

Untersuchung eines neuartigen 2D- gestützten 3D-PMD-Bildverarbeitungssystems

vom Fachbereich Elektrotechnik und Informatik
der Universität - Gesamthochschule Siegen
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
(Dr.-Ing.)

genehmigte Dissertation

von

Diplom-Ingenieur Detlef Justen

1. Gutachter: Prof. Dr.-Ing. Rudolf Schwarte
2. Gutachter: Prof. Dr.-Ing. Otmar Loffeld

Tag der mündlichen Prüfung: 26. Januar 2001

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Zentrum für Sensorsysteme (ZESS) der Universität Siegen.

An dieser Stelle möchte ich meinem Doktorvater Herrn Prof. Dr.-Ing. R. Schwarte für die wissenschaftliche Betreuung meiner Arbeit, Herrn Prof. Dr.-Ing. O. Loffeld für die Übernahme des Korreferats und Herrn Prof. Dr.-Ing. W. Düchting für die Leitung und Durchführung des Promotionsverfahrens Danken.

Des weiteren möchte ich mich bei allen Kollegen und allen Mitarbeitern des Zentrums für Sensorsysteme und des Instituts für Nachrichtenverarbeitung ohne namentliche Nennung für die vielen Anregungen und Hilfestellungen bedanken.

Meinen ganz besonderen Dank gilt meiner Frau Nina Justen, die mich während des lang andauernden Werdeganges dieser Arbeit tatkräftig unterstützt hat und viel Verständnis für die vielen der Dissertation gewidmeten Wochenenden und Feierabendstunden gezeigt hat. Nicht zu vergessen sind meine beiden Kinder Pia und Amy, welche in ihren ersten zwei bzw. erstem Lebensjahr ihren Vater hauptsächlich vor dem Computer zu sehen bekommen haben (wenn sie denn nicht schliefen).

Siegen, im März 2001

Inhaltsverzeichnis

1	Einleitung	7
1.1	Aufgabenstellung und Zielsetzung	8
1.2	Motivation	9
1.3	Gliederung.....	11
2	Abstract	13
3	Übersicht über flächendeckende 3D-Aufnahmeverfahren.....	15
3.1	Grundlagen	15
3.1.1	Optische Formerfassung.....	15
3.1.2	Vergleich zwischen Grauwert- und Tiefenkamera.....	16
3.1.3	Definition 2½D / 3D	17
3.2	3D-Aufnahmeverfahren.....	18
3.2.1	Triangulation	19
3.2.1.1	Statische Stereoanalyse.....	19
3.2.1.2	Dynamische Stereoanalyse.....	21
3.2.1.3	Räumlich kodierter Lichtansatz	22
3.2.2	Photometrische Verfahren.....	23
3.2.2.1	Shape-from-Shading	23
3.2.2.2	Photometrische Stereoanalyse	24
3.2.3	Laufzeitverfahren	25
3.2.3.1	Scannende Systeme	25
3.2.3.2	Pockels-Zelle.....	26
3.2.3.3	Bildverstärker	27
3.3	Zusammenfassung.....	28
4	3D-Bildverarbeitung.....	31
4.1	Aufgaben und Anwendungsgebiete der 3D-Bildverarbeitung.....	31
4.2	Allgemeine Vorgehensweise	32
4.2.1	Vorverarbeitung.....	33
4.2.2	Segmentierung.....	35
4.2.2.1	Kantentypen	35
4.2.2.2	Regionen- und kantenbasierte Segmentierung	38

4.2.3	Zuordnung.....	38
4.3	Ausblick.....	40
5	Beschreibung der Systemkomponenten.....	41
5.1	Photonic Mixer Device.....	41
5.1.1	Laufzeitentfernungsmessung	41
5.1.2	CMOS-PMD	42
5.1.3	Optimierte PMD-Pixel-Strukturen	44
5.2	2D/3D-Kamerakzept.....	46
5.2.1	Strahlteiler-Kamerasystem	46
5.2.2	MixedPixel-Kamerasystem.....	47
5.3	Anordnungsmöglichkeiten der vereinzelteten Tiefenwerte.....	48
5.3.1	Gitteranordnung	48
5.3.2	Orientierungsbestimmung mit vereinzelteten Tiefenwerten	49
5.3.3	Füllfaktor	49
5.3.4	Anwendungsbeispiele	50
5.4	Beleuchtung	52
5.4.1	Eigenbeleuchtung	52
5.4.2	Fremdlicht und Eigenbeleuchtung.....	53
5.5	Visualisierung.....	54
5.6	Zusammenfassung.....	55
6	Grundlagen zur 2D-gestützten 3D-Bildverarbeitung	57
6.1	Arbeitsumgebung.....	57
6.1.1	Simulationsdaten.....	58
6.1.2	Einschränkungen hinsichtlich der Simulation	59
6.2	Untersuchung des Aufnahmesystems.....	61
6.2.1	Koordinatensystem	61
6.2.2	Kameramodell	62
6.2.3	XYZ-Korrektur	64
6.2.3.1	Berechnung der tatsächlichen Objektentfernung	65
6.2.3.2	Berechnung der Position im Raum.....	69
6.2.4	Zusammenfassung.....	69
6.3	Flächenapproximation.....	71
6.3.1	Polynomapproximation.....	71

6.4	Flächennormale	73
6.5	Lokale Orientierung	74
6.6	Reflexionsmodelle	75
6.6.1	Grundlagen	75
6.6.2	Beleuchtungs- und Reflexionsmodelle	76
6.6.3	Diffuse Reflexion	77
6.6.4	Spiegelnde Reflexion	78
6.7	Shape-from-Shading	78
6.7.1	Lösungsansätze	80
6.7.1.1	Globale Verfahren	80
6.7.1.2	Lokale Verfahren	81
6.7.1.3	Propagationsverfahren	82
7	Datenfusion und -integration	83
7.1	Motivation	83
7.2	Grundlagen	83
7.3	Datenfusion in der Bildverarbeitung	85
7.4	Datenfusion von Tiefen- und Grauwertdaten	86
7.4.1	Fusion auf Pixelebene	87
7.4.2	Fusion auf Bildebene	88
7.4.2.1	Kantenklassifizierung	89
7.4.2.2	Kantenfusionierung	91
7.4.3	Fusion auf Symbolebene	92
8	3D-Bildverarbeitung mit vereinzelt Entfernungswerten	93
8.1	Graubildverarbeitung mit vereinzelt Entfernungswerten	95
8.1.1	Allgemein	95
8.1.1.1	Objekterkennung	96
8.1.1.2	Bestimmung der Objektgröße, Objektentfernung und Lage	97
8.1.1.3	Texturierte Regionen	97
8.1.1.4	Flächeninterpretation	98
8.1.1.5	Kantenklassifizierung	99
8.1.2	Applikationsbeispiele	100
8.1.2.1	Geschwindigkeitsmessung von Fahrzeugen	100
8.1.2.2	Innenraumüberwachung von Kraftfahrzeugen	102

8.2	Ausweitung der vereinzelt Entfernungswerte auf eine Fläche.....	105
8.2.1	Flächeninterpolation.....	107
8.2.1.1	Rekonstruktion einer Halbkugel.....	108
8.2.1.2	Rekonstruktion einer komplexen Szene.....	109
8.2.2	Shape-from-Shading.....	110
8.2.2.1	Voraussetzungen.....	112
8.2.2.2	SFS-Ansatz von Bichsel & Pentland.....	112
8.2.2.3	SFS-Ansatz von Pentland.....	115
8.2.2.4	SFS-Ansatz von Tsai & Shah.....	118
8.2.2.5	Vergleich der SFS-Verfahren.....	122
8.2.2.6	Zusammenfassung der SFS-Verfahren.....	124
8.2.3	Lambert-Verfahren.....	125
8.2.3.1	Allgemein.....	126
8.2.3.2	Lambert-Z.....	127
8.2.3.3	Lambert-N.....	132
8.2.3.4	Zusammenfassung der Lambert-Verfahren.....	135
8.2.4	Zusammenfassung der Verfahren zur Ausweitung der vereinzelt Entfernungswerte.....	136
9	Zusammenfassung und Ausblick.....	138
	Literaturverzeichnis.....	141
	Abkürzungsverzeichnis.....	150
	Formelverzeichnis.....	152
	Pseudo-Programme.....	155

1 Einleitung

Die Erkennung und Lokalisierung von Objekten in Bildern gehört zu den Hauptaufgaben eines Bildverarbeitungssystems. Wurden bisher vor allem Grauwertbilder (intensity, brightness, graylevel) verarbeitet, nimmt der Einsatz von Tiefenbildern (depth, range) in der 3D-BV aufgrund der darin enthaltenen Abstandsinformation ständig zu. Die Tiefeninformation sowohl mit als auch ohne Grauwertdaten ermöglicht eine einfache und effiziente Lösung von bisher nur aufwendig oder gar nicht lösbaeren Problemen.

In Abhängigkeit von der gestellten Meßaufgabe stehen aufgrund der Vielzahl unterschiedlicher 3D-Meßsysteme verschiedene Kriterien für die Auswahl des geeigneten Meßsystems zur Diskussion, wie z.B.:

- ∅ die geforderte laterale Auflösung (lateral resolution) / Anzahl der 3D-Werte
- ∅ die Meßgeschwindigkeit (speed)
- ∅ die geforderte Tiefenauflösung (depth resolution)
- ∅ die Kosten (cost)
- ∅ die geforderte Genauigkeit (accuracy)
- ∅ die Sicherheit (safety)
- ∅ der Leistungsverbrauch (power consumption)
- ∅ das Verhalten des Sensors in Abhängigkeit der Oberfläche (ruggedness)
- ∅ die Kompaktheit des Gerätes (compactness)
- ∅ die Robustheit des Gerätes (robustness)
- ∅ die Augensicherheit

Der wichtigste Punkt dieser Aufzählung ist die Anzahl der notwendigen 3D-Werte. Die Applikation gibt dabei im wesentlichen die Anzahl und Position der notwendigen Tiefenwerte pro Bild vor, sie bestimmt also, ob die Szene dicht (dense), mittel (medium) oder schwach (sparse) aufgenommen werden muß. Eine geringe Anzahl von Entfernungsdaten kann beispielsweise bereits in einem industriellen Vermessungssystem zur Prozeßkontrolle ausreichen, wenn dafür die Genauigkeit und die Geschwindigkeit entsprechend hoch sind. Die Kosten haben in einem solchen Anwendungsfall nur eine untergeordnete Bedeutung. Andererseits sind für ein fahrerloses Transportsystem viele aber nicht unbedingt genaue Tiefendaten zur Wegplanung und Hinderniserkennung erforderlich. In diesem Falle haben die Leistungsaufnahme, die Meßgeschwindigkeit und das Gewicht eine größere Relevanz. Innerhalb eines Systems zur Aufnahme und Speicherung von

Objekten in einem CAD-System sind neben der Meßzeit die Anzahl und Auflösung der Entfernungswerte die Hauptkriterien zur Auswahl des zu verwendenden 3D-Aufnahmesystems.

Wie zu erkennen ist, bestimmen verschiedene Faktoren die Wahl des optimalen 3D-Aufnahmesystems. Die Anzahl der Tiefenwerte hat dabei eine übergeordnete Funktion. Eine geringe Anzahl von Tiefenwerten wirkt sich z.B. vorteilhaft auf die Verarbeitungsgeschwindigkeit, die Kosten, den Leistungsverbrauch u.a. aus.

1.1 Aufgabenstellung und Zielsetzung

Grundlage dieser Arbeit ist die Einführung eines 2D/3D-Kamerakzeptes auf Basis einer flächendeckenden Grauwertkamera mit Unterstützung vereinzelter Entfernungswerte. Die Einführung dieses neuen Kamerakzeptes wurde erst durch die von Prof. Schwarte 1996 [Scw96] eingeführte neue PMD-Technologie ermöglicht. Bereits mit Hilfe von vereinzelt Entfernungswerten in Verbindung mit einem flächendeckenden Grauwertbild wird die 3D-Bildverarbeitung wesentlich vereinfacht bzw. erst ermöglicht (z.B. 3D-Objektrekonstruktion, Bewegungsanalyse ...).

Diese Arbeit bezieht sich vorwiegend auf die Untersuchung von bisher bekannten Grundlagen der 2D- und 3D-BV und auf deren Tauglichkeit hinsichtlich der oben genannten neuen Sensoranordnung. Der Schwerpunkt der ausgewählten Verfahren liegt dort, wo normale 2D-BV Konzepte Schwierigkeiten haben oder infolge der fehlenden Tiefeninformation versagen. Mit den nur vereinzelt Tiefenwerten in Verbindung mit einem flächendeckenden Grauwertbild stellt das Aufnahme- und Verarbeitungssystem eine einfache Möglichkeit zur effektiven Nutzung der Tiefeninformation zur Verfügung. Auf Basis der herkömmlichen 2D-BV und mit vereinzelt Tiefenwerten kann unter Berücksichtigung des Abtastproblems mit einfachen Mitteln:

- ∅ eine Klassifizierung von Objekten erfolgen
- ∅ eine Entfernungs-, Lage- und Größenbestimmung erfolgen
- ∅ die Oberflächenorientierung bestimmt werden
- ∅ eine Flächeninterpretation erfolgen
- ∅ eine Klassifizierung von texturierten Regionen erfolgen
- ∅ eine Objektverfolgung (Geschwindigkeit, Richtung) erfolgen
- ∅ ...

Neben der direkten Nutzung der vereinzelt Tiefenwerte an mißverständlichen bzw. mehrdeutigen Stellen im Grauwertbild können die wenigen vorhandenen Tiefendaten auch dazu genutzt werden, die fehlenden Tiefenwerte mit Hilfe der vorhandenen Grauwertdaten zu rekonstruieren. Dazu werden Verfahren eingesetzt, welche eine Beschreibung der Zusammenhänge zwischen den Grauwert- und Tiefenbildern nutzen (z.B. Shape-from-Shading). Die real vorliegenden Tiefenwerte werden dabei nur für die Initialisierungsphase genutzt. Als Ergebnis erhält man ein ‚vollständiges‘ rekonstruiertes Tiefenbild.

Ungeachtet der vielfältigen Einsatzmöglichkeiten stellen die hier beschriebenen Ansätze und Verfahren keinen vollständigen Ersatz für ein vollwertiges 3D-Aufnahmesystem dar. Dennoch ermöglichen bereits vereinzelt Tiefenwerte in Verbindung mit einem flächendeckenden Grauwertbild eine für viele Fälle ausreichende 3D-BV. Durch die Nutzung der vereinzelt Tiefenwerte in Verbindung mit dem Grauwertbild wird neben der schnelleren Tiefenbildaufnahme auch eine schnellere 3D-Bildverarbeitung im Vergleich zu der reinen Nutzung von Grauwert- bzw. Tiefenbildern ermöglicht. Im gleichen Sinne werden damit neue Einsatzgebiete, die infolge der bisherigen ‚langsamen‘ Tiefenaufnahmeverfahren (siehe Laserscanner, räumlich kodierter Lichtansatz ...) nicht möglich waren, erschlossen.

Aufgrund dessen, daß die Arbeit parallel zur Entwicklung der zugrundeliegenden PMD-Technologie erfolgte, konnten nur wenige Untersuchungen an realen Daten durchgeführt werden. Zur frühzeitigen Validierung des vorgeschlagenen Kamerakzeptes wurde daher häufig auf synthetisch erzeugte Tiefen- und Grauwertdaten ausgewichen.

1.2 Motivation

Die Bildverarbeitung auf Basis vereinzelter Entfernungswerte läßt sich sehr gut mit dem menschlichen Sehen vergleichen. Das räumliche Sehen des Menschen basiert unter vorwiegender Nutzung von Grauwertbildern sowohl auf dem binokularen- als auch auf dem monokularen Sehen (siehe Hauske [Ha94]). Innerhalb des binokularen 3D-Sehens wird die Entfernung aus der Überlappung der zu den beiden Augen gehörenden Sehfeldern und den Augenwinkeln bestimmt. Im Gegensatz dazu wird beim monokularen 3D-Sehen die Information über die Tiefe auf Basis des aufgenommenen Grauwertbildes bestimmt. Hierbei wird in erster Linie versucht, die Szene zu interpretieren, Gegenstände vom Hintergrund zu separieren und relevante Daten von nicht relevanten zu trennen.

Aus

- Ø der Wissensbasis
- Ø der absoluten Objektgröße
- Ø dem Größenvergleich von Objekten
- Ø der Verdeckungsanalyse
- Ø der Fokussierung
- Ø dem Schatten und Glanzwurf
- Ø Shape-from-X (z.B. Shading, Texture)
- Ø Bewegungsparallaxe
- Ø ...

wird anschließend mit einfachen Mitteln ein relatives Entfernungsbild aufgebaut. Je mehr Faktoren zur Gewinnung von Tiefenbildern genutzt werden können, desto stärker wird der Tiefeindruck (siehe Hauske [Ha94]).

All diesen Verfahren ist gemeinsam, daß sie kein flächendeckendes, absolutes Tiefenbild liefern. Selbst der Stereoansatz, die Augenwinkelposition und die Fokussierung liefern nur vereinzelte Tiefeninformationen im Fokusbereich. Reichen diese nicht aus, so werden sie mit einem der zuvor genannten relativen Ansätze rekursiv erweitert. Objekte außerhalb des Fokusbereiches werden dabei beispielsweise in Relation zum Entfernungswert im Fokuspunkt gesetzt. Liegen Informationen aus dem Vorgeschehen vor, werden diese ebenfalls mit in den Interpretationsprozeß einbezogen. Im gesamten Interpretationszyklus entscheidet dabei die Wissensbasis über die Wahl der zu nutzenden Verfahren. Dies ist sehr gut anhand von unbekanntem Objekten oder Oberflächen (Texturen) zu erkennen. In einem solchen Fall benötigen wir einen entsprechend längeren Zeitraum zur Erkennung und folglich zur Bestimmungsbestimmung. Die Wissensbasis ist somit der zentrale Punkt des menschlichen Sehens.

Auf Basis dieser menschlichen Aufnahme- und Vorgehensweisen sollen in dieser Arbeit einige Aspekte der menschlichen Bildverarbeitung auf das dem menschlichen Sehen entsprechende Bildaufnahmesystem (flächendeckendes Grauwertbild mit vereinzelten Tiefenwerten) übertragen werden. Dabei kann die Verarbeitung gleichermaßen nur mit den vereinzelten Entfernungswerten als auch über die Ausweitung der vereinzelten Entfernungswerte auf eine größere Fläche erfolgen. Auch hier muß für eine sinnvolle

Interpretation der Szene eine Wissensbasis aufgebaut werden, welche allerdings zunächst auf die gestellte Aufgabenstellung beschränkt bleibt.

1.3 Gliederung

Für ein besseres Verständnis der 3D-BV werden in **Kapitel 3** nach einer allgemeinen Einführung in die 3D-Bildaufnahme verschiedene Kamerakonzepte zur Aufnahme von 3D-Daten vorgestellt. Aufgrund der großen Anzahl von Sensoren zur Aufnahme von Entfernungswerten wird dabei das Augenmerk auf flächendeckende Aufnahmesysteme gelegt, die die Szene mit Ausnahme der scannenden Systeme parallel aufnehmen. Aus der verbleibenden großen Zahl von Sensorsystemen werden repräsentativ einige Verfahren aus den drei stellvertretenden Kategorien (Time-of-Flight, photometrischer Ansatz und Triangulation) vorgestellt.

Aufgrund der Komplexität der 3D-BV wird in **Kapitel 4** für einen einfacheren Einstieg in die 3D-BV und für ein besseres Verständnis dieser Arbeit zunächst ein Überblick in die allgemeine Vorgehensweise der 3D-BV gegeben. Dabei liegt das Augenmerk auf der zu der 2D-BV unterschiedlichen Datenart und Vorgehensweise. Die Beschreibung selbst ist unabhängig von der Anzahl der zur Verfügung stehenden Tiefendaten.

Im **Kapitel 5** wird das in dieser Arbeit zugrundeliegende Kamerakonzept beschrieben. Dazu wird zunächst das PMD-Element zur Aufnahme von Entfernungs- und Grauwertdaten beschrieben. Auf Basis dieses Sensorelementes wird anschließend das 2D/3D-Kamerakonzept zur Aufnahme von Grauwert- und vereinzelt Tiefendaten dargestellt.

Kapitel 6 beschäftigt sich mit den Grundlagen der 2D-gestützten 3D-BV. Hier werden neben der Arbeitsumgebung das Aufnahmesystem und sich daraus ergebende Fragestellungen untersucht. Anschließend werden weiterführende Grundlagen bezüglich der nachfolgenden 3D-BV beschrieben. Dazu gehören unter anderem die Flächenapproximationsverfahren, die Bestimmung der Flächennormale und der lokalen Orientierung.

Das in dieser Arbeit vorgestellte Sensorkonzept liefert sowohl ein Grauwertbild als auch ein Tiefenbild. Daher wird nach einer Einführung in die Beschreibung der Zusammenhänge zwischen Grauwert- und Tiefendaten (Reflexionsmodelle) das Shape-from-Shading Verfahren vorgestellt, welches auf dieses Grundwissen aufsetzt.

Kapitel 7 ist eine Einführung in die Datenfusion und –integration von Grauwert- und Tiefendaten. Zum besseren Verständnis wird dazu die Datenfusion vom menschlichen Sehen abgeleitet und auf die Bildverarbeitung übertragen. Auf Basis der Datenfusion von Tiefen- und Grauwertdaten werden drei verschiedenen Ebenen der Datenfusion beschrieben.

Innerhalb des **Kapitels 8** werden auf Basis des erarbeiteten Wissens Möglichkeiten zur Bildverarbeitung von Grauwertbildern mit einer geringeren Anzahl von Entfernungswerten vorgestellt. Durch die Möglichkeit, die PMD-Elemente in verschiedenen Anordnungen zu platzieren, ergeben sich vielversprechende Möglichkeiten zur Bildverarbeitung. Wird für einen bestimmten Anwendungsfall ein fein aufgelöstes Entfernungsbild benötigt, so kann das ebenfalls mit diesem Sensorkonzept erfolgen. Die fehlende Tiefeninformation kann beispielsweise auf Basis der Flächenapproximation, Shape-from-Shading oder den Lambert-Verfahren aus den Grauwertpixeln und/oder den vereinzelt Tiefenwerten rekonstruiert werden.

In **Kapitel 9** werden die Ergebnisse dieser Arbeit zusammengefaßt und es wird ein Ausblick auf zukünftige Entwicklungen gegeben.

Hinsichtlich des einfacheren Verständnisses werden hier alle mathematischen Betrachtungen im kartesischen Koordinatensystem durchgeführt. Die Achsenbeschriftung der verschiedenen Koordinatensysteme erfolgt mit Großbuchstaben und angehängtem Index. Die eigentlichen Komponenten im Raum werden zur einfacheren Unterscheidung mit Kleinbuchstaben gekennzeichnet. Vektoren und Matrizen sind fett gedruckt und Einheitsvektoren sind mit einem kleinen e im Index gekennzeichnet.

2 Abstract

The recognition and localization of objects within images is one of the main tasks of an image processing system. Whereas it was mainly greyscale images (intensity, brightness, grey level) being processed before, the use of depth images (depth, range) for the three-dimensional image-processing is steadily increasing mainly due to the comprised distance information. The depth information both with and without grey scale data offers easier and more efficient solutions to problems that have so far been only possible - if at all - on a larger scale.

Basis of this thesis is the introduction of a 2D/3D-camera concept based on a surface-covering greyscale camera, supported by individual depths data. The introduction of this new camera concept was made possible only by the new PMD-technology introduced by Prof. Schwarte in 1996 [Scw96]. Already the use of only individual depths data in combination with a complete greyscale image simplifies a 3D-image processing considerably or makes it actually possible (e.g. 3D-object reconstruction, motion analysis ...).

The thesis presented here is mainly based on the research into already existing theories on 2D- and 3D-image processing and their value in regard to the above mentioned sensor arrangement. Due to the direct spatial relation between greyscale and depths data the main focus lies on the registered data sources (synchronization of the scanning elements). Possible applications for the herein introduced new 3D-image processing are to be found wherever standard 2D-image processing concepts encounter difficulties or fail due to the missing depth information. Using only individual depths data in combination with a surface covering greyscale image, the recording and processing system offers a simple possibility for the effective utilization of the depths information. Already individual depths data offer simple means to:

- ∅ classify objects
- ∅ determine distance, orientation and size
- ∅ determine the surface orientation
- ∅ ...

Next to the direct utilization of the individual depths data in such spots in the greyscale image that might lead to misinterpretation or have ambiguous meaning, the real depths

data can also be used to reconstruct missing depths values by using the available greyscale data. Thus such procedures are applied which use a description of the relation between the greyscale and depths images (e.g. Shape-from-Shading).

Despite the manifold possibilities for applications the approaches and procedures described here, do not at all present a complete alternative for a full 3D-recording system. Nevertheless, already individual depths data make 3D-image processing possible, which suffices in most of the cases. Compared to solely 2D- or 3D-approaches, the utilization of the individual depths data in combination with a surface-covering makes both faster depths image recording and faster 3D-image processing possible. Along the same lines new possible applications are made accessible, which were formerly not possible because of the 'slow' 3D-recording processes.

3 Übersicht über flächendeckende 3D-Aufnahmeverfahren

Im Gegensatz zu den traditionell verwendeten Grauwertbildern, bei denen jeder Bildpunkt die von einer Kamera aufgenommene Lichtintensität darstellt, repräsentieren die einzelnen Werte eines Tiefensensors zusätzlich den Abstand des Punktes vom Sensor zum Objekt und gegebenenfalls weiter zur Lichtquelle. Gegenüber Grauwertbildern liegt der größte Vorteil von Tiefenbildern in der expliziten Darstellung der Form dreidimensionaler Objekte. In direktem Zusammenhang damit steht die Unempfindlichkeit von Tiefenbildern gegenüber Faktoren wie Beleuchtung, Schattenwurf und Verschmutzung von Objektoberflächen, welche die Analyse von Grauwertbildern erheblich erschweren. Im Gegensatz zu den Grauwertbildern besteht jedoch weiterhin ein indirekter Zusammenhang zwischen dem aufgenommenen Wert und der Position der Lichtquelle, der Oberflächeneigenschaft und –orientierung der Objekte und der Betrachterposition.

Die Erfassung von Daten zur dreidimensionalen Beschreibung von Objekten mit optoelektronischen Methoden ist bereits Gegenstand jahrelanger Forschung. Es wurden bereits eine Vielzahl von verschiedenartigen Lösungsansätzen innerhalb der Literatur vorgestellt, und es bestehen darüber hinaus noch viele weitere Entwicklungsmöglichkeiten. Um diese Arbeit innerhalb der 3D-Aufnahme und 3D-BV Konzepte besser einordnen zu können, sollen dazu in diesem Kapitel aus der Vielzahl der verschiedenen Aufnahmeverfahren einige wenige vorgestellt werden. Dabei soll das Augenmerk nach einer kurzen Grundlagenbeschreibung auf die flächendeckende Aufnahme von 3D-Bildern gerichtet werden.

3.1 Grundlagen

Im folgenden sollen einige grundlegende Aspekte der 3D-Aufnahmetechnik (optische Formerfassung) vorgestellt werden. Dabei liegt das Augenmerk auf dem Unterschied zwischen den Grauwert- und den Tiefenkameras.

3.1.1 Optische Formerfassung

Aufgabe der optischen Formerfassung ist es, zu vermessende (möglicherweise sehr komplexe) Objekte durch ein Meßgerät so aufzunehmen, daß die gestellte Meßaufgabe gelöst werden kann (siehe Abb. 3.1). Hierzu nimmt das Meßgerät die Koordinaten einer genügenden Anzahl von Punkten der Objektoberfläche auf.

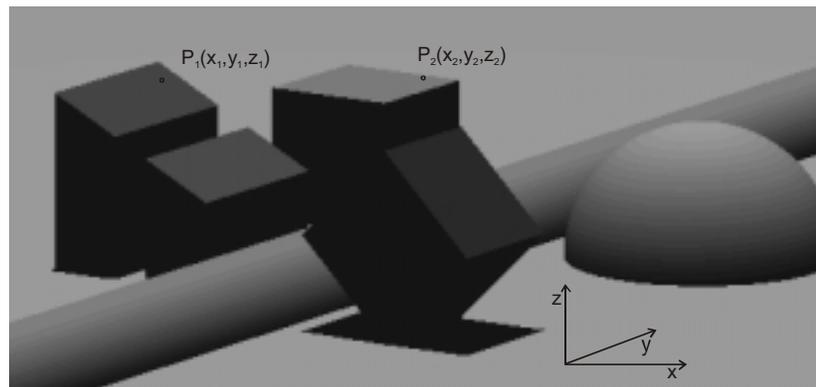


Abb. 3.1: 3D-Musterszene

3.1.2 Vergleich zwischen Grauwert- und Tiefenkamera

Mit einer Grauwertkamera (im folgenden auch 2D-Kamera genannt) wird die Helligkeit des Objektpunktes aufgenommen. Die Objektentfernung geht dementsprechend verloren, d.h. sie ist vieldeutig (Abb. 3.2 a). Zu jedem Bildpunkt auf der Sensorfläche gibt es genau eine Gerade im Raum, deren Punkte P in den Punkt P' der Bildebene abgebildet werden. Zur eindeutigen Bestimmung aller drei Koordinaten eines Raumpunktes werden somit zusätzliche Informationen benötigt.

Im Gegensatz zu der Grauwertkamera ist eine Tiefenkamera ein Meßinstrument, welches zusätzlich die Entfernungen von einer Objektoberfläche zur Kamera mißt. Zur Darstellung einer kompletten Szene werden dazu mehrere Entfernungsmessungen (parallel oder sequentiell) im Raster der ebenen Sensorfläche mit konstantem Gitterabstand oder konstantem Winkel durchgeführt (Abb. 3.2 b). In Anlehnung an die Bezeichnung Pixel (Picture Element) in der 2D-BV wird der Punkt im Raum in der 3D-BV als Voxel (Volume Element) bezeichnet.

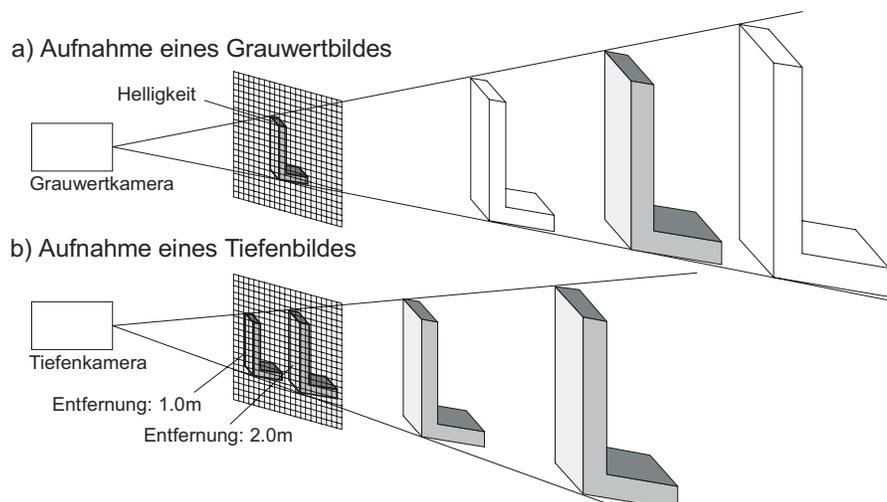


Abb. 3.2: Vergleich zwischen Grauwert- und Tiefenkamera

Viele Tiefenkameras geben zu der Tiefeninformation die Helligkeit des entsprechenden Objektpunktes als zusätzliche Information mit an. Dabei muß in Anlehnung an das Aufnahmeverfahren zwischen ‚registered‘ und ‚non-registered‘ unterschieden werden. Registered bedeutet im Falle der 3D-BV, daß der Tiefenwert und der Helligkeitswert zum gleichen Meßpunkt gehören. Durch diese Korrespondenz können beispielsweise in der nachfolgenden Bildverarbeitung die tatsächlichen Kanten aus dem Grauwert- und aus dem Tiefenbild zuverlässiger erkannt werden (z.B. können im Grauwertbild vorhandene Texturkanten durch Hinzunahme des Tiefenbildes als Falschkanten klassifiziert werden). Die hinzugekommene Tiefeninformation ermöglicht somit eine Klassifizierung der Kanten, was zu einer vereinfachten Objekterkennung bzw. Unterscheidung führt. Werden statt dessen z.B. getrennte Kameras zur Aufnahme von Tiefen- und Helligkeitswerten genutzt, so entfällt die einfache Zuordnung zwischen diesen Werten. Die nachfolgende Verarbeitung muß dies bei Bedarf aufwendig rekonstruieren.

Wird der Helligkeitswert aus der Tiefenaufnahme abgeleitet, so kann in Abhängigkeit der verwendeten Wellenlänge des Sensors eine weitere Unterscheidung erfolgen. Nach Hoover [Ho96] wird die Intensität des reflektierten Lichtes als Reflexionswert bezeichnet, wenn nicht sichtbares Licht (z.B. von einem Laserscanner) verwendet wurde. Andernfalls wird von einem Helligkeitsbild gesprochen. Da es sich in beiden Fällen um eine Helligkeitsinformation des Bildes handelt, werden in der nachfolgenden Beschreibung die Begriffe Helligkeit, Intensität, Reflexionswert und Grauwert gleichermaßen verwendet. In diesem Sinne sollen Tiefen- und Entfernungswert ebenso gleichbedeutend angewendet werden.

3.1.3 Definition 2½D / 3D

In der Literatur werden für Daten, welche mit einer Entfernungskamera aufgenommen wurden, gleichermaßen die Begriffe 2½D und 3D verwendet. Eine klare Zuordnung der Dimension zu einem bestimmten Aufnahmeverfahren ist jedoch nicht zu finden.

Der Begriff 2½D wird oftmals verwendet, um hiermit Höhendaten zu beschreiben, d.h. Tiefenwerte in Abhängigkeit eines XY-Gitters [Ho96] oder unvollständige räumliche Informationen über ein Objekt (siehe Klette & Koschan [KIKo96]). Analog dazu wird der Begriff 3D mit den Volumen- oder Oberflächendaten verbunden. Die Daten liegen in diesem Falle in Abhängigkeit von XYZ-Werten vor. Das bedeutet für Volumendaten, daß für ein räumliches Gitter z.B. ein Dichtewert des Volumenpunktes in der Speicherposition

(XYZ-Angabe) abgelegt wird. Im Falle der Oberflächendaten existiert für jeden einzelnen Meßpunkt eine eigene XYZ-Angabe.

Eine weitere Definitionsmöglichkeit geht von der Anzahl der veränderlichen Variablen aus. Demnach ist ein Höhenbild $z=f(x,y)$ gleichzusetzen mit einem Grauwertbild $I=f(x,y)$, also einem 2D-Bild. Entsprechend gilt für Punktwolken die Definition 3D, da die Daten in Abhängigkeit von XYZ gespeichert werden. Allerdings kann bei dieser Definition die Ordnung sehr schnell in die Höhe gehen, wenn weitere Abhängigkeiten, wie z.B. die Temperatur, der Aufnahmezeitpunkt, die Farbe bzw. das Spektrum oder andere Variablen berücksichtigt werden.

Der gebräuchlichste Ansatz ist die generelle Zuordnung des Begriffes 3D zu Daten, welche sowohl Grauwert- als auch Tiefeninformationen enthalten. Da innerhalb der Literatur keine eindeutige Definition der Begriffe 2½D und 3D existiert, wird in dieser Arbeit diese gebräuchliche Definition auf alle Daten mit Tiefeninformationen angewendet. In diesem Sinne wird der Begriff 2D-Daten für reine Grauwertbilder verwendet.

3.2 3D-Aufnahmeverfahren

Systeme zur flächendeckenden Aufnahme von Entfernungswerten einer Szene lassen sich gemäß ihrer Arbeitsweise in verschiedene Kategorien einteilen. Die drei wichtigsten Kategorien sind **Time-of-Flight (TOF)**, **photometrischer Ansatz** und **Triangulation**.

Systeme, welche mittels **Time-of-Flight** arbeiten, senden ein Signal (z.B. Laser, Ultraschall) aus, das vom Meßobjekt reflektiert und vom Empfänger empfangen und verarbeitet wird. Aus der Zeit zwischen Senden und Empfangen des Impulses oder der Phasenverschiebung des Signals läßt sich der zurückgelegte Weg berechnen. Die interferometrischen Verfahren nutzen die Auslöschung bzw. Verstärkung des reflektierten kohärenten Lichtes durch einen Referenzstrahl aufgrund der unterschiedlichen Laufzeit aus. Durch die sehr kurze Wellenlänge des Lichtes eignen sich solche Verfahren sehr gut zur Messung von kleinsten Abmaßen und nur bedingt zur Messung von größeren Entfernungen.

Innerhalb der **Triangulation** kann zwischen aktiven und passiven Verfahren unterschieden werden. In den Bereich der **passiven Triangulation** fallen Verfahren, die nicht durch eine aktive Quelle beleuchtet werden. Dazu werden mehrere Aufnahmen benötigt, aus denen die Tiefeninformation korrespondierender Punkte mit Hilfe der Triangulation berechnet wird. Die beiden Hauptvertreter dieser Kategorie sind die statische Stereoanalyse, welche mindestens zwei ortsfeste Kameras benötigt, und die dynamische Stereoanalyse. Hierfür wird eine Bildsequenz benötigt, in welcher sich die Kamera, das

Objekt oder beide bewegen. Die **aktive Triangulation** beruht auf mindestens einer aktiven Lichtquelle zur direkten Abstandsmessung. Eine effiziente Triangulationstechnik stellt der räumlich kodierte Lichtansatz dar.

Innerhalb des **photometrischen Ansatzes** wird die Entfernung anhand der Helligkeitsveränderung in Abhängigkeit der Flächennormalen des Meßobjektes bestimmt. Dies entspricht im wesentlichen dem Verfahren, mit dem der Mensch eine Fotografie auswertet. Stellvertreter dieser Kategorie sind (ungeachtet der Namensgebung) die photometrische Stereoanalyse und Shape-from-Shading (SFS).

Weitere Verfahren, wie z.B Shape-from-Textur (SFT) oder Rekonstruktion auf Basis einer Wissensbasis, können nicht direkt zu einer der obigen Kategorien gezählt werden. Sie werten zum Teil den Informationsinhalt des Bildes direkt aus oder versuchen die Tiefeninformation durch Interpretation und Zuordnung zu bekannten Objekten zu rekonstruieren.

In den nachfolgenden Kapiteln soll zur besseren Einordnung dieser Arbeit eine Auswahl von gebräuchlichen Sensoren zur flächendeckenden Aufnahme von Tiefenwerten vorgestellt werden. Die Auswahl ist nicht vollständig. Ansätze basierend z.B. auf interferometrischen oder taktilen Verfahren werden hier nicht besprochen.

3.2.1 Triangulation

Bei der Triangulation wird der Abstand eines Oberflächenpunktes mittels der anliegenden Winkel einer bekannten optischen Basis bestimmt. Alle Verfahren, welche auf dem Triangulationsansatz beruhen, leiden daher unter diesem notwendigen räumlichen Abstand der Lichtquelle und der Kamera bzw. den Kameras.

3.2.1.1 Statische Stereoanalyse

Nur wenige Probleme innerhalb der 3D-Bildgewinnung wurden tiefer untersucht als der Stereoansatz. Im Mittelpunkt eines Stereoverfahrens steht die Zuordnung korrespondierender Merkmale in einem Stereobildpaar. Ein Merkmal im linken und rechten Bild wird als korrespondierend bezeichnet, wenn es Projektionen derselben Struktur in der Szene repräsentiert. Im allgemeinen stellt diese Korrespondenzanalyse das schwierigste Teilproblem in einem Stereoverfahren dar, wobei die Schwierigkeit vor allem daher rührt, daß für ein Merkmal des einen Bildes in der Regel mehrere Merkmale im anderen Stereobild als Kandidaten in Frage kommen. Das menschliche visuelle System hat damit im

allgemeinen nur geringe Schwierigkeiten, wenn von periodischen Mustern (z.B. kariertes Stoff) abgesehen wird.

Zur Lösung des Korrespondenzproblems werden in der Literatur zwei verschiedene Verfahren vorgeschlagen, die **merkmalsbasierte** und die **intensitätsbasierte** Korrespondenzanalyse [KIKo96]. Die Auswahl des Verfahrens hängt von der Anzahl der benötigten Tiefenwerte und der Geschwindigkeit ab.

Innerhalb der **intensitätsbasierten** Korrespondenzanalyse in Stereobildern (area-based stereo) [KIKo96] wird die Annahme getroffen, daß korrespondierende Pixel einen ähnlichen Intensitätswert besitzen. Diese Annahme allein ist jedoch nicht ausreichend für eine eindeutige Zuordnung zwischen den Pixeln, da in der Regel eine große Anzahl identischer Intensitätswerte in einem Bild auftreten können. Zudem können korrespondierende Punkte in beiden Bildern aufgrund unterschiedlicher Blickrichtungen und einem unterschiedlichen Kameraabgleich nie den selben Intensitätswert besitzen. Die Disparität, also die Pixelverschiebung, muß somit aufgrund der Korrelation zwischen mehreren benachbarten Intensitätswerten getroffen werden. Diese aufwendige und schwierige Korrespondenzanalyse bedingt folglich eine langsame, aber dafür fast vollständige Rekonstruktion des Tiefenbildes.

Wird die Korrespondenzanalyse anstatt mit Intensitätswerten mit ausgewählten Bildmerkmalen durchgeführt, so wird diese Technik als **merkmalsbasierte** Korrespondenzanalyse (feature-based stereo) [KIKo96] bezeichnet. Hierbei werden besonders markante Teile des Bildes, wie z.B. Kanten oder Intensitätssprünge, als Merkmale herangezogen. Die Zuordnung erfolgt in der Regel aufgrund ausgewählter Eigenschaften dieser Merkmale, wie z.B. Orientierung oder Länge der Kanten. Durch die geringere Anzahl von Merkmalszuordnungen kann davon ausgegangen werden, daß die Geschwindigkeit des Algorithmus gewöhnlich wesentlich schneller und aufgrund der besseren Lokalisierbarkeit der Merkmale genauer als bei dem intensitätsbasierten Verfahren stattfindet. Als Nachteil muß dafür die nicht vollständige Rekonstruktion des Tiefenbildes in Kauf genommen werden. Liegen innerhalb einer Szene nicht genügend Merkmale vor, so können sie mit Hilfe einer strukturierten Beleuchtung künstlich erzeugt werden.

3.2.1.2 Dynamische Stereoanalyse

Die Wahrnehmung und Interpretation von Bewegungen spielen eine wichtige Rolle innerhalb des menschlichen Sehsystems. Unser Sehsystem ist sehr empfindlich für Bewegungen, was sich an unserer Neigung zur Fokussierung von sich bewegenden Objekten erkennen läßt. Bewegungslose Objekte hingegen können nur schwer aus der Menge an Bildinformationen extrahiert werden. Diese Eigenschaft wird häufig in der Tierwelt zur Tarnung genutzt. Aus der Bewegung können wir Objekte und deren Geschwindigkeit erkennen und darauf aufbauend die relative Entfernung, den zurückgelegten Weg, die Bewegungsrichtung und die Objektgröße extrahieren.

Die Tiefengewinnung mit Hilfe von Bildsequenzen entspricht dem statischen Stereoansatz mit dem Unterschied, daß anstatt zwei oder mehrerer Bilder einer Szene eine Bildsequenz aufgenommen wird, in welcher die Kamera, das Objekt oder beide sich bewegen. Die Entfernung bzw. Geschwindigkeit wird entsprechend dem statischen Ansatz aus der Verschiebung zweier identischer Merkmale bestimmt. Dazu werden in einem ersten Schritt aus den Bildern die Verschiebungsvektoren ($V V$) bestimmt und diese darauf aufbauend interpretiert.

Zur Bestimmung von $V V$ (also Bewegungen) sind Änderungen innerhalb des Grauwertbildes notwendig [Jä93]. Der Umkehrschluß, daß Grauwertänderungen von Bewegungen im Bild verursacht werden, ist aber nicht richtig. Zu große Bewegungen oder zu große Zeitabstände zwischen den Aufnahmen können dazu führen, daß die korrespondierenden Punkte aus den Bildern nicht mehr eindeutig zuzuordnen sind. Besonders schwierig ist das Problem der Verdeckung. Gegenstände können hinter anderen verschwinden bzw. wieder auftauchen. Zu den neu auftauchenden bzw. verschwindenden Objektpunkten gibt es keine korrespondierenden Punkte im anderen Bild.

Innerhalb des gesamten Verarbeitungsprozesses zur Bewegungs- und Entfernungsbestimmung können bereits kleine Fehler z.B. in der Merkmalszuordnung große Probleme in der nachfolgenden Verarbeitung bewirken. Einschränkungen bzw. a priori Informationen bezüglich der Bewegungsrichtung, der Geschwindigkeit, der Objektgeometrie und andere können helfen, dieses komplexe System einzuschränken und damit die Lösung einfacher, schneller und sicherer zu machen.

3.2.1.3 Räumlich kodierter Lichtansatz

Eine bekannte und effiziente Methode zur Gewinnung von Tiefenbildern ist der räumlich kodierte Lichtansatz (siehe Robmann [Ro94], Jiang & Bunke [JiBu93], Scheffe [Se95]). Hierbei wird eine Folge von Streifenmustern auf die Szene projiziert und von einer Kamera aufgenommen (Abb. 3.3). Der Tiefenwert kann nach durchlaufen aller Lichtmuster aus der sich ergebenden Hell-Dunkel-Sequenz eines jeden binärisierten Kamerapixels mit Hilfe der Triangulation berechnet werden. Bei vollständiger Beleuchtung der Szene ergibt sich ein zu dem Tiefenbild bildpunktweise übereinstimmendes Grauwertbild.

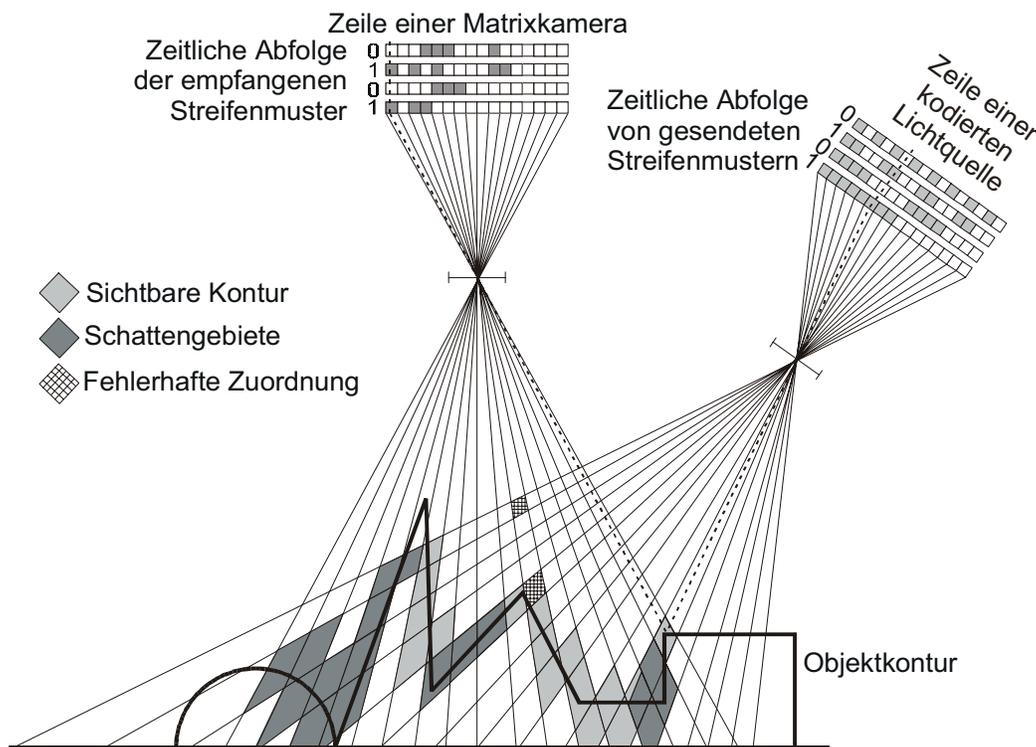


Abb. 3.3: Funktionsprinzip des räumlich kodierten Lichtansatzes (Schnitt)

Die Genauigkeit der Tiefenberechnung eines beleuchteten Punktes hängt entscheidend von der richtigen Bestimmung der aufgenommenen Hell-Dunkel-Sequenz ab [JiBu93]. Bereits eine falsche Zuordnung eines Helligkeitswertes führt zu einem fehlerhaften Tiefenwert. Als mögliche Fehlerquelle für diesen Zuordnungsfehler kommen mehrere Möglichkeiten in Betracht. Liegt z.B. der betrachtete Bildpunkt genau auf einem Übergang zwischen einem beleuchteten und einem nichtbeleuchteten Lichtstreifen (Abb. 3.3), kann die nachfolgende Bildverarbeitung keine genaue Zuordnung treffen. Weitere Fehler werden z.B. durch Diskontinuitäten der Meßobjekte, durch Spiegelungen, durch zu dunkle Oberflächen, durch Bewegungen bzw. Vibrationen oder durch nicht konstante Umge-

bungsbeleuchtung erzeugt. Der Fehler selbst hängt von der räumlichen Distanz zwischen der fehlerhaft erkannten Sequenz und der richtigen Sequenz ab. Je kleiner diese Distanz ist, desto geringer ist der Tiefenfehler. Folglich kann mit der Wahl des projizierten Codes (z.B. Gray-Code anstatt des Binär-Codes) der Fehler in Abhängigkeit der Applikation verringert werden.

3.2.2 Photometrische Verfahren

So wie der Mensch relative Entfernungen und Orientierungen von Objekten anhand der Helligkeitsverteilung erkennt, basieren die photometrischen Verfahren auf der Grundlage, die Oberflächenorientierung und/oder Entfernung der Meßpunkte anhand der Grauwertverteilung zu bestimmen. 1977 führte Horn [Ho77] ausgehend von früheren Arbeiten in der Photogrammetrie erstmals eine quantitative Beschreibung für die aufgenommene Bildintensität in Abhängigkeit der Oberflächenorientierung unter Berücksichtigung der Anordnung von Kamera und Lichtquelle ein:

$$I(x, y, z) = \kappa \cdot I_i(x, y, z) \cdot \Phi(\mathbf{n}(x, y, z), \mathbf{s}(x, y, z), \mathbf{v}(x, y, z)) \quad (3.1)$$

Mit dieser Formel wird die von einer Kamera aufgenommene Intensität I in Abhängigkeit der Kamerakonstanten κ der Eingangsintensität I_i und der Reflexionsfunktion Φ an einem räumlichen Punkt $P(x, y, z)$ beschrieben (Abb. 3.4). Die Reflexionsfunktion selbst ist dabei abhängig von der lokalen Oberflächenorientierung $\mathbf{n}(x, y, z)$, der Beleuchtungsrichtung $\mathbf{s}(x, y, z)$ und der Aufnahmerichtung $\mathbf{v}(x, y, z)$ im Meßpunkt. Diese Formel ist somit die Voraussetzung zur photometrischen Rekonstruktion dreidimensionaler Objektmerkmale. Hauptvertreter dieser Kategorie sind Shape-from-Shading und die photometrische Stereoanalyse.

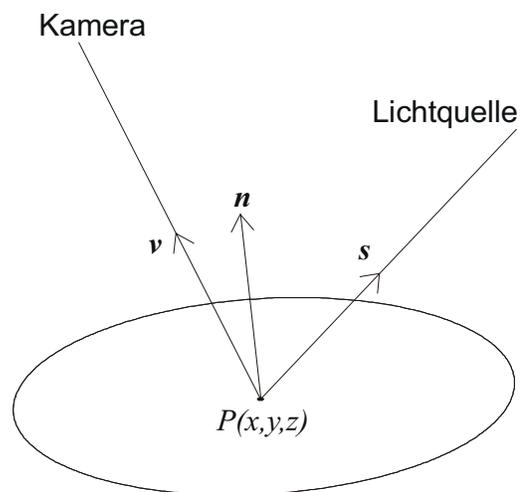


Abb. 3.4: Geometrie der Bildentstehung

3.2.2.1 Shape-from-Shading

Das Ziel von Shape-from-Shading (SFS) ist die Rekonstruktion der Kontur eines Objektes von einem gegebenen 2D-Grauwertbild. Bereits 1970 wurde diese Problematik von Horn

in seiner Doktorarbeit wie folgt beschrieben [Ho70]: „Give an intensity image of a continuous surface with constant, known reflectance and illumination, recover the shape of that reflectance“.

Zur Bestimmung der Oberflächenkontur muß eine nichtlineare Gleichung gelöst werden. Aufgrund dessen sind Restriktionen bei der Lösung des Problems notwendig. Die erste und weitreichendste Einschränkung ist die Annahme einer diffusen Reflexion. Das bedeutet, die Oberfläche reflektiert das einfallende Licht nur in Abhängigkeit der Flächennormalen und der Beleuchtungsrichtung (Lambert-Reflexion). Die zweite Vereinfachung nimmt an, daß die Beleuchtung der Szene von einer einzigen Punktlichtquelle erfolgt, welche unendlich weit entfernt positioniert ist. Weitere Vereinfachungen zur Lösung der Problematik verlangen, daß die Beleuchtungsposition bekannt ist und das die zu rekonstruierende Oberfläche stetig oder stetig differenzierbar ist. In einigen Fällen wird zudem vorausgesetzt, daß Höhenwerte bzw. Oberflächenorientierungen in gewissen (singulären) Punkten bekannt sind. All diese Restriktionen vereinfachen die Problemstellung bis zu einem gewissen Grad. Mit jeder Einschränkung jedoch sinkt der praktische Nutzen von SFS.

Nach Aufzählung dieser Nachteile des SFS-Verfahrens sollten die Vorteile nicht unterschlagen werden. SFS rekonstruiert im allgemeinen aus der Aufnahme eines einzigen Bildes eine dicht besetzte Tiefenmatrix einer homogenen Oberfläche.

3.2.2.2 Photometrische Stereoanalyse

Die Idee der photometrischen Stereoanalyse geht auf die Arbeit von Woodham [Wo80] zurück. Unter photometrischer Stereoanalyse wird die Erweiterung des im vorherigen Kapitel angesprochenen SFS-Verfahrens zur schattierungsbasierten Oberflächenrekonstruktion mit zwei oder mehr Bildern verstanden. Durch die größere Datenmenge werden die Rekonstruktionsergebnisse verbessert und gleichzeitig die Zahl der notwendigen Restriktionen verringert.

Zur Durchführung der photometrischen Stereoanalyse wird das stillstehende Objekt nacheinander von mehreren Beleuchtungsquellen bestrahlt und von einer ebenfalls stillstehenden Kamera aufgenommen. Damit können einem abgebildeten Oberflächenpunkt mehrere Irradianzen zugeordnet werden, ohne mit einem Korrespondenzproblem, wie bei den herkömmlichen Stereoansätzen, konfrontiert zu werden. Die photometrische Stereoanalyse setzt im allgemeinen stetige und differenzierbare Intensitätsverteilungen voraus,

d.h. man geht von Oberflächen ohne Textur, Ecken und Kanten aus bzw. man muß diese durch eine entsprechende Segmentation von der Analyse ausschließen.

Zur Bestimmung der Oberflächenorientierung werden Angaben zur Position, Orientierung und Leuchtstärke der einzelnen Lichtquellen benötigt. Jede Ungenauigkeit dieser Werte führt zu einer Beeinträchtigung der Rekonstruktion der Oberflächennormalen bzw. Objektdistanz. Dies kann unterdrückt werden, indem mehr Bilder als notwendig aufgenommen werden oder der Einfluß der wahrscheinlich fehlerhaften Parameter reduziert wird. Die LookUp Technik kommt ohne Kenntnisse dieser Parameter aus. Dieser Verarbeitungsweg basiert auf einer vorangegangenen Justierphase des Systems mit Hilfe eines bekannten Kalibrierobjektes mit konstanten Oberflächeneigenschaften. Folglich führt die LookUp Technik zu einer einfachen und schnellen Tiefenbestimmung mit der Einschränkung, daß die Oberflächeneigenschaften des Meßobjektes dem des Kalibrierobjektes entsprechen müssen.

3.2.3 Laufzeitverfahren

Die Laufzeitverfahren arbeiten nach dem Prinzip der Laufzeitmessung eines ausgesendeten Signals (z.B. Laser, Mikrowellen oder Ultraschall), das vom Meßobjekt reflektiert und vom Empfänger, der sich im allgemeinen nahe des Sensors befindet, empfangen und verarbeitet wird. Aus der Zeit zwischen Senden und Empfangen des Impulses oder der Phasenverschiebung zwischen dem gesendeten und empfangenen Signal läßt sich der zurückgelegte Weg berechnen.

3.2.3.1 Scannende Systeme

Das wohl bekannteste Verfahren unter den Laufzeitverfahren zur Aufnahme einer flächendeckenden Szene ist der Laserscanner. Aufgrund der teuren, aufwendigen und großen Sende-Empfangselektronik benutzen die scannenden Systeme nur einen einzigen Entfernungssensor (Abb. 3.5). Zur flächendeckenden Vermessung der Szene wird das Signal daher über eine Spiegeleinheit sowohl horizontal als auch vertikal abgelenkt (siehe z.B. Nitzan & Brain & Duda [NiBr77], Perceptron [Pe93] und Weber [We98]).

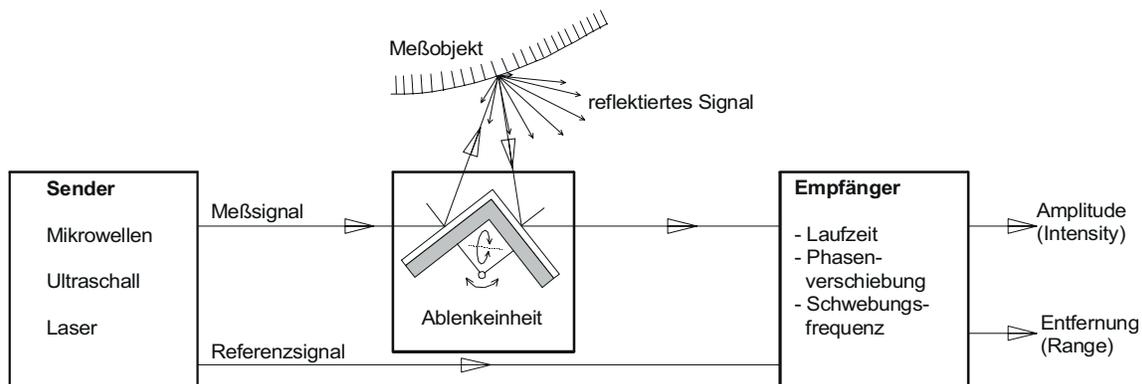


Abb. 3.5: Schema eines Laserscanners [NiBr77]

Als Auswahlkriterien für das zu verwendende Signal sind nach Jiang & Bunke [JiBu97] im wesentlichen drei verschiedene Eigenschaften zu berücksichtigen:

- Ø Bündelungsfähigkeit (laterale Auflösung)
- Ø Ausbreitungsgeschwindigkeit (Tiefenauflösung)
- Ø Reflexionseigenschaft

Die mechanische Ablenkung des Sendestrahls bedingt eine zeitaufwendige Vermessung der gesamten Szene, innerhalb derer sich das Meßobjekt nicht bewegen darf. Aufgrund der in den meisten Scannern vorhandenen parallelen Sende- und Empfangsstrahlen tritt keine Schattenbildung auf. Ebenso kann neben der doppelten Objektentfernung aus der empfangenen Leistung des Empfangssignals ein Intensitäts- bzw. Reflexionsbild ermittelt werden.

3.2.3.2 Pockels-Zelle

Ein wesentlicher Nachteil der meisten Laufzeitsysteme ist ihre große Anfälligkeit für Störungen (Rauschen, Temperaturdrift ...) aufgrund der diskreten Empfangs- und Demodulationseinheit. Innerhalb des Ansatzes der Pockels-Zelle wird dieses Problem gelöst, indem die notwendige Modulation und Demodulation des Meßsignals bereits auf optischem Wege erfolgt [HeXu97]. Der Einsatz von optischen Modulatoren stellt eine quasi drifffreie Referenzmethode dar. Das eigentliche Empfängerelement, in diesem Falle eine herkömmliche CCD- oder CMOS-Matrixkamera, empfängt nur noch das niederfrequente demodulierte Signal.

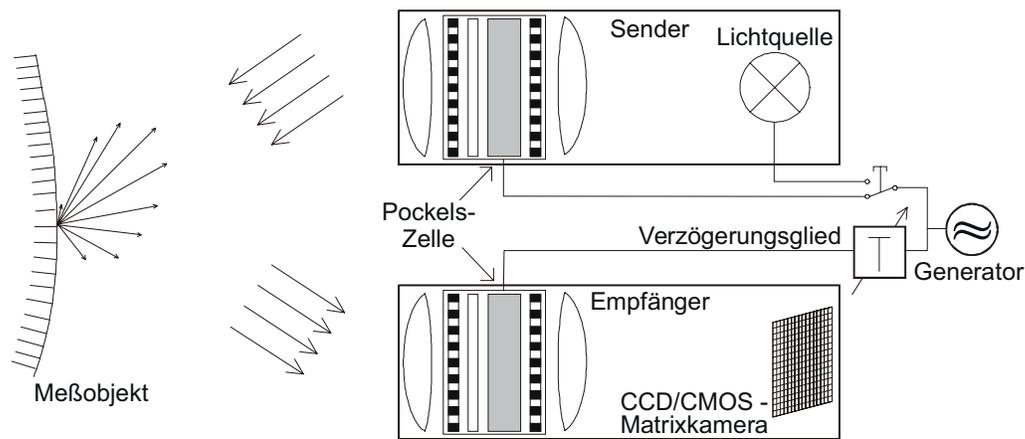


Abb. 3.6: 3D-Kamerasystem mit Pockels-Zellen Demodulator [HeXu97]

Die optische Modulation und Demodulation hat neben den bereits erwähnten Vorteilen bzgl. der Pockels-Zelle den Nachteil, daß sie aufgrund des notwendigen polarisierten Lichtes einen Großteil der optischen Leistung verbraucht. Durch die gesetzliche Begrenzung der Laserleistung muß das somit schwache optische Empfangssignal aufwendig aufbereitet werden. Dazu bietet sich neben der Auswahl von entsprechend empfindlichen Empfangssensoren auch die Verlängerung der Integrationszeit an. Mit steigender Integrationszeit jedoch wächst auch die Beeinflussung durch Fremdlicht. Der dadurch erzeugte Offset reduziert die Genauigkeit und die Tiefenauflösung.

Weitere Nachteile der Pockels-Zelle sind ihre nichtlinearen Eigenschaften und die hohe notwendige Modulationsspannung [XuSc98]. Die Abhängigkeiten der Pockels-Zelle von der optischen Wellenlänge und dem Betrachtungswinkel beeinträchtigen weiterhin die Modulationstiefe, die Bildqualität und die Tiefenauflösung.

3.2.3.3 Bildverstärker

Die Entfernungsmessung auf Basis eines Bildverstärkers ist ein weiterer Vertreter der Time-of-Flight Technik. Wie der Name bereits andeutet, wird das zurückgestreute Licht der Szene auf einen Bildverstärker (Image Intensifier) abgebildet, welcher das einfallende Licht unter Erhaltung der lateralen Bildinformation und in Abhängigkeit eines Modulations-signals verstärkt. Das entfernungsproportionale Signal wird anschließend von einer herkömmlichen Kamera aufgenommen und verarbeitet.

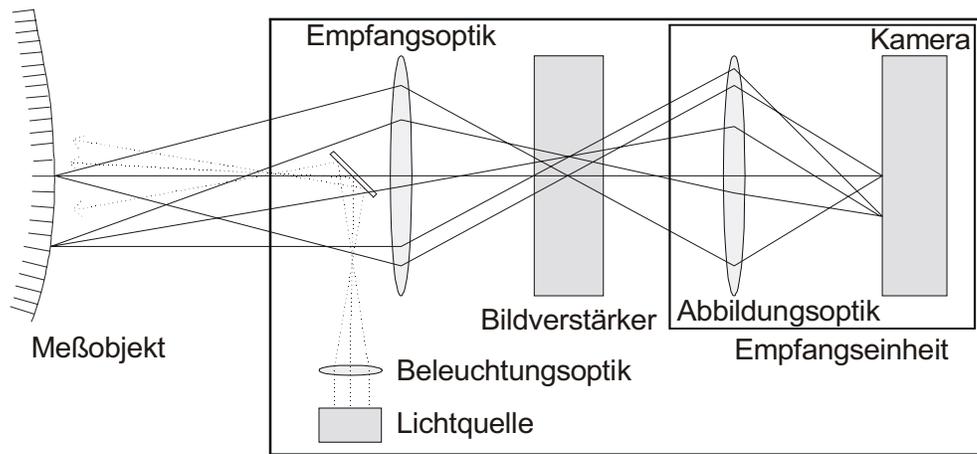


Abb. 3.7: Funktionsprinzip der 3D-Kamera auf Basis eines Bildverstärkers [HöJe98]

Die Genauigkeit und die Auflösung des Entfernungsmesssystems auf Basis eines Bildverstärkers werden durch eine Reihe von Parametern bestimmt. So beeinflusst der Laser, der zur Beleuchtung eingesetzt wird, das Verhalten des Systems durch den Speckle Effekt (hervorgerufen durch die kohärente Natur des Laserlichtes). Die Dynamik der Kamera legt fest, mit welcher Auflösung die Intensität der Pixel gemessen werden können. Die laterale Auflösung wird durch den Abbildungsmaßstab und die Anzahl der Pixel des CCD-Chips bestimmt. Zur Gewährleistung einer homogenen Auflösung ist ein nahezu konstanter Modulationsgrad des Bildverstärkers über die gesamte Apertur notwendig.

3.3 Zusammenfassung

Obwohl die hier aufgeführten Verfahren nur Trends der Arbeiten zur automatischen Erfassung von 3D-Objektdaten darstellen, wird deutlich, in welcher Breite versucht wird, auf diesem Gebiet Lösungen zu finden. Aus den vorgestellten Verfahren ist ersichtlich, daß es kein perfektes 3D-Aufnahmesystem gibt. Alle Verfahren besitzen ihre spezifischen Vor- und Nachteile. Sie arbeiten nur für bestimmte Objektklassen und Randbedingungen optimal. Gemäß den 2D-Bildverarbeitungssystemen muß auch im 3D-Fall die Kamera und Beleuchtung optimal auf die Problemstellung abgestimmt werden.

Zur Bewertung der einzelnen Verfahren bzw. zur Bestimmung der applikationsspezifischen Anforderungen können folgende Kriterien angewendet werden:

- ∅ Aufnahmezeit pro Szene bzw. Meßpunkt
- ∅ Rechenzeit
- ∅ Tiefenauflösung / Genauigkeit
- ∅ Laterale Auflösung
- ∅ Meßbereich

- Ø Berechnungskomplexität
- Ø Signal / Rausch Verhalten (Fehleranfälligkeit)
- Ø Fremdlichtabhängigkeit
- Ø Oberflächenbeschaffenheit (Homogen / Texturiert)
- Ø Berechnung der Orientierung und/oder Entfernung
- Ø Schattenproblematik
- Ø Anzahl der notwendigen Komponenten (Beleuchtung, Kamera, usw.)
- Ø Kosten
- Ø Sicherheit / Augensicherheit
- Ø Leistungsverbrauch
- Ø Kompaktheit des Gerätes
- Ø ...

Aus dieser Aufzählung wird deutlich, daß keines der Verfahren alle Punkte abdeckt, bzw. daß die benötigte Anwendung alle Punkte erfordert. Wurde ein Verfahren für eine Applikation ausgewählt, sollte ferner die Fehleranfälligkeit der Tiefenwerte berücksichtigt werden. Keines der hier vorgestellten Verfahren liefert ein fehlerfreies Tiefenbild. Dies muß in Folge dessen von der nachfolgenden BV berücksichtigt werden. Einzelfehler können bereits mit einfachen Filtern entfernt bzw. ausgeglichen werden. Flächenhafte Fehler sind nur schwer erkennbar. Hier können nur a priori Informationen weiterhelfen.

Unter dem Aspekt, daß keines der aufgeführten Verfahren ein Optimum an Qualität aufweist, kann evtl. eine Kombination mehrerer Verfahren, angelehnt beispielsweise an den menschlichen Wahrnehmungsprozeß, eine erhöhte Qualität der Rekonstruktion erzeugen. Zur vollständigen Beschreibung eines Bildinhaltes wären z.B. denkbar:

- Ø Kombination von photometrischen Verfahren mit der Stereoanalyse auf Basis der merkmalsbasierten Korrespondenzanalyse (siehe Cryer & Shah [CrTs92])
- Ø Verknüpfung von merkmalspezifischen und photometrischen Symbolen in einem Graph zur Objektidentifikation (siehe Straub [St91])
- Ø Erfassung der Szene durch mehr als 2 Bildaufnahmen aus prinzipiell beliebigen Richtungen (siehe Straub [St91])
- Ø Kombination von statischer und dynamischer Stereoanalyse
- Ø Verifizierung und ggf. erneute Analyse der Daten mit Hilfe einer Wissensdatenbank
- Ø ...

Eine weitere und zur Zeit häufig anzutreffende Methode zur Steigerung der Qualität ist die Verwendung von Farbinformationen. Die gleichzeitige Aufnahme von drei Farbbildern anstatt eines Grauwertbildes vereinfacht beispielsweise die Zuordnungsproblematik innerhalb der Stereoanalyse wesentlich (siehe Klette & Koschan [KIKo95]). Im Bereich der Triangulation reduziert sich durch die farbige Beleuchtung und Aufnahme der Szene die Anzahl der kodierten Beleuchtungssequenzen. Im Idealfall kann bereits aus einer Aufnahme die vollständige Tiefeninformation rekonstruiert werden (siehe Schubert [Scu96]). Neben diesen Vorteilen der Farbinformation dürfen die Nachteile nicht vergessen werden. Abgesehen von dem erhöhten Hardwareaufwand sollte die zu vermessende Szene möglichst einfarbig sein. Die Erfassung beliebig farbiger Objekte ist problematisch.

4 3D-Bildverarbeitung

Unter dem Begriff Bildverarbeitung wird die Aufgabe verstanden, aus Bildern eine Beschreibung der aufgenommenen Szene abzuleiten. Die zu extrahierenden Merkmale hängen dabei stark von der Aufgabenstellung ab. Innerhalb einer Gut/Schlecht Überprüfung wird z.B. die aufgenommene Szene nach vordefinierten Fehlermerkmalen (Maßhaltigkeit, Oberflächenprüfung, abgebrochene Kante, usw.) untersucht. Im einfachsten Fall kann dies bereits mit einer einfachen Schwellwertoperation erfolgen. Im Gegensatz hierzu wird für die Objekterkennung eine möglichst vollständige Beschreibung der Szene benötigt. Hier können einfache Operatoren nicht angewendet werden. Vielmehr wird eine aufwendige Klassifizierung der Einzelsegmente mit anschließender Zuordnung benötigt.

Innerhalb dieses Kapitels soll ein kurzer Überblick über prinzipielle Anwendungsgebiete, Strategien und Verfahren der 3D-BV gegeben werden. Hierbei soll das Augenmerk hauptsächlich auf Änderungen von der 2D-BV zur 3D-BV gelegt werden bzw. was in Bezug auf die 3D-BV zu berücksichtigen ist. Für eine tiefergehende Einführung in die 3D-BV wird auf folgende Literatur verwiesen: Besl & Jain [BeJa85]; Faugeras [Fa93]; Franke & Gaßmann [FrGa96]; Jiang & Bunke [JiBu93]; Shirai [Sh87]; Suk & Bhandarkar [SuBh92]; Zhang [Zh95]. Ein interessantes Nachschlagewerk ist dabei der interaktive BV-Kurs im Internet [MaPr96].

4.1 Aufgaben und Anwendungsgebiete der 3D-Bildverarbeitung

Wie die 2D-BV kann auch die 3D-BV vielfältig eingesetzt werden. Aufgrund der Aufnahmeart jedoch bieten 3D-Kameras bedeutende Vorteile gegenüber herkömmlichen 2D-Bildaufnahmesystemen. Die tatsächlichen Eigenschaften der Szene wie Tiefe und Oberflächenorientierung vereinfachen die Segmentierung, Klassifizierung und die Objekterkennung grundlegend. Diese Eigenschaften können aufgrund der mehrdeutigen Projektion der 3D-Szene auf eine Ebene nur schwer aus den Daten einer 2D-Kamera bestimmt werden. Folglich werden Entfernungsdaten eine in zunehmendem Maße wichtige Quelle der Informationen für eine Vielzahl von Anwendungen.

Die Formprüfung ist für die 3D-BV ein Paradebeispiel. Ihre Aufgabe ist die Überprüfung der Übereinstimmung zwischen Objekt- und Sollform. Ohne die Entfernungsinformation ist mit der 2D-BV die Aufnahme mehrerer Bilder zur vollständigen Prüfung der Maßhaltigkeit notwendig. Im Bereich der Objektsortierung wird erst mit Hilfe der Tiefeninformation die

Erkennung und Lokalisierung von Objekten, welche auf einem Haufen liegen, ermöglicht. Das zum Greifen am günstigsten liegende Objekt kann anschließend von einem Roboter aussortiert werden.

Ein immer aktueller werdendes Anwendungsgebiet für die 3D-BV ist die Navigation von fahrerlosen Fahrzeugen bzw. die Navigationshilfe in Kraftfahrzeugen. Nur im dreidimensionalen Raum kann eine Verfolgung der Straße und ein Erkennen von Hindernissen (Art, Größe und Entfernung) sinnvoll durchgeführt werden. Zusätzliche Angaben zur Umgebung ermöglichen dem nachfolgenden Verarbeitungsprozeß ein vorausschauendes Handeln. Weiterhin fordern die gestiegenen Sicherheitsanforderungen im Straßenverkehr neue Lösungen, welche z.T. nur mit der Aufnahme und Verarbeitung der räumlichen Umgebung außerhalb und innerhalb des Kraftfahrzeuges gelöst werden können.

4.2 Allgemeine Vorgehensweise

Die allgemeine Vorgehensweise der 3D-BV entspricht im wesentlichen der Vorgehensweise der 2D-BV. Nach [JiBu97] wird das allgemeine Schema eines 3D-Bildanalyse-Systems wie folgt beschrieben (Abb. 4.1):

Der erste Schritt eines 3D-Bildanalyse-Systems besteht in der Aufnahme der Szene und damit in der räumlichen und wertmäßigen Abtastung bzw. Diskretisierung. Hierauf folgt üblicherweise die Vorverarbeitung, die als Zielsetzung eine Verbesserung der Tiefen- und Grauwertqualität hat.

Eine Szenenbeschreibung setzt sich meistens aus Objekten in einer Hierarchie verschiedener Abstraktionsstufen zusammen [JiBu97]. Die Objekte auf den untersten Stufen bilden hierbei Merkmale wie Kanten und Flächen. Diese werden in einem Segmentierungsschritt extrahiert. Zusammen mit

ihren Attributen ergibt sich daraus eine merkmalsbasierte bzw. symbolische Szenenbeschreibung, die einem Zuordnungsprozeß (Merkmalsextraktion/ Klassifikation) zugeführt wird. In diesem werden die Merkmale gruppiert, und es wird versucht, auf Basis einer Wissensbasis ein Objekt einer höheren Abstraktionsebene zuzuordnen. Fehlende Merkmale werden dabei in diesem Schritt ergänzt, bzw. fehlerhafte Merkmale werden korrigiert. Als Ergebnis erhält man eine Klasse von Objekten aus einer Wissensbasis. Neben der

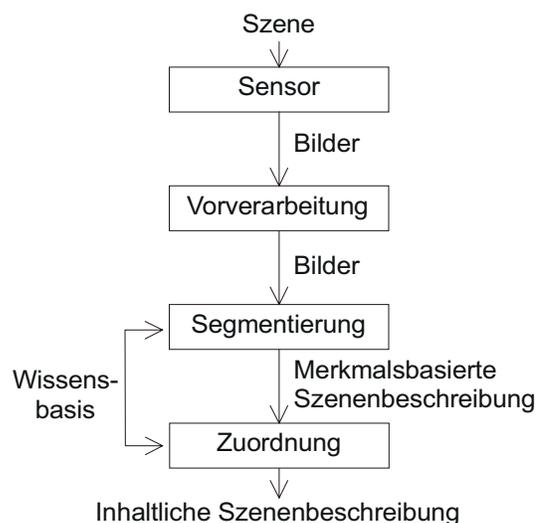


Abb. 4.1: Allgemeines Schema eines Bildanalyse-Systems

Nutzung dieses Wissens im Zuordnungsprozeß kann dieses a priori Wissen auch zur Verbesserung der Segmentierung herangezogen werden. Durch das Umdrehen der Reihenfolge, d.h. Rekonstruktion des vorangegangenen Schrittes, können einzelne Verarbeitungsschritte verifiziert werden.

In den nachfolgenden Kapiteln sollen auf Basis dieser Grundlage besondere Aspekte der Vorverarbeitung und Segmentierung bezüglich der 3D-BV kurz angesprochen werden. Aufgrund der Komplexität der Zuordnung wird dieses Verfahren nur sehr allgemein diskutiert. Zu beachten ist, daß die erläuterten Methoden von der Applikation und dem Aufnahmeverfahren abhängig sind. Nicht jede Methode ist für jede Datenbeschreibung nutzbar.

4.2.1 Vorverarbeitung

Unter Bildvorverarbeitung versteht man eine Bildmodifikation, die den Informationsgehalt des Bildes im Hinblick auf die spätere Auswertung nicht wesentlich verändert [Ab94]. Von der Art der Verarbeitung her ist die Bildvorverarbeitung eine Bildtransformation, bei der das Ergebnis wieder ein Bild ist.

Aufgrund der räumlichen Abtastung, der Meßwertdiskretisierung und der Abhängigkeit vom 3D-Aufnahmeverfahren, weisen 3D-Daten Eigenschaften wie Positionsrauschen, Häufungspunkte, Ausfälle, Überschwinger, Ausreißer u.a. auf. Für die weitere Verarbeitung erweisen sich zudem Datensätze als besonders problematisch, in denen die Punkte in keinem festen Grundraster angeordnet sind und die eine unregelmäßige Dichte der Punkte aufweisen. Nach Franke & Gaßmann [FrGa96] kann eine Einteilung der Verfahren bzgl. des Anliegens der Verarbeitung, aber auch nach der Art der angewandten Methoden erfolgen:

- Ø Beseitigung von Ausreißern
- Ø Glättung / Rauschminderung
- Ø Ergänzung von fehlenden Daten
- Ø Regularisierung
- Ø Minderung oder Beseitigung von Redundanzen
- Ø Kompensation des Tasterradius (nur bei mechanischen Tastern)
- Ø Kombination von Teilansichten (Zuordnung, Überdeckung, Ausrichtung optimieren)

Bedingt aus der Aufgabenstellung kann die 3D-Vorverarbeitung der 2D-Vorverarbeitung prinzipiell gleichgesetzt werden. Einige Methoden der 2D-Vorverarbeitung können

dementsprechend direkt auf 3D-Daten angewendet werden. Als Beispiele seien hier die Beseitigung von Ausreißern und die Glättung erwähnt. Im Falle von Ausreißern einzelner Meßwerte können analog der 2D-BV schwellwertbasierte Verfahren oder Filter (Tiefpaß, Median, Morphologische ...) eingesetzt werden. Entsprechend den Grauwertbildern enthalten auch Tiefenbilder aufgrund von Quantisierungs- und Meßfehlern immer Störungen. In Bezug auf 3D-Daten ist allerdings neben der Glättung besonders auf die Erhaltung der Kanteneigenschaften zu achten, da in diesen wesentliche Informationen gespeichert sind. Schließen aufgrund der Restriktionen herkömmliche Verfahren fehl, müssen speziell angepasste Verfahren, wie z.B. approximierende Verfahren, angewendet werden. Eine besondere Art der Filterung von Tiefenbildern ist die Glättung der Flächennormalen. Hierbei werden Unebenheiten bei stetigen Flächen herausgefiltert.

Spezielle 3D-Vorverarbeitungsverfahren werden für die Fälle benötigt, welche in herkömmlichen 2D-Aufnahmesystemen nicht auftreten. Dies wäre beispielsweise eine Ergänzung fehlender Daten bzw. eine Datenregularisierung, wie sie z.B. bei optischen Verfahren an stark reflektierenden Stellen oder an Stellen starker Steigung auftreten können. Zum Ausgleich dieser Störungen kommen interpolierende Verfahren in Frage. Sowohl interpolierende als auch approximierende Verfahren können eingesetzt werden, wenn Tiefenwerte von einem unregelmäßigen Raster auf ein regelmäßiges Raster überführt werden sollen. Bei beiden Verfahren kann durch die notwendige Vorgabe der Abtaststellen die Lage der neuen 3D-Punkte bzw. die Punktdichte bewußt gesteuert werden [FrGa96].

Aus diesen Beispielen wird ersichtlich, daß ein Großteil der angepassten 3D-Vorverarbeitungen mit approximierenden bzw. interpolierenden Methoden erfolgen können. Bei der Verfahrensauswahl sollte neben meßverfahrensspezifischen Eigenschaften gegebenenfalls die Art und Qualität der digitalisierten Objektoberflächen (Rauhigkeit, Objektkante) berücksichtigt werden. Nach Franke & Gaßmann [FrGa96] lassen sich die einzelnen Algorithmen nach den Kriterien der Anwendbarkeit (Verteilung der Punkte im 3D-Raum: 3D-Punktwolken, Volumen- oder Oberflächendaten) und Lokalität des Verfahrens (globale oder lokale Methoden) weiter spezifizieren. Bei globalen Methoden hängt das Ergebnis eines Algorithmus von allen gegebenen Datenpunkten ab (lange Rechenzeit). Demgegenüber ist das Resultat bei lokalen Methoden nur von den Punkten der lokalen Umgebung abhängig. Der Vorteil der kurzen Rechenzeit wird durch Probleme an den Rändern der lokalen Gebiete ausgeglichen (wobei stetige Übergänge erwünscht sind). Die

Stetigkeitsprobleme können entfallen, wenn die lokalen Methoden gleitend auf sich überlappende Gebiete angewendet werden.

4.2.2 Segmentierung

Die Segmentierung bewirkt die Trennung der zu untersuchenden Objekte vom übrigen Bildinhalt, die Trennung von sich berührenden Objekten und die Zerlegung einzelner Objekte in zusammenhängende Regionen. Unter Regionen können dabei im Falle der intensitätsbasierten BV sowohl durch Kanten getrennte Gebiete, als auch Gebiete mit konstanten oder sich schwach ändernden Grauwertflächen verstanden werden. Für den 3D-Fall muß letztere Bedingung auf Flächen abgeändert werden. Diese werden üblicherweise als zusammenhängend definiert, wenn dieser Bereich sich durch eine Flächenfunktion aus einer Menge von vorgegebenen Funktionen approximieren läßt.

Aus dieser Aufgabenbeschreibung ergibt sich, daß man sowohl kanten- als auch regionenbasiert vorgehen kann. Während letzterer Ansatz versucht, direkt nach homogenen Regionen bezüglich der Funktionsapproximation zu suchen, zielt eine kantenbasierte Segmentierung auf die Ermittlung von Grenzen zwischen den homogenen Bildregionen ab. In beiden Fällen spielen somit die Kanten eine Schlüsselfunktion. Innerhalb der regionenbasierten Vorgehensweise fallen diese als Abfallprodukte an bzw. werden zur Segmentierung für das kantenbasierte Verfahren als Grundvoraussetzung benötigt. Aufgrund der Vielfältigkeit und der zu den grauwertbasierten Verfahren unterschiedlichen Detektionsart soll daher vor der Beschreibung der genannten Segmentierungsarten auf die 3D-Kantentypen eingegangen werden.

4.2.2.1 Kantentypen

Kanten können sowohl aus Grauwert- als auch aus Entfernungsbildern extrahiert werden. In bezug auf die Aufnahmeart sind in beiden Quellen unterschiedliche Kantentypen erkennbar. Innerhalb von Grauwertbildern sind Kanten Kurven, auf welchen sich die Grauwerte oder deren Ableitungen stark ändern. Sichtbare Kanten sind dabei Trennlinien zwischen Objektoberflächen, die unterschiedlich beleuchtet werden bzw. reflektieren. In Abhängigkeit der photometrischen und geometrischen Eigenschaften können aber auch Kanten unsichtbar bleiben oder durch Schattenwurf und Texturen zusätzliche Kanten erzeugt werden.

Nach Schneider [Sc90] können Kanten in Grauwertbildern prinzipiell in die folgenden drei Klassen unterteilt werden:

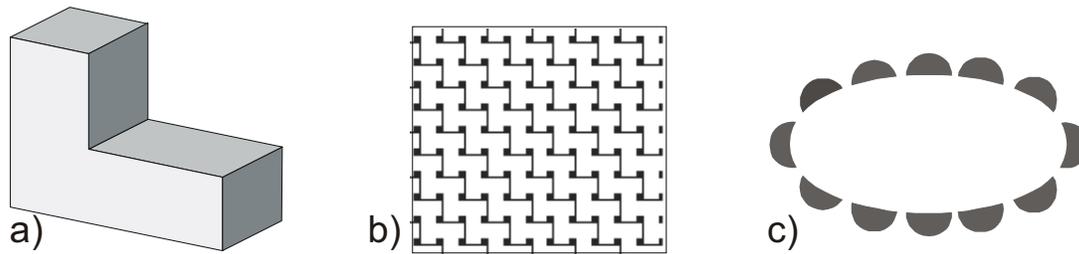


Abb. 4.2: Kanten in Grauwertbildern: a) Grauwertkanten; b) Texturkanten; c) Phantomkanten

- ∅ **Grauwertkanten** (Abb. 4.2 a) sind Trennlinien zweier homogener Gebiete mit unterschiedlichen Grauwerten. Diese Trennlinien können sowohl durch einen abrupten Grauwertsprung als auch durch einen stetigen Grauwertübergang entstehen.
- ∅ **Texturkanten** (Abb. 4.2 b) sind Trennlinien zweier Gebiete unterschiedlicher Mikrostruktur (Textur).
- ∅ **Phantomkanten** (Abb. 4.2 c) sind Trennlinien, die der menschliche Beobachter aufgrund der räumlichen Anordnung von Figuren in einer Szene zu sehen glaubt (Ellipse).

Im Gegensatz zu den Entstehungsmöglichkeiten von Grauwertkanten werden Kanten innerhalb von Tiefenbildern ‚nur‘ durch Änderungen der Oberflächenorientierung erzeugt. Dabei kann ebenfalls zwischen drei Kantentypen unterschieden werden (Abb. 4.3):

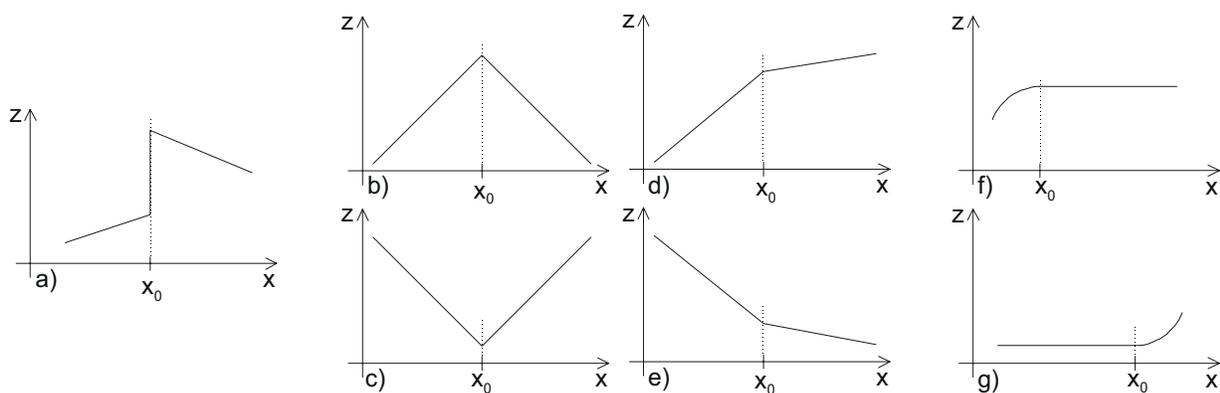


Abb. 4.3: Kanten in Tiefenbildern: a) Sprungkante; b) positive Dachkante; c) negative Dachkante; d) positive Knickkante; e) negative Knickkante; f) positive glatte Kante; g) negative glatte Kante

- ∅ **Sprungkanten** (jump edges) (Abb. 4.3 a) werden durch Diskontinuitäten in den Tiefenwerten erzeugt. Sie entstehen beispielsweise, wenn ein Objekt ein anderes Objekt verdeckt oder sich Teile eines Objektes selbst überdecken.
- ∅ **Schnittkanten** (Abb. 4.3 b-e) werden erzeugt, wenn zwei Oberflächen zusammentreffen bzw. sich die Oberflächenorientierungen abrupt ändern. Innerhalb der Schnittkanten kann in Abhängigkeit der Lage des Knickpunktes eine weitere Unterscheidung erfolgen. Liegt der Knickpunkt in einem Extremum, so handelt es sich um Dachkanten (roof edges), ansonsten werden diese Kanten als Knickkanten (crease edges) bezeichnet.
- ∅ **Glatte Kanten** (smooth edges) (Abb. 4.3 f-g) sind Kanten mit glattem Übergang, also ohne abrupte Änderung der Oberflächennormale. Sie entstehen beispielsweise an Verbindungsstellen zwischen einer ebenen und einer gekrümmten Fläche.

Aufgrund der unterschiedlichen Natur der verschiedenen Kantentypen werden unterschiedliche Techniken zu Erkennung angewendet. Während bei den grauwertbasierten Verfahren die Kantendetektion auf der Suche nach Extremen oder Nulldurchgängen innerhalb der ersten oder zweiten Ableitung beruht, werden kompliziertere Algorithmen für Tiefenbilder benötigt. Ein oft angewendetes Verfahren ist die Kantendetektion basierend auf Änderungen in den Oberflächennormalen. Weitere Vorgehensweisen arbeiten beispielsweise mit der morphologischen Residuenanalyse [GuKr91] [JiBu97] oder versuchen die Tiefenkante einem vorgegebenen Modell anzupassen [GM94] [JiBu97]. Während eine Reihe von Verfahren zur Detektion von Sprung- und Schnittkanten bekannt sind, existieren nur wenige Verfahren zur Erkennung von glatten Kanten. Diese sind aufgrund des sanften Überganges nur sehr schwer detektierbar und können folglich kaum von homogenen Regionen unterschieden werden.

Neben dem prinzipiellen Problem der schwierigen Zuordnung der berechneten Kante zum tatsächlichen geometrischen Ort stellt die Schwellwertbildung zur Extraktion der eigentlichen Kanten ein weiteres Problem dar. Durch ungünstige Wahl der Schwelle können entweder Kanten als breite Linien im resultierenden Bild erscheinen oder im Kontrast hierzu zum Teil oder ganz herausgefiltert werden. Beide Möglichkeiten sollten sowohl durch eine Kantenverdünnung unter Berücksichtigung der tatsächlichen Kantenposition, als auch durch ein Verfahren zur Schließung von Lücken korrigiert werden.

Von der Auswahl des Kantendetektors hängt primär die Qualität der nachfolgenden Verfahren und des Gesamtergebnisses ab. Für die Erkennung von Kanten wurden daher von Canny [Ca86] folgende auch auf Tiefenbilder anwendbare Kriterien aufgestellt:

- Ø **Gute Erkennung** (good detection); Trotz Rauschen und anderen Störungen innerhalb des Bildes sollten alle tatsächlichen Kanten möglichst als solche erkannt und nicht vorhandene Kanten nicht erkannt werden.
- Ø **Gute Lagebestimmung** (good localization); Erkannte Kanten sollten so nah wie möglich an dem Mittelpunkt der tatsächlichen Kante liegen.
- Ø **Einzelkantenerkennung** (only one response to a single edge); Liegt mehr als eine Möglichkeit zur Kantendetektion vor, kann nur eine richtig sein.

4.2.2.2 Regionen- und kantenbasierte Segmentierung

Die Segmentierung auf Basis von Tiefenbildern entspricht, wie bereits erwähnt, im wesentlichen der Segmentierung von Grauwertbildern. In beiden Fällen kann eine Unterscheidung der Segmentierungsverfahren zwischen regionen- und kantenbasierten Verfahren erfolgen. Während kantenbasierte Verfahren darauf beruhen, zusammengehörige Kanten im Tiefenbild zu finden, geht es bei den regionenbasierten Verfahren darum, Gruppen benachbarter Bildpunkte zu finden, die bezüglich eines bestimmten Kriteriums homogen sind. Dazu sollte das Homogenitätskriterium so formuliert sein, daß die daraus entstehenden Regionen typischerweise einer Fläche eines Objektes entsprechen [JiBu97]. In realen Bildern jedoch sind Oberflächen über ihr gesamtes Gebiet nicht glatt bzw. lassen sich nur schwer durch mathematische Formulierungen beschreiben. Rauschen erschwert ferner das Finden von zusammenhängenden Regionen. Dieser rechenaufwendige Prozeß steht somit im Kontrast zu der aufwendigen Bestimmung von zusammenhängenden Kanten. Auch hier verursacht das Rauschen Lücken innerhalb der Kontur. Keiner der beiden Ansätze kann folglich als optimal bezeichnet werden. Infolgedessen werden zunehmend Methoden entwickelt, welche die Vorteile beider Verfahren miteinander kombinieren (siehe Zhang & Zhao [ZhZh95]).

4.2.3 Zuordnung

Der Zuordnungsprozeß ist innerhalb der Bildverarbeitung der aufwendigste und komplizierteste Arbeitsschritt. Der Prozeß selbst kann in die beiden Kategorien Merkmalsextraktion und Klassifikation unterteilt werden.

Bei der Merkmalsextraktion handelt es sich um einen Arbeitsschritt, welcher aus einer Bildmatrix einen Merkmalsvektor erzeugt, der Informationen über wichtige Eigenschaften des Bildes oder der Bildobjekte enthält [Ab94]. Somit wird innerhalb der Merkmalsextraktion der entscheidende Schritt zur Datenreduktion durchgeführt.

Die aus der Merkmalsextraktion erzeugten Eigenschaften werden nachfolgend der Klassifikation zugeführt, welche die Zuordnung dieser Eigenschaften zu einer vorgegebenen Objektklasse durchführt. Als Ergebnis der Merkmalsanalyse erhält man eine Liste von signifikanten Merkmalen mit Gewichts- oder Merkmalsbewertungsfaktoren, mit deren Hilfe man eine Trennfunktion berechnen und die einzelnen Objekte in ihre Objektklasse einteilen kann. Typische Ergebnisse einer Klassifikation sind somit die Benennung und die Positions- und Orientierungsbestimmung von Objekten. Weiterführende Ziele können beispielsweise die Bewegungsanalyse von Objekten sein, die Identifizierung von Objekten nach Funktionsgruppen (z.B. Behälter, Tasse, Flasche...) oder die Beschreibung der Szene im allgemeinen Sinne (z.B. Büro, Wohnzimmer ...).

Die prinzipiellen Verfahren für die 3D-Klassifizierung unterscheiden sich dabei nur geringfügig von der 2D-Variante. In beiden Fällen wird versucht, aus einer gegebenen Liste von Merkmalen eine Zuordnung zu einem Objekt durchzuführen. Der Hauptunterschied besteht in der Beschreibung der Merkmale. Innerhalb der 3D-Szene besitzen die Merkmale bedingt durch die Tiefeninformation mehrere und eindeutigere Eigenschaften. D.h. durch somit gewonnene skalierte Informationen kann die Zuordnung einfacher und sicherer erfolgen.

Das zentrale Element der Klassifizierung bildet genau wie beim Menschen die Wissens- oder Datenbasis (erstellt aus einem Lernprozeß oder externen Daten (z.B. CAD)). Während der Mensch ohne offensichtliche Schwierigkeiten in der Lage ist, mit Hilfe seiner Wissensbasis seine Umgebung wahrzunehmen, darin beliebige Gegenstände oder Objekte zu erkennen, diese zu greifen und begrifflich zu erfassen, ist diese scheinbar mühelos erlernte menschliche Leistung jedoch nur schwer durch Algorithmen nachzuahmen. Die Kunst dabei ist, eine Wissensbasis zu erstellen, in der möglichst alle relevanten Informationen für ein Objekt abgelegt sind und gleichzeitig eine Suchmaschine bereitzustellen, welche nicht vollständige, ungenaue oder gar fehlerhafte Beschreibungen dem richtigen Objekt zuordnet.

Wie bereits erwähnt, ist die Klassifizierung einer der aufwendigsten Schritte innerhalb der BV. Zur Vereinfachung können z.B. Einschränkungen bezüglich der Anzahl und Art der Objekte erfolgen. Im Gegensatz dazu kann aber auch weiteres Wissen über die zugrunde liegenden Objekte und die Szene eine Verbesserung der Klassifizierung ermöglichen. Sind neben der Objektgeometrie noch Daten bezüglich der stabilen Lage eines Objektes bekannt, kann die Zuordnung schneller erfolgen. Ebenso können zusätzliche Annahmen zur einfacheren Klassifizierung herangezogen werden, solange sie nicht bestätigt oder zurückgewiesen werden. Ein erkanntes Hausdach muß beispielsweise mit einem Quader angepaßter Größe mit dem Boden verbunden sein.

Für einen tieferen Einblick in die Merkmalsextraktion und Klassifikation wird an dieser Stelle auf folgende Literatur verwiesen: Besl & Jain [BeJa85]; Faugeras [Fa93]; Jiang & Bunke [JiBu97]; Otterbach [Ot95]; Ritter [Ri96]; Rosenfeld [Ro86]; Shirai [Sh87]; Suk & Bhandarkar [SuBh92].

4.3 Ausblick

In jüngster Zeit erweitern Verfahren der Fuzzy-Theorie und der Neuroinformatik die grundlegenden Methoden der klassischen Bildverarbeitung. Inspiriert durch die Leistungsfähigkeit biologischer Seh- und Erkennungsprozesse werden Algorithmen und Verfahren entwickelt, welche komplexe Systeme nach biologischem Vorbild aufbauen. Die Verfahren an sich sind nicht neu und haben sich bereits in vielen Applikationen bewährt. Durch die Nachahmung der biologischen Informationsverarbeitung kommen sie innerhalb der Bildverarbeitung sowohl in der Vorverarbeitung als auch in der abschließenden Interpretation des Bildinhaltes zum Einsatz. Als Eingabedaten können neben den Tiefen- und/oder Grauwertdaten auch vorverarbeitete und segmentierte Bilder genutzt werden.

Ein weiterer Vorteil neuronaler Netze gegenüber konventionellen Klassifikationsverfahren liegt neben ihrer einfachen Adaptierbarkeit und Generalisierungsfähigkeit darin, daß sie aufgrund ihrer Einzelelemente mit nichtlinearen Übertragungsfunktionen auch nichtlineare Verknüpfungen in den Daten aufspüren können [EP98]. Dies ist gerade dann von Vorteil, wenn viele Daten parallel ausgewertet werden müssen.

5 Beschreibung der Systemkomponenten

Basis dieser Arbeit ist im wesentlichen ein neuartiges Sensorelement zur gleichzeitigen Aufnahme von Entfernungs- und Grauwertdaten, das Photonic Mixer Device (PMD). Auf Basis dieses Elementes wird ein Kamerakzept zur Aufnahme von vereinzelt Entfernungsdaten mit einem vollständigen Grauwertbild vorgestellt. Dieses Konzept dient im folgenden als Grundlage für die weiteren Untersuchungen innerhalb dieser Arbeit.

5.1 Photonic Mixer Device

Die erste veröffentlichte PMD-Struktur, an dem das Funktionsprinzip durch Simulation nachgewiesen wurde, war das sogenannte CCD-PMD [Scw96a]. Im Rahmen dieser Arbeit soll ein weiteres Mitglied der PMD-Familie auf Basis der CMOS-Technologie [BuScRi98] vorgestellt werden. Zuvor jedoch erfolgt für ein besseres Verständnis der Technologie eine kurze Einführung in die Grundlagen der Laufzeitentfernungsmessung.

5.1.1 Laufzeitentfernungsmessung

Der Abstand oder die Tiefe l wird bei den Laufzeitverfahren über die Echolaufzeit τ eines von einem Sender gerichtet abgestrahlten, vom Meßobjekt reflektierten und vom Empfänger empfangenen Strahls, bestimmt. Neben der direkten Messung der Laufzeit eines Pulses (Pulslaufzeitverfahren) stehen weitere Verfahren zur Verfügung. Dabei ist zur Bestimmung der Laufzeit τ sowohl auf der Sender- als auch auf der Empfängerseite eine entsprechende Modulation notwendig, d.h. aus dem vom Sender ausgestrahlten modulierten Licht wird durch eine entsprechende Korrelation des Empfangssignals mit einem geeigneten Referenzsignal die Laufzeit τ bestimmt. Als mögliche Modulationsverfahren stehen auf Basis der Phasenlaufzeitmessung beispielsweise die Continuous Wave (CW)-Modulation (siehe [OI97]) und die Pseudo Noise (PN)-Modulation (siehe [KI93]) zur Verfügung. Innerhalb der CW-Modulation sind mindestens drei Meßwerte mit unterschiedlichen Referenzphasen zur Entfernungsbestimmung notwendig. Auf der Sender- und Empfängerseite ist eine analoge Modulation bzw. Mischung erforderlich. Im Gegensatz dazu wird bei der PN-Modulation auf der Sender- und Referenzseite nur mit binären (bzw. ternären) Signalen gearbeitet. Dem vereinfachten Schaltungsaufbau wird eine durch die Modulation verursachte feinere Abtastung des Korrelationssignals entgegengesetzt, das bedeutet deutlich mehr Messungen zur Bestimmung der Objektentfernung als bei der CW-Modulation. Durch diesen erhöhten Meßaufwand besteht dafür die Möglichkeit, Mehrfachziele in

gewissen Grenzen separieren zu können. Beiden Verfahren ist gemeinsam, daß sie nur einen begrenzten Eindeutigkeitsbereich besitzen. Dieser ist abhängig von der Modulationsfrequenz und bei der PN-Modulation zusätzlich von der Sequenzlänge. Eine Erhöhung der Modulationsfrequenz verbessert in gleichem Maße die Tiefenauflösung und reduziert den Eindeutigkeitsbereich.

5.1.2 CMOS-PMD

Das PMD-Element ist ein Halbleiterelement, welches alle notwendigen Komponenten zur Aufnahme und Verarbeitung eines Entfernungswertes in einem einzigen Element vereint:

- Ø Optische Detektion
- Ø Demodulation innerhalb des quasi optischen Bereiches
- Ø Gegentaktmischung
- Ø Integration
- Ø Signalverarbeitung

Mit Hilfe dieser Eigenschaften kann eine direkte Mischung eines amplitudenmodulierten optischen Eingangssignals mit einem elektrischen Gegentakt-Bezugssignal durchgeführt werden [ScHe97]. Da die Mischung bereits während der Detektion durchgeführt wird, entfällt der sonst notwendige Hochfrequenzempfänger. Das aufintegrierte Ergebnis der Mischung (Korrelationsergebnis) erhält man durch Abtasten der beiden Ausgangssignale am Ende der Integrationsperiode. Die Differenz der beiden Ausgangssignale ist dabei frei von Gleichanteilen und stellt ein entfernungsproportionales Signal dar. Aus der Summe der beiden Ausgangssignale kann ein der herkömmlichen Photodioden äquivalentes Helligkeitssignal direkt entnommen werden.

Das Arbeitsprinzip des PMD-Elementes kann unter Zuhilfenahme der Abb. 5.1 wie folgt beschrieben werden:

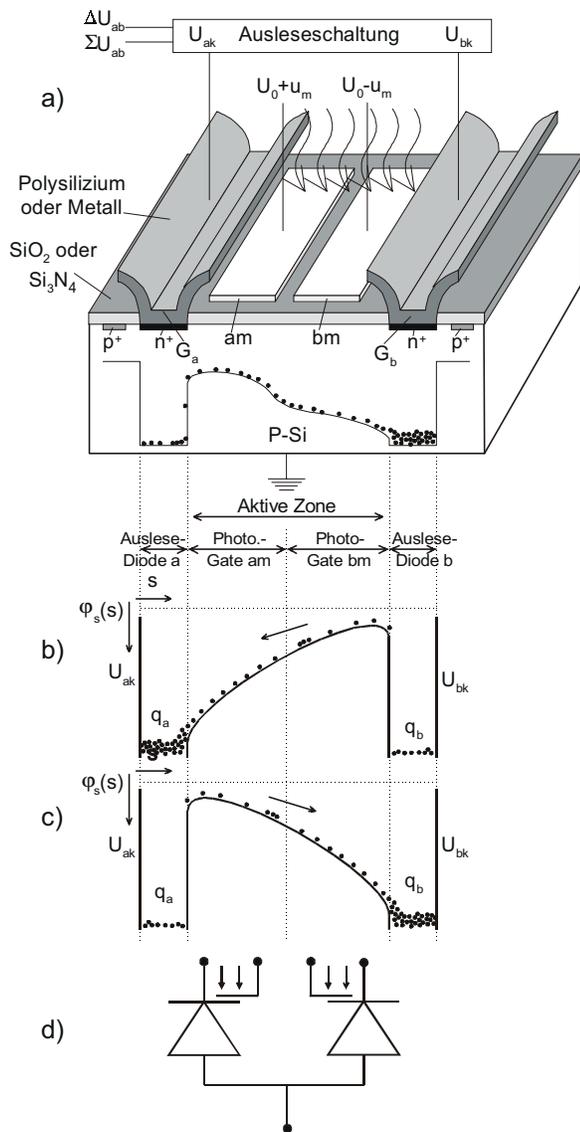


Abb. 5.1: a) Querschnitt- b,c) Potentialverlauf- d) Symbol eines PMD-Elementes [ScHe97]

In Abb. 5.1a) ist der Querschnitt eines PMD-Elementes dargestellt. Die weißen Streifen im mittleren Bereich des Elementes stellen die beiden transparenten Photogates bzw. Modulationselektroden am und bm dar und bilden den lichtempfindlichen Pixelbereich. In der Raumladungszone unterhalb dieser Schicht entstehen proportional zur einfallenden Lichtleistung Photoelektronen, die nach rechts und links zu den vorgespannten n^+ -Zonen wandern. Diese Elektroden (G_a und G_b) stellen die Kathoden (Abb. 5.1d) der beiden angrenzenden gesperrten Auslesedioden dar. Das P-Substrat bildet die gemeinsame Anode der Dioden.

Wird an den Modulationselektroden am und bm eine offsetbehaftete Gegentakt-Modulationsspannung $u_o \pm u_m$ angelegt, so entsteht in der darunter liegenden Raumladungszone ein sich in Abhängigkeit der Modulationsspannung u_m änderndes Potentialgefälle, das die Photoelektronen mehr nach rechts oder links driften läßt (Abb. 5.1b,c). Der Mischeffekt besteht nun darin, daß die Photoelektronen des modulierte

einfallenden Lichtes im Takt der Modulationsspannung wie mit einer „Ladungsschaukel“ auf die beiden Auslesedioden verteilt werden. Die Auswertung der vom PMD-Pixel aufintegrierten Ladungspakete erfolgt mit einer angepaßten Ausleseschaltung im Stromauslesemodus, die das Korrelationsergebnis in Form einer Differenzspannung ΔU_{ab} und die Pixelhelligkeit bzw. den Grauwert als eine Summenspannung ΣU_{ab} liefert.

Das elektrische Symbol dieses Photomischers in Abb. 5.1d) entspricht vereinfacht der Struktur aus Abb. 5.1a) mit einem optischen und einem elektrischen Gegentakteingang und dem Gegentaktmischerausgang.

5.1.3 Optimierte PMD-Pixel-Strukturen

Aufbauend auf dem im vorherigen Abschnitt beschriebenen grundlegenden PMD-Element soll hier eine Weiterentwicklung vorgestellt werden. Das in Abb. 5.1a) dargestellte Element ist aufgrund der großen lichtempfindlichen Fläche und der damit verbundenen großen Kapazität auf eine Arbeitsfrequenz im unteren MHz Bereich beschränkt. Zur Steigerung der Modulationsfrequenz in den für die Entfernungsmessung empfehlenswerten Frequenzbereich von 10-100 MHz und höher wird daher die Fingerstruktur vorgeschlagen [ScHe98], wie sie in Abb. 5.2 zu sehen ist. In dieser Anordnung werden viele schmale PMD-Streifen parallel betrieben. Die kurze Gatelänge von wenigen μm bewirkt eine dementsprechend kurze Laufzeit der Ladungsträger und damit eine Erhöhung der Grenzfrequenz. Durch eine Parallelschaltung der PMD-Streifen kann die effektive Pixelgröße ‚beliebig‘ vergrößert werden, ohne dabei die max. Arbeitsfrequenz zu beeinträchtigen.

In Abb. 5.2 stellen die hellen Streifen die transparenten Photogates und die dunklen Streifen die Kathodenstreifen der Auslesedioden dar. Letztere haben in dieser Anordnung eine doppelte Funktion: Sie sammeln gleichzeitig die Photoelektronen von beiden angrenzenden Photogates und erhöhen somit den optischen Füllfaktor. Der neu hinzugekommene mittlere helle Streifen bewirkt eine bessere Linearisie-

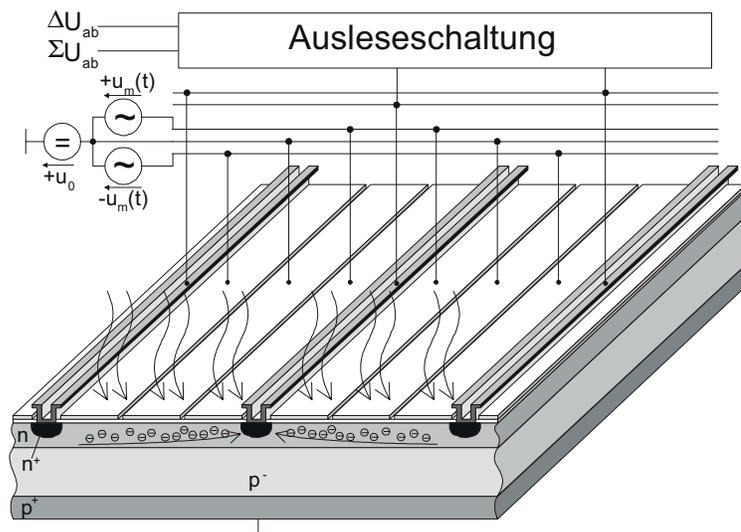


Abb. 5.2: PMD-Element in Fingerstrukturtechnik [XuSc98]

rung des Potentialverlaufes der Ladungsverteilung und kann ggf. entfallen. Für eine Entfernungsmessung auf Basis der CW-Modulation sind, wie bereits erwähnt, mindestens drei Messungen mit unterschiedlichen Phasenwerten notwendig. Diese können wahlweise durch zwei aufeinanderfolgende Messungen mit einem einzigen PMD-Element oder durch gleichzeitige Verwendung von zwei benachbarten Elementen auf einmal erfolgen. In diesem Falle wird die Kombination der beiden Elemente als IQ-PMD (entsprechend der Sinusmodulation Inphase/Quadraturphase) bzw. 2-Q-IQ-PMD (für 2-Quadranten IQ-PMD) bezeichnet. Die hieraus gewonnenen 4 Meßwerte ermöglichen, wie der Name schon andeutet, die direkte Bestimmung der Phase bzw. der Verzögerung aus einer Messung.

In Abb. 5.3 ist eine weitere Variante des soeben erwähnten Prinzips dargestellt, das 4-Q (Vier-Quadranten)-IQ-PMD. Hierbei werden 4 einzelne Fingerstruktur-PMD-Elemente durch Drehung und Spiegelung so miteinander kombiniert, daß elektromagnetische Interferenzen und Störungen möglichst gering ausfallen. Auch hier stellen die hellen Streifen die transparenten Photogates und die dunklen Streifen die Kathodenstreifen der Auslese-dioden dar.

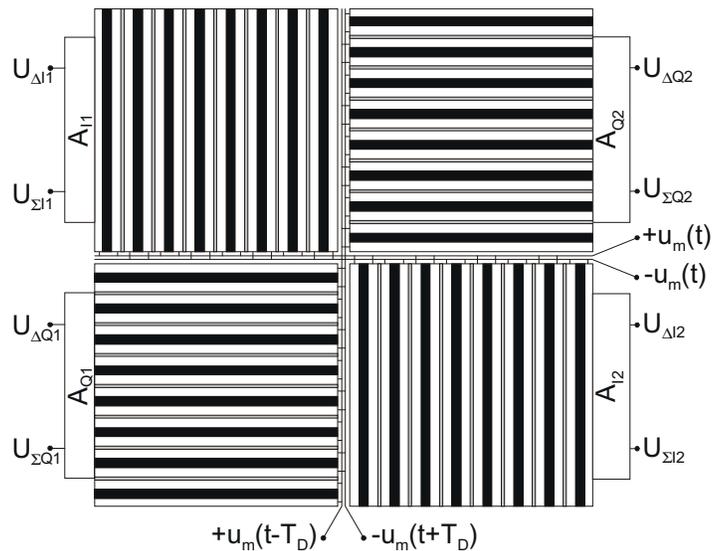


Abb. 5.3: 4-Q-IQ-PMD [XuSc98]

Diese Anordnung von vier PMD-Elementen ermöglicht verschiedene Anwendungsmöglichkeiten. Für einige Applikationen bietet dieses orthogonale IQ-PMD die Option einer getrennten Auswertung der vier PMD-Quadrate an. Auf diese Weise erhält man benachbarte IQ-Kombinationen, die vier Distanzwerte der vier jeweils zugehörigen Flächenschwerpunkte und zusätzlich den Mittelwert dieser vier Resultate für das gesamte Pixel.

Durch um 90° phasenverschobenes Ansteuern von zwei benachbarten Elementen wird die Meßzeit halbiert, wobei hierfür eine homogene Lichtverteilung über die gemeinsam genutzten Pixel notwendig ist. Liegt diese Bedingung nicht vor, so kann eine Mikrolinse (angebracht über dem 4-Q-IQ-PMD) neben der Erhöhung des optischen Füllfaktors eine scharfe Abbildung von Hell/Dunkel-Übergängen innerhalb des Pixel durch Defokussierung vermeiden. Allerdings geht hiermit ein weiterer Vorteil dieser Anordnung verloren. Mit Hilfe von drei getrennt abbildenden Elementen kann zusätzlich zur Entfernung die Oberflächenorientierung aus der Verhältnisbildung der Elemente bestimmt werden. Eine Erhöhung der Genauigkeit kann sowohl über eine räumliche (vier mögliche Dreierkombinationen) als auch über einen zeitlichen Mittelwert erfolgen. Mit diesem 4-Q-IQ-PMD-Pixel steht somit erstmalig ein Sensorelement zur gleichzeitigen Aufnahme von Intensität, Objektentfernung und Oberflächenorientierung zur Verfügung.

5.2 2D/3D-Kamerakzept

Die Zielsetzung der PMD-Architektur ist u.a. die Matrixkamera zur flächendeckenden Aufnahme von Helligkeits- und Entfernungswerten, die durch das PMD-Element erstmals möglich wird. Ein Vorstadium der flächendeckenden Anordnung ist die vereinzelte oder zeilenmäßige Plazierung von PMD-Elementen, verteilt auf den zur Verfügung stehenden Matrixbereich.

Aufbauend auf diesem PMD-Konzept wird im folgenden das dieser Arbeit zugrundeliegende 2D/3D-Kamerakzept zur Aufnahme eines vollständigen Grauwertbildes mit vereinzelten Entfernungswerten vorgestellt. Zur Realisierung bietet sich dabei sowohl die getrennte Nutzung der PMD- und CMOS-Elemente (Strahlteiler-Kamerasystem) als auch die Kombination von beiden an (MixedPixel-Kamerasystem). Alle nachfolgenden Untersuchungen innerhalb dieser Arbeit beziehen sich zwar vordergründig, aber nicht ausschließlich auf dieses Konzept.

5.2.1 Strahlteiler-Kamerasystem

Der direkteste Weg zur Aufnahme eines Grauwertbildes mit vereinzelten Tiefenwerten ist die Kombination von PMD-Sensoren mit einer CMOS- oder CCD- Kamera. Bei den PMD-Sensoren kann es sich dabei sowohl um einzeln angebrachte Elemente, um eine Zeilenanordnung als auch um ein Array mit vereinzelten PMD-Elementen handeln. Werden diese PMD-Elemente dabei unabhängig zum CMOS- oder CCD-Array montiert, sind aufwendige Zuordnungsstrategien erforderlich, um die Pixel beider Elemente in Verbindung zu setzen (siehe Weber [We98]). Mit geringerem Aufwand kann dies erfolgen, wenn beide Elemente über einen Strahlteiler kombiniert werden (siehe Abb. 5.4).

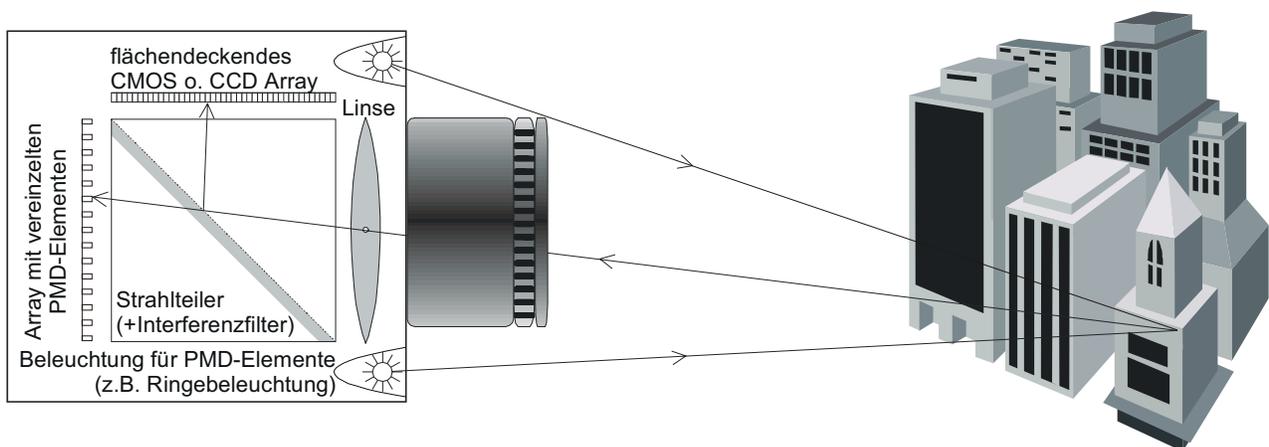


Abb. 5.4: Beispiel einer möglichen Sensoranordnung am Beispiel eines Strahlteiler-Kamerasystems

Neben den variablen Anordnungsmöglichkeiten der PMD-Elemente und der Unabhängigkeit von unterschiedlichen Pixelgrößen bietet dieses Strahlteiler-Konzept die Möglichkeit der spektralen Trennung von PMD- und CMOS- bzw. CCD-Beleuchtung. Auf diesem Wege kann die Beleuchtung der Szene optimal auf jeden Sensortyp angepaßt werden.

5.2.2 MixedPixel-Kamerasystem

Die genutzte CMOS-Technologie ermöglicht es, neben den PMD-Elementen weitere Elemente auf dem Chip zu integrieren. Zusammen mit der möglichen Einzel- oder Zeilenanordnung von PMD-Elementen können mit Hilfe dieser Technologie beispielsweise verfügbare Photosensoren (z.B. Aktive Pixel Sensor-Elemente APS) in den freien Flächen zwischen den PMD-Elementen platziert werden (siehe Abb. 5.5). Die unterschiedlich großen lichtempfindlichen Bereiche der PMD- und APS-Elemente und die zu jedem Element gehörige Ausleseelektronik können bei Bedarf über Mikrolinsen, angeordnet über jedes einzelne Element, kompensiert werden. Gemeinsam mit den ebenfalls für Grauwertaufnahmen geeigneten PMD-Pixel kann auf dieser Basis eine Standardvideokamera geschaffen werden, welche gleichzeitig zu den flächendeckenden Intensitätswerten vereinzelte Entfernungswerte bzw. bei Verwendung von 4-Q-IQ-PMD-Elementen Oberflächenorientierungen ausgibt.

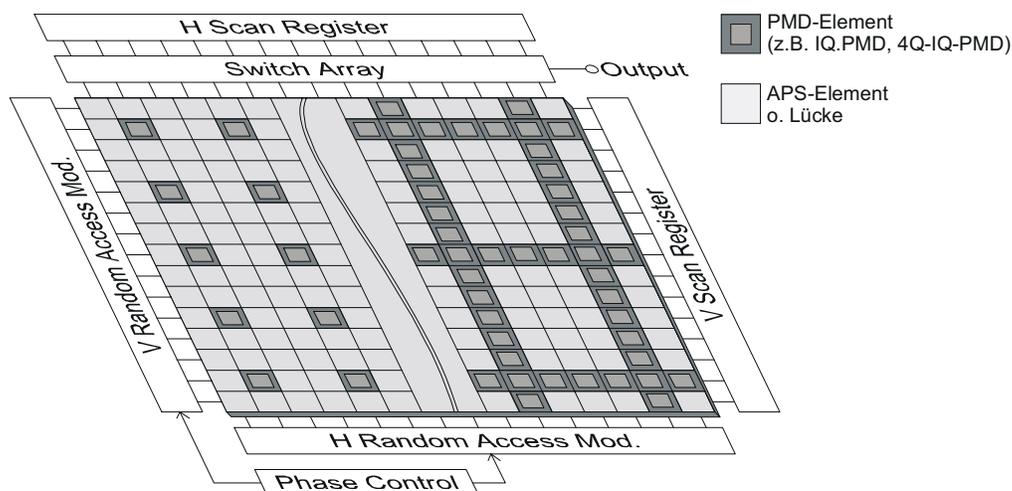


Abb. 5.5: APS-Matrixstruktur mit vereinzelt platzierten PMD-Elementen

Im Vergleich der CMOS-Technologie zur CCD-Technologie müssen allerdings einige ‚Nachteile‘ in Kauf genommen werden. Die Tatsache, daß ein kostengünstiger ‚Standard‘ CMOS-Prozeß genutzt wird, bedingt, daß die unterschiedliche Empfindlichkeit der einzelnen Photoelemente durch aufwendige Analogtechnik korrigiert werden muß. Zudem muß

im Vergleich zur CCD-Technologie, welche die Ladung des Lichtstroms fast verlustfrei transportieren kann, eine Schaltung zur Auslesung aller Elemente (ähnlich der von Speichern) integriert werden. Im Gegensatz zur CCD-Technologie kann dafür im CMOS-Prozeß weitere Logik am Rande der Chipfläche platziert werden. Durch diese Integration kann die notwendige Ausleseschaltung der Entfernungs- und Grauwertdaten der CMOS-Photodioden und der PMD-Elemente incl. einer Korrektur der unterschiedlich empfindlichen Elemente am Chiprand realisiert werden. Weiterhin ist innerhalb dieses Bereiches der Einbau weiterer Steuerungswerke, der Modulations- und Demodulationstechnik, von AD-Wandlern und anderen Komponenten möglich. Kundenspezifische Schaltungswünsche (z.B. auf VHDL-Basis) sind ebenso denkbar. So können die Kosten für die notwendige externe Elektronik, Aufbau und Gehäuse der Chips drastisch reduziert werden. Damit ergeben sich für den Hersteller z.B. digitaler Kameras wesentliche Kosten- und Systemvorteile.

5.3 Anordnungsmöglichkeiten der vereinzelt Tiefenwerte

Entsprechend der Meßaufgabe kommen verschiedene Gitteranordnungen zur Anordnung der vereinzelt PMD-Elemente in Frage. Aus der Vielzahl der möglichen Anordnungen werden im folgenden einige wenige Strukturen und deren Vor- und Nachteile vorgestellt. Einige Anwendungsbeispiele sollen dem Leser abschließend eine bessere Vorstellung des Einsatzbereiches solcher Strukturen geben.

5.3.1 Gitteranordnung

Als Beispiele für die Gitteranordnung von vereinzelt Tiefenwerten sollen hier die Anordnungen aus Abb. 5.6 vorgestellt werden. Das Punkt-, Zeilen- und Kreuzgitter stellen innerhalb dieser Beispiele die einfachsten Formen dar und können ungeachtet dessen vielfältig eingesetzt werden. Ähnliches gilt auch für das Karogitter. Durch den regelmäßigen Abstand der Linien stellt es an allen Stellen die gleiche laterale Auflösung zur Verfügung. Im Gegensatz hierzu besitzen das Kreis-, Stern- und Strahlengitter eine unterschiedliche laterale Auflösung. Als Anwendungsgebiete solcher Gitter kommen Fälle in Frage, in denen eine höhere Auflösung in bestimmten Regionen gefordert ist oder die Analyse von Objekten mit Hilfe dieser Gitter vereinfacht wird.

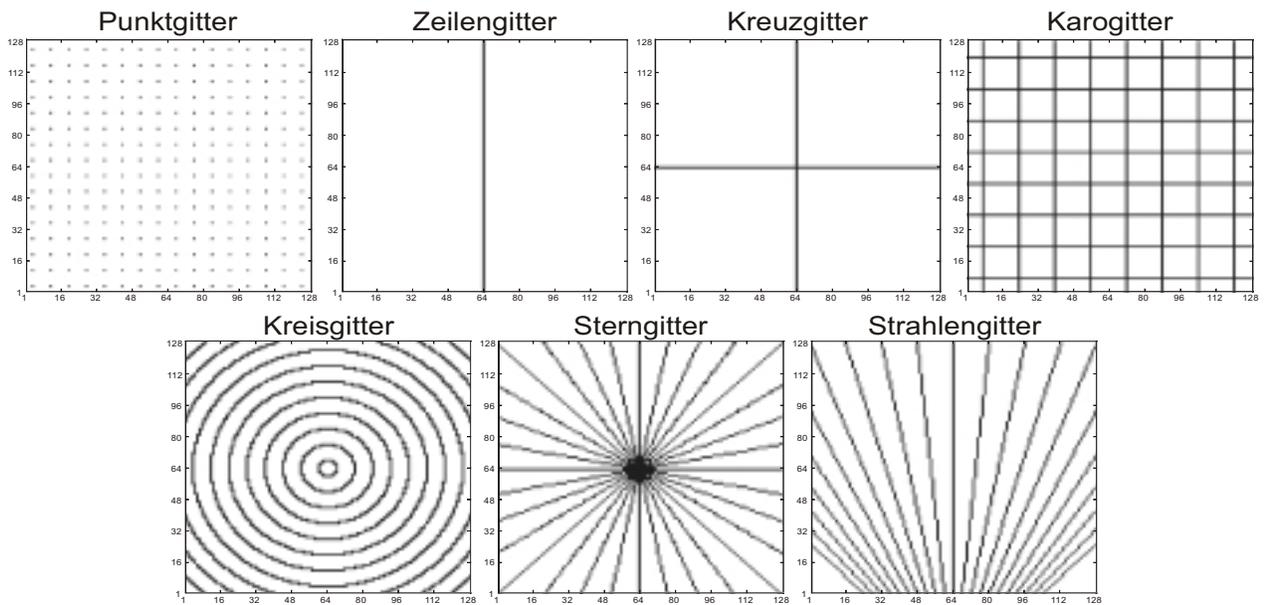


Abb. 5.6: Mögliche Gitteranordnungen für die Tiefenwerte

5.3.2 Orientierungsbestimmung mit vereinzelt Tiefenwerten

Das Punktgitter stellt die Gitteranordnung mit der geringsten Anzahl benötigter Elemente dar. Oberflächenorientierungen können folglich nur bedingt bestimmt werden. Sind dennoch Oberflächenorientierungen gefordert (Lagebestimmung von Objekten, Vereinfachung der Interpretation und Klassifikation), so können diese beispielsweise durch Nutzung der 4-Q-IQ-PMD-Pixel (siehe Kap. 5.1.3) gewonnen werden.

Linienstrukturen besitzen gegenüber dem Punktgitter nicht nur den Vorteil, daß sie deutlich mehr Elemente besitzen, sondern auch, daß aus der Aneinanderreihung der einzelnen Elemente direkt eine Komponente der Oberflächenorientierung bestimmt werden kann. Reicht diese nicht aus, so kann die zweite Komponente durch eine Doppelreihe erzeugt werden. Dieser geringe zusätzliche Aufwand bringt den Vorteil der vollständigen Orientierungsbestimmung mit sich.

5.3.3 Füllfaktor

Unter der Definition des Tiefenfüllfaktors (FF) wird in dieser Arbeit das Verhältnis der Anzahl der Tiefenwerte zur Anzahl der Grauwerte verstanden:

$$FF = \frac{\text{Anzahl der Tiefenwertelemente}}{\text{Anzahl der Grauwertelemente}} [\%] \quad (5.1)$$

Der Füllfaktor ist dabei von der Gitteranordnung abhängig. Aufgrund der geringen Anzahl der PMD-Elemente bei dem Punktgitter liegt der Füllfaktor im unteren Prozentbereich. Linienanordnungen hingegen besitzen einen größeren Füllfaktor.

In Abhängigkeit der Verteilung der PMD-Elemente kann eine zusätzliche Angabe zum Füllfaktor über die Verteilung der Elemente Auskunft geben. Beim Punktgitter und Karogitter gibt z.B. das Verhältnis zwischen horizontalem und vertikalem Abstand eine genauere Auskunft über die Verteilung der Elemente. Dementsprechend können beim Kreisgitter die Anzahl der Ringe, beim Sterngitter der Winkelabstand und beim Strahlengitter der Öffnungswinkel zusätzliche Informationen enthalten.

5.3.4 Anwendungsbeispiele

Das **Punktgitter** bedingt durch seine geringe Anzahl von Entfernungselementen eine grobe laterale Tiefenauflösung des Bildes. Dadurch kann eine sehr schnelle Bildverarbeitung realisiert werden. Diese Anordnung ist sinnvoll, wenn Objekte mit einer kleinen Anzahl von unterschiedlichen Flächen untersucht werden sollen, wie z.B. bei der Inspektion von einfachen Objekten oder Werkstücken (siehe Abb. 5.7 links). Ist die Geometrie und die Lage der zu vermessenden Objekte a priori bekannt, kann durch eine vorangegangene Segmentierung des Grauwertbildes eine Höhenkontrolle der Objekte erfolgen, bzw. kann an den fragwürdigen bzw. wichtigen Objekten die Tiefeninformation zur weiteren Verarbeitung und Klassifizierung genutzt werden.

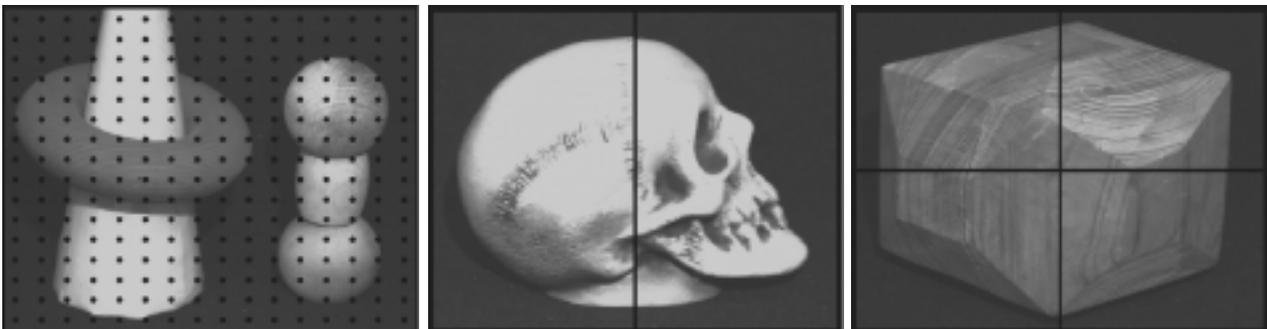


Abb. 5.7: Anwendungsbeispiele für das Punkt-, Zeilen- und Kreuzgitter [WwwMa99]

Das **Zeilen-** und **Kreuzgitter** sind entsprechend dem Punktgitter ebenfalls sehr einfache Anordnungsmöglichkeiten der PMD-Elemente. Sowohl mit als auch ohne flächendeckendes Grauwertbild können mit diesen beiden Gitterstrukturen bereits eine Vielzahl von 3D-BV Ansätzen gelöst werden (z.B. Konturüberprüfung, Vermessung von Objekten, Größenbestimmung, Lagebestimmung). Neben dem einfachen Aufbau und der einfachen Herstellung des Sensors wirken sich Oberflächenreflexionen (hervorgerufen durch die Beleuchtung der Szene für die Tiefenmessung) nicht so störend aus wie bei einer flächendeckenden Anordnung und Beleuchtung der PMD-Elemente. Aufgrund der Einzelzeile

beeinflussen sich die benachbarten Elemente nur geringfügig. Entsprechend dem einfachen Layoutprozeß für die PMD-Elemente fällt auch die Optik zur Erzeugung der sender- und empfängerseitigen Strahlengänge sehr einfach aus.

Das **Karogitter** ist aufgrund seiner rechtwinklig angeordneten Linien und seiner ‚relativ‘ großen Anzahl von Tiefenwerten für alle Aufgabentypen gleichermaßen gut geeignet. Die rechtwinklige Linienstruktur ermöglicht die Rekonstruktion der Oberflächenkontur und der Lage von Objekten. Die Auswertung von dynamischen Szenen, in denen sich die Kamera und/oder die Objekte bewegen, ist ebenfalls denkbar.

Das **Kreis-** und das **Sterngitter** sind zueinander komplementär und können folglich für die gleichen Aufgaben eingesetzt werden. Als ein einfaches Anwendungsbeispiel sei hier die Inspektion von runden bzw. kugelförmigen Objekten dargestellt (siehe Abb. 5.8 links). Sowohl mit dem Kreis als auch mit dem Sterngitter können z.B. rotationssymmetrische Objekte ausreichend genau vermessen werden, sofern sie sich in der Nähe des Zentrums befinden. Neben diesem einfachen Anwendungsbereich besteht mit diesen Gitteranordnungen auch die Möglichkeit, eine räumliche Umgebungskontrolle durchzuführen (siehe Abb. 5.8 Mitte). Die Anhäufung der Linien bzw. Kreise im Bildmittelpunkt ermöglicht eine hohe Auflösung im interessanten Meßbereich. Mit zunehmendem Abstand vom Mittelpunkt reduziert sich entsprechend dem menschlichen Sehen die laterale Auflösung.

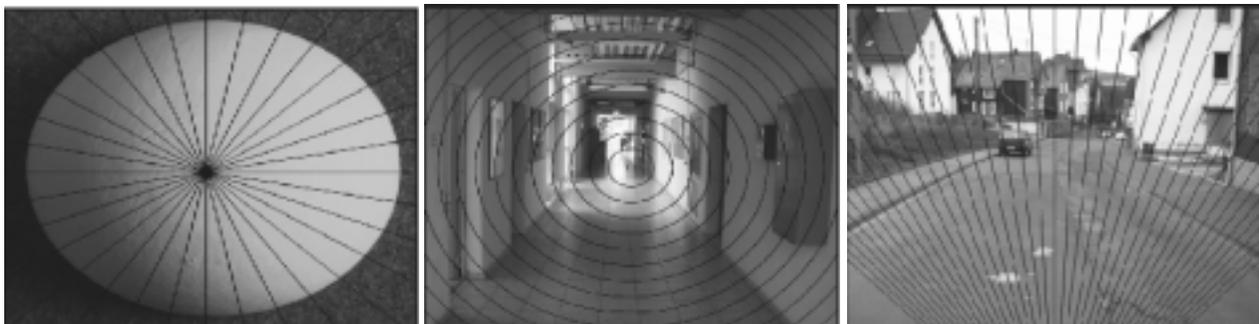


Abb. 5.8: Anwendungsbeispiele für das Stern- Kreis- und Strahlengitter

Für die Hinderniserkennung und die Straßenverfolgung von autonomen Fahrzeugen bzw. von Kraftfahrzeugen ist es erforderlich, über eine ebene Fläche (z.B. Straße) in Fahrtrichtung voraus zu blicken. Dies wird mit dem **Strahlengitter** ermöglicht (siehe Abb. 5.8 rechts). Abhängig vom Öffnungswinkel des Strahlengitters kann in Fahrtrichtung vorausschauend eine feine laterale Tiefenauflösung von Hindernissen in der Nahzone und eine

grobe Auflösung der Straßenkontur und von Hindernissen in der Fernzone erfolgen. Der Öffnungswinkel ist dabei abhängig vom Anwendungsgebiet. Bei schneller Fahrt oder vorrangiger Hinderniserkennung in Fahrtrichtung empfiehlt sich ein kleiner Öffnungswinkel. Fährt man dagegen nur mit geringer Geschwindigkeit oder sind auch aufgrund von evtl. Ausweichmanövern Hindernisse abseits der Fahrtrichtung von Interesse, sollte ein großer Öffnungswinkel gewählt werden.

5.4 Beleuchtung

Als ein grundlegendes Element der Bildverarbeitung ist die Beleuchtung zu nennen. Eine dem Objekt angepaßte Beleuchtung ermöglicht in vielen Fällen erst die Lösung von Bildverarbeitungsproblemen. Je nach den Eigenschaften eines Objektes (Material, Oberflächenstruktur, Farbe, Transparenz, Absorption, Streu- und Reflexionsverhalten usw.) sollte eine günstige Verteilung der Lichtintensität auf einer Objektoberfläche durch die Beleuchtung erzielt werden, so daß sich ein maximaler Kontrast (Helligkeitsmessung) bzw. max. Empfangshelligkeit (Tiefenmessung) ergibt. Aufgrund dieser unterschiedlichen Bedingungen wird im folgenden zwischen dem Fremdlicht (FL) und der Eigenbeleuchtung (EB) unterschieden. Als Fremdlicht wird dabei die indirekte Beleuchtung der Szene durch die Umgebungshelligkeit oder durch nicht modulierte Quellen verstanden. Die Eigenbeleuchtung ist eine aktive modulierte Beleuchtung, welche für die Entfernungsmessung notwendig ist.

5.4.1 Eigenbeleuchtung

Neben den PMD-Elementen ist zur laufzeitbasierten Entfernungsbestimmung die aktive Eigenbeleuchtung ein wichtiges Element. Die Entfernung bestimmt sich, wie bereits besprochen, aus der Impulslaufzeit eines ausgesendeten Impulses bzw. der Phasenlaufzeit eines modulierten ausgesendeten Signals. Aus nicht beleuchteten Oberflächenpunkten kann somit keine Entfernung bestimmt werden. Ähnliches gilt für zu schwach beleuchtete, zu weit entfernte bzw. schlecht reflektierende Objekte.

Ähnlich einem CCD-Element besitzt auch das PMD-Element eine wellenlängenabhängige Empfindlichkeit. Für eine optimale Beleuchtung der Szene sollte der Wellenlängenbereich mit der größten Empfindlichkeit des PMD-Elementes ausgewählt werden. Weitere Anforderungen an die Beleuchtung ist ihre Modulierbarkeit und ihr Öffnungswinkel. Wie bereits angesprochen, werden hohe Frequenzen zur genauen Vermessung der Szene gefordert.

Die optimale Beleuchtungsquelle wäre somit eine Punktquelle, von der alle Strahlen zum gleichen Zeitpunkt ausgesendet werden. Dies wiederum steht im Widerspruch zur Sendeleistung. Entsprechend dem geforderten Meßbereich werden unterschiedlich starke Beleuchtungsquellen benötigt. Da die Leuchtstärke eines einzelnen Elementes beschränkt ist, muß diese ggf. durch eine Flächenanordnung wieder vergrößert werden. Mit all diesen Anforderungen an die Beleuchtungsquelle sind derzeit Hochgeschwindigkeits-LEDs, Laserdioden und HF-modulierte Plasma-Entladungsröhren als Beleuchtungsquellen denkbar. Bezüglich der Laserdioden müssen die Laserschutzklassen berücksichtigt werden.

Eine Reduzierung der Leuchtstärke für die Tiefenmessung kann erreicht werden, wenn anstatt der flächendeckenden Beleuchtung der gesamten Szene nur der Meßbereich der einzelnen PMD-Elemente beleuchtet wird. Hierbei muß sichergestellt werden, daß die Sendestrahlen über den gesamten Meßbereich mit den Empfangsstrahlen übereinstimmen. Als optimale Optik zur Realisierung einer solchen Verteilung bietet sich bei Nutzung von kohärentem Licht das ‚diffractive pattern generator‘ [ESC98] an. Mit Hilfe dieses optischen Elements wird ein einfallender Lichtstrahl fast verlustfrei auf die gewünschte Anordnung verteilt. Die projizierten Linien weisen jedoch aufgrund des Specklerauschens keine gleichmäßige Helligkeitsverteilung auf.

Neben der Reduzierung der Lichtleistung bietet die diffraktive Optik den Vorteil, daß Reflexionen von benachbarten beleuchteten Regionen reduziert werden. Mehrfachreflexionen treten durch die vereinzelt oder zeilenmäßige Plazierung von PMD-Elementen und entsprechender Beleuchtung nur noch geringfügig auf.

5.4.2 Fremdlicht und Eigenbeleuchtung

Neben den PMD-Elementen benötigen gegebenenfalls auch die Photosensoren eine Beleuchtung. Reicht die Hintergrundbeleuchtung (Fremdlicht) für diese Elemente nicht aus bzw. ist sie ungeeignet, bietet sich die flächendeckende Beleuchtung der PMD-Elemente und der Photosensoren mit einer Lichtquelle an. Für einen optimalen Betrieb beider Sensoren empfiehlt sich jedoch die getrennte Beleuchtung, wobei eine gegenseitige Beeinflussung vermieden werden muß, sofern sie nicht durch Soft- und Hardware korrigiert werden kann. Diese kann entweder durch Ausnutzung der sensorspezifischen Empfindlichkeiten der verschiedenen Sensoren, durch spektrale Trennung direkt an den lichtempfindlichen Sensoren, am Strahlteiler oder durch externe Filter bewerkstelligt werden.

5.5 Visualisierung

Für die Fälle, in denen keine Möglichkeit der 3D-Visualisierung (Rot-Grün-Darstellung, Polarisationsbrillen) zur Verfügung steht, kann die Visualisierung von 3D-Daten (d.h. Rückprojektion auf eine ebene Fläche) auf verschiedene Arten erfolgen. Aufwendige Ansätze visualisieren z.B. die Tiefeninformation anhand der Projektion der Tiefenszene mit Hilfe von vorgegebenen Reflexionseigenschaften der Oberflächen (z.B. diffuse oder spiegelnde Oberfläche) zurück auf eine ebene Fläche (Raytracing, Radiosity). Mit Hilfe einer beliebigen Beobachtersposition, der frei positionierbaren Beleuchtungsquellen und der Vorgabe von Oberflächentexturen können verschiedene Problemstellen der 3D-Szene gut untersucht werden. Gute Rekonstruktionsprogramme erzeugen mit dieser Vorgehensweise Bilder, die einer Aufnahme der Tiefenszene mit einer frei positionierbaren Grauwertkamera sehr nahe kommen.

Einfachere und schnellere Ansätze erzeugen die Grafik z.B. anhand der Zentral- oder Parallelprojektion. Über eine Farbkodierung der Höhenwerte (Falschfarben-Höhenbild) oder aus den zu den Tiefenwerten gehörigen Grauwerten wird ein räumlicher Eindruck erzeugt. In Abb. 5.9 sind zur Verdeutlichung zwei Tiefenbilder dargestellt. Beide Bilder wurden mit einem PMD-Element und einer Spiegeleinheit zur XY-Ablenkung aufgenommen.

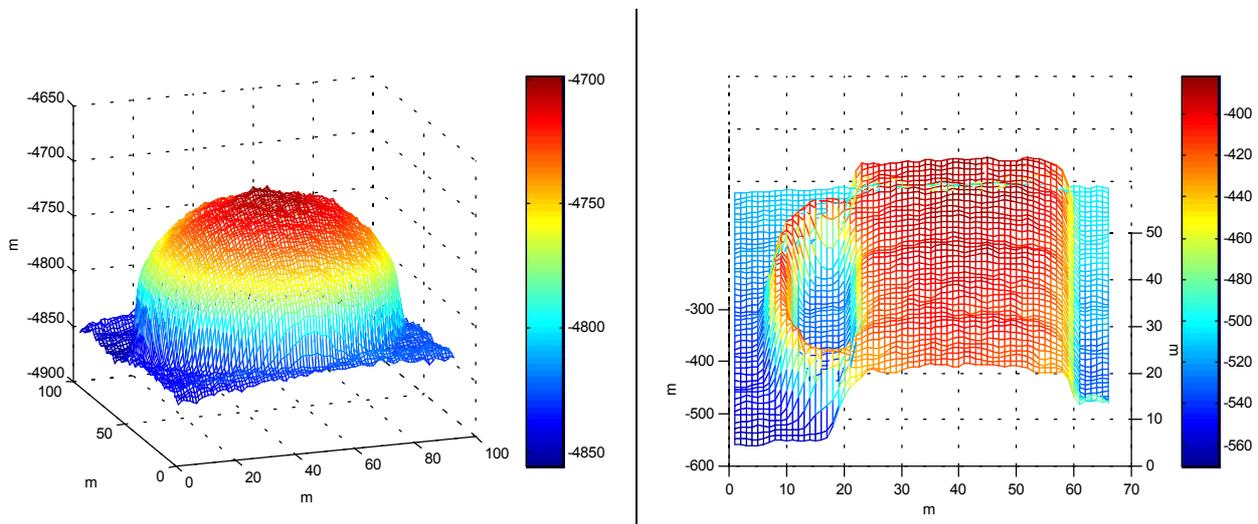


Abb. 5.9: Visualisierung einer Halbkugel und einer Tasse auf Basis der Zentralprojektion mit Falschfarbendarstellung.

Hinsichtlich der vereinzelt Tiefenwerte sind die zuvor beschriebenen Möglichkeiten zur Visualisierung der Tiefendaten nur begrenzt anwendbar (siehe Abb. 5.10 links). Aufgrund der aufwendigen und wenig aussagekräftigen Darstellungsart zeigen nachfolgende

Grafiken die Szene von oben, d.h. aus der Blickrichtung der Kamera. Dies entspricht der Darstellung eines normalen Grauwertbildes, wobei die vereinzelt Tiefenwerte durch schwarze Punkte oder Linien angedeutet werden (siehe Abb. 5.10 rechts).

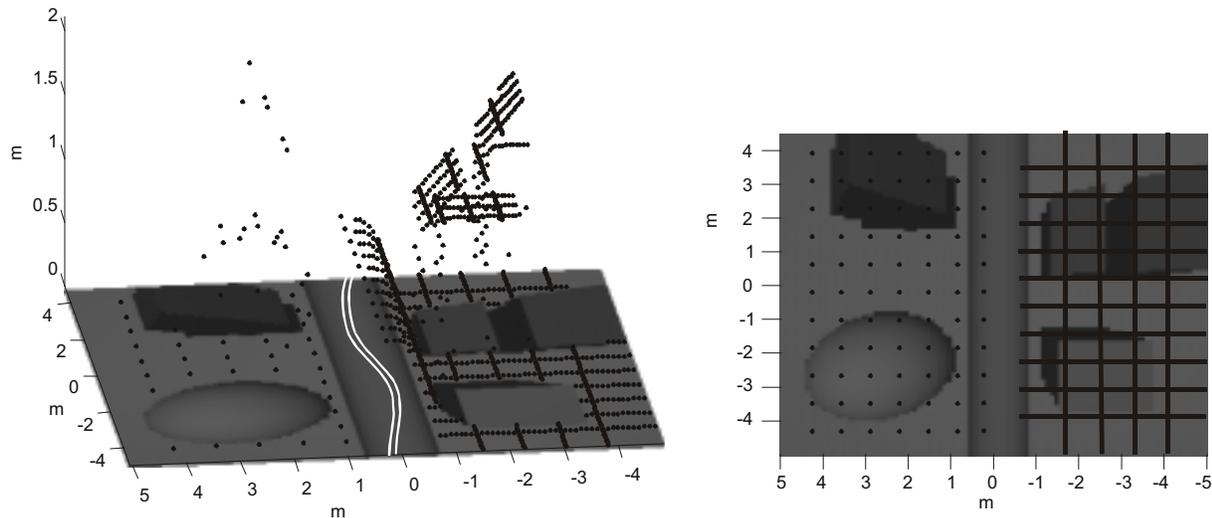


Abb. 5.10: Visualisierung von vereinzelt Tiefendaten mit einem flächendeckenden Grauwertbild

5.6 Zusammenfassung

Mit Hilfe des hier vorgestellten 2D/3D-Kamerakzeptes werden bestehende Einschränkungen derzeit erhältlicher 3D-Meßsysteme umgangen. Die Modulation und Detektion im quasi optischen Bereich vereinfacht die Handhabbarkeit von PMD-Elementen und führt überdies zu einer enormen Stabilität der Phasenmessung. Für die weitere Optimierung, speziell für flächendeckende PMD-Anordnungen mit einer großen Pixelanzahl und hohen Tiefenauflösung bis in den Sub-mm-Bereich, ist in der Zukunft weiterhin viel Arbeitskraft notwendig. Allerdings kommen die stetigen Entwicklungserfolge der CMOS-Kamera und die Entwicklung der CMOS-Technologie hin zu Strukturgrößen bis in den Sub- μm -Bereich hinein der Entwicklung des Photomischdetektors zugute.

Mit Hilfe des PMD-Kzeptes wird nach Schwarte [Scw99b] und Heinol [HeXu99] eine Fülle neuartiger Lösungen erschlossen. Aufgrund der markanten Vorteile

- Ø Kleines Bauvolumen, verglichen mit konventionellen Lösungen
- Ø Reduzierung der Kosten und Fehler der bislang benötigten Empfangstechnik durch hohe Integrationsfähigkeit und die Fertigung in Standard-CMOS Technologie
- Ø Keine HF-Empfangsverstärker notwendig, geringes Rauschen, hohe Dynamik
- Ø Hohe Integrierbarkeit als „Smart Pixel“ – PMD-Array hoher Funktionalität
- Ø Kleine Modulationsspannungen und hohe Modulationsbandbreite (GHz)

- ∅ Verwendbar für beliebige Lichtquellen von UV bis FARED
- ∅ Dynamiksteuerung durch Unterdrückung nichtkorrelierter Hintergrundbeleuchtung und des Gleichanteils der modulierten Beleuchtung durch adaptive Integrationszeiten und adaptive Modulationsamplituden u_m
- ∅ Verminderung von Mehrfachreflexionen durch Beleuchtung von vereinzelteten Punkten
- ∅ Verminderung des Störeinflusses durch Mehrfachreflexionen bei PN-Modulation, wenn Störentfernung größer als die halbe Bitbreite T_B ist

erschließt die PMD-Technologie mit einer kleinen, flexiblen Solid-State-3D-Kamera eine unüberschaubare Vielfalt neuer Anwendungsgebiete z.B.:

- ∅ Steuerung autonomer Transportfahrzeuge
- ∅ Einfache und sichere Geschwindigkeitsmessung von Fahrzeugen (siehe 8.1.2.1)
- ∅ Aktive Sicherheit im Straßenverkehr (siehe 8.1.2.2)
- ∅ Gefahrenraumüberwachung
- ∅ Roboter Vision in der Fertigung, Qualitätskontrolle
- ∅ Sicherheitstechnik und „Intelligentes Haus“
- ∅ Personenidentifikation, Authentifizierung
- ∅ Medizintechnik, Endoskopie
- ∅ CDMA (Code-division-multiple-access) - Technik für optische Kommunikation
- ∅ Optischer Computer

6 Grundlagen zur 2D-gestützten 3D-Bildverarbeitung

Aufbauend auf das im vorherigen Abschnitt vorgestellte 2D/3D-Kamerakzept sollen in diesem Abschnitt grundlegende Verfahren zur Verarbeitung der mit Hilfe dieser Kamera gewonnenen Daten untersucht werden. Dabei sollen gleichermaßen mathematische und algorithmische Grundlagen behandelt werden, welche unabhängig von der nachfolgenden Verarbeitung sind. Zuvor jedoch werden die Arbeitsumgebung und die Einschränkungen hinsichtlich der Simulation kurz erläutert.

6.1 Arbeitsumgebung

Für die Untersuchung von 3D-BV Ansätzen wurde eine Arbeitsumgebung aufgebaut, innerhalb derer der gesamte Arbeitsablauf von Einlesen und Aufbereitung über Vorverarbeitung und Auswertung bis hin zur Visualisierung der 3D-Daten untersucht werden kann.

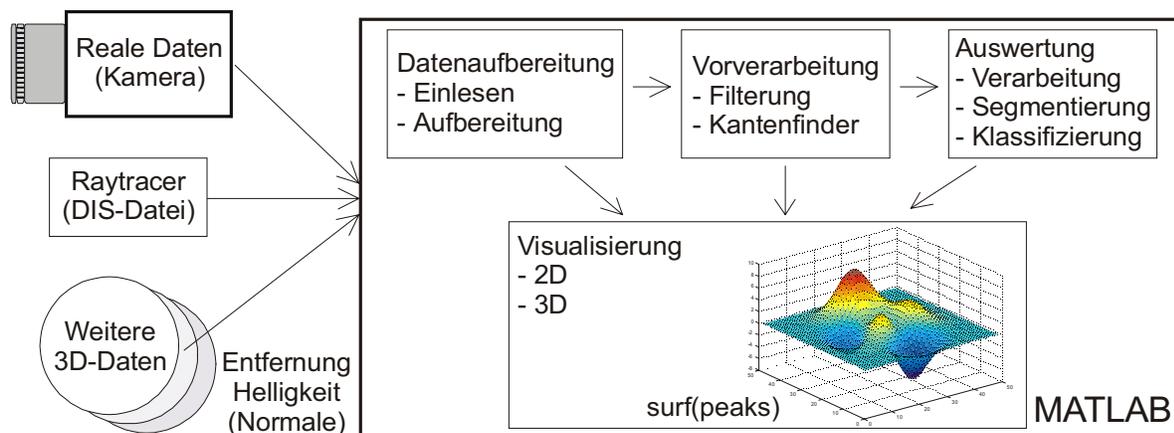


Abb. 6.1: Arbeitsumgebung zur Simulation der 3D-BV

In Abb. 6.1 sind die Arbeitsumgebung und der dazugehörige Ablauf dargestellt. Datensätze mit Entfernungs- und Helligkeitswerten stellen dabei die Grundlage einer jeden 3D-Untersuchung dar. Diese können wahlweise von einem realen Sensor oder synthetisch von einem Programm erzeugt werden. Da jede Datenquelle ein eigenes Datenformat verwendet, müssen die Daten zunächst mit Hilfe der Datenaufbereitung in ein in der Simulationsumgebung allgemeingültiges Datenformat konvertiert werden. Auf Basis dieser Daten kann nachfolgend die Vorverarbeitung und Auswertung erfolgen. Als Arbeitsumgebung für diesen gesamten Zyklus wurde die Programmiersprache MATLAB (Version 5.x) gewählt. Neben der einfachen numerischen, symbolischen Berechnung und den optimierten

Matrixfunktionen bietet diese Sprache einen großen Befehlssatz zur Visualisierung von zwei- und dreidimensionalen Daten.

Alle im Rahmen dieser Arbeit geschriebenen Programme sind in Bezug auf Benutzerfreundlichkeit und Geschwindigkeit nicht optimiert. Es wurde vielmehr darauf geachtet, eine klare Strukturierung der einzelnen Programme für eine bessere Lesbarkeit und ein besseres Verständnis einzubringen. Innerhalb der gesamten Programmstruktur wurde versucht, einen weitgehend modularen Aufbau zu verwirklichen, um einen einfachen Austausch einzelner Module und somit die Adaption der Verfahren an weiterführende Fragestellungen zu ermöglichen.

6.1.1 Simulationsdaten

Als Grundlage jeder Simulation sind unter anderem die Datensätze anzusehen. Nur mit einer reichhaltigen Auswahl von Datensätzen können verlässliche Aussagen über die Qualität und die Randbedingungen der zu untersuchenden Verfahren gemacht werden. Durch unterschiedliche Signal/Rauschverhältnisse der Daten können die Verfahren auf ihre Robustheit und ihre Anwendungstauglichkeit hin untersucht werden. Speziell auf die Bedürfnisse eines Programms ausgewählte Daten führen zwar zu einer Aussage über die Funktionsfähigkeit des Algorithmus, was aber nicht gleichbedeutend mit der Funktionsfähigkeit des Programms für allgemeine Anwendungsfälle ist.

Simulationsdaten können auf zweierlei Arten erzeugt werden: a) mittels eines realen Sensors oder b) durch synthetische Generierung durch ein Programm. Daten von einem realen Sensor (im folgenden reale Daten) sollten dabei immer bevorzugt werden. Synthetisch generierte Daten sind in den Fällen einzusetzen, in denen keine realen Daten zur Verfügung stehen oder spezielle Anforderungen an die Daten gestellt werden. Diese Anforderungen sind beispielsweise der einfachere und schnellere Vergleich von verschiedenen BV-Ansätzen, die Bestimmung der Genauigkeit der Ansätze, die Reproduzierbarkeit der Daten oder die Möglichkeit der einfachen Änderung eines Parameters (Blickwinkel, Entfernung, Oberflächeneigenschaften usw.) bei der Generierung der Daten.

Innerhalb dieser Ausarbeitung wurden synthetisch generierte Daten vereinzelt als Ersatz für Tiefendaten genutzt. Die synthetischen Daten stammen dabei wahlweise aus einem angepassten Raytracer-Programm (Object-Oriented Ray Tracer OORT [Wi94]), welches zusätzlich zu den üblichen Farbwerten auch die innerhalb des Programms zur Verfügung

stehenden Entfernungsdaten vom Objekt zur Sensorfläche und die Oberflächenorientierung ausgibt oder von beliebigen 3D-Daten, welche z.B. im Internet zu finden sind (siehe [WwwMa99], [FtpMsu99] und [FtpUtah99]).

6.1.2 Einschränkungen hinsichtlich der Simulation

Im Rahmen einer Simulation wird ein Ausschnitt der Realität in ein Simulationsmodell überführt. Bei dem Übergang von der Realität in die Simulation sind Einschränkungen notwendig. Komplexe Abhängigkeiten lassen sich beispielsweise nur sehr schwer erfassen und noch schwieriger in der Simulation rekonstruieren. Zu komplexe Modelle führen obendrein zu aufwendigen und langsamen Simulationen, sofern sie überhaupt erfaßbar sind. Somit ergibt sich, daß das Simulationsmodell immer nur ein Teil der Realität repräsentiert. Gleiches gilt auch für die Simulationsdaten. Auch hier sind Einschränkungen bei synthetischen Daten hinsichtlich der Generierung notwendig.

Innerhalb dieser Ausarbeitung wurden bezüglich des Simulationsmodells und der Simulationsdaten folgende Einschränkungen bzw. Voraussetzungen festgesetzt:

- ∅ Es wird eine Lochkamera verwendet:
 - ∅ Das Abbildungsverhältnis ist bekannt
 - ∅ Die optische Genauigkeit des Sensors ist hinreichend
 - ∅ Das Bild wird nicht optisch verzerrt
 - ∅ Die Positionen der Helligkeitspixel und der vereinzelter Entfernungs pixel stimmen überein, d.h. es erfolgt keine Zuordnung zwischen Helligkeits- und Entfernungswerten in Abhängigkeit der Objektentfernung
- ∅ Annahme, daß alle Pixel die gleiche Größe besitzen und kein unbelichteter Zwischenraum zwischen den Pixeln besteht
- ∅ Alle Helligkeitswerte liegen bereits als korrigierte Daten vor, d.h. es wird keine Korrektur der unterschiedlichen Empfindlichkeit der einzelnen Photosensoren bzw. der PMD-Elemente und der Photosensoren benötigt
- ∅ Unabhängigkeit von der verwendeten Modulation:
 - ∅ Keine Berücksichtigung des Eindeutigkeitsbereiches
 - ∅ Unabhängigkeit von der Objektoberfläche, d.h. zu stark oder zu schwach reflektierte Oberflächen beeinflussen den Entfernungswert nicht
 - ∅ Keine Untersuchung der möglichen Mehrfacherkennungsmöglichkeiten bezüglich der PN-Modulation

- ∅ Alle Entfernungswerte liegen bereits als korrigierte Daten vor, d.h. nichtlineare Abhängigkeiten hinsichtlich der verwendeten Modulation bzw. der unterschiedlichen Eigenschaften der PMD-Elemente sind bereits korrigiert
- ∅ Keine Berücksichtigung der unterschiedlichen Laufzeiten auf dem PMD-Array
- ∅ Keine Berücksichtigung von Reflexion oder Abschattung:
 - ∅ Es wird immer mit dem Entfernungswert gearbeitet, welcher die kürzeste Entfernung zur Kamera besitzt
 - ∅ Keine Beeinträchtigung des Entfernungswertes durch Reflexion oder Abschattung
 - ∅ Keine Berücksichtigung der Linsenreflexion
- ∅ Unabhängigkeit von der Beleuchtung:
 - ∅ Verwendung einer einzigen Punktlichtquelle. Daraus folgt, daß keine Interferenzen durch flächenhafte Strahler oder von mehreren Lichtquellen berücksichtigt werden
 - ∅ Die Punktlichtquelle besitzt eine homogene Lichtverteilung über die gesamte zu belichtende Fläche
 - ∅ Keine Berücksichtigung der Leistungsabnahme in Abhängigkeit der Entfernung

Einige der oben aufgeführten Punkte können bereits im ersten Schritt der Datenaufbereitung korrigiert werden oder durch entsprechende Auswahl von Arbeitsmitteln und Aufnahmebedingungen ausgeschlossen werden und sind somit keine realen Einschränkungen. Beispielsweise können die nichtlinearen Abhängigkeiten der Entfernungswerte oder die ungleichmäßigen Helligkeitswerte durch einen Einmeßprozeß korrigiert werden oder Unter- bzw. Überbelichtungen durch geeignete Einstellung der Belichtungszeit vermieden werden. Die Korrektur der Linsenverzerrung wurde bereits in vielen Ausarbeitungen ausführlich behandelt und soll somit hier nicht weiter berücksichtigt werden. Die Einschränkungen hinsichtlich der Reflexion wurden eingeführt, um den Simulationsaufwand gering zu halten. Die Annahme der gleichen Pixelgröße wurde festgelegt, da zum Zeitpunkt der Ausarbeitung die letztendliche Sensorstruktur noch nicht feststand.

Innerhalb der oben erwähnten Einschränkungen sind keine Aussagen über die Datenqualität gemacht worden. Auf eine Anpassung der Daten durch eine Quantisierung und Beaufschlagung mit Rauschen wird bei Verwendung in den nachfolgenden Verfahren gesondert hingewiesen. Ansonsten wird von ‚idealen‘ Daten ausgegangen, welche obigen Vereinbarungen entsprechen.

6.2 Untersuchung des Aufnahmesystems

Grundlage für jede 3D-BV ist die Beschreibung des Aufnahmesystems, mit dem sich die Raumpunkte aus dem Weltkoordinatensystem (WKS) über das Kamerakoordinatensystem (KKS) und das Sensorkoordinatensystem (SKS) in das Rechnerkoordinatensystem (RKS) hin- und rücktransformieren lassen. Ausgangspunkt für diese Beschreibung ist die Unterteilung des Gesamtsystems in mehrere Einzelkomponenten. Dazu sollen zunächst das WKS, das KKS und die dazugehörigen Transformationen erläutert werden. Auf Basis der Abbildungseigenschaft wird anschließend die Transformation des KKS in das SKS und RKS beschrieben. Für Grauwertbilder endet mit diesem Schritt die Transformationskette. Aufgrund dessen, daß Tiefenbilder als Entfernungswerte von der Sensorebene über das Objekt zur Lichtquelle aufgenommen werden, ist eine weitere Transformation für Tiefenbilder notwendig. Diese transformiert die gemessenen Tiefenwerte mit den Angaben aus dem SKS zurück in das KKS.

6.2.1 Koordinatensystem

In Abb. 6.2 ist die allgemeine Aufnahmesituation skizziert. Die zu vermessende Szene befindet sich im Weltkoordinatensystem, welches im folgenden mit dem Index w gekennzeichnet wird. Zusätzlich zu diesem $X_w Y_w Z_w$ -Weltkoordinatensystem wird ein $X_k Y_k Z_k$ -Kamerakoordinatensystem angenommen, welches an der Kamera ausgerichtet ist.

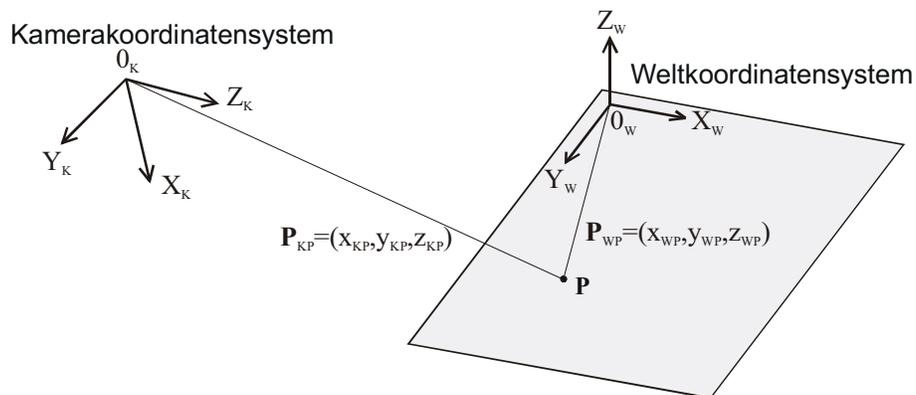


Abb. 6.2: Welt- und Kamerakoordinatensystem

Die Überführung des Weltkoordinatensystems in das Kamerakoordinatensystem erfolgt durch eine Translation und Rotation. Im homogenen Koordinatensystem dargestellt lautet diese Transformation wie folgt [Ge94]:

$$\begin{pmatrix} x_{KP} \\ y_{KP} \\ z_{KP} \\ 1 \end{pmatrix} = \begin{pmatrix} \cos \varphi \cos \theta & \cos \varphi \sin \theta \sin \psi - \sin \varphi \cos \psi & \cos \varphi \sin \theta \cos \psi - \sin \varphi \sin \psi & x_t \\ \sin \varphi \cos \theta & \sin \varphi \sin \theta \sin \psi - \cos \varphi \cos \psi & \sin \varphi \sin \theta \cos \psi - \cos \varphi \sin \psi & y_t \\ -\sin \theta & \cos \theta \sin \psi & \cos \theta \cos \psi & z_t \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{WP} \\ y_{WP} \\ z_{WP} \\ 1 \end{pmatrix} \quad (6.1)$$

Hierbei stellt der Vektor $(x_t, y_t, z_t)^T$ den Translationsvektor zwischen dem WKS und dem KKS im Kamerakoordinatensystem dar. Die Winkel ψ , θ und φ beschreiben die Rotation

Ø um die X-Achse (ψ)

Ø um die Y-Achse (θ)

Ø um die Z-Achse (φ)

im Kamerakoordinatensystem.

6.2.2 Kameramodell

Innerhalb der Kamera erfolgt die Projektion eines Raumpunktes auf eine ebene Sensorfläche, wobei aufgrund der Abbildung die Tiefeninformation verloren geht. Als Kameramodell wird ein vereinfachtes mathematisches Modell angenommen. Entsprechend einer Lochblenden-Kamera (pin hole camera) wird auch hier eine sehr kleine Öffnung der Blende vorausgesetzt. Das Objektiv weist somit keine Abbildungsfehler auf [Lu93]. Der Ursprung des KKS wird in das Zentrum der Blende gesetzt und die optische Achse in Richtung der positiven Z-Achse des KKS gelegt (Abb. 6.3).

Abbildungsfehler, wie z.B. radiale Linsenverzerrungen, Verkippung der Sensorebene u.a., werden, wie bereits erwähnt, innerhalb dieser Ausarbeitung nicht berücksichtigt. Für weiterführende Informationen wird auf folgende Literatur verwiesen: Gerdes [Ge94], Weber [We98], Klette & Koschan [KIKo96], Lenz [Le90] und Fraser & Shortis [FrSh95]

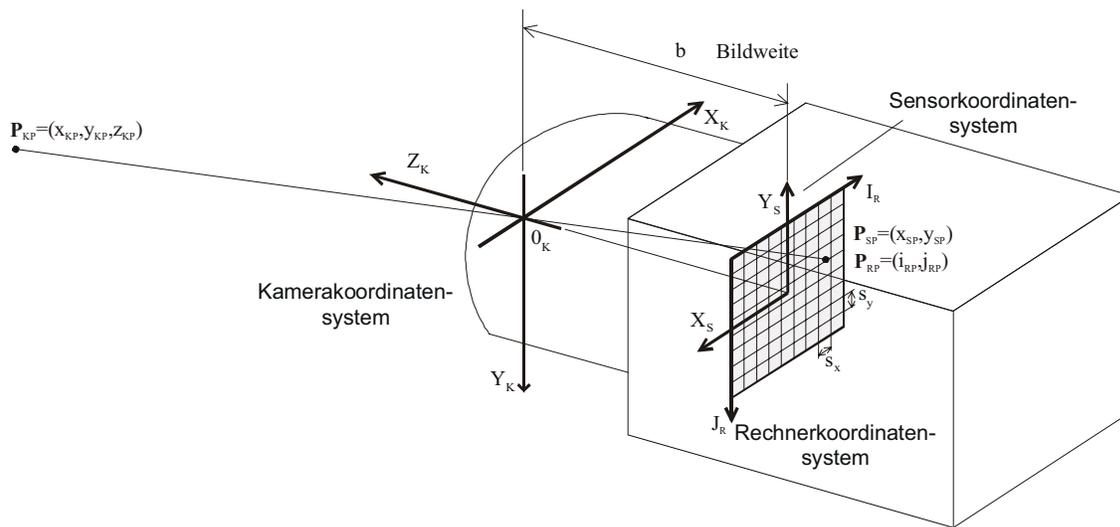


Abb. 6.3: Kameramodell mit geometrischer Optik

Das SKS (und damit die ebene Sensorfläche) befindet sich im Abstand der Bildweite b parallel hinter dem KKS. Aufgrund der Abbildungseigenschaft ist auch dieses System um 180° um die Z -Achse gedreht. Ausgehend vom KKS wird der räumliche Punkt $P_{KP} = (x_{KP}, y_{KP}, z_{KP})$ in Abhängigkeit der Bildweite b in den Punkt $P_{SP} = (x_{SP}, y_{SP})$ wie folgt abgebildet:

$$x_{SP} = b \cdot \frac{x_{KP}}{z_{KP}} \quad y_{SP} = b \cdot \frac{y_{KP}}{z_{KP}} \quad (6.2)$$

In homogenen Koordinaten ausgedrückt ergibt sich:

$$\begin{pmatrix} s \cdot x_{SP} \\ s \cdot y_{SP} \\ s \end{pmatrix} = \begin{pmatrix} b & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_{KP} \\ y_{KP} \\ z_{KP} \\ 1 \end{pmatrix} \quad (6.3)$$

Die weitere Überführung des SKS in das RKS ist abhängig von der verwendeten Pixelstruktur und der Zugriffstechnik der Software auf die einzelnen Elemente im Speicher. Die notwendige Diskretisierung der lateralen Daten erfolgt dabei direkt durch den Sensor, die der Helligkeits- bzw. Entfernungsdaten durch einen AD-Wandler.

Bei üblichen rechteckigen Pixelelementen fällt diese Transformation sehr einfach aus. Der Zugriff auf die Daten erfolgt dabei entsprechend der Anordnung in einer Matrixstruktur, wobei die Zeilen i der Matrix gewöhnlich der Y -Richtung und die Spalten j der X -Richtung entsprechen.

Die Matrixposition im RKS $P_{RP}=(i_{RP}, j_{RP})$ läßt sich unter der Annahme, daß der Nullpunkt des RKS in positiver XY-Richtung des SKS liegt, der Vorgabe der Pixelbreite s_X bzw. s_Y und der maximalen Anzahl der Elemente i_{max} bzw. j_{max} .wie folgt berechnen:

$$i_{RP} = \text{round}\left(\frac{i_{max}}{2} - \frac{y_{SP}}{s_Y}\right) \quad j_{RP} = \text{round}\left(\frac{j_{max}}{2} - \frac{x_{SP}}{s_X}\right) \quad (6.4)$$

In dieser Formel stellt die Funktion ‚round‘ entsprechend der Funktionalität der Programmiersprache MATLAB eine Funktion zur Überführung von FloatingPoint-Zahlen in Integer-Zahlen dar.

In homogenen Koordinaten lautet obige Formel wie folgt:

$$\begin{pmatrix} i_{RP} \\ j_{RP} \\ 1 \end{pmatrix} = \text{round} \left\{ \begin{pmatrix} -\frac{1}{s_Y} & 0 & \frac{i_{max}}{2} \\ 0 & -\frac{1}{s_X} & \frac{j_{max}}{2} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} y_{SP} \\ x_{SP} \\ 1 \end{pmatrix} \right\} \quad (6.5)$$

6.2.3 XYZ-Korrektur

Die Daten, welche eine Tiefenkamera aufnimmt, sind Entfernungswerte von der Lichtquelle zur Objekt Oberfläche und weiter zum Sensorelement. Nur in dem Fall, in dem Sensorelement und Beleuchtungsquelle auf der ‚gleichen‘ Position liegen, entspricht die Objektentfernung dem halben Meßwert. Andernfalls ist eine aufwendige Umrechnung des Meßwertes notwendig. Auf Grundlage dieser Objektentfernung kann anschließend die Berechnung der XYZ-Koordinaten im Kamerakoordinatensystem erfolgen.

Zur Verdeutlichung der verschiedenen Entfernungswerte für die Zentralprojektion sind in Abb. 6.4 ein nicht korrigierter Datensatz (links), ein teil-korrigierter Datensatz (Objektentfernung) (Mitte) und ein XYZ-korrigierter Datensatz (rechts) dargestellt. Die Position der Beleuchtungsquelle befindet sich links von der Kameraposition.

In der rechten Abbildung können alle Elemente der Szene (vorne: Rechteck; links: Halbkugel; rechts: zwei Rechtecke mit schräg abgeschnittener Oberfläche; hinten: auf einer Kante stehendes Rechteck) gut erkannt werden. Anders sieht es im Falle der gemessenen Entfernungsdaten (links) und der Objektentfernung (Mitte) aus. (Zur besseren Interpretation sind diese Daten negiert dargestellt, d.h. der Entfernungswert -12 ist die kürzeste Objektentfernung zwischen Kamera und Beleuchtung). Aufgrund dessen, daß die Beleuchtungsquelle im Gegensatz zur Kameraposition nach links versetzt ist (oberhalb der

Halbkugel), können aus dem linken Bild die beschriebenen Objekte nur sehr schwer erkannt werden. Die gemessenen Entfernungen nehmen mit größer werdendem Abstand von der Beleuchtungsposition zu. Einfacher, aber nicht viel besser, kann das mittlere Bild interpretiert werden. Hier wurde aus der gemessenen Entfernung von der Kamera über das Objekt bis zur Beleuchtungsquelle der Anteil der Beleuchtungsquelle herausgerechnet. Deutlich zu erkennen ist die Zunahme der Entfernung von der Mitte zum Rande des Bildes aufgrund der zunehmenden Distanz vom Sensor zum Objekt.

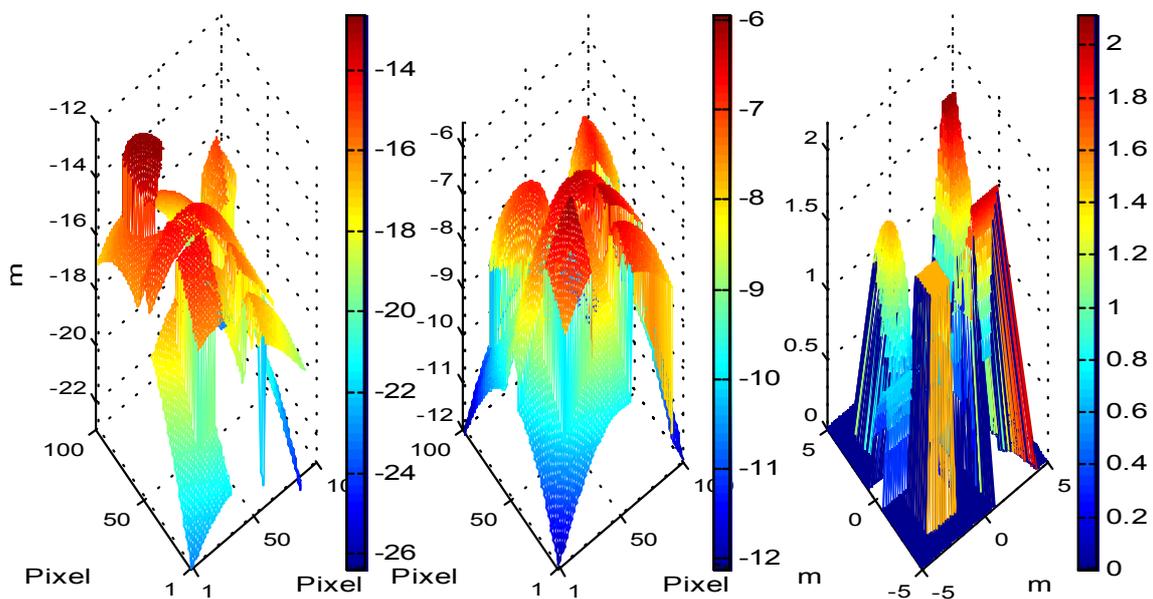


Abb. 6.4: Vergleich von nicht korrigierten (gemessenen) Daten (links), teil-korrigierten Daten (Objektentfernung) (Mitte) und XYZ korrigierten Daten (rechts) (Kamera befindet sich über dem Mittelpunkt und Beleuchtungsquelle über der Halbkugel)

6.2.3.1 Berechnung der tatsächlichen Objektentfernung

Unter der tatsächlichen Objektentfernung l_P wird in dieser Arbeit die Entfernung vom Nullpunkt des KKS zum Meßpunkt am Objekt verstanden. Der gemessene Tiefenwert l_{SENSOR} jedoch entspricht der Weglänge von der Sensorfläche über den Nullpunkt des KKS $l_{P'}$ zum Meßobjekt l_P und weiter zur Beleuchtungsquelle l_{PL} .

$$l_{SENSOR} = l_{P'} + l_P + l_{PL} \quad (6.6)$$

Durch den Beleuchtungsanteil l_{PL} ergeben sich, wie in Abb. 6.4 links zu erkennen, nichtlineare Abhängigkeiten zwischen der tatsächlichen Objektentfernung und dem gemessenen Anteil. Diesen gilt es daher aus der gemessenen Wegstrecke heraus zu rechnen.

Ausgangspunkt der Berechnung der tatsächlichen Objektentfernung l_P und damit des Meßpunktes $\mathbf{P}_P=(x_P,y_P,z_P)$ ist neben der gemessenen Entfernung l_{SENSOR} (bzw. der um die Wegstrecke vom Sensorelement zum Nullpunkt reduzierte Weg l_{SENSOR}') die Position der Beleuchtungsquelle $\mathbf{P}_L=(x_L,y_L,z_L)$ und die Sensorposition $\mathbf{P}_{P'}=(x_{P'},y_{P'},-b)$. Da die Berechnungen im Kamerakoordinatensystem erfolgen, müssen oben genannte Positionen ggf. zuerst in das KKS transformiert werden.

Umrechnung für $\mathbf{P}_L=\mathbf{P}_{P'}$

Dieser theoretisch unmögliche Fall kann angenommen werden, wenn sich die Sende- und Empfangseinheit im Verhältnis zur Objektentfernung nah genug zusammen befinden.

Unter der Annahme, daß sich die Beleuchtungsquelle im Nullpunkt des KKS befindet, ergibt sich die tatsächliche Objektentfernung aus dem halben Meßwert abzüglich der Strecke zwischen dem betrachteten Sensorelement und dem Nullpunkt:

$$l_P = \frac{l_{SENSOR}}{2} - \sqrt{x_{P'}^2 + y_{P'}^2 + b^2} \quad \text{für} \quad \mathbf{P}_L=\mathbf{P}_{P'} \quad (6.7)$$

Umrechnung für $P_L \neq P_{P'}$

Dieser allgemeine Anwendungsfall soll anhand der nachfolgenden Abbildung beschrieben werden.

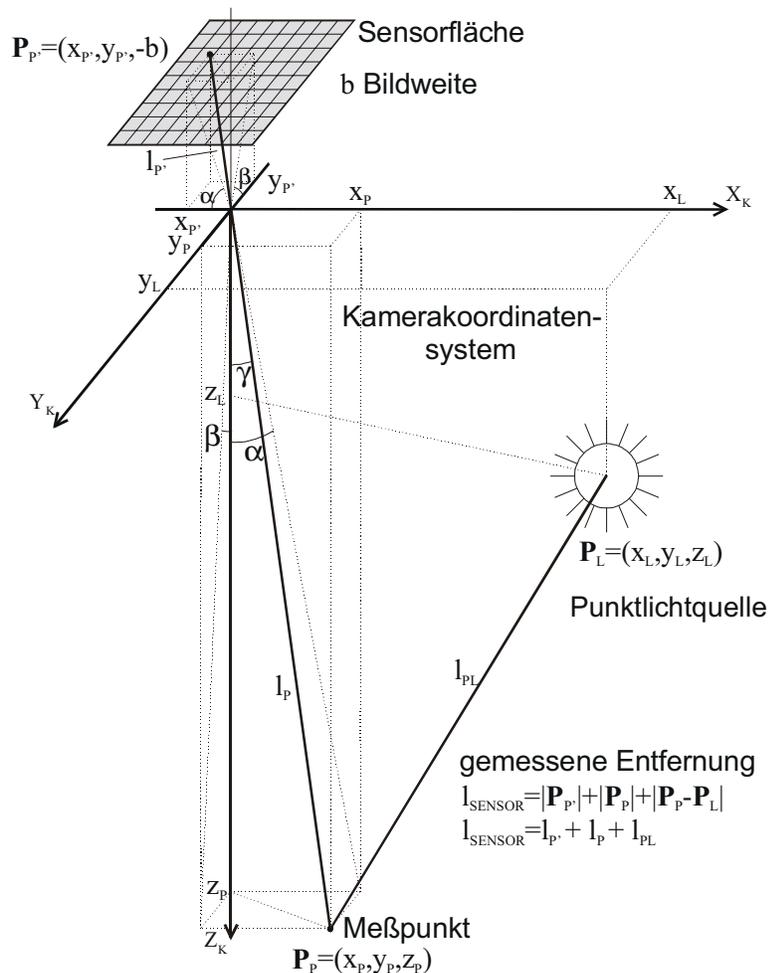


Abb. 6.5: Aufnahmegeometrie zur Bestimmung der tatsächlichen Objektentfernung

Ausgangspunkt der eigentlichen Berechnung ist die Wegstrecke l_{PL} der Punktlichtquelle zum Meßpunkt. Dafür ist es zunächst hilfreich, aus der gemessenen Entfernung l_{SENSOR} die Entfernung vom Nullpunkt bis zur Sensorposition herauszurechnen und drei im weiteren benötigte Winkel zu definieren:

$$l_{SENSOR'} = l_{SENSOR} - \sqrt{x_{P'}^2 + y_{P'}^2 + b^2} \quad (6.8)$$

$$\alpha = \arctan\left(\frac{x_{P'}}{b}\right) = \arctan\left(\frac{x_P}{z_P}\right) = \arcsin\left(\frac{x_{P'}}{\sqrt{x_{P'}^2 + b^2}}\right) = \arcsin\left(\frac{x_P}{\sqrt{x_P^2 + z_P^2}}\right) \quad (6.9)$$

$$\beta = \arctan\left(\frac{y_{P'}}{b}\right) = \arctan\left(\frac{y_P}{z_P}\right) = \arcsin\left(\frac{y_{P'}}{\sqrt{y_{P'}^2 + b^2}}\right) = \arcsin\left(\frac{y_P}{\sqrt{y_P^2 + z_P^2}}\right) \quad (6.10)$$

$$\gamma = \arccos\left(\frac{b}{\sqrt{x_{P'}^2 + y_{P'}^2 + b^2}}\right) = \arccos\left(\frac{z_P}{\sqrt{x_P^2 + y_P^2 + z_P^2}}\right) = \arccos\left(\frac{z_P}{l_P}\right) \quad (6.11)$$

Auf Basis der Gleichung (6.8) ergibt sich die Entfernung zwischen der Punktlichtquelle und dem Meßpunkt zu:

$$l_{PL} = l_{SENSOR'} - l_P = |\mathbf{P}_L - \mathbf{P}_P| \quad (6.12)$$

Nach Auflösung der Betragsgleichung und Quadrieren erhält man:

$$(l_{SENSOR'} - l_P)^2 = (x_L - x_P)^2 + (y_L - y_P)^2 + (z_L - z_P)^2 \quad (6.13)$$

Zur Reduzierung der Zahl der Unbekannten werden die Gleichungen (6.9) und (6.10) nach x_P und y_P aufgelöst und in Gleichung (6.13) eingesetzt:

$$(l_{SENSOR'} - l_P)^2 = (x_L - z_P \tan(\alpha))^2 + (y_L - z_P \tan(\beta))^2 + (z_L - z_P)^2 \quad (6.14)$$

Der Austausch der Unbekannten z_P durch Gleichung (6.11) und die Auflösung der Quadrierung ergibt:

$$l_P^2 \cdot \left(1 - \cos^2(\gamma) \tan^2(\alpha) - \cos^2(\gamma) \tan^2(\beta) - \cos^2(\gamma)\right) + \quad (6.15)$$

$$l_P^1 \cdot 2 \left(-l_{SENSOR'} + x_L \cos(\gamma) \tan(\alpha) + y_L \cos(\gamma) \tan(\beta) + z_L \cos(\gamma)\right) +$$

$$l_P^0 \cdot \left(l_{SENSOR'}^2 - x_L^2 - y_L^2 - z_L^2\right) = 0$$

Betrachtet man nur den Term a nach Einsetzen der Winkel in Abhängigkeit der Sensorkoordinaten, so löst sich dieser nach Null auf:

$$a = 1 - \frac{b^2}{x_{P'}^2 + y_{P'}^2 + b^2} \left(\frac{x_{P'}^2}{b^2} + \frac{y_{P'}^2}{b^2} + \frac{b^2}{b^2} \right) = 0 \quad (6.16)$$

Mit dieser Vereinfachung läßt sich Gleichung (6.15) einfach nach l_P umstellen. Gleichzeitiges Austauschen der Winkel in Abhängigkeit der vorgegebenen Sensorkoordinaten ergibt die gesuchte Objektentfernung:

$$l_P = \frac{\left(x_L^2 + y_L^2 + z_L^2 - l_{SENSOR'}^2\right) \sqrt{x_{P'}^2 + y_{P'}^2 + b^2}}{2 \left(x_L x_{P'} + y_L y_{P'} + z_L b - l_{SENSOR'} \sqrt{x_{P'}^2 + y_{P'}^2 + b^2}\right)} \quad \text{für } \mathbf{P}_L \neq \mathbf{P}_P \quad (6.17)$$

6.2.3.2 Berechnung der Position im Raum

Für die Verarbeitung der Meßdaten ist es vorteilhaft, die Objektentfernung in eine Position im KKS zu transformieren. Im Falle der Zentralprojektion kann diese über die Gleichungen (6.9), (6.10) und (6.11) wie folgt berechnet werden:

$$P_p = l_p \cdot \frac{(x_{p'}, x_{p'}, b)}{\sqrt{x_{p'}^2 + y_{p'}^2 + b^2}} \quad (6.18)$$

Im Falle der Parallelprojektion entspricht die Z-Komponente der tatsächlichen Entfernung und die XY-Position ergibt sich aus einer einfachen linearen Operation aus den Pixelkoordinaten.

Die Korrektur der gemessenen Entfernung in das kartesische Kamerakoordinatensystem ist in dem MATLAB-Programm *,dis2xyz.m'* für den Fall der Zentralprojektion implementiert worden.

6.2.4 Zusammenfassung

Das in dieser Arbeit beschriebene 2D/3D-Kamerakzept erzeugt ein Tiefen- und ein Grauwertbild. Unabhängig von den vereinzelt Tiefeninformationen soll im nachfolgenden ohne Verletzung der Allgemeinheit die laterale Auflösung des Tiefenbildes der lateralen Auflösung des Grauwertbildes gleichgesetzt werden. Damit ergibt sich ein Tiefenbild mit der Dimension $m \times n$:

$$\mathbf{T}=(l_{i,j}) \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (6.19)$$

Die aufgenommenen Tiefenwerte entsprechen dabei der Wegstrecke vom Sensor über das Objekt zur Lichtquelle und liegen aufgrund der Digitalisierung als diskrete Zahlen vor. Die kleinste und größte Zahl sind dem minimalen und dem maximalen Meßbereich des Sensors zugeordnet.

Für die weitere Verarbeitung der 3D-Daten ist ggf. die XYZ-Korrektur der Entfernungsdaten notwendig. Dabei werden die gemessenen Entfernungsdaten in eine Position im Raum des KKS umgerechnet und liegen fortan als 3D-Punktwolken ($0=F(x,y,z)$) vor:

$$\mathbf{T}=(x_{i,j}, y_{i,j}, z_{i,j}) \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (6.20)$$

Aus dieser Überführung ergibt sich eine Verdreifachung der Datenmenge und gleichermaßen eine Konvertierung des Datentyps. Die ehemals als Integer-Zahlen vorliegenden Tiefendaten werden mit der Konvertierung aufgrund der beliebigen Objektposition im Raum in FloatingPoint-Zahlen überführt. Die benötigte Kapazität zur Speicherung eines

Bildes erhöht sich somit auf die 6- bis 12-fache Speichermenge. Weiterführende Bildoperationen müssen fortan mit drei FloatingPoint-Zahlen durchgeführt werden, was neben einer Vergrößerung des Speicherbedarfs auch zu einer Reduzierung der Geschwindigkeit führen kann¹.

Aufgrund der schweren Handhabbarkeit dieser Datenmenge und den aufwendigen Algorithmen zur Bearbeitung der 3D-Daten kann es sinnvoll sein, diese 3D-Punktwolken nachfolgend in einen Datensatz mit äquidistantem XY-Gitter zu überführen:

$$\mathbf{T}=(Z_{i,j}) \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (6.21)$$

Hierbei sind die X- und Y-Werte implizit durch den Zeilen- bzw. Spaltenindex gegeben und lassen sich aus

$$x_{SP} = \frac{j_{\max}}{2} - j_{RP} \cdot s_X \quad y_{SP} = \frac{i_{\max}}{2} - i_{RP} \cdot s_Y \quad (6.22)$$

errechnen, wobei (i_{\max}, j_{\max}) der Anzahl der Matrixelemente entspricht und s_x und s_y die Pixelbreite in X- bzw. Y-Richtung repräsentieren. Die ehemals große Datenmenge wird hiermit zurück auf ein Drittel reduziert. Aufgrund der bekannten XY-Positionen können die Daten nun als Höhendaten ($z=Z(x,y)$) interpretiert werden. Eine weitere Reduzierung der Datenmenge kann man durch das Einbeziehen von Fehlern erreichen, wenn die Floating-Point-Zahlen zurück in Integer-Zahlen konvertiert werden.

Unter Umständen kommt es vor, daß einzelne Punkte im Meßfeld vom Sensor nicht gemessen werden können (schlechtes Reflexionsverhältnis, Abschattung usw.). Im Tiefenbild können derartige Punkte mit einem speziellen Wert – beispielsweise 0 oder NaN (Not a Number) – gekennzeichnet werden, der sonst als Tiefenwert nicht vorkommt. Andernfalls muß das Tiefenbild um ein Markierungsfeld ergänzt werden.

Die Datenspeicherung der Grauwertdaten erfolgt entsprechend der Tiefenwerte nach folgendem Prinzip:

$$\mathbf{T}=(I_{i,j}) \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (6.23)$$

Nur im Falle von registrierten Daten kann die XYZ-Position eines Grauwertpunktes direkt vom korrespondierenden Tiefenwert entnommen werden. Andernfalls befindet sich der Grauwertpunkt auf einer Geraden im Raum. Das mit dem Tiefenbild korrespondierende

¹ Neue Rechnerarchitekturen auf Basis von MMX (Intel), ISSE (Intel), AltiVec (Motorola), VIS (Sun) oder MVI (Dec) ermöglichen die parallele Verarbeitung von mehreren Datensätzen. Auf Basis dieser Technologie kann die genannte Einschränkung aufgehoben werden.

Grauwertbild muß somit den Konvertierungsschritten des Tiefenbildes folgen, falls es übereinstimmend zum Tiefenbild bleiben soll. Dies trifft in dem Fall zu, in dem sich die Matrixorganisation des Tiefenbildes ändert, d.h. die Überführung der Tiefendaten in ein äquidistantes Gitter durchgeführt wird. Für die Umrechnung der Entfernungswerte in 3D-Punktwolken wird keine gesonderte Konvertierung der Grauwertdaten benötigt.

6.3 Flächenapproximation

Die Flächenapproximation wird für eine Vielzahl von Anwendungen innerhalb der 3D-BV benötigt. Als Anwendungsbeispiele seien hier aus der Menge der Möglichkeiten

- ∅ die regionenbasierte Segmentierung
- ∅ die Kantenfindung
- ∅ die Bestimmung der Flächennormalen und
- ∅ die Überführung von nicht äquidistanten XY-Daten in äquidistante XY-Daten genannt.

Ziel der Flächenapproximation ist es, die Parameter einer Oberflächenfunktion innerhalb einer homogenen Region anhand der zur Verfügung stehenden Stützpunkte zu bestimmen. Neben der Rekonstruktion bzw. Annäherung der Oberflächenfunktion erfolgt somit auch gleichzeitig eine Rauschunterdrückung. Aufgrund der großen Bedeutung der Flächenapproximation für die 3D-BV wird nachfolgend ein Approximationsverfahren auf Basis von Polynomen beschrieben.

6.3.1 Polynomapproximation

Zur Berechnung der Polynomapproximation wird innerhalb einer Umgebung U nach einer Funktion $Z(x,y)$ gesucht, bei der die Abweichung von den gemessenen Tiefenwerten minimal ist. Als Oberflächenfunktion wird das Polynom des Grades m verwendet:

$$Z(x, y) = \sum_{k+l \leq m} a_{kl} x^k y^l \quad (6.24)$$

Zur Berechnung der Parameter von $Z(x,y)$ muß die Summe der quadratischen Fehler ε^2 innerhalb der Umgebung U minimiert werden:

$$\varepsilon^2 = \sum_{(i,j) \in U} (Z(x_{i,j}, y_{i,j}) - z_{i,j})^2 \quad (6.25)$$

Für die Lösung dieser Gleichung muß die Anzahl der Stützpunkte (XYZ-Koordinaten) größer gleich dem Grad der Flächenfunktion sein. Entspricht die Anzahl der Stützpunkte

dem Grad der Funktion, so verlaufen alle Punkte der Fläche durch die vorgegebenen Stützpunkte. Ansonsten wird eine Fläche mit dem minimalsten Fehler durch die Stützpunkte gelegt. Aufgrund der Vorgabe von XYZ-Werten ist eine vorangegangene XYZ-Transformation der gemessenen Daten notwendig.

Die Auswahl des Grades m der Flächenfunktion ist abhängig von der Szene. Besteht die Szene beispielsweise nur aus polyedrischen Objekten, so kann eine bilineare Flächenfunktion angewendet werden:

$$Z(x, y) = a_2x + a_1y + a_0 \quad (6.26)$$

Ansonsten müssen Funktionen höheren Grades verwendet werden, mit der alle Flächen der Objekte am besten beschrieben werden können.

Zur Festlegung der Umgebung U der Flächenapproximation stehen verschiedene Möglichkeiten zur Verfügung. Die einfachste Art ist eine vorangegangene kantenbasierte Segmentierung des Bildes. Die Flächenapproximation wird dann auf die einzelnen zusammenhängenden Regionen angewendet. Steht kein segmentiertes Bild zur Verfügung kann der umgedrehte Weg genutzt werden. Hierbei wird die Flächenapproximation mit einer kleinen Initialfläche gestartet und nach jedem Durchgang wird ein weiterer Stützpunkt hinzugenommen. Weicht der quadratische Fehler von einer oberen Schwelle ab oder ist die Differenz zwischen einem wahren und einem aus der Funktion berechneten Stützwert zu groß, ist eine Grenzposition gefunden. Dieser rechenintensive Vorgang wird sol oft wiederholt, bis alle Grenzen der Regionen getestet worden sind.

Ein Nachteil des Ansatzes des ‚quadratischen Fehlers‘ ist dessen schlechtes Verhalten in Bezug auf Ausreißer. Liegt bereits ein einzelner Punkt weit abseits der restlichen Vorgabewerte, kann dieser den gesamten Flächenausgleich stark verfälschen. Zur Vermeidung dieser Problematik sollten daher die Eingabedaten mit einer Ausreißerkontrolle (wie z.B. Median-Filter) vorbehandelt werden.

Da, wie bereits erwähnt, die Flächenapproximation eine Grundvoraussetzung für die 3D-BV ist, wurde die Polynomapproximation in dem MATLAB-Programm ‚*ausgleich.m*‘ implementiert. Viele weitere Verfahren greifen auf diese Funktion zurück.

6.4 Flächennormale

Neben den Grauwert- und Tiefenwerten nehmen die Flächennormalen eine gesonderte Stellung innerhalb der 3D-BV ein. Mit Hilfe der Flächennormale wird ein einfaches und sicheres Finden von Kanten in Tiefenbildern ermöglicht. Ähnliches gilt auch für die regionenbasierte Segmentierung. Nur auf Basis der Flächennormalen ist eine sinnvolle Trennung der Regionen möglich.

Für eine partiell nach x, y und z differenzierbare Bestimmungsgleichung $F(x, y, z) = 0$ bzw. eine partiell nach x und y differenzierbare Flächenfunktion $z = Z(x, y)$ ergibt sich die Flächennormale \mathbf{n} durch die partiellen Ableitungen wie folgt:

$$\mathbf{n} = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right)^T \quad (6.27)$$

$$\mathbf{n} = \left(\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y}, -1 \right)^T$$

Ergänzend zu der Flächennormale wird ebenfalls häufig die Angabe des Gradienten genutzt, welche für eine Oberflächenfunktion $z = Z(x, y)$ wie folgt definiert ist:

$$\text{grad}(Z) = \left(\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y} \right) = (p, q) \quad (6.28)$$

Hieraus ist ersichtlich, daß der Gradient und die Normale bis auf einen Skalierungsfaktor gleichbedeutend sind. Im Gradienten sind die Ableitungen einer Flächenfunktion in X- und Y-Richtung zusammengefaßt.

Unter Nutzung von mehr als drei Vorgabewerten kann mit Hilfe der Flächenapproximation eine lineare Fläche den Vorgabewerten angenähert werden. Die Flächennormale ergibt sich dabei aus der partiellen Ableitung nach x und y der approximierten Funktion.

Liegen die Funktionen $F(x, y, z)$ bzw. $Z(x, y)$ zur Beschreibung der Oberfläche nicht vor, so kann die partielle Ableitung durch eine diskrete Approximation ersetzt werden. Falls die benachbarten XY-Werte auf der Ebene parallel zu dem XY-Koordinatensystem angeordnet sind, ergibt sich die Flächennormale aus der Differenz der Höhenwerte:

$$\mathbf{n}_{i,j} = \left(\frac{z_{i+1,j} - z_{i,j}}{x_{i+1,j} - x_{i,j}}, \frac{z_{i,j+1} - z_{i,j}}{y_{i,j+1} - y_{i,j}}, -1 \right) \quad (6.29)$$

Der im 2D-Fall häufig vernachlässigte Term im Nenner darf hier aufgrund der durch die Abbildung hervorgerufenen nicht äquidistanten Schrittweiten nicht vernachlässigt werden. Aufgrund der Approximation ist das Ergebnis entsprechend mit Fehlern behaftet. Nachfolgende Verarbeitungsschritte sollten in diesem Falle von verrauschten Eingabedaten ausgehen.

Liegen die diskreten XY-Werte nicht parallel zum XY-Koordinatensystem, so kann die Flächennormale über das Vektorprodukt bestimmt werden:

$$\mathbf{n}_{ij} = (\mathbf{P}_{i+1,j} - \mathbf{P}_{i,j}) \times (\mathbf{P}_{i,j+1} - \mathbf{P}_{i,j}) \quad (6.30)$$

Im Rahmen der Untersuchungen wurde zur Erzeugung der Flächennormalen die Formel (6.27) unter Nutzung der Flächenapproximation und (6.29) für den reinen diskreten Fall in dem MATLAB-Programm *normal.m* implementiert.

6.5 Lokale Orientierung

Unter der lokalen Orientierung bzw. linearen Symmetrie in einer kleinen Bildumgebung wird die Richtung einer Kante verstanden. Zur Beschreibung dieser Eigenschaft werden zwei Komponenten benötigt. Die erste Komponente gibt an, in welche Richtung die Orientierung verläuft und die zweite, wie stark sie, gemessen z.B. an der Höhe oder Steigung der Kante, ausgeprägt ist (siehe Jähne [Jä93]). Diese beiden Informationen gehören unmittelbar zusammen, da nur bei einer signifikanten Kante die Richtungsinformation zuverlässig ist. Flächen mit konstanten Daten oder irregulären Mustern beispielsweise besitzen unterschiedliche Orientierungen mit minimaler Ausprägung innerhalb der gesamten Fläche.

Zur Bestimmung der lokalen Orientierung stehen mehrere Möglichkeiten zur Auswahl. Das einfachste Verfahren beruht auf der Ableitung des Bildes in X- und Y-Richtung. Die Steilheit der Kante ergibt sich dabei aus dem Betrag der beiden Ableitungen:

$$|\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (6.31)$$

und die Richtung aus dem Arcustangens der Verhältnisbildung:

$$\varphi = \arctan\left(\frac{\partial I}{\partial y} / \frac{\partial I}{\partial x}\right) \quad (6.32)$$

Ein weiteres Verfahren geht von Überlegungen im Fourierraum mit anschließender Überführung in den Ortsraum aus und erlaubt eine direkte Berechnung aus einfachen horizontalen und vertikalen Ableitungen.

Das Verfahren selbst wird mit Faltungsoperatoren beschrieben. Dabei stellen D_X und D_Y diskrete Ableitungsoperatoren und B eine frei wählbare Glättungsmaske an einem Bildpunkt dar. Durch die Faltungsoperation mit der Glättungsmaske wird die Maske über das ganze Bild geschoben und damit die Orientierung überall berechnet (siehe Jähne [Jä93]). In Operatorschreibweise ergibt sich für die Steilheit:

$$|\nabla I| = \sqrt{(B(D_Y \bullet D_Y - D_X \bullet D_X))^2 + (2B(D_X \bullet D_Y))^2} \quad (6.33)$$

Die Richtung ist nur im Bereich von 0° bis 180° eindeutig und ergibt sich aus:

$$\varphi = \frac{1}{2} \arctan \left(\frac{2B(D_X \bullet D_Y)}{B(D_Y \bullet D_Y - D_X \bullet D_X)} \right) \quad (6.34)$$

Aufgrund der nichtlinearen Operation darf die Reihenfolge der Operatoren nicht vertauscht werden. Für eine genaue mathematische Abhandlung wird auf [Jä93] verwiesen.

Beide Ansätze wurden in dem MATLAB-Programm *,orientierung.m'* implementiert.

6.6 Reflexionsmodelle

Viele Methoden zur Rekonstruktion dreidimensionaler Objektmerkmale wie Tiefe oder Oberflächenorientierung aus Grauwertbildern setzen gewisse Reflexionseigenschaften der zu analysierenden Objektoberflächen voraus. Aus diesem Grund sollen hier einige Gesetze und Modelle behandelt werden, die bei der Analyse von Grauwertbildern von Interesse sind. Bevor jedoch auf diese Eigenschaften eingegangen wird, sollen zunächst einige Bezeichnungen hierzu definiert werden.

6.6.1 Grundlagen

Die von einer Kamera oder einem Beobachter empfangene Lichtstärke hängt unter anderem von der Art und Position der Lichtquellen, der Orientierung und Oberflächeneigenschaften des Objektes und der Position des Betrachters ab. Zur Verdeutlichung der Abhängigkeiten ist in Abb. 6.6 ein einfaches Modell dargestellt. Eine unendlich weit entfernte Punktlichtquelle beleuchtet aus der Richtung s einen Oberflächenausschnitt mit der Flächennormale n . Der so beleuchtete Punkt wird von einem unendlich weit entfernten Beobachter aus der Richtung v betrachtet. Aufgrund der jeweils unendlichen Entfernung

der Punktlichtquelle und des Beobachters werden die dazugehörigen Vektoren als Normalenvektoren angegeben. Infolgedessen reichen bereits zwei Werte zur Spezifizierung der Vektoren aus. In der 3D-BV werden für diese zwei Werte neben der Festlegung $((\delta F/\delta x)/(\delta F/\delta z), (\delta F/\delta y)/(\delta F/\delta z), 1)$ und den Gradienten (p, q) auch häufig die Winkel Deklination σ (slant) und Azimut θ (tilt) verwendet. Die Deklination ist dabei der Winkel zur z-Achse ($\arccos(z)$) und der Azimut ist der Winkel des Punktes in der Ebene ($\arctan(y/x)$).

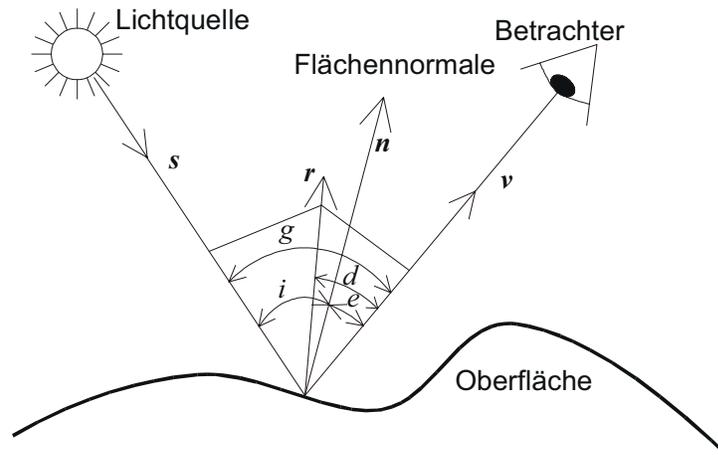


Abb. 6.6: Geometrie des Aufnahmesystems

Für die Bestimmung der Reflexionsabhängigkeiten sind die Winkel zwischen den Vektoren von besonderem Interesse. Üblich ist dabei die Angabe der photometrischen Winkel i , e , d und g (siehe Abb. 6.6). Dabei ist i der Winkel zwischen der Oberflächennormalen und der Beleuchtungsrichtung (incident angle), e der Winkel zwischen der Oberflächennormalen und der Beobachtungsrichtung (emitance angle, emergent angle), d der Winkel zwischen der Beobachtungsrichtung und der gerichteten Reflexion der Beleuchtung und g der Winkel zwischen der Beleuchtungsrichtung und der Beobachtungsrichtung (phase angle).

6.6.2 Beleuchtungs- und Reflexionsmodelle

Die Lichtmenge, die als Grauwert in einem Punkt eines digitalen Bildes codiert ist, kann als das Ergebnis einer Vielzahl von Interaktionen zwischen Materialien und Beleuchtungen verstanden werden. Diese Abhängigkeiten wurden bereits vielfach im Bereich der Computergrafik [FoDa94], [PöCl94], [Wa93] oder optischen Physik [BeSc93], [He89] untersucht und sollen hier im Sinne der 3D-BV betrachtet werden.

Viele gravierende Veränderungen finden bei der Interaktion von Licht an Grenzflächen zwischen zwei Medien statt. An ihnen können drei Effekte eintreten: Das Licht wird reflektiert, transmittiert oder absorbiert. Die Transmission und Absorption erzeugen in unserem

Falle nur eine reine Abschwächung des uns interessierenden reflektierten Signals und sollen somit nicht weiter berücksichtigt werden. Innerhalb der Reflexion gibt die Oberflächeneigenschaft die Art der Reflexion vor. Rauhe Oberflächen erzeugen eine diffuse Reflexion des Lichtes und glatte Oberflächen eine spiegelnde Reflexion. Weitere Abhängigkeiten sind beispielsweise der entfernungsabhängige Helligkeitsabfall, die Abhängigkeit von der Wellenlänge und die Art der Lichtquelle. Mit Hilfe einer ambienten Lichtquelle werden beispielsweise alle indirekten Beleuchtungen einer Szene zusammengefaßt. Selbst wenn Lichtquellen verdeckt sind oder nur indirekte Beleuchtungen in einer Szene vorhanden sind, erscheinen Objekte in einer bestimmten Grundhelligkeit. Diese Vielzahl von Abhängigkeiten können nur sehr schwer durch eine mathematische Formel zusammengefaßt werden. Zur Vereinfachung schlagen daher Klette & Koschan [KIKo96] vor, daß möglichst die Anforderungen der Einfachheit und der physikalischen Plausibilität erfüllt werden. Dies führt folglich zurück auf die von Horn bereits im Rahmen der Photometrischen Verfahren (siehe Kap. 3.2.2) aufgestellte Formel:

$$I(x, y, z) = \kappa \cdot I_i(x, y, z) \cdot \Phi(\mathbf{n}(x, y, z), \mathbf{s}(x, y, z), \mathbf{v}(x, y, z)) \quad (6.35)$$

Für den praktischen Anwendungsfall der 3D-BV kann diese Formel weiterhin vereinfacht werden. Hier wird im allgemeinen davon ausgegangen, daß sich die Reflexion aus der diffusen- und der spiegelnden Reflexion additiv zusammensetzt. Sie werden daher in den folgenden beiden Abschnitten kurz erläutert. Tiefergehende Reflexionsmodelle ergänzend zu dieser Ausarbeitung sind beispielsweise in Horn [Ho77] und Zhang [Zh95] zu finden.

6.6.3 Diffuse Reflexion

Diffuse Reflexion oder auch matte Reflexion tritt an stumpfen oder matten Flächen (z.B. Kreide) auf. Solche Flächen erscheinen unter allen Betrachtungswinkeln gleich hell, da sie das Licht mit gleicher Intensität in alle Richtungen reflektieren. Die beobachtete Intensität ist proportional zum Kosinus des Winkels zwischen der Flächennormalen und der Beleuchtungsrichtung und lautet:

$$\Phi = \alpha_d \cdot \cos(i) \quad (6.36)$$

Sind beide Vektoren normiert, so kann der Kosinus des Winkels durch das Skalarprodukt zwischen den beiden Vektoren ersetzt werden:

$$\Phi = \alpha_d \cdot (\mathbf{n}_e \cdot \mathbf{s}_e) \quad (6.37)$$

Aufgrund der genannten Eigenschaften wird die diffuse Reflexion im allgemeinen durch den Lambertschen Kosinus beschrieben. Daher wird im folgenden diese Art der Reflexion sowohl als diffuse als auch als Lambertsche Reflexion bezeichnet.

6.6.4 Spiegelnde Reflexion

Im Gegensatz zur diffusen Reflexion existieren für die Beschreibung der spiegelnden Reflexion (specular reflection) eine Vielzahl von Modellen (Beckmann-Spizzichino-Modell, Torrance-Sparrow-Modell, Phong-Modell, ...), von denen hier nur das einfachste und gebräuchlichste Modell von Phong beschrieben werden soll.

Eine ideale spiegelnde Reflexion reflektiert das einfallende Licht nur in Richtung des reflektierten Beleuchtungsvektors und vermindert dabei dessen Intensität. Das bedeutet, daß nur dann eine Intensität wahrgenommen werden kann, wenn der Blickpunkt genau in Richtung des reflektierten Beleuchtungsvektors liegt. Phong [Po75] hat dieses Modell in der Hinsicht erweitert, daß auch ein gewisser Nachbarschaftsbereich mit einbezogen wird. Je größer die Rauigkeit der Oberfläche, desto breiter wird dieser Nachbarschaftsbereich. Dieses empirische Modell ergänzt somit die Parameter einer Fläche durch eine Helligkeitsexponente, die angibt, wie matt oder glänzend eine Oberfläche erscheinen soll. Mathematisch ausgedrückt lautet diese Abhängigkeit wie folgt:

$$\Phi = \alpha_s \cdot \cos^n(d) \quad (6.38)$$

Der Grad der spiegelnden Reflexion wird mit n angegeben. Je größer n ist, desto schmaler wird der Nachbarschaftsbereich und um so größer die Spiegelwirkung.

Im Falle von normierten Vektoren kann die Gleichung wie folgt geschrieben werden :

$$\Phi = \alpha_s \cdot (\mathbf{n}_e \cdot \mathbf{r}_e)^n \quad \text{mit} \quad \mathbf{r} = 2\mathbf{n}(\mathbf{n} \cdot \mathbf{s}) - \mathbf{s} \quad (6.39)$$

6.7 Shape-from-Shading

Ausgangspunkt von Shape-from-Shading ist die Beschreibung der Bildintensität in Abhängigkeit der Reflexionseigenschaft, wie sie bereits in Abschnitt 6.6.2 (Beleuchtungs- und Reflexionsmodelle) beschrieben wurde:

$$I(x, y) = \kappa I_i \Phi(\mathbf{n}, \mathbf{s}, \mathbf{v}) \quad (6.40)$$

SFS versucht auf Basis dieser Gleichung über die gemessene Bildintensität I die relative Objektentfernung oder Oberflächenorientierung durch eine angenommene Beschreibung der Reflexionseigenschaft der Objektoberfläche in Abhängigkeit der Flächennormalen, der Beleuchtungs- und evtl. der Betrachterposition zu rekonstruieren. Diese Eigenschaft

bezogen auf eine homogene Fläche wird als Reflektanzkarte R (reflectance map) oder Reflektanzfunktion bezeichnet:

$$I(x, y) = R(\mathbf{n}, \mathbf{s}, \mathbf{v}) \quad (6.41)$$

Aus dieser Gleichung wird deutlich, daß das Rekonstruktionsproblem aus einem einzigen Grauwertbild ohne weitere Einschränkungen unterbestimmt ist. Für eine beobachtete Intensität I existieren demnach unendlich viele Lösungen. Zur Vereinfachung des Problems sind daher unter anderem Einschränkungen bezüglich der Reflektanzkarte notwendig. Als Unbekannte werden beispielsweise nur die Flächenfunktion $z=Z(x,y)$ oder deren partielle Ableitungen, die Flächennormale bzw. die Gradienten, zugelassen:

$$\mathbf{n} = \left(-\frac{\partial Z}{\partial x}, -\frac{\partial Z}{\partial y}, 1 \right) = (-p, -q, 1) \quad (6.42)$$

Die Beleuchtungsposition und die Beobachterposition werden als bekannt vorausgesetzt. Liegen beide in unendlich weiter Entfernung, reicht bereits die Angabe der Richtung dieser Vorgabewerte aus. Die aufwendige Bestimmung dieser Orientierungen in Abhängigkeit des zu rekonstruierenden Punktes kann folglich auch entfallen. Damit ergibt sich folgende häufig angetroffene Gleichung des SFS-Problems:

$$I(x, y) = R(p, q) \quad (6.43)$$

Mit Hilfe dieser Gleichung wird die Intensität in Abhängigkeit der Flächennormalen beschrieben. Wird die Beleuchtungsrichtung \mathbf{s} durch die Gradienten $(-p_s, -q_s, 1)$ angegeben und als Reflexionsfunktion die Lambert-Reflexion angenommen, so ergibt sich für die gemessene Intensität:

$$I(x, y) = \rho \cdot \frac{1 + p_s p + q_s q}{\sqrt{1 + p_s^2 + q_s^2} \sqrt{1 + p^2 + q^2}} \quad (6.44)$$

wobei κI_i zu dem Term ρ , auch Albedo genannt, zusammengefaßt wurden. Eine ebenfalls häufig angetroffene Formel gibt die Beleuchtungsposition auf Basis der Winkel Deklination und Azimut an:

$$I(x, y) = \rho \cdot \frac{\cos(\sigma) + p \cos(\theta) \sin(\sigma) + q \sin(\theta) \sin(\sigma)}{\sqrt{1 + p^2 + q^2}} \quad (6.45)$$

Zur Lösung dieser nichtlinearen Gleichung muß diese aufgrund der Anwendung auf eine Bildmatrix diskretisiert werden. Der Zeitpunkt dieser Überführung hängt von der Wahl des Lösungsansatzes ab. Ansätze von Szeliski [Sz91] oder Tsai & Shah [TsSh92] diskretisieren die Formel bereits im frühen Stadium. Im Gegensatz dazu arbeiten Ikeuchi & Horn

[IkHo81] und Oliensis [OI91] bis zum Schluß mit der kontinuierlichen Form. Erst bei der Überführung des Lösungsansatzes auf das Bildgitter wird diese diskretisiert.

Weitere Unterscheidungsmerkmale des Lösungsansatzes sind beispielsweise die Art der rekonstruierten Daten. Die auf Horns Grundidee [Ho70] basierenden Ansätze oder die Ansätze von Lee & Rosenfeld [LeRo85] rekonstruieren nur die Flächennormale im Gegensatz zur geläufigen Höhenbestimmung. Wird aufgrund der Aufgabenstellung die Höhe benötigt, muß diese in einem weiteren Integrationsschritt berechnet werden. Eine Kombination beider Ausgabedaten ist in dem Verfahren von Zheng & Chellappa [ZhCh91] zu finden. Liegen bereits Initialinformationen, wie z.B. bekannte Höhenwerte oder Oberflächenorientierungen vor, können diese zur schnelleren und besseren Rekonstruktion führen (Ikeuchi & Horn [IkHo81], Horn & Brooks [HoBr85] oder Bichsel & Pentland [BiPe92] [PeBi94]). Eine weitere Möglichkeit der Geschwindigkeitssteigerung ist die Nutzung von verschiedenen Auflösungsstufen (Zheng & Chellappa [ZhCh91], Zhang & Shah [ZhSh94] und Terzopoulos [Te84]). Solche Verfahren arbeiten sich von einem groben Gitter auf die feinste Auflösungsstufe vor und ermöglichen so eine einfache Art der Geschwindigkeitssteigerung.

Für eine genauere Beschreibung der SFS-Verfahren sei abgesehen von den Beschreibungen der entsprechenden Autoren auf a) das Buch von Klette & Koschan [KIKo96], welche einige ausgewählte Verfahren beschreiben, b) die Dissertation von Zhang [Zh95], welcher sieben ausgewählte Verfahren beschreibt und c) auf das Buch von Brooks & Horn [HoBr85] verwiesen, welches neben einer allgemeinen Beschreibung auch ausführliche Aufsätze der einzelnen Verfahren enthält.

6.7.1 Lösungsansätze

Nach Zhang [Zh95] gibt es neben den bereits aufgeführten Unterscheidungsmerkmalen drei verschiedene Richtungen zur Lösung der SFS Problematik: **Globale-, Lokale- und Propagationsverfahren.**

6.7.1.1 Globale Verfahren

Die globalen Verfahren beruhen auf dem Prinzip der Minimierung einer Energiefunktion über die gesamte Fläche. Mit einer Repräsentation der Oberflächenorientierungen $p(x,y)$ und $q(x,y)$ im Gradientenraum hat die Energiefunktion die Form:

$$\varepsilon = \iint (I(x, y) - R(p(x, y), q(x, y)))^2 dx dy \quad (6.46)$$

Diese Formel besagt, daß die Reflektanzkarte R mit der gemessenen Bildintensität I in allen Bildpunkten möglichst ähnlich sein muß. Da das Rekonstruktionsproblem in dieser Form nach wie vor unterbestimmt ist, werden weitere Einschränkungen in diese Gleichung mit aufgenommen. Eine Möglichkeit besteht im Vergleich der Ähnlichkeiten benachbarter Orientierungen (smoothness constraint). Je unterschiedlicher zwei benachbarte Orientierungen sind, desto unwahrscheinlicher ist die rekonstruierte Oberfläche. Eine glatte Oberfläche bedingt somit, daß die partiellen Richtungsableitungen $\partial p/\partial x$, $\partial p/\partial y$, $\partial q/\partial x$ und $\partial q/\partial y$ betragsmäßig klein sind. Eingebettet in die Formel (6.46) ergibt sich:

$$\varepsilon = \iint \left[(I(x, y) - R(p(x, y), q(x, y)))^2 + \gamma \left(\left(\frac{\partial p}{\partial x} \right)^2 + \left(\frac{\partial p}{\partial y} \right)^2 + \left(\frac{\partial q}{\partial x} \right)^2 + \left(\frac{\partial q}{\partial y} \right)^2 \right) \right] dx dy \quad (6.47)$$

wobei γ die relativen Gewichte der beiden Fehlerterme zum Ausdruck bringt. Weitere Einschränkungen bezüglich der Oberflächenform sind beispielsweise das ‚brightness constraint‘, das ‚brightness derivative constraint‘ und das ‚integrability constraint‘. Eine ausführliche Beschreibung der einzelnen Einschränkungen ist in Zhang [Zh95] und Klette & Koschan [KIKo96] zu finden. Neben diesen Einschränkungen sind zur Lösung der Gleichung Bildpunkte mit bekannter Forminformation (Höhenwerte oder Gradienten) erforderlich.

Die Lösung der Gleichung erfolgt zumeist über einen iterativen Ansatz. Die vielen dazu notwendigen Iterationsschritte, angewendet auf das gesamte Bild, führen zu einer genauen, aber sehr rechenaufwendigen Rekonstruktion des Tiefen- bzw. Oberflächenorientierungsbildes. Neben dem Grundansatz von Horn [Ho70] wenden Ikeuchi & Horn [IkHo81], Horn & Brooks [HoBr85], Szeliski [Sz91] und andere ebenfalls den globalen Ansatz zur Lösung des SFS-Problems an.

6.7.1.2 Lokale Verfahren

Innerhalb der lokalen Verfahren wird die Oberfläche anhand von Einschränkungen bezüglich der lokalen Nachbarschaft der Oberfläche oder der Reflexionseigenschaft berechnet. Diese Verfahren sind gegenüber den globalen Verfahren wesentlich einfacher und aufgrund ihrer geringeren Iterationsanzahl oder des nicht iterativen Ansatzes wesentlich schneller. Dafür ist ihre Rekonstruktionseigenschaft im Gegensatz zu den globalen Ansätzen schlechter.

Lokale SFS-Verfahren nahmen ihren Anfang durch die Arbeit von Pentland [Pe88], der die Reflektanzkarte linear approximiert. Zur Lösung dieses Verfahrens ist eine zweidimensionale FFT sowohl zur Hin-, als auch zur Rücktransformation notwendig. Ähnlich dem Ansatz von Pentland linearisieren Tsai & Shah [TsSh92] die Reflektanzkarte. Dazu nutzen sie die diskrete Approximation von p und q zur Linearisierung der Reflektanzkarte in Abhängigkeit von z . Lee & Rosenfeld [LeRo85] schränken das SFS-Problem in der Hinsicht ein, daß jeder Punkt der Objektoberfläche lokal durch eine Kugelfläche approximiert wird.

6.7.1.3 Propagationsverfahren

Einen Mittelweg zwischen den lokalen und globalen Verfahren bilden die Propagationsverfahren. Sie entwickeln die Oberfläche ausgehend von Punkten bekannter Oberflächenorientierung oder Höhenwerte über die Objektregionen des Bildes. Der Wachstumsprozeß kann dabei entweder iterativ oder entlang bestimmter linearer oder gekrümmter Pfade im Bild erfolgen. Mit jedem Iterationsschritt wird im iterativen Fall der rekonstruierte Oberflächenanteil ausgedehnt. Ansätze basierend auf den Propagationsverfahren wurden z.B. von Oliensis & Dupuis [Ol91], [OlDu91] oder Bichsel & Pentland [BiPe92], [PeBi94] veröffentlicht.

7 Datenfusion und -integration

Bis zu diesem Zeitpunkt wurden die Grauwert- und Tiefendaten jeweils als eigenständige, getrennte Daten angesehen. Es ist aber unbestritten, daß die Kombination dieser Daten zahlreiche Vorteile bietet. Motiviert vom menschlichen Wahrnehmungsvermögen sollen daher im folgenden einige Konzepte der Datenfusion und -integration von Grauwert- und Tiefendaten vorgestellt werden.

7.1 Motivation

Die Kombination von verschiedensten Eingangsinformationen ist naheliegend. Sie wird vom Menschen täglich auf vielfältige Art und Weise genutzt. Neben der Kombination der Daten aus verschiedensten Sinnesorganen (akustischen, visuellen, taktilen und thermischen Informationen) können auch mehrere Informationen aus einer Quelle fusioniert werden. Das menschliche Tiefensehen basiert beispielsweise auf der Kombination von binokularen Verfahren (Stereosehen, Augenwinkel) und monokularen Verfahren (Größenvergleich, Fokussierung, Tiefe aus Textur, Bewegungsanalyse, Verdeckungsanalyse, Schatteninterpretation, Wissensbasis usw.). Unabhängig von der Informationsquelle werden fehlende Daten durch anderweitig zur Verfügung stehende Informationen ergänzt und gleiche bzw. ähnliche Daten in Abhängigkeit der Informationsquelle unterschiedlich stark gewichtet (siehe Hauske [Ha94]). Als Ergebnis erhalten wir eine zuverlässigere und aussagekräftigere Beschreibung unserer Umgebung.

7.2 Grundlagen

Die Auswertung von Informationen (Meßdaten), die von mehreren gleichartigen oder unterschiedlichen Sensoren stammen, ist eine Möglichkeit, um komplexe Meßsysteme zu realisieren. Dabei muß prinzipiell zwischen der Datenintegration und Datenfusion unterschieden werden (siehe Abb. 7.1). Innerhalb der Datenintegration werden die Meßdaten zur Gewinnung jeweils einer Information verwendet. Im Gegensatz dazu ergeben sich die einzelnen Informationen bei der Datenfusion aus der Kombination mehrerer Meßdaten.

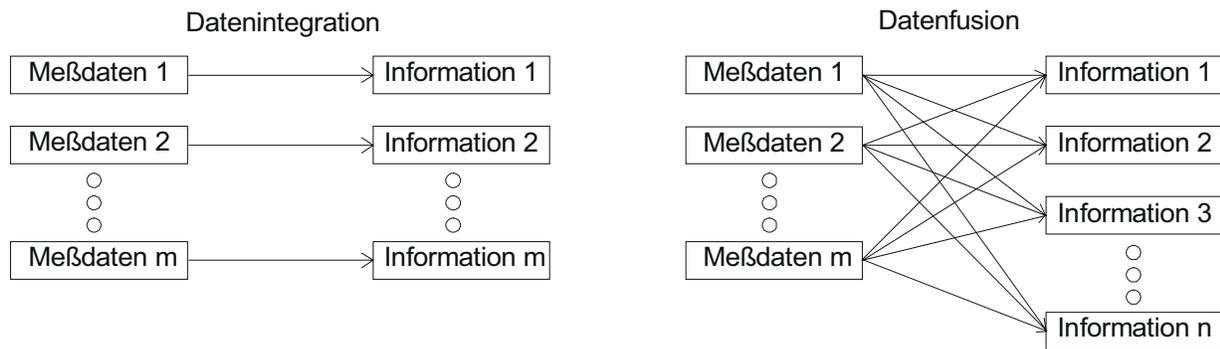


Abb. 7.1: Datenintegration und -fusion

Auf Basis der Integration werden innerhalb der 3D-BV beispielsweise die einzelnen Meßdaten neben der direkten Nutzung auch dazu genutzt, innerhalb einer weiteren Quelle Daten zu suchen. In den Bereich der Datenfusion fällt überwiegend die gleichzeitige Nutzung von Tiefen- und Grauwerten von registrierten Bildern, weswegen sie im folgenden näher betrachtet werden soll.

Als Vorteile der Datenfusion sind neben einer zuverlässigeren Informationsgewinnung und einer besseren (Wieder-) Erkennungsrate nach Hall [Hal92] folgende Punkte zu nennen:

- ∅ Robusteres Arbeiten (robust operational performance)
- ∅ Erweiterte räumliche Überdeckung (extended spatial coverage)
- ∅ Erweiterte zeitliche Überdeckung (extended temporal coverage)
- ∅ Erhöhte Sicherheit (increased confidence)
- ∅ Weniger Unklarheiten (reduced ambiguity)
- ∅ Sicherere Erkennung (improved detection)
- ∅ Erhöhte räumliche Auflösung (enhanced spatial resolution)

Um diese Vorteile zu erreichen, müssen wir die Physik und Art der jeweiligen Sensoren verstehen. Dabei liegt das Augenmerk nach Aggarwal & Chu [AgCh89] auf folgenden vier Punkten:

- ∅ Wie ist der Sensor und das dazugehörige Rauschen modelliert
- ∅ Wie stehen die Informationen der verschiedenen Sensoren in Verbindung und wie können sie kombiniert werden
- ∅ Wie können die einzelnen erkannten Merkmale der verschiedenen Sensoren verglichen werden
- ∅ Wie wählt man eine optimale Strategie für ein BV-System, um solche Informationen zu nutzen

Aufgrund der anwendungsbezogenen Betrachtungen der Datenfusion sind diese Abhängigkeiten nicht Gegenstand dieser Arbeit. Ebenso wird hier keine stochastische Interpretation der Daten berücksichtigt. Für tiefergehende Informationen in diese Richtung wird auf Loffeld [Lo90b] und [Lo90c] verwiesen.

7.3 Datenfusion in der Bildverarbeitung

Nach Chu & Aggarwal [ChAg91] gibt es zwei wesentliche Vorteile zur Nutzung der Datenfusion im Bereich der Bildverarbeitung. Erstens, solche Systeme können Daten unabhängig von der Sensibilität eines einzelnen Sensors aufnehmen (Regen, Nebel, Dunkelheit ...) und zweitens, sie ermöglichen die bessere und sicherere Erkennung von Objekten und deren Eigenschaften. Das Haupteinsatzgebiet liegt folglich im Bereich der Navigation, Inspektion und der Zielerkennung, wo entsprechend robuste und zuverlässige Aufnahmesysteme gefordert sind.

In der BV können im Bereich der Datenfusion prinzipiell drei Unterscheidungen bezüglich der Sensorkonfiguration getroffen werden, wobei diese sowohl getrennt als auch kombiniert angewendet werden können:

- 1) Datenfusion auf Basis unterschiedlicher Merkmale aus einem einzelnen Bild
- 2) Datenfusion auf Basis von unterschiedlichen Blickrichtungen und/oder Positionen, aufgenommen mit dem gleichen Sensortyp
- 3) Datenfusion auf Basis von unterschiedlichen Sensoren

Zum ersten Punkt gehören alle Verfahren, die aufgrund unterschiedlicher Strategien gleiche Informationen aus einem Bild extrahieren, wie z.B. Shape-from-X (Textur, Shading) Verfahren. Diese Techniken sind dabei sehr stark vom Vorwissen der Szene abhängig und können nur unter gewissen Einschränkungen angewendet werden.

Der Stereoansatz (statisch und dynamisch) stellt den Hauptvertreter des zweiten Punktes dar. Bei diesen wird aus bekannten Aufnahmegeometrien und übereinstimmenden Szenenpunkten die räumliche Information bestimmt.

Der für diese Arbeit interessanteste Punkt ist die Datenfusion auf Basis von unterschiedlichen Sensoren. Hierunter fallen alle Verfahren, welche Daten unterschiedlicher Informationsquellen derselben Szene fusionieren oder diese zur gegenseitigen Verifizierung nutzen. Dabei stellen unterschiedliche Sensoren nicht nur unterschiedliche Arten von Informationen, sondern auch verschiedene Beobachtungen derselben Information zur

Verfügung. Beispielsweise können Objektkonturen von Tiefendaten und Grauwertdaten (über die Reflexionseigenschaft) sowohl zur sichereren Erkennung als auch zur Verifizierung der jeweiligen anderen Information genutzt werden.

Innerhalb der Punkte 2) und 3) treten bezüglich der Datenfusion auf Bildebene verschiedene Problematiken auf, die den weiteren Fusionsprozeß bestimmen [DuNi79]:

- ∅ Unterschiedliche Bildauflösung hinsichtlich der verschiedenen Sensoren
- ∅ Unterschiedliche Quantisierung, Quantisierungsfehler und Rauschen
- ∅ Unterschiedliche Beleuchtungseigenschaften (Schattierung, Schatten, Reflexionen, ...)
- ∅ Unterschiedliche Bewertung der Oberflächeneigenschaften (Markierungen, Verschmutzungen ...)
- ∅ Unterschiedliche Aufnahmerichtungen, Entfernungen und Projektionsarten und damit unterschiedliche Sicht und Verdeckungen der Objekte

7.4 Datenfusion von Tiefen- und Grauwertdaten

Das in dieser Arbeit vorgestellte 2D/3D-Kamerakzept liefert sowohl ein Grauwertbild als auch ein Tiefenbild. Daher werden im folgenden einige Verfahren zur Fusion dieser beiden Datenquellen beschrieben. Die laterale Auflösung des Tiefensensors hat dabei eine untergeordnete Bedeutung, d.h. das Tiefenbild muß nicht in einer vergleichbar hohen Auflösung wie das Grauwertbild vorliegen.

Bestehende Modelle zur Fusion von Grauwert- und Tiefenbildern lassen sich grob in drei Klassen oder Ebenen unterteilen: In der **Pixelebene** (Data Level) werden auf der untersten Basis die Bildpunkte des Grauwert- und Tiefenbildes direkt in Verbindung gesetzt. Auf der nächst höheren **Bildebene** erfolgt ebenfalls die Fusion auf Basis jedes einzelnen Pixels, wobei anstatt der direkten Nutzung der Grauwert- und Tiefeninformation bildpunktweise extrahierte Merkmale aus beiden Quellen (z.B. Kantenfinder) fusioniert werden. Die höchste Ebene stellt die Fusion auf **Symbol-** oder **Merkmalsebene** (Feature Level) dar. Hier werden zunächst die Merkmale aus jeder Informationsquelle getrennt extrahiert (z.B. Kantenstücke), klassifiziert und interpretiert. Erst im nachfolgenden Schritt erfolgt die Fusionierung der Merkmale.

Aus dieser Aufzählung ist zu erkennen, daß in Abhängigkeit der Ebenen bereits Operationen unabhängig von der Datenfusion auf die jeweiligen Bilder angewendet werden, angefangen bei der untersten Ebene, wo evtl. nur eine simple Filteroperation durchgeführt

wird, über die mittlere Ebene, wo ein Kantenfinder auf jedes Bild einzeln angewandt wird bis hin zur höchsten Ebene, wo jedes Bild unabhängig interpretiert und evtl. klassifiziert wird. Diese Operationen können zum einen - an die jeweilige Datenquelle angepaßt - Informationen herausfiltern, aber auch durch falsche oder fehlerhafte Annahmen nützliche Informationen unterdrücken oder falsch interpretieren. Je früher die Fusion angesetzt wird, desto komplexer, aufwendiger und rechenintensiver werden die Fusionsprozesse. Mit späterem Einstieg in den Fusionsprozeß können bereits nützliche Informationen verloren gegangen sein, oder sie sind nur schwer zu interpretieren.

Aus der spärlichen Anzahl der Veröffentlichungen mit der Zielrichtung der Fusion von Grauwert- und Tiefendaten sollen im folgenden einige Ansätze in Abhängigkeit der FusionsEbene vorgestellt werden. Das Augenmerk liegt dabei auf der Anwendbarkeit der Verfahren auf das in dieser Arbeit vorgestellte 2D/3D-Kamerakzept.

7.4.1 Fusion auf Pixelebene

Die Datenfusion auf Pixelebene basiert auf der direkten Fusion der einzelnen Datenquellen anhand spezieller Operatoren, welche die Verknüpfung dieser beiden Datenquellen beschreibt. Die Originalbilder werden dazu pixelweise miteinander fusioniert, d.h. sie müssen registriert sein. Diese Aussage bedeutet nicht gleichzeitig, daß beide Bilder in der selben lateralen Auflösung vorhanden sein müssen. Die Fusion kann auch auf vereinzelte Punkte angewendet werden.

Standardvertreter der Datenfusion auf Pixelebene sind die Shape-from-Shading Verfahren, welche ein Tiefenbild ausgehend von einem Grauwertbild und Tiefenwerten an singulären Stellen des Grauwertbildes rekonstruieren. Durch die Initialisierung der zu rekonstruierenden Tiefenmatrix durch Vorgabewerte wird die Möglichkeit des Abdriftens der Konvergenz während des Rekonstruktionsprozesses verringert und gleichzeitig die Rechenzeit verkürzt. Folgende SFS-Verfahren nutzen oder benötigen Initialwerte: Ikeuchi & Horn [IkHo81], Horn & Brooks [HoBr85], Bichsel & Pentland [BiPe92] [PeBi94] u.a. .

Ein weiterer Ansatz zur Datenfusion auf Pixelebene basiert ebenfalls auf dem Prinzip von Shape-from-Shading. Dabei werden anstelle der vorrangigen Betrachtung der Grauwertdaten innerhalb der SFS-Gleichung die realen Tiefenwerte in die Minimierungsgleichung mit einbezogen [PiGa95], [ShPiGa96]:

$$\varepsilon = \iint \left\{ (Z - D)^2 + \gamma^2 (I - R)^2 \right\} dx dy \quad (7.1)$$

Auf Basis dieser Gleichung wird das zu rekonstruierende Tiefenbild bzw. die zu rekonstruierende Flächennormale so bestimmt, daß beide Bedingungen ($(Z-D)$ für die Tiefen- und $(I-R)$ für die Intensitätsdaten) gleichermaßen minimiert werden. Die Variable γ beschreibt dabei einen Faktor zur unterschiedlich starken Gewichtung beider Terme. Entsprechend der herkömmlichen globalen SFS-Verfahren kann auch diese Gleichung um entsprechende Energiefunktionen, wie z.B. smoothing constraint oder brightness constraint (siehe Kap. 6.7.1.1), ergänzt werden. Aufgrund des globalen SFS-Ansatzes ist auch dieses Verfahren sehr rechenintensiv. Ein Anwendungsbeispiel für diesen Ansatz sind die scannenden Systeme. Durch den hohen Rauschanteil sowohl in den Tiefen- als auch in den Grauwertdaten sind beide Datenquellen sehr stark gestört.

7.4.2 Fusion auf Bildebene

Innerhalb der Fusion auf Bildebene werden bereits vorverarbeitete Bilder als Eingabedaten für die Fusion genutzt. Die Vorverarbeitung bezieht sich dabei auf der Extraktion von Merkmalen, wie z.B. Objektflächen, Kanten oder Ecken. Robmann faßt für die gebräuchlichste Kantenfusion folgende Vorteile zusammen [Ro93]:

- ∅ Die problematische Kantendetektion im Tiefenbild wird durch das korrespondierende Grauwertbild zuverlässiger gemacht
- ∅ Kanten, die im Grauwertbild nicht detektiert werden konnten, können z.T. im korrespondierenden Tiefenbild extrahiert und integriert werden

Diese Eigenschaften werden beispielsweise beim räumlich kodierten Lichtansatz für die Erkennung von nicht beleuchteten Gebieten (Schattenregionen) genutzt. Die durch den Triangulationsansatz hervorgerufenen Schattenregionen können die nachfolgende Objekterkennung bei falscher Detektion deutlich erschweren. Das gleiche gilt für falsch erkannte Kanten. Hierdurch werden Objekte ungewollt in mehrere kleinere Objekte geteilt.

7.4.2.1 Kantenklassifizierung

In Grauwert- oder Tiefenbildern detektierte Kanten haben aufgrund ihrer Entstehung unterschiedliche Bedeutungen und müssen infolgedessen nicht gemeinsam auftreten (siehe Kap. 4.2.2.1). Eine sinnvolle Klassifizierung von Kanten sollte daher sowohl mit Hilfe des Grauwert- als auch des Tiefenbildes erfolgen. Nach Zhang & Austin & Wallace [ZhAu92] können die Kanten wie folgt klassifiziert werden:

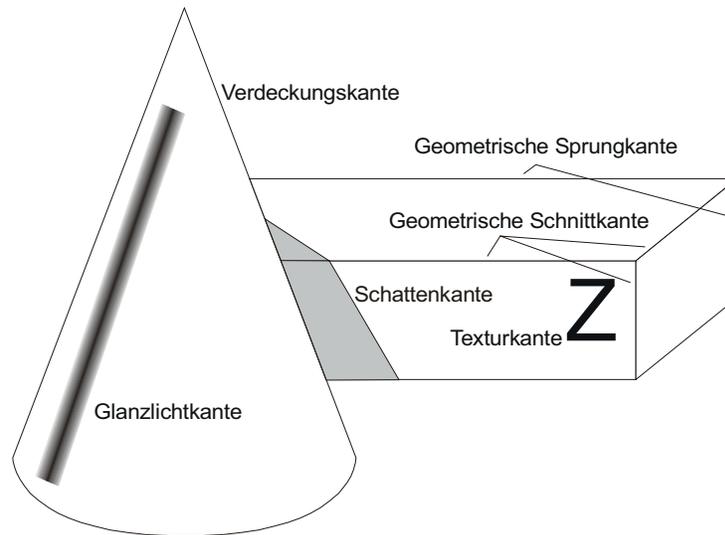


Abb. 7.2: Kantenklassifizierung

- ∅ **Geometrische Sprungkanten** (blade edges) entstehen beim Übergang von einem Objekt zum Hintergrund bzw. zu einem anderen Objekt oder wenn sich Teile eines Objektes selbst überdecken (→ Sprungkante)
- ∅ **Geometrische Schnittkanten** (fold edges) entstehen durch Veränderungen in der Geometrie von Objektoberflächen (→ Schnittkante)
- ∅ **Verdeckungskanten** (extremal edges) entstehen durch Abgrenzungen von Objekten gegen den Hintergrund bzw. gegen weitere Objekte und hängen von der Beobachtungsposition ab (→ glatte Kante)
- ∅ **Schattenkanten** (shadow edges) sind Kanten, die durch einen Schatten eines davor liegenden Objektes erzeugt werden
- ∅ **Texturkanten** (mark edges) sind Kanten, welche durch eine abrupte Änderung des Reflexionswertes erzeugt werden
- ∅ **Glanzlichtkanten** (specular edges) werden durch Oberflächenspiegelungen erzeugt
- ∅ **Falschkanten** (no-edges) werden diejenigen Kanten genannt, die zwar detektiert oder berechnet, aber nicht klassifiziert werden können (Bsp. Kratzer)

Einige der oben aufgeführten Kantenklassifizierungen können direkt aus den Tiefen- bzw. Grauwertbildern entnommen werden. Zur vollständigen Zuordnung wird jedoch die Fusionierung beider Informationsquellen benötigt, welches im folgenden anhand von Abb. 7.3 erklärt werden soll.

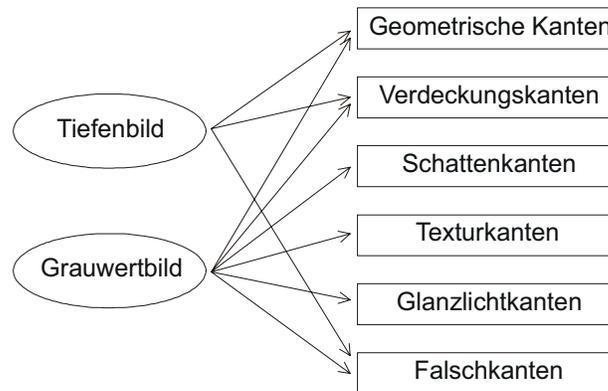


Abb. 7.3: Abhängigkeit der Kantentypen von der Datenquelle

Zur Bestimmung von Geometrischen- und Verdeckungskanten muß eine Kante in beiden Informationsquellen erkannt werden. Die Unterscheidung selbst erfolgt mit Hilfe des Kantentyps aus dem Tiefenbild. Sprungkanten weisen dabei auf geometrische Kanten und Schnittkanten auf Verdeckungskanten hin. Schatten-, Textur- und Glanzlichtkanten treten nur in Grauwertbildern auf. Eine eindeutige Unterscheidung ist nicht ohne zusätzliche Information möglich. Texturkanten können beispielsweise durch Änderung der Beleuchtungsposition identifiziert werden, da sie im Gegensatz zu Glanzlicht- und Schattenkanten an ihrer Position verbleiben.

Aus diesen Bedingungen und den Eigenschaften des Aufnahmesystems ergibt sich, daß nicht in jedem Fall alle Kantentypen erkannt und damit klassifiziert werden können. Gleichfalls kann in Abhängigkeit des Aufnahmesystems eine leicht abgeänderte Zuordnung erfolgen.

Für Verfahren, welche mit Hilfe der Kantenkorrespondenz eine Rekonstruktion der Tiefeninformation durchführen (z.B. Stereoansatz), ergibt sich aus den Erkenntnissen der Kantenklassifizierung eine Zuordnungsproblematik. Bei der Stereoanalyse dürfen nur Geometrische-, Schatten- und Texturkanten berücksichtigt werden. Glanzlichtkanten und Verdeckungskanten dürfen nicht zugeordnet werden, da sie von der Betrachterposition abhängen. Die Unterscheidung zwischen zuzuordnenden und nicht zuzuordnenden Kanten ist allerdings rein aus der Information zweier Grauwertbilder ohne eine qualitative

3D-Rekonstruktion schwierig. Diese basiert jedoch ihrerseits auf der Zuordnung der Kanten, wobei für diese Zuordnung die Klassifizierung verwendet werden soll.

7.4.2.2 Kantenfusionierung

Auch für die eigentliche Kantenfusionierung müssen entsprechend der Fusion auf Pixel-ebene beide Quellen miteinander registriert sein. Dabei ist allerdings zu bedenken, daß unterschiedliche Kantenfinder auch unterschiedliche Positionen für die gefundenen Kanten zurückgeben. Damit fallen einfache Operationen, wie die pixelweise „AND“ oder „OR“ Verknüpfung für die beiden Quellen heraus. Die pixelweise „AND“ Verknüpfung würde in einem solchen Falle für vereinzelte Kantenpunkte und die pixelweise „OR“ Verknüpfung für viele kleine Regionen sorgen.

Aufgrund der unterschiedlichen Lage der detektierten Kantenpunkte und der unterschiedlichen Empfindlichkeit der auf die jeweilige Datenquelle angepaßten Kantendetektoren werden in der Literatur verschiedene Lösungsansätze zur Kantenfusionierung vorgeschlagen. Ein häufig genutzter Lösungsansatz wird von Wei & Levi & Rembold [WeLe89] genutzt. Sie extrahieren zunächst nur die Kanten aus dem Intensitätsbild und nutzen Stichproben aus dem Tiefenbild zur Klassifizierung der Kanten. Nur wenn Anzeichen von fehlenden Kanten vorliegen, wird der entsprechende Bereich im Tiefenbild untersucht. Robmann [Ro93] nutzt ebenfalls das Intensitätsbild als Grundlage und ergänzt das aus diesem erzeugte Kantenbild um nicht korrespondierende Sprung- und Schnittkanten aus dem Tiefenbild.

Gil & Mitiche [GiMi83] und Magee & Aggarwal [MaAg85] binärisieren die detektierten Kantenpunkte aus beiden Bildern und integrieren diese durch eine globale oder lokale „AND“ Verknüpfung. Bei der lokalen „AND“ Verknüpfung wird von einer Kante im Grauwertbild ausgehend eine Kante in der $(m \times m)$ Nachbarschaft im Tiefenbild gesucht. Eine ganz andere Vorgehensweise wird in der globalen Methode verfolgt. Hier werden ausgehend vom Intensitätsbild gleichermaßen zusammenhängende Kantensegmente im Kantenbild des Tiefenbildes gesucht. Dies hat den Vorteil, daß auch kurze Aussetzer überbrückt werden. Ein der lokalen „AND“ Methode verwandter Ansatz wird von Robmann & Bunke [RoBu86] verfolgt. Sie nutzen anstelle der „AND“ die „OR“ Verknüpfung.

7.4.3 Fusion auf Symbolebene

Die Datenfusion auf der Symbol- oder Merkmalsebene basiert auf den aus Tiefen- und Grauwertdaten extrahierten Merkmalen. Die im Vergleich zu den anderen Fusionsstrategien notwendige pixelweise Übereinstimmung der beiden Datenquellen ist hier nicht gefordert. Entsprechend den Bearbeitungsebenen müssen hier die zu fusionierenden Merkmale übereinstimmen.

Bei der Fusion auf der Symbolebene werden die extrahierten Merkmale in Regionen (Oberflächen), Kanten und eventuell Eckpunkte unterteilt und weiter klassifiziert bzw. symbolisch dargestellt. Nach der Segmentierung des Bildes können aus den Regionen Merkmale wie Oberflächennormale, Parameter der approximierten Oberflächenfunktion, Oberflächenkrümmungen usw. hergeleitet werden. Für eine Fusion wird nach Robmann [Ro93] angestrebt, die Forminformationen aus beiden Bildern zu erhalten und zu vergleichen. Wei [We89], [WeLe90] untersucht die segmentierten Regionen durch eine „globale“ Krümmungsschätzung auf Inkonsistenz. Unstetigkeiten in der Krümmung auf einem Netz von Prüfpunkten deuten auf eine nichtdetektierte Kante hin. Sind außerdem Kantentyp und Kantenverlauf des Randes eines Bereiches bekannt, kann die Bereichskrümmung einfacher aus Hypothesen aufgrund der Kontur in den Bereichsecken hergeleitet, bzw. können Inkonsistenzen detektiert werden.

8 3D-Bildverarbeitung mit vereinzelt Entfernungen

Der Begriff 3D-Bildverarbeitung wird häufig mit einem hochauflösenden Entfernungsbild (große Anzahl von Entfernungswerten) entsprechend der 2D-Bildverarbeitung verbunden. Hierbei stellt sich allerdings die Frage, ob die hohe Pixelanzahl für die Weiterverarbeitung notwendig ist. Betrachtet man das menschliche Sehen, so ist erkennbar, daß der Mensch kein flächendeckendes Tiefenbild seiner Umgebung erstellt (siehe Kap. 1.2). In den meisten Fällen arbeitet er rein auf Basis eines Grauwertbildes. Sind für bestimmte Aktionen Entfernungswerte notwendig, so entnimmt er diese in Abhängigkeit der Szene aus einer beliebigen Fusion von Stereosehen, Shape-from-Shading, Shape-from-Textur, Größenvergleich von Objekten, Bewegungsanalyse und anderen. Dabei liefert keines der Verfahren ein vollständiges Tiefenbild zurück.

Entsprechend der menschlichen Vorgehensweise in Abhängigkeit der Szene die Anzahl der notwendigen Entfernungswerte zu variieren, soll dies anhand einiger Beispiele auf die 3D-BV übertragen werden. Dabei gibt die Applikation im wesentlichen die Anzahl der notwendigen Tiefenwerte vor:

- 1) Innerhalb der Objekterkennung oder der Anwesenheitskontrolle steht die Klassifizierung von Objekten im Vordergrund. Die Tiefeninformation wird nur in den Fällen benötigt, in denen die Klassifizierung innerhalb des Grauwertbildes alleine nicht möglich ist oder die Entfernung der Objekte von Interesse sind. Für beide Fälle reichen bereits eine geringe Anzahl von Entfernungswerten aus.
- 2) Im Bereich des Reverse Engineerings, der konstruktiven Überarbeitung (Ergänzung von Montageelementen), der Qualitätssicherung (SOLL-IST Vergleich) und anderen Ansätzen wird ein Bild mit möglichst vielen Tiefenwerten benötigt. Je feiner die laterale Auflösung, desto genauer ist die Rekonstruktion, bzw. desto präzisere Aussagen sind möglich.
- 3) Im Falle der Navigation von z.B. fahrerlosen Transportfahrzeugen (FTS) oder der Hinderniserkennung ist ein Mittelweg gefragt. Gegenstände in der Ferne erfordern nur eine grobe Erkennung bzw. Klassifizierung im Gegensatz zum Nahbereich, in dem eine genaue Klassifizierung notwendig ist. Da sich das Objekt jedoch in der Nähe des Sensors befindet, reichen für beide Fälle eine mittlere Anzahl von Tiefenwerten zur Klassifizierung des Szeneninhaltes aus.

Mit Hilfe des dieser Arbeit zugrundeliegenden 2D/3D-Kamerakzeptes können ungeachtet der niedrigen lateralen Auflösung der Tiefenwerte alle drei genannten Punkte ‚gleichmaßen‘ gelöst werden. Neben der direkten Nutzung der Tiefenwerte an mißverständlichen bzw. mehrdeutigen Stellen im Grauwertbild kann z.B. auch durch eine geschickte Anordnung der vereinzelt Tiefenwerte eine Navigation ermöglicht werden. Für Anwendungsfälle, in denen eine hohe Auflösung gefordert ist, besteht die Möglichkeit der Rekonstruktion der fehlenden Tiefendaten.

Aufbauend auf dieser Einteilung sollen in diesem Kapitel Lösungsansätze für die ersten beiden oben genannten Punkte vorgestellt werden. Das Hauptaugenmerk liegt dabei auf den Verfahren zur Ausweitung der vereinzelt Tiefenwerte auf Basis von Shape-from-Shading, der Flächenapproximation und den Lambert-Verfahren. Zuvor jedoch werden Lösungsansätze vorgestellt, bei denen die vereinzelt Tiefenwerte eine einfachere und sicherere Interpretation der Szene im Gegensatz zur konventionellen 2D-BV ermöglichen. Die verloren gegangene Tiefeninformation aus Grauwertbildern kann in vielen Fällen bereits mit wenigen Tiefenwerten ausgeglichen werden. Hinsichtlich der BV mit einer mittleren Anzahl von Tiefenwerten wird hier auf die Beschreibung der Anordnungsmöglichkeiten der vereinzelt Tiefenwerte verwiesen. Wie bereits in Kap. 5.3 dargestellt wurde, kann durch die Auswahl der richtigen Gitterstrukturen die laterale Auflösung der Tiefenwerte an die Aufgabenstellung angepaßt werden.

Voraussetzungen:

Zur Vereinfachung der auf dieses Sensorzept angewendeten Untersuchungen wird davon ausgegangen, daß die Pixel des Grauwertbildes und des Tiefenbildes eine einheitliche Größe besitzen und auf einem regelmäßigen Gitter verteilt sind. Wird die Gitteranordnung aufgrund der evtl. größeren PMD-Elemente geändert oder gestört, so muß dies in den entsprechenden Verfahren berücksichtigt werden. Neben der Vereinheitlichung der Pixelgröße wird weiterhin angenommen, daß sowohl die PMD- als auch die Photosensoren die selbe Empfindlichkeit und Auflösung für die Aufnahme des Helligkeitswertes besitzen.

8.1 Graubildverarbeitung mit vereinzelt Entfernungen

Auf Basis der in Kap. 5.3 (Anordnungsmöglichkeiten der vereinzelt Tiefenwerte) vorgestellten unterschiedlichen Gitteranordnungen sollen im folgenden einige Lösungsansätze bzw. Anwendungsbeispiele für die 3D-BV mit vereinzelt Entfernungen vorgestellt werden. Die Durchführung eines vollständigen 3D-BV Lösungsansatzes, wie er in Kap. 4.2 vorgestellt wurde, ist nicht Bestandteil dieses Abschnittes. Vielmehr sollen diese Beispiele zeigen, daß auch ohne ein flächendeckendes Tiefenbild eine 3D-BV durchgeführt werden kann. Die Grundlage bilden dabei Verfahren, an denen die herkömmliche 2D-BV versagt und die Hinzunahme von vereinzelt Tiefeninformationen eine einfachere und sicherere Interpretation der Szene ermöglichen, als die alleinige Nutzung von Grauwertdaten. Als Beispiele sind zu nennen:

- ∅ Bestimmung des Objektes
- ∅ Bestimmung der Objektgröße, Lage und Entfernung
- ∅ Flächeninterpretation
- ∅ Klassifizierung von Kanten
- ∅ ...

Für ein besseres Verständnis der 3D-BV Verfahren mit vereinzelt Tiefenwerten wird zunächst eine allgemeine Vorgehensweise beschrieben. Aufbauend auf diesem Wissen werden anschließend anhand der verschiedenen Gitteranordnungen einige Anwendungsgebiete vorgestellt. Abschließend folgen zwei interessante Anwendungsbeispiele.

8.1.1 Allgemein

Grundlage für die meisten nachfolgend vorgestellten 3D-BV Verfahren ist eine vollständige Segmentierung des Grauwertbildes auf herkömmliche Weise (siehe Kap. 4.2.2). Bereits in diesem Schritt kann a priori Wissen über die erwartete Szene die Segmentierung vereinfachen. Auf dieser Grundlage folgt die Interpretation bzw. Klassifikation der segmentierten Regionen. Diese kann sowohl mit einer kleinen als auch mit einer großen Anzahl von zur Verfügung stehenden Tiefenwerten in Abhängigkeit des Informationsgehaltes der Szene erfolgen. Vorhandene Oberflächenorientierungen erhöhen die Anzahl der nützlichen Informationen und führen zur Verbesserung der Interpretation.

Werden mehrere vereinzelt Tiefenwerte zur Interpretation der Regionen hinzugezogen, sollte bereits Wissen über die mögliche Verteilung der Tiefenwerte vorliegen. Z.B. wird bei

homogenen Grauwertflächen eine ebene Tiefenfläche oder bei einer runden Fläche sowohl eine Kugel als auch eine ebene Scheibe erwartet. Stimmen die erwarteten Flächenfunktionen nicht mit den vorgegebenen Tiefenwerten überein, so müssen entweder erweiterte Beschreibungen genutzt werden oder die Segmentierung muß innerhalb dieses Bereiches verfeinert werden. Folglich ist es ohne dieses a priori Wissen schwierig, die vereinzelt Tiefenwerte richtig zuzuordnen.

Wenige Tiefenwerte stellen keine große Hilfe für die Klassifizierung dar. Mit ihnen können nur allgemeine Aussagen (wie z.B.: die Objektentfernung, –orientierung und –größe) getroffen werden. Gleiches gilt für nicht klassifizierte Grauwertregionen. Auch hier können die Tiefendaten nur für allgemeine Aussagen genutzt werden.

Neben der Hilfestellung zur Klassifizierung von Objekten können die Tiefendaten auch zur Verifizierung genutzt werden. Dabei werden die klassifizierte Objekte aus dem Grauwertbild in ein räumliches Bild umgerechnet und mit den Tiefendaten verglichen.

8.1.1.1 Objekterkennung

Die Objekterkennung basiert vorwiegend auf der Interpretation des Grauwertbildes. Die nur wenigen vorhandenen Tiefenwerte können bei Unsicherheit bzw. bei mehrdeutiger Klassifizierung zur Hilfe herangezogen werden. Als einfache und schnelle Unterscheidungskriterien dienen beispielsweise die Objekthöhe und die Abmaße. Weitere Unterscheidungsmerkmale sind unter anderem im Grauwertbild nicht erkennbare Oberflächenformen oder –konturen (siehe Abb. 8.1).

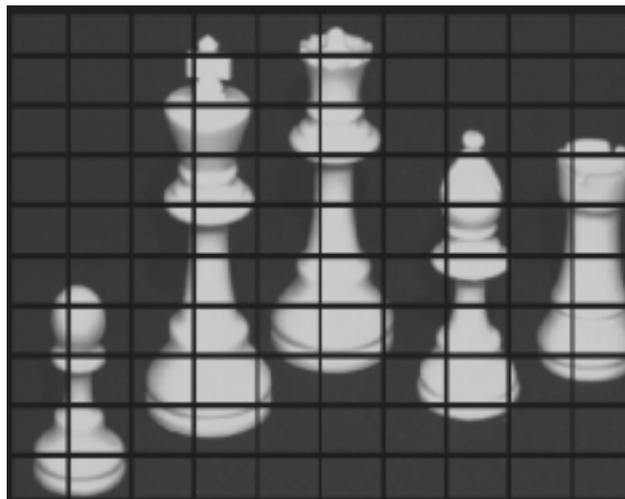


Abb. 8.1: Beispiel einer Objekterkennung mit Hilfe vereinzelt Tiefenwerte anhand der Objekthöhe, des min. und max. Durchmessers und der Oberflächenkonturen

8.1.1.2 Bestimmung der Objektgröße, Objektentfernung und Lage

Aufgrund der bei der Projektion einer Szene auf eine ebene Fläche verlorengegangenen Tiefeninformation können Abmessungen und Entfernungen aus einem Grauwertbild nur sehr aufwendig bestimmt werden. Nur wenn die Größe des Objektes und die Abbildungseigenschaft des Objektivs bekannt sind, ist eine Entfernungsangabe realisierbar. Im Gegensatz dazu ermöglichen bereits einige vereinzelte Tiefenwerte die Ermittlung der Größe, der Entfernung und der Lage von Objekten. Zwingende Voraussetzung dazu ist eine vorangegangene Objekterkennung. Ohne diese können nur allgemeine Aussagen, wie beispielsweise die minimale und maximale Entfernung einer Region, angegeben werden.

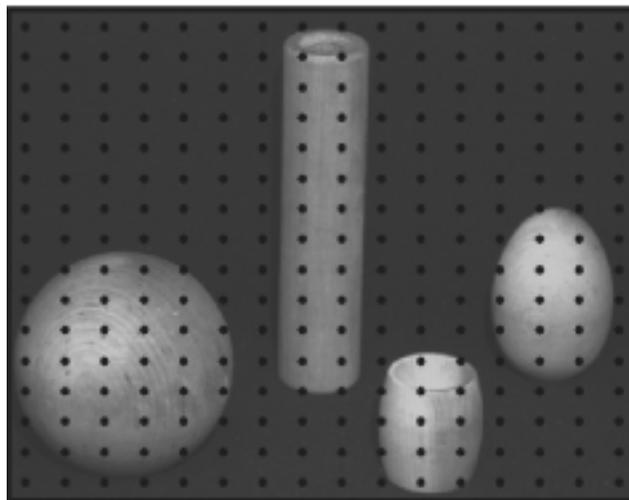


Abb. 8.2: Bestimmung der Objektgröße und –position mit Hilfe vereinzelter Tiefenwerte anhand einiger Entfernungswerte auf bereits klassifizierten Objekten

Mit mindestens einem Entfernungswert und einem Oberflächenorientierungswert oder zweier Entfernungswerte können die Entfernung und Orientierung von einfachen und bekannten Objekten im Raum bestimmt werden. Über die Größe des Objektes im Grauwertbild, der relativen Meßposition des Tiefenwertes im Grauwertbild und dem Öffnungswinkel des Objektivs kann die tatsächliche Objektgröße aus diesen wenigen Tiefendaten bestimmt werden.

8.1.1.3 Texturierte Regionen

Die Rekonstruktion von Tiefenwerten aus texturierten Grauwertregionen ist sehr problematisch. Nur bei bekannter Texturierung kann über den Shape-from-Textur (SFT) Ansatz (siehe z.B. [ChKa90], [Ga90] und [JiBu97]) die Oberflächenorientierung bestimmt werden.

Ist ferner die Größe der texturierten Elemente und der Öffnungswinkel des Objektivs bekannt, kann auf die Entfernung der Oberfläche geschlossen werden. Die Anzahl der Oberflächenorientierungen und der Tiefenwerte ist abhängig von der Form und Größe der texturierten Elemente. Über die Oberflächenorientierung und ggf. den Tiefenwert kann nachfolgend eine Segmentierung der Regionen erfolgen. Bei unbekannt texturierten Oberflächen versagt dieser Ansatz. Eine Segmentierung des texturierten Grauwertbildes ist dann mit einfachen Mitteln nicht möglich.

In Regionen ohne bekannte Texturierung können vereinzelt Tiefenwerte (und evtl. Orientierungswerte) genutzt werden, um den umgekehrten Weg zu gehen. Auf Basis der Tiefenwerte und der Flächenapproximation wird eine Initialfläche expandiert, bis der Fehler der Flächenapproximation eine festgelegte Schranke überschreitet. Aufgrund der vereinzelt Tiefenwerte kann auf dieser Basis nur eine grobe Segmentierung des Bildes erfolgen. Die feine Abgrenzung der Regionen kann anschließend mit vereinfachten Mitteln auf Basis des Grauwertbildes erfolgen.

8.1.1.4 Flächeninterpretation

Bereits 1984 wurde von Pentland [Pe84] ein Verfahren zur Bestimmung von charakteristischen Flächenmerkmalen aus Grauwertbildern vorgestellt. Anhand von Ableitungen in X- und Y-Richtung können die fundamentalen Flächentypen (Elliptisch, Zylindrisch, Ebene und Sattelfläche) geschätzt werden. Zur Bestimmung der Wölbung der Oberfläche reichen die Daten aus den Ableitungen jedoch nicht aus (siehe Abb. 8.3). Nur auf Basis von Schatten- bzw. Spiegelregionen und bekannter Beleuchtungsposition oder a priori Wissen über das zu interpretierende Objekt kann die Konvexität rein aus dem Grauwertbild bestimmt werden.

Ein direkterer Weg zur Bestimmung der Flächentypen stellt die Flächenapproximation dar. Liegen nur einfache Flächenfunktionen vor, wie sie in künstlichen Umgebungen üblich sind, reichen bereits einige wenige Tiefenwerte zur Bestimmung der Fläche aus. Ist zudem über die Grauwertverteilung oder a priori Wissen die zu erwartende Flächenfunktion bekannt, kann zum einen die Rekonstruktion auf die erwartete Funktion angepaßt oder das tatsächliche Ergebnis mit dem erwarteten kontrolliert werden.

Die Flächeninterpretation erfolgt über die in Abschnitt 6.3 beschriebene Flächenapproximation. Die eigentliche Interpretation kann dabei beispielsweise direkt über die Parameter

der ermittelten Funktion, der Krümmung, den Fundamentalgrößen (siehe Jiang & Bunke [JiBu97]) oder anderen Eigenschaften erfolgen.

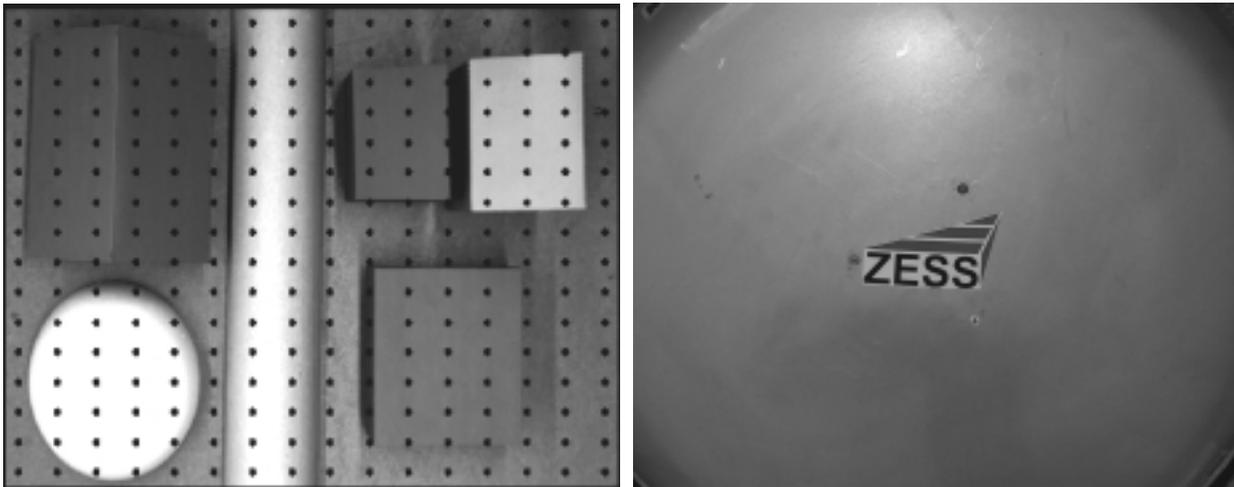


Abb. 8.3: Nutzung der Flächeninterpretation zur Bestimmung der Oberflächenform (Ebene, Halbzylinder, Halbkugel), der Orientierung und der Wölbung

8.1.1.5 Kantenklassifizierung

Kanten in Grauwertbildern werden durch Änderungen sowohl von geometrischen als auch photometrischen Bedingungen erzeugt. Ohne Hinzunahme von Tiefeninformationen können, wie bereits in Kap. 4.2.2.1 beschrieben, nur zwischen den Grauwertkantentypen Grauwert-, Textur- und Phantomkante unterschieden werden. Für eine eindeutige Klassifizierung der Kanten wird zusätzlich die Tiefeninformation an den Kantenübergängen benötigt.

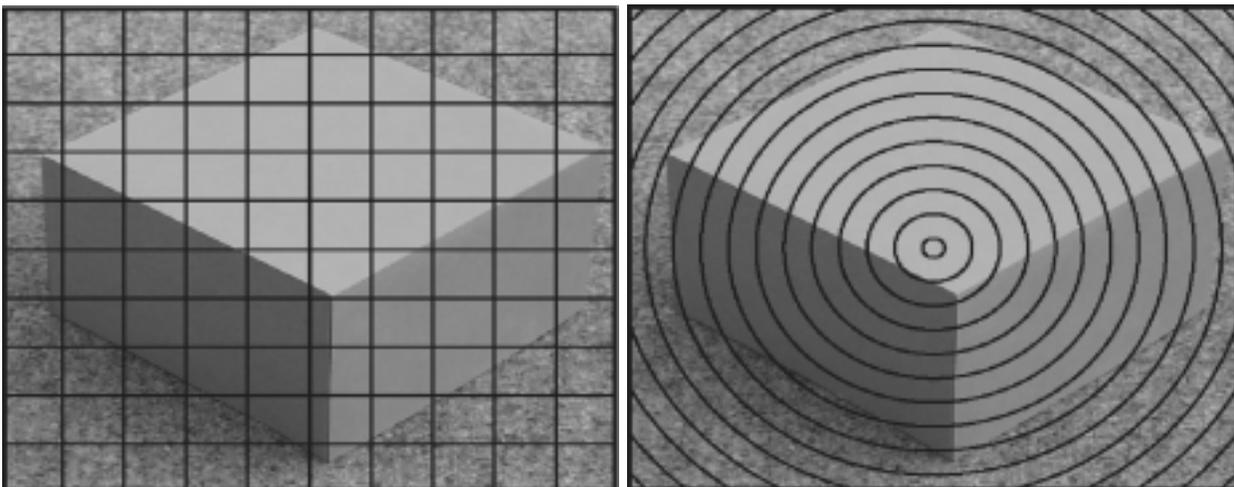


Abb. 8.4: Kantenklassifizierung mit dem Karo- und dem Kreisgitter

Zur Klassifizierung der Kante ist eine unabhängige Bestimmung des Kantentyps im Tiefenbild notwendig. Mit Hilfe der Linienstruktur (siehe Abb. 8.4) kann ein zusammenhängender Bereich im Tiefenbild untersucht werden. Aus einem kleinen Linienausschnitt im Bereich einer detektierten Grauwertkante können die Tiefenkantentypen Glatte-, Sprung-, und Schnittkante bestimmt werden. Der Winkel zwischen der Kante und der Linie ist von untergeordneter Bedeutung und beeinflusst nur die Länge des Ausschnittes. Neben geraden Liniensegmenten können auch gekrümmte Linien und/oder nicht XYZ transformierte Tiefendaten genutzt werden. Der eigentliche Kantenübergang wird hierdurch nicht verändert. Der zu untersuchende Kantenübergang ist in beiden Fällen vorhanden, nur die Kantenumgebung wird unterschiedlich dargestellt. Liegen neben den Tiefenwerten auch die Oberflächenorientierungen der einzelnen Punkte vor, kann mit diesen Informationen eine einfachere Bestimmung des Kantentyps erfolgen.

Unabhängig von einer vorangegangenen Bestimmung des Kantentyps im Grauwertbild kann die Kantenklassifizierung auch ausgehend vom Tiefenbild erfolgen. Dabei werden alle Tiefenwerte, die sich auf einer Linie befinden, auf Unstetigkeiten hin untersucht und die Grauwertdaten in der Kantennachbarschaft zur Klassifizierung herangezogen. Ist ein Vorwissen über die zu untersuchenden Objekte vorhanden, können die vereinzelt gefundenen Kanten verbunden werden.

8.1.2 Applikationsbeispiele

Nach dieser allgemeinen Einführung in die Anwendungsmöglichkeiten der Graubildverarbeitung mit Hilfe einzelner Tiefeninformationen sollen hier zwei Beispiele für die praktische Anwendung des Sensorkonzeptes vorgestellt werden. Diese Realisierungen sind heuristisch entstanden und wurden schrittweise durch Überlegungen entwickelt. Die in den vorherigen Abschnitten erwähnten Vorgehensweisen dienten dabei als Grundlage. Entsprechend den vorangegangenen Lösungsvorschlägen wurden diese Funktionen ebenfalls nicht realisiert.

8.1.2.1 Geschwindigkeitsmessung von Fahrzeugen

Heutige transportable Systeme zur Geschwindigkeitsmessung von Fahrzeugen basieren überwiegend auf der Messung der Geschwindigkeit über einen einzigen Meßstrahl, welcher genau auf das zu vermessende Objekt zu richten ist. Bei Benutzung des Gerätes von ungeschultem Personal können dadurch leicht Fehlmessungen auftreten. Mit Hilfe

einer Zeilenkamera und einiger Entfernungs- und ggf. Orientierungswerte kann die Messung der Geschwindigkeit wesentlich sicherer erfolgen. Ein zur Zeilenanordnung außerdem vorhandenes flächendeckendes Grauwertbild stellt zusätzliche Sicherheiten zur Verfügung. Die dazugehörige Meßszene ist in Abb. 8.5 dargestellt. Das Zeilenarray ist dabei senkrecht zur Straße ausgerichtet. Zur sicheren Bestimmung der Geschwindigkeit muß die vor dem Fahrzeug liegende Straße, das Fahrzeug selbst und der Horizont über dem Fahrzeug aufgenommen werden.

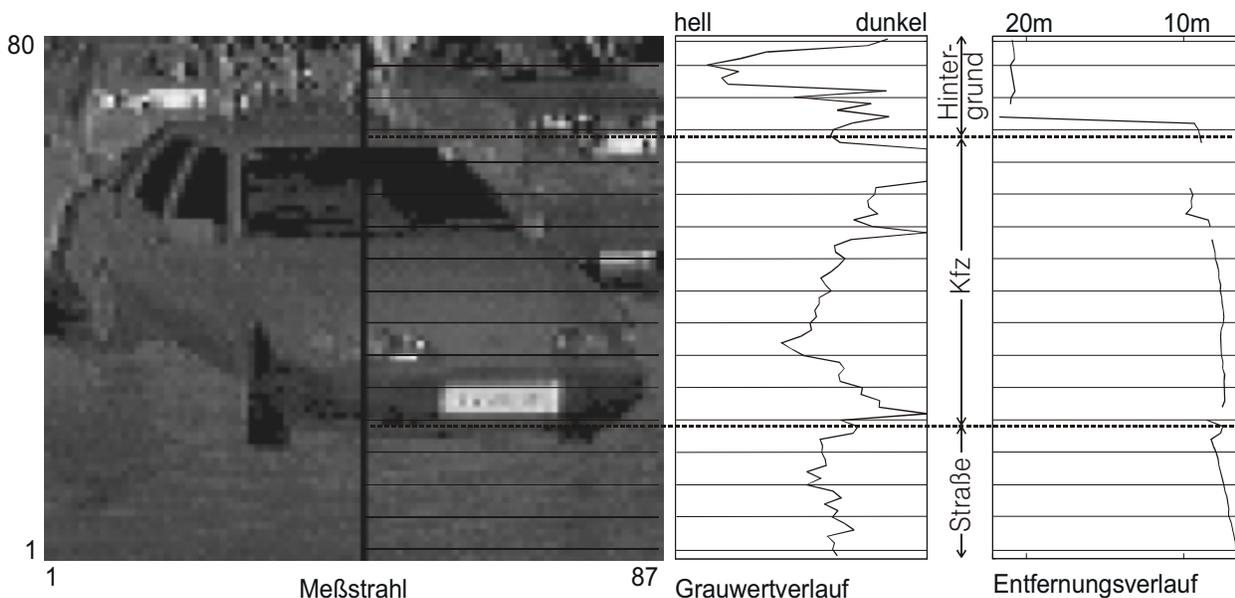


Abb. 8.5: Geschwindigkeitsmessung auf Basis einer Zeilenkamera mit vereinzelt Entfernungenssensoren¹

Die Auswertung der aufgenommenen Szene erfolgt zunächst innerhalb der Grauwerteile von unten nach oben. Der unterste Bereich der Zeile ist, wie bereits erwähnt, eine Abbildung der Straße. Innerhalb des Grauwertbildes muß aufgrund des vorwiegend gleichmäßigen Straßenbelags dieser ebenfalls als fast gleichmäßiger Grauwertverlauf abgebildet werden. Über die vereinzelt Tiefen- und Oberflächenorientierungswerte aus dem korrespondierenden Tiefenbild kann dieser Bereich zusätzlich kontrolliert werden. Die Verbindung der einzelnen Tiefenwerte muß dabei eine Gerade ergeben, die evtl. vorhandenen Orientierungswerte weisen keine oder nur geringe XY-Komponenten auf. Zur sicheren Erkennung des Fahrzeuges wird anschließend das Dach des Fahrzeuges gesucht. Dazu werden die vereinzelt Tiefenwerte beginnend vom oberen Zeilenende nach einer geometrischen Sprungkante (Horizont) abgesucht. Das gesuchte Fahrzeug liegt zwischen dem Ende der Straße und dem Beginn des Horizonts. Über den Grauwertverlauf und den

¹ Mit freundlicher Genehmigung der ‚Riegler Laser Measurement Systems GmbH‘, Austria

vorhandenen Tiefen- und Orientierungswerten innerhalb dieses Bereiches können Rückschlüsse auf die Kontur des Fahrzeuges und des -typs gezogen werden.

Die Bestimmung der Geschwindigkeit kann auf Basis des klassifizierten Fahrzeugtyps sowohl mit einem als auch mit mehreren Tiefenwerten erfolgen. Bei Wahl eines Entfernungswertes muß dieser in seiner Position bezüglich des Fahrzeuges konstant sein. Die Position sollte so gewählt werden, daß er eine ebene und senkrechte Fläche nutzt. Im Bereich der Windschutzscheibe können aufgrund von Spiegelungen und Reflexionen Fehlmessungen auftreten (siehe Abb. 8.5). Genauso ist der Kühlerbereich bei einigen Fahrzeugen zu vermeiden. Zur Erhöhung der Meßsicherheit können auch mehrere Tiefenwerte genutzt werden. Dabei erfolgt die Messung der Geschwindigkeit über eine Konturverfolgung oder die Fusion einzelner Meßergebnisse. Fehlerhafte Messungen werden durch die Interpretation der gesamten Daten frühzeitig erkannt und dem nachfolgenden Verarbeitungsschritt nicht zugeführt.

8.1.2.2 Innenraumüberwachung von Kraftfahrzeugen

Mit steigenden Sicherheitsanforderungen an ein Kraftfahrzeug wächst auch die Notwendigkeit von Sensoren zur Erkennung von Gefahrensituationen. Der alleinige Einbau von aktiven und passiven Sicherheitssystemen im Kfz bedeutet nicht gleichzeitig eine Verbesserung der Insassensicherheit. Die Verwendung von seitlichen Airbags oder die Unterteilung der Airbags in ein Mehrkammersystem bringt keine Verbesserung, sofern diese zusätzlichen Schutzmaßnahmen nicht gezielt angewendet werden. Aktiver Insassenschutz kann nur sichergestellt werden, wenn diese an die jeweilige Fahrsituation und aktuelle Sitzbelegung angepaßt ist. Nicht in allen Fällen ist beispielsweise die Auslösung des vollständigen Airbags notwendig. In vielen Situationen reicht bereits eine Teilzündung aus. Es muß nur sichergestellt sein, daß diese Teilzündung den Insassen nicht mehr gefährdet, als sie ihn schützt. Dies trifft auch auf die Seitenairbags zu. Bei Zündung eines Seitenairbags in der falschen Situation kann der Airbag anstatt des Aufbaus eines Luftpolsters die zu schützende Person aus dem Nutzungsbereich des vorderen Airbags herausschleudern. Entsprechend dürfen die Airbags bei Belegung des Sitzes mit einem Kindersitz oder einer Fehlbelegung ebenfalls nicht gezündet werden.

Ein Verfahren zum Schutz des Beifahrers kann in zwei verschiedene Bereiche unterteilt werden. Zum ersten Bereich gehört die **Sitzbelegungserkennung**. Dieser statische

Anwendungsfall bestimmt die Belegung des Beifahrersitzes. Hierunter fällt die Klassifizierung der Sitzbelegung, z.B. in Personen, Kindersitz oder Gegenstände. Zur gezielten Steuerung von Mehrkammerairbags gehört neben diesen Punkten auch die Gewichts-, Größen- und Volumenbestimmung. Das dynamische Erkennen der aktuellen Personenhaltung fällt in den zweiten Bereich zum Schutz des Beifahrers (**Out-of-Position Detektor OOP**). Beugt sich beispielsweise die Person nach vorne oder befinden sich die Füße auf dem Armaturenbrett, darf der Airbag nicht gezündet werden.

Zur Überwachung der beiden oben genannten Anwendungsfälle bietet sich das 2D/3D-Kamerakzept an, wobei durch die unterschiedliche Natur der Grauwert- und Tiefendaten neben einer Vereinfachung der Lösungsproblematik auch eine Erhöhung der Auswertesicherheit erreicht wird. Mit Hilfe eines Grauwertarrays mit vereinzelt Tiefensensoren im Punktgitter, angebracht z.B. im Bereich des Rückspiegels, kann der gesamte Beifahrerbereich überwacht werden. Da mit einem optischen System keine Gewichtsbestimmung möglich ist, müssen bestehende Konzepte wie die Gewichtsmatte oder kapazitive Meßverfahren mit dem optischen System zu einem Gesamtsystem verbunden werden. Neben der reinen Gewichtsbestimmung bietet dieses System auch eine gewisse Redundanz an. Damit können beide Systeme zur Erhöhung der Sicherheit genutzt werden.

Das optische Sensorsystem muß so ausgerichtet sein, daß es sowohl die gesamte Rückenlehne des Beifahrersitzes, unabhängig von der Sitzstellung, als auch den vorderen Bereich bis hin zum Armaturenbrett überwacht. Um eine spezielle Anpassung des Systems an einen bestimmten Fahrzeugtyp zu vermeiden, kann der Meßbereich entsprechend größer gewählt werden. In Abhängigkeit des Fahrzeugtyps und der Positionierung des Systems wird nur ein Teilbereich der gesamten Sensorfläche genutzt. Um die notwendige Rechenleistung gering zu halten, kann beispielsweise ein Grauwertsensor mit 128 x 128 Punkten Auflösung und ein Tiefensensor mit 16 x 16 Punkten Auflösung eingesetzt werden. Ein mögliches Aufnahmebild eines solchen Sensors ist in den nachfolgenden Abbildungen dargestellt.

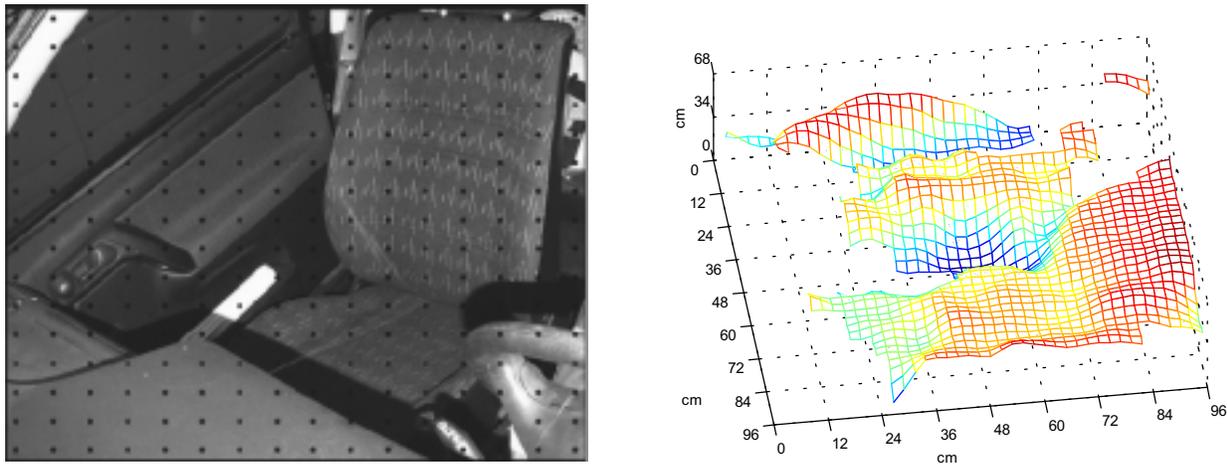


Abb. 8.6: Bilder eines leeren Beifahrersitzes
 links: Grauwertbild mit angedeuteten Tiefenwerten im 16 x 16 Punktgitter¹;
 rechts: flächendeckendes Tiefenbild (Unten: Sitz; Oben: A-Säule)²

Sitzbelegungserkennung

Innerhalb der Sitzbelegungserkennung ist unabhängig von der Sitzposition festzustellen, ob der Sitz mit einem Menschen, einem Kindersitz oder mit einem anderen Gegenstand (Handtasche, Einkaufstüten usw.) belegt ist. Im Falle des Menschen ist neben der Größe der Person auch das Volumen von Interesse. Aufgrund der Montage der Kindersitze gegen die Fahrtrichtung fällt dieses Objekt mit in die Klasse der Gegenstände. Der Airbag darf in solchen Fällen nicht gezündet werden.

Die Sitzbelegungserkennung basiert wie bei allen Verfahren dieses Kapitels vorwiegend auf den Grauwertdaten. Über diese wird zunächst auf Basis eines homogenen Grauwertverlaufes oder einer texturierten Oberfläche (siehe Abschnitt 8.1.1.3) versucht, einen unbelegten Sitz und dessen Position zu bestimmen. Wurde ein freier Sitz gefunden, kann dieser Bereich über die Tiefenwerte kontrolliert werden. Im Falle eines unbekanntes Bezuges oder Überzuges erfolgt die umgekehrte Vorgehensweise. Auf Basis der Tiefenwerte wird versucht, zwei ebene Flächen zu finden, welche eine Orientierung nach oben bzw. nach vorne haben.

Bei einem vollständig freien Sitz endet hier die Sitzbelegungserkennung, andernfalls versucht das System im nächsten Schritt herauszufinden, ob der Sitz mit einem Gegenstand oder einer Person belegt ist. Dazu wird zunächst die gesamte Beifahrertür als

¹ Mit freundlicher Unterstützung der Firma CARGLASS; Siegen-Eiserfeld

² Mit freundlicher Unterstützung der Firma Temic; Ingolstadt, Abteilung System Simulation

Referenzebene ausgewertet. Auch die Tür stellt eine ‚quasi‘ ebene Fläche mit bekannten Richtungsnormalen dar. Am rechten Ende dieser Fläche fängt der evtl. belegte Sitz an. Ausgehend von dieser Grenze kann der erkannte Sitzbereich einer Klassifizierung zugeführt werden. Dazu bieten sich beispielsweise Verfahren mit neuronalen Netzen, der Erkennung von Gesichtern, der Erkennung des Gurtes und der Kopfstütze, der Erkennung von technischen Flächen und anderen Verfahren an. Zur Erhöhung der Sicherheit des Systems können diese Verfahren nacheinander angewendet werden. Die Sitzbelegung ergibt sich abschließend aus der Fusion der einzelnen Ergebnisse. Stehen Daten wie ein angelegter Gurt oder das Personengewicht zur Verfügung, können diese ebenfalls in den Fusionsprozeß integriert werden.

Aufgrund der statischen Belegung ist eine Auswertung in Echtzeit nicht gefordert. Der gesamte Rechenprozeß kann wiederholt angewendet und die einzelnen Ergebnisse jedes Durchlaufes können ebenfalls einer Fusion zugeführt werden.

Out-of-Position Detektor

Innerhalb des Out-of-Position (OOP) Detektors werden Tiefenbereiche definiert, bei deren Überschreitung das Auslösen des Airbags unterbunden wird. In Abhängigkeit der Art und Anzahl der Airbags können diese Bereiche in mehrere Zonen aufgeteilt werden. Die Meßbereiche selbst werden nur auf Basis der Tiefenwerte kontrolliert. Befindet sich ein Objekt innerhalb des Gefahrenbereiches eines Meßstrahls, signalisiert dieser eine Gefährdung der Person. Im Zusammenhang mit der Sitzbelegungserkennung wird ein Signal zur Freischaltung des Airbags erzeugt. Nur bei Belegung mit einer Person in ‚normaler‘ Sitzhaltung darf der Airbag in einer Unfallsituation ausgelöst werden. Eine evtl. Teilzündung des Mehrkammerairbags hängt zusätzlich von der aktuellen Fahrsituation und der Haltung des Beifahrers ab.

Aufgrund der schnellen Änderung der Situation muß die OOP-Erkennung mit einer hohen Wiederholrate durchgeführt werden. Die max. Meßrate wird dabei durch die Meßgeschwindigkeit der Entfernungssensoren und der Taktrate der Airbag-Signalverarbeitung vorgegeben.

8.2 Ausweitung der vereinzelt Entfernungen auf eine Fläche

Innerhalb dieses Abschnittes werden verschiedene Möglichkeiten zur Rekonstruktion der fehlenden Tiefeninformationen ausgehend von bekannten Tiefeninformationen und/oder einem flächendeckendem Grauwertbild betrachtet (siehe Abb. 8.7). Entsprechend der

vorangegangenen Beschreibung können auch für diese Anwendungen unterschiedliche Punkt- und Linienstrukturen zur Aufnahme der Tiefeninformation gewählt werden. Die Auswahl hängt dabei nicht nur von der Applikation, sondern auch von dem verwendeten Rekonstruktionsverfahren ab. Nicht in allen Fällen können alle vorhandenen Tiefenwerte genutzt werden. Generell gilt jedoch, je mehr Initialwerte zur Verfügung stehen, desto besser und genauer ist die rekonstruierte Szene im 3D-Raum.

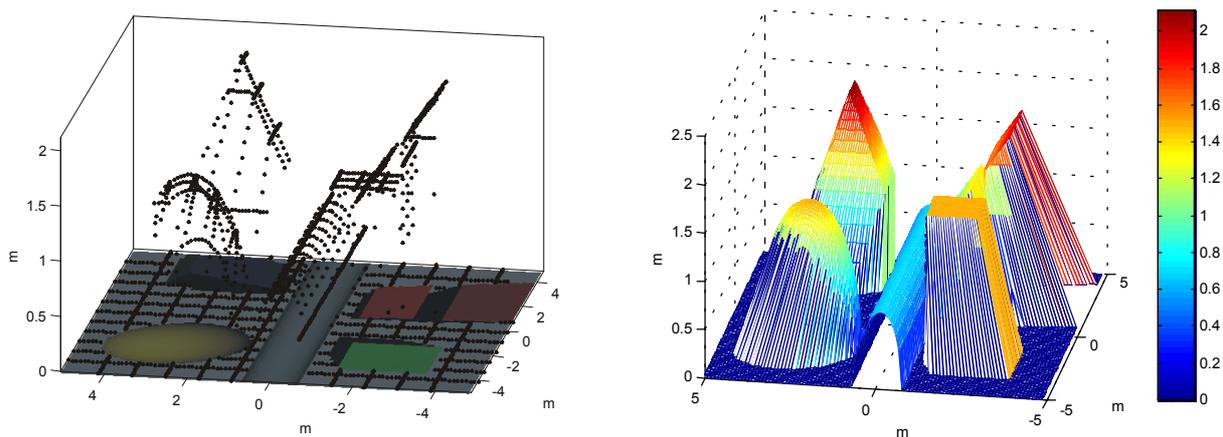


Abb. 8.7: Prinzip der Ausweitung von vereinzelt Entfernungswerten
links: Ausgangsbild; rechts: Rekonstruktionsziel

Eine Hauptrichtung der Rekonstruktion der fehlenden Tiefenwerte aus einem Grauwertbild und den vereinzelt zur Verfügung stehenden Tiefenwerten ist die Fusion auf Pixelebene (siehe Kap. 7.4.1). Auf Basis der bildpunktweisen Übereinstimmung der Datenquellen wird eine Beschreibung der Abhängigkeiten zwischen dem zusammengehörigen Tiefen- und Grauwert benötigt. Dazu wird häufig das Lambert'sche Reflexionsgesetz (siehe Kap. 6.6.3) genutzt. Diese einfache Verknüpfung beschreibt die wahrgenommene Helligkeit in Abhängigkeit der Oberflächennormalen und der Beleuchtungsposition. Die Verfahren auf Basis von Shape-from-Shading sind die Hauptvertreter dieses Grundgedankens und sollen hier hinsichtlich des Rekonstruktionsproblems genauer untersucht werden. Ein weitergehender Ansatz nutzt die Idee der Lambert-Reflexion von der Unabhängigkeit der Beobachterposition. Auf Basis dieses Grundgedankens werden ebenfalls zwei Verfahren vorgestellt.

Neben der Fusion auf Pixelebene stellt die Flächeninterpolation eine weitere Möglichkeit zur Rekonstruktion der fehlenden Tiefenwerte dar. Aufgrund der einfachen und effizienten Vorgehensweise soll dieses Verfahren vor den oben genannten fusionsbasierten Verfahren untersucht werden.

8.2.1 Flächeninterpolation

Mit Hilfe der Flächeninterpolation können mit einfachen Mitteln die Parameter einer vorzugebenden Flächenfunktion anhand vereinzelter Tiefenwerte bestimmt werden. Auf Basis der berechneten Flächenfunktion werden die zu rekonstruierenden Tiefenwerte an allen benötigten Stellen berechnet. Die Rekonstruktion kann dabei wahlweise mit Bezug auf die Bildkoordinaten oder eines konstanten XY-Abstandes durchgeführt werden.

Die Flächeninterpolation erfolgt über die in Abschnitt 6.3 beschriebene Flächenapproximation. Dazu ist eine vorangegangene Segmentierung des Bildes auf Basis der Grauwerten notwendig. Stehen nur wenige Stützstellen innerhalb einer Region zur Verfügung, ist das Ergebnis in Abhängigkeit der Genauigkeit der Stützpunkte nur bedingt brauchbar. Abhilfe kann in diesem Fall die Hinzunahme weiterer Stützwerte, wie z.B. die Flächennormale, schaffen. Dazu wird Gleichung (6.25) durch die Quadrate der Fehler der Normalenableitungen erweitert.

$$\varepsilon^2 = \sum_{(i,j) \in U} (Z(x_{i,j}, y_{i,j}) - z_{i,j})^2 + \sum_{(i,j) \in U'} \left\{ \left(\frac{\partial Z(x_{i,j}, y_{i,j})}{\partial x} - \frac{\partial z_{i,j}}{\partial x_{i,j}} \right)^2 + \left(\frac{\partial Z(x_{i,j}, y_{i,j})}{\partial y} - \frac{\partial z_{i,j}}{\partial y_{i,j}} \right)^2 \right\} \quad (8.1)$$

$\partial z / \partial x$ und $\partial z / \partial y$ stellen die diskreten Flächennormalen dar, welche möglichst genau zur Verfügung gestellt werden müssen. Die Werte an den Randgebieten sollten daher nicht mit in die Minimierungsformel einbezogen werden, da an diesen Stellen der Normalenvektor durch evtl. vorhandene Objektkanten verfälscht werden könnte. Die Umgebung U muß infolgedessen auf die Umgebung U' verkleinert werden (z.B. über die morphologische Dilatation).

Hinsichtlich der Einteilung in Regionen muß auf eine zuverlässige Segmentierung des Grauwertbildes geachtet werden, da fehlerhaft segmentierte Regionen Problemzonen darstellen. Eine Einteilung der Szene in zu viele Bereiche führt zu einer schlechten oder fehlenden Tiefenrekonstruktion aufgrund der geringen Anzahl der für jede Region zur Verfügung stehenden Tiefenwerte. Ebenso führen unterbrochene oder nicht detektierte Kanten zwischen zwei Regionen zu einem Problem bei der Funktionsanpassung. Kanten oder Tiefensprünge zwischen den beiden Regionen können mit der Flächenapproximation nicht rekonstruiert werden.

8.2.1.1 Rekonstruktion einer Halbkugel

Im Rahmen von Untersuchungen wurde die Flächeninterpolation in dem MATLAB-Programm ‚*fapprox.m*‘ implementiert. Ein Ergebnis der Flächeninterpolation, angewendet auf eine synthetische Halbkugel, ist in der Abb. 8.8 dargestellt. Als Gitter wurde das Punktgitter mit 11 x 11 Punkten, gleichmäßig verteilt über die gesamte Szene, gewählt. Die zu approximierende Funktion ist die biquadratische Funktion.

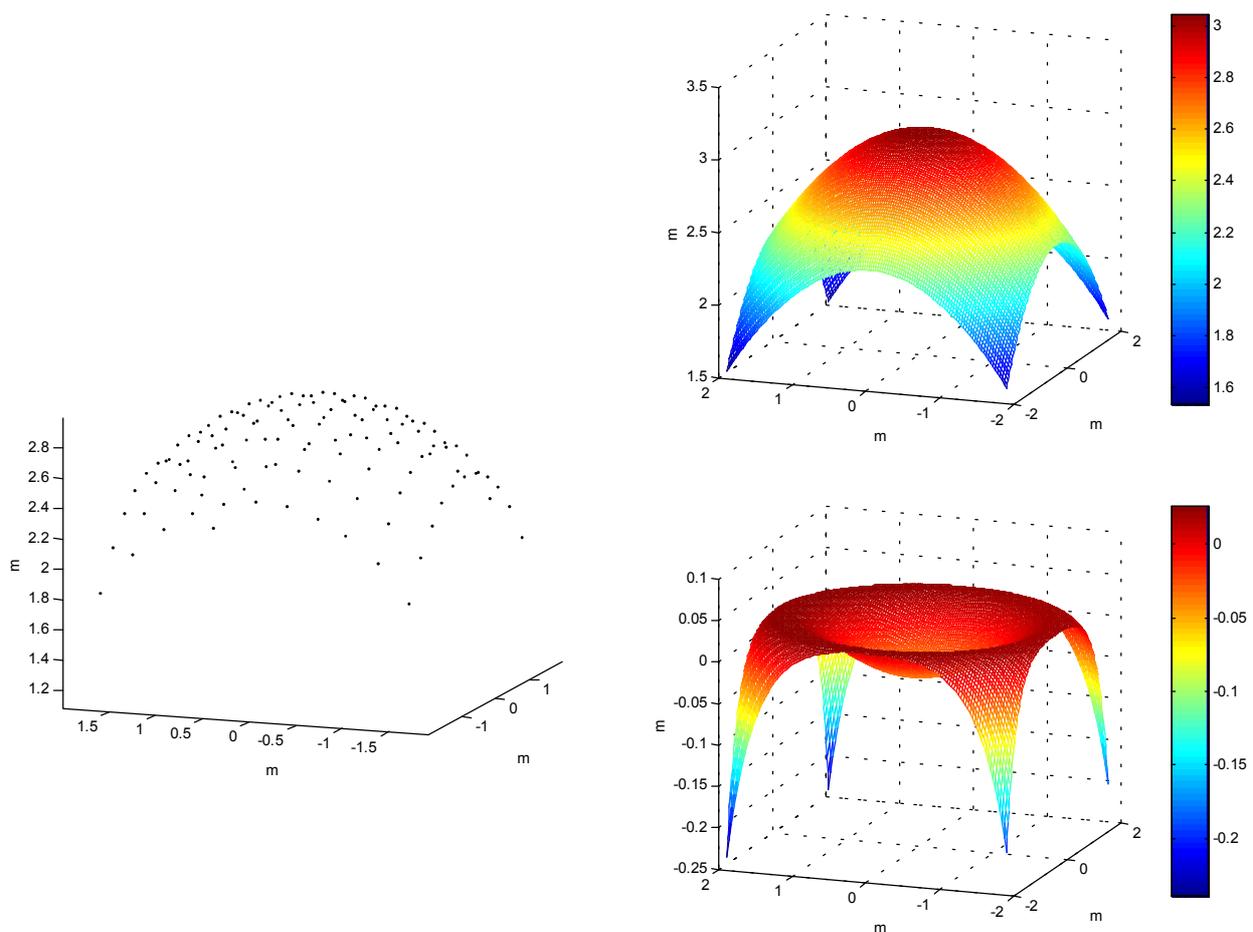


Abb. 8.8: Ausweitung von vereinzelt Tiefenwerten mit Hilfe der Flächenapproximation
links: Ausgangsbild (Punktgitter mit 11 x 11 Punkten);
rechts oben: rekonstruiertes Tiefenbild; rechts unten: absoluter Rekonstruktionsfehler

Das Rekonstruktionsergebnis der Halbkugel (oben rechts) sieht auf den ersten Blick sehr gut aus. Erst eine Interpretation des absoluten Fehlers (unten rechts) zeigt die Problemstellen der Flächenapproximation. Aufgrund der Anpassung der gegebenen Stützpunkte an eine biquadratische Funktion weist der absolute Fehler eine Trichterfunktion auf. Ebenso wächst der Fehler an den Ecken des Bildes sehr stark an. Zur Verbesserung der Rekonstruktion ist es daher ratsam, eine Flächenfunktion höheren Grades zu wählen. Das

eigentliche Problem hinsichtlich der fehlerhaften Annahme der Flächenfunktion bleibt jedoch bestehen. Mit der Erhöhung des Grades wächst gleichermaßen die Anzahl der notwendigen Stützpunkte.

8.2.1.2 Rekonstruktion einer komplexen Szene

Als weiteres Beispiel für die Rekonstruktionseigenschaft der Flächenapproximation wurde eine real aufgenommene Szene gewählt (siehe Abb. 8.9 links oben) aus der Tiefenwerte in Anordnung des Liniengitters mit 10×10 Linien als Eingabedaten für den Rekonstruktionsprozeß entnommen wurden (siehe Abb. 8.9 links unten). Die zu approximierende Funktion ist auch hier die biquadratische Funktion.

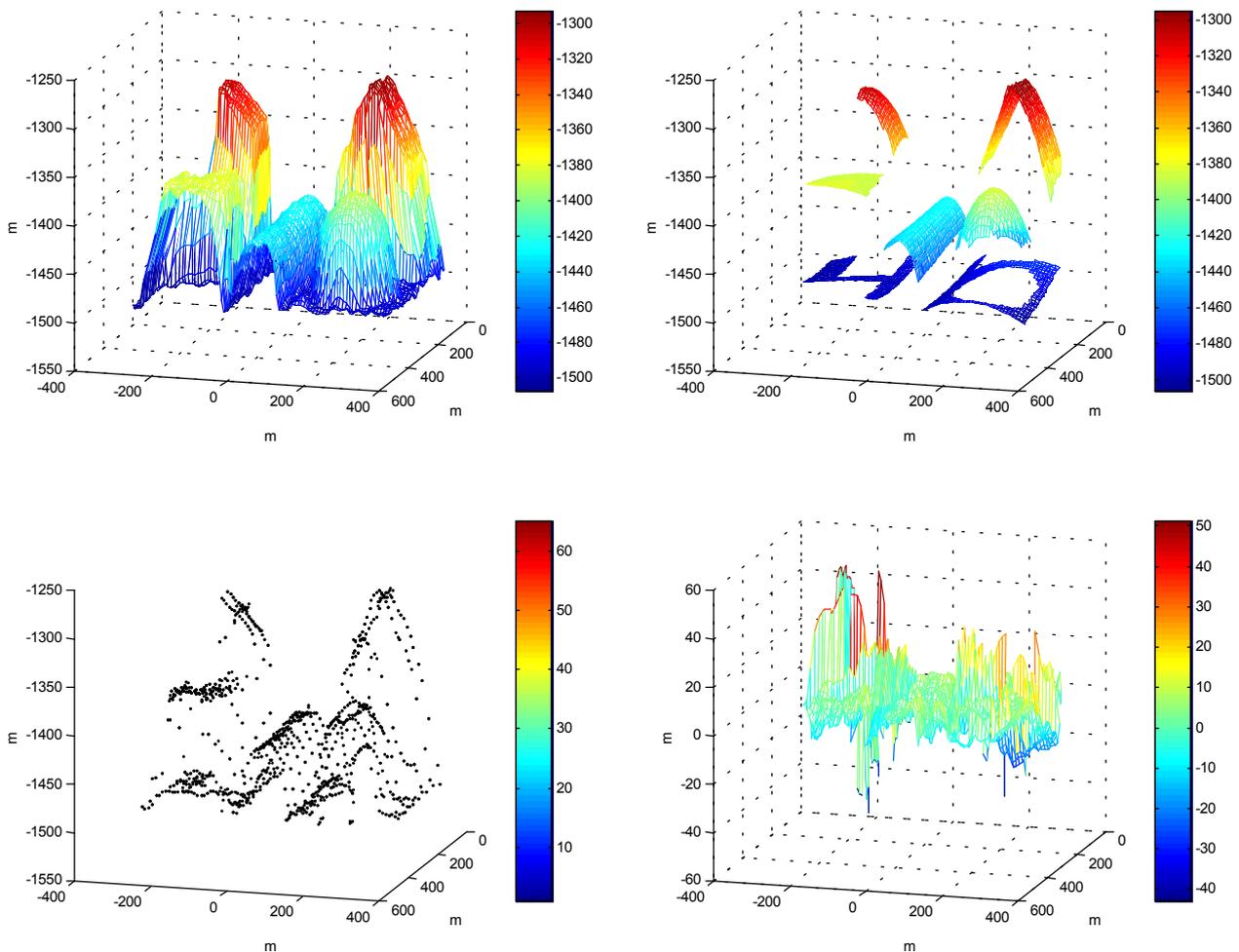


Abb. 8.9: Rekonstruktion einer realen Szene (siehe Abb. 3.1);
 links oben: Flächendeckendes Ausgangsbild (für Vergleich);
 links unten: Eingabedaten der Flächenapproximation (10×10 Liniengitter);
 rechts oben: rekonstruiertes Tiefenbild;
 rechts unten: absoluter Rekonstruktionsfehler

In Abb. 8.9 rechts oben ist die rekonstruierte Szene dargestellt. Bis auf zu kleine oder seitliche Flächen, an denen nicht genügend Stützpunkte für die Flächenapproximation vorliegen, wurden alle Flächen rekonstruiert. Der durchschnittliche Höhenfehler (rechts unten) zwischen dem Ausgangsbild und der rekonstruierten Szene liegt bei 8%.

Bei näherer Betrachtung des Rekonstruktionsergebnisses fallen die unebenen Bodenflächen und das Rechteck vorne links auf. Jedoch liegen diese unter Berücksichtigung des Fehlerbildes noch gut im Toleranzbereich. Eine Verbesserung kann erreicht werden, wenn für solche Flächen anstatt der biquadratischen eine bilineare Funktion genutzt wird.

Ein Nebeneffekt der Flächenapproximation ist die Rauschunterdrückung, wie aus dem Vergleich zwischen Ausgangsbild und rekonstruiertem Ergebnis sichtbar wird. Die Eingangsdaten brauchen daher nicht vorbearbeitet zu werden.

Weitere Untersuchungen mit anderen Szenen ergaben für technische Flächen, wie sie hier untersucht wurden, ein ähnliches Ergebnis. Natürliche Flächen besitzen gegenüber den technischen Flächen eine wesentlich höhere Ordnung, falls sie überhaupt mathematisch beschreibbar sind. Damit lassen sich solche Szenen mit diesem Verfahren nicht rekonstruieren.

8.2.2 Shape-from-Shading

Das SFS-Verfahren stellt auf den ersten Blick aufgrund seiner Zielsetzung, aus einem vorhandenen 2D-Grauwertbild (entsprechend der menschlichen Vorgehensweise) ein Höhen- bzw. Orientierungsbild zu rekonstruieren, eine effektive Lösung der hier gestellten Problematik dar. Die zusätzliche Möglichkeit der Nutzung von Tiefenwerten an singulären Stellen als Initialisierungswerte für den Rekonstruktionsprozeß deckt sich mit dem 2D/3D-Kamerakzept.

Seit den ersten Veröffentlichungen von SFS-Verfahren wurden viele Verbesserungen von bestehenden, aber auch neue SFS-Verfahren vorgestellt. Aus den Veröffentlichungen ist die Qualität der Verfahren nur sehr schwer erkennbar. Jeder Autor benutzt für den Test seines Verfahrens nur Bilder, die der Aufgabenstellung entsprechen bzw. den optimalen Einsatz seines Verfahrens erlauben. Insbesondere werden innerhalb der Untersuchungen hauptsächlich synthetisch generierte Bilder verwendet und nur wenige reale Bilder. Komplexere Szenen kommen so gut wie nicht vor.

Für die Nutzung der SFS-Verfahren hinsichtlich der 2D-BV mit vereinzelt Tiefenwerten können anhand der Beschreibungen der Autoren folglich keine SFS-Verfahren direkt ausgewählt werden. Somit sind eigene Untersuchungen betreffend der Nutzbarkeit der SFS-Verfahren für das 2D/3D-Kamerakonzept notwendig [JuHa99a]. Zur Auswahl eines Verfahrens aus der Vielzahl der veröffentlichten Publikationen wurden daher verschiedene Kriterien ausgewählt, welche entsprechend ihrer Wichtigkeit für die Nutzung im 2D/3D-Kamerakonzept geordnet sind:

- 1) Geschwindigkeit
- 2) Nutzung von (Tiefen)-Vorgabewerten
- 3) Empfindlichkeit gegenüber Rauschen
- 4) Genauigkeit
- 5) Verhalten an Unstetigkeiten

Innerhalb dieser Aufzählung spielen die ersten beiden Punkte eine wesentliche Rolle. Neben der Nutzung von (Tiefen)-Vorgabewerten ist die Verarbeitungsgeschwindigkeit ein wichtiges Auswahlkriterium, da eine Verarbeitung in Videoechtzeit angestrebt wird. Globale Ansätze kommen selbst bei der heutigen CPU Rechenleistung nicht an die Sekundengrenze heran. Dementsprechend wurde die Auswahl der Verfahren auf die lokalen und die Propagationsverfahren beschränkt. Unter diesen wurden in Abhängigkeit der Aussagen der Autoren über die Zuverlässigkeit, die Nutzung von Vorgabewerten und die Geschwindigkeit folgende drei Verfahren ausgewählt, die nachfolgend genauer untersucht werden:

- Ø Ansatz von Bichsel & Pentland: ‚A simple algorithm for Shape from Shading‘ [BiPe92]
- Ø Ansatz von Pentland: ‚Linear Shape-from-Shading‘ [Pe88]
- Ø Ansatz von Tsai & Shah: ‚A fast linear Shape from Shading‘ [TsSh92]

Nur die Verfahren von Bichsel & Pentland und Pentland nutzen die Vorgabe von Tiefenwerten. Das Verfahren von Tsai & Shah wurde aufgrund der Aussage über die Geschwindigkeit und Zuverlässigkeit ausgewählt.

Für eine einfachere Bewertung und einen direkten Vergleich der verschiedenen Rekonstruktionsergebnisse wurde bei allen Verfahren eine reale und eine synthetisch generierte Halbkugel verwendet. Zur Beurteilung der Rekonstruktionseigenschaften für komplexe

Szenen wurden zudem mehrere reale Aufnahmen untersucht. *Aufgrund der unterschiedlichen Rekonstruktionseigenschaften jedes einzelnen SFS-Verfahrens wurde im Rahmen der Vorstellung repräsentativ diejenige reale Szene ausgewählt, mit der das jeweilige SFS-Verfahren am besten arbeitet. Im Anschluß an die Beschreibung der einzelnen Verfahren sind die Ergebnisse aller Verfahren für einen besseren Vergleich anhand einer kleinen Auswahl von Bildern nochmals zusammengefaßt dargestellt (siehe 8.2.2.5).*

8.2.2.1 Voraussetzungen

Grundvoraussetzungen für jedes SFS-Verfahren sind homogene Grauwertregionen. Eine Änderung der Schattierung bedeutet eine Änderung des Albedos (siehe Kap. 6.7). Da dieser Wert innerhalb der Regionen als konstant angenommen wird, ist eine Segmentierung des Bildes auf Basis des Grauwertbildes in homogene Regionen notwendig. Texturierte Oberflächen lassen sich folglich nicht mit dem SFS-Ansatz bearbeiten. Neben dieser allgemeingültigen Einschränkung gelten in Abhängigkeit des gewählten Verfahrens weitere Einschränkungen, welche hier nochmals (siehe Kap. 6.7) aufgezählt werden sollen:

- ∅ Lichtquelle und Kamera befinden sich in unendlich weiter Entfernung
- ∅ Die Position der Lichtquelle ist bekannt
- ∅ Es ist nur eine Lichtquelle erlaubt
- ∅ Bekannte Reflexionseigenschaft der Oberfläche

8.2.2.2 SFS-Ansatz von Bichsel & Pentland

Aufbauend auf den Ansatz von Oliensis & Dupuis [OI91] [OIDu91] entwickelten Bichsel & Pentland [BiPe92] [PeBi94] ein effizientes Verfahren, welches den Tiefenwert eines betrachteten Bildpunktes aus einer kleinen Nachbarschaftsumgebung unter der Nutzung eines Grauwertbildes und Tiefenwerten an singulären Stellen des Grauwertbildes berechnet (siehe Anhang: Pseudo-Programme). Dazu bestimmt das Verfahren zunächst in einem Vorverarbeitungsschritt die möglichen Oberflächensteigungen in jedem 45° Segment des Bildes. Aus den darin im allgemeinen enthaltenen zwei Extremwerten wird diejenige Steigung ausgewählt, bei der sich die Oberfläche in inverser Beleuchtungsrichtung entwickelt. Zur Vermeidung von unendlichen Steigungen in Umgebungen kleiner Irradianzen wird das Grauwertbild und die vorhandenen Tiefenwerte dazu derart um einen Winkel α rotiert, das eine der diskreten Richtungen mit dem Azimut der Beleuchtungsrichtung möglichst gut übereinstimmt.

Im Anschluß an diesen Vorverarbeitungsschritt folgt die eigentliche Tiefenbestimmung. Dazu wird in jedem Schritt des iterativen Prozesses vorausgesetzt, daß die Oberfläche bereits bis zu einem gewissen Grad rekonstruiert worden ist. Zur Berechnung eines neuen Tiefenwertes wird geprüft, mit welcher der acht zuvor berechneten Steigungen die Oberfläche am stärksten in Beleuchtungsrichtung propagiert wird. Aus der Addition von diesem und dem alten Tiefenwert ergibt sich der neue Tiefenwert. Nach der Berechnung des Tiefenbildes muß dieses aufgrund der anfänglichen Rotation des Grauwertbildes wieder zurückgedreht werden.

Der hier vorgestellte Algorithmus von Bichsel & Pentland wurde in dem MATLAB-Programm *,bichsel.m'* implementiert. Für die Darstellung des Rekonstruktionsergebnisses wurden jeweils eine reale und synthetische Halbkugel und eine reale Szene ausgewählt. Innerhalb der synthetisch generierten Halbkugel stehen zu allen Grauwertpunkten die dazugehörigen Tiefenwerte zur Verfügung. Daher konnten an allen singulären Stellen des Grauwertbildes die korrespondierenden Tiefenwerte für die Initialisierung genutzt werden. Im Gegensatz dazu lagen keine Tiefenwerte für die Initialisierung der realen Bilder vor. Bichsel & Pentland schlagen in diesem Fall vor, alle singulären Stellen mit einem konstanten Wert zu belegen.

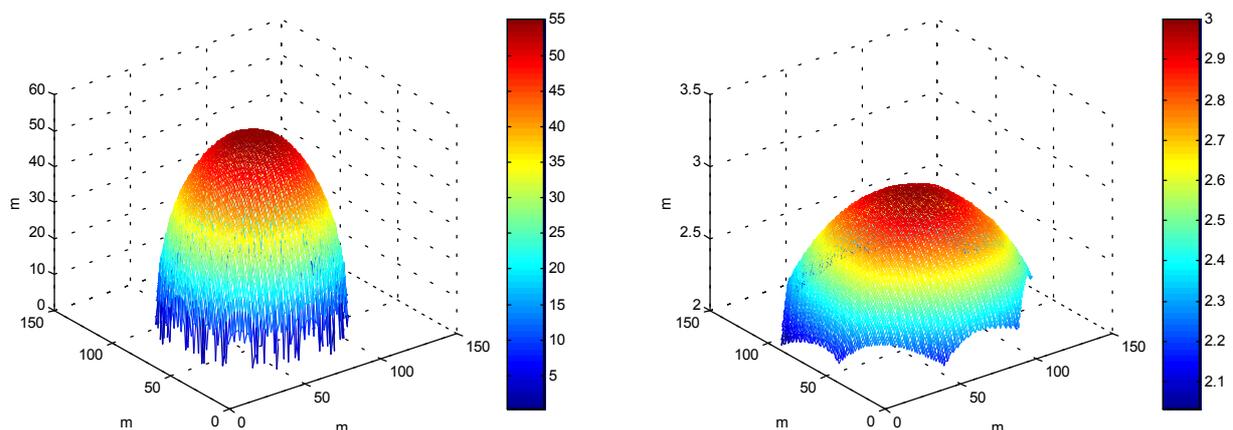


Abb. 8.10: Links: Rekonstruktion einer realen Halbkugel (keine Vorgabewerte);
Rechts: Rekonstruktion einer synthetisch generierten Halbkugel (Nur eine singuläre Stelle vorhanden)

In der linken Abbildung von Abb. 8.10 ist die Rekonstruktion der realen Halbkugel dargestellt. Die Anzahl der Iterationsschritte beträgt hier, wie bei allen anderen hier vorgestellten Rekonstruktionen, vier Schritte. Ein Fehler kann aufgrund der fehlenden Tiefenwerte nicht angegeben werden.

Die Halbkugel auf der rechten Seite von Abb. 8.10 wurde aus einem synthetisch generierten Bild erzeugt. An den Rändern der Rekonstruktion kann ein Problem des Ansatzes sehr gut erkannt werden. Durch die notwendige Hin- und Rückrotation des Bildes wird in Abhängigkeit des Winkels ein Teilgebiet abgeschnitten. Aufgrund von nur einem vorhandenen singulären Punkt der Halbkugel wird das Rekonstruktionsergebnis durch diese Vorgabe nicht beeinflusst. Dieser eine Wert wirkt sich nur auf den Offset aus. Die Steigung bleibt unabhängig vom Vorgabewert konstant. Wie bei den meisten SFS-Verfahren stimmen ungeachtet der Initialwerte der Offset und die Amplitude des rekonstruierten Bildes nicht mit dem Vorgabebild überein. Für eine Fehlerbestimmung wurde daher das rekonstruierte Bild an die Vorgabe angepaßt. Der absolute Fehler liegt trotz der nur vier Iterationsschritte unter ± 0.15 m (siehe Abb. 8.17).

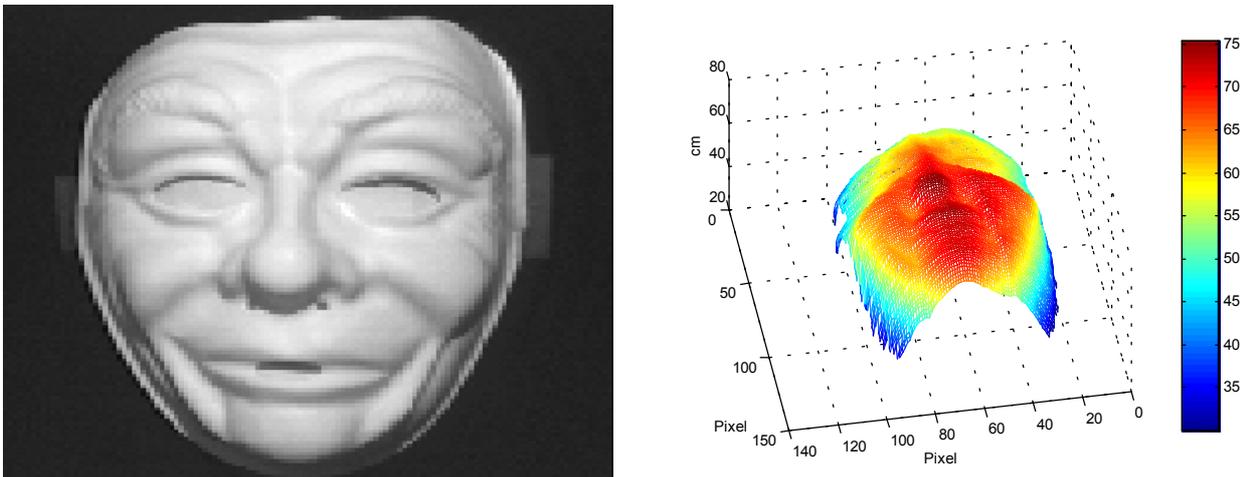


Abb. 8.11: Links: Darstellung des realen Ausgangsbildes (Mask) [FtpUcf99];
Rechts: Rekonstruiertes Tiefenbild der linken Abbildung (keine Nutzung von Tiefenvorgabewerten)

Aufbauend auf die guten Rekonstruktionsergebnisse der Halbkugel wurden weitere Tests an realen Bildern durchgeführt. Dabei lieferte keines der Testbilder, mit Ausnahme des Bildes aus Abb. 8.11, ein brauchbares Ergebnis. Nur der vordere Bereich des Tiefenbildes (rechts) kann schemenhaft im Originalbild (links) erkannt werden. Der gesamte hintere Bereich wurde unkenntlich rekonstruiert.

Weiterführende Untersuchungen an verschiedenen Szenen mit z.B. einer größeren Iterationszahl brachten keine wesentlichen Verbesserungen des Rekonstruktionsergebnisses. Ebenso schlugen Untersuchungen auf Basis von zusätzlichen Initialwerten unabhängig der singulären Stellen fehl. Auf Grundlage der synthetisch generierten Halbkugel wurden alle Tiefenwerte der jeweiligen Gitteranordnungen als Initialwerte für die Rekonstruktion verwendet. Neben einem deutlich schlechteren Rekonstruktionsergebnis

wurde kein vollständiges Bild mehr rekonstruiert. Das Ergebnis wies, im Gegensatz zur Vorgabe von Tiefenwerten, einzig an singulären Stellen Lücken innerhalb des gesamten rekonstruierten Tiefenbildes auf. Die vereinzelt Tiefenwerte aus einer beliebigen Gitteranordnung können somit nicht vollständig genutzt werden. Nur wenn singuläre Punkte mit den einzelnen Elementen der Gitteranordnung übereinstimmen, können diese verwendet werden.

Als Vorgabewert für die Beleuchtung nutzt dieser Algorithmus wie fast alle SFS-Verfahren die Deklination und den Azimut, d.h. er arbeitet nur mit weit entfernten Beleuchtungsquellen. Untersuchungen auf Basis der Beleuchtungsrichtung ergaben, daß dieser Algorithmus stark von der Beleuchtungsrichtung abhängig ist. Die besten Rekonstruktionseigenschaften wies der Algorithmus auf, wenn die Beleuchtung von der Seite erfolgte.

8.2.2.3 SFS-Ansatz von Pentland

In dem SFS-Ansatz von Pentland [Pe88],[Pe90] wird ein Tiefenbild auf Basis eines Grauwertbildes und ggf. von Tiefeninitialwerten und Randbedingungen unter Voraussetzung einer linearen Approximation der Reflektanzfunktion in p und q rekonstruiert. Dazu wendet Pentland die Taylorreihenentwicklung auf die Reflektanzfunktion unter Vernachlässigung der höherwertigen Terme und Berechnung in der Umgebung $p=p_0$ und $q=q_0$ an:

$$I(x, y) \approx R(p_0, q_0) + (p - p_0) \frac{\partial R(p, q)}{\partial p} \Big|_{p=p_0, q=q_0} + (q - q_0) \frac{\partial R(p, q)}{\partial q} \Big|_{p=p_0, q=q_0} \quad (8.2)$$

Innerhalb eines kleinen Bereiches auf einer glatten Oberfläche um p_0 und q_0 ist diese Approximation ausreichend genau. Darüber hinaus vergrößert sich dieser Bereich mit größer werdendem Beleuchtungswinkel. Demzufolge gilt für eine begrenzte Region, daß die echte Reflektanzfunktion durch eine lineare Reflektanzfunktion genau genug approximiert werden kann. Damit werden nach Pentland Verfahren zur Lösung der SFS-Problematik ermöglicht, welche effizienter und allgemeiner sind als solche Techniken, die das Problem mit dem globalen Ansatz lösen.

Als Grundlage für die Reflektanzfunktion nutzt Pentland, wie viele seiner Kollegen, die Lambert-Reflexion. Unter Angabe der Beleuchtungsrichtung als Einheitsnormale $(\mathbf{L}_e = (x_{Le}, y_{Le}, z_{Le}) = (\cos(\theta)\sin(\sigma), \sin(\theta)\sin(\sigma), \cos(\sigma)))$ sieht diese wie folgt aus:

$$R(p, q) = \frac{-x_{Le}p - y_{Le}q + z_{Le}}{\sqrt{1+p^2+q^2}} = \frac{-\cos(\theta)\sin(\sigma)p - \sin(\theta)\sin(\sigma)q + \cos(\sigma)}{\sqrt{1+p^2+q^2}} \quad (8.3)$$

Die lineare Approximation (8.2) dieser Reflektanzfunktion für $p_0=q_0=0$ ergibt:

$$I(x, y) \approx \cos(\sigma) + p \cos(\theta) \sin(\sigma) + q \sin(\theta) \sin(\sigma) \quad (8.4)$$

Durch eine Fouriertransformation dieser Gleichung wird die Lösung des Problems stark vereinfacht. Dieses läßt sich am einfachsten über die Transformation von p und q beschreiben. Aufgrund der partiellen Ableitung von $z(x,y)$ nach x bzw. y ergeben sich für p und q folgende Fouriertransformationen:

$$F_p(f, \varphi) = 2\pi \cos(\varphi) \cdot f \cdot e^{i\pi/2} F_z(f, \varphi) \quad (8.5)$$

$$F_q(f, \varphi) = 2\pi \sin(\varphi) \cdot f \cdot e^{i\pi/2} F_z(f, \varphi) \quad (8.6)$$

$F_z(f, \varphi)$ stellt die komplexe Fouriertransformation von $z(x,y)$ dar und f und φ die radiale Frequenz bzw. die Orientierung. Die nicht zu interpretierenden Terme $\cos(\varphi)$ bzw. $\sin(\varphi)$ sind nur in der Ausarbeitung von Pentland aufgeführt, werden dort aber nicht weiter erklärt. In weiterführenden Arbeiten zu diesem Ansatz sind diese Terme nicht vorhanden.

Auf Basis der Transformationen (8.4) und (8.5), eingesetzt in die Fouriertransformierte Gleichung (8.4), ergibt sich unter Vernachlässigung des ersten Terms der rechten Seite für die Fouriertransformierte des Grauwertbildes F_I folgende Gleichung [Pe88]:

$$F_I(f, \varphi) = F_z(f, \varphi) \cdot 2\pi f \cdot e^{i\pi/2} [\cos(\theta) \sin(\sigma) \cos(\varphi) + \sin(\theta) \sin(\sigma) \sin(\varphi)] \quad (8.7)$$

Mit dieser Formel wird die Abhängigkeit des Grauwertbildes vom Tiefenbild im Fourierraum beschrieben. Die gesuchte Oberflächenkontur ergibt sich demnach aus der inversen Fouriertransformation von F_z . Unter Hinzunahme von Initialwerten und Randbedingungen kann weiterhin das Ergebnis an schwer rekonstruierbaren Stellen verbessert werden. Auch ohne diese Vorgabe werden laut Pentland [Pe90] gute Rekonstruktionsergebnisse erwartet (siehe Anhang: Pseudo-Programme).

Für eine Bewertung dieses Ansatzes wurde die Gleichung (8.7) in dem MATLAB-Programm ‚pentland.m‘ implementiert. Dazu wird innerhalb dieses Programms anfangs das gefilterte und normierte Grauwertbild in den Fourierraum transformiert. In einer zweidimensionalen Schleife wird anschließend die genannte Gleichung auf die Transformation angewendet. Das Tiefenbild ergibt sich in einem letzten Schritt aus der Rücktransformation in den Ortsbereich. Für die Untersuchung der Auswirkung der sin/cos-Terme auf die Gleichung wurde jede Simulation sowohl mit als auch ohne diese Terme durchgeführt.

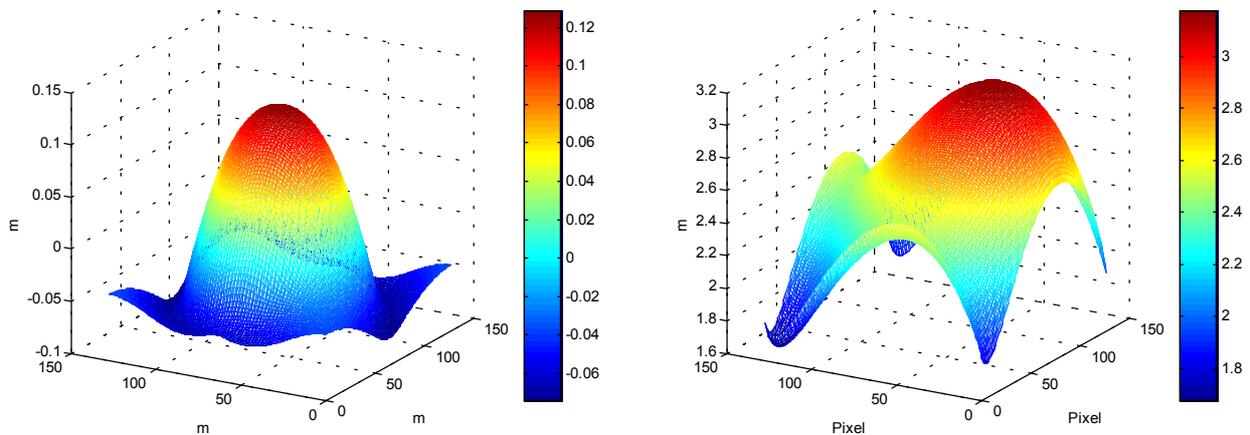


Abb. 8.12: Links: Rekonstruktion einer realen Halbkugel;
Rechts: Rekonstruktion einer synthetisch generierten Halbkugel

Innerhalb des linken Bildes von Abb. 8.12 ist das Rekonstruktionsergebnis der realen Halbkugel abgebildet. Die Qualität der Rekonstruktion ist überraschenderweise besser als das Ergebnis des synthetisch generierten Bildes (rechts). Ungeachtet der Darstellung wurde auch das linke Bild nicht perfekt rekonstruiert. In dieser Ansicht kann die Verzerrung des Bildes nur sehr schwer erkannt werden. Da für die realen Daten kein korrespondierendes Tiefenbild vorliegt, können keine Fehlerangaben für dieses Bild angegeben werden. Eine Fehlerangabe aus dem synthetisch generierten Bild ist aufgrund der fehlgeschlagenen Rekonstruktion nicht sinnvoll (siehe Abb. 8.17).

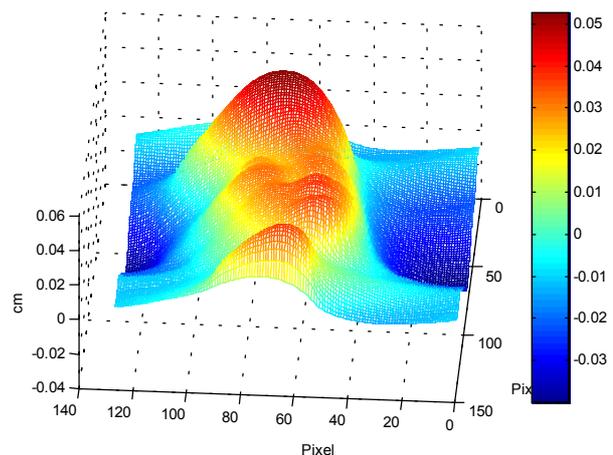


Abb. 8.13: Links: Darstellung des realen Ausgangsbildes (Manneq);
Rechts: Rekonstruiertes Tiefenbild der linken Abbildung [FtpUcf99]

Unabhängig von diesem schlechten Ergebnis wurden weitere Untersuchungen an realen Bildern durchgeführt. In Abb. 8.13 ist das Resultat einer solchen Rekonstruktion (rechts) mit dem dazugehörigen Grauwertbild (links) abgebildet. Das Ergebnis ist wie bei vielen anderen untersuchten Bildern nur bedingt gebrauchsfähig. Aus dem Tiefenbild kann nur schemenhaft das Ausgangsbild erkannt werden.

In keinem der getesteten Bilder konnte ein gravierender Unterschied zwischen der Nutzung bzw. Vernachlässigung der sin/cos-Terme festgestellt werden. Das Rekonstruktionsergebnis ist immer gleich ‚schlecht‘. Untersuchungen auf Basis von unterschiedlichen Beleuchtungspositionen brachten ebenso keine Verbesserung des Rekonstruktionsergebnisses.

Auf Grundlage dieser schlechten Ergebnisse wurden keine weiterführenden Untersuchungen unter Nutzung von Tiefenwerten als Initialwerte durchgeführt. Diese versprachen bei den getätigten Untersuchungen keine Verbesserung des Ergebnisses.

In den Ergebnissen aus der Arbeit von Pentland wurden im Gegensatz zu den hier getätigten Untersuchungen auch bei komplexeren Bildern (Gesicht, Relief, Münze) sehr gute Ergebnisse erzielt. Ein Unterschied kann in der Wahl des Bildmaterials liegen. Auf Grundlage der linearen Approximation können nur lineare Flächen gut rekonstruiert werden. Bei Ausnutzung dieser Eigenschaft wird laut Pentland ein gutes Ergebnis erzielt.

8.2.2.4 SFS-Ansatz von Tsai & Shah

Der SFS-Ansatz von Tsai & Shah [TsSh92], [TsSh94] benötigt als Vorgabewert für die Rekonstruktion nur das Grauwertbild. Im Gegensatz zur Methode von Pentland, welche die Reflektanzfunktion nach p und q diskret approximiert, approximieren Tsai & Shah p und q diskret nach z und linearisieren die Reflektanzfunktion. Als Reflektanzfunktion nutzen sie ebenfalls die Lambert-Reflexion. Zur Vereinfachung der nachfolgenden Schritte wird die

Beleuchtungsrichtung in Abhängigkeit von $p_s = \frac{\cos(\theta) \sin(\sigma)}{\cos(\sigma)}$ und $q_s = \frac{\sin(\theta) \sin(\sigma)}{\cos(\sigma)}$

angegeben:

$$\begin{aligned}
 I(x, y) &= R(p, q) & (8.8) \\
 &= \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2} \sqrt{1 + p_s^2 + q_s^2}}
 \end{aligned}$$

Unter Nutzung der diskreten Approximation für $p=Z(x,y)-Z(x-1,y)$ und $q=Z(x,y)-Z(x,y-1)$ kann die Reflektanzfunktion wie folgt umgeschrieben werden:

$$\begin{aligned} 0 &= f(I(x, y), Z(x, y), Z(x-1, y), Z(x, y-1)) \\ &= I(x, y) - R((x, y) - Z(x-1, y), Z(x, y) - Z(x, y-1)) \end{aligned} \quad (8.9)$$

Für einen bestimmten Punkt (x,y) und ein gegebenes Grauwertbild I ergibt die lineare Approximation (Taylorreihenentwicklung bis zum ersten Grad) folgende Gleichung:

$$\begin{aligned} 0 &= f(I(x, y), Z(x, y), Z(x-1, y), Z(x, y-1)) \\ 0 &\approx f(I(x, y), Z^{n-1}(x, y), Z^{n-1}(x-1, y), Z^{n-1}(x, y-1)) \\ &+ (Z(x, y) - Z^{n-1}(x, y)) \frac{\partial}{\partial Z(x, y)} f(I(x, y), Z^{n-1}(x, y), Z^{n-1}(x-1, y), Z^{n-1}(x, y-1)) \\ &+ (Z(x-1, y) - Z^{n-1}(x-1, y)) \frac{\partial}{\partial Z(x-1, y)} f(I(x, y), Z^{n-1}(x, y), Z^{n-1}(x-1, y), Z^{n-1}(x, y-1)) \\ &+ (Z(x, y-1) - Z^{n-1}(x, y-1)) \frac{\partial}{\partial Z(x, y-1)} f(I(x, y), Z^{n-1}(x, y), Z^{n-1}(x-1, y), Z^{n-1}(x, y-1)) \end{aligned} \quad (8.10)$$

Für ein $m \times m$ Bild existieren m^2 solcher linearen Gleichungen. Zur Lösung dieses Gleichungssystems wird von Tsai & Shah das iterative Verfahren nach Jacobi vorgeschlagen.

Dabei ergibt sich folgende einfache Gleichung:

$$\begin{aligned} 0 &= f(Z(x, y)) \\ &\approx f(Z^{n-1}(x, y)) + (Z(x, y) - Z^{n-1}(x, y)) \frac{d}{dZ(x, y)} f(Z^{n-1}(x, y)) \end{aligned} \quad (8.11)$$

Für $Z(x,y)=Z^n(x,y)$ kann die Tiefenmatrix bezüglich der n-ten Iteration wie folgt bestimmt werden:

$$Z^n(x, y) = Z^{n-1}(x, y) + \frac{-f(Z^{n-1}(x, y))}{\frac{d}{dZ(x, y)} f(Z^{n-1}(x, y))} \quad (8.12)$$

wobei der Nenner folgendermaßen definiert wird:

$$\frac{df(Z^{n-1}(x, y))}{dZ(x, y)} = \frac{(p+q)(pp_s + qq_s + 1)}{\sqrt{(1+p^2+q^2)^3} \sqrt{1+p_s^2+q_s^2}} - \frac{p_s+q_s}{\sqrt{1+p^2+q^2} \sqrt{1+p_s^2+q_s^2}} \quad (8.13)$$

Unter Nutzung von Gleichung (8.12) und (8.13) ergibt sich iterativ die gesuchte Tiefenmatrix, wobei bereits 5 bis 10 Iterationen ausreichen. Als Startbedingung für die Iteration werden alle Entfernungswerte $Z^0(x,y)$ auf Null gesetzt. Abschließend sollte das Ergebnis nach Aussage von Tsai & Shah gefiltert werden (siehe Anhang: Pseudo-Programme).

Zur Bestimmung der Tiefenmatrix wurde das MATLAB-Programm ‚tsai.m‘ entwickelt. Dieses berechnet iterativ in einer zweidimensionalen Schleife alle Matrixelemente auf Basis der Gleichung (8.12). Zur Verbesserung des Rekonstruktionsergebnisses wurde neben dem resultierenden Tiefenbild auch das Eingangsbild gefiltert. Wenn alle zusammengehörigen Berechnungen als eine Zeile angesehen werden, besteht das gesamte Programm insgesamt aus ca. 10 Zeilen. Aufgrund der ‚relativen‘ Unabhängigkeit des iterativen Berechnungsschrittes von der Nachbarschaft kann dieser Algorithmus sehr leicht auf ein paralleles Computersystem angepaßt werden.

Entsprechend der beiden vorangegangenen Untersuchungen soll auch hier die Qualität dieses Verfahrens anhand der beiden Halbkugeln und hier zusätzlich anhand von zwei realen Bildern besprochen werden.

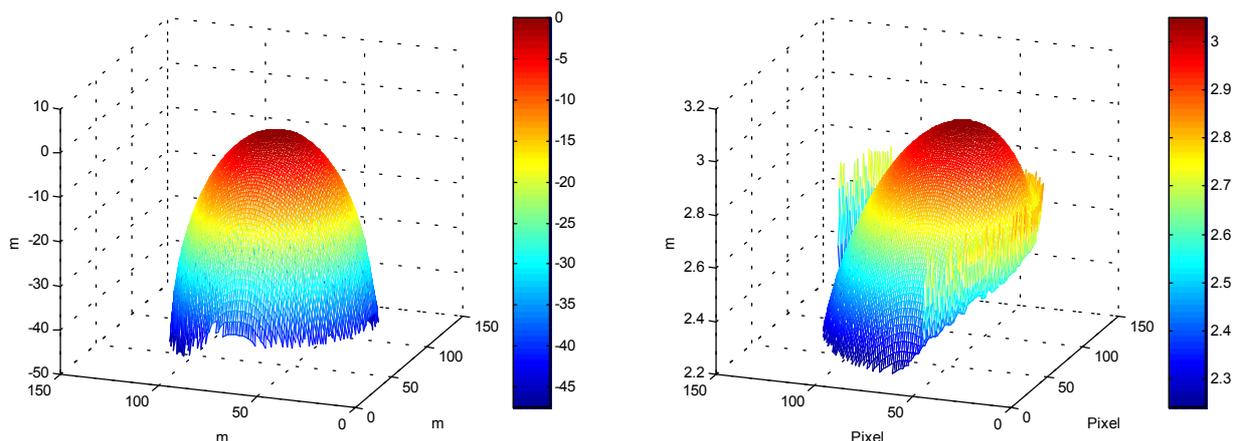


Abb. 8.14: Links: Rekonstruktion einer realen Halbkugel;
Rechts: Rekonstruktion einer synthetisch generierten Halbkugel

Wie bei allen vorangegangenen Ansätzen wurde die reale Halbkugel sehr gut rekonstruiert. Auch die synthetisch generierte Halbkugel sieht auf den ersten Blick, mit Ausnahme des zerfransten Randes sehr gut aus. Nähere Betrachtungen zeigen jedoch, daß der höchste Punkt der Kugel nicht mit der Vorgabe übereinstimmt. Innerhalb dieser Rekonstruktion wurde der aus der Beleuchtungsposition resultierende hellste Punkt ebenso zum höchsten Punkt der Rekonstruktion. Damit ist eine Fehlerangabe auf Basis dieser Rekonstruktion nur unter Vorbehalt möglich. Die Differenz von der Vorgabe der Rekonstruktion ergibt wiederum eine Halbkugel (siehe Abb. 8.17). Die Fläche selbst ist sehr glatt, was auf einen kleinen Fehler hinweist. Die Wölbung der Fehlerkugel weist jedoch zum Rand hin einen Fehler von 35% auf.

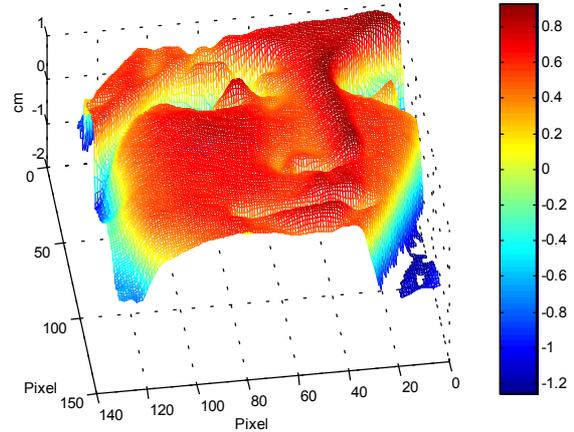


Abb. 8.15: Links: Darstellung des realen Ausgangsbildes (Lenna);
Rechts: Rekonstruiertes Tiefenbild der linken Abbildung [FtpUcf99]

Als erste Testszene für ein reales Bild wurde Abb. 8.15 ausgewählt. Nur mit diesem SFS-Verfahren wurde dieses Bild einigermaßen kenntlich rekonstruiert. Alle anderen Verfahren versagten mit dieser Vorgabe. Bei näherer Betrachtung des Tiefenbildes (rechts) fallen die Problemstellen des Verfahrens auf. Die Nase des Frauengesichtes wurde nur sehr schmal rekonstruiert und der Augenbereich ist ungeachtet der Einbuchtung komplett falsch. Das liegt allerdings an dem nicht konstanten Albedo. Der gesamte Augenbereich inkl. den Augenbrauen weist keinen konstanten Reflexionsfaktor auf.

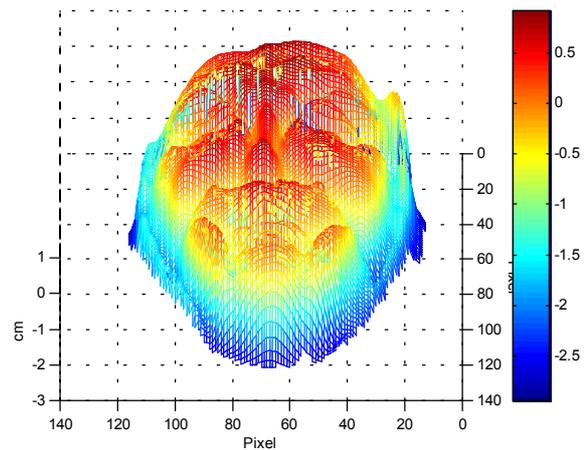


Abb. 8.16: Links: Darstellung des realen Ausgangsbildes (Mask);
Rechts: Rekonstruiertes Tiefenbild der linken Abbildung [FtpUcf99]

Neben dem zuvor vorgestellten ‚guten‘ Rekonstruktionsergebnis soll hier im Gegensatz zu den anderen Verfahren ein weiteres bereits bekanntes Bild mit konstantem Albedo vorgestellt werden. An diesem können sehr gut die Problemstellen des Verfahrens erkannt werden. Im Tiefenbild (rechts) von Abb. 8.16 sind alle Konturen sehr stark ausgeprägt. Aufgrund dessen ist ohne das Originalbild (links) nur sehr schwer eine Zuordnung zu den einzelnen Gesichtsbereichen möglich. Nach Tsai & Shah hat ihr Ansatz Probleme mit Eigenschatten und starken Intensitätsänderungen, wie an Nase oder Wangen erkennbar ist.

Ungeachtet dessen, daß für dieses Verfahren Vorgabewerte nicht vorgesehen sind, wurden dennoch Untersuchungen auf Basis der Startinitialisierung für die Iteration getätigt. Hierbei wurden auf Grundlage der synthetisch generierten Halbkugel alle Tiefenwerte aus der gewählten Gitteranordnung als Initialwerte genutzt. Eine Verbesserung des Rekonstruktionsergebnisses wurde hiermit jedoch nicht erzielt. Das liegt an der diskreten Approximation von z nach x und y . An den Vorgabepositionen gibt die Ableitung an diesen Stellen einen fehlerhaften Wert zurück, mit dem das Verfahren nicht zurechtkommt.

Der Ansatz von Tsai & Shah ist ein einfacher, unkomplizierter und effizienter Ansatz. Die Tiefe wird ungeachtet der Randbedingungen iterativ berechnet. Bereits mit wenigen Iterationsschritten (in den hier vorgestellten Bildern fünf) wird ein gutes Ergebnis erzielt. Besonders an glatten Flächen, wie in Abb. 8.15 zu sehen ist, werden sehr gute Ergebnisse erzielt. Gebiete mit starken Wölbungen schlagen dagegen fehl. Hier macht sich der Eigenschatten und die starke Intensitätsänderung negativ bemerkbar.

Wie bei allen Verfahren muß auch hier die Beleuchtung in unendlich weiter Entfernung angebracht sein. Dabei ist das Rekonstruktionsergebnis um so besser, je kleiner der Winkel zwischen Beobachter- und Beleuchtungsrichtung ist. Ein weiterer Vorteil des Ansatzes ist seine einfache Adaptierbarkeit an weitere Reflektionsmodelle. Tsai & Shah haben in ihrer Ausarbeitung [TsSh94] neben dem Reflexionsmodell von Lambert auch eine Rechenvorschrift für spiegelnde Oberflächen aufgestellt.

8.2.2.5 Vergleich der SFS-Verfahren

Für einen besseren Vergleich der einzelnen SFS-Verfahren werden im folgenden einige Ergebnisse der Verfahren parallel dargestellt. Entsprechend der Vorgehensweise in den einzelnen Verfahren soll auch hier mit der Betrachtung der synthetischen Halbkugel

begonnen werden. In Abb. 8.17 sind dazu die Rekonstruktionsergebnisse der synthetischen Halbkugel mit den dazugehörigen absoluten Fehlern dargestellt. Für eine genauere Beschreibung der Ergebnisse und der verwendeten Startbedingungen wird auf die Erläuterung innerhalb der einzelnen Verfahren verwiesen.

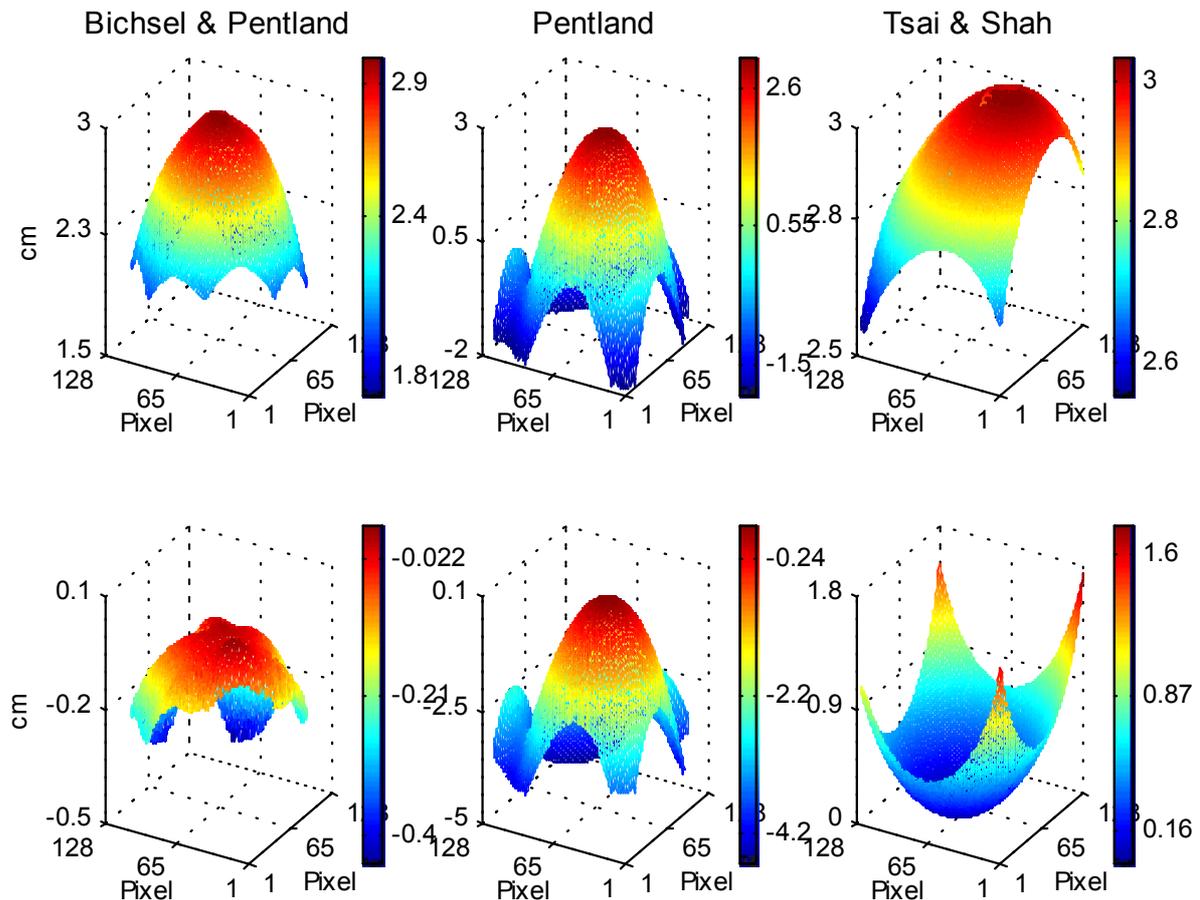


Abb. 8.17: Vergleich der SFS-Verfahren anhand einer synthetischen Halbkugel
obere Zeile: rekonstruiertes Tiefenbild; untere Zeile: absoluter Rekonstruktionsfehler

Im Anschluß an die Rekonstruktionen der Halbkugeln sind in Abb. 8.18 einige Ergebnisse von realen Bildern dargestellt. Die dazugehörigen Grauwertbilder sind aus Platzgründen hier nicht enthalten und können bei Bedarf für die obere Zeile aus Kap. 8.2.2.2 (Mask) und die untere Zeile aus Kap. 8.2.2.3 (Manneq) entnommen werden. Eine Darstellung des Fehlers konnte aufgrund der nicht vorhandenen Tiefendaten nicht erfolgen.

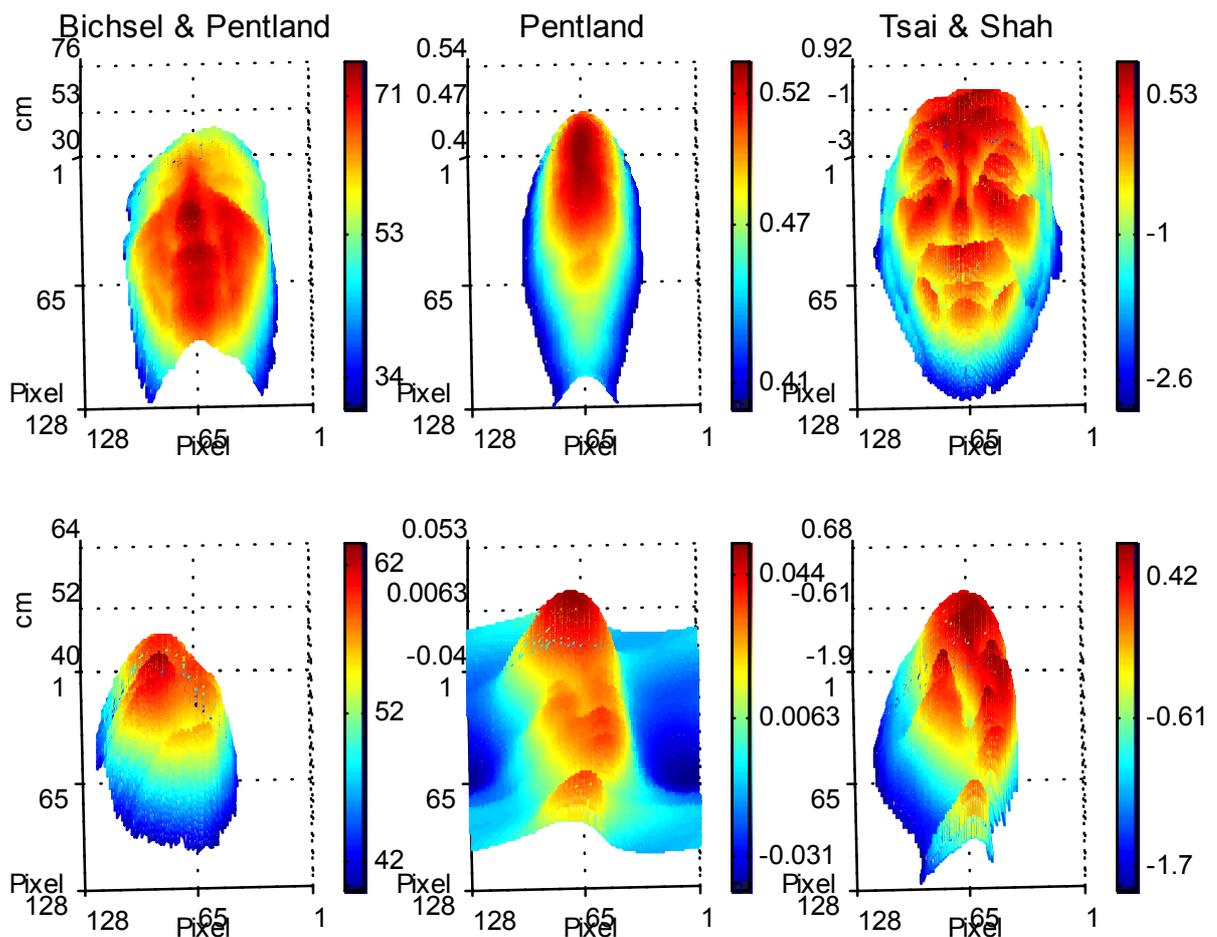


Abb. 8.18: Vergleich der SFS-Verfahren anhand von Mask (obere Zeile) und Manneq (untere Zeile)

Wie in den Darstellungen von Abb. 8.18 gut zu erkennen ist, sind die meisten Rekonstruktionen fehlgeschlagen. Nur die in den Verfahren bereits vorgestellten Bilder wurden zufriedenstellend rekonstruiert. Alle anderen Rekonstruktionen sind nicht zu gebrauchen und sollen daher hier nicht weiter diskutiert werden. Die Ergebnisse der zufriedenstellenden Rekonstruktionen wurden bereits in den entsprechenden Abschnitten erläutert.

8.2.2.6 Zusammenfassung der SFS-Verfahren

Von den drei hier untersuchten SFS-Verfahren konnte keines bezüglich seiner Rekonstruktionseigenschaften überzeugen. In Hinsicht auf die gestellte Aufgabe besitzt jeder Ansatz seine spezifischen Vor- und Nachteile und hat somit seinen speziellen Anwendungsfall. Der aus der Halbkugel gewonnene Fehlerwert ist nur bedingt aussagekräftig. Mit Hilfe dieses Wertes kann keine Aussage über die allgemeine Anwendbarkeit des Ansatzes gemacht werden, wie aus dem Vergleich der Rekonstruktionsergebnisse in Kap. 8.2.2.5 deutlich wurde.

Der Ansatz von Bichsel & Pentland liefert für die vorgestellten Halbkugeln das beste Rekonstruktionsergebnis, schlug dafür aber fehl bei den anderen getesteten Bildern. Im Gegensatz zur Aussage der Autoren ist dieser Ansatz der langsamste, was an seiner Zugehörigkeit zu den Propagationsverfahren liegt. Ungeachtet dessen wurde dieser Algorithmus mit in die Auswahl einbezogen, da er vereinzelt Tiefenwerte als Vorgabewerte unterstützt. Allerdings stellte sich heraus, daß die zur Verfügung stehenden Stützpunkte aus der Gitteranordnung aufgrund der Nutzung einzig an singulären Stellen nur begrenzt genutzt werden konnten. Die vorhandenen Tiefenwerte von der Gitteranordnung liegen nur im seltensten Fall an den singulären Stellen.

Neben den Verfahren von Bichsel & Pentland können auch bei dem SFS-Verfahren von Pentland Vorgabewerte genutzt werden. Dies wurde allerdings aufgrund der schlechten Rekonstruktionsergebnisse nicht weiter untersucht. Wie auch bei den anderen Verfahren ist das vorgestellte Rekonstruktionsergebnis das Beste aus der Vielzahl der getesteten Bilder. Das dennoch ‚schlechte‘ Ergebnis kann mit der zugrundeliegenden Vereinfachung zusammenhängen. Der Ansatz ist für lineare Oberflächen, welche von einem großen Beleuchtungswinkel bestrahlt werden, optimiert. Hinsichtlich der Geschwindigkeit liegt das SFS-Verfahren von Pentland im Mittelbereich. Die notwendige 2D-FFT (Hin- und Rücktransformation) benötigt viel Rechenzeit.

Der letzte untersuchte Ansatz von Tsai & Shah ist der schnellste und beste Ansatz aller hier getesteten Verfahren. Dennoch hat auch er seine Schwachstellen. Er arbeitet nur an homogenen Grauwertverläufen gut. Bei großen Grauwertänderungen versagt dieser Ansatz.

Im Rahmen der Untersuchungen wurden infolge der ‚schlechten‘ Ergebnisse keine weiterführenden Untersuchungen auf Basis von verrauschten Bildern durchgeführt. Eine Verschlechterung des Ausgangsbildes würde zu einer unbrauchbaren Rekonstruktion führen.

8.2.3 Lambert-Verfahren

In den nachfolgenden Abschnitten sollen zwei vom Autor entwickelte Verfahren zur Rekonstruktion von Tiefenwerten vorgestellt werden, welche die Reflexionseigenschaften eines Vorgabewertes auf einen kleinen Nachbarschaftsbereich übertragen. Ausgehend von der Helligkeit und dem Tiefenwert eines Startpunktes werden allen benachbarten Punkten mit einer ähnlichen Helligkeit die Reflexionseigenschaften des Ausgangspunktes übertragen. Aufbauend auf diese Eigenschaften können die gesuchten Tiefen- bzw.

Flächennormalen ausgehend von der Tiefeninformation vom Vorgabewert rekonstruiert werden.

Die Idee dieses Ansatzes stammt von den Level-Set-Verfahren oder Niveaumengen-Methoden ab. Diese erstmals von Bruckstein 1988 [Br88] auf die SFS-Problematik angewendete Methode beschreibt einen rekursiven Ansatz, welcher die ‚equal-height‘ oder ‚level contours‘ einer Oberfläche von einem Startwert ausgehend bestimmt. In den nachfolgenden Jahren wurde dieser Ansatz insbesondere von Kimmel & Bruckstein [KiBr95a] [KiBr95b] [KiKa95] [ShKi96] verbessert. Durch den großen notwendigen Rechenaufwand ist diese Methode nicht direkt für die Aufgabenstellung dieser Arbeit geeignet. Einzig die zugrundeliegende Idee des Verfahrens wurde vom Autor aufgegriffen und auf das 2D/3D-Kamerakonzept angepaßt.

Im folgenden sollen zwei auf dieser Basis entwickelte Verfahren vorgestellt werden. Im Abschnitt Lambert-Z wird dabei ein Verfahren zur Rekonstruktion der Objekthöhe und in Lambert-N ein Verfahren zur Rekonstruktion der Flächennormale vorgestellt. Zuvor jedoch werden einige allgemeine Voraussetzungen, welche gleichermaßen beide Verfahren betreffen, erläutert.

8.2.3.1 Allgemein

Grundvoraussetzung für beide nachfolgenden Verfahren ist ein vollständiges Grauwertbild mit vereinzelt Tiefenwerten als Startwerte. Ausgehend von diesen Startwerten wird deren Reflexionseigenschaft auf Basis der Lambert-Reflexion in einem kleinen Nachbarschaftsbereich ausgeweitet. Unter der Voraussetzung eines konstanten Grauwertes und der konstanten Beleuchtungsposition wird für jedes benachbarte Element mit einem ähnlichen Grauwert wie dem des Vorgabewertes die resultierende Flächennormale bzw. Objekthöhe berechnet. Aufgrund der dazu notwendigen kleinen Toleranz innerhalb des Grauwertkriteriums werden für eine flächendeckende Rekonstruktion möglichst viele Startwerte benötigt. Die Linienstruktur wird dieser Anforderung am besten gerecht. Durch die benachbarte Anordnung der einzelnen Punkte werden viele Startwerte zur Verfügung gestellt. Ungeachtet der Anzahl und Anordnung der Tiefenvorgabewerte werden im Gegensatz zu den bereits vorgestellten Verfahren alle Tiefenwerte effektiv genutzt.

Voraussetzung für eine gute Rekonstruktion ist eine Szene mit möglichst diffusen Reflexionseigenschaften. Für eine einwandfreie Funktion sollten am besten nur homogene Oberflächen genutzt werden. Ein sich langsam ändernder Albedo führt zu einer fehler-

haften Rekonstruktion. Durch die Einschränkung des zulässigen Toleranzbandes für die Grauwertsuche kann dennoch ein gewisser Fehler toleriert werden. Je kleiner die Toleranz ist, desto kleiner wird der rekonstruierte Fehler. Zwangsläufige Änderungen der Flächennormale trotz konstanter Helligkeit fallen folglich geringer aus. Allerdings steigt mit kleiner werdender Toleranz die Anzahl der notwendigen Stützpunkte und somit die Rechenzeit.

Ein Vorteil dieser Methode ist die Unabhängigkeit von einer vorangegangenen Segmentierung. Infolge der Nutzung des konstanten Grauwertes wird ein Bereich entweder bis zur frei definierbaren Bereichsgrenze oder bis zur Änderung des Grauwertes ausgeweitet. Die sonst notwendige Segmentierung der Oberfläche kann folglich entfallen.

In Abhängigkeit der Oberflächenbeschaffenheit und des Gitterabstandes kann es vorkommen, daß für einen zu rekonstruierenden Punkt mehrere Berechnungen von verschiedenen Startpunkten aus erfolgen oder aufgrund der großen Grauwertabweichung gar nicht berechnet werden. Daher ist es hilfreich, eine zweite Matrix mit der Anzahl der Rekonstruktionen für einen Punkt mitzuführen. Besitzt der Zähler einen Wert von Null, so wurde die entsprechende Position nicht rekonstruiert. Bei Werten größer als eins kann z.B. der resultierende Entfernungswert über eine Mittelwertbildung bestimmt werden.

8.2.3.2 Lambert-Z

Ausgehend von der Lambert-Reflexion mit $I = \rho \cdot \cos(i)$ wird angenommen, daß in einem kleinen Nachbarschaftsbereich um einen Startwert bei gleicher wahrgenommener Helligkeit auch der Albedo und der Winkel i zwischen der Flächennormalen und der Beleuchtungsrichtung konstant sind. Unter dieser Voraussetzung werden für alle benachbarten Punkte innerhalb einer kleinen Umgebung mit demselben Grauwert wie dem Startwert angenommen, daß die Entfernung von der Kamera über den Meßpunkt bis zur Beleuchtungsquelle ebenfalls konstant ist. Diese Bedingung wird um so besser erfüllt, je näher sich die Beleuchtungsquelle und die Kamera an dem Meßpunkt befinden. Mit größer werdendem Abstand wirkt sich ein Helligkeitsabfall in Abhängigkeit der Entfernung geringer aus.

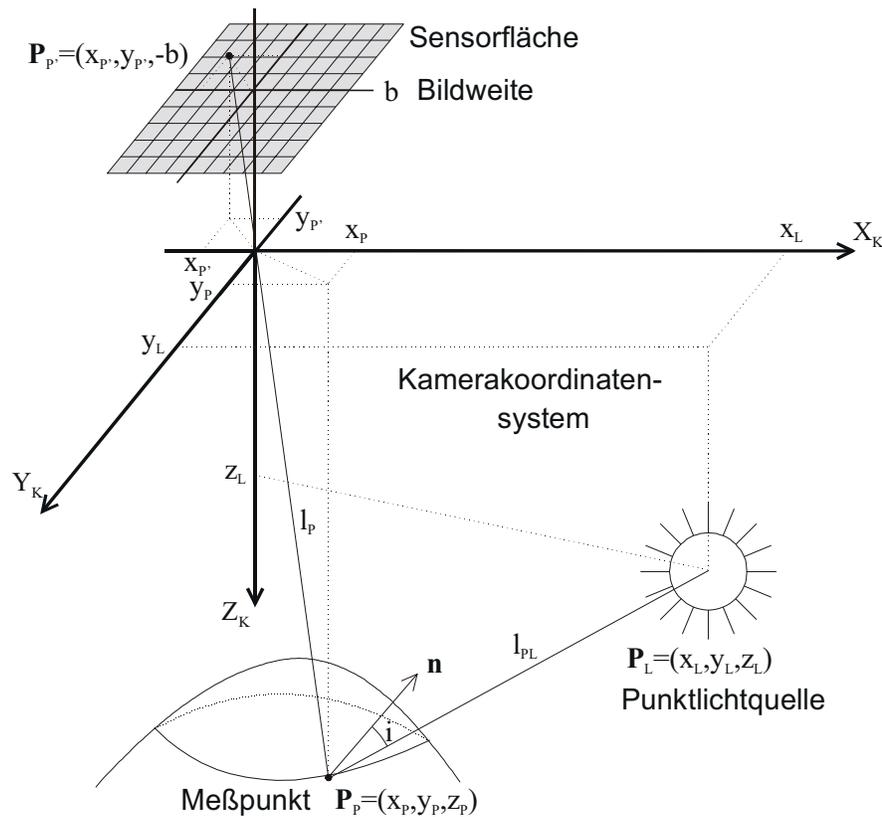


Abb. 8.19: Aufnahmegeometrie zur Bestimmung der Objekthöhe

Ausgehend von der Aufnahmegeometrie in Abb. 8.19 wird die Entfernung $l_{SENSOR'}$ von der Lochblende der Kamera (Nullpunkt) über den Meßpunkt P_P zur Beleuchtungsquelle P_L berechnet:

$$l_{SENSOR'} = l_P + l_{PL} = |P_P| + |P_L - P_P| \quad (8.14)$$

Mit Hilfe dieser Formel kann zum einen die Vorgabelänge $l_{SENSOR'}$ und zum anderen die Position P_P der zu rekonstruierenden Punkte berechnet werden. Dazu wird ein Strahl, beginnend vom gesuchten Pixel $P_{P'}$ der Sensorfläche, über den Nullpunkt berechnet, auf dem sich der zu rekonstruierende Punkt befinden muß. Durch diesen Ansatz schränkt sich der Suchbereich für die gesuchte Position im Raum von drei auf eine Unbekannte ein. Für den Strahl ergibt sich für die zu bestimmende Sensorposition folgende Geradengleichung:

$$G(\gamma) = \gamma P_{P'} \quad (8.15)$$

Durch Einsetzen dieser Gleichung in (8.14) verbleibt γ als einzige Unbekannte in der Gleichung:

$$l_{SENSOR'} = |\gamma P_{P'}| + |P_L - \gamma P_{P'}| \quad (8.16)$$

Nach Auflösung der Betragsgleichung fällt der dabei anfallende quadratische Anteil von γ heraus. Übrig bleibt eine einfache lineare Gleichung in Abhängigkeit von γ :

$$\gamma = \frac{x_L^2 + y_L^2 + z_L^2 - l_{SENSOR'}^2}{2(x_L x_{P'} + y_L y_{P'} + z_L b - \sqrt{x_{P'}^2 + y_{P'}^2 + b^2})} \quad (8.17)$$

Die gesuchte Position im Raum kann nun durch das Einsetzen des Wertes γ in Gleichung (8.15) bestimmt werden. Für eine Geschwindigkeitssteigerung können z.B. die konstanten Werte aus Gleichung (8.17) im Vorfeld berechnet werden. Damit ergibt sich eine einfache und schnelle Methode zur Rekonstruktion der gesuchten Position.

Für ein besseres Verständnis der Vorgehensweise dieses Ansatzes ist der Algorithmus in einer Pseudo-Programmiersprache im Anhang beschrieben. Dieser dient als Grundlage für das zur Untersuchung dieses Ansatzes entwickelte MATLAB-Programm *lambertz.m*. Entsprechend den SFS-Untersuchungen soll auch hier das Rekonstruktionsergebnis an einer Halbkugel demonstriert werden. Aufgrund der fehlenden Tiefenwerte der realen Szene wurde für den zweiten Test die bekannte Testszene als weiteres Beispiel herangezogen.

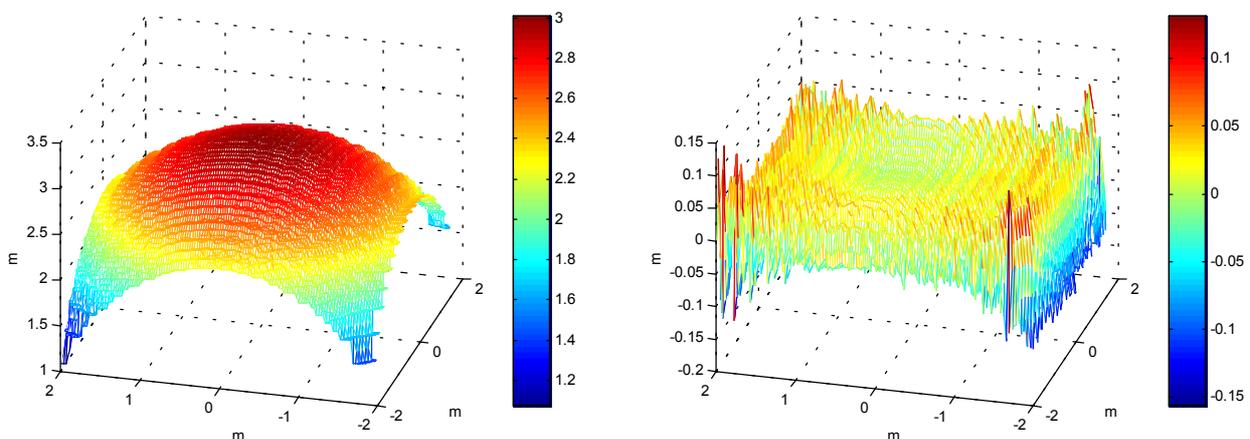


Abb. 8.20: Links: Rekonstruktion einer synthetisch generierten Halbkugel bei Vorgabe von 20 x 20 Tiefenwerten und 128 x 128 Grauwerten; Rechts: absoluter Rekonstruktionsfehler

Unabhängig von der Beleuchtungsposition rekonstruierte das Verfahren die Halbkugeln gut bis sehr gut. Bereits unter Nutzung der Punktstruktur (20 x 20 Punkte) als Vorgabe für die Tiefenwerte wurde, wie in Abb. 8.20 (links) zu sehen ist, ein sehr gutes Rekonstruktionsergebnis erzielt. Als absoluter Fehler wurde mit dieser Vorgabe im mittleren

Bildbereich eine Abweichung von ± 0.05 m und zum Rand hin eine Abweichung von ± 0.1 m erreicht.

In Abhängigkeit der Szene und der Vorgabe der Gitteranordnung kann sich die Vorgabe auch ‚negativ‘ auswirken. Bei Nutzung der Stern- oder Kreisstruktur kann diese im rekonstruierten Bild erkannt werden. Dennoch erhöht sich der absolute Fehler nur geringfügig. Für dieselbe Szene und dieselben Randbedingungen ergibt sich bei Vorgabe der Stern- und Kreisstruktur ein Fehler im mittleren Bildbereich von ± 0.05 m, der sich zum Rand hin auf ± 0.25 m erhöht.

Ein klarer Vorteil dieses Verfahrens ist die Übergabe der Szene ohne eine vorangegangene Segmentierung des Bildes, wie sie für die SFS-Verfahren normalerweise notwendig ist. Dieses Verfahren nutzt aufgrund seiner Arbeitsweise nur Daten mit konstantem Grauwert, welche folglich ein und derselben Region angehören. Dies ist sehr gut anhand der Abb. 8.21 zu erkennen. Diese Szene wurde dem Verfahren ohne eine Vorverarbeitung übergeben. Als Gitteranordnung wurde dabei die Punktstruktur mit 20×20 Punkten (linke Spalte) und die Kreisstruktur mit 10 Ringen (rechte Spalte), jeweils ausgehend von einem Grauwertbild mit 128×128 Punkten, genutzt. Der absolute Rekonstruktionsfehler (untere Zeile) liegt bei beiden Vorgaben zwischen ± 0.1 m und ± 0.25 m. Eine Ausnahme bilden die seitlichen Begrenzungsflächen. Hier liegt der Fehler bei $+0.8$ m.

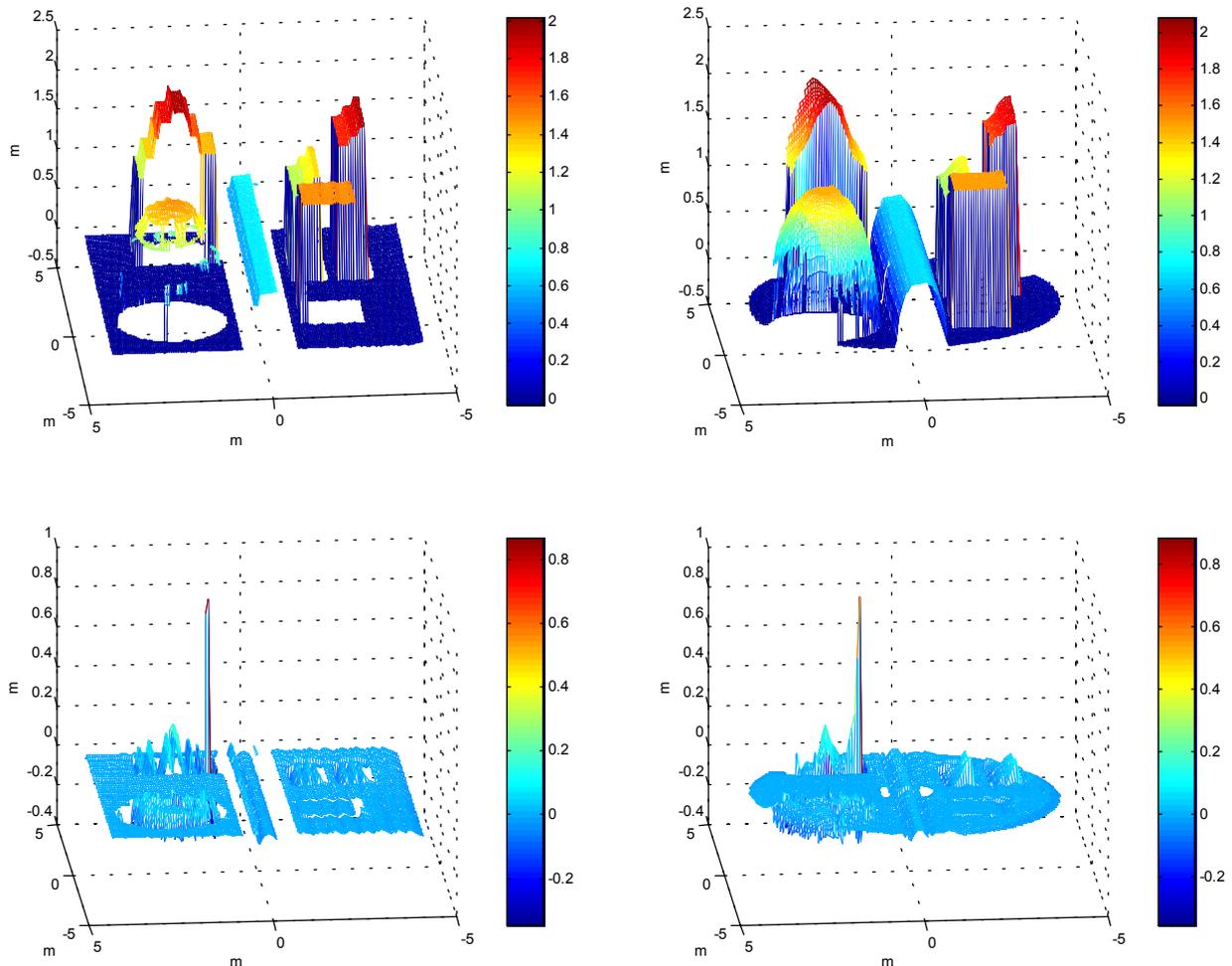


Abb. 8.21: obere Zeile: Rekonstruktion von 128 x 128 Tiefenwerten unter Vorgabe von 20 x 20 Tiefenwerten (links) und Tiefenwerten in Form von 10 Ringen (rechts)
untere Zeile: dazugehöriger Rekonstruktionsfehler

Aufgrund der wenigen Vorgabewerte wurde die linke Abbildung nur zum Teil rekonstruiert. Die runde Kreisfläche in der rechten Abbildung rührt von dem letzten zur Verfügung gestellten Kreisring her. Durch die hohe Anzahl von Vorgabewerten ist die rechte Rekonstruktion im Gegensatz zur linken, in der viele Lücken im Ergebnis enthalten sind, sehr gut. Die Halbkugel und der Halbzylinder in der linken Abbildung wurden nur zum Teil rekonstruiert. Ebenso sind in den hinteren Objekten Treppenstufen sichtbar. Ungeachtet dieser Problemstellen können alle Objekte gut erkannt werden, besonders in der rechten Abbildung. Hier wurde die Szene fast vollständig rekonstruiert, und es sind nur vereinzelte Aussetzer zu erkennen.

Durch die Einschränkung des Suchbereiches des Zielpunktes auf eine Gerade, welche durch das Pixelelement und die Lochblende der Kamera geht, werden alle rekonstruierten

Werte entsprechend dem vorgestellten Aufnahmeprinzip (siehe Kap. 6.2.3) rekonstruiert. Die berechneten Werte sind im Gegensatz zu den SFS-Verfahren absolute Werte und können wahlweise als Objektentfernung l_p oder als XYZ-Koordinaten abgespeichert werden.

Als Vorgabewerte werden für das Lambert-Z Verfahren neben den Startwerten aus der Gitteranordnung die Kamera- und Beleuchtungspositionen benötigt. Diese sind die gleichen Vorgaben wie sie für die XYZ-Korrektur benötigt werden. Aufgrund des von den Pixelelementen ausgehenden Ansatzes rekonstruiert das Verfahren einzelne Punkte äquivalent zu der Aufnahmeart der Kamera. Die Formel zur Rekonstruktion (8.17) ähnelt dementsprechend der Gleichung zur XYZ-Korrektur (6.17). Daher ist es ratsam, das Verfahren gleichzeitig mit der XYZ-Korrektur der vereinzelt Punkte anzuwenden. Als Ergebnis erhält man mit diesem Verfahren unter Nutzung von vereinzelt Tiefenwerten ein fast vollständig rekonstruiertes Tiefenbild.

8.2.3.3 Lambert-N

Entsprechend dem Lambert-Z Verfahren wird auch hier angenommen, daß in einem kleinen Nachbarschaftsbereich um einen Startpunkt bei gleicher wahrgenommener Helligkeit auch der Albedo und der Winkel i zwischen der Flächennormalen und der Beleuchtungsrichtung konstant sind. Unter Nutzung dieser Annahme wird versucht, in einem benachbarten Punkt mit gleicher Helligkeit wie dem des Ausgangspunktes den Winkel zwischen der Flächennormale und der Beleuchtungsquelle auf den Wert i zu setzen.

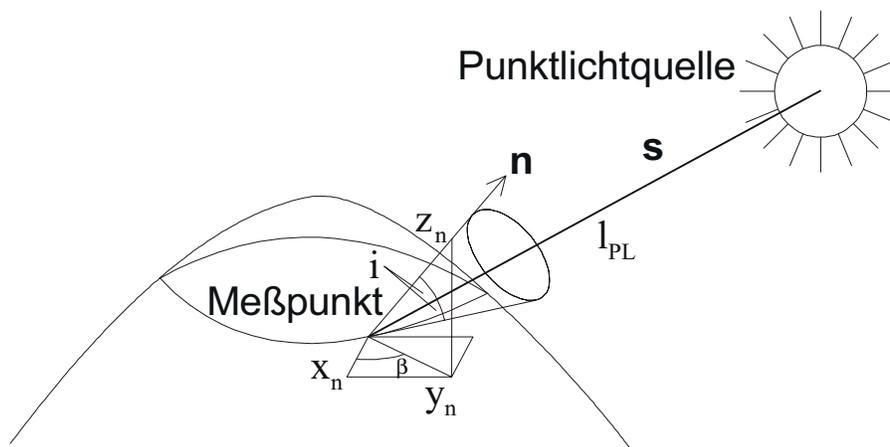


Abb. 8.22: Aufnahmegeometrie zur Bestimmung der Oberflächennormale

Ausgehend von der Aufnahmegeometrie in Abb. 8.22 wird die Lambert-Reflexion unter der Annahme der Beleuchtungsrichtung als Einheitsnormale $s_e = (x_{Se}, y_{Se}, z_{Se})$ wie folgt aufgestellt:

$$\begin{pmatrix} x_{Se} \\ y_{Se} \\ z_{Se} \end{pmatrix} \cdot \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \cos(i) \cdot \sqrt{x_n^2 + y_n^2 + z_n^2} \quad (8.18)$$

Für jeden Startpunkt wird mit dieser Formel der Vorgabewinkel i berechnet. Anschließend wird für jeden benachbarten Punkt mit ähnlicher Helligkeit angenommen, daß dieser Winkel konstant bleibt. Somit befindet sich die gesuchte Flächennormale auf einer Trichteroberfläche mit dem Öffnungswinkel i zur Beleuchtungsrichtung (siehe Abb. 8.22). Infolge der großen Lösungsanzahl wird zur Einschränkung des Suchbereiches weiterhin angenommen, daß der Winkel β sich entsprechend der lokalen Grauwertorientierung (siehe Kap. 6.5) um die selbe Winkeldifferenz analog dem des Start- und Zielpunktes dreht. Die nun noch fehlende Z-Komponente der neuen Flächennormalen ergibt sich nach Umstellung der Gleichung (8.18) unter Nutzung der berechneten XY-Komponenten:

$$\begin{aligned} z_n^0 * (x_{Se}^2 x_n^2 + y_{Se}^2 y_n^2 + 2x_{Se} x_n y_{Se} y_n - \cos^2(i) \cdot x_n^2 - \cos^2(i) \cdot y_n^2) + \\ z_n^1 * (2x_{Se} x_n z_{Se} + 2y_{Se} y_n z_{Se}) + \\ z_n^2 * (z_{Se}^2 - \cos^2(i)) = 0 \end{aligned} \quad (8.19)$$

Damit ergeben sich zwei Lösungen, wobei diejenige ausgesucht wird, welche sich näher an der Z-Komponente der Ausgangsnormalen befindet. Als Ergebnis erhält man, anstelle der Objekthöhe, die Rekonstruktion der Flächennormalen. Wird die Höhendarstellung benötigt, so muß diese z.B. mit einem parallel mitlaufenden Lambert-Z Verfahren oder über einen zusätzlichen Integrationsschritt berechnet werden.

Eine Beschreibung des Ansatzes in einer Pseudo-Programmiersprache ist im Anhang zu finden. Für eine praktische Untersuchung wurde das Verfahren in dem MATLAB-Programm `,lambertn.m'` implementiert.

Für die Bestimmung der lokalen Orientierung wird in der Initialisierungsphase das bereits besprochene Programm `,orientierung.m'` auf das Grauwertbild angewendet. Der Rückgabeparameter `,Phase'` ist nur gültig, wenn der Rückgabeparameter `,Betrag'` einen unteren Grenzwert überschreitet. Sonst kann davon ausgegangen werden, daß der zu rekonstruierende Bereich eine homogene Fläche darstellt. In diesem Fall wird die gesuchte Flächennormale auf den Wert der Startnormalen gesetzt. Wird im Gegensatz zur unteren Grenze

ein oberer Grenzwert überschritten, so kann die Phase aufgrund eines möglichen abrupten Überganges des Phasenwertes nicht genutzt werden. Die Rekonstruktion dieses Punktes ist somit nicht möglich.

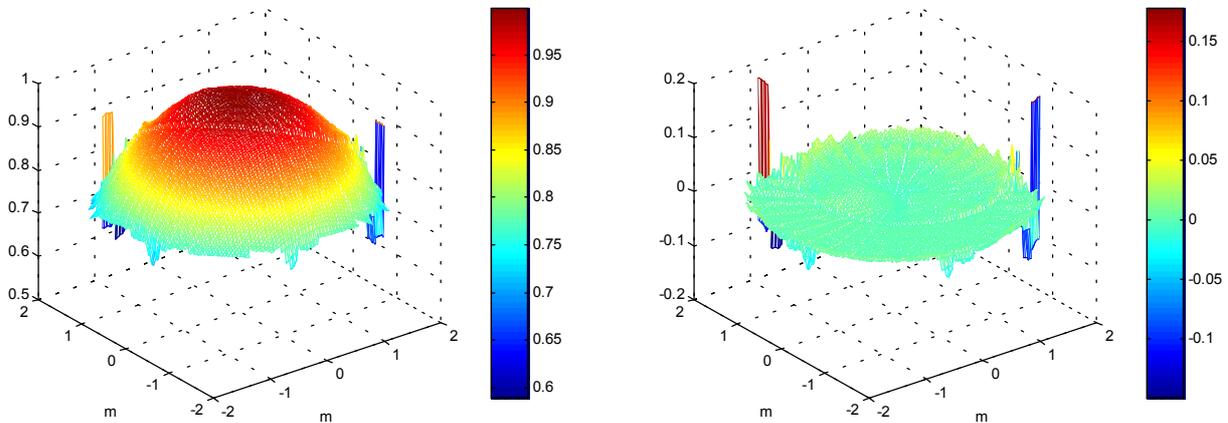


Abb. 8.23: Links: Rekonstruierte Z-Komponente der Flächennormale einer synthetisch generierten Halbkugel mit 128×128 Werten bei Nutzung des Sterngitters mit 16 Strahlen; Rechts: absoluter Rekonstruktionsfehler der Z-Komponente

In der Abb. 8.23 ist die Z-Komponente der Flächennormale einer rekonstruierten synthetischen Kugel (links) und deren absoluter Fehler (rechts) dargestellt. Als Gitteranordnung für die Tiefenwerte wurde das Sterngitter mit 16 Tiefenlinien vorgegeben. Die X- und Y-Komponente der Flächennormalen entsprechen einer ebenen Fläche und sind somit nicht so aussagekräftig wie die Z-Komponente. Unter Vernachlässigung der zwei Fehlstellen am Rande der Halbkugel beträgt der Fehler 2%. Eine Erhöhung der Anzahl der Stützpunkte führt nur bedingt zu einer Verbesserung des Rekonstruktionsergebnisses. Abhängig von der Beleuchtungsposition und der Anzahl der Stützpunkte ist das Ergebnis ein wenig schlechter als das hier präsentierte und weist vereinzelte Löcher auf. Besonders bei der Vorgabe eines Kreismusters schlägt das Verfahren fehl. Hier werden entsprechend der Vorgabe nur vereinzelte Ringe rekonstruiert.

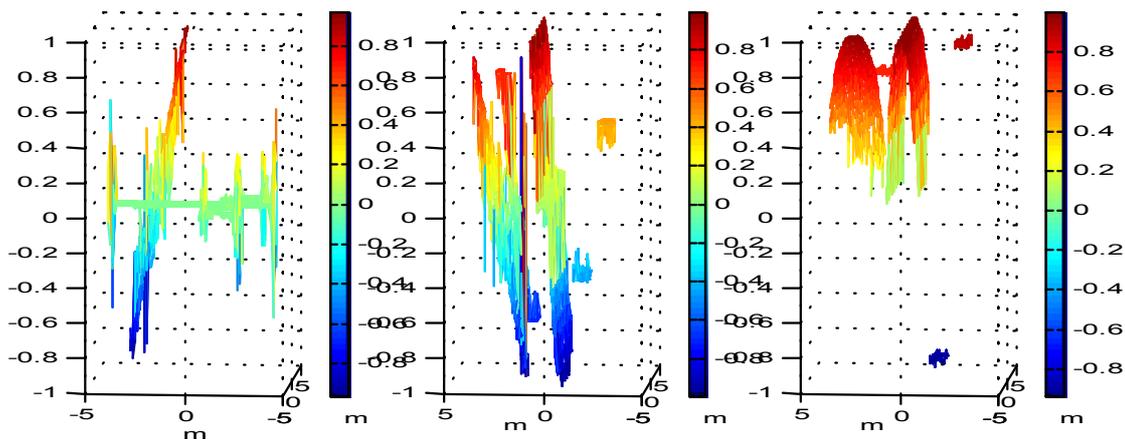


Abb. 8.24: Rekonstruktion von 128×128 Flächennormalen unter Vorgabe von 10 Ringen mit Flächennormalen (Links: X-Komponente; Mitte: Y-Komponente; Rechts: Z-Komponente)

Für die Darstellung des Rekonstruktionsergebnisses wurde in Anlehnung an das Lambert-Z Verfahren dieselbe Szene mit derselben Betrachter- und Beleuchtungsposition ausgewählt. Als Gitteranordnung wurde ebenfalls die Kreisstruktur mit 10 Ringen genutzt. In Abb. 8.24 sind alle drei Komponenten der Flächennormale dargestellt. Zur besseren Interpretation ist die X-Komponente um 90° gedreht worden.

Wie in den Abbildungen zu sehen ist, arbeitet das Verfahren an komplexen Szenen nur begrenzt gut. Die ebene Grundfläche wurde gar nicht rekonstruiert. Nur Teile der Objekte sind vorhanden. Dies liegt an der Orientierungsbestimmung. Innerhalb der Kantenübergänge zwischen den Objekten zeigt der Orientierungswert aufgrund des abrupten Überganges in eine falsche Richtung und der Betrag weist ein Maximum auf. Diese Bereiche können folglich nicht rekonstruiert werden.

Die X- und Y-Komponenten des rekonstruierten Bildes sind sehr stark gestört. Auch die Z-Komponente weist Störungen auf. Allerdings fallen diese im Vergleich zu den anderen Komponenten geringer aus. Eine Weiterverarbeitung dieser Daten ist folglich nur begrenzt möglich. Alle anderen untersuchten Szenen mit unterschiedlichen Betrachter- und Beleuchtungspositionen weisen ein noch schlechteres Ergebnis als das hier vorgestellte auf. Werden anstatt der idealen Vorgabedaten reale Daten genutzt, sind die Ergebnisse folglich nicht für eine Weiterverarbeitung geeignet. In Anbetracht dieser Tatsache wurden keine weiteren Tests mit diesem Verfahren durchgeführt.

8.2.3.4 Zusammenfassung der Lambert-Verfahren

Die Lambert-Verfahren versuchen auf Grundlage der Übertragung bestimmter Eigenschaften eines Startpunktes auf die Nachbarpunkte, den Tiefen- bzw. Orientierungswert in

einem kleinen Nachbarschaftsbereich zu rekonstruieren. Ungeachtet dieser ‚bedingt‘ fehlerhaften Annahme sind die Rekonstruktionsergebnisse des Lambert-Z Verfahrens sehr gut. Auch ohne eine vorangegangene Segmentierung wird die komplexe synthetische Szene sehr gut rekonstruiert. Im Falle von realen Daten wird nach einer entsprechenden Filterung des Grauwertbildes ebenfalls ein gutes Ergebnis erwartet. Der Rekonstruktionsfehler steigt jedoch mit größer werdendem Rauschen. Als Anwendungsbeispiel eignet sich das Lambert-Z Verfahren am besten für technische Flächen oder Gebiete mit einer geringen Grauwertdifferenz. Große Grauwertdifferenzen, wie sie in natürlichen Umgebungen vorkommen, können ohne dazugehörige Startwerte nicht überbrückt werden.

Unabhängig von der ähnlichen zugrundeliegenden Idee beider Verfahren weist das Lambert-N Verfahren nicht die guten Rekonstruktionseigenschaften des Lambert-Z Verfahrens auf. Aufgrund der schlechten Ergebnisse mit synthetischen Bildern ist eine Nutzung des Lambert-N Verfahrens mit realen Bildern nicht empfehlenswert. Bereits kleine Störungen im Grauwertbild bzw. in der vorgegebenen Flächennormalen führen zu einem unbrauchbaren Ergebnis.

Unabhängig von der Qualität der Rekonstruktionen weisen beide Verfahren folgende Vor- und Nachteile auf:

- + Effiziente Nutzung der Vorgabewerte
- + Keine Segmentierung des Bildes notwendig
- + Beschränkung des zulässigen Fehlers durch Wahl des Suchbereiches
- + Lambert-N Verfahren ist unabhängig von der Beobachterposition
- + Auch auf andere Reflexionsmodelle anwendbar
- Nur eine Punktlichtquelle erlaubt
- Nicht geeignet für Schattenregionen
- Homogene Oberfläche mit konstantem Albedo notwendig
- Geschwindigkeit abhängig von der Anzahl der Stützpunkte und des Suchbereiches

8.2.4 Zusammenfassung der Verfahren zur Ausweitung der vereinzelt Entfernungen

Innerhalb dieses Kapitels wurden drei verschiedene Vorgehensweisen zur Rekonstruktion der Tiefenwerte untersucht. Das direkteste Verfahren, die Flächenapproximation, nutzt dabei nur Tiefendaten. Bei den anderen beiden Verfahren werden die fehlenden Tiefeninformationen durch Datenfusion auf Pixelebene rekonstruiert.

Aufgrund der Datenfusion und der dazu genutzten Lambert-Reflexion sind für das SFS-Verfahren und die Lambert-Verfahren die Angabe der Position bzw. Richtung der Beleuchtungsquelle notwendig. Als weiteren Übergabeparameter benötigt das Lambert-Z Verfahren die bereits bekannte Kameraposition. Ein nicht konstanter Albedo führt in beiden Gruppen direkt zu einer Verfälschung des Ergebnisses. Aufgrund der Nutzung von konstanten Grauwerten zur Ausbreitung der Tiefeninformation können im Gegensatz zu den SFS-Verfahren mit dem Lambert-Verfahren abrupte Übergänge, wie sie z.B. bei Texturen auftreten, verarbeitet werden. Ein weiterer Nachteil der SFS-Verfahren ist deren Rekonstruktionsergebnis. Sie geben trotz der Vorgabe von Startwerten nur relative Werte zurück, welche ggf. noch zusätzlich in der Amplitude angepaßt werden müssen. Hier sind das Flächenapproximations- und das Lambert-Z Verfahren klar im Vorteil. Sie berechnen die absolute Entfernung.

Aus den Untersuchungen folgt, daß die SFS-Verfahren nur unter einer großen Zahl von Einschränkungen zufriedenstellend arbeiten. Für den generellen Einsatz sind sie nicht empfehlenswert. Auch die Lambert-Verfahren unterliegen Einschränkungen. Dennoch rekonstruiert das Lambert-Z Verfahren die Testszene unter allen getesteten Verfahren am besten. Ungeachtet dessen sollte auch hier das Ergebnis mit Vorsicht behandelt werden. In Abhängigkeit der genutzten Gitteranordnung und der Applikation können Rekonstruktionsfehler auftreten. Das Lambert-N Verfahren ist im Gegensatz zu diesen guten Rekonstruktionsergebnissen nur in speziellen Fällen anwendbar. Es arbeitet nur unter Vorgabe von idealen Daten zufriedenstellend.

Mit Hilfe der Flächenapproximation kann bereits ohne Nutzung der Grauwertdaten die Szene mit gewissen Einschränkungen rekonstruiert werden. Für eine gute Rekonstruktion ist eine zuverlässige Segmentierung des Bildes notwendig. Werden Bereiche aufgrund eines schlecht erkannten Kantenüberganges nicht getrennt, so liefert das Flächenapproximationsverfahren fehlerhafte Parameter für diese Region zurück. Infolge der notwendigen Vorgabe der zu rekonstruierenden Flächenfunktion eignet sich das Flächenapproximationsverfahren nur für technische Flächen.

9 Zusammenfassung und Ausblick

Die vorliegende Arbeit beschäftigt sich mit der Untersuchung von Konzepten zur Aufnahme und Verarbeitung von flächendeckenden Grauwertdaten mit vereinzelt Tiefendaten. Als Grundlage diente dabei das PMD-Element zur gleichzeitigen Aufnahme von Grauwert- und Tiefendaten. Durch Kombination bzw. Integration einer herkömmlichen CMOS-Matrixkamera mit mehreren dieser PMD-Elemente (verteilt über die gesamte empfindliche Kamerafläche) können Grauwert- und Entfernungsbilder in Echtzeit aufgenommen und verarbeitet werden. Daraus ergibt sich eine neue Art der 3D-Bildverarbeitung. Objekterkennungsaufgaben, Anwesenheitskontrolle und andere Meßaufgaben profitieren direkt von den einzelnen hinzugekommenen Entfernungsdaten. Somit kann unter gewissen Umständen eine vollständige Objektrekonstruktion unter nur geringfügig gesteigerter Rechenleistung erfolgen. Reine grauwertbasierte Verfahren sind hier durch ihre aufwendige Rekonstruktion der verlorengegangenen Tiefeninformation klar im Nachteil. Ist ungeachtet dieser vielfältigen Einsatzmöglichkeiten dennoch ein flächendeckendes Tiefenbild notwendig, so können die vorhandenen Informationen zur Ausweitung der Tiefenwerte genutzt werden.

Zur Erfassung einer Szene mit Hilfe einer Grauwertkamera mit vereinzelt Tiefenwerten stehen verschiedene Anordnungsmöglichkeiten der vereinzelt Tiefenwerte zur Auswahl. Die Anordnung und die Dichte der Elemente sind dabei von der Aufgabenstellung und der nachfolgenden Verarbeitung abhängig. Zur Verdeutlichung dieser Problematik wurden neben der Vorstellung des 2D/3D-Kamerakonzeptes verschiedene Anordnungsbeispiele der vereinzelt Tiefenwerte dargestellt und anhand von praxisbezogenen Problemen diskutiert. Mit Hilfe des Punktgitters können bereits viele einfache Problemstellungen gelöst werden. Komplexere Aufgabenstellungen erfordern demgegenüber dichtere Gitterstrukturen. Dies wurde anhand verschiedener Gitterstrukturen erläutert.

Aufbauend auf die Untersuchung der Gitterstrukturen wurden verschiedene Lösungsstrategien für die 3D-BV mit vereinzelt Tiefenwerten vorgestellt. Hier lag das Augenmerk unter anderem auf solchen Problemstellungen, bei denen die herkömmliche Graubildverarbeitung versagt (mißverständliche bzw. mehrdeutige Stellen) und die Hinzunahme von vereinzelt Tiefeninformationen eine einfachere und sicherere Interpretation der Szene ermöglicht. Als einfachstes Beispiel kann dabei die Konkav/Konvex-Problematik angesehen werden. Bereits mit einigen wenigen Tiefenwerten kann dieses Problem gelöst

werden. Auch die in Grauwertbildern verlorengegangene Entfernungs- und Größenbestimmung kann mit Hilfe einiger weniger Tiefenwerte gelöst werden. Im einfachsten Fall reichen hierzu bereits zwei Tiefenwerte zur Bestimmung der Lage und Orientierung von einfachen und bekannten Objekten im Raum aus. Über die Flächeninterpretation stehen weiterführende Möglichkeiten zur Verfügung. Neben der einfacheren Bestimmung des Oberflächentyps kann hiermit auch die Oberflächenorientierung rekonstruiert werden.

Für Anwendungen, in denen eine hohe laterale Anzahl von Entfernungswerten notwendig sind, wurden entsprechende Möglichkeiten zur Rekonstruktion der Tiefeninformation an Stellen ohne Entfernungsinformation untersucht. Ausgehend von bekannten Tiefen- und/oder Grauwertinformationen wird dabei versucht, die fehlenden Tiefeninformationen zu rekonstruieren. Das einfachste Verfahren stellt dabei die Flächenapproximation dar. Ohne direkte Nutzung der Grauwertdaten werden die fehlenden Tiefenwerte auf Basis der Anpassung einer Flächenfunktion an die vorhandenen Tiefenwerte rekonstruiert.

Eine weitere Rekonstruktion kann über die SFS-Verfahren erfolgen, deren Ziel die Rekonstruktion der Oberflächenorientierung bzw. Objektentfernung ist. Für eine genauere Aussage über die Einsatzfähigkeit der Verfahren wurden dazu repräsentativ drei SFS-Verfahren implementiert. Ein den SFS-Verfahren ähnlicher Ansatz nutzt die Reflexionseigenschaften der Oberflächen direkt aus. Ausgehend von der Helligkeit des Vorgabepunktes werden dabei allen benachbarten Punkten mit derselben Helligkeit die Reflexionseigenschaften des Vorgabewertes übertragen. Aufbauend auf diese Eigenschaften können die gesuchten Tiefen- bzw. Flächennormalen mit Hilfe der Tiefeninformation ausgehend vom Vorgabewert rekonstruiert werden.

Die Untersuchungen auf Basis der Ausweitung der vereinzelt Tiefenwerte zeigten unterschiedliche Ergebnisse. Keines der hier untersuchten Verfahren ist für alle Aufgabenstellungen geeignet. Jede Gruppe von Verfahren und jedes Verfahren selbst hat seine speziellen Anwendungsgebiete. Eine falsche Wahl des Rekonstruktionsverfahrens kann zu fehlerhaften Tiefenwerten führen. Keine der beschriebenen Methoden erreicht eine hundertprozentige Rekonstruktion der Tiefenwerte. Dies gilt insbesondere für die SFS-Verfahren. Sie arbeiten nur unter einer großen Zahl von Einschränkungen zufriedenstellend. Für den generellen Einsatz sind sie somit nicht empfehlenswert. Ein ebenfalls schlechtes Ergebnis wies das Lambert-N Verfahren auf. Dieses arbeitet nur unter sehr guten Voraussetzungen befriedigend. Die einzigen zufriedenstellenden Rekonstruktionsergebnisse

wiesen das Lambert-Z Verfahren und die Flächenapproximation auf. Dennoch sollte auch hier das Rekonstruktionsergebnis mit Vorsicht behandelt werden.

Zusammengefaßt kann gesagt werden, daß die 3D-BV auf Basis eines Grauwertbildes mit vereinzelt Tiefenwerten viele neue Lösungsansätze ermöglicht, an denen die herkömmliche Graubildverarbeitung scheitert. Bereits mit Hilfe von vereinzelt Entfernungswerten in Verbindung mit einem vollständig vorliegenden Grauwertbild wird die 3D-BV wesentlich vereinfacht bzw. erst ermöglicht. Im Gegensatz zu diesen guten Ergebnissen stellt die Ausweitung der vereinzelt Tiefenwerte auf Gebiete ohne Tiefenwerte nur ein befriedigendes Ergebnis dar. Solche Verfahren können nur unter idealen Bedingungen und bekannten Szenen genutzt werden.

Ausblick

In dieser Arbeit konnte nur ein Teil der 3D-BV dargestellt werden. Viele Teilaspekte, wie beispielsweise die Komprimierung der Tiefendaten, die Objekt- oder die Szeneninterpretation, wurden hier nicht weiter berücksichtigt. Neben den zwei in Kap. 8.1.2 vorgestellten Verfahren zur Geschwindigkeitsmessung und zur Innenraumüberwachung bietet das 2D/3D-Kamerakzept viele weitere interessante Lösungsansätze, wie z.B. die Gesichtserkennung mit Tiefenwerten, die Bewegungsanalyse von 3D-Objekten oder die Konturvermessung von bewegten Objekten.

Im Vergleich zum menschlichen Sehen ist die 2D-BV wie auch die 3D-BV immer auf eine bestimmte Aufgabenstellung zugeschnitten. Änderungen beispielsweise des Szeneninhaltes, der Objektformen oder der Beleuchtungsverhältnisse erschweren umgehend die Interpretation der aufgenommenen Daten. Ein Ausweg könnte die Künstliche Intelligenz (KI) bieten. Durch eine Beschreibung der (Teil-)Objekte anhand einer Wissensbasis kann mit Hilfe der KI eine Vorabsegmentierung und Interpretation der Szene erfolgen. Unbekannte oder neue Objekte können durch einen einfachen Lernprozeß hinzugefügt werden. Die eigentlichen Vermessungsaufgaben, die Entfernungbestimmung, die Hinderniserkennung oder die Auswertung des Szeneninhaltes können aufbauend auf die Ergebnisse der KI in einem anschließenden Prozeß auf herkömmliche Weise erfolgen. Diese Kombination von KI und herkömmlicher BV ermöglicht folglich die Bereitstellung einer universellen Grundstruktur zur Lösung der meisten Bildverarbeitungsproblematiken.

Literaturverzeichnis

- [Ab94] W.Abmayr; "Einführung in die digitale Bildverarbeitung"; B.G.Teubner Stuttgart; 1994
- [AgCh89] J.K.Aggarwal, C.C.Chu; "The Issues, Analysis, and Interpretation of Multi-Sensor Images"; Multisensor Fusion for Computer Vision; Springer Verlag; Page 37-62; 1989;
- [BeJa85] P.J.Besl, R.C.Jain; "Three-Dimensional Object Recognition"; Computer Surveys; Vol. 17; No. 1; Page 76-145; 1985
- [BeJa88] P.J.Besl, R.C.Jain; "Segmentation through Variable-Order Surface Fitting"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(10); No. 2; Page 167-192; 1988
- [BeSc93] Bergmann, Schaefer; "Optik, Lehrbuch der Experimentalphysik"; Walter de Gruyter Verlag; Band 3; Auflage 9; 1993
- [BiPe92] M.Bichsel, A.P.Pentland; "A Simple Algorithm for Shape from Shading"; IEEE Computer Vision and Pattern Recognition; CVPR-92; Page 459-465; 1992
- [Br88] A.M.Bruckstein; "On Shape from Shading"; Computer Vision Graphics and Image Processing; CVGIP(44); No. 2; Page 139-154; 1988
- [BrSe81] I.N.Bronstein, K.A.Semendjajew; "Taschenbuch der Mathematik"; Verlag Harri Deutsch; 1981
- [BuScRi98] B.Buxbaum, R.Schwarte, T.Ringbeck, H.Heinol, Z.Xu, J.Olk, W.Tai, Z.Zhang, X.Luan; "A new approach in optical broadband communication systems: A highly integrated optical phase locked loop based on a mixing and correlating sensor, the Photonic Mixer Device (PMD)"; Tagungsband Opto98 - Applications of Optical Measuring Techniques; Kap. 3.3; 1998
- [Ca86] J.Canny; "A Computational Approach to Edge Detection"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(8); No. 6; Page 679-698; 1986
- [CeSh95] C.Cédras, M.Shah; "Motion-Based Recognition: A Survey"; Image and Vision Computing; IVC(13); No. 2; Page 129-155; 1995
- [ChAg91] C.Chu, J.K.Aggarwal; "The Interpretation of Laser Radar Images by a Knowledge Based System"; Machine Vision and Applications; MVA(4); No. 3; Page 145-163; 1991
- [ChHi95] S.Christie; S.L.Hill, B.Bury, J.O.Gray, K.M.Booth; "Design and Development of a Multi-Detecting Two-Dimensional Ranging Sensor"; Meas. Sci. Technol.; Vol. 6; IOP Publishing Ltd; Page 1301-1308; 1995
- [ChKa90] Y.Choë, R.L.Kashyap; "Shape from Texture and Shaded Surface"; International Conference on Pattern Recognition; ICPR-90; Vol-I; Page 294-296; 1990
- [ChNa90] C.C.Chu, N.Nandhakumar, J.K.Aggarwal; "Image Segmentation using Laser Radar Data"; Pattern Recognition; PR(23); No. 6; Page 569-581; 1990
- [Cl71] M.B.Clowes; "On Seeing Things"; Artificial Intelligence; AI(2); No. 1; Page 79-116; 1971

- [CoTi95] M.Corbato, S.Tinonin, E.Trucco, V.Roberto; "Dense Depth Maps from Motion using Dynamic Data Fusion"; Image Processing and its Applications; Conference Publication No. 410; Page 153-157; 1995
- [CrTs92] J.E.Cryer, P.S.Tsai, M.Shah; "Combining Shape from Shading and Stereo Using Human Vision Model"; Department of Computer Science; Technical Report; CS-TR-92-25; 1992
- [CrTs95] J.E.Cryer, P.S.Tsai, M.Shah; "Integration of Shape from Shading and Stereo"; Pattern Recognition; PR(28); No. 7; Page 1033-1043; 1995
- [DuNi79] R.O.Duda, D.Nitzan, P.Barret; "Use of Range and Reflectance Data to Find Planar Surface Regions"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(1); No. 3; Page 259-271; 1979;
- [EP98] Elektronik Praxis; "Intelligent Kombiniert / Neuronale Netze und Fuzzy Verfahren lösen komplexe Aufgaben"; Elektronik Praxis Nr. 23; Seite 90-94; 1998
- [ESC98] Edmund Scientific Company; "Produktkatalog / Laser Accessories"; Seite 200; 1998
- [Fa93] O.D.Faugeras; "Three-Dimensional Computer Vision / A Geometric Viewpoint"; MIT Press; Seite 663 ff.; 1993
- [FaMe87] T.J.Fan, G.G.Medioni, R.Nevatia; "Segmented Description of 3D-Surfaces"; IEEE Trans. Robotics and Automation; RA(3); No. 6; Page 527-538; 1987
- [FoDa94] J.Foley, A.v.Dam, S.Feiner, J.Hughes, R.Phillips; "Grundlagen der Computergraphik / Einführung, Konzepte, Methoden"; Addison-Wesley Publishing Company; 1994
- [FrGa96] K.H.Franke, F.Gaßmann, D.Kollhoff, T.Machleidt; "Verfahren zur Bearbeitung von 3D-Daten"; Symposium 'Aktuelle Entwicklungen und industrieller Einsatz der Bildverarbeitung'; Aachen; MIT GmbH; 5/6.9.1996
- [FrSh95] C.S.Fraser, M.R.Shortis, G.Ganci; "Multisensor system self-calibration"; SPIE; Videometrics IV; Vol. 2598; Page 2-18; 1995
- [FtpMsu99] Tiefenbildersammlung ftp://ftp.cps.msu.edu/pub/prip/data/range_images; 1999
- [FtpUcf99] Bildersammlung <ftp://eustis.cs.ucf.edu/pub/source/shading/ucfimgs>; 1999
- [FtpUtah99] Tiefenbildersammlung <ftp://ftp.cs.utah.edu/pub/range-database>; 1999
- [Ga90] J. Gårding; "Shape from Texture and Contour by Weak Isotropy"; International Conference on Pattern Recognition; ICPR-90; Vol-I; Page 324-330; 1990
- [Ge94] R.Gerdes; "Dreidimensionales Messen mit einer Videokamera unter Ausnutzung geometrischer Randbedingungen"; Dissertation Universität-GH Siegen; VDI Verlag; Reihe 8; Nr 435; 1994
- [GiMi83] B.Gil, A.Mitiche, J.K.Aggarwal; "Experiments in Combining Intensity and Range Edge Maps"; Computer Vision, Graphics, and Image Processing; CVGIP(21); No. 3; Page 395-411; 1983
- [GM94] S.Ghosal, R.Mehrotra; "Detection of Composite Edges"; IEEE Trans. Image Processing; IP(3); No. 1; Page 14-25; 1994

- [GoLe89] G.D.Godin, M.D.Levine; "Structured Edge Map of Curved Objects in a Range Image"; IEEE Conference on Computer Vision and Pattern Recognition; CVPR-89; Page 276-281; 1989
- [GuKr91] S.Gupta, R.Krishnapruam; "Morphologic Edge Detection in Range Images"; SPIE; Image Algebra and Morphological Image Processing II; Vol. 1568; Page 335-346; 1991
- [Hal92] D.L.Hall; "Mathematical Techniques in Multi-Sensor Data Fusion"; Artech House Inc; 1992
- [Han92] A.Hanczak; "Fusion of Range and Intensity Data in a Scanning Sensor"; SPIE; Optics, Illumination and Image Sensing for Machine Vision VI; Vol. 1614; Page 231-239; 1992
- [Ha94] G.Hauske; "Systemtheorie der visuellen Wahrnehmung"; Teubner Stuttgart; 1994
- [He89] E.Hecht; "Optik"; Deutsche Übersetzung Addison-Wesley Publishing Company; 1989
- [HeXu97] H.Heinol, Z.Xu, R.Schwarte, T.Ringbeck; "Experimentelle Ergebnisse eines neuartigen schnellen 3D-Meßsystems mit modifizierten elektrooptischen Modulationskomponenten"; Tagungsband 'Optische Formerfassung'; DGfFP-VDI/VDE-GMA; GMA-Bericht 30; Seite 99-103; 1997
- [HeXu99] H.Heinol, Z.Xu, J.Schulte; "Laufzeitbasierende 3D-Kamerasysteme – Smart-Pixel-Lösungen"; Tagungsband 'Optische Formerfassung'; DGfFP-VDI/VDE-GMA; GMA-Bericht 70; Beiheft; 1999
- [Ho70] B.K.P.Horn; "Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View"; Ph.D. Dissertation of B.K.P.Horn ; Department of Electrical Engineering; MIT; 1970
- [Ho77] B.K.P.Horn; "Understanding Image Intensities"; Artificial Intelligence; AI(8); No. 2; Page 201-231; 1977
- [Ho90] B.K.P.Horn; "Height and Gradient from Shading"; International Journal of Computer Vision; IJCV(5); No. 1; Page 37-76; 1990
- [Ho96] A.Hoover; "The Space Envelope Representation for 3D Scenes"; Ph.D Dissertation of Adam Hoover; Visual Computing Lab; University of South Florida; 1996
- [HoBr85] B.K.P.Horn, M.J.Brooks; "Shape and Source from Shading"; International Joint Conference on Artificial Intelligence; IJCAI-85; Page 932-936; 1985
- [HoBr89] B.K.P.Horn, M.J.Brooks; "Shape from Shading"; Cambridge: MIT Press; 1989
- [HoBr89b] B.K.P.Horn, M.J.Brooks; "The Variational Approach to Shape from Shading"; Computer Vision, Graphics and Image Processing; CVGIP(33); No. 2; Page 174-208; 1989
- [HoJa87] R.Hoffman, A.K.Jain; "Segmentation and Classification of Range Images"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(9); No. 5; Page 608-620; 1987
- [HöJe98] H.Höfler, V.Jetter, E.Wagner; "3D-Profiling by Optical Demodulation with an Image Intensifier"; Tagungsband Opto98 – Optische Sensorik Meßtechnik & Elektronik; Kap. 8.1; Page 191-195; 1998

- [Hu71] D.A.Huffman; "Impossible Objects as Non-Sense Sentences"; Machine Vision; MI(6); Page 295-323; Edinburgh University Press; 1971
- [IkHo81] K.Ikeuchi, B.K.P.Horn; "Numerical Shape from Shading and Occluding Boundaries"; Artificial Intelligence; AI(17); Page 141-184; 1981
- [Jä93] B.Jähne; "Digitale Bildverarbeitung"; Springer Verlag; 1993
- [JäHaGe99] B.Jähne, H.Haußecker, P.Geißler; "Handbook of Computer Vision and Applications"; Academic Press; 1999
- [Jan93] D.Jansen; "Optoelektronik: Grundlagen, Bauelemente, Übertragungstechnik, Netzwerke und Bussysteme"; Friedr. Vieweg & Sohn Verlagsgesellschaft; 1993
- [Jar83] R.A.Jarvis; "A Perspective on Range Finding Techniques for Computer Vision"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(5); No. 2; Page 122-139; 1983
- [Jar93] Jarvis; "Range Sensing for Computer Vision"; 3D Object Recognition Systems, Advances in Communication 1; Elsevier Science Publishers B.V.; 1993
- [JeAy91] J.L.Jezouin, N.Ayache; "Three-Dimensional Fusion from a monocular sequence of images"; Multisensor Fusion for Computer Vision; Springer Verlag; Page 155-167; 1991
- [JiBu93] X.Y.Jiang, H.Bunke; "Range Data Acquisition by Coded Structured Light: Error Characteristics of Binary and Gray Projection Code"; SPIE; Optical 3D Measurement techniques II; Vol. 2252; Page 386-393; 1993
- [JiBu97] X.Jiang, H.Bunke; "Dreidimensionales Computersehen / Gewinnung und Analyse von Tiefenbildern"; Springer Verlag; 1997
- [JuHa99a] D.Justen, K.Hartmann; "3D-Image Processing through Grayscale Images and Distance Data Related to Individual Pixels"; SPIE; Visual Information Processing VIII; Vol. 3716; 1999
- [JaHa99b] D.Justen, K.Hartmann, R.Schwarte; "3D-Bildverarbeitung auf Basis eines Grauwertbildes mit vereinzelt Entfernungsdaten"; Tagungsband 'Optische Formfassung'; DGFfP-VDI/VDE-GMA; GMA-Bericht 70; Seite 249-258; 1999
- [Ka88] T.Kasvand; "Extraction of Edges in 3D Range Images to Subpixel Accuracy"; International Conference on Pattern Recognition; ICPR-88; Page 93-98; 1988
- [KiBr95a] R.Kimmel, A.M.Bruckstein; "Tracking Level Sets by Level Sets: A Method for Solving the Shape from Shading Problem"; Computer Vision and Image Understanding; CVIU(62); No. 1; Page 47-58; 1995
- [KiBr95b] R.Kimmel, A.M.Bruckstein; "Global Shape from Shading"; Computer Vision and Image Understanding; CVIU(62); No. 3; Page 360-369; 1995
- [KiKa95] R.Kimmel, K.Siddiqi, B.B.Kimia, A.M.Bruckstein; "Shape from Shading: Level Set Propagation and Viscosity Solutions"; International Journal of Computer Vision; IJCV(16); No. 2; Page 107-133; 1995
- [KI93] R.Klein; "Ein laseroptisches Entfernungsmeßverfahren mit frequenzvariabler Pseudo-Noise-Modulation"; Dissertation Universität-GH Siegen; 1998

- [KIKo95] R.Klette, A.Koschan; "Surface Reconstruction based on Visual Information"; Technical Report 95/6; Computer Vision Group FR 3-11; Department of Computer Science; 1995
- [KIKo96] R.Klette, A.Koschan; "Computer Vision / Räumliche Information aus digitalen Bildern"; Vieweg Technik; 1996
- [Le90] R.Lenz; "Grundlagen der Videometrie, angewandt auf eine ultrahochauflösende CCD-Farbkamera"; Technisches Messen; Nr. 57; Oldenbourg Verlag; 1990
- [LeRo85] C.H.Lee, A.Rosenfeld; "Improved Methods of Estimating Shape from Shading Using the Light Source Coordinate System"; Artificial Intelligence; AI(26); Page 125-143; 1985
- [Lo90a] O.Loffeld; "Estimationstheorie I / Grundlagen und stochastische Konzepte"; Oldenbourg Verlag; 1990
- [Lo90b] O.Loffeld; "Estimationstheorie II / Anwendungen – Kalman-Filter"; Oldenbourg Verlag; 1990
- [Lo90c] O.Loffeld; "Multisensorielle Signalverarbeitung / Data-Fusion und Kalman-Filter"; Hochschulinterne Interne Veröffentlichung Uni-GH-Siegen; 1990
- [Lu93] J.Lu; "Ein dreidimensionales Bildverarbeitungssystem für die Automatisierung visueller Prüfvorgänge"; IPA-IAO 'Forschung und Praxis' Band 180; Springer Verlag; 1993
- [Ma87] J.Malik; "Interpreting Line Drawings of Curved Objects"; International Journal of Computer Vision; IJCV(1); No. 1; Page 73-103; 1987
- [MaAg85] M.J.Magee, J.K.Aggarwal; "Using Multisensory Images to Derive the Structure of Three-Dimensional Objects: A Review"; Computer Vision, Graphics, and Image Processing; CVGIP(32); No. 2; Page 145-157; 1985
- [MaBo85] M.J.Magee, B.A.Boyter, C.H.Chien, J.K.Aggarwal; "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(7); No. 6; Page 629-637; 1985
- [MaPr96] "The Marble Project / Interactive Vision"; Interactive computer-based vision course; Joint project of Edinburgh, Heriot Watt, and Napier Universities; <http://www.icbl.hw.ac.uk/marble/vision/>
- [MaSz91] L.Matthies, R.Szeliski, T.Kanade; "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences"; Multisensor Fusion for Computer Vision; Springer Verlag; Page 87-125; 1991
- [MaWo91] T.A.Mancini, L.B.Wolff; "3D-Shape and Source Location from Depth and Reflectance"; SPIE; Optics, Illumination, and Image Sensing for Machine Vision VI; Vol. 1614; Page 87-98; 1991
- [NaJa92] S.G.Nadabar, A.K.Jain; "Edge Detection and Labeling by Fusion of Intensity and Range Images"; SPIE; Applications of Artificial Intelligence X; Vol. 1708; Page 108-119; 1992
- [NiBr77] D.Nitzan, A.E.Brain, R.O.Duda; "The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis"; Proceedings of the IEEE; PIEEE(65); No. 2; Page 206-220; 1977

- [OI91] J.Oliensis; "Shape from Shading as a Partially Well-Constrained Problem"; Computer Vision Graphics and Image Processing; CVGIP(54); No. 2; Page 163-183; 1991
- [OI97] J.Olk; "Untersuchung von Laufzeitentfernungsmeßsystemen unter besonderer Berücksichtigung des Referenzproblems"; Dissertation Universität-GH Siegen; Shaker Verlag; ZESS-Forschungsberichte; Bd. 5; 1997
- [OIDu91] J.Oliensis, P.Dupuis; "Direct Method for Reconstructing Shape from Shading"; SPIE; Geometric Methods in Computer Vision; Vol. 1570; Page 116-128; 1991
- [Ot95] R.Otterbach; "Robuste 3D-Objekterkennung und Lagebestimmung durch Auswertung von 2D-Bildfolgen"; Dissertation Universität-GH Siegen; VDI Verlag; Reihe 8; Nr. 481; 1995
- [Pe84] A.P.Pentland; "Local Shading Analysis"; IEEE Trans. Pattern Anal. Machine Intelligence; PAMI(6); No. 2; Page 170-187; 1984
- [Pe88] A.P.Pentland; "On The Extraction of Shape Information from Shading"; National Conference on Artificial Intelligence; AAAI-88; Page 826-830; 1988
- [Pe90] A.P.Pentland; "Linear Shape from Shading"; International Journal of Computer Vision; IJCV(4); No. 2; Page 153-162; 1990
- [Pe93] Perceptron Inc.; "LASAR Hardware Manual"; 23855 Research Drive; Farmington Hills; Michigan 48335; 1993
- [PeBi94] A.P.Pentland, M.Bichsel; "Extracting Shape from Shading"; Handbook of Pattern Recognition and Image Processing; HPRIP-CV94; Page 161-183; 1994
- [PiGa95] H.H.Pien, J.M.Gauch; "A Variational Approach to Multisensor Fusion of Images"; Applied Intelligence; ApplIntel(5); No. 3; Page 217-235; 1995
- [Po75] B.T.Phong; "Illumination for Computer Generated Images"; Communication of the ACM(18); No. 6; Page 311-317; 1975
- [PöCl94] J.Pöpsel, U.Claussen, R.Klein, J.Plate; "Computergrafik / Algorithmen und Implementierung"; Springer Verlag; 1994
- [Ri96] D.Ritter; "Merkmalsorientierte Objekterkennung und -lokalisierung im 3D-Raum aus einem einzelnen 2D-Grauwertbild und Referenzmodellvermessung mit optischen 3D-Sensoren"; Dissertation Universität Erlangen-Nürnberg; 1996
- [RiSt96] R.Richmond, R.Stettner, H.Bailey; "Laser Radar Focal Plane Array for Three-Dimensional Imaging"; SPIE; Laser Radar Technology and Applications; Vol. 2748; Page 61-67; 1996
- [RiSt97] R.Richmond, R.Stettner, H.Bailey; "Laser Radar Focal Plane Array for Three-Dimensional Imaging: Update"; SPIE; Laser Radar Technology and Applications II; Vol. 3065; Page 177-181; 1997
- [Ro86] A.Rosenfeld; "Techniques for 3D Machine Perception / Machine Intelligence and Pattern Recognition 3"; Elsevier Science Publishers B.V.; 1986
- [Ro93] R.Robmann, H.Bunke; "Ein kantenorientierter Ansatz zur Integration von Grauwert und Tiefenbildern"; Technical Report; IAM 93-016; 1993

- [Ro94] R.Robmann; "Interpretation von Schatten- und Intensitätskanten in integrierten Grauwert- und Tiefenbildern"; Technical Report; IAM 94-007; 1994
- [RoBu86] R.Robmann, H.Bunke; "An Edge Labeling Scheme for Polyhedra in incomplete Range images"; Proceedings of the Scandinavian Conference on Image Analysis; Page 53-64; 1986
- [Sc82] R.Schark; "Vektoranalysis und Funktionentheorie / Eine Einführung für Ingenieure und Naturwissenschaftler"; Verlag Chemie GmbH; 1982
- [Sc90] N.Schneider; "Kantenhervorhebung und Kantenvorfolgung in der industriellen Bildverarbeitung"; Vieweg Verlag; 1990
- [ScHe97] R.Schwarte, H.Heinol, Z.Xu, J.Olk, W.Tai; "Schnelle und einfache optische Formerfassung mit einem neuartigen Korrelations-Photodetektor-Array"; Tagungsband 'Optische Formerfassung'; DGFFP-VDI/VDE-GMA; GMA-Bericht 30; Seite 199-209; 1997
- [ScHe98] R.Schwarte, H.Heinol, B.Buxbaum, Z.Xu, T.Ringbeck; "Neuartige 3D-Visionssysteme auf der Basis Layout-optimierter PMD-Strukturen"; tm-Technisches Messen 65 (1998); H. 7-8; Seite 264-271; 1998
- [Scu96] E.Schubert; "Mehrfach farbcodierte Triangulationsverfahren zur topometrischen Erfassung und Vermessung von 3D-Objekten"; Dissertation Universität-GH Siegen; Shaker Verlag; ZESS-Forschungsberichte; Bd. 1; 1996
- [Scw96] R.Schwarte; "Verfahren und Vorrichtung zur Bestimmung der Phasen- und/oder Amplitudeninformation einer elektromagnetischen Welle"; Patentschrift; Schwarte Rudolf Prof. Dr. Ing. (DE); 1996
- [Scw96a] R.Schwarte; "Eine neuartige 3D-Kamera auf Basis eines 2D-Gegentaktkorrelator-Arrays"; Tagungsband Symposium 'Aktuelle Entwicklungen und industrieller Einsatz der Bildverarbeitung'; MIT GmbH; Seite 273-278; Aachen; 1996
- [Scw97] R.Schwarte; "Überblick und Vergleich aktueller Verfahren der optischen Formerfassung"; Tagungsband 'Optische Formerfassung'; DGFFP-VDI/VDE-GMA; GMA-Bericht 30; Seite 1-12; 1997
- [Scw99a] R.Schwarte; "3D-Bildsensorik"; Tagungsband '10. Heidelberger Bildbearbeitungsforum' Schwerpunktthema: 3D-Bildaufnahme und -verarbeitung; Siegen; 1999;
- [Scw99b] R.Schwarte; "Ein Jahrzehnt 3D-Bilderfassung- Zielsetzungen, Fortschritte, Zukunftsperspektiven"; Tagungsband 'Optische Formerfassung'; DGFFP-VDI/VDE-GMA; GMA-Bericht 70; Beiheft; 1999
- [Se95] B.Scheffe; "Schnelle dreidimensionale Objektvermessung durch die Auswertung strukturierter Lichtmuster mit bildverarbeitenden Methoden"; Studienarbeit Universität-GH-Siegen, Zentrum für Sensorsysteme; 1995
- [Sh87] Y.Shirai; "Three-Dimensional Computer Vision"; Springer Verlag; 1987
- [ShKi96] I.Shimshoni, R.Kimmel, A.M.Bruckstein; "Dialogue: Global Shape from Shading"; Computer Vision and Image Understanding; CVIU(64); No. 1; Page 188-189; 1996

- [ShPiGa96] J.Shah, H.H.Pien, J.M.Gauch; "Recovery of Surfaces with Discontinuities by Fusing Shading and Range Data within a Variational Framework"; IEEE Trans. Image Processing; IP(5); No. 8; Page 1243-1251; 1996
- [St91] B.Straub; "Ein Verfahren zur Rekonstruktion von dreidimensionalen Objektmodellen aus digitalen Bildern"; Dissertation; Deutsche Geodätische Kommission; Verlag der Bayerischen Akademie der Wissenschaften in Kommission bei der C.H.Beckschen Verlagsbuchhandlung München; 1991
- [SuBh92] M.Suk, S.M.Bhandarkar; "Three-Dimensional Object Recognition from Range Images"; Springer Verlag; 1992
- [Sz91] R.Szeliski; "Fast Shape from Shading"; Computer Vision Graphics and Image Processing; CVGIP(53); No. 2; Page 129-153; 1991
- [Te84] D.Terzopoulos; "Multiresolution Algorithms in Computational Vision"; Image Understanding; IU-84; Page 225-262; 1984
- [TeCh89] C.H.Teh, R.T.Chin; "On the Detection of Dominant Points on Digital Curve"; IEEE Trans. Pattern Analysis and Machine Intelligence; PAMI(11); No. 8; Page 859-872; 1989
- [TsSh92] P.S.Tsai, M.Shah; "A fast linear Shape from Shading"; IEEE Computer Conference on Computer Vision and Pattern Recognition; CVPR-92; Page 734-736; 1992
- [TsSh94] P.S.Tsai, M.Shah; "Shape from Shading Using Linear Approximation"; Image and Vision Computing; IVC(12); No. 8; Page 487-498; 1994
- [Wa75] D.Waltz; "Understanding Line Drawings of Scenes with Shadows"; The psychology of Computer Vision; Page 19-91; McGraw-Hill; 1975
- [Wa84] F.M.Wahl; "Digitale Bildsignalverarbeitung / Grundlagen, Verfahren, Beispiele"; Springer Verlag; Nachrichtentechnik; Bd. 13; 1984
- [Wa93] A.Watt; "3D Computer Graphics"; Addison-Wesley Publishers Ltd.; 1993;
- [We89] J.Weil; "Modellgesteuerte Szenen-Interpretation durch Fusion von Intensitäts- und Abstandsbildern"; Dissertation von J.Weil; Universität Karlsruhe; Fakultät für Informatik; 1989
- [We98] J.Weber; "Ein visuell unterstütztes, laseroptisches Multisensorsystem zur automatisierten Erfassung dreidimensionaler Objekte"; Dissertation Universität-GH Siegen; Shaker Verlag; ZESS-Forschungsberichte; Bd. 13; 1998
- [WeLe89] J.Weil, P.Levi, U.Rembold; "Fusion of Range and Intensity Image Data for Recognition of 3D object surfaces"; Multisensor Fusion for Computer Vision; Springer Verlag; Page 171-194; 1989
- [WeLe90] J.Weil, P.Levi; "Entwicklung dreidimensionaler Fusionstechniken für Abstands- und Intensitätsbilder auf Basis eines doppelt synchronisierten Laser-Scanners"; Mustererkennung 1990; 12. DAGM-Symposium; Seite 316-323; 1990
- [Wi94] N.Wilt; "Object-Oriented Ray Tracing in C++"; Wiley Professional Computing; 1994

- [Wo80] R.J.Woodham; "Photometric Method for Determining Surface Orientation from Multiple Images"; Optical Engineering; OptEng(19); No. 1; Page 139-144; 1980
- [WwwMa99] Tiefenbildersammlung <http://marathon.csee.usf.edu/range/seg-comp/SegComp.html>; Range image segmentation comparison; 1999
- [XuSc98] Z.Xu, R.Schwarte, H.Heinol, B.Buxbaum, T.Ringbeck; "Smart Pixel – Photometric Mixer Device (PMD) / New System Concept of a 3D-Imaging Camera-on-a-Chip"; 5th International Conference on Mechatronics and Machine Vision in Practice; Page 259-264; Nanjing; 1998
- [Zh95] R.Zhang; "Shape from Photomotion"; Ph.D. Disseration of R.Zhang; University of Central Florida; 1995
- [ZhAu92] G.Zhang, W.J.Austin, A.M.Wallace; "Combining Depth and Intensity Processing for Scene Description and Object Recognition"; SPIE; Sensor Fusion V; Vol. 1828; Page 49-59; 1992
- [ZhCh91] Q.Zheng, R.Chellappa; "Estimation of Illuminant Direction, Albedo, and Shape from Shading"; IEEE Trans. Pattern Analysis and Machine Intelligence; PAMI(13); No. 7; Page 680-702; 1991
- [ZhSh94] R.Zhang, M.Shah; "Height Recovery from Intensity Gradient"; IEEE Computer Vision and Pattern Recognition; CVPR-94; Page 508-513; 1994
- [ZhTs94] R.Zhang, P.S.Tsai, J.E.Cryer, M.Shah; "Analysis of Shape from Shading Techniques"; IEEE Computer Vision and Pattern Recognition; CVPR-94; Page 377-384; 1994
- [ZhZh95] X.Zhang,D.Zhao; "Range Image Segmentation via Edges and Critical Points"; Spie; Visual Communications and Image Processing '95; Vol. 2501; Page 1626-1637; 1995

Abkürzungsverzeichnis

Abkürzung	Bedeutung
2-Q-IQ-PMD	2-Quadranten Inphase/Quadraturphase PMD
4-Q-IQ-PMD	4-Quadranten Inphase/Quadraturphase PMD
2D/3D-Kamerasystem	Kamerasystem bestehend aus einem flächendeckenden Graubildaufnahme und vereinzelt Tiefenwerten
AD-Wandler	Analog-Digital Wandler
APS	Aktiver Pixel Sensor
BV	Bildverarbeitung
CAD	Computer Aided Design
CCD	Charge Coupled Device
CDMA	Code-division-multiple-access
CMOS	Complementary Metal Oxide Semiconductor
CW-Modulation	Continuos-Wave Modulation
EB	Eigenbeleuchtung
FF	Füllfaktor
FFT	Fast Fourier Transformation
FL	Fremdlicht
FTS	Fahrerloses Transportsystem
IQ-PMD	Inphase/Quadraturphase PMD
KI	Künstliche Intelligenz
KKS	Kamerakoordinatensystem
LRP	Laser-Radar-Processor
OOP	Out-of-Position
OORT	Object-Oriented Ray Tracer
Pixel	Picture Element
PMD	Photonic Mixer Device
PN-Modulation	Pseudo-Noise Modulation
RKS	Rechnerkoordinatensystem
SFS	Shape-from-Shading

Abkürzung	Bedeutung
SFT	Shape-from-Textur
SKS	Sensorkoordinatensystem
TOF	Time of Flight
VHDL	Verilog Hardware Description Language
Voxel	Volume Element
V V	Verschiebungsvektoren
WKS	Weltkoordinatensystem

Formelverzeichnis

Abkürzung	Bedeutung	Kap.
x, y, z	Koordinate im 3D-Raum	3,6,8
I	Intensität	3,6,8
I_i	Eingangsintensität	3,6
κ	Kamerakonstante	3,6
Φ	Reflexionsfunktion	3,6
$\mathbf{n}(x, y, z)$	Flächennormale	3,6,8
$\mathbf{s}(x, y, z)$	Beleuchtungsrichtung	3,6,8
$\mathbf{v}(x, y, z)$	Aufnahme- bzw. Beobachtungsrichtung	3,6,8
l	Entfernung	3,5
(p, q)	Gradienten	3,6,8
τ	(Echo)Laufzeit	5
a_m, b_m	Modulationsgates des PMD-Elementes	5
G_a, G_b	Kathoden des PMD-Elementes	5
u_o	Offsetspannung	5
u_m	Modulationsspannung	5
ΔU_{ab}	PMD-Differenzspannung	5
ΣU_{ab}	PMD-Summenspannung	5
X_W, Y_W, Z_W	Weltkoordinatensystem	6
X_K, Y_K, Z_K	Kamerakoordinatensystem	6,8
X_S, Y_S, Z_S	Sensorkoordinatensystem	6
I_R, J_R	Rechnerkoordinatensystem	6
$\mathbf{P}_{WP} = (x_{WP}, y_{WP}, z_{WP})$	Punkt im 3D-Weltkoordinatensystem	6
$\mathbf{P}_{KP} = (x_{KP}, y_{KP}, z_{KP})$	Punkt im 3D-Kamerakoordinatensystem	6
$\mathbf{P}_{SP} = (x_{SP}, y_{SP})$	Punkt im 2D-Sensorkoordinatensystem	6
$\mathbf{P}_{RP} = (i_{RP}, j_{RP})$	Punkt im 2D-Rechnerkoordinatensystem	6
$(x_b, y_b, z_b)^T$	Translationsvektor zwischen dem WKS und dem KKS	6
ψ, θ, φ	Rotationswinkel um die XYZ-Achse im Kamerakoordinatensystem	6

Abkürzung	Bedeutung	Kap.
b	Bildweite	6,8
i_{max}, j_{max}	Anzahl der Elemente im Rechnerkoordinatensystem	6
s_x, s_y	Pixelbreite im Sensorkoordinatensystem	6
l_{SENSOR}	gemessene Objektentfernung von der Sensorfläche über den Meßpunkt bis zur Beleuchtungsquelle	6,8
l_{SENSOR}'	Objektentfernung vom Nullpunkt des KKS über den Meßpunkt bis zur Beleuchtungsquelle	6,8
$l_{P'}$	Entfernung von der Sensorfläche bis zum Nullpunkt des KKS	6,8
l_P	Entfernung vom Nullpunkt des KKS bis zum Meßpunkt (tatsächliche Objektentfernung)	6,8
l_{PL}	Entfernung vom Meßpunkt zur Lichtquelle	6,8
$P_L=(x_L, y_L, z_L)$	Position der Beleuchtungsquelle	6,8
$P_P=(x_P, y_P, z_P)$	Position des Meßpunktes	6,8
$P_{P'}=(x_{P'}, y_{P'}, -b)$	Position des auf die Sensorfläche projizierten Meßpunktes	6,8
α, β, γ	Winkelangaben zur XYZ-Korrektur	6
$T=(l_{i,j})$	Entfernungsmatrix	6
$T=(x_{i,j}, y_{i,j}, z_{i,j})$	XYZ korrigierte Tiefenmatrix	6
$T=(Z_{i,j})$	Tiefenmatrix mit äquidistantem XY-Gitter	6
$T=(I_{i,j})$	Graubildmatrix	6
ε	Fehler	6,7
U	Umgebung	6
m	Polynomgrad	6
$z=Z(x,y)$	Explizite Funktion Z	3,6
$0=F(x,y,z)$	Implizite Funktion F	6
$\mathbf{n}=(x_n, y_n, z_n)$	Flächennormale	6
$ \nabla I $	Steilheit einer Kante	6
φ	Orientierung einer Kante	6
B	Frei wählbare Glättungsmaske	6
D_X, D_Y	Diskrete Ableitungsoperatoren	6
σ	Deklination (slant); Winkel eines Punktes zur z-Achse	6,8

Abkürzung	Bedeutung	Kap.
θ	Azimut (tilt); Winkel eines Punktes in der Ebene	6,8
i	Winkel zwischen der Oberflächennormalen und der Beleuchtungsrichtung (incident angle)	6,8
e	Winkel zwischen der Oberflächennormalen und der Beobachtungsrichtung (emitance angle, emergent angle)	6
d	Winkel zwischen der Beobachterrichtung und der gerichteten Reflexion der Beleuchtung	6
g	Winkel zwischen der Beleuchtungsrichtung und der Beobachtungsrichtung (phase angle)	6
α_d	Diffuser Reflexionskoeffizient	6
α_s	Ambienter Reflexionskoeffizient	6
n	Grad der spiegelnden Reflexion	6
R	Reflektanzfunktion	6,8
(p_s, q_s)	Gradienten der Beleuchtungsrichtung	6,8
ρ	Albedo	6,8
γ	Gewichtungsfaktor	6,7
$(I-R)$	Minimierungsfunktion für Intensitätswerte	7
$(Z-D)$	Minimierungsfunktion für Tiefenwerte	7
FF	Füllfaktor	8
U'	verkleinerte Umgebung von U	8
$F_p(f, \varphi)$	Fouriertransformierte von $p(x, y)$	8
$F_q(f, \varphi)$	Fouriertransformierte von $q(x, y)$	8
$F_I(f, \varphi)$	Fouriertransformierte von $I(x, y)$	8
$F_Z(f, \varphi)$	Fouriertransformierte von $Z(x, y)$	8
$G(\gamma)$	Geradenfunktion in Abhängigkeit von γ	8
$\mathbf{s}=(x_s, y_s, z_s)$	Beleuchtungsrichtung	8

Pseudo-Programme

SFS - Algorithmus von Bichsel & Pentland

Übergabe: Grauwertbild -> Flächendeckendes Grauwertbild

Tiefenbild -> Vereinzelte Z-Tiefenwerte

Beleuchtung -> XYZ-Position der Beleuchtungsquelle

Rückgabe: z -> Rekonstruiertes Tiefenbild

Funktion: Berechnung eines Tiefenbildes auf Basis des SFS-Verfahrens von
Bichsel & Pentland

```
function z=bichsel_pentland(Grauwertbild,Tiefenbild,Beleuchtung)

//Höhenkarte Initialisieren
z=Tiefenbild(Singulären Stellen);

//Rotationswinkel bestimmen
alpha=GetAlpha(Grauwertbild);

//Grauwert- und Tiefenbild um den Winkel alpha rotieren
Grauwertbild=rotate(Grauwertbild,alpha);
z           =rotate(z           ,alpha);

//MxNx8-Bild mit Richtungsableitungen in 8er-Nachbarschaft generieren
dz=DirectionalSlope(Grauwertbild,alpha);

for (Anzahl_Iterationen)
  begin
  for (jeden Bildpunkt vom GrauwertBild)
    begin
    //Inkrementiere Z-Wert mit bester Richtungsableitung
    z=GetHeight(z);
    end;
  //Abtastrichtung des Bildes ändern
  end;

//Tiefenbild um den Winkel alpha zurück rotieren
z=rotate(Grauwertbild,-alpha);

return
```

SFS - Algorithmus von Pentland

Übergabe: Grauwertbild -> Flächendeckendes Grauwertbild

Beleuchtung -> XYZ-Position der Beleuchtungsquelle

Rückgabe: z -> Rekonstruiertes Tiefenbild

Funktion: Berechnung eines Tiefenbildes auf Basis des SFS-Verfahrens von Pentland

```
function z=pentland(Grauwertbild,Beleuchtung)

//Grauwertbild Fourier transformieren
Fi=FFT2(Grauwertbild);

//Konstanten initialisieren
k2=cos(tilt)*sin(slant);
k3=sin(tilt)*sin(slant);

for (jeden Bildpunkt von Fi)
  begin
    //Abstand und Frequenz des zu rekonstruierenden Punktes berechnen
    p=GetAbst();
    f=GetFreq();

    //Transformationsformel anwenden
    Fz(f,p)=Fi(f,p)/(2*pi*f*(exp(1)^(i*pi/2))*(k2*cos(thet)+k3*sin(thet)));
  end;

//Berechnetes Tiefenbild zurück transformieren
z=IFFT2(Fz);

return;
```

SFS - Algorithmus von Tsai & Shah

Übergabe: Grauwertbild -> Flächendeckendes Grauwertbild

Beleuchtung -> XYZ-Position der Beleuchtungsquelle

Rückgabe: z -> Rekonstruiertes Tiefenbild

Funktion: Berechnung eines Tiefenbildes auf Basis des SFS-Verfahrens von
Tsai & Shah

```
function z=Tsai_Shah(Grauwertbild,Beleuchtung)

//Startwerte setzen
Zn1=zeros(size(Grauwertbild),1);

//Max. Anzahl der Iterationen
for (Anzahl_Iterationen)
    begin
        for (jeden Bildpunkt vom Grauwertbild)
            begin
                //Rechenvorschrift anwenden
                Zn(x,y) = Zn1(x,y) - fZ1(x,y) / dfZ1(x,y)
            end;
            //Tiefenmatrix zwischenspeichern
            Zn1=Zn;
        end;
    end;

z=Zn;

return
```

Lambert - Z Verfahren

Übergabe: Grauwertbild -> Flächendeckendes Grauwertbild
 Tiefenbild -> Vereinzelte XYZ-Tiefenwerte
 Beleuchtung -> XYZ-Position der Beleuchtungsquelle

Rückgabe: xyz -> Rekonstruiertes Tiefenbild

Funktion: Berechnung eines Tiefenbildes auf Basis des Lambert-Z Verfahrens

```

function xyz=lambertz(Grauwertbild,Tiefenbild,Beleuchtung)

xyz=zeros(size(Grauwertbild),3); //Für X,Y,Z
anz=zeros(size(Grauwertbild),1); //Für Mittelwertbildung

//Normierung des Grauwertbildes
GrauNorm=(Grauwertbild-Min(Grauwertbild))/(Max(Grauwertbild)-Min(Grauwertbild));

//Alle vorhandenen Tiefenwerte durchlaufen
for (lauf=1 to Anzahl(Tiefenbild))
  begin
    //XY-Matrixposition des Tiefenwertes holen
    MatrixPos=MatrixpositionXY(Tiefenbild(lauf));

    //XYZ-Raumposition des Ausgangswertes holen
    Pp=RaumpositionXYZ(Tiefenbild(lauf));

    //Vorgabewert ausgehend vom Ausgangspunkt berechnen
    lsensor'=abs(Pp)+abs(Beleuchtung-Pp);

    //Alle Punkte innerhalb eines kleinen Nachbarschaftsbereiches durchlaufen
    for (x=MatrixPos.X-5 to MatrixPos.X+5)
      begin
        for (y=MatrixPos.Y-5 to MatrixPos.Y+5)
          begin
            //Grauwert des aktuellen Punktes innerhalb des Toleranzbereiches?
            if((GrauNorm(x,y) > (GrauNorm(MatrixPos.X,MatrixPos.Y))-0.05) and
              (GrauNorm(x,y) < (GrauNorm(MatrixPos.X,MatrixPos.Y))+0.05)) then
              begin
                //Matrixposition des zu rekonstruierenden Punktes in Sensorposition
                //umrechnen
                (xp',yp')=SensorpositionX(x,y);

                //Abstand vom Nullpunkt berechnen
                gamma=(Beleuchtung.X^2+Beleuchtung.Y^2+Beleuchtung.Z^2-lsensor'^2)/
                  (2*(Beleuchtung.X*xp'+Beleuchtung.Y*yp'+Beleuchtung.Z*b-
                    sqrt(xp'^2+yp'^2+b^2)));
                //Ergebnis in Ausgabematrix speichern
                xyz(x,y)=gamma*(xp',yp',-b);
                anz(x,y)=anz(x,y)+1;
              end; //if(..)
            end; //for(y..)
          end; //for(x..)
        end; //for(lauf..)
      end;
    end;

//Mittelwerte aller rekonstruierten Entfernungswerte berechnen
xyz()/=xyz()/anz;
return;

```

Lambert - N Verfahren

Übergabe: Grauwertbild -> Flächendeckendes Grauwertbild

Tiefenbild -> Vereinzelte XYZ-Tiefenwerte

Normalbild -> Vereinzelte Oberflächennormalen

Beleuchtung -> XYZ-Position der Beleuchtungsquelle

Rückgabe: nor -> Rekonstruiertes Orientierungsbild

Funktion: Berechnung eines Tiefenbildes auf Basis des Lambert-N Verfahrens

```

function nor=lambertn(Grauwertbild,Tiefenbild,NormalenBild,Beleuchtung)

//Initialisierung allgemeiner Daten
nor=zeros(size(Grauwertbild),3); //Für Oberflächenorientierung
anz=zeros(size(Grauwertbild),1); //Für Mittelwertbildung

//Normierung des Grauwertbildes
GrauNorm=(Grauwertbild-Min(Grauwertbild))/(Max(Grauwertbild)-Min(Grauwertbild));

//Orientierungsbild aus Grauwertbild berechnen
OrientBild=Orientierung(Grauwertbild);

//Alle vorhandenen Tiefenwerte durchlaufen
for (lauf=1 to Anzahl(NormalenBild))
  begin
    //XY-Matrixposition des Tiefenwertes holen
    MatrixPos=MatrixpositionXY(NormalenBild(lauf));

    //XYZ-Raumposition des Ausgangswertes holen
    Pp=RaumpositionXYZ(TiefenBild(lauf));

    //Phasenwert innerhalb des gültigen Bereiches?
    if((OrientBild.Betrag(MatrixPos) > Schranke) then
      begin
        //Normalenvektor vom aktuellen Meßpunkt zur Beleuchtungsquelle bestimmen
        pl=Normierung(Beleuchtung-Pp);

        //Winkel zwischen Lichtrichtung und Flächennormale bestimmen
        cos_alpha=dot(pl,NormalenBild(lauf));

        //Ausgangswinkel berechnen
        beta_start=atan2(NormalenBild(lauf).Y,NormalenBild(lauf).X);

        //Alle Punkte innerhalb eines kleinen Nachbarschaftsbereiches um den
        //Ausgangspunkt durchlaufen
        for (x=MatrixPos.X-5 to MatrixPos.X+5)
          begin
            for (y=MatrixPos.Y-5 to MatrixPos.Y+5)
              begin

```

```

//Grauwert des aktuellen Punktes innerhalb des Toleranzbereiches?
if((GrauNorm(x,y) > GrauNorm(MatrixPos.X,MatrixPos.Y)-0.05) and
    (GrauNorm(x,y) < GrauNorm(MatrixPos.X,MatrixPos.Y)+0.05)) then
  begin
    //Betrag der Orientierungsberechnung kleiner als Schwelle?
    //->Homogene Grauwertverteilung
    if((OrientBild.Betrag(MatrixPos)< Schwelle) and
        (OrientBild.Betrag(x,y) < Schwelle)) then
      begin
        nor(x,y)=nor(x,y)+NormalenBild(lauf);
        anz(x,y)=anz(x,y)+1;
      end
    else
      begin
        //Winkeldifferenz zwischen Bezugspunkt und zu berechnendem Punkt
        //bestimmen
        beta=beta_start-(OrientBild.Phase(MatrixPos) -
                          OrientBild.Phase(x,y) );

        //XY-Normalenkomponenten des neuen Orientierungswertes berechnen
        erg.x=cos(beta);
        erg.y=sin(beta);

        // Z-Normalenkomponente des neuen Orientierungswertes berechnen
        a=pl.x^2*erg.x^2+pl.y^2*erg.y^2+2*pl.x*erg.x*pl.y*erg.y-
          cos_alpha^2*(erg.x^2+erg.y^2);
        b=2*pl.x*erg.x*pl.z+2*pl.y*erg.y*pl.z;
        c=pl.z^2-cos_alpha^2;

        //Quadratische Gleichung hat zwei Ergebnisse
        erg.z1=(-b+sqrt(b^2-4*a*c))/(2*c);
        erg.z2=(-b-sqrt(b^2-4*a*c))/(2*c);

        //Gültigkeitsabfrage
        if(abs(NormalenBild(lauf).Z-erg.z1) >
            abs(NormalenBild(lauf).Z-erg.z2)) then
          nor(x,y)=nor(x,y)+(erg.x,erg.y,erg.z2)
        else
          nor(x,y)=nor(x,y)+(erg.x,erg.y,erg.z1)
          anz(x,y)=anz(x,y)+1;
        end; //if (..)
      end; //if (..)
    end; //for(y..)
  end; //for(x..)
end; //for(lauf..)

//Mittelwerte aller rekonstruierten Entfernungswerte berechnen
nor()=nor()/anz;

return;

```