlappende Aufnahmen einer ebenen Holzplatte durchgeführt. In Abbildung 3.38 ist eine Übersicht zur durchgeführten Messung dargestellt.

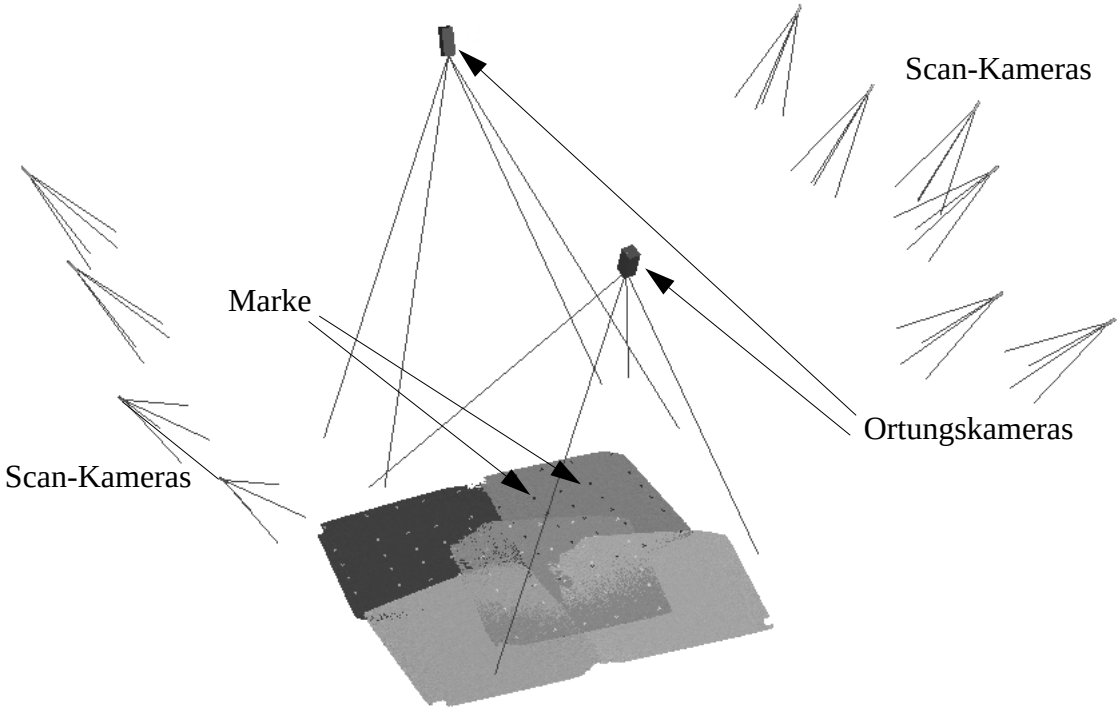


Abbildung 3.38: Ebene Fläche mit zwei Ortungskameras

Die beiden hervorgehoben dargestellten Kameras entsprechen den Ortungskameras, von der linken Seite wurden zwei, von der rechten Seite wurden drei Einzelmessungen durchgeführt, die Scan-Kameras sind jeweils klein dargestellt. Scan-Kameras und Projektor waren auf einer festen Halterung montiert, bei der der Projektor in der Mitte der Kameras positioniert war. Der Basisabstand zwischen den Scan-Kameras lag bei etwa 70 Zentimetern. Die erfasste Gesamtfläche hatte einen Durchmesser von etwa 80 Zentimetern, mit einer Einzelmessung konnten etwas mehr als 40 Zentimetern erfasst werden. Die erreichte Genauigkeit ist in Abbildung 3.39 und Abbildung 3.40 dargestellt.

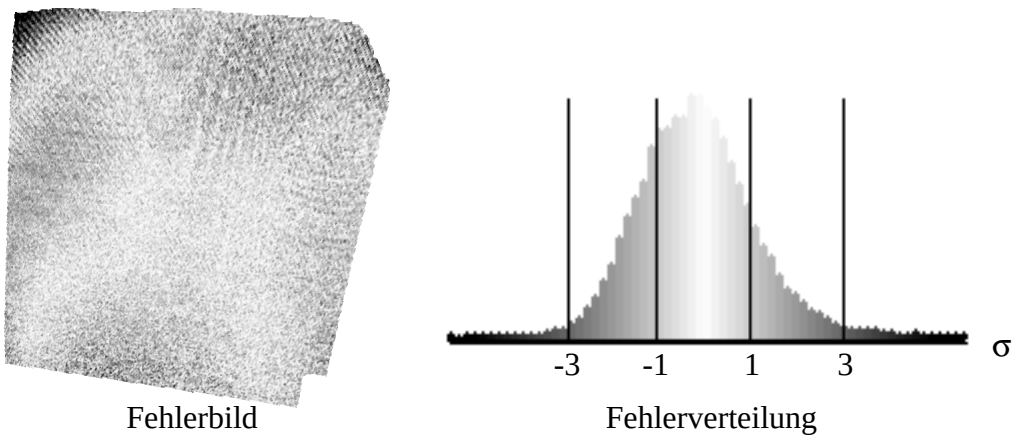


Abbildung 3.39: Erreichte Genauigkeit einer Einzelmessung

Die Kennzahl der Einzelmessungen als Maß der absoluten Genauigkeit lag zwischen 4950 und 6740, was einer mittleren Genauigkeit (1σ) von 0,1 bis 0,07 Millimetern entspricht. Im Feh-

lerbild ist zu erkennen, dass sich die Messfehler gleichmäßig über den gesamten Messbereich verteilen. Allein oben links im Bild ist eine systematische Verschlechterung der Daten über einen kleinen Bereich zu erkennen. Die Holzplatte wurde in einem Winkel von etwa 50 bis 60 Grad vom Projektor angestrahlt, so dass ein vernünftiger Kompromiss zur Scharfstellung des Projektors gefunden werden musste. Die scharf abgebildeten Bereiche im oberen Bereich des Fehlerbildes führen zu Moirée-Effekten, in der Mitte und im unteren Bereich wird dagegen eine sehr homogene Oberflächenstruktur erreicht. Die im Histogramm dargestellte Fehlerverteilung ist Gauß-förmig, was darauf hindeutet, dass die Stereokalibrierung ein Ergebnis erzielt hat, bei dem im Wesentlichen nur noch das stochastische Rauschen der Daten als Fehler übrig geblieben ist.

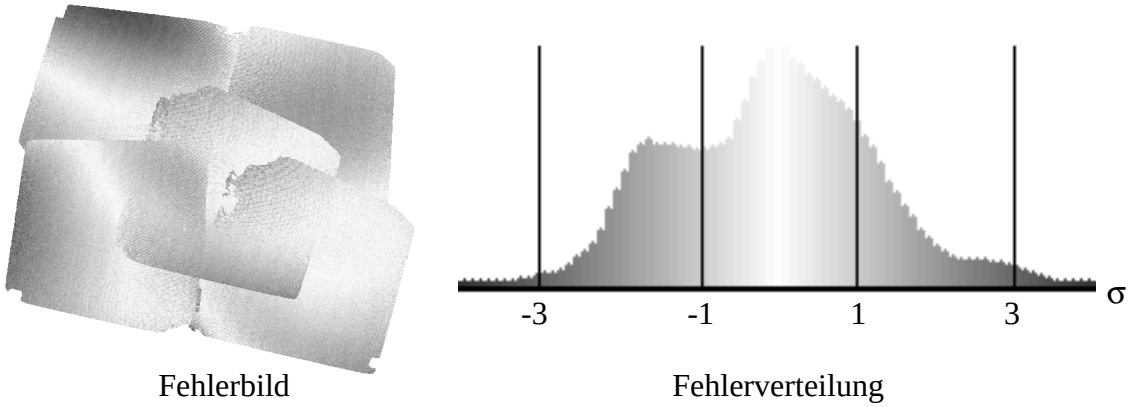


Abbildung 3.40: Erreichte Gesamtgenauigkeit mit zwei Ortungskameras

Für die Gesamtmessung wurde eine mittlere Genauigkeit (1σ) von 0,22 Millimetern erreicht, was auf das Volumen gerechnet einer Kennzahl von 4500 entspricht. Die größte Klaffung im Überlappungsbereich zwischen zwei Einzelansichten entspricht mit 0,5 Millimetern etwa 2σ. Aus dem Histogramm ist zu ersehen, das die Registrierung insgesamt suboptimal verlaufen ist. Die Verteilung der Fehlerwerte zeigt deutlich die Überlagerung zweier Gauß-Kurven. Der Abstand der beiden Maxima beträgt dabei in etwa 2σ. Die globale Markengenauigkeit ist mit einem Wert von 0,13 Millimetern erstaunlich hoch und liegt damit in etwa in der Größenordnung der Genauigkeit der Einzelmessung. Die Ortungskameras hatten zwar ein deutlich größeres Sichtfeld als die Scan-Kameras, allerdings wurde dieser Nachteil durch einen großen Basisabstand wieder ausgeglichen. Für die Genauigkeit der Registrierung konnte ein Wert von 0,5 Millimetern ermittelt werden. Für die Genauigkeit des vorliegenden Beispiels war also im Wesentlichen die Registrierung der Einzelansichten ausschlaggebend.

Beim nächsten Beispiel wurden vier Ortungskameras eingesetzt. Die Anordnung der Scan-Kameras zum ABW-Projektor war ähnlich dem ersten Beispiel, so dass auch die beschriebenen Eigenschaften der Einzelmessungen wieder gelten. Das Testobjekt war wieder die Holzplatte, die dieses mal mit drei Einzelmessungen erfasst wurde. Eine Übersicht über den durchgeführten Versuch ist in Abbildung 3.41 dargestellt.

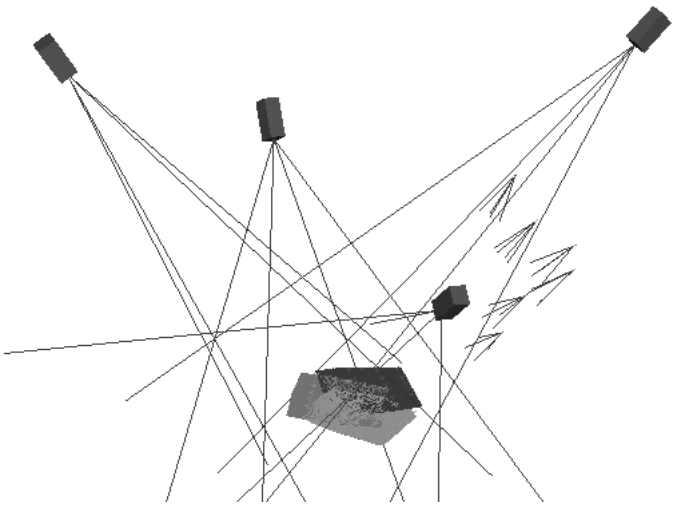


Abbildung 3.41: Ebenen Fläche mit vier Ortungskameras

Die Ortungskameras sind um das Objekt angeordnet. Die beiden inneren Kameras hatten wie im ersten Versuch eine Brennweite von 6 Millimetern, die zusätzlich aufgestellten äußeren Kameras hatten eine Brennweite von 12,5 Millimetern. Um auch mit den äußeren Kameras ein ausreichend großes Sichtfeld zu erhalten, hatten sie einen Abstand von 2,2 Metern zueinander. Die drei Einzelansichten wurden jeweils von rechts aufgenommen, wie es aus dem Übersichtsbild zu entnehmen ist. Die erfasste Gesamtfläche war etwas kleiner, als im vorherigen Versuch. Die erreichte globale Genauigkeit ist in Abbildung 3.42 dargestellt.

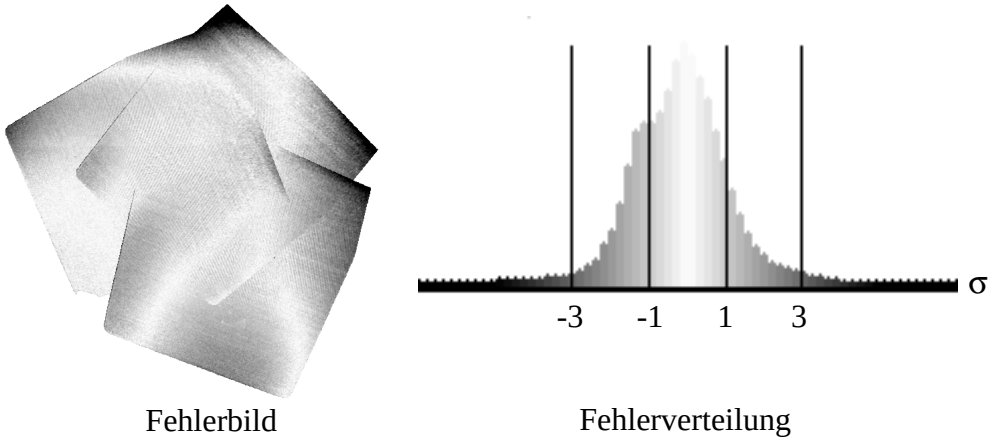


Abbildung 3.42: Erreichte Genauigkeit mit vier Ortungskameras

Die mittlere globale Genauigkeit liegt mit 0,1 Millimetern im Bereich der Genauigkeiten der Einzelmessungen. Auf das Volumen bezogen entspricht diese Genauigkeit einer Kennzahl von 7650. Das Fehlerbild zeigt zwar deutliche systematische Fehler in einigen Ecken der Gesamtmessung, allerdings rühren diese Fehler aus Ungenauigkeiten der Einzelmessungen her und hatten keine Auswirkung auf die Registrierung. Zudem ist zu beachten, dass der maximale Fehler von 4σ einer Ungenauigkeit von etwa 0,4 Millimetern entspricht. Die Klaffungen in den Überlappungsbereichen sind in der Regel deutlich geringer. Dass die Registrierung nahezu optimal abgelaufen ist, lässt sich auch dem Histogramm entnehmen, das einer fast idealen Gauß-Kurve entspricht. Für die globale Markengenauigkeit wurde ein Wert von 0,52 Millime-

tern berechnet, dabei hat der maximale Fehler unter allen Marken 0,7 Millimeter betragen. Die globale Markengenauigkeit ist zwar relativ schlecht, allerdings zeigt der nur geringfügig höhere maximale Fehler, dass die Qualität der Markenlokalisierung sehr gut ist. Die Fehler aus allen Ortungskameras haben sich also im Wesentlichen gegenseitig ausgelöscht. Die Registrierungsqualität ist zwar mit einem Wert von 0,34 Millimetern ebenfalls schlechter als die globale Genauigkeit (0,1mm), aber auch daran ist zu sehen, dass auch die Registrierung noch einmal eine deutliche Fehlerkorrektur erreicht hat.

Im Vergleich zum ersten Versuch, bei dem nur zwei Ortungskameras eingesetzt wurden, fallen zwei Dinge auf:

- die globale Markengenauigkeit ist mit vier Ortungskameras schlechter als mit zwei Ortungskameras,
- die Registrierungsqualität ist mit vier Ortungskameras besser als mit zwei Ortungskameras.

Mit nur zwei Ortungskameras ist die Kalibrierung der Ortungskameras sicherer zu berechnen, allerdings kann die Fehlerauslöschung nicht so hoch sein, wie beim Einsatz von vier Ortungskameras. Die interne Genauigkeit der Markenpositionen ist also relativ hoch, die tatsächliche Genauigkeit der Marken ist aber relativ schlecht, da fast alle Fehler, die auf ungenaue interne Kameraparameter oder Rauschen in den Daten zurückzuführen sind, voll in das Ergebnis durchschlagen. Beim Einsatz von vier Ortungskameras löschen sich gerade diese Fehler gegenseitig aus, so dass die Registrierung mit höherer Genauigkeit abläuft.

Im Folgenden werden drei Beispiele dargestellt, die die praktische Einsetzbarkeit von *HandyScanCLM* demonstrieren sollen. Zum Einsatz kamen jeweils nur zwei Ortungskameras, so dass nur beim Telefon eine Rundumerfassung durchgeführt wurde. Neben einem Übersichtsbild, das die für die Vermessung verwendete Sensorkonfiguration darstellt, werden jeweils noch die Ergebnisse und Detailansichten der gescannten Objekte abgebildet. Die externe Genauigkeit wurde nicht gemessen, die Ergebnisse sind aber mit den bereits vorgestellten Beispielen vergleichbar. Bei der globalen Markengenauigkeit und der Genauigkeit der Registrierung wurde ebenfalls Ergebnisse erzielt, die in der Größenordnung der bereits dokumentierten Beispiele liegen. Um Abschattungen weitestgehend zu vermeiden, wurde der Basisabstand der Scan-Kameras von 70 auf 45 Zentimeter verkürzt.

Apple PowerBook

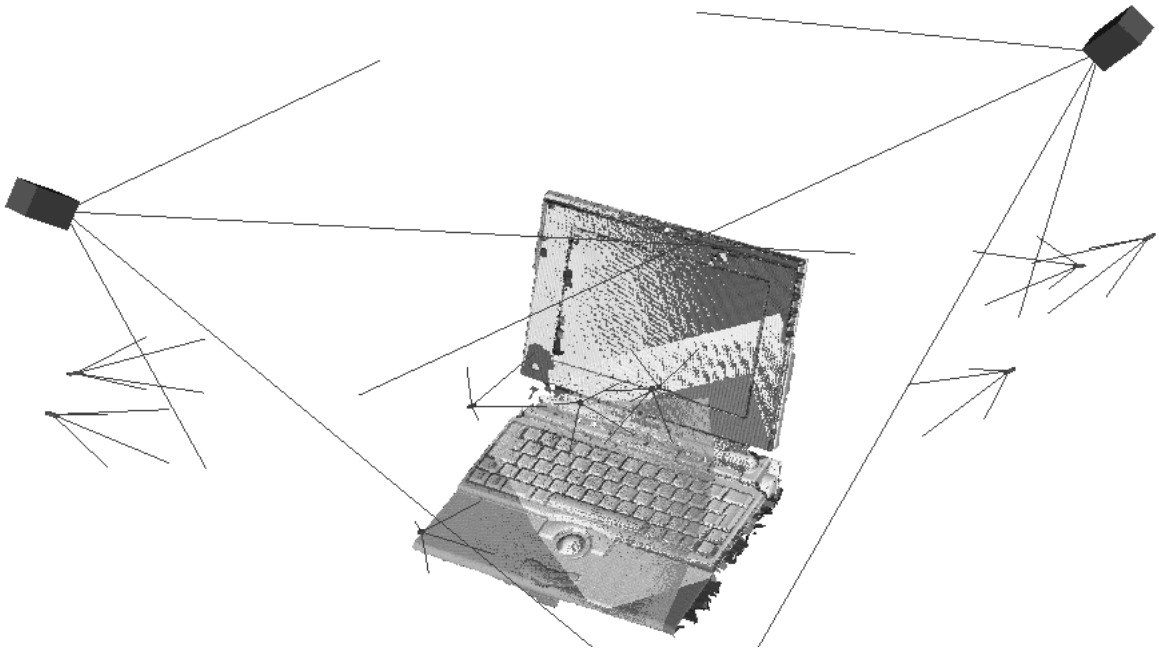


Abbildung 3.43: PowerBook mit zwei Ortungskameras

Das Apple PowerBook wurde mit vier Einzelansichten erfasst, dabei wurde eine Gesamtzahl von etwa 650.000 Punkten aufgenommen. Über jede Einzelmessung wurden etwa 50 Prozent des PowerBooks erfasst, mit allen vier Messungen konnte die gesamte Front aufgenommen werden. Im Übersichtsbild (Abbildung 3.43) sind die Einzelansichten unterschiedlich eingefärbt, um die Qualität der Überdeckung zu verdeutlichen. Im unteren linken Bereich des Displays lag eine Einzelansicht vor den anderen Ansichten, im Großen und Ganzen wurden die Daten aber im Rahmen der Sensorgenauigkeit zur Deckung gebracht.

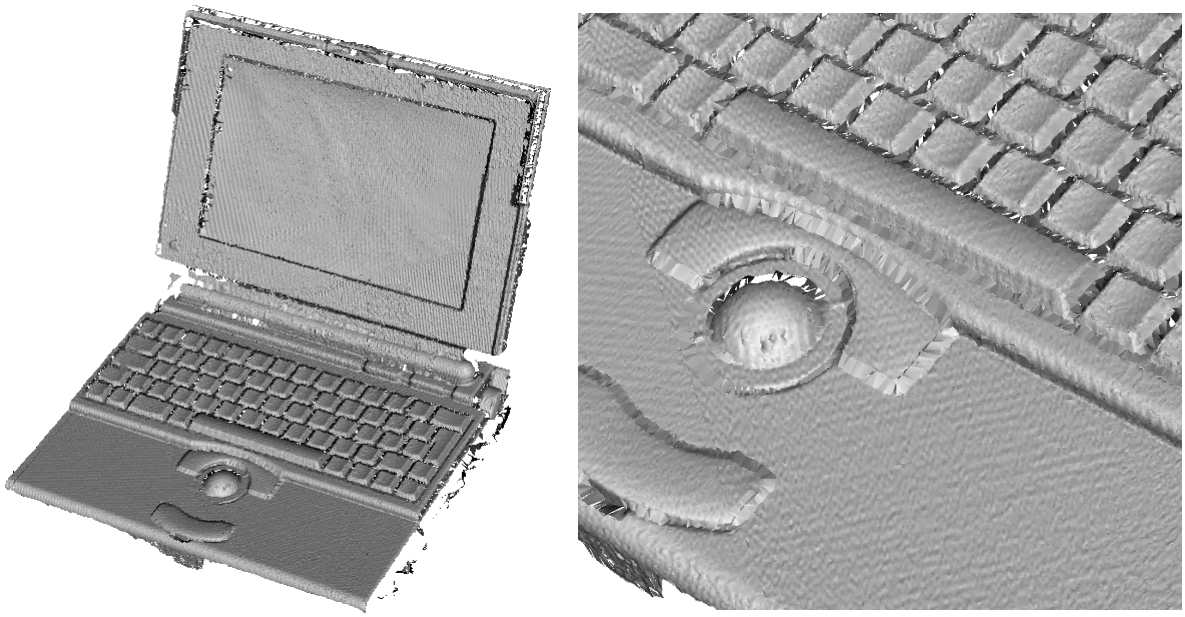


Abbildung 3.44: Scan-Ergebnis des PowerBooks (mit Detailansicht)

Der Überlappungsfehler auf dem Bildschirm beträgt nur wenige zehntel Millimeter und ist in Abbildung 3.44 nicht mehr zu erkennen. Die Frontpartie des PowerBooks wurde bis auf wenige kleine Lücken vollständig erfasst. Sogar die Zwischenräume auf der Tastatur wurden aufgelöst, da das PowerBook aus zwei Richtungen erfasst wurde.

Telefon

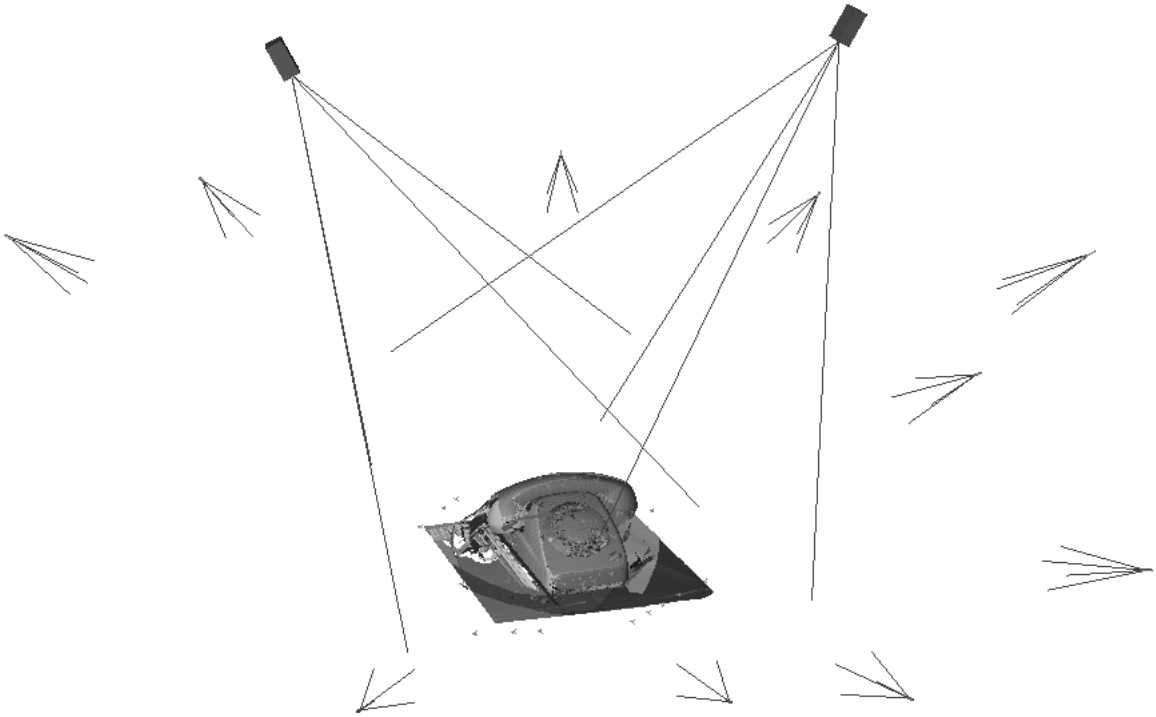


Abbildung 3.45: Telefon mit zwei Ortungskameras

Das bereits aus Kapitel 3.2.4 (Beispiele) bekannte Telefon wurde mit *HandyScanCLM* in fünf Ansichten von allen Seiten aufgenommen. Eine Ansicht wurde in einem Winkel von etwa 45 Grad aufgenommen, die anderen Ansichten wurden in einem Winkel von weniger als 20 Grad zur Tischplatte aufgenommen, auf dem das Telefon plaziert wurde.

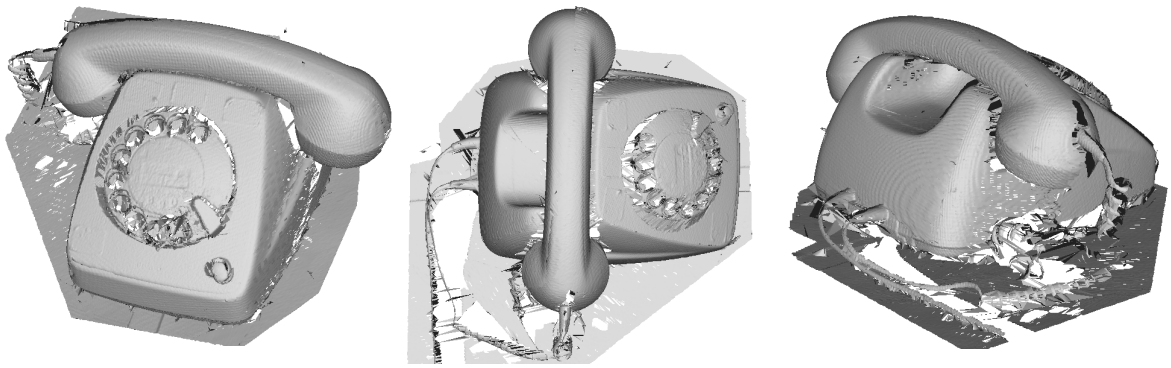


Abbildung 3.46: Ansichten des Telefons

Im Gegensatz zu dem Ergebnis, das mit *HandyScanTT* erzielt wurde, konnte die Erfassung mit *HandyScanCLM* sehr viel vollständiger durchgeführt werden, obwohl eine Messung weniger

nötig war. Vor allem der Haltegriff des Telefons konnte sehr viel tiefer erfasst werden, wie aus Abbildung 3.46 ersichtlich wird. Bei der Vermessung des Telefons war besonders der Vorteil von *HandyScanCLM* von Nutzen, der die individuelle Ausrichtung des Scanners auf das Objekt zulässt. Die Vermessung aus vier Richtungen konnte in einem sehr flachen Winkel durchgeführt werden, so dass der Haltegriff auf der Rückseite und die Seiten des Telefons optimal aufgenommen werden konnten. Die Oberseite des Telefonhörers wurde in einer zusätzlichen fünften Messung von oben aufgenommen. Als nachteilig hat sich auch bei *HandyScanCLM* die glänzende Oberfläche des Telefons erwiesen. Durch Reflexionen auf dem Telefon konnten einige optische Marken nur ungenau erfasst werden, so dass die Registrierung insgesamt nur befriedigend abgelaufen ist. Die größte Klaffung zwischen zwei Ansichten beträgt etwa 2 Millimeter.

AMR Phoenix

Das letzte Beispiel demonstriert die Vermessung eines relativ großen Objektes. Der in der Arbeitsgruppe verwendete AMR[1] *Phoenix* hat eine Höhe von etwa 1,2 Metern bei einer Breite von etwa 0,65 Metern. Wie das Übersichtsbild (Abbildung 3.47) zeigt, wurden zwei Ortungskameras verwendet, um insgesamt acht Aufnahmen von der Vorderseite des Roboters zusammen zuführen. Die Messung hat etwa 1,6 Millionen Scanpunkte ergeben, der Basisabstand zwischen den beiden Ortungskameras betrug 1,6 Meter. Die Aufnahme und Registrierung der acht Aufnahmen war insgesamt problemlos. Die Genauigkeit der Registrierung lag wieder etwa im Bereich der Sensorgenauigkeit.

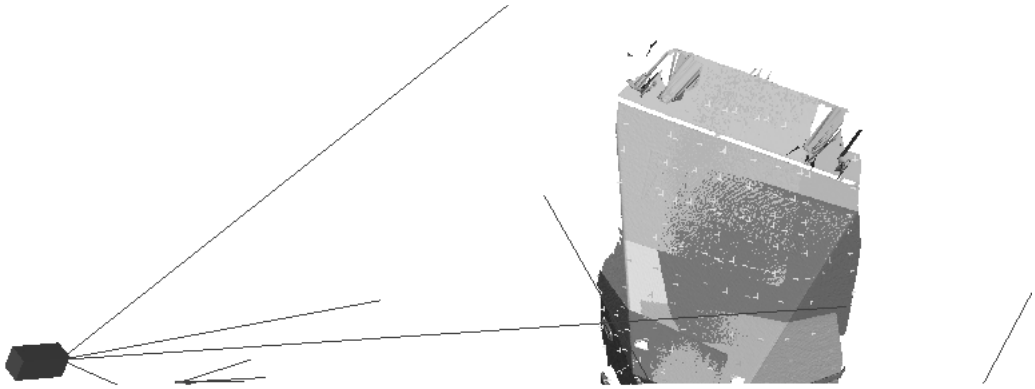


Abbildung 3.47: AMR Phoenix mit zwei Ortungskameras

1. Autonomer Mobiler Roboter

4 Resumee und Perspektiven

Die dreidimensionale Erfassung komplexer Objekte ist zur Zeit nur für jeweils auf eine spezifische Problemstellung optimierte Sensoren möglich. Damit erfordert jedes Problem in vielen Fällen eine neue Konzeption mit der erst die Objekterfassung möglich wird. Hauptsächlich aus diesem Grund sind im Laufe der letzten Jahre eine Vielzahl unterschiedlichster Systeme entwickelt und weiterentwickelt worden. Das Maß der Flexibilität eines Sensorsystems wird dabei zum einen durch das Volumen begrenzt, das mit dem Sensor erfasst werden kann, zum anderen durch die Genauigkeit im Verhältnis zu diesem Volumen. Eine Auffälligkeit ist, dass das erfassbare Volumen sich in den meisten Fällen nur über einen sehr engen Bereich erstreckt. Bei einer Vielzahl von Verfügbaren Systemen ist das Beispielsweise der Bereich von zehn bis dreißig Zentimetern. Größere Objekte sind nicht mehr erfassbar, bei kleineren Objekten wird das Verhältnis zwischen Volumen und Genauigkeit zu schlecht.

Vor diesem Hintergrund untersucht die vorliegende Arbeit die Möglichkeiten, die bestehenden Grenzen vieler Sensorsysteme wenigstens zum Teil zu durchbrechen. Dazu werden die gängigsten Konzepte zur dreidimensionalen Objekterfassung, die laserbasierten Triangulationsverfahren, die Verfahren nach dem codierten Licht Ansatz und Stereo-Verfahren unter einem umfassenden System vereint. Dabei ist es möglich, unter gleichzeitiger Elimination der Nachteile dieser Systeme die Vorteile zu kombinieren, so dass das entwickelte Konzept sich vor allem durch seinen skalierbaren Messraum und seine hohe Genauigkeit auszeichnet. Das Basisverfahren, dass sich inhaltlich der aktiven Stereoskopie zuordnen lässt, garantiert im Weiteren durch seine selbstkalibrierende Eigenschaft die einfache Einsetzbarkeit des Sensorsystems und die flexible Adaption an neue Problemstellungen.

Die echtzeitfähige dreidimensionale Erfassung von Objekten wird durch *HandyScan3D*, dem Basissystem, realisiert. Im einfachsten Fall werden zwei Kameras auf das Objekt gerichtet und das Objekt wird mit Hilfe eines einfachen Laserpointers gescannt. Erweiterungen, im Wesentlichen zur Erhöhung der Datenrate sehen den Einsatz eines zur Linie aufgeweiteten Lasers oder einen Linienprojektor vor. Sowohl die Kalibrierung des Sensorsystems, als auch die Visualisierung des aktuellen Stand der Objekterfassung in Form einer 3D-Rekonstruktion im Computer finden während der eigentlichen Vermessung statt und werden in Echtzeit durchgeführt. Der große Nachteil von *HandyScan3D* besteht in der prinzipbedingten fehlenden Skalierung der 3D-Daten. Dieser Parameter muss nötigenfalls über externe Hilfen berechnet werden. Das kann beispielsweise durch das manuelle Ausmessen einer Objektgröße geschehen, die auch in den 3D-Daten wiedergefunden wird.

Das Problem der Skalierung wurde bei der Realisierung von *HandyScanTT* durch einen einfachen Kalibrierstab gelöst. Das auf *HandyScan3D* aufbauende System ermöglicht durch den Einsatz eines Drehtellers zusätzlich die Rundumerfassung und vollständige dreidimensionale Rekonstruktion von Objekten. Der Drehteller und der Kalibrierstab sind so einfach wie möglich gestaltet und erfordern keine elektronische Kopplung mit einem Computern, so dass trotz

eines nur minimalen Zusatzaufwandes eine deutliche funktionelle Aufwertung von *HandyScanTT* gegenüber *HandyScan3D* erreicht wird. Der verwendete Versuchsaufbau war jedoch ungenau gefertigt und die Halterung für den Kalibrierstab im Drehteller war recht wakkelig. Eine letztendliche Bewertung der erreichbaren Genauigkeit, die mit *HandyScanTT* erreicht werden kann, steht daher noch aus. Die Fertigung eines präziseren Drehtellers und Kalibrierkörpers sollte aber im Weiteren kein großes Problem darstellen. Zudem stellt die nötige Benutzerinteraktion zur Einstellung der Drehtellerposition im Rechner einen Unsicherheitsfaktor dar, da falsche Eingaben schnell zu unbrauchbaren Ergebnissen führen. Hier kann grundsätzlich über die elektronische Übertragung der Winkelstellung des Drehtellers an den Computer nachgedacht werden. Sinnvoller wäre aber beispielsweise die Anbringung von permanenten Marken am Drehteller, die von den Scan-Kameras erfasst werden können und mit denen ebenfalls die Drehlage des Tellers bestimmt werden kann. Wenn sichergestellt werden kann, dass die Marken während der Vermessung nicht vom Objekt verdeckt werden, kann durch eine einfache Erweiterung auf die manuelle Interaktion verzichtet werden.

Das umfassendste, aber auch aufwändigste Konzept auf Basis von *HandyScan3D* stellt *HandyScanCLM* dar. Mit einem Linienprojektor und Kameras werden über die Projektion von wenigen Linienbilder Bereiche auf einem Objekt mit einem Durchmesser von etwa 40 Zentimetern mit einer Genauigkeit von 0,1 Millimetern vermessen. Zusätzlich zur Vermessung eines Teilbereichs der Oberfläche werden mit dem Projektor noch Marken auf das Objekt projiziert. Ortungskameras, die um das Objekt aufgestellt sind erfassen ebenfalls diese Marken und ermöglichen die Berechnung einer Transformation der Scandaten in das globale Koordinatensystem dieser Kameras. Ebenso können natürlich auch alle weiteren 3D-Aufnahmen des Objektes in dieses globale Koordinatensystem transformiert werden. Prinzipiell ist durch dieses Verfahren die Vermessung nahezu beliebiger Objekte mit der Genauigkeit des Linienprojektors möglich. Im Rahmen dieser Arbeit war die Untersuchung der optimalen Anordnung der Ortungskameras nicht mehr möglich. Die Berechnung einer Markenpositionen wird zwar besser, abhängig von der Anzahl der Ortungskameras, zuviele Ortungskameras stellen aber einen Aufwand dar, der sich nicht mehr in ein entsprechend besseres Ergebnis umsetzen lässt. Auch die Form und Zahl der projizierten Marken kann Auswirkungen auf die Gesamtgenauigkeit haben, die noch zu untersuchen bleibt. Obwohl die algorithmische Realisierung des Verfahrens eine Unterstützung für sehr vieler Ortungskameras vorsieht, hat die begrenzte technische Ausstattung im Rahmen der Forschungsarbeit nur Versuche mit maximal vier Ortungskameras zugelassen. So bleibt auch offen, inwieweit das implementierte Verfahren zur Kalibrierung der Ortungskameras Fehlerfortpflanzungen über mehrere Kameras hinweg verursacht oder tolerant gegenüber fehlerhaften Daten ist.

Die durchgeführten Testszenarien haben insgesamt für alle realisierten Sensorsysteme den Nachweis der praktischen Einsetzbarkeit erbracht. Die 3D-Vermessung eines Objektes mit *HandyScan3D* ist kinderleicht, da auf alle aufwändigen Arbeitsgängige, wie Kalibrierung oder Einrichtung des Sensorsystems, verzichtet werden kann, der Laser selber wird per Hand geführt. Das Softwarekonzept hat sich zusätzlich bei der Entwicklung von *HandyScanTT* und *HandyScanCLM* bewährt, so dass zu erwarten ist, dass auch weitere Entwicklungen von neuen Sensorsystemen auf Basis von *HandyScan3D* in kurzer Zeit möglich sind. Allein der Einsatz der vorgestellten Systeme im industriellen Umfeld als letzte Bewährungsprobe steht noch aus.

A Abbildungsverzeichnis

1 Dreidimensionale Objekterfassung

1 Klassifizierung von 3D-Sensoren 13
2 Lichtpunkttriangulation 14
3 Lichtlaufzeit 15
4 Phasenmodulationverfahren 16
5 Lichtschnittverfahren (einfache Triangulation) 17
6 Lichtschnittverfahren mit zwei Kameras (Doppeltriangulation) 18
7 Linienzuordnung beim Mehrlinienverfahren 19
8 Flächig kodiertes Muster 20
9 Kodierung der Tiefe 21
10 Gray-Code Verfahren 22
11 Volumenschnittverfahren 26
12 Passive Stereoskopie 27
13 Epipolargeometrie 28
14 TOPAS-Scanner 33
15 VITUS 3D-Scanner 34
16 Cyberware WB4 35
17 Tecmath Contour 35
18 Cyberware Desktop Scanner Modell 15 36
19 3D-Dimension Scanbook 37
20 Objektrekonstruktion mit Facade 37
21 Riegl Scanner mit Gebäude und erzeugtem Tiefenbild 38
22 3D-Scanners ModelMaker 39
23 ATOS Sensor 40
24 Softwaresystem SLIM 41

2 Kalibrierung von 3D-Sensoren

1 CCD-Sensor 44
2 CMOS-Sensor 45
3 Lochkamera- und Mattscheibenmodell 47
4 Kameramodell nach Tsai 49
5 Kalibriersystem 50
6 Kalibrierung eines Lichtschnittsensors 53
7 Interpolierende Kalibrierung 55
8 Modell des Stereosensors 58
9 Lösungen der Stereokalibrierung 62

3 Aktive Stereoskopie

1 Hardware von HandyScan3D und Messobjekt (Tasse) 64
2 Rekonstruiertes Objekt (Tasse) 64
3 Projektionsmuster des Linienprojektors ABW 320 66
4 Differenzbildverfahren 68

5 Laserpunktsuche 69
6 Stereoorientierungen 70
7 Korrespondenzsuche beim Einsatz eines Linienlasers 72
8 Auswertung von Gray-Code und Phase 74
9 Bildverarbeitung beim Gray-Code Verfahren 76
10 Unterteilung des Kamerabildes zur Auswahl von Korrespondenzen 79
11 Rauschverhalten 81
12 Berechnung von Raumkoordinaten 83
13 Die beiden regelmäßigen Netztopologien 86
14 Aufzählungsregeln zur Visualisierung von Dreiecksnetzen 86
15 Triangulation des Rasters (Typ A) 87
16 Triangulation eines spärlich besetzten Rasters 88
17 Skalierung von Mehrfachansichten 89
18 Rekonstruktion eines Kartons 90
19 Rekonstruktion einer Tastatur 91
20 Rekonstruktionsqualität 92
21 Unterschiedliche Ausdünnungen 93
22 Unterschiedliche Grade von Lückenfüllern 94
23 Unterschiedliche Darstellungsattribute 95
24 Erweiterte Ansichten durch Umstellen von Kameras 95
25 Drehteller und Kalibrierstab 98
26 Markendetektion 99
27 Koordinatensysteme von HandyScanTT 100
28 Rekonstruiertes Objekt 102
29 Virtuelle Kamerapaare 102
30 Berechnung von Korrekturvektoren 104
31 Berechnung der Korrekturfunktion 105
32 Beleuchtungsartefakte in texturierter Darstellung und korrigierte Textur 106
33 Ananas 107
34 Banane 108
35 Pinguin 109
36 Telefon 109
37 Startwertberechnung bei der Mehrkamerakalibrierung 116
38 Ebene Fläche mit zwei Ortungskameras 121
39 Erreichte Genauigkeit einer Einzelmessung 121
40 Erreichte Gesamtgenauigkeit mit zwei Ortungskameras 122
41 Ebenen Fläche mit vier Ortungskameras 123
42 Erreichte Genauigkeit mit vier Ortungskameras 123
43 PowerBook mit zwei Ortungskameras 125
44 Scan-Ergebnis des PowerBooks (mit Detailansicht) 125
45 Telefon mit zwei Ortungskameras 126
46 Ansichten des Telefons 126
47 AMR Phoenix mit zwei Ortungskameras 127
48 Scan-Ergebnis des AMR Phoenix 128

B Literaturverzeichnis

[Acuity] http://www.acuityresearch.com

[Atkinson89] Atkinson, K. E.: *An Introduction to Numerical Analysis*, 2nd Edition, John Wiley and Sons, New York 1989

[Berg99] Berg, M.; Jörg, K.-W.; Paulick, J.-P.: *3D Ultraschallentfernungsmessung mit Pseudo-Random Sequenzen*, 15. Fachgespräch Autonome Mobile Systeme (AMS‘99) München, Springer-Verlag, Reihe "Informatik Aktuell", pp.78-87, Nov. 1999

[Bergmann97] Bergmann, D.;Galanulis, K.; Winter, D.: *Advanced 3D-Fringe-Projection-Systems*, GOM mbH, Braunschweig, 1997

[Besl92] Besl, P.J.; McKay, N.D,: *A Method for Regitration of 3-D Shapes*, Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.14, No.2, pp.239-256, 1992

[Busch91] Busch, H.: *Ein Verfahren zur Oberflächenmodellierung dreidimensionaler Objekte aus Kamerabildfolgen*, Ph.D. Thesis, Universität Hannover, 1991

[Casott99] Casott, N; Prenting, J.: *Genauigkeitssteigerung beim Korrelationsverfahren in der digitalen Bildverarbeitung*, Photogrammetrie Fernerkundung Geoinformation, S.101-112, Heft 2/199

[Caspar97] Caspar, K.: *Konzeption und Realisierung eines robusten und hochgenauen Lichtschnitttriangulationssystems*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, Oktober 1997

[Contour] http://www.hs.tecmath.com/deutsch/htp.shtml

[Debevec96] Debevec, P.E.; Taylor, C.J.; Malik, J.: *Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach*, In SIGGRAPH'96, August 1996

[Dieterich98] Dieterich, E.: *Entwicklung eines 3D-Handscanners*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, Juni1998

[Digiclops] http://www.ptgrey.com/products/digiclops/index.htm

[Faro] http://www.faro.com

[Ferner97] Ferner, L.: *Automatische Klassifikation von Kleinobjekten anhand von 3D-Scandaten*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, September 1997

[Foley92] Foley, J.; van Dam, A.; Feiner, S.; Hughes, J.; Phillips, R.: *Grundlagen der Computergrafik; Einführung, Konzepte, Methoden*, Addison-Wesley-Verlag, 1992

[Haberäcker95] Haberäcker, P.: *Praxis in der digitalen Bildverarbeitung und Mustererkennung*, Carl Hanser Verlag, München, Wien, 1995

[Hamfeld98] Hamfeld, H.; Hansen, G.; v. Puttkamer, E.; Trieb, R.: *3D Menschvermessung im TOPAS-Projekt*, 5. ABW-Workshop über optische 3D Formerfassung, Esslingen, Jan. 1998

[Hamfeld98b] Hamfeld, H.: *Dreidimensionale Vermessung großer Objekte, Entwicklung von Methoden zur automatisierten Digitalisierung großer Objekte*, Forschungsantrag bei der Stiftung Rheinland-Pfalz für Innovation, März 1998

[Hansen96] Hansen, G.: *Konzeption und Realisierung eines Sensorsystems zur dreidimensionalen Erfassung anthropometrischer Daten*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, April 1996

[Hartley92] Hartley, R.I., *Estimation of Relative Camera Positions for Uncalibrated Cameras*, In Proc. of ECCV92, pp.579-587, Springer-Verlag, 1992

[Hartley95] Hartley, R. I.: *In Defence of the 8-Point Algorithm*, In 5th Inter. Conf. on Computer Vision, p.1064-1070, MIT, Cambridge, Massachusetts, IEEE Comp.Soc.Press, June 1995

[Heiduk96] Heiduk, B.†: *Generierung kalibrierter Konturinformationen aus Schattenrißaufnahmen*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, April 1996

[Hightech00] Hightech Report 2000, Vernetzte Welten, Reverse Engineering, *Rückkehr in den Rechner*, S.54-55

[Hofmann00] Hofmann, K.: *Erweiterung eines 3D-Handscanners um eine Laserlinie zur Detektion von Objekten*, Projektarbeit im Fachbereich Informatik der Universität Kaiserslautern, Oktober 1997

[Huang86] Huang, T.S.: *Determining Three-Dimensional Motion and Structure from Two Perspective Views*, Handbook of Pattern Recognition and Image Processing, Academic Press, 1986

[Jörg94] Jörg, K.W.: *Echtzeitfähige Multisensorintegration für autonome mobile Roboter,* Dissertation, im Fachbereich Informatik der Universität Kaiserslautern, B.I. Wissenschaftsverlag, Januar 1994

[Karbacher97] Karbacher, S.: *Rekonstruktion und Modellierung von Flächen aus Tiefenbildern*, Shaker Verlag, Aachen, 1997

[Karbacher99] Karbacher, S.; Häusler, G.; Schönfeld H.: *Reverse Engineering Using Optical Range Sensors*, In Handbook of Computer Vision ans Applications, Volume 3, Systems ans Applications, 1999

[Karbacher00] Karbacher, S.; Schön, N.; Schönfeld, H.; Häusler, G.: *Digitizing 3D Objects for Reverse Engineering and Virtual Reality*, In Principles of 3D Image Analysis and Synthesis, S.336-347, Kluwer Academic Publishers, Boston-Dodrecht-London, 2000

[Krattenthaler93] Krattenthaler, W.; Mayer, K.; Duwe, H.: *3d-Surface Measurement with Coded Light Approach*, presented at 17th ÖAGM Workshop and 4th Int. Workshop for Digital Image Processing and Computer Graphics, Graz, Austria, June 1993

[LaserDesign] http://www.laserdesign.com

[Lerios95] Lerios, A.; Garfinkle, C.D.; Levoy, M.: *Feature-Based Volume Metamorphosis*, Proceedings of SIGGRAPH '95 (Los Angeles, CA, '95), In Computer Graphics Proceedings, Annual Conference Series, pp.449-456. 1995

[Levoy99] Levoy, M.: *The Digital Michelangelo Project*, Proceedings of the Second International Conference on 3D Digital Imaging and Modeling, Ottawa, Canada, October 1999

[Liu98] Liu, Y.-F.: *A Unified Approach to Image Focus and Defocus Analysis*, Ph.D. Thesis, State University of New Yorc, December 1998

[Longuet81] Longuet-Higgins, H.C.: *A Computer algorithm for Reconstructing a Scene from two Projection*s, Nature, 293:133-135, September 1981

[Mezger96] Mezger, O.:*LICURA - Light Cut Range Sensor*, Diplomarbeit im Fachbereich Informatik der Universität Kaiserslautern, September 1996

[MinPack] http://www.netlib.org/minpack

[Neugebauer98] Neugebauer, P.J.;Klein, K.: *Adaptive Triangulierung komplexer Oberflächen*, 5. ABW Workshop 3D-Bildverarbeitung, Technische Akademie Esslingen, Jan. 1998

[Osten91] Osten, W.: *Digitale Verarbeitung und Auswertung von Interferenzbildern*, Informatik Lehrbücher und Monographien, Akademie Verlag, 1991

[Ottmann90] Ottmann, T.:*Algorithmen und Datenstrukturen*, Mannheim, BI-Wissenschafts-Verlag, 1990

[Pollefeys99] Pollefeys, M.: *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, Ph.D. Thesis, University of Leuven, May 1999

[Press86] Press, W. H.; Flannery, S. A.; Teukolsky, S. A.; Vetterling, W. T.: *Numerical recipes in C: The art of scientific computing*, Cambridge University Press, 1986

[Ranger] http://www.ivp.se

[Ristow98] Ristow, B.; Diener, H.; Köthe, U.; Stelbe, U.; Schreyer, M.: *ERSO - Acquisition, Reconstruction and Simulation of Real Objects*, Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, Vol. 4, Piscataway, NJ, USA, IECON ‘98

[Sasse94] Sasse, R.: *Bestimmung von Entfernungsbildern durch aktive stereoskopische Verfahren*, Vieweg Verlag Braunschweig, 1994

[Schröder80] Schröder, G.: *Technische Optik: Grundlagen und Anwendungen*, Vogel-Verlag, Würzburg, 3.Auflage, 1980

[Sick] http://www.sick.de

[Stoddart96] Stoddart, A.J.; Hilton, A.: *Registration of Multiple Point Sets, 13th Int. Conf. of Pattern Recognition*, Vienna, Austria, 1996

[Trieb95] Trieb, R.: *Erfassung, Analyse und Synthese anthropometrischer Körperbaudaten zur Automatisierung von Design- und Fertigungsprozessen bei Individualprodukten*, Forschungsantrag bei der Stiftung Rheinland-Pfalz für Innovation, 1995

[Tsai87] Tsai, R.Y.:*A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses*, IEEE Journal of Robotik and Automation, Vol. RA-3, No. 4, S. 323-344, August 1987

[Tsai88] Tsai, R.Y.: *Real Time versatile robotics hand/eye calibration using 3D machine vision*, Proceedings - IEEE International Conference on Robotics and Automation, 1:544-556, April 1988

[TsaiCode] http://www.cs.cmu.edu/~rgw/TsaiCode.html

[Vuylsteke90] Vuylsteke, P.; Oosterlinck, A.: *Range Image Acquisition with a Single Binary-Encoded Light Pattern*, IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 12, No. 2, pp. 148-587, Springer-Verlag, 1992

[Wahl86] Wahl, F.: *A Coded Light Approach for Depth Map Acquisition*, 8. DAGM-Symposium, Paderborn 1986, Springer Verlag, 1986

[Weber01] Weber, J.; Franken, L.; Jörg, K.-W.; Schmitt, K.; v. Puttkamer, E.: *An Integrative Framework for Global Self-Localization*, 2001 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2001), Baden-Baden, August 2001

[Wei94] Wei, T.-C.:*Three Dimensional Machine Vision Using Image Defocus*, Ph.D. Thesis, State University of New Yorc, December 1994

[Wichert95] v. Wichert, G.; Kleiner, K.: *Selbstorganisierende Bildanalyse für die Navigation autonomer mobiler Roboter*, Autonome Mobile Systeme (AMS ‘95), Springer-Verlag, pp.152-161, 1995

[Winter97] Winter, D.; Reich, C.: *Video-3D-Digitalisierung komplexer Objekte mit frei beweglichen, topometrischen Sensoren*, GOM mbH Braunschweig, Institut für Meßtechnik und Experimentelle Mechanik, TU Braunschweig, 1997

[YuvColor] http://www.webartz.com/fourcc/fccyuv.htm

[Zhang93] Zhang, Z.; Luong, Q.-T.; Faugeras, O.D.: *Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction*, Technical report No.2079, October 1993

[Zhang96] Zhang, Z.: *Determining the Epipolar Geometry and its Uncertainty: A Review*, Technical Report No.2927, July 1996

[Zheng94] Zheng, J.Y.: *Acquiring 3-D Models from Sequences of Contours*, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.16, No.2, February 1994

C Index

Numerisch

3D
-Scanner 11
-Sensor 10
-Sensorsystem 11
3D-Dimension Scanbook 36
3D-Rekonstruktion 82, 100
3D-Sensoren
aktive optische ~ 13
Eindimensional messende ~ 13
Klassifizierung von ~ 12
passive optische ~ 24
räumlich messende ~ 19
zweidimensional messende ~ 17

A

Abbildungseigenschaft 46
Abschattung 14, 18
ABW 320 ..... 66, 74
AccuRange3000LV 16
Acht-Punkte Verfahren 77
AMR 25
ATOS Sensor 39
Auflösung 11
laterale ~ 11
mittlere ~ 11
mittlere Tiefen~ 11
Tiefen~ 11

B

Basisabstand 14, 17
Beleuchtungsartefakte 106
Bildkoordinatensystem 71
Bildübertragung 45
Bildverarbeitung 67
Binär-Codes 75
Bresenham 71

C

CameraLink 45
Canoma 30
CCIR-Norm 45
Cyberware Desktop Scanner 36
Cyberware Scanner 34

D

Differenzbild 68, 71, 99
Drehteller 26, 98
~zentrum 100
Dreiecksnetz 85
Dual-Head Grafikkarte 67

E

EIA-Norm 45
Einzelansichten 111
Elektronische Kameras 43
Entscheidungsbereich 75
Epipol 59
Epipolar
~einschränkung 28, 64
~geometrie 28, 57, 70, 71, 78
~linie 28, 59, 70, 71
~matrix 58, 77, 78
Ersatzpunkt 87
Extraktion
~ des Gray-Codes 73
~ einer Linie 70
~ eines Punktes 68

F

Facade 38
Faro-Arm 32
feature based visibility 29
feature extraction 29
FireWire 66
Fokus-Defokus Analyse 24
Framegrabber 45

G

Genauigkeit 11
~ der Einzelmessung 120
mittlere ~ 11
Genauigkeitssteigerung 103
Gewichtungsfaktor 105
Gray-Code Verfahren 22, 76

H

Halogenprojektor 66
HandyScan3D 63
HandyScanCLM 110
HandyScanTT 97

Hash-Tabelle 75

I
IEEE1394 ..... 45, 66
Interferenz 21
Interpolation 55
  lineare ~ 75

K
Kalibrier
  ~körper 49
  ~marken 49
  ~stab 98
  ~system 50
  ~wand 50
Kalibrierung 12, 28, 100, 113
  ~ der Ortungskameras 115
  ~ nach Tsai 49
  approximierende ~ 56
  Drehteller~ 100
  interpolierende ~ 55
  Kamera~ 12
  Mehrkamera~ 116
  modellbasierte ~ 53
  modellfreie ~ 54
  Stereo~ 12, 57, 77
Kamera
  ~kalibrierung 43
  ~modelle 47
  ~objektive 46
  CCD ~ 44
  CMOS ~ 44
  Master ~ 71, 75
  Ortungs~ 112
  Slave ~ 71, 75
Kameramodell nach Tsai 49
Kennzahl 11, 110
Kissenverzerrung 71
Kodierung
  ~ der Fläche 20
  ~ der Tiefe 21
  ~ über die Zeit 22
Koordinatenmessmaschinen 9
Koordinatensystem
  Drehteller~ 100
  globales ~ 100
  Sensor~ 118
  Stereo~ 100
Korrektur
  ~ der Textur 106
Korrekturfunktion 103
Korrekturvektor 103
Korrespondenz 27
Korrespondenzpunktsuche 29

L
Laser
  ~linie 70
  ~punkt 68
Laserpunktsuche 69
Least-Square 50, 56, 60, 77, 103
Lichtgeschwindigkeit 15
Lichtlaufzeitmessung 15
Lichtpunkttriangulation 14
Lichtschnittsensor 53
Lichtschnittverfahren 17, 18
Linienbildern 74
Lochkameramodell 47
Lokalität 69, 71

M
Marken
  ~genauigkeit 120
  ~koordinaten 118
  ~ortung 112
  kodierte ~ 111
Markendetektion 99
Mattscheibenmodell 47
Mehrfachansichten 88
Mehrlinienverfahren 19
Modellbasierte Bildauswertung 25
ModelMaker 39
Moirée-Verfahren 21
Monoskopie 24

N
Netztopologie 86
nichtlineare Optimierung 61, 116

O
OpenGL 85
Ortungskamera 111

P
pattern matching 29
PCEye4 65
Phasen
  ~bild 74
  ~modulation 15
  ~werte 74
  ~winkel 75
Phasen-Shift Verfahren 21

Philips Hopper XG20 67
Pixelkoordinatensystem 71

Q
Qualität
  externe ~ 79
  interne ~ 79

R
RAMSIS 33
Rapid-Prototyping 10
Referenzbild 68, 74
Referenzkamera 115
Registrierung 111, 113, 117
  Genauigkeit der ~ 120
Riegl Scanner 38
Rundumerfassung 99

S
Schattenrissverfahren 25
Schwellwertverfahren 68
Sehstrahl 83
Sick Sensor 16
Skalierung 59, 118
  relative ~ 88
SLIM 40
Sony DFW V500 66
SONY XC55 ..... 65, 80
Speckles 73, 77
Stereo
  ~kalibrierung 12, 57, 77
  ~sensor 58
Stereoorientierung 70
Stereoskopie
  aktive ~ 23
  passive ~ 27
Streifenordnung 21
Structure from Motion 57

T
Taktile Systeme 31
Tecmath Contour-System 35
TOPAS-Scanner 32
TriangleFan 86
TriangleStrip 86
TRITOP System 39

U
Ultraschall 31
USB 45

V
Videobeamer 67
Visualisierung 85
VITUS Scanner 33
Volumenmodell 26
Volumenschnitt 25
Voxelraum 26

Y
YUV411 66