

Lehrstuhl für Realzeit-Computersysteme

**Bewegungsstereo in Echtzeit
zur
Exploration und Lokalisation**

Gunter Magin

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.rer.nat. M. Lang

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. G. Färber
2. Univ.-Prof. Dr.rer.nat B. Radig

Die Dissertation wurde am 07.02.2000 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 21.08.2000 angenommen.

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Assistent am Lehrstuhl für Prozeßrechner (heute Lehrstuhl für Realzeit-Computersysteme) der Technischen Universität München. Das Thema ergab sich aus den Arbeiten im interdisziplinären Forschungsprojekt "Informationsverarbeitung in autonomen, mobilen Handhabungssystemen" der Deutschen Forschungsgemeinschaft.

Mein besonderer Dank gilt Prof. Dr.-Ing. G. Färber für die Unterstützung in allen Phasen dieser Arbeit, für seine Geduld, und nicht zuletzt für die Zeit, in der ich am LPR mitarbeiten durfte. Herr Prof. Dr.rer.nat. B. Radig übernahm freundlicherweise das Koreferat.

Sehr hilfreich waren die Diskussionen, die sich mit den damaligen Kollegen am Lehrstuhl entwickelt haben. Dem gesamten Robot-Vision-Team des LPR gebührt Dank für die kollegiale Unterstützung.

J. Quade bin ich zu besonderem Dank verpflichtet.

München, im Februar 2000

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	2
1.2. Begriffsdefinitionen	3
1.3. Gliederung	5
2. Bekannte Bewegungsstereo-Verfahren und -Systeme	6
2.1. Verfahren für Bewegungsstereo	6
2.1.1. Rekonstruktion der Tiefeninformation	6
2.1.2. Interpretation der Bewegung im Bildraum	7
2.1.3. Korrespondenzfindung	8
2.1.4. 3D-Rekonstruktion	9
2.2. Systeme mit Bewegungsstereo	10
2.2.1. Bewegungsstereosystem für punktförmige Szenenmerkmale	10
2.2.2. Bewegungsstereosystem zur Exploration von Innenräumen	11
2.2.3. Rekonstruktion durch axiales Bewegungsstereo	13
2.2.4. Bewegungsstereosystem durch Messung der Verschiebungsgeschwindigkeit ..	13
2.2.5. Exploration mit Bewegungsstereo und Merkmalsverfolgung	14
3. Algorithmen für Online-Bewegungsstereo	16
3.1. Merkmalsextraktion	16
3.1.1. Anforderungen	16
3.1.2. Merkmale in Grauwertbildern	18
3.1.3. Prinzip der Konturverfolgung	19
3.1.4. Konturverfolgung mit einem physikalischen Analogon	21
3.1.5. Feldberechnung mit Subpixelauflösung	28
3.1.6. Abbildung der Feldberechnung auf klassische Operatoren	32
3.1.7. Startpunktsuche und Nutzung von Vorwissen	36
3.1.8. Eckpunkterkennung und Polygonapproximation	38
3.1.9. Bildverzerrung	39
3.1.10. Zusatzattribute	40
3.1.11. Echtzeitaspekte	41
3.2. Korrespondenzfindung	42
3.2.1. Prinzip	42
3.2.2. Merkmalsraum	43
3.2.3. Prädiktion	49
3.2.4. Nachbarschaftsbeziehungen	52
3.3. 3D-Rekonstruktion	54
3.3.1. Schnitt der Sichtgeraden	54
3.3.2. Schnitt der Sichtebenen	59
3.3.3. Strategie	60
3.4. Systemarchitektur	62
4. Anwendungen des Bewegungsstereo	65
4.1. Exploration	65
4.2. Lokalisation	67
4.2.1. Systemmodell	68
4.2.2. Meßmodell	71
4.2.3. Sequentielle Innovation	72
4.2.4. Beiträge des Bewegungsstereo	73

5. Ergebnisse und Bewertung	76
5.1. Merkmalsextraktion	76
5.1.1. Güte der Extraktion	76
5.1.2. Zeitanalyse innerhalb der Vorverarbeitung	78
5.2. Korrespondenzfindung	80
5.2.1. Methodik zur Bewertung der Korrespondenzfindung	80
5.2.2. Korrespondenzfindung mit Lagemerkmale	81
5.2.3. Nutzen der Feldwerte als Zusatzattribute	82
5.2.4. Distanzmaß	84
5.3. 3D-Rekonstruktion	86
5.4. Lokalisation	88
5.5. Implementierung und Verarbeitungszeiten	95
6. Zusammenfassung und Ausblick	98
A. Anhang	100
A.1. Herleitung der Feldmatrizen	100
A.2. Felder eines synthetischen Bildes	104
A.3. Bildbeispiele aus den Experimenten	104
A.4. Kalman-Filter	105
A.5. Koordinatensysteme und Nomenklatur	110
B. Literatur	113

Abbildungsverzeichnis

Abb. 1:	Informationsfluß bei der Sensordatenauswertung	1
Abb. 2:	Auswirkung von Relativbewegungen auf den optischen Fluß	7
Abb. 3:	axiales Stereosystem nach Lavest/Rives/Dhome	13
Abb. 4:	Berechnung der Objektdistanz nach Huber/Gräfe	14
Abb. 5:	Physikalische Ursachen von Kanten in Grauwertbildern	19
Abb. 6:	Grauwertbild	23
Abb. 7:	Betrag des E-Felds	23
Abb. 8:	vertikales E-Feld	24
Abb. 9:	horizontales E-Feld	24
Abb. 10:	H-Feld	24
Abb. 11:	Konturpunkte	24
Abb. 12:	Grauwertprofil für vertikale Rampenkanten unterschiedlicher Breite	25
Abb. 13:	Synthetisches Grauwertbild Rampenkante mit 7 Stufen	25
Abb. 14:	vertikales E-Feld-Profil	26
Abb. 15:	H-Feld-Profil	26
Abb. 16:	Konturen einer künstlichen, vertikalen Rampenkante	28
Abb. 17:	2-dimensionale, lineare Interpolation	29
Abb. 18:	Aufteilung der Vollpixel in Subpixel	29
Abb. 19:	H-Feld-Profil mit Subpixel	31
Abb. 20:	vertikales E-Feld-Profil mit Subpixel	31
Abb. 21:	Konturpunkte mit/ohne Subp-Auflösung	32
Abb. 22:	horizontale und vertikale Kirschmatrizen	33
Abb. 23:	E-Feldverlauf mit Kirsch-Matrizen	33
Abb. 24:	Koeffizienten der H-Feld Matrix	34
Abb. 25:	Form der H-Feld Matrix und Marr-Hildreth-Matrix	35
Abb. 26:	H-Feldverlauf mit Marr-Hildreth-Matrizen	35
Abb. 27:	Startpunktsuche	37
Abb. 28:	Gleitende Mittelwertbildung zur Eckpunkterkennung	38
Abb. 29:	Alternativen der Kantendarstellung	44
Abb. 30:	Aperturproblem	45
Abb. 31:	Äquidistanzlinien im 2-dimensionalen Merkmalsraum	48
Abb. 32:	Schnitt der Projektionsgeraden	55
Abb. 33:	Zentralprojektion	56
Abb. 34:	Schnitt windschiefer Geraden	57
Abb. 35:	Schnitt der Projektionsebenen	59
Abb. 36:	Das Experimentiersystem MARVIN	62
Abb. 37:	Datenfluß zwischen den Modulen	63
Abb. 38:	Bandbreite der sensorgestützten Modelländerung	65
Abb. 39:	Struktogramm zum prinzipiellen Ablauf der schritthaltenden Lokalisation	68
Abb. 40:	Fahrzeugkinematik und Kreisbogenapproximation	69
Abb. 41:	Struktogramm zur sequentiellen Innovation	74
Abb. 42:	Korrespondenzfindung bei der Lokalisation	75
Abb. 43:	Lageveränderung extrahierter Kanten in Bildfolgen konstanten Inhalts	76
Abb. 44:	Überschwingen an scharfen Richtungsänderungen einer Kontur	77
Abb. 45:	globales Distanz-Histogramm (euklidisch, Kantenmittelpunkte, normiert)	81
Abb. 46:	globales Distanz-Histogramm gezoomt (eukl, Mittelpunkte, normiert)	81
Abb. 47:	lokale Distanzen (eukl, Mittelpunkte, normiert)	82
Abb. 48:	lokale Distanzen (eukl, Mittelpunkte und Länge, normiert)	83
Abb. 49:	lokale Distanzen (eukl, Mittelpunkte und Winkel, normiert)	83

Abb. 50: lokale Distanzen (eukl, Mittelpunkte, Länge und Winkel, normiert)	84
Abb. 51: lokale Distanzen (horizontales E-Feld)	84
Abb. 52: globale Distanzen (horizontales E-Feld)	85
Abb. 53: lokale Distanzen (H-Feld)	85
Abb. 54: lokale Distanzen (Manhattan, Mittelpunkte, normiert)	86
Abb. 55: Reihenfolge der Transformationen bei der 3D-Rekonstruktion	87
Abb. 56: XY-Plot eines Lokalisationsexperimentes	88
Abb. 57: Grauwertbild aus der Meßreihe mk_lok4 zur Lokalisation	89
Abb. 58: Abweichung x/y zwischen Schätzung und Odometrie	90
Abb. 59: Verlauf des Fahrzeugwinkels f bei Odometrie und Schätzung	91
Abb. 60: Ausschnittsvergrößerung XY-Plot	91
Abb. 61: Anzahl der Merkmale und Korrespondenzen	92
Abb. 62: Teilbilder a-c: extrahiertes und erw. Merkmalsbild, Korrespondenzen	93
Abb. 63: Zeitdiagramm und Pipelining	96
Abb. 64: Transformation Grauwerte in Dipole	100
Abb. 65: Felderzeugende Dipolladungen	101
Abb. 66: horizontales Leitelement mit stromerzeugenden Ladungen	102
Abb. 67: Beitrag zum Magnetfeld eines horizontalen Leitelements	103
Abb. 68: Felder eines idealen synthetischen Grauwertbildes	105
Abb. 69: Rohbild und Kantenbild aus der Sequenz bf_0000	106
Abb. 70: Rohbild und Kantenbild aus der Sequenz mk6	107
Abb. 71: Halbbild aus einer Lokalisationsfahrt	108
Abb. 72: Signalfluß beim Kalman-Filter	109

1. Einleitung

Die Automatisierungstechnik trägt heute dazu bei, viele Vorgänge in Fertigung und Service rationeller durchführen zu können und den Menschen von gefährlichen Tätigkeiten zu befreien [26]. Durch die Kombination von *Autonomie* und *Mobilität* kann die Effektivität, Flexibilität und der Anwendungsbereich der Systeme erweitert werden. *Autonome Systeme* sind gegenüber Abweichungen im Umfeld tolerant und können sich auf solche Änderungen einstellen, ohne auf externe Hilfe zurückgreifen zu müssen. Sind solche autonomen Systeme zusätzlich *mobil*, ergeben sich weitere anspruchsvolle Anforderungen an die Informationsverarbeitung. Derartige Systeme enthalten unter anderem Sensoren, mit denen die Umgebung erfaßt wird, ein geometrisches Modell, in dem die Realität dieser Umgebung abgebildet ist, sowie eine Auswertung, die Sensor- und Modelldaten miteinander vergleicht. Die taktische und strategische Aktionsplanung, die Verbesserung der Modellierung und die Lokalisation basieren auf den Ergebnissen der Sensordatenauswertung. (aus [66])

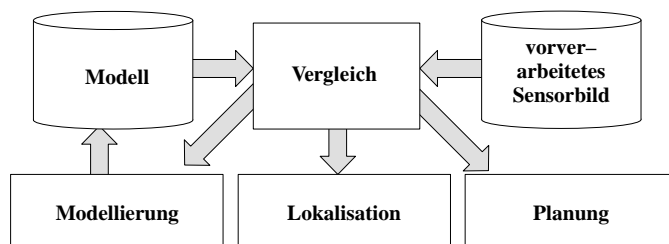


Abb. 1: Informationsfluß bei der Sensordatenauswertung

Diese Arbeit ist innerhalb des Sonderforschungsbereichs SFB331 "Informationsverarbeitung in autonomen mobilen Handhabungssystemen" der Deutschen Forschungsgemeinschaft an der Technischen Universität München entstanden. Bisherige Arbeiten in diesem Umfeld (Kap. 3.13 und 4.3 in [94]) haben gezeigt, daß die geometrische Modellierung der Operationsumgebung in Form einer polygonalen Objektbeschreibung die Auswertung der Sensordaten für die typischen Anwendungen Navigation, Objekterkennung und Schutzraumüberwachung vereinfacht und robuster macht. Von zentraler Bedeutung ist eine schnelle Modellanfrage der lokal sichtbaren, sensorspezifischen Merkmale.

Die Szeneninformation des verwendeten Modells wurde bisher aus CAD-Daten transformiert und mit erheblichem Aufwand manuell an die verwendete Sensorik angepaßt. Im Fall fehlender CAD-Daten mußte die komplette Operationsumgebung aufwendig vermessen werden.

1.1. Motivation

Ziel der vorliegenden Arbeit war es zu untersuchen, inwieweit anhand eines einfachen und preiswerten passiven Sensors, nämlich einer monokularen Grauwert-Videokamera, Beiträge zur weitgehend automatischen Erstellung des Umgebungsmodells geliefert werden können, sowie die hierbei verwendeten Verfahren auf Eignung auch für die *Lokalisation* zu prüfen.

Dazu sind bekannte Verfahren hinsichtlich ihrer Verwendbarkeit untersucht worden. Da diese Verfahren vorwiegend offline operieren, oder sich nur durch einen massiven Hardware-Einsatz in einer realen Anwendung einsetzen lassen, jedoch die Forderung besteht, mit geringem HW-Aufwand eine schritthaltende Verarbeitung - also *echtzeitfähig* - durchzuführen, wurde ein eigenes Verfahren entwickelt und hinsichtlich seiner Tauglichkeit verifiziert. Es konnte erreicht werden, daß die komplette Auswertung eines Sensorbildes innerhalb eines vorher festgelegten Zeitraums abgeschlossen wird. Das Sensorbild ist Teil einer Bildsequenz einer *realen* Szene, die während einer *realen* Fahrt aufgenommen wurde. Die Zeitanforderungen werden dabei durch die Eigengeschwindigkeit festgelegt.

Die erforderliche Auswertzeit einer Aufnahme hängt von der Komplexität und Struktur der aufgenommenen Szene ab. Für Bilder von komplexen Szenen ist bei der schritthaltenden Verarbeitung mit äquidistanter Abtastung die zur Verfügung stehende Auswertzeit nicht ausreichend, um eine Vollbildanalyse durchzuführen. Die Beschränkung auf relevante Bildausschnitte und deren Auswahl ist daher ein wichtiges Teilproblem.

Die in Frage kommenden Bildverarbeitungsverfahren, und zwar die untersuchten Algorithmen aus der Literatur, als auch das in dieser Arbeit entwickelte Verfahren, nutzen die systembedingte Eigenbewegung des mobilen Systems für die Rekonstruktion der bei bildgebenden Sensoren fehlenden Tiefeninformation. Im Gegensatz zum binokularen Stereosehen, das auch der Mensch mit seinen zwei Augen in vielen Situationen verwendet, entsteht beim sogenannten *Bewegungsstereo* die zweite Ansicht mit dem selben Aufnahmesystem zeitlich versetzt von einer anderen Beobachungsposition aus. Durch die Analyse einer zeitlich dichten Bildfolge und durch Ausnutzung des zeitlichen Kontextes vereinfacht sich das Problem der *Korrespondenzfindung*, das bei diskretem Bewegungsstereo erheblichen Aufwand verursacht.

Um eine Szene mit einem Videosensor auflösen zu können, sind Grauwertunterschiede im Bild erforderlich. Die Art der Szene, und damit die Art der Grauwertunterschiede im Bild hat entscheidenden Einfluß auf die erforderliche Sensordaten-Verarbeitung. Das Abbild von natürlicher Umgebung, z.B. der Krone eines belaubten Baumes, unterscheidet sich strukturell vom Abbild künstlicher, menschengeschaffener Umgebung, z.B. eines Büroraums oder eines

Labors. Letztere sind die Umgebungen, für die die Untersuchungen vorgenommen wurden. Die dort vorherrschenden langen, linienförmigen Konturen lassen sich in den Bildern gut detektieren und als komprimierte Repräsentanten (*Merkmale*) der aufgenommenen Szene weiterverarbeiten.

Desweiteren wurden ausschließlich quasistatische Szenen betrachtet. Das heißt, die Umgebung darf sich grundsätzlich ändern, nicht aber, während die Szene beobachtet wird. Dadurch können Szenen bearbeitet werden, in denen Gegenstände ihre Position oder ihre Stellung ändern. Das komplexe Problem der Interpretation von Fremdbewegung bei gleichzeitiger Eigenbewegung wird jedoch ausgeklammert.

1.2. Begriffsdefinitionen

Im folgenden werden die wesentlichen Begriffe definiert, die zum Verständnis der Arbeit notwendig sind.

Exploration	Unter Exploration wird hier die Gewinnung von geometrischem Wissen über eine vollständig unbekannte Szene verstanden.
Kante	Eine Kante ist eine Kontur ohne Krümmung. Bei idealer Abbildung kann von einer Kante im Bild auf eine Kante in der Szene geschlossen werden.
Kantenextraktion	Eine von mehreren Möglichkeiten, um vom Bildraum in einen Merkmalsraum zu gelangen, ist die Kantenextraktion. Dabei wird aus einem Graubild ein Kantenbild, also ein Merkmalsbild, erzeugt.
Kartographierung	Vorgang zur Detaillierung der vorhandenen geometrischen Information über eine Szene, z.B. durch konsistentes Einfügen der interpretierten Sensorwahrnehmungen.
Kontur	Hier wird mit Kontur der Übergang zwischen zwei oder mehreren Bereichen eines Videobilds unterschiedlicher Intensität oder Farbe bezeichnet.
Korrespondenzfindung	Der Vorgang zur Identifikation der gleichen Merkmale in unterschiedlichen Bildern derselben Szene wird Korrespondenzfindung genannt.

Lochkamera	Bei dem vereinfachten Abbildungsmodell der Lochkamera werden Punkte des Urbilds durch ein punktförmiges Abbildungszentrum auf die Bildebene projiziert. Es gelten die Gesetze der Strahlenoptik.
Lokalisation	Unter Lokalisation versteht man die Bestimmung der Position und der Orientierung eines Objektes innerhalb einer Szene. Einschränkend wird hier die Bestimmung der Eigenposition und der Orientierung des bewegten Beobachters als Lokalisation bezeichnet. Bei einem autonomen System kann hierzu beispielsweise die →Odometrie verwendet werden. Eine andere Methode ist die Referenzierung an künstlichen oder natürlichen Landmarken, deren örtliche Lage bekannt ist. Durch →Triangulation der Sehstrahlen zu diesen Marken wird die Eigenposition bestimmt.
Merkmal	Merkmale sind Kennzeichen der Szenenelemente und dienen zu deren Unterscheidung und Erkennung. Merkmale werden durch Vorverarbeitung und Abstraktion der Sensor-Rohdaten gewonnen. Sie stellen informationstechnisch betrachtet eine komprimierte Repräsentation der relevanten Sensorinformationen dar, indem die für die weitere Interpretation irrelevante Sensorinformation weggelassen wurde. Typische Merkmale, die aus einem Grauwertbild gewonnen werden können, sind Kanten, Ecken oder Flächen.
Navigation	Die Navigation ist die Durchführung und Überwachung einer geplanten Fahrt durch die Szene. Bei Abweichung vom geplanten Kurs wird Antrieb und Lenkung des Fahrzeugs nachgeregelt.
Odometrie	Die Odometrie ist eine Vorrichtung an einem Fahrzeug, die durch fortgesetzte Aufsummierung der Radbewegungen die aktuelle Position und die Orientierung des Fahrzeugs fortschreibt. Aufgrund von Meßungenauigkeiten und Problemen mit der Radhaftung driftet die berechnete Fahrzeugposition bereits nach kurzer Strecke von der wahren Position weg.
Stereo	”Stereo” ist das aus dem griechischen stammende ”Bestimmungswort von Zusammensetzungen mit der Bedeutung ’starr, fest, massiv; räumlich, körperlich’” [72]. Hier wird damit das Verfahren zur Rekonstruktion der bei der Projektion auf einen Flächensensor

verloren gegangenen Tiefeninformation, in der Regel durch →Triangulation bezeichnet. Wird an mehreren Positionen gleichzeitig beobachtet, spricht man von binokularem, triokularem oder multiokularem Stereo, beim Beobachten aus der Bewegung heraus vom Bewegungstereo oder Motionstereo.

Triangulation

Triangulation ist ein Verfahren zur Bestimmung der Koordinaten eines Punkts durch Geradenschnitt. In der Stereobildverarbeitung sind die zu schneidenden Geraden zwei oder mehrere Sehstrahlen, mit denen der zu bestimmende Punkt auf die jeweiligen Projektionsflächen abgebildet wird. Die Basis der Triangulationsdreiecke ist die Strecke zwischen den Projektionszentren, die Winkel der Sehstrahlen ergeben sich aus der Lage des Bildpunkts auf der Bildebene sowie der Ausrichtung des Beobachtungssystems.

Umgebungsmodell

Das hier verwendete Umgebungsmodell enthält Wissen über die Geometrie der Umgebung, sowie über sensorspezifische Eigenschaften, z.B. Reflexions- und Absorptionsvermögen. Mit Hilfe von Abfrageverfahren kann eine sensornahe Ansicht der Szene errechnet werden.

1.3. Gliederung

In **Kapitel 2.** findet sich die Diskussion der allgemeinen Prinzipien und der bekannten Verfahren des Bewegungstereo aus der Literatur. Ausgangspunkt der Top-Down-artigen Erörterung ist die Anwendung "Szenenanalyse aus der Bewegung", was die Auswahl der betrachteten Verfahren einschränkt.

Das in dieser Arbeit entwickelte Bewegungstereo-Verfahren wird in **Kapitel 3.** vorgestellt und die drei erforderlichen Einzelschritte ausführlich dargelegt. Die Beschreibung erfolgt nach dem Bottom-Up-Prinzip, d.h. in der Reihenfolge der Bearbeitungsschritte.

Die Anwendung des entwickelten Verfahrens ist Thema des **4. Kapitels.** Hierzu werden Beispiele aus dem Bereich Exploration und Lokalisation herangezogen.

Die mit dem Verfahren erzielten Ergebnisse werden in **Kapitel 5.** präsentiert und qualitativ und quantitativ bezüglich Rauschempfindlichkeit, Genauigkeit und Zeitverhalten bewertet.

Die Arbeit schließt in **Kapitel 6.** mit einer Zusammenfassung.

2. Bekannte Bewegungsstereo-Verfahren und -Systeme

2.1. Verfahren für Bewegungsstereo

2.1.1. Rekonstruktion der Tiefeninformation

Bei Verwendung eines bildgebenden Sensors, wie z.B. der Videokamera, geht bei der Abbildung der Szene auf die Projektionsfläche die Tiefeninformation verloren. Dies ermöglicht bei Einzelbildern lediglich die Auswertung in zwei Dimensionen.

Faßt man mehrere zweidimensionale Bilder in einer Aufnahmesequenz zusammen, wird die Messung um eine dritte Dimension erweitert. Die hinzukommende dritte Dimension kann räumlich sein, wie z.B. bei der Auswertung von Schnittbildern in der Tomographie[75], oder bei der Auswertung von Aufnahmen der gleichen Szene aus unterschiedlichen Positionen bei multi-okularem Stereo[27]; sie kann zeitlich sein bei der Analyse dynamischer Szenen von einem festen Standpunkt aus [51]; sie kann sich auf den spektralen Übertragungskanal beziehen, beispielsweise bei der Auswertung von Infrarot- und Tageslichtaufnahmen; sie kann sich sogar auf die Art des bildgebenden Sensors beziehen, z.B. eine Videokamera und ein Laser-scanner [103]; sie kann aber auch eine Mischform sein, wie sie beim Bewegungsstereo vorliegt: durch die Eigenbewegung der Kamera in 6 räumlichen Freiheitsgraden (3 translatorische und 3 rotatorische) wird eine Trajektorie im 7-dimensionalen Raum-Zeit-Kontinuum erzeugt. Entlang dieser Trajektorie werden die Videobilder aufgenommen. Dadurch entsteht sowohl ein zeitlicher, als auch ein räumlicher Versatz zwischen den Aufnahmeorten.

Ziel der Bildfolgenauswertung, wie eigentlich der gesamten Bildverarbeitung, ist es, durch die Beobachtung eine oder mehrere Unbekannte zu bestimmen. Meist ist diese Unbekannte die Struktur der Szene, d.h. die Art und Lage der Objekte sowohl relativ zueinander, als auch bezüglich eines ortsfesten Koordinatensystems. Die Strukturanalyse enthält dabei qualitative und quantitative Aspekte, beispielsweise die Frage, welches das nächstgelegene Objekt ist, oder die Aufgabe, die Breite einer Durchfahrt zu vermessen.

Ist die Szene hinsichtlich Qualität und Quantität hinreichend bekannt, kann das Ziel der Bildfolge-Analyse sein, die unbekannte eigene Bewegungstrajektorie zu rekonstruieren (Lokalisation). Bei autonomen mobilen Systemen hat dies eine herausragende Bedeutung. Die Problemstellungen statische Kamera – bewegtes Objekt und bewegte Kamera – statisches Referenzobjekt sind äquivalent, solange nur die relative Lagebeziehung hergestellt werden soll.

Der zeitliche Versatz hat bei statischen Szenen keine Auswirkung, wohl aber bei bewegten Szenen, und erfordert eine separate Betrachtung von Szenenteilen, die unterschiedliche Bewegungen ausführen. Die Eliminierung der Eigenbewegung und die Zuordnung von Bildprimiven zu gleichartig bewegten Objekten und Objektteilen ist eine anspruchsvolle Aufgabe, die auch unter dem Begriff Objekt-Hintergrund-Trennung bekannt ist.

2.1.2. Interpretation der Bewegung im Bildraum

Die Relativbewegung zwischen Szene und Beobachter verursacht eine Bewegung der abgebildeten Szene auf der Bildebene des Sensors. Ausgehend von der Bildhelligkeitsfunktion $I(x, y, t)$, die die Intensität eines jeden Pixels in Abhängigkeit von der Zeit und der Position beschreibt, wird der sog. optische Fluß über die folgende Beziehung definiert:

$$I_x u + I_y v + I_t = 0 \quad I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}, \quad I_t = \frac{\partial I}{\partial t}$$

Mit dem Flußfeld wird jedes Pixel mit einem zeitabhängigen Geschwindigkeitsvektor u, v attribuiert.

Abb. 2 zeigt die qualitative Auswirkung verschiedener Arten von Relativbewegungen zwi-

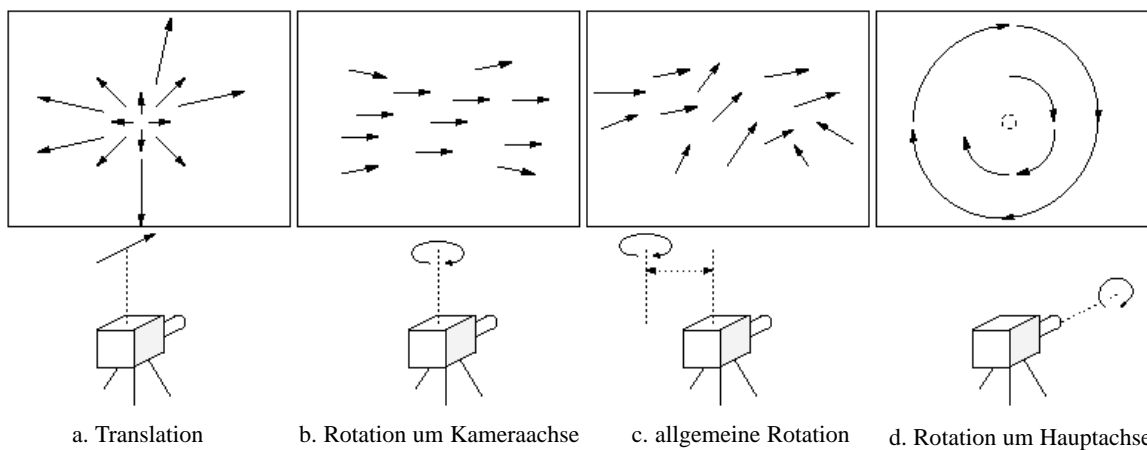


Abb. 2: Auswirkung von Relativbewegungen auf den optischen Fluß

schen Beobachter und Szene auf den optischen Fluß, die nachstehend erläutert sind:

a)	translatorische Bewegung in Richtung Hauptachse	radiale Flußvektoren
b)	Rotation um eine zur Projektionsebene parallele Achse, die durch das Projektionszentrum verläuft	Hyperbeln

c)	Rotation um eine zur Projektionsebene parallele Achse, die nicht durch das Projektionszentrum verläuft	Superposition von Hyperbeln und Linien
d)	Rotation um die Hauptachse	konzentrische Kreise

Liegt der optische Fluß vor, kann durch Invertierung der Flußformel das dreidimensionale Bewegungsfeld bestimmt werden, und so auf die Szene rückgeschlossen werden.

Technisch wird der optische Fluß durch eine 2-dimensionale Korrelation von zeitlich versetzt aufgenommenen Einzelbildern bestimmt [1][42]. Im Idealfall kann zu jedem Pixel ein Verschiebungsvektor angegeben werden [42] [80].

In der Praxis ist die Berechnung des Verschiebungsvektors durch Korrelation sehr aufwendig und ohne HW-Unterstützung nicht für die Online-Auswertung geeignet.

In [98] wird ein System beschrieben, das ein Verschiebungsfeld mit Hilfe von sog. Motion-Estimation-Prozessoren errechnet, wobei jedoch der Bewegungsvektor nicht für jedes Pixel, sondern lediglich für Blöcke von Pixeln bestimmt wird.

2.1.3. Korrespondenzfindung

Die Errechnung der Korrelation entspricht aus funktionaler Sicht der bei Stereosystemen erforderliche Korrespondenzfindung. Neben diesem pixelorientierten Verfahren gibt es weitere Verfahren, die sich durch den Grad der Analyse innerhalb der Einzelbilder unterscheiden, bevor die Bestandteile der Einzelbilder in Beziehung gesetzt werden. Die Palette der Verfahren reicht von der erwähnten Korrelation bis zu Verfahren auf Objektebene. Dazwischen sind merkmalsorientierte Verfahren angesiedelt. Bei diesen werden zuerst prägnante Elemente aus den Einzelbildern extrahiert, die dann in einem zweiten Schritt in Beziehung gesetzt werden. Im Grunde ist der Unterschied zu den pixelbasierten Verfahren die Vertauschung der Bearbeitungsschritte Korrespondenzfindung und Segmentierung.

Mathematisch läßt sich das Korrespondenzproblem bei den merkmalsbasierten Verfahren folgendermaßen beschreiben[60]: Gegeben sei ein Bild n mit den Merkmalen A_i , sowie ein Bild $n+1$ mit den Merkmalen B_j . Gesucht ist eine Permutation $j = \sigma(i)$, mit der jedem Merkmal A_i ein Merkmal $B_{\sigma(i)}$ optimal zugeordnet wird.

Bei dieser Formulierung bleibt allerdings unberücksichtigt, daß in einzelnen Bildern der Folge Merkmale durch Verdeckung oder Nichtentdecken ausfallen können, kurz darauf aber wieder auftauchen, und daß Merkmale aus dem Blickfeld verschwinden und durch Annähe-

rung erstmalig entdeckt werden können. Die Anzahl der Merkmale pro Bild ist also nicht konstant, und für einige Merkmale aus A_i und B_j lassen sich möglicherweise keine Korrespondenzen finden. Die Etablierung von Korrespondenzen bei merkmalsbasierten Verfahren ist im Regelfall also eine aufwendige Suchoperation in den Merkmalslisten der beiden Bilder, bei der durchaus falsche Korrespondenzen entstehen können, die dann zu Fehlern in der Szenenrekonstruktion führen.

Eine Schwierigkeit bei der Zuordnung ist eine große Stereodisparität. Die Stereodisparität ist umso größer, je weiter die Stereobasis ist. Werden zwischen den Aufnahmepositionen eines Stereopaars Zwischenaufnahmen gemacht, kann die Bewegung eines Merkmals von der Position im ersten Stereobild über die Positionen in den Zwischenaufnahmen zu der Position im anderen Stereobild verfolgt werden [74]. Der Vorteil liegt in der nahezu kontinuierlichen Änderung der Merkmalspositionen in der über die Stereobasis verteilten Bildfolge, gegenüber der stufigen Änderung bei den Bildpaaren beim binokularen Stereo. Diese Kontinuität kann zur Verfolgung der Merkmale genutzt werden, um auch bei weiteren Stereobasen Korrespondenzen sicherer zu bestimmen.

Bei den Aufnahmen aus der Bewegung entstehen zeitlich dichte Aufnahmen, in denen die Kontinuität der Merkmalsbewegung in der Regel gewährleistet ist.

Ullman [100] formuliert die Korrespondenzfindung erstmalig als Problem der minimalen Zuordnung. Eine korrekte Korrespondenz ist gefunden, wenn die Summe der "Kosten" aller Zuordnungen minimal ist. Als Kostenfunktion wird der negative Logarithmus der Wahrscheinlichkeit für eine angenommene Merkmalsgeschwindigkeit verwendet.

[70] ist ein weiteres Beispiel für ein Korrespondenzverfahren mit Kostenfunktion. In die Kostenfunktion geht hier nicht nur die Lage des Merkmals, sondern auch Attribute wie "Breite" der Kante, Kontrast, Steilheit, Mittelwert der 10% hellsten Pixel, usw. bei.

In einem anderen Verfahren [93] wird die dynamische Programmierung [7] für die Korrespondenzfindung von Konturen verwendet. Auch hier wird eine Kostenfunktion benötigt. Für die Bestimmung von Korrespondenzen zwischen Liniensegmenten verwendet das in [63] beschriebene Verfahren zusätzlich zur dynamischen Programmierung eine Transformation der Linien in den Hough-Raum.

2.1.4. 3D-Rekonstruktion

Für die 3D-Rekonstruktion gibt es im wesentlichen die klassischen Photogrammetrieverfahren und die Schätzverfahren. Beiden gemeinsam ist das Prinzip der Triangulation, d.h. dem Schnitt von Sichtlinien.

Aus der Photogrammetrie ist die Triangulation unter dem Namen räumlicher Vorwärtsschnitt bekannt: Bei bekannten inneren und äußeren Kameraparametern, – d.h. Abbildungseigenschaften, Eigenposition und -orientierung sind bekannt – ist aus zwei gemessenen Bildkoordinatenpaaren eines korrespondierenden Punktes dessen Weltkoordinaten zu bestimmen. Das Problem wird durch den Bündelausgleich der Sichtlinien gelöst. Dieses ist ein Ausgleichsproblem zur Minimierung der Fehlerquadrate, das durch die Aufstellung eines Normalgleichungssystems mit drei Unbekannten gelöst wird.[49]

Auch das Lokalisationsproblem ist in der Photogrammetrie ein Standardproblem, das unter dem Begriff räumlicher Rückwärtsschnitt bekannt ist. Dahinter steckt die Bestimmung der Eigenposition in sechs Freiheitsgraden aus einem Bild, in dem mindestens drei Paßpunkte, also Punkte mit bekannten Weltkoordinaten, sichtbar sind. Auch hier erfolgt die Lösung durch Aufstellung eines überbestimmten Gleichungssystems, das durch Ausgleichsverfahren in den sechs Koordinaten gelöst wird.

Bei den Schätzverfahren wird die unbekannte 3D-Position des beobachteten Punktes in der Szene als Systemzustand interpretiert, der durch die Merkmalskoordinaten im Bild beobachtbar wird. Im Prädiktionsprozeß des Minimum-Varianz-Schätzers wird über die Abbildungsgleichungen aus dem Zustandvektor ein Prädiktionsvektor gebildet, der den erwarteten Merkmalskoordinaten im Bild entspricht. Die Differenz zwischen Prädiktion und Beobachtung wird zurückgespeist. Diese Differenz korrigiert und verbessert den geschätzten Systemzustand. Der Vorgang läuft iterativ ab, und wertet die Beobachtungen jeder Einzelaufnahme aus.

2.2. Systeme mit Bewegungstereo

2.2.1. Bewegungstereosystem für punktförmige Szenenmerkmale

P. Schaeren [91] hat ein Bewegungstereo-System für punktförmige Szenenmerkmale entwickelt, in dem die für ein solches System anfallenden Probleme, nämlich Grauwertverarbeitung, Tracking, Korrespondenzfindung und 3D-Rekonstruktion unter massivem Hardware-Einsatz gelöst werden. Es wird eine Bildrate von 50Hz erreicht.

Die Auswahl der Bildpunkte, die an kein festes Raster gebunden sind, erfolgt durch Analyse der Autokovarianz der Grauwerte mit Hilfe einer 4-stufigen Bild-Pyramide, mit der kaskadenartig jedes Rohbild der Folge unterabgetastet wird, um mit einer hierarchischen Suche beginnend bei der auf 32x32 Pixel am niedrigsten gerasterten Ebene geeignete Bildpunkte zu finden.

Aus Performancegründen wird zur Punktauswahl keine Kovarianz berechnet, sondern der Betrag und die Richtung des lokalen Gradienten. Dies geschieht durch gleichzeitige Anwendung von 8 3x3 Kirsch-Kompaß-Matrizen an jedem Bildpunkt, Maximumsbildung über die 8 Richtungen und abschließender Binarisierung.

Die Korrespondenzen der ausgewählten Bildpunkte werden durch die Kreuzkorrelation eines 15x15 großen Bildausschnitts um den zu verfolgenden Bildpunkt etabliert. Die Implementierung auf einem Signalprozessor erlaubt 71 zu verfolgende Punkte in einem Videotakt von 20ms. Durch Extrapolation der Punkteposition mit Hilfe von Geradenstücken kann in einzelnen Bildern auf die Verifikation dieser Punkte verzichtet werden, wodurch Rechenkapazität für zusätzlich zu verfolgende Punkte frei wird.

Die 3D-Rekonstruktion erfolgt durch Schnitt der Projektionsgeraden der korrespondierenden Bildpunkte, sobald die Stereobasis ausreichend groß ist. Die jeweilige Eigenposition zum Aufnahmezeitpunkt ist durch die Ansteuerung der Roboterhand, die die Kamera bewegt, bekannt. Es entsteht eine Wolke von 3D-Positionen, die durch Heuristiken von Ausreißern befreit wird. Nach einer Plausibilisierung durch die Varianz wird die angenommene 3D-Position durch Mittelwertbildung bestimmt.

Probleme entstehen hier, wie bei jedem korrelationsbasierten Ansatz, bei Drehung des Korrelationsmusters in der Bildebene, die durch bestimmte Eigenbewegungen entstehen können, denn die Korrelation erfolgt nur über x und y , nicht aber über den Winkel.

Viel Aufwand wird auch hier für die Merkmalsextraktion getrieben: Spezialhardware in Form von einigen ASICs für die Korrelationen, FIR-Filter, sowie ein Transputernetzwerk bewältigen die erforderlichen 2 Gigaops/sec für die Bildverarbeitung.

Die gezeigten Experimente beschränken sich auf künstliche Szenen, z.B. planare Bildmuster. Es wird eine hohe Genauigkeit bei den rekonstruierten 3D-Punkten erreicht (ca. 1 mm bei 60 mm Entfernung), wenn ausreichend viele Bilder (Größenordnung einige hundert) aus unterschiedlichen Positionen vorliegen.

2.2.2. Bewegungstereosystem zur Exploration von Innenräumen

Lebeque et. al. stellen in [56],[57],[58] das Robotersystem "Robotex" zur Exploration und Kartographierung von Innenräumen vor. Ziel ist die exakte Vermessung mit Video, um die Szene semantisch interpretieren zu können. Daher muß die Qualität des zu erstellenden Umgebungsmodells höher sein, als für Navigationszwecke erforderlich.

Aus der Grauwert-Videosequenz, die während der Fahrt durch die Szene aufgenommen wird, werden nur die Kanten extrahiert, die in 3D parallel zu den in der Szene vorherrschenden

Raumrichtungen liegen. Bei Standardräumen sind das beispielsweise die vertikale Achse sowie die beiden dazu orthogonalen Raumrichtungen.

Die Projektionen paralleler 3D-Kanten schneiden sich in einem Fluchtpunkt, der für jede 3D-Vorzugsrichtung unter Verwendung der Kamera-Parameter berechnet werden kann. Liegt der Fluchtpunkt in Richtung einer 2D-Kante, ist die Kante ein Kandidat für die Raumrichtung, die dieser Fluchtpunkt repräsentiert.

Die inneren Kamera-Parameter werden durch Kalibrierung bestimmt, die äußeren mit Hilfe der on-Bord-Sensorik (Odometrie mit Korrektur durch Video und Inclinometer für die Kamerawinkel) gemessen. Es ist mit dem Meßverfahren auch möglich, eine Eigenpositionshypothese zu verifizieren.

Die implementierte Kantenextraktion ist klassisch und orientiert sich an [11]. Sie beinhaltet die Schritte

- Bestimmung eines Richtungs-Feldes für jeden Fluchtpunkt
- Unterdrückung der Non-Maxima
- Verketten der Pixel
- Geradenapproximation und
- globale Richtungprüfung

Das Richtungsfeld besteht für jedes Pixel aus der Differenz zwischen Richtung zum Fluchtpunkt und Richtung des lokalen Grauwertgradienten. Durch Schwellwerte werden bei der Geradenbildung nur Pixel mit entsprechender Raumrichtung berücksichtigt. Zur Verkettung der Pixel wird ein lokaler Suchmechanismus verwendet, und die 2D-Gerade wird mit einem Least-Square-Algorithmus approximiert.

Die Bestimmung der 3D-Kanten aus der Bildfolge erfolgt wie in [15] mit einem rekursiven Schätzverfahren. Der Zustandvektor der Kante ist die Lage, nachdem die Richtung durch die besondere Art der Kantenextraktion bereits feststeht. Kanten parallel zur Bewegungsrichtung des Beobachters werden aufgrund der unsicheren Endpunkte ausgenommen.

Ein extremes Weitwinkel-Objektiv ermöglicht bei der vorwiegenden Bewegung in Blickrichtung Kanten länger zu beobachten, bevor sie an den Bildrändern verschwinden. Durch den stärker variierten Beobachtungswinkel einer Kante und die geringere Unsicherheit der Meßwerte im Nahbereich erhöht sich die 3D-Genauigkeit.

Das implementierte System läuft auf einer HP-730 Workstation und benötigt 1.2 Sekunden pro Videobild (512x480) für die Kantenextraktion, sowie 0.2 Sekunden für die Korrespon-

denzfindung und die 3D-Schätzung. Hinzu kommen 0.3 Sekunden für die Bildverzerrung auf Rohbildebene.

2.2.3. Rekonstruktion durch axiales Bewegungstereo

Lavest/Rives/Dhome stellen in [54] ein System vor (Abb. 3), das axiales Bewegungstereo

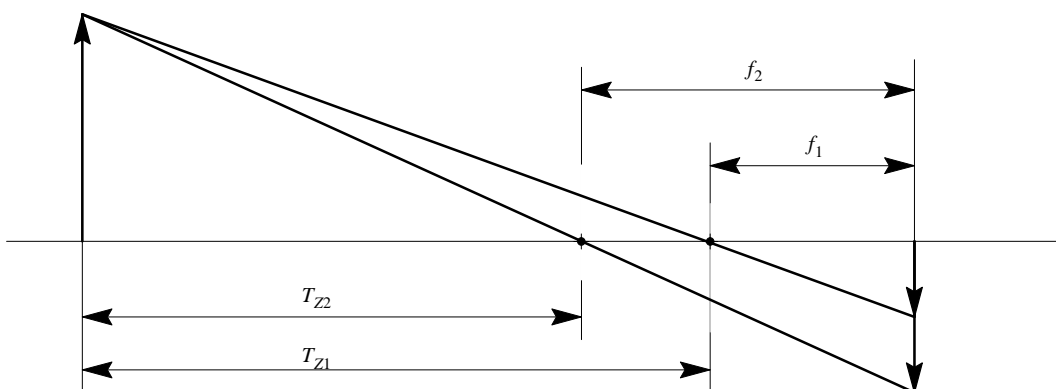


Abb. 3: axiales Stereosystem nach Lavest/Rives/Dhome

zur Rekonstruktion der Szene verwendet, jedoch ohne die Eigenposition zu verändern: die Bewegung im Bild wird mit Hilfe eines Zoom-Objektivs erzeugt, das gleichzeitig die Bildweite und die Gegenstandsweite verändert. Im Lochkamera-Modell, das die tatsächlichen Verhältnisse eines Zoomobjektivs nur sehr vereinfacht wiedergibt, entspricht das Zoomen einer Verschiebung der Lage des Kamerabezugspunkts zwischen der Bild- und Gegenstandsebene, wodurch sowohl eine Vergrößerung, als auch eine minimale Eigenpositionsverschiebung entsteht. Die erforderlichen Schritte zur 3D-Rekonstruktion sind die Konturextraktion, die Korrespondenzfindung und die Triangulation. Neben einer exakten Kalibrierung der Kameraparameter ist aufgrund der kurzen Stereobasis eine subpixelgenaue Extraktion der projizierten Bildpunkte für exakte 3D-Koordinaten kritisch. Die Experimente wurden an künstlichen Objekten durchgeführt. In [55] wird der Spezialfall dargestellt, wie mit Hilfe von Zoom-Objektiven die Vermessung von rotationssymmetrischen Objekten vorgenommen werden kann.

2.2.4. Bewegungstereosystem durch Messung der Verschiebungsgeschwindigkeit

Ein anderes Bewegungstereosystem ist von Huber/Gräfe in [44] beschrieben: Unter der Annahme der Bewegung in Z-Richtung (Blickrichtung) wird der Abstand zum Objekt ohne Kenntnis der Objektgröße und der inneren Abbildungsparameter der Videokamera aus dem Versatz der Bildpunkte berechnet (vgl. Abb. 4). Voraussetzung wird, daß die beiden Objektpunkte $P_1 = (X_1, Y_1, Z_1)$ und $P_2 = (X_2, Y_2, Z_2)$ von der Kamera gleichweit entfernt sind ($Z_1 = Z_2 = Z$)

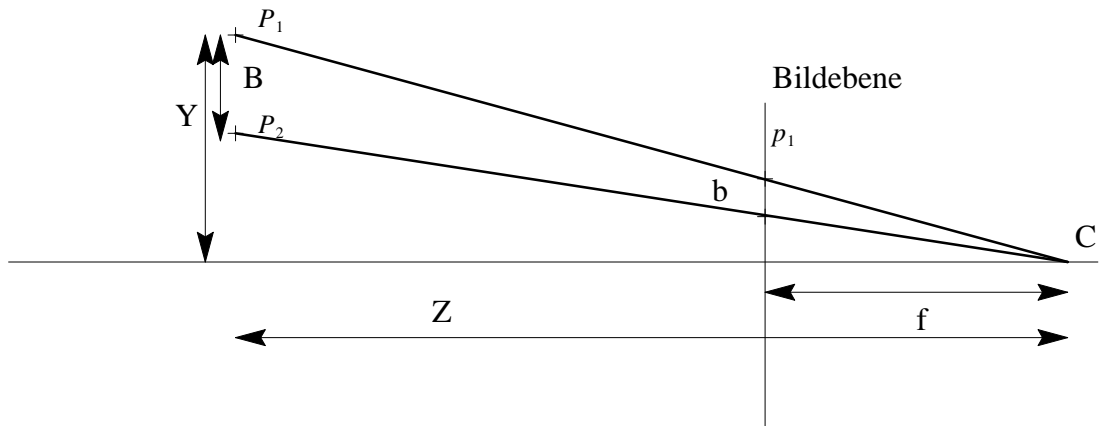


Abb. 4: Berechnung der Objektdistanz nach Huber/Gräfe

Der Punkt p_i ist die Projektion eines Szenenpunkts P_i auf die Bildebene. Es gilt exemplarisch für die y-Richtung:

$$B = |Y_2 - Y_1| \quad b = |y_2 - y_1|$$

$$b = \frac{f \cdot B}{Z}$$

$$\frac{db}{dt} = \frac{db}{dZ} \cdot \frac{dZ}{dt} = -\frac{f \cdot B}{Z^2} = -\frac{b}{Z} \cdot \frac{dZ}{dt}$$

$$Z = -\frac{b}{\frac{db}{dZ}}$$

Damit kann die Tiefe ohne Kenntnis der Abbildungseigenschaften der Kamera allein aus der Stereodisparität b und deren Änderung db , sowie aus dem Versatz in der Blickrichtung dZ bestimmt werden.

Die Fehleranalyse ergibt, daß die Punkte im Bild subpixelgenau vermessen werden müssen, damit die 3D-Koordinaten ausreichend genau sind. Ein geeignetes Verfahren für die subpixelgenaue Feinvermessung von Kanten wird angegeben. Es fehlen jedoch Angaben, wie Korrespondenzen etabliert werden.

2.2.5. Exploration mit Bewegungsstereo und Merkmalsverfolgung

Crowley et.al. stellen in [15] ein Bewegungsstereosystem für die Szenenanalyse mit einem bewegten Beobachter vor. Das System besteht aus zwei formalen Zustandsschätzern (Kalman-Filtern), die für jede Aufnahme die Sequenz Prädiktion, Vergleich und Korrektur durchführen.

Der erste Zustandsschätzer führt die Merkmalsverfolgung im 2D-Bildraum durch. Als Merkmale werden Geradenstücke verwendet. Das Bewegungsmodell besteht aus der Position der

Kantenmitte in erster und zweiter Ableitung, Winkel und halbe Kantenlänge. Beim Vergleich werden zunächst die Winkel der prädizierten Kante und der Bildkante geprüft, danach findet ein Überlappungstest statt. Das korrespondierende Paar wird durch eine Form der Mahalanobis-Distanz bestimmt. Die Abweichung von Prädiktion und Beobachtung wird für die Korrektur des Bewegungsmodells verwendet. Dieser Vorgang läuft bei jedem Bild ab.

Der zweite Zustandsschätzer führt die 3D-Rücktransformation durch, wobei der Zustandsvektor diesmal die Koordinaten der 3D-Kante enthält. Die Bestimmung der äußeren Kameraparameter, also der Beobachterposition, erfolgt durch Schnitt der durch die korrespondierenden Kanten aufgespannten Sichtebenen. Der Zustandsschätzer für die 3D-Rekonstruktion wird in den Experimenten nur alle 5 Einzelbilder aufgerufen, ohne daß für dieses Vorgehen Gründe angegeben werden.

Die Kantenextraktion erfolgt offline außerhalb des Schätzmechanismus.

Das Verfahren wird an verschiedenen Beispielen zur 3D-Analyse von Objekten demonstriert, die von einer Bewegungstrajektorie beobachtet werden.

3. Algorithmen für Online-Bewegungsstereo

In diesem Kapitel werden die Algorithmen vorgestellt, die im Rahmen dieser Arbeit entwickelt, untersucht und erprobt wurden.

Voraussetzung für die Online-Bildfolgenverarbeitung unter Annahme der differentiellen Bewegung der Merkmale ist eine effiziente, schritthaltende Merkmalsextraktion. Wie die Echtzeitfähigkeit mit relativ bescheidenem Hardwareeinsatz zu erreichen ist, wird in Abschnitt 3.1. erläutert. Danach wird in Abschnitt 3.2. beschrieben, wie zwischen den Merkmalen einer Bildfolge Korrespondenzen etabliert werden. Aus diesen Korrespondenzen und der Aufnahmeposition, die entweder aus der Odometrie oder der Lokalisation entstammt, werden die Merkmale im ortsfesten Koordinatensystem rekonstruiert. Dies ist in Abschnitt 3.3. dargestellt. In Abschnitt 3.4. wird schließlich das Experimentiersystem beschrieben, auf dem die dargestellten Algorithmen implementiert sind.

3.1. Merkmalsextraktion

3.1.1. Anforderungen

Für das Bewegungsstereo in den genannten Anwendungen werden besondere Anforderungen an die Merkmalsextraktion gestellt:

- Nutzung von Vorwissen:
Aus vielen Quellen im System steht Vorwissen über die erwarteten Merkmale im aktuellen Bild zur Verfügung. Eine effiziente Merkmalsextraktion nutzt dieses Wissen, um die Suche nach diesen Merkmalen zu vermeiden und um die nachfolgende Korrespondenzfindung zu vereinfachen. Steht kein Vorwissen zur Verfügung, sollen trotzdem alle vorhandenen Merkmale im Bild gefunden werden.
- Selektivität:
Die Merkmalsextraktion als erste Stufe der Informationsreduktion sollte zusätzlich eine Auswahl der Merkmale treffen und nicht jedes entdeckte Merkmal melden. Damit wird vermieden, daß nachfolgende Instanzen durch die Verarbeitung irrelevanter Information belastet werden. Gute Merkmale für die Lokalisation sind z.B. lange, herausragende, gut sichtbare Kanten, die zeitlich stabil sind.
- Merkmalseigenschaften:
Für die Korrespondenzfindung sind über die Einzelbilder einer Sequenz zeitlich stabiler Merkmalseigenschaften mit zu extrahieren, um die Dimensionalität des Korrespondenzraums zu erhöhen und die Korrespondenzfindung robuster zu machen.

– Nachbarschaftsbeziehungen:

Bei der Bildinterpretation auf abstrakterer Ebene sind Nachbarschaftsbeziehungen, wie sie ja in der Vorverarbeitung leicht gewonnen werden können, bzw. ohne Zusatzaufwand abfallen, hilfreich. Diese sollen nicht ungenutzt verworfen werden, sondern den nachfolgenden, auswertenden Modulen erhalten bleiben.

– Rauschempfindlichkeit:

Die Merkmalsextraktion soll möglichst unempfindlich gegenüber dem Amplitudenrauschen des Sensors sein. Dieses Problem tritt bei Einzelbildauswertungen nicht auf, wohl aber bei der Bildfolgenanalyse. Insbesondere ist die Rauschempfindlichkeit so gering zu halten, daß extrahierte Kanten in einer Sequenz mit statischer Szene und ruhender Kamera stets detektiert werden, und ihre Lage nicht ändern.

– Echtzeitfähigkeit:

Die Merkmalsextraktion soll schritthaltend mit der Bildrate von 5–10 Hz durchgeführt werden, um bei den zu erwartenden Fahrzeuggeschwindigkeiten nur kleine Merkmalsverschiebungen in der Bildebene zu erhalten. Dabei sind die rotatorischen Geschwindigkeiten erheblich kritischer als die translatorischen. Der angestrebte Zeithorizont erfordert einen ressourcensparenden Extraktionsalgorithmus und eine effiziente Implementierung. Echtzeit heißt in diesem Zusammenhang allerdings nicht nur hohe Geschwindigkeit. Da der Rechenaufwand für die Merkmalsextraktion von der Anzahl der im Bild vorkommenden Merkmale abhängt, kann nicht ausgeschlossen werden, daß bei einigen Bildern die Zeitbedingung verletzt wird. Es müssen daher Vorkehrungen getroffen werden, daß in diesen Fällen die Merkmalsextraktion auch mit unvollständigen Ergebnissen rechtzeitig beendet wird. Dabei bietet es sich an, durch eine Priorisierung sicherzustellen, daß die für die Anwendung wichtigen Merkmale sicher extrahiert werden, und bei einem möglichen Abbruch nur weniger wichtige Bildinhalte verloren gehen. Für die Wichtigkeit bestimmter Bildbereiche kann wieder Vorwissen herangezogen werden.

– Einschrittige Verarbeitung:

Als Weg zur geforderten Effizienz der Merkmalsextraktion sollen rekursive und iterative Lösungen vermieden werden. Hierzu zählt auch die aufwendige, nachträgliche Fusion von zerfallenen Kanten. Ziel ist, die Extraktion möglichst auf Anhieb korrekt durchzuführen.

Darüber hinaus sollen die bekannten Anforderungen an die Kantenextraktion gemäß Canny [14] ebenfalls bestmöglich eingehalten werden:

- Vollständigkeit und Eindeutigkeit:

Die Wahrscheinlichkeit, eine vorhandene Kante im Bild zu extrahieren (Entdeckungswahrscheinlichkeit) sollte maximal sein, und die Wahrscheinlichkeit, Kanten zu detektieren, die gar nicht vorhanden sind (Falschalarmrate), soll minimal sein. Das ist äquivalent zur Forderung, daß aus jeder Kante im Rohbild exakt eine Kante extrahiert wird.

- Genauigkeit:

Die extrahierten Kanten sollen so nah wie möglich an den tatsächlichen Kanten im Rohbild liegen.

Es hat sich gezeigt, daß die in Software-Paketen verfügbaren, klassischen Verfahren (z.B. Canny in VISTA [2], Deriche in HORUS [24]) diese Anforderungen nicht in ihrer Gesamtheit erfüllen konnten. Hauptgrund sind die Ausführungszeiten: Beispielsweise liegt die Kantenextraktion eines 512x512 Bildes bei der Implementierung des Canny-Algorithmus im Programmpaket VISTA [2] in der Größenordnung von 3-5 sec auf einem Pentium75-System [12], wobei die Zeiten zur Bereitstellung des Bildes noch nicht berücksichtigt sind. Diese Zeiten reichen nicht aus, differentielles Motionstereo durchzuführen¹.

Darüberhinaus fehlt den Standardpaketen die Fähigkeit, zusätzliche Merkmalsattribute und Nachbarschaftsbeziehungen mit zu extrahieren, die für die Korrespondenzfindung zu diesem Zeitpunkt als vorteilhaft angesehen wurde.

Aus diesem Grund wurde eine eigene Vorverarbeitung auf der Basis der Konturverfolgung entwickelt, die die oben genannten, sich zum Teil ausschließenden Anforderungen bestmöglich erfüllt.

Bevor sich Absatz 3.1.3. mit dem implementierten Verfahren beschäftigt, wird in 3.1.2. zunächst auf die zu extrahierenden Merkmale eingegangen.

3.1.2. Merkmale in Grauwertbildern

Als Merkmale in den Grauwertbildern werden lange, gerade Kanten verwendet, die in künstlich geschaffenen Umgebungen überwiegen. Kanten sind Linien in der Bildebene, an denen sich der Helligkeitsverlauf abrupt ändert. Da ein Videobild durch die Projektion der Strahlung in der Szene entsteht, können Kanten verschiedene Ursachen haben, wobei nur eine davon die Existenz einer physikalischen Kante in der dreidimensionalen Szene ist (Abb. 5, aus Faugeras [27] S. 70, Nalwa [76] S. 77))

1. Aufgrund der hohen Datenrate und der verfügbaren Speicherbandbreiten zum Massenspeicher können bei realen Fahrexperimenten keine Rohdaten-Meßreihen mit Bildraten im 5 Hz-Bereich aufgezeichnet werden, um darauf eine Offline-Auswertung durchzuführen. Wäre dies möglich, könnte für die experimentelle Verifikation der Bewegungstereo-Algorithmen auf die aufwendige Realisierung einer eigenen, realzeitfähigen Merkmalsextraktion verzichtet werden. Ein Online-Bewegungstereo-System könnte damit jedoch nicht erstellt werden.

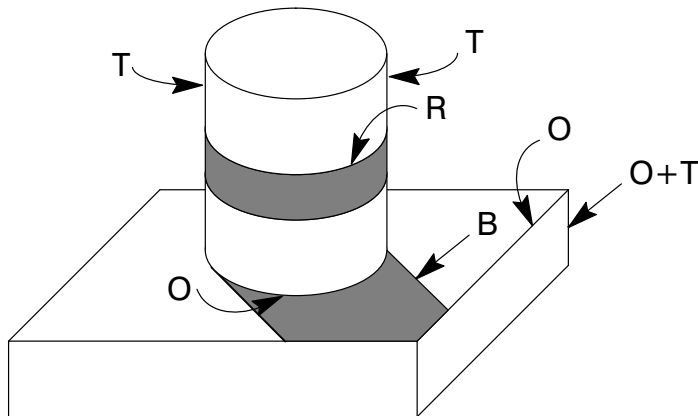


Abb. 5: Physikalische Ursachen von Kanten in Grauwertbildern

- (B) Unstetigkeiten in der Beleuchtung (Schatten, Glanzreflexionen, Lichtquellen),
- (R) Unstetigkeiten in den Texturen und Reflexionseigenschaften auf Oberflächen,
- (T) Unstetigkeiten in der Tiefe, sowie
- (O) Unstetigkeiten in der Oberflächenorientierung

Um die Struktur einer Szene zu analysieren, interessieren nur Kanten vom Typ T und O. Allerdings ist es schwierig, die Kanten des Typs B und R herauszufiltern, nicht zuletzt deshalb, weil üblicherweise diese sehr viel stärker heraustreten. Bei der Interpretation des Kantenbildes besteht dann die Gefahr von Mehrdeutigkeiten und Täuschungen, die ein menschlicher Beobachter oft nur durch sein enormes Hintergrundwissen und seine Erfahrung auflösen kann.

Gelegentlich sind Kanten, die durch Schlagschatten oder Muster auf Oberflächen entstehen, auch von Vorteil, dann nämlich, wenn die Kanten lang, herausragend, und zeitlich stabil sind. Es entstehen dann im Dreidimensionalen komplanare Oberflächenkanten. Damit besitzt dann eine ansonsten schwach strukturierte Szene eine höhere Merkmalsdichte, die die Lokalisation verbessern hilft. Nachdem derartige Kanten von der Merkmalsextraktion nicht von Kanten des Typs T und O unterschieden werden können, ergibt sich dafür auch keine Sonderbehandlung. Kanten können unabhängig von ihrem Typ ihre Lage verändern. Dies muß in der Auswertung berücksichtigt werden².

3.1.3. Prinzip der Konturverfolgung

Die Konturverfolgung gehört zur Klasse der selbststeuernden, sequentiellen Bildsegmentation [95], bei der Ergebnisse vorhergehender Segmentationschritte beim aktuellen Schritt verwendet werden und dieser dadurch weniger Aufwand erfordert. Das Grundprinzip besteht

2. In der vorliegenden Arbeit wurde eine Bewegung der Kanten während einer Aufnahmesequenz durch Einschränkung auf quasistatische Umgebungen ausgeschlossen, und Bewegungen zwischen verschiedenen Aufnahmesequenzen dem Problem Modellanpassung zugeordnet.

darin, einen Bildpunkt zu finden, der auf einer Kontur liegt, und aus der lokalen Umgebung einen Folgepunkt zu bestimmen, der ebenfalls auf der Kontur liegt. Die Konturpunkte werden anschließend gruppiert und mit einfachen geometrischen Formen approximiert, z.B. durch Geradenstücke, Kreisbögen oder Splines. Vorteil dieses Verfahrens ist, daß nur an aussichtsreichen Bildpunkten die lokale Umgebung untersucht wird, so daß letztlich nur ein geringer Teil der Bildpunkte betrachtet werden muß.

Die lokale Betrachtung bei der Extraktion von Kanten ist im übrigen sehr gut der Problemstellung der Kantenextraktion angepaßt, denn Kanten sind per Definition durch *lokale* Gradientenextrema gekennzeichnet.

Insbesondere für die Echtzeitanwendung ist von Nachteil, daß der Rechenaufwand pro Bild vom Inhalt abhängig ist, wodurch eine Zeitabschätzung erschwert wird: Je strukturierter das Bild ist, d.h. je mehr Kanten sich im Bild befinden, desto mehr Konturpunkte müssen bearbeitet werden, und desto länger dauert die Konturextraktion.³ Allerdings wird der Rechenaufwand niemals größer sein als bei der klassischen Vollbildverarbeitung, bei der grundsätzlich jeder Bildpunkt betrachtet wird, so daß die Konturverfolgung aus Sicht des Rechenaufwands stets besser abschneidet.

Das Prinzip der Konturverfolgung geht zurück auf die Beschreibung von 2-dimensionalen Kurvenformen durch Kettencodes (Freeman [28]). Auf die Anwendbarkeit der Konturverfolgung in der Bildverarbeitung ist schon länger hingewiesen worden, (Rosenfeld/Kak [86], Shirai [95]), und auch die notwendigen Schritte bei der Verfolgung einer Kontur sind wohlbekannt:

- Suche eines Startpunkts
- Prädiktion des Folgepunktes
- Prädiktionsunterstützte Detektion des Folgepunktes

Allerdings fehlen in beiden Quellen konkrete Angaben oder Querverweise zur Bestimmung des Folgepunktes, oder über Verfahren zur Startpunktsuche.

Konturverfolgung wird in der Bildverarbeitung im Rahmen der Kettencode-Erzeugung seit langem erfolgreich für binarisierte Bilder angewendet, z.B. bei der Schrifterkennung oder der Digitalisierung technischer Zeichnungen, also Anwendungen, bei denen eine Schwellwertbildung auf einfache Art möglich ist. Hierfür existiert sogar eine Hardware-Implementierung [45]. Weitere typische Anwendungen sind die Extraktion von Flächenmerkmalen, wie z.B. Schwerpunkt, Momente, Formparameter [28],[62].

3. Diese Aussage muß bei der Diskussion der Startpunktsuche in 3.1.7. eingeschränkt werden.

Die direkte Anwendung der Konturverfolgung auf Grauwertbilder ohne vorherige Schwellwert-Binarisierung ist seltener. Martelli [69] verwendet in seinem Verfahren den lokalen Gradienten, um den nächsten Konturpunktkandidaten zu bestimmen. Überschreitet dort der lokale Gradient einen Schwellwert, ist ein neuer Konturpunkt gefunden. Andernfalls wird anhand des mittleren Grauwertes um den Konturpunktkandidat festgestellt, ob man sich innerhalb oder außerhalb des "Objekts" befindet, und der Gradiententest an einem anderen Konturpunktkandidat durchgeführt, der sich vermutlich näher an der Kontur befindet. Dieser Algorithmus vermeidet zwar die Schwellwertbildung für *alle* Pixel, ist aber für die Verfolgung der Konturen von kontrastreichen Objekten ausgelegt.

Raubenheimer stellt in [83] ein Verfahren vor, das mit Hilfe lokal integrierender, anisotroper Bildvorverarbeitung auf optomechanischer Basis⁴ bzw. optoelektronischer Basis⁵ den Bewegungszustand eines Punktes, der sich entlang einer Kontur bewegt, beobachtet und fort schreibt. Die Konturen werden dabei durch ein Kreisbogenmodell im Rahmen der Kalman-Filter-Theorie modelliert. Vor allem aufgrund des nicht-realzeitfähigen Meßverfahrens eignet sich diese Methode primär nur für offline-Bildauswertungen.

In Ballard&Brown [4] ist die Konturverfolgung mit Hilfe der dynamischen Programmierung (DP) beschrieben: Die Gütefunktion bewertet die möglichen benachbarten Folgepunkte, wobei ein lokaler Kantenoperator den Betrag und die Richtung der Kante bestimmt. Im eigentlichen Verfolgungsschritt wird dann mit Hilfe des DP-Algorithmus [7] der optimale Pfad durch die Konturpixel, und damit die Konturkette, gefunden. Allerdings wird die Bewertungsfunktion an sehr viel mehr Bildpunkten berechnet, als dann tatsächlich Konturpunkte entstehen, im Extremfall muß der Kantenoperator auf das komplette Grauwertbild angewendet werden. Dies verletzt die Forderung nach einem effizienten Algorithmus.

3.1.4. Konturverfolgung mit einem physikalischen Analogon

Zur Bestimmung des nächsten Konturpunktes und als Kriterium für die Startpunktsuche wird bei dem implementierten, neuen Verfahren eine Analogie aus der Elektrodynamik verwendet [46]: Ein geladenes Teilchen wird durch Kräfte in der 2-dimensionalen Bildebene entlang einer Kontur bewegt, wobei die erzeugte Spur der zu detektierenden Kontur entspricht. Die Kräfte entstehen aus der Interaktion des Teilchens mit einem elektrischen und einem magnetischen Feld. Diese Felder sind keine Felder im physikalischen Sinn und haben auch nur die Eigenschaften gemeinsam, die ausdrücklich erwähnt sind. Allerdings werden sie der Einfachheit halber weiterhin als elektrisches und magnetisches Feld bezeichnet.

4. Rotieren einer Blende mit keilförmigem Spalt
5. Random-Access-Kamera, bei der an jedem Bildpunkt durch Variation der Ablenkspannungen ein sternförmiges Auslesemuster erzeugt wird, entlang dessen Strahlen die Ladungsmenge gemessen wird

Die Felder werden durch elektrische Dipole und Ströme erzeugt, deren Lage durch das Pixelraster festgelegt ist, und die durch die Grauwerte in den Pixeln parametrisiert sind. Die Feldberechnung erfolgt nach dem Coulomb-Gesetz und dem Biot-Savart-Gesetz, und bestimmt sich aus den Ladungen, und damit den Grauwerten in der näheren Umgebung eines Konturpunktes, und den Abständen des Teilchens zu den Ladungen.

Auf das Teilchen wirken eine horizontale und eine vertikale Komponente des elektrischen Feldes, sowie der Anteil des magnetischen Feld senkrecht zur Bildebene. Komponenten des magnetischen Feldes parallel der Bildebene brauchen nicht berücksichtigt zu werden, denn die resultierende Lorentz-Kraft auf das Teilchen steht senkrecht zum Feldvektor und dem Geschwindigkeitsvektor. Da sich das Teilchen nur in der Bildebene bewegen kann, ist nur die Komponente der Kraftwirkung in der Bildebene relevant, und die wird gerade durch den Anteil des magnetischen Felds senkrecht zur Bildebene ermittelt.

$$\vec{F} = e \cdot \left(\vec{E} + \mu_0 \mu_r \cdot \vec{H} \times \vec{v} \right) \quad (1)$$

Die exakte Umsetzung des Grauwertbildes in elektrische Ladungen und Ströme, die Herleitung der Kraftwirkung auf das Teilchen, und die Implementierung der Feldberechnung aufgrund der lokalen Umgebung mit Hilfe von drei 5x5 Faltungskernen wird im Anhang A.1. auf Seite 100 dargelegt.

An einem gegebenen Konturpunkt $\vec{x}(i)$, an dem das Teilchen die Geschwindigkeit $\vec{v}(i)$ hat, soll der Punkt $\vec{x}(i + 1)$ bestimmt werden. Die Geschwindigkeitsänderung, die durch die angreifenden Kräfte bewirkt wird, bestimmt sich zu:

$$\Delta \vec{v}(i) \approx \frac{1}{m_e} \cdot \int_{t_i}^{t_{i+1}} F(i) dt = \frac{e}{m_e} \cdot T \cdot \left[\vec{E}(i) + H(i) \cdot \mu_0 \mu_r \cdot \begin{pmatrix} -v_y(i) \\ v_x(i) \end{pmatrix} \right] \quad (2)$$

wobei $\frac{e}{m_e}$ der Quotient von Teilchenladung und Teilchenmasse darstellt, und eine einstellbare Konstante ist. Das Vorzeichen von $\frac{e}{m_e}$ kann wechseln, um eine positive oder negative Ladung des Teilchens zu modellieren. Dadurch ist es möglich, auf einfache Weise die Verfolgungsrichtung zu ändern. T ist die Integrationszeit für jeden Verfolgungsschritt. Die Geschwindigkeit des Teilchens wird normiert auf:

$$\vec{v}(i + 1) = \frac{\vec{v}(i) + \Delta \vec{v}(i)}{|\vec{v}(i) + \Delta \vec{v}(i)|} \cdot v_s \quad (3)$$

wobei $v_s \cdot T$ eine feste Schrittweite ist. Sie entspricht der gewünschten Distanz zwischen diskreten Konturpunkten.

Der nächste Konturpunkt berechnet sich demnach zu:

$$\vec{x}(i + 1) = \vec{x} + T \cdot \vec{v}(i + 1) \quad (4)$$

Dort wird anhand der berechneten Felder und eines Labelbildes geprüft, ob das Ende der Kontur erreicht ist, ob auf eine bereits extrahierte Kontur getroffen wurde, oder ob ein neuer Konturpunkt bestimmt werden muß.

Die folgenden Bilder zeigen den 4-fach vergrößerten Ausschnitt von 184x143 Pixel des na-



Abb. 6: Grauwertbild



Abb. 7: Betrag des E-Felds

türlichen Grauwertbildes mk6_00001 (736x574 Pixel) aus Anhang A.3., sowie die daraus resultierenden Felder $|E| = \sqrt{E_h^2 + E_v^2}$, E_v , E_h , und H , die die Konturverfolgung steuern.

Bei der Darstellung des H , E_h , E_v -Feldes bedeutet Grau einen Amplitudenwert des jeweiligen Feldes von 0, weiß einen positiven Wert, und schwarz einen negativen. Die resultierende Spur des Teilchens ist in Abb. 11 dargestellt.

Der Absolutbetrag des elektrischen Feldes in Abb. 7 ist augenscheinlich vergleichbar mit dem Gradientenbetrag, der bei anderen Verfahren der Kantenextraktion verwendet wird [11], [47]. In 3.1.6. wird gezeigt, daß ohne Qualitätseinbuße in der Konturverfolgung tatsächlich klassische Gradientenmatrizen zur E -Feldberechnung, und mit Einschränkungen auch für die H -Feldberechnung verwendet werden können. Im übrigen entspricht der E -Feldbetrag bereits grob den zu extrahierenden Konturen: Er ist vergleichbar mit dem Zwischenergebnis, das nach der Schwellwertbildung der Filterantwort bei [11] oder [2] entsteht: In nachfolgenden

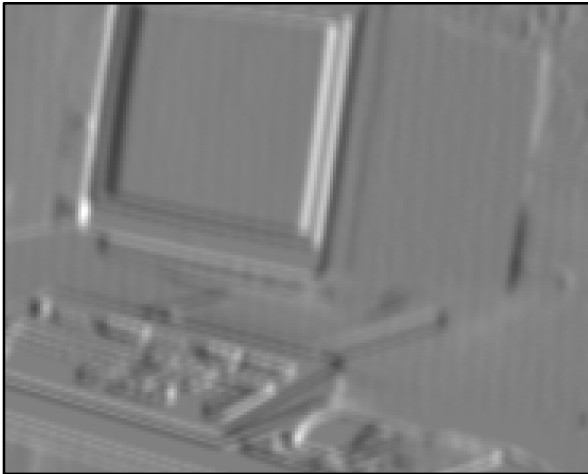


Abb. 8: vertikales E-Feld

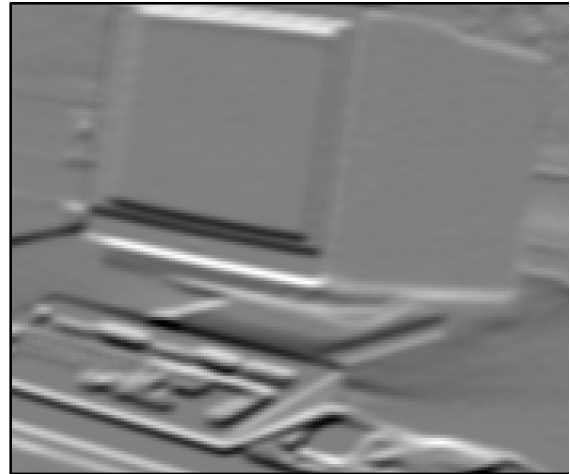


Abb. 9: horizontales E-Feld



Abb. 10: H-Feld

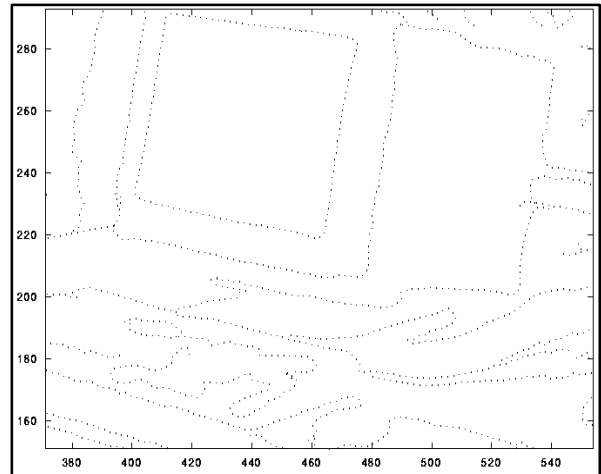


Abb. 11: Konturpunkte

Schritten werden dort dann Nebenmaxima unterdrückt resp. die Kontur ausgedünnt, Kantenpunkte verkettet und anschließend mit Geradenstücken approximiert.

Der resultierende elektrische Feldvektor liegt in der Nähe von Konturen parallel zur Tangente an die Kontur. Abseits von Konturen bestimmt primär das Bildrauschen den Feldvektor, und macht ihn für Auswertungen ungeeignet. Das stört aber nicht weiter, denn man will ja gerade die Konturen extrahieren. Darüber hinaus ist der Feldbetrag an Konturen um eine Größenordnung größer.

Die Diskussion der Feldberechnung erfolgt anhand von idealen, vertikalen Rampenkanten unterschiedlicher Breite, wie sie in Abb. 12 dargestellt sind. Abb. 13 zeigt exemplarisch einen Ausschnitt der Rampenkante mit 7 Stufen.

Die E-Feldberechnung aus den Grauwerten ist so gestaltet, daß entlang einer Kontur ein Feldprofil entsteht, bei dem in Konturmitte der größte Feldbetrag auftritt, der sich zum Rand hin

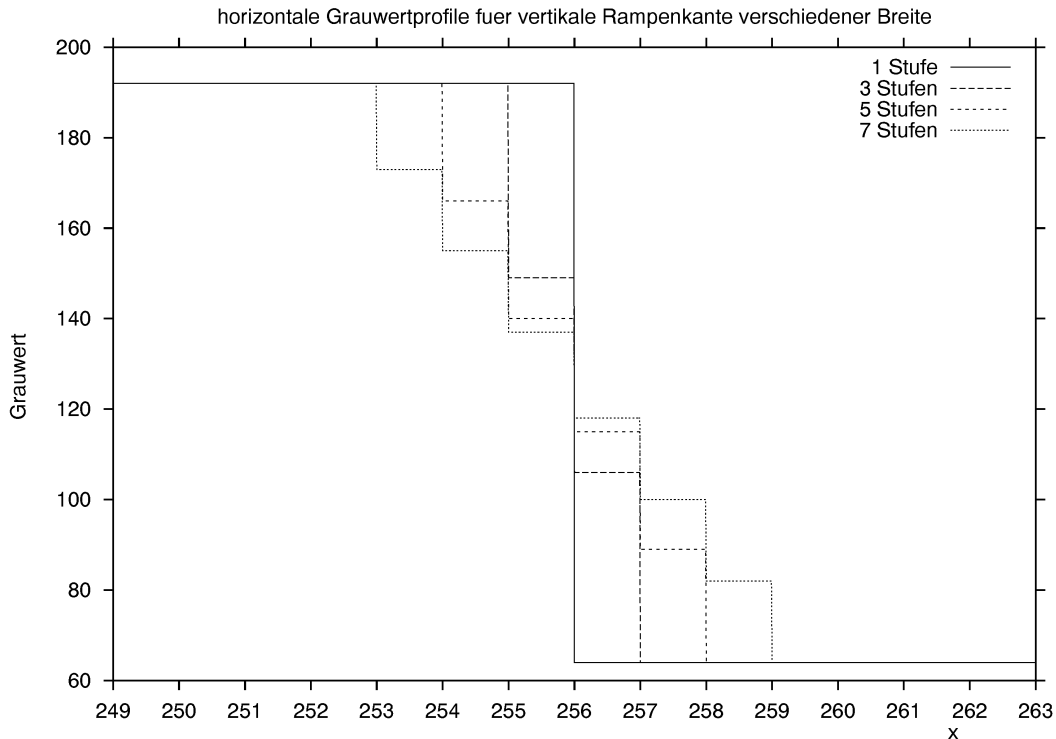


Abb. 12: Grauwertprofil für vertikale Rampenkanten unterschiedlicher Breite



Abb. 13: Synthetisches Grauwertbild Rampenkante mit 7 Stufen

reduziert. Abb. 14 zeigt die vertikale E-Feldkomponente. Die horizontale E-Feldkomponente dieser Profile ist konstant Null. Für nicht-achsenparallele Konturen gilt die nachfolgende Betrachtung ebenfalls, allerdings muß dann von tangentialen und transversalen E-Feld-Beiträgen gesprochen werden.

Das E-Feld bewirkt, daß ein Teilchen entlang einer Kontur vorwärts bewegt wird, und auch Konturkrümmungen folgen kann. Damit ist jedoch nicht sichergestellt, daß das Teilchen sich auf dem Grat der Kontur, also entlang des maximalen Gradienten bewegt, und daß es bei starken Krümmungen die Kontur nicht verläßt. Diese Stabilisierungsaufgabe wird mit Hilfe des magnetischen Feldes gelöst.

Das Profil des magnetischen Feldes der idealen Rampenkanten aus Abb. 12 zeigt Abb. 15. Man kann den Verlauf in erster Näherung als zweite Ableitung des Grauwertes interpretieren: Bis zu einem gewissen Abstand vom Gradientenmaximum wird der H-Feld-Beitrag größer.

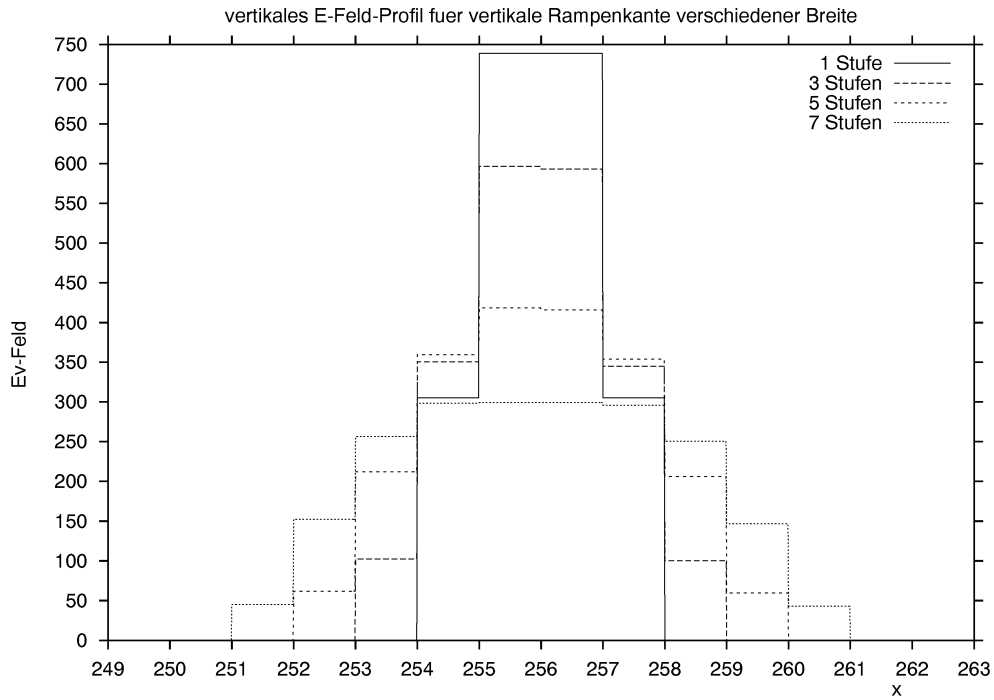


Abb. 14: vertikales E-Feld-Profil

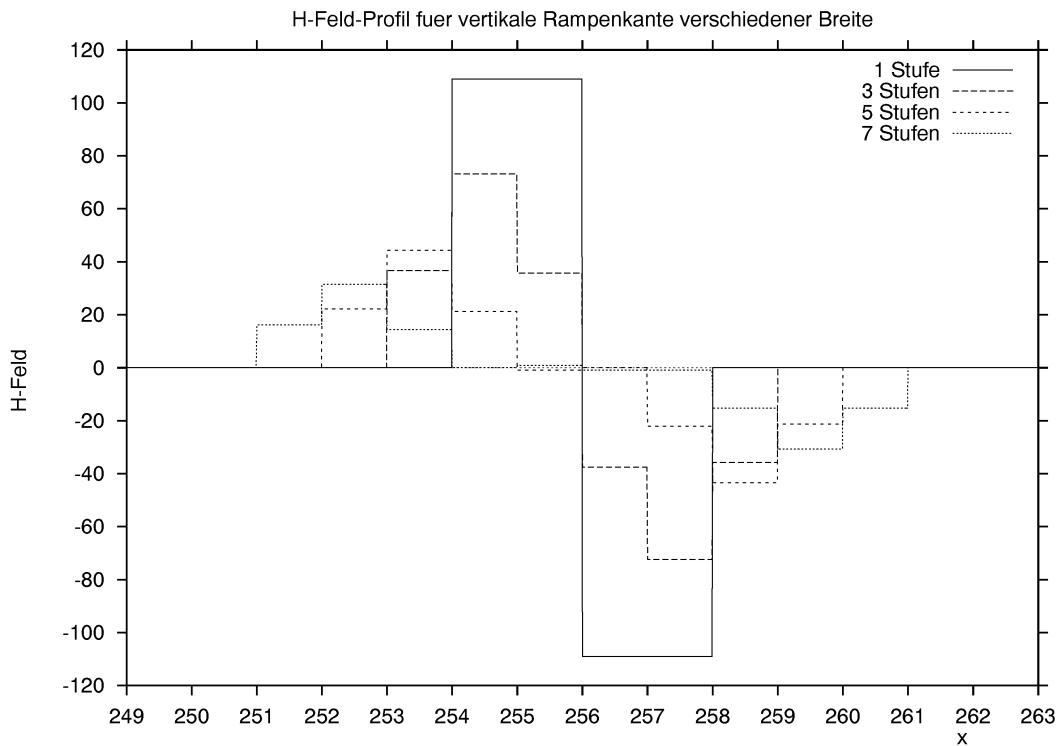


Abb. 15: H-Feld-Profil

Dieser bewirkt zusammen mit der Eigengeschwindigkeit des Teilchens nach (2), daß eine Geschwindigkeitsänderung in Richtung Konturmaximum erfolgt. Diese Beschleunigung ist um so höher, je größer der Abstand ist. Sie ist weiterhin unabhängig von der Verfolgungsrichtung, denn Ladungsvorzeichen und Vorzeichen des Geschwindigkeitsvektors heben sich in (2) auf.

Der nach außen abnehmende Betrag des E-Feldes verstärkt über die Summation der angreifenden Kräfte am Teilchen und der anschließenden Normierung des Betrags der Geschwindigkeit den Einfluß des magnetischen Feldes und trägt zur weiteren Stabilisierung der Bahn bei.

Bei dem implementierten Verfahren handelt es sich um die Querregelung der Bewegung: Entlang der Kontur nimmt das Teilchen an diskreten Orten Proben der kontinuierlichen, ortsvarianten, lokalen Felder und regelt seine Lage in der von der Kontur erzeugten Potentialrinne. Konturkrümmungen werden als Störungen interpretiert, die bewirken, daß das Teilchen aus seiner Sollage im Konturmaximum ausgelenkt wird, und durch die Regelung wieder dorthin zurückgeführt werden soll.

Zur Identifikation der Art des Reglers wird das H-Feld als die zweite Ableitung angesehen, obwohl das nicht ganz korrekt ist, wie in 3.1.6. gezeigt wird. Die Art der Querregelung hängt dann wiederum vom Grauwertprofil der Kontur ab: Das von einer idealen Stufenkante erzeugte H-Feld-Profil nach Abb. 15 bewirkt eine Regelung mit einem 2-Punkt-Regler ohne Hysterese, eine Rampenkante mit 3 Stufen im Bereich von ± 1.5 Pixel um das Konturmaximum eine Regelung mit einem P-Regler, usw. Kurvenformen natürlicher Kanten, die sich aus Stufen und Rampen zusammensetzen lassen, erzeugen dann Mischformen dieser Regler. Ein sich im Verlauf der Kontur änderndes Grauwertprofil erzeugt dann bei Bedarf ein unterschiedliches Querregelungsverhalten. Allen Regelungen gemeinsam ist eine Totzeit, die der Schrittlänge entspricht. Der Betriebsbereich für die Querregelung ist relativ klein; er beträgt ± 2.5 Pixel und entspricht genau der Hälfte der Matrixgröße. Außerhalb dieses Bereichs wird der Regler instabil, und das Teilchen verläßt die Kontur.

Der Entwurf eines Bahnreglers für Potentialfelder am Beispiel eines Gleitreglers wurde für die Bahnplanung von holonomen Robotern von Guldner[32] bereits ausführlich untersucht, und kann prinzipiell auf das hier vorliegende Problem angewendet werden, um die implementierte, funktionsfähige Lösung weiter zu verbessern.

Abb. 16 zeigt einen Ausschnitt aus der Bahn des Teilchens entlang zweier Rampenkanten (1-stufig und 5-stufig) aus Abb. 12. Der Startpunkt liegt für beide Konturen bei $y=508$. Bei der Konturpunktfolge der 5-stufigen Rampe im linken Teil von Abb. 16 ist eine Ablage von der Konturmitte bei $x=256.0$ von einem Pixel zu erkennen. Wie aus den Feldprofilen in Abb. 14 und Abb. 15 hervorgeht, verschwindet für $x > 255.0$ das H-Feld nahezu. Numerische Auslöschung aufgrund des differentiellen Charakters der H-Feldmatrix bewirken im Bereich $255.0 \leq x < 256.0$ einen negativen Feldwert. Dadurch entsteht ein schwach instabiles Verhalten, wodurch der Konturpunkt nach außen abwandert. Wird $x = 255.0$ unterschritten, bewirkt das an dieser Stelle sprunghaft angestiegene, positive H-Feld eine Korrektur von mehreren Zehntel Pixel in Richtung Konturmitte, wodurch der sägezahnartige Konturpunktverlauf resultiert.

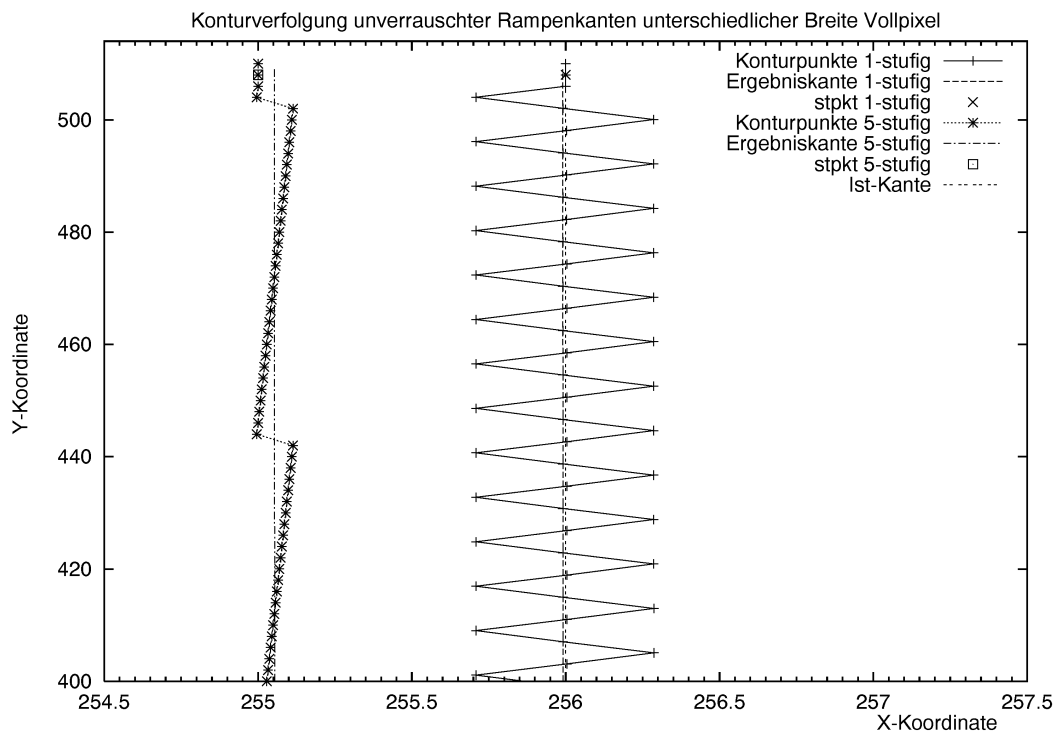


Abb. 16: Konturen einer künstlichen, vertikalen Rampenkante

Für die 1-stufige Rampe ist das deutlich erkennbare Schwingen des Konturpunkts um das jeweilige Konturmaximum ein weiterer nennenswerter Effekt. Dies ist eine Wirkung der starken Amplitudendiskretisierung der Feldberechnung. Innerhalb des Querregelungsbereiches von ± 2.5 Pixel gibt es aufgrund der 5×5 -Matrizen lediglich vier diskrete Amplitudenwerte pro Feldgröße, die zur Bahnregelung verwendet werden können. Dadurch entstehen Regelschwingungen und möglicherweise ein Gleichanteil bei der Ablage von der wahren Kontur. In Abb. 16 rechts treten nur zwei der vier möglichen Amplitudenwerte auf, und aufgrund der idealen Verhältnisse entsteht keine nennenswerte Ablage von der Kontur. Die Ausgleichsgerade durch die Konturpunkte (Abschnitt 3.1.8.) mittelt bei ausreichend langen Kanten das Schwingen heraus. Bei kurzen Kanten ist jedoch mit Fehlern zu rechnen. Abhilfe schafft die Einführung der Subpixelberechnung, mit der bei nahezu gleichem Online-Rechenaufwand eine beliebig feine Diskretisierung der Feldamplituden erreicht wird, und die im folgenden Abschnitt erläutert wird.

3.1.5. Feldberechnung mit Subpixelauflösung

Die Berechnung der Feldgrößen durch Faltung des Grauwertbildes mit 5×5 -Matrizen ist genau genommen nur in einer kleinen Umgebung des Pixelmittelpunktes korrekt. Zu den Pixelrändern hin vergrößern sich die Fehler, und an der Pixelgrenze entstehen Unstetigkeiten, die eine korrekte Bahnregelung erschweren.

Zur Erzeugung subpixelgenauer Feldgrößen werden die Größen zwischen den 4 umliegenden Vollpixel-Mittelpunkten zweidimensional interpoliert (Abb. 17). Hierzu wird das Vollpixel-

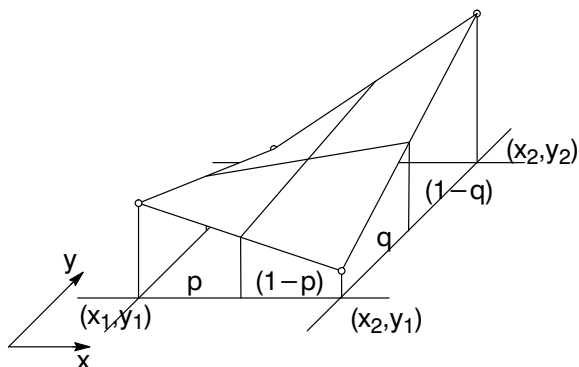


Abb. 17: 2-dimensionale, lineare Interpolation

$$f(x_1 + p, y_1) = f(x_1, y_1) + p * (f(x_2, y_1) - f(x_1, y_1)) \quad (5)$$

$$f(x_1 + p, y_2) = f(x_1, y_2) + p * (f(x_2, y_2) - f(x_1, y_2)) \quad (6)$$

$$f(x_1 + p, y_1 + q) = f(x_1 + p, y_1) + q * (f(x_1 + p, y_2) - f(x_1 + p, y_1)) \quad (7)$$

raster in Subpixel aufgeteilt, deren Anzahl der Einfachheit halber eine Zweierpotenz ist. Dabei muß eine Phasenverschiebung um 0.5 Subpixeleinheiten berücksichtigt werden, damit die Berechnung des Feldwertes im Zentrum des Subpixels stattfindet, und die Subpixel konform zum Vollpixelraster zu liegen kommen. Abb. 18 zeigt die Aufteilung am Beispiel zweifacher

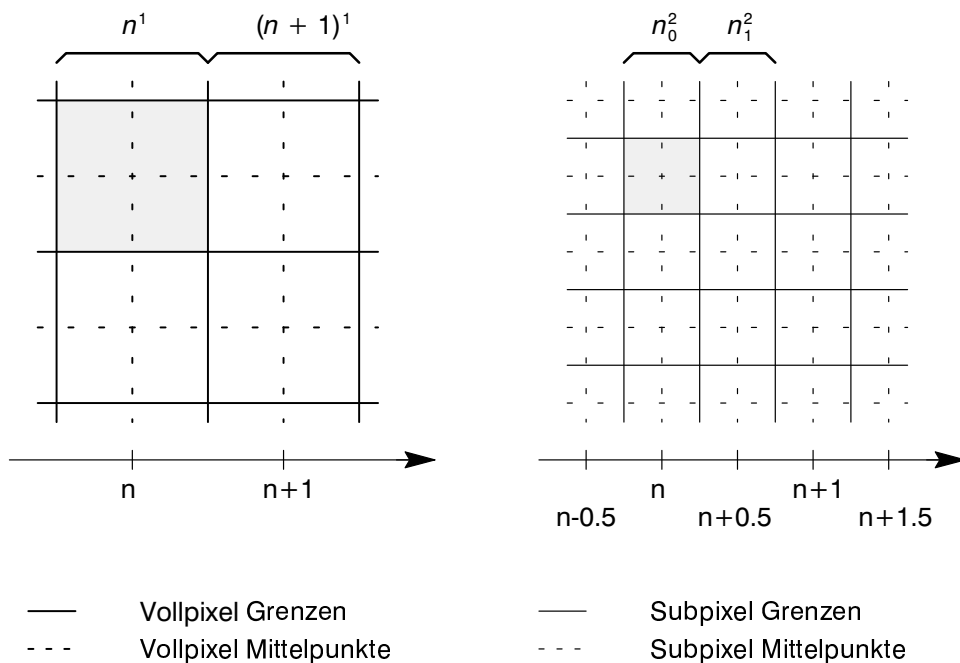


Abb. 18: Aufteilung der Vollpixel in Subpixel

Subpixel-Auflösung.

Die Abbildung von kontinuierlichen Bildkoordinaten in Vollpixelkoordinaten und Subpixelindizes errechnet sich demnach wie folgt:

$$t = 0.5 - \frac{1}{2 \cdot subp} \quad (8)$$

$$i = \lfloor f - t \rfloor \quad (9)$$

$$s = \lfloor (f - t - i) \cdot subp \rfloor \quad (10)$$

Dabei ist f der kontinuierliche Positionswert, i die ganzzahlige Vollpixelposition, an der die Faltungsmatrix angewendet wird, s der Subpixel-Index, $subp$ die eingestellte Anzahl der Subpixel innerhalb eines Vollpixels, $\%$ der Modulo-Operator und $\lfloor \cdot \rfloor$ die Entier-Funktion (Abschneiden zur nächstkleineren ganzen Zahl). Die 2-dimensionale, lineare Interpolation erfolgt dann gemäß (5) bis (7). Dabei entsprechen die Gewichte p und $(1-p)$ dem Teilungsverhältnis der auf 1 normierten Abstände zwischen Vollpixelpositionen und berechnen sich aus dem Quotient von Subpixelindex und Subpixelanzahl.

Eine Interpolation der Feldwerte der vier umliegenden Vollpixel zur Laufzeit bedeutet mehr als den vierfachen Zeitaufwand, denn es muß für jeden Feldwert eine separate Faltung durchgeführt werden. Zur Beibehaltung der Effizienz wird die Interpolation offline vorbereitet: Für jede Subpixelposition innerhalb eines Vollpixels wird ein eigener Matrizensatz erzeugt, der aus den beiden E-Feld- und einer H-Feldmatrix besteht. Die Subpixel-Matrizen entstehen durch die gewichtete Überlagerung der verschobenen Vollpixelmatrizen. Die Gewichte sind die gleichen, wie bei der Interpolation der vier Vollpixel-Feldgrößen. Zur Laufzeit wird aufgrund des Subpixelindex s ein Matrizensatz ausgewählt und damit die Faltung an der errechneten Vollpixelposition i durchgeführt.

Durch geeignetes Runden nach Gleichungen (8), (9), (10) kann die Vergrößerung der Matrizen von ursprünglich 5x5 bei Vollpixelauflösung auf 7x7 verhindert werden, zumal dann die erste bzw. letzte Zeile oder Spalte immer Nullelemente enthalten würde. Aber auch bei 6x6-Matrizen gibt es von den $(subp \cdot subp)$ erforderlichen Matrizensätzen $(2 \cdot subp + 1)$ Sätze, bei denen die letzte Zeile oder Spalte nur Nullelemente enthält, nämlich an den Positionen, die für x oder y der Vollpixelposition entspricht. Will man bei der Feldberechnung bei diesen Sonderfällen eine Einsparung durch Berechnung mit 5x5-, 5x6- oder 6x5-Matrizen erzielen, muß an jedem Konturpunkt eine Fallunterscheidung vorgenommen werden. Der hierfür erforderliche Aufwand ist größer als die mögliche Einsparung, die zudem mit steigender Subpixel-Auflösung weniger wahrscheinlich wird. Das führt dazu, daß grundsätzlich immer mit einer 6x6-Matrix gerechnet wird.

Abb. 19 zeigt exemplarisch für die anderen Feldgrößen den stückweise konstanten H-Feldverlauf bei unterschiedlichen Subpixelauflösungen anhand des Profils der synthetischen

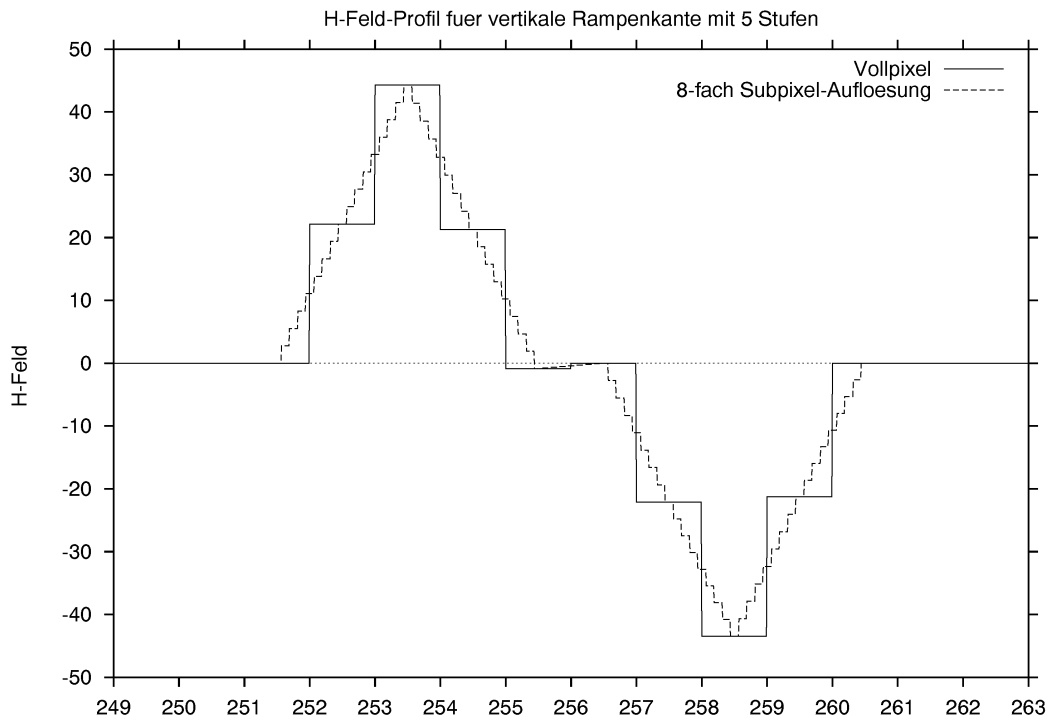


Abb. 19: H-Feld-Profil mit Subpixel

Rampenkante mit 5 Stufen aus Abb. 12. Man erkennt die feinere Granulierung, die ein reduziertes Überschwingverhalten bei der Konturverfolgung bewirkt. Abb. 20 zeigt die Verhält-

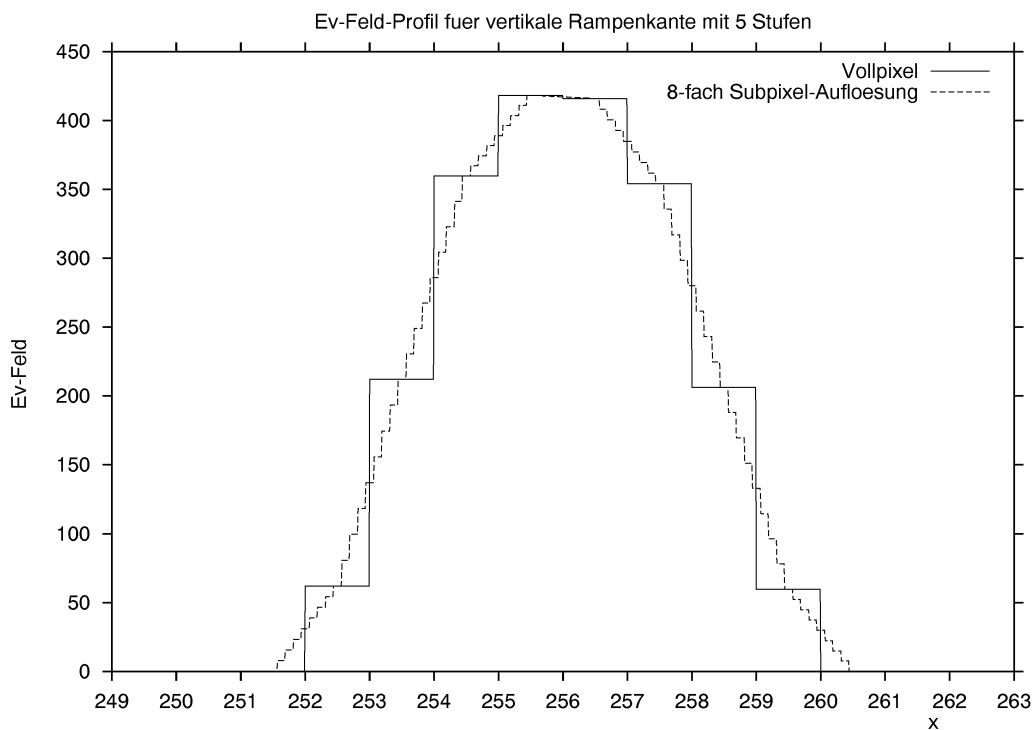


Abb. 20: vertikales E-Feld-Profil mit Subpixel

nisse im Verlauf des E-Feldes. Die Unsymmetrien bei $x=256.0$ entstehen durch die o.g. numerischen Effekte bei der Faltung.

In Abb. 21 ist die Auswirkung der subpixelgenauen Berechnung der Feldgrößen auf die Kon-

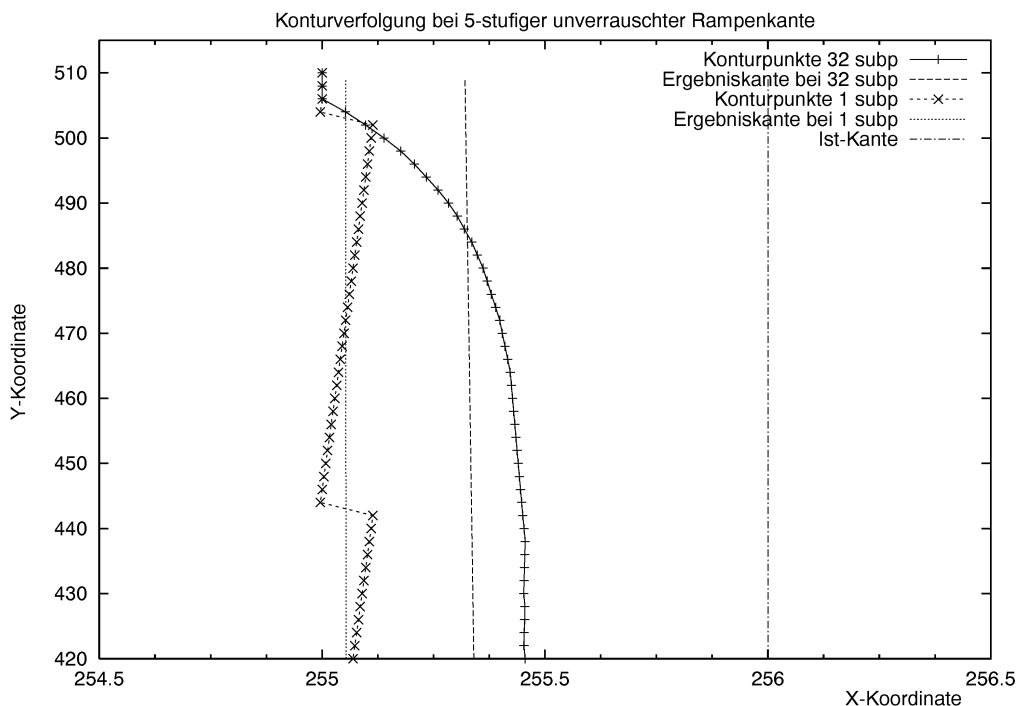


Abb. 21: Konturpunkte mit/ohne Sub-Auflösung

turpunktfolge zu erkennen. Für beide Subpixel-Auflösungen liegt der Startpunkt bei $y=508$. Allerdings konvergiert bei der höheren Auflösung die resultierende Kante zu $x = 255.5$. Ursache ist die Kompensation des Vorzeichenfehlers im Bereich $255.0 \leq x < 255.5$ aufgrund der linearen Interpolation. Für den Bereich $255.5 \leq x < 256.0$ ist auch hier der Rauschfehler negativ, und bewirkt den Ablagefehler.

Der Vorteil der subpixelgenauen Feldberechnung ist der reduzierte Bereich, in der wegen des verschwindenden H-Feldes keine Querstabilisierung erfolgt, obwohl eine signifikante Ablage von der Konturmitte vorliegt.

3.1.6. Abbildung der Feldberechnung auf klassische Operatoren

Die elektrodynamischen Felder und Kräfte, die den Konturpunkt entlang einer Kontur weiterbewegen, entsprechen augenscheinlich Größen, die in der klassischen Bildverarbeitung wohlbekannt sind: Das horizontale und vertikale elektrische Feld entspricht dem lokalen horizontalen und vertikalen Gradienten, und das magnetische Feld der zweiten Richtungsableitung. Aus diesem Grund wurde versucht, die Feldberechnung alternativ mit aus der Literatur bekannten Matrizen durchzuführen. Aus Gründen der Vergleichbarkeit wurden hierzu ebenfalls 5x5 Matrizen verwendet. Die subpixelgenaue Berechnung nach 3.1.5. erfordert dann wieder 6x6 Matrizen.

Die Wahl für die Gradientenmatrizen fiel auf den Kirsch-Operator für horizontale und vertikale Kanten, der folgendes Aussehen hat (vgl. [4], S. 79):

1	1	1	1	1	1	1	0	-1	-1
1	1	1	1	1	1	1	0	-1	-1
0	0	0	0	0	1	1	0	-1	-1
-1	-1	-1	-1	-1	1	1	0	-1	-1
-1	-1	-1	-1	-1	1	1	0	-1	-1

Abb. 22: horizontale und vertikale Kirschmatrizen

Abb. 23 zeigt repräsentativ einen Teil eines vertikalen Schnitts durch den horizontalen E-

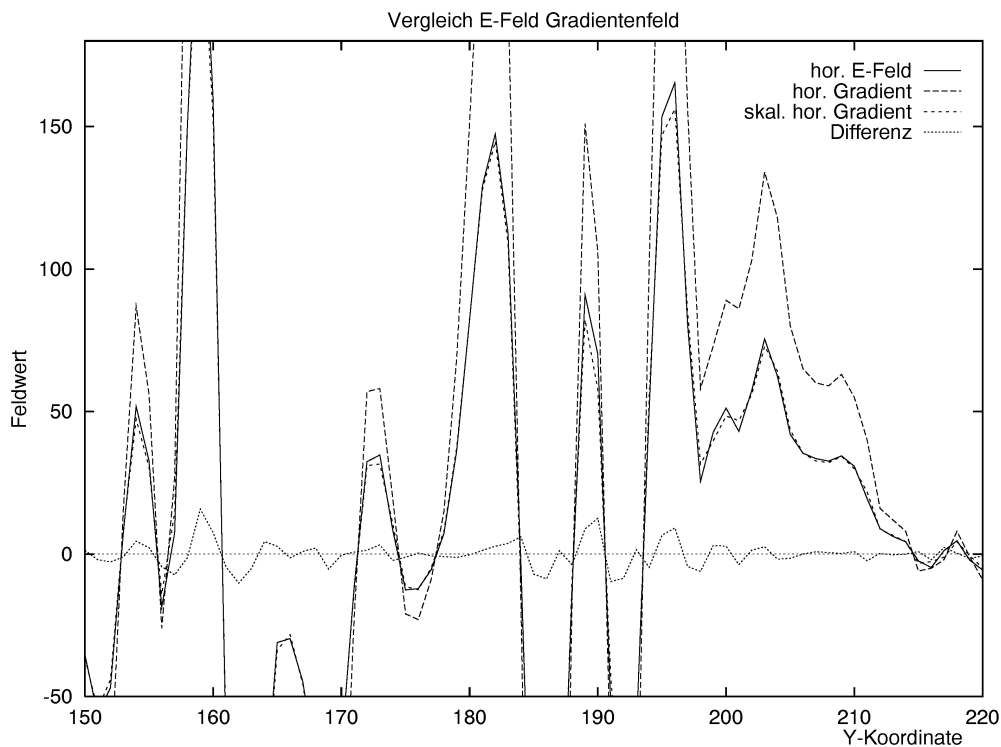


Abb. 23: E-Feldverlauf mit Kirsch-Matrizen

Feld-Verlauf und dem horizontalen Gradientenverlauf bei $x=70$ des Testbildes bf_00001 in Anhang A.3.. Die resultierenden Feldverläufe unterscheiden sich nur durch einen Skalar, der sich aus dem Mittelwert der pixelweise gebildeten Quotienten von E-Feld-Bild und Gradientenbild bestimmt. Die absolute Skalierung der Gradienten ist für die Konturverfolgung irrelevant, vielmehr ist das Verhältnis zum magnetischen Feld wichtig. Abb. 23 zeigt auch den korrigierten Gradientenverlauf gegenüber dem E-Feldverlauf, sowie die Differenz von E-Feld und skaliertem Gradientenfeld. Bis auf Übersteuerungseffekte, die mit der Evaluierungsmethode zusammenhängen, unterscheiden sich E-Feld und skaliertes Gradientenfeld nur sehr geringfügig. Die Kirsch-Matrizen haben daher die gleiche Wirkung, wie die E-Feld-Matrizen. Signifi-

kanter Vorteil der ternären Kirsch-Operatoren ist die sehr einfache und effiziente Implementierbarkeit: im Gegensatz zu 5x5 Fließkomma-Multiplikationen- und Additionen kann bei den Kirsch-Matrizen mit Integerwerten gerechnet werden, und es sind nur Additionen erforderlich. Je nach Zielarchitektur kann eine Leistungssteigerung von bis zu 50% erzielt werden. Allerdings verfügen moderne Prozessorarchitekturen über Floating-Point-Rechenwerke, deren Geschwindigkeit der der Integer-Rechenwerke entspricht, oder bei superskalaren Architekturen deren Geschwindigkeit übertrifft. Die Fließkomma-Berechnung ist dann also nicht langsamer, und der qualitative Vorteil der Subpixelberechnung kann genutzt werden.

Für die H-Feld-Berechnung wurde der Marr-Hildreth-Operator (MH-Operator), der wegen der Form des Operators auch 'Mexican Hat' genannt wird (z.B. Faugeras [27], S. 79ff.) als klassischer Vergleichsoperator gewählt. Ausgehend von einer 2-dimensionalen Gaußfunktion, die den Rauschanteil σ im Bild glättet, wird die zweite Ableitung gebildet und in einer Faltungsmatrix dargestellt. Mit steigendem σ wird der Verlauf der Gaußkurve flacher, dementsprechend auch die zweite Ableitung. Um den Gesamtfehler bei der Abbildung der kontinuierlichen Funktion in eine Faltungsmatrix gering zu halten, müßte die Matrixgröße dem Rauschanteil des Bildes angepaßt werden. Eigene Untersuchungen ergaben bei dem eingesetzten Kamera/Framegrabber-System eine Standardabweichung in der Signalamplitude in der Größenordnung 2 bis 2.5 Amplitudendiskretisierungsstufen. Diesem σ entspräche im kontinuierlichen Fall bei einem angenommenen zulässigen Fehler von 10^{-4} ein Radius von 6.2 Pixel. Dies würde mindestens eine Faltungsmatrix der Größe 14x14 erfordern, wobei Fehler durch die Diskretisierung des Ortes nicht berücksichtigt sind. Da aus Vergleichsgründen hier jedoch eine 5x5 Matrix erzeugt werden soll, wurde σ auf $\sqrt{2}$ (entspricht Radius 4.85) festgelegt, und überstehende Matrixelemente abgeschnitten. Die dadurch erhaltene Faltungsmatrix hat folgende Elemente:

-0.406	-0.430	-0.368	-0.430	-0.406
-0.430	0.000	0.389	0.000	-0.430
-0.368	0.389	1.000	0.389	-0.368
-0.430	0.000	0.389	0.000	-0.430
-0.406	-0.430	-0.368	-0.430	-0.406

Abb. 24: Koeffizienten der H-Feld Matrix

Abb. 25 zeigt die H-Feld-Matrix im Vergleich zur MH-Matrix, wobei auch hier eine Skalierung vorgenommen wurde, um vergleichen zu können. Gut zu erkennen ist die ähnliche Form der Matrizen. Beide enthalten strukturelle 0-Elemente. Bei der MH-Matrix sind diese durch die Wahl von σ entstanden.

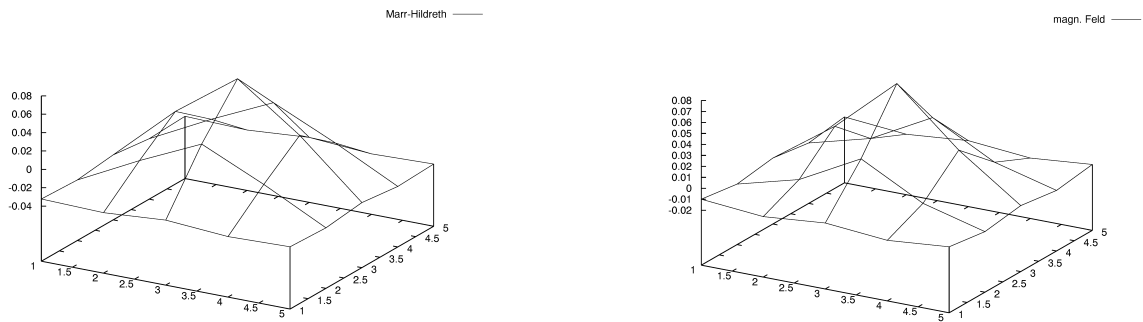


Abb. 25: Form der H-Feld Matrix und Marr-Hildreth-Matrix

Die Analyse der resultierenden Feldverläufe erfolgt anhand von Abb. 26: Aufgetragen sind

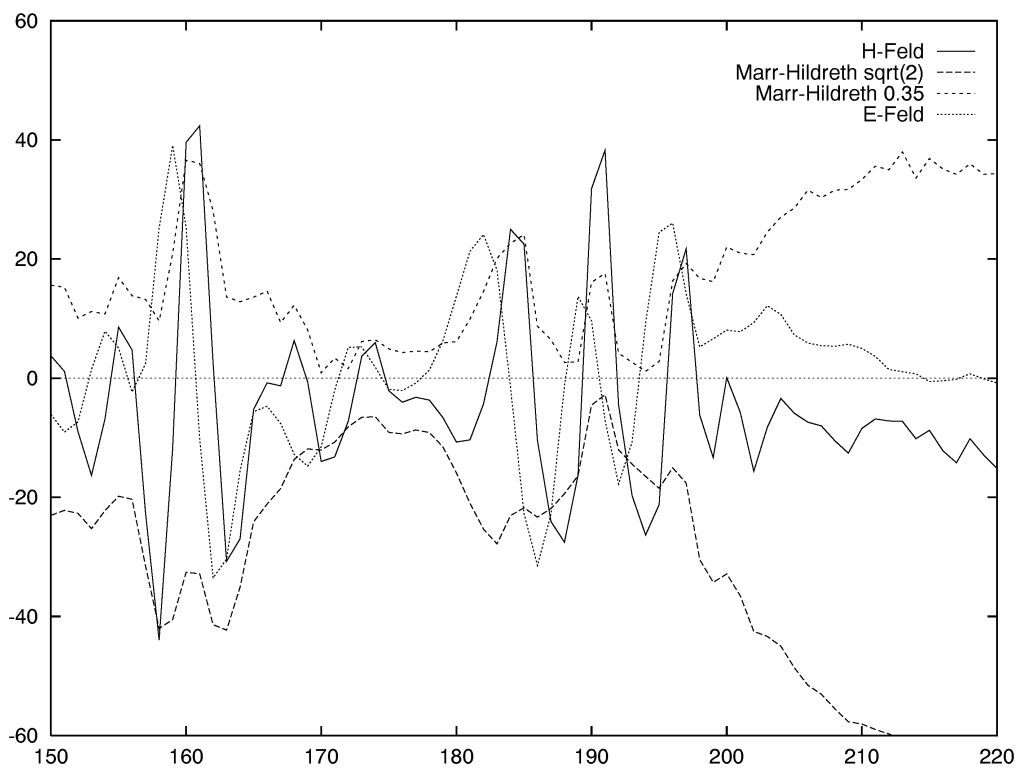


Abb. 26: H-Feldverlauf mit Marr-Hildreth-Matrizen

wieder ein vertikaler Schnitt des 2-dimensionalen H-Feld bei $x=70$ und im Vergleich hierzu der (skalierte) Feldverlauf, der mit dem MH-Operator und $\sigma = \sqrt{2}$ erzeugt wurde.

Es ergibt sich ein signifikanter Unterschied in den Feldverläufen: Der MH-Operator hat einen deutlichen (negativen) Gleichanteil, und die erforderliche differenzierende Wirkung ist kaum erkennbar. Der Gleichanteil kann durch Wahl von $\sigma=0.35$ eliminiert werden. Kriterium für dieses spezielle σ war, daß gerade keine Nicht-Null-Elemente abgeschnitten werden müssen. Der somit bestimmte Feldverlauf (Marr-Hildreth-035) ist für die Zwecke der Konturverfolgung ebenfalls nicht geeignet, weil auch hier die differenzierende Wirkung fehlt, die sich bei der Konturverfolgung in regelmäßigem Verlassen der Kontur auswirkt. Zum Vergleich mit Abb. 23 ist skaliert der Verlauf des E-Feldes mit aufgetragen.

Der aufgetretene Fehler beim MH-Operator wird zurückgeführt auf die starke Diskretisierung bei der Abbildung der kontinuierlichen zweiten Ableitung der Gaußfunktion auf die kleine 5x5-Matrix. Diese Erkenntnis deckt sich mit Aussagen in der Literatur [6] S. 24, [33], [31], [34], daß vernünftige Ergebnisse mit dem MH-Operator mit Matrixgrößen ab 10x10 aufwärts erzielt werden können. Diese Matrixgrößen sind jedoch ungeeignet, um die für differentielles Motion-Stereo erforderlichen Zeitschranken einhalten zu können.

3.1.7. Startpunktsuche und Nutzung von Vorwissen

Voraussetzung für die Anwendung des Prinzips der Konturverfolgung ist es, einen Punkt zu kennen, der auf einer Kontur liegt. Die lokalen Feldwerte, die durch Auswertung der lokalen Umgebung entstehen, bestimmen dann den jeweiligen Folgepunkt. Eine Auswertung der Umgebung von Bildpunkten, die nicht auf einer Kontur liegen, kann entfallen, so daß letztlich nur ein geringer Anteil an Bildpunkten analysiert werden muß. Dies ist der Grund für die hohe Effizienz der Konturverfolgung. Problematisch ist jedoch die Suche nach den Startpunkten für die Kontursuche, damit die bei der Konturverfolgung erzielten Einsparungen bei der Startpunktsuche vernichtet werden. Bei der nachfolgend beschriebenen Realisierung wurde dieses Problem gelöst.

Zur Bestimmung des ersten Konturpunktes wird das Grauwertbild horizontal und vertikal gitterartig in gewissen Abständen traversiert, wobei Nulldurchgänge des H-Feldes gesucht werden, an denen noch keine Kontur extrahiert wurde. Dabei muß, um rauschbedingte Falschalarme zu vermeiden, der Nulldurchgang eine gewisse Steilheit besitzen. Ist ein solcher Nulldurchgang gefunden (Suchfenster 2 in Abb. 27), wird dieser Punkt als Startpunkt einer Kontur interpretiert. Diese Kontur wird dann nach links verfolgt, bis entweder der Bildrand erreicht ist (Suchfenster 3 in Abb. 27), auf eine bereits extrahierte Kontur getroffen wird, oder der Gradient der Kontur unter einen Schwellwert sinkt. Anschließend erfolgt die Konturverfolgung ab dem Startpunkt nach rechts, um auch den anderen Ast zu extrahieren. Wurde bei der Extraktion des linken Teilastes der Startpunkt wieder erreicht, liegt eine zyklische Kontur vor, und die Extraktion des rechten Teilastes kann entfallen.

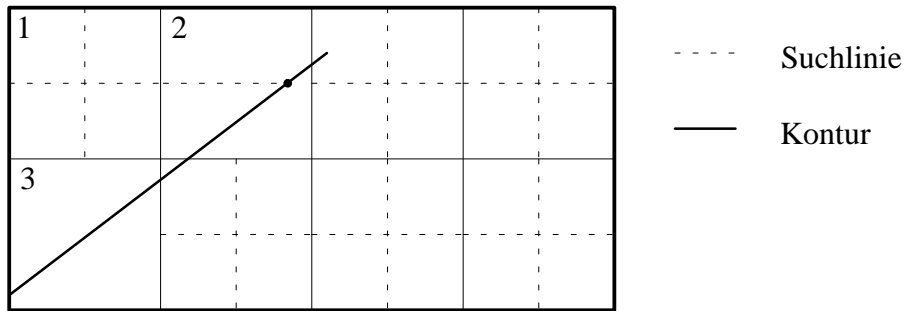


Abb. 27: Startpunktsuche

Verzweigungen werden nur in einer Richtung weiterverfolgt. Der andere Teilast wird von einem anderen Startpunkt aus extrahiert. Diese Konturverfolgung endet dann an der Verzweigungsstelle. Welcher Teilast an einer Verzweigung weiterverfolgt wird, hängt von den Krümmungswinkeln und den Kontrasten der vom Konturpunkt wegführenden Teiläste ab und ist nicht mit einfachen Mitteln vorhersagbar.

Aufgrund statistischer Untersuchungen der Merkmals- und Konturpunktverteilung bei Bildern von realen Szenen kann eine weitere Optimierung der Startpunktsuche vorgenommen werden: Wird innerhalb eines Suchfensters ein Startpunkt gefunden (Suchfenster 2 in Abb. 27), wird dieses Suchfenster als bearbeitet markiert, und die Startpunktsuche nach der Extraktion der gefundenen Kontur in einem anderen Fenster fortgesetzt. Darüberhinaus wird ein Suchfenster ebenfalls als bearbeitet markiert, wenn ein gewisser Anteil (Größenordnung wenige Prozent) der darin enthaltenen Pixel bei vorhergehenden Konturverfolgungen als Konturpunkte identifiziert wurden (Suchfenster 3 in Abb. 27). In diesem Fall ist es bei hinreichend kleiner Suchfenstergröße (z.B. 16×16 Pixel) unwahrscheinlich, eine weitere Kontur darin zu entdecken. Befindet sich darin tatsächlich noch eine zweite Kontur, die eine gewisse Mindestlänge hat, wird ein Startpunkt zwar nicht in diesem, jedoch in einem anderen Suchfenster erkannt, und der Konturfolger läuft diese Kontur ebenfalls ab, nur ausgehend von einem anderen Startpunkt.

Die Einsparung gegenüber der vollständiger Betrachtung der horizontalen und vertikalen Suchlinien beträgt typischerweise zwischen 20 und 40% je nach Suchfenstergröße und Mindestanteil an Konturpunkten. Diese Zahlen sind aber auch abhängig von der Strukturierung der Szene: Eine weiße Wand enthält keine Konturen, so daß die Optimierung bei der Startpunktsuche nicht greifen kann. Allerdings gibt es dann aber auch keine Konturverfolgung. Bei einer stark strukturierten Szene ist hingegen nach der Verfolgung einer einzigen Kontur bereits eine große Anzahl von Suchfenstern als bearbeitet markiert.

Steht Vorwissen über die erwartete Lage der Kanten zur Verfügung, kann dieses vorteilhaft zur Einschränkung der Startpunktsuche verwendet werden: Nicht mehr im gesamten Bild,

sondern zunächst nur innerhalb eines Toleranzbereiches um die erwarteten Kanten wird horizontal und vertikal nach Startpunkten gesucht. Dabei wird die Mindeststeilheit des Nulldurchgangs reduziert, um auch schwache, erwartete Kanten zu extrahieren, die z.B. von höheren Schichten hypothetisch erwartet werden, und von der Vorverarbeitung verifiziert werden sollen [13].

3.1.8. Eckpunkterkennung und Polygonapproximation

Während der Verfolgung einer Kontur wird an jedem Konturpunkt untersucht, ob dieser ein Eckpunkt ist. Wichtig dabei ist, daß trotz der verrauschten Konturpunktkette das resultierende Polygon keine große Abweichung von den Konturpunkten besitzt, daß aber auch nicht zu viele kleinen Geradenstücke entstehen.

Das entwickelte Verfahren ermöglicht es, einen Eckpunkt nur aufgrund des Konturstücks seit dem letzten Eckpunkt bzw. Startpunkt zu erkennen. Dabei wird ein gleitender Mittelwert über die Winkel der Kontursekanten zum letzten Eckpunkt bzw. Startpunkt gebildet (Abb. 28). Ist

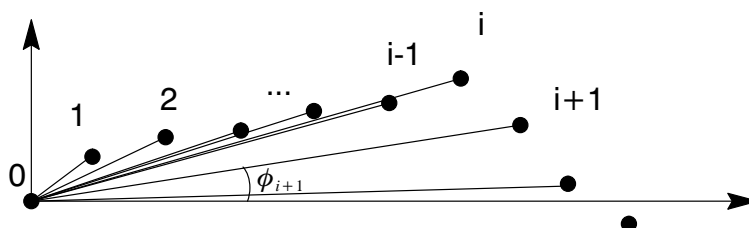


Abb. 28: Gleitende Mittelwertbildung zur Eckpunkterkennung

die Differenz zwischen aktuellem Sekantenwinkel und gleitendem Mittelwert größer als eine Schwelle, wird ein Eckpunkt erkannt. Diese Schwelle reduziert sich mit zunehmender Konturabschnittslänge i gemäß:

$$|\bar{\phi}_i - \phi_{i+1}| > \frac{\pi}{4i} \quad (11)$$

Hiermit wird berücksichtigt, daß der Einfluß eines "Ausreißer"-Konturpunktes auf den Mittelwert mit zunehmender Länge des Konturabschnittes sinkt. Außerdem wird durch die mit der Konturabschnittslänge sinkende Schwelle sichergestellt, daß bei schwach gekrümmten Konturen überhaupt ein Eckpunkt gefunden wird.

An jedem Eckpunkt wird aus den Konturpunkten eine Regressionsgerade berechnet [92], [84] S. 172, um den Fehler durch das Schwingen der Querregelung und durch Bildrauschen zu reduzieren. Die tatsächlichen Eckpunkte werden dann durch Schnitt der Regressionsgera-

den bestimmt. Aufgrund der extrahierten Geradenstücke wird in dem schon mehrfach angesprochenen Labelbild eine Spur mit der Kennung des extrahierten Geradenstücks angelegt.

Nach der Verfolgung des linken und rechten Konturastes wird nachfolgend geprüft, ob der Winkel zwischen benachbarten Geradenstücken kleiner als ein Schwellwert ist. An solchen falschen Eckpunkten, deren Ursache wiederum vor allem die Schwingungen der Querregelung ist, werden die angrenzenden Geradenstücke verschmolzen. Auch in diesem Zusammenhang ist die subpixelgenaue Berechnung des Folgepunktes von Vorteil, weil durch die reduzierten Schwingungen des Konturpunktes weniger "falsche" Eckpunkt auftreten, die anschließend wieder elimiert werden müssen.

Im Vergleich zum klassischen Split-Verfahren nach Duda/Hart [21] (siehe auch [6], Kap. 6), bei dem der komplette Konturzug benötigt wird, bevor rekursiv bis zum Erreichen eines Abbruchkriteriums immer neue Eckpunkte eingefügt werden, vermeidet der hier vorgestellte Algorithmus den erheblichen Aufwand, für jeden Konturpunkt in jeder Rekursionsstufe den Abstand zu einer Sekante zu berechnen. Der vorgestellte Polygonapproximationsalgorithmus ist der Klasse der sog. "Scan-Along-Algorithmen", wie z.B. [104], zuzuordnen, bei denen ausgehend von einem bekannten Eckpunkt ein neuer Eckpunkt gesetzt wird, wenn das Minimum der Abstände der Konturpunkte zur Sekante einen Schwellwert übersteigt. Letzteres wird durch den Schnitt von Kreissektoren erreicht[104], um den Abstand eines Konturpunktes zur jeweils aktuellen Sekante nicht immer wieder zu berechnen. Ein sehr schnelles, sequentielles Verfahren ist in [102] beschrieben, bei dem ein Eckpunkt gesetzt wird, wenn der Betrag der Differenzfläche zwischen Kontur und Sekante eine gewisse Schwelle überschreitet. Aufwand und Qualität sind vergleichbar mit dem vorgestellten Verfahren. In [22] wird ein anderer Scan-Along-Algorithmus beschrieben, der mit Hilfe der dynamischen Programmierung optimale Approximationen im Sinne minimaler Anzahl von Sekanten bei gegebenem maximalen Approximationsfehler erreicht. Das Ergebnis der suboptimalen Approximationen ist qualitativ niedriger zu bewerten, als das der optimalen [81][22], dafür ist der Aufwand mit $O(n)$ bei dem hier implementierten Verfahren oder [102] der Echtzeitanforderung besser angepaßt als $O(n^2)$ in [81].

3.1.9. Bildentzerrung

Für quantitative Anwendungen des Rechnersehens müssen die a-priori bekannten Sensorfehler korrigiert werden. Diese Fehler entstehen durch radiale Verzerrungen der Optik und Toleranzen in der Herstellung und Montage des CCD-Chips. Typische Kamera-Framegrabber-Systeme haben dabei Abbildungsfehler in der Größenordnung von einigen Pixeln.

Nach Tsai-Lenz [99] können die internen und externen Kameraparameter durch eine Offline-Kalibrierung hinreichend genau bestimmt werden, so daß die Korrektur der Bildkoordinaten

zur Laufzeit möglich wird. Hier werden nur die internen Kameraparameter Hauptpunktverschiebung C_x und C_y , Skalierung P_x , und P_y , Verzeichnungskoeffizient κ und Bildweite f zur Entzerrung verwendet. Diese Parameter wurden mit dem in [53] veröffentlichten Verfahren bestimmt.

$$x_v = (x_{pixel} - C_x) \cdot P_x \quad y_v = (y_{pixel} - C_y) \cdot P_y \quad (12)$$

$$x^b = \frac{x_v}{f \cdot (1 + \kappa \cdot (x_v^2 + y_v^2))} \quad y^b = \frac{y_v}{f \cdot (1 + \kappa \cdot (x_v^2 + y_v^2))} \quad (13)$$

Normalerweise müßte jeder Konturpunkt gemäß Gleichung (12) und (13) entzerrt und normiert werden, bevor die Geradenapproximation vorgenommen wird. Der Aufwand ist dabei mit 12 Fließkomma-Operationen pro Konturpunkt [85] erheblich. Stattdessen wird die Tatsache genutzt, daß die zu extrahierenden Kanten in 3D Geraden sind, und gekrümmte Kanten auch bei anderen Extraktionsverfahren durch Geradenstücke approximiert werden. Daher wird zunächst ein Konturabschnitt zwischen zwei Eckpunkten in den verzerrten Koordinaten zu einem Geradenstück zusammengefaßt, wie in Absatz 3.1.8. beschrieben. Die Endpunkte des Geradenstücks werden anschließend entzerrt. Schließlich werden alle benachbarten Geradenstücke noch einmal im entzerrten Bildkoordinatensystem überprüft, ob der Schnittwinkel kleiner einem Schwellwert ist, und eine weitere Verschmelzung vorgenommen werden kann.

Der Gefahr, daß im Verlauf einer Kontur der Maximalabstand zwischen approximierender Geraden und Konturpunkten zu groß wird, kann durch geeignete Wahl der Winkelschwellwerte für die beiden Verschmelzungsoperationen begegnet werden. Schwach gekrümmte Konturen, die durch die Linsenverzerrungen im verzerrten Bild als gerade erscheinen, werden allerdings mit diesem Verfahren nicht in separate Geradenstücke aufgetrennt. In diesem in der Praxis noch nicht beobachteten Fall stimmt dann die extrahierte Kante nicht mit der 3D-Szene überein.

Bei der Entzerrung werden die Kantenkoordinaten ohne Mehraufwand auf eine Kammerkonstante von $f=1$ normiert. Das vereinfacht die nachfolgende Triangulation insofern, als die Behandlung der internen Kameraparameter in der Vorverarbeitung konzentriert ist und weiterverarbeitende Instanzen davon unabhängig sind.

3.1.10. Zusatzattribute

Bei der Diskussion der Korrespondenzfindung in Abschnitt 3.2. wird gezeigt, daß die Robustheit dieses Prozesses verbessert wird, je mehr unkorrelierte Merkmalskomponenten von der

Vorverarbeitung geliefert werden. Als zusätzliche Merkmalskoordinaten werden deshalb die über den Kantenverlauf gemittelten lokalen Feldgrößen verwendet. Die Ergebnisse dieses Ansatzes finden sich in Absatz 5.2.3..

Desweiteren werden die bei der Extraktion gefundenen Nachbarschaftsbeziehungen explizit modelliert: sobald der Konturfolger auf eine bereits extrahierte Kontur trifft, steht mit Hilfe eines Labelbildes eine Kennung des aus der Kontur entstandenen Geradenstücks bereit. Verzweigungspunkte können dann auf einfache Weise klassifiziert werden, z.B. als T, λ , Y, V, X-Verbindung ([76] S. 146) Diese 2D-Beziehungen werden über die Triangulation in 3D-Beziehungen transformiert, und vereinfachen die räumlichen Interpretation.

Bei benachbarten Geradenstücken sind darüber hinaus die Endpunkte mit den Kennungen des jeweils angrenzenden Geradenstücks attribuiert. Diese Information steht ebenfalls für die Korrespondenzfindung zur Verfügung.

3.1.11. Echtzeitaspekte

Die Art der Startpunktsuche innerhalb von Suchfenstern ermöglicht eine Einflußnahme auf die Reihenfolge der Kantenextraktion, wenn die Abarbeitungsreihenfolge über eine Umsetztabelle festgelegt wird. Die Belegung dieser Tabelle kann mit Hilfe von Vorwissen über die Lage, die Signifikanz und die Prägnanz einer Kante erfolgen, wodurch die Kanten in ihrer Abarbeitung priorisiert werden. Bleiben anschließend noch Suchfenster übrig, kann die Abarbeitung beliebig vorgenommen werden.

Auch die Beschränkung der Merkmalsextraktion auf einen oder mehrere Bildbereiche kann über die Look-Up-Tabelle vorgenommen werden, indem nur Suchfenster in den interessierenden Gebieten eingetragen werden. Zwar werden zusammenhängende Konturen über die Grenzen dieser Interessensgebiete weiterverfolgt, die Startpunktsuche bleibt jedoch begrenzt, und damit im Normalfall auch der Gesamtaufwand.

Unter der Annahme, daß die Zeit zur Extraktion einer einzelnen Kontur klein ist gegenüber der vorgegebenen Zykluszeit, wird zwischen der Abarbeitung zweier Suchfensters die verfügbare Restzeit bis zum Erreichen der Zeitschranke überprüft, und gegebenenfalls ein Abbruch mit suboptimalem Extraktionsergebnissen vorgenommen. Prinzipiell könnte diese Überprüfung auch an jedem Konturpunkt oder an jedem gefundenen Eckpunkt durchgeführt werden, allerdings steigt der Aufwand, und die erreichte Zeitgranularität reicht ohnehin schon aus.

3.2. Korrespondenzfindung

3.2.1. Prinzip

Ein zentrales Problem bei allen Stereoverfahren ist die Korrespondenzfindung. Die Aufgabe besteht darin, in zwei Sensorbildern zusammengehörende Merkmale zu identifizieren. Die erkannten Merkmalsdiskrepanzen sind dann Ausgangspunkt für die weitere Interpretation und den eigentlichen Stereoalgorithmus, der die Tiefeninformation rekonstruiert.

Der Vorgang der Korrespondenzfindung wird gelegentlich auch unter dem normalerweise weiterfassenden Begriff der Sensordatenfusion gesehen, bei dem die Wahrnehmungen unterschiedlicher Sensoren in Beziehung gebracht werden. Dabei können sich die zu vergleichenden Aufnahmen in folgenden Aspekten unterscheiden, wobei Kombinationen möglich sind [71]:

- Aufnahmezeitpunkt

Die Änderungen in den Einzelaufnahmen resultieren zum einen aus dem unterschiedlichen Aufnahmezeitpunkt und zum anderen aus der Bewegung in der Szene oder aus Änderungen in der Beleuchtung. Bei nicht-stationärer Kamera bewirkt die Eigenbewegung einen unterschiedlichen Aufnahmeort, wodurch sich das Bild der Szene durch die unterschiedlichen Blickwinkel ändert.

- Physikalisches Prinzip

Die Aufnahmen zeigen zwar die gleiche Szene, und sind zum gleichen Zeitpunkt und am gleichen Ort aufgenommen, werten allerdings ein unterschiedliches Spektrum aus (Farbfilter, Infrarot, usw.)

- Sensorparameter

Die Aufnahmen unterscheiden sich durch andere Sensorparameter, z.B. die Brennweite der Optik (Nahkamera und Fernkamera bei bifokalem Stereo[20], Zoom-Objektive [54])

- Aufnahmeort und Blickrichtung

Zwei ansonsten identische Sensoren nehmen die Szene von unterschiedlichen Positionen auf (binokulares Stereo). Der Unterschied zu Korrespondenzen aufgrund unterschiedlicher Aufnahmezeit ist die fixe relative Lage der Kameras.

- aktuelles und akkumuliertes Sensorbild

Dabei handelt es sich um den Vergleich von aktueller Sensorinformation und Modellinformation, die als akkumulierte, zeitlich stabilisierte Sensorbilder angesehen werden können.

Merkmale in unterschiedlichen Bildern, die zusammen gehören, sind zwar nicht gleich, müssen aber, um als zusammengehörig erkennbar zu sein, eine gewisse Ähnlichkeit aufweisen. Diese Ähnlichkeit ist für den menschlichen Betrachter eines Bildpaares unmittelbar sichtbar, nicht jedoch für den Rechner, für den die beiden Bilder zunächst nur zwei Matrizen mit Werten von 0 bis 255 sind. Die Herstellung von Korrespondenzen durch menschliche Beobachter ist ein intuitiver und iterativer Vorgang, bei dem die Extraktion und Erkennung von charakteristischen Szenenelementen und Teilszenen (Segmentierung), die Aufstellung und Verifikation von Hypothesen, sowie die semantische Bildinterpretation unter Nutzung von großen Mengen Hintergrundwissen eine Rolle spielen [43]. Dieses Vorgehen ist aufgrund der Komplexität und des fehlenden Formalismus derzeit nicht auf einfache Weise auf Maschinen portierbar. Allerdings werden Anstrengungen gemacht, mit Hilfe von semantischen Netzen [79], relationalen Objektstrukturen [82], und anderen Methoden aus dem Bereich der künstlichen Intelligenz diese Vorgänge nachzubilden.

3.2.2. Merkmalsraum

Für eine maschinengerechtere Lösung des Korrespondenzproblems können Methoden der Mustererkennung [78] verwendet werden, die für Zuordnungsaufgaben von zu analysierenden Proben auf a-priori bekannten Sätzen von Mustern entwickelt wurden. Die aus der Probe extrahierten Primärmerkmale spannen einen mehrdimensionalen Merkmalsraum auf. Jede Probe repräsentiert ein Punkt in diesem Merkmalsraum, und jedes Muster belegt ein bestimmtes Volumen. Die Zuordnung von Probe zu einem Muster erfolgt über eine Metrik, die die Distanzen zwischen zwei Punkten im Merkmalsraum mißt. In [77] finden sich Beispiele für die Anwendung dieses Ansatzes in der visuellen Mustererkennung.

Die komplexe und schwer zu quantifizierende Problematik der Unterscheidbarkeit und Ähnlichkeit von Merkmalen wird mit Hilfe der Abstandsbildung auf den Vergleich von Skalaren abgebildet. Merkmalspaare mit der größten Ähnlichkeit haben dann die minimale Distanz.

Um dieses Verfahren auf die Korrespondenzfindung bei Bildfolgen anwenden zu können, sind nachfolgend mehrere Fragen zu klären:

- Was entspricht den *Mustern* bei der Korrespondenzfindung in Bildfolgen?
- Welche *Merkmalskoordinaten* spannen den Merkmalsraum auf?
- Wie lautet das *Distanzmaß*?

Das *Muster* entspricht dem ersten Vergleichspartner aus Bild n bei der vollständigen Analyse aller Merkmalskombinationen zwischen Bild n und Bild $n + 1$. Nachdem es in diesem Fall

kein Volumen im Musterraum gibt, entfällt die übliche Schwerpunktsbestimmung für den Referenzpunkt eines Musters.

Denkbar wäre die Akkumulation der bisher aufgelaufenen korrespondierenden Merkmale zu einem Cluster im Sinne der Mustererkennung, das z.B. durch Schwerpunkt und Kovarianz charakterisiert ist. Es stellt sich jedoch das Problem der zeitlichen Konstanz der Merkmalskoordinaten. Diese Eigenschaft besitzen die Merkmale einer zeitvarianten Bildfolge in der Regel nicht, denn gerade die zeitlichen Änderungen im Merkmalsraum sind der Ausgangspunkt für die weitere Interpretation. Allerdings kann die Ableitung einer Merkmalskoordinate konstant sein. Wenn man sich nun auf einen Unterraum des Merkmalsraumes beschränkt, der durch die zeitinvarianten Merkmalskoordinaten aufgespannt wird, lassen sich Cluster bilden, und damit auch schrittweise statistische Aussagen zur Merkmalsverteilung aufbauen. Die entstehenden Kovarianzen lassen sich dann verwenden, um über die Eigenvektor-Berechnung und eine Merkmalstransformation zu einer optimalen Beschreibung der separierenden Hyperebenen zu gelangen, wie es in der Clusteranalyse [78][87] üblich ist.

Die Koordinaten des *Merkmalsraumes* bestehen zunächst aus der geometrischen Position der extrahierten Kante, die durch Anfangs- und Endpunkt repräsentiert ist. Für punktförmige Merkmale ist der 2-dimensionale Merkmalsraum identisch mit der 2D-Bildebene, für Kantenmerkmale ist der Merkmalsraum 4-dimensional.

Eine alternative Beschreibungsform für 2D-Kanten setzt sich aus Winkel, Länge und Mittelpunkt der Kante zusammen [19],[18]. (vgl. Abb. 29) Diese läßt sich aus Anfangs- und End-

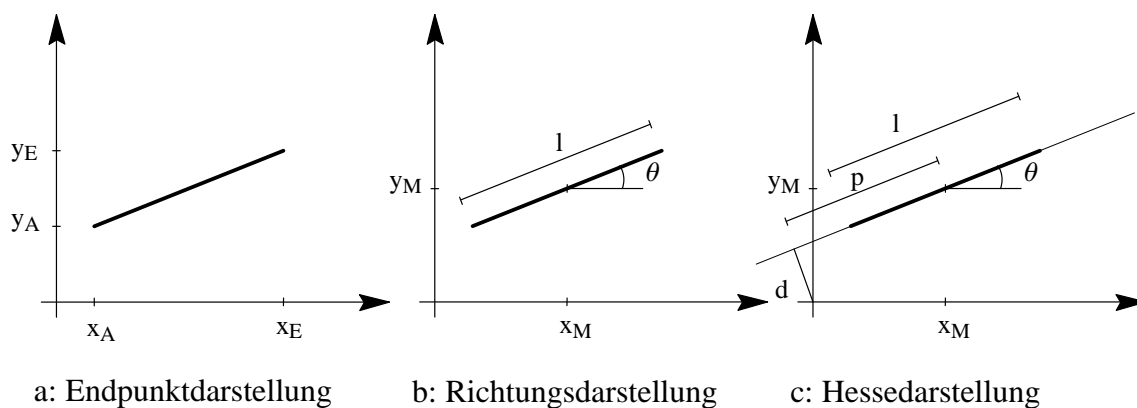


Abb. 29: Alternativen der Kantendarstellung

punktskoordinaten gemäß folgender Formeln berechnen[19]:

$$\theta = \text{atan} \frac{y_E - y_A}{x_E - x_A} \quad (14)$$

$$l = \pm \sqrt{(x_E - x_A)^2 + (y_E - y_A)^2} \quad (15)$$

Diese alternative Form kann besser die bei der Kantenextraktion auftretenden Unsicherheiten in den Merkmalskoordinaten berücksichtigen. Beispielsweise bewirkt das Pixelrauschen ein gelegentliches Verkürzen der Kante, das sich in einer hohen Unsicherheit der Kantenlänge und der Lage der Kante in Kantenrichtung auswirkt. Dagegen ist die Lage der Kante senkrecht zur Kantenrichtung sehr exakt zu bestimmen, meist mit Genauigkeiten unter einem Pixel. Die Richtung der Kante hat dann eine Genauigkeit, die ungefähr dem Verhältnis von senkrechter Genauigkeit und Länge entspricht. Bei der Transformation der geometrischen Merkmalskoordinaten entsteht keine neue Information, stattdessen können die Unsicherheiten der einzelnen Merkmalskoordinaten aufgrund des Wissens über die Merkmalsextraktion exakter modelliert werden. Dies ist auch das Ziel der oben angesprochenen Transformation bei der Clusteranalyse.

Ein weiterer Grund für die Transformation der geometrischen Merkmalskoordinaten ist das Aperturproblem (Abb. 30): Beobachtet man durch eine Öffnung, d.h. ohne Kenntnis der End-

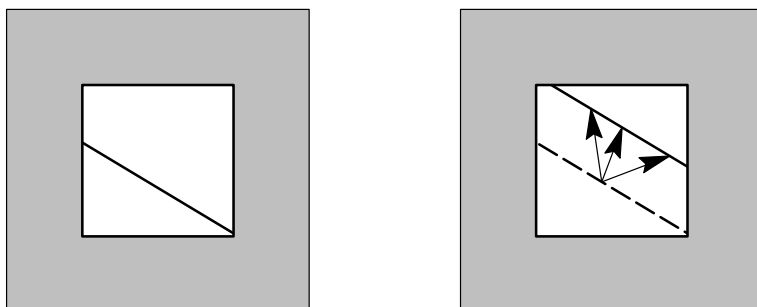


Abb. 30: Aperturproblem

punkte, die Bewegung einer Kante in der Bildebene, so läßt sich aus zwei zu unterschiedlichen Zeitpunkten aufgenommen Kanten nur eine Aussage über die senkrecht zur Kante liegende Verschiebungskomponente machen, nicht jedoch über die Komponente in Richtung der Kante. Nach der Transformation der Merkmalskoordinaten in die Hessedarstellung nach Abb. 29-c läßt sich durch Gewichten der Tangentialkomponente die Tatsache der fehlenden oder unsicheren Information berücksichtigen. Die Transformationsvorschrift aus der Endpunktdarstellung nach Abb. 29-a lautet[19]:

$$d = \frac{x_E * y_A - x_A * y_E}{l} \quad (16)$$

$$p = \frac{(x_E - x_A) * (x_A + x_E) + (y_E - y_A) * (y_A + y_E)}{2 * l} \quad (17)$$

Diese Kantendarstellung ist so noch nicht eindeutig. In der Endpunktdarstellung entstehen durch Vertauschen von Start- und Endpunkt zwei unterschiedliche Punkte im Merkmalsraum. Eine Einschränkung, z.B. $x_A < x_E$, stellt diese Eindeutigkeit wieder her. Diese Einschrän-

kung entspricht in den anderen beiden Darstellungen $l > 0$ und $-\frac{\pi}{2} \leq \theta < \frac{\pi}{2}$. Problematisch für die Korrespondenzfindung sind die dabei entstehenden Unstetigkeiten bei $\pm \frac{\pi}{2}$. Hier muß durch die zyklische Interpretation der θ -Dimension bei der Distanzbildung der Äquivalenz von $-\frac{\pi}{2}$ und $+\frac{\pi}{2}$ Rechnung getragen werden.

Muß man die Distanzen im Merkmalsraum weiter erhöhen, um Mehrdeutigkeiten auflösen zu können, benötigt man neue, unkorrelierte Merkmalsattribute. Wenn man den Kantenextraktionsprozeß als eine Form der Informationsverdichtung ansieht, heißt das, daß zusätzliche Eigenschaften aus dem Rohbild in die Merkmalsrepräsentation hinübergerettet werden müssen. Diese Unterscheidungsattribute können nach Levi in [82] z.B. die Art des Kantenprofils (Stufe, Rampe, Dach, Stiftkante, und deren Linearkombinationen), sowie die Stärke (Gradientenbetrag) und die Richtung (Gradientenwinkel) sein. McIntosh/Mutch [70] verwenden für die Zuordnung Attribute, die bei der dort verwendeten Art der Kantenextraktion [11] zusätzlich zu den geometrischen Attributen abfallen: "Breite" der Kante, Kontrast, Steilheit, Mittelwert der 10% hellsten Pixel, die zur Kante beitragen, und Mittelwert der 10% dunkelsten Pixel. Diese Attribute sind zum Teil korreliert, wie sich aus der Berechnungsvorschrift ergibt, ohne die gute Wirkung auch bei der Zuordnung von stärker veränderten Kanten zu beeinträchtigen [70].

Wie in Absatz 3.1.10. beschrieben, wird bei dem hier implementierten Ansatz ebenfalls versucht, Kanteneigenschaften zur Erhöhung der Dimensionalität des Merkmalsraums zu verwenden, nämlich die über den Kantenverlauf gemittelten Felder E_v , E_h , und H , die bei der Kantenextraktion an jedem Konturpunkt sowieso berechnet werden müssen. Die Art des Kantenprofils wird in der Kantenextraktion nicht analysiert, und steht als Diskriminator nicht zur Verfügung.

Die letzte Frage zur Anwendung der Methoden der Mustererkennung bei der Korrespondenzfindung betraf die Auswahl eines *Distanzmaßes*, das die besonderen Eigenschaften einer Metrik besitzen muß:

Die Funktion $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ ist eine Metrik im Raum \mathcal{M} , wenn gilt:

$$\forall \vec{x}, \vec{y}, \vec{z} \in \mathcal{M}$$

$$d(\vec{x}, \vec{y}) = 0 \Leftrightarrow \vec{x} = \vec{y} \tag{18}$$

$$d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x}) \quad (\text{Symmetriebedingung}) \tag{19}$$

$$d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}) \quad (\text{Dreiecksungleichung}) \tag{20}$$

Die einfachsten Metriken sind die City-Block-Distanz (Manhattan-Distanz) bzw. der euklidische Abstand, bei denen die Eigenschaften der einzelnen Merkmalsdimensionen nicht berücksichtigt werden:

$$d_{City}(\vec{x}, \vec{y}) = \sum_i |y_i - x_i| \quad i = 1, \dots, < \text{Anzahl der Dimensionen} > \quad (21)$$

$$d_{Euklid}(\vec{x}, \vec{y}) = \sqrt{\sum_i (x_i - y_i)^2} = \sqrt{(\vec{y} - \vec{x})^T (\vec{y} - \vec{x})} \quad (22)$$

Diese Distanzmaße sind einfach zu implementieren, gewichten jedoch nicht die Dimensionen entsprechend der Skalierung und der Merkmalsdichte der einzelnen Koordinaten. In [70] geschieht die Gewichtung mit Hilfe von speziellen, auf die Anwendung angepaßte Heuristiken, die auf Erfahrungen basieren.

Die formale Berücksichtigung der statistischen Eigenschaften der bisher gefundenen korrespondierenden Merkmale einer Bildfolge führt zur Mahalanobis-Distanz:

$$d_{Mahalanobis}(\vec{x}, \vec{y}) = \sqrt{(\vec{y} - \vec{x})^T C^{-1} (\vec{y} - \vec{x})} \quad (23)$$

C ist die Kovarianzmatrix, deren Inverse als Gewichtsmatrix verwendet wird. Sie entsteht nebenbei ohne zusätzlichen Aufwand, wenn zum Verfolgen eines Merkmals auf der Bildebene ein Minimum-Varianz-Schätzer Anwendung findet, wie z.B. in [15], [19]. Die Linien konstanter Mahalanobisdistanz um einen Punkt im Merkmalsraum sind, bei angenommener Gaußscher Verteilung der bisher akkumulierten Merkmalskoordinaten, konzentrische, mehrdimensionale Ellipsoide, deren Hauptachsen durch die Eigenvektoren der Kovarianzmatrix bestimmt sind. Wie aus der Skizze der Äquidistanzlinien der drei diskutierten Distanzmaße am Beispiel des 2-dimensionalen Merkmalsraums (Abb. 31) erkennbar ist, ist bei der Mahalanobis-Distanz auch die oben angesprochene Merkmalstransformation implizit im Distanzmaß enthalten.

Für die Beurteilung der Qualität der verschiedenen Distanzmaße muß ein Qualitätsmaß eingeführt werden. Hierzu werden die folgenden Maßzahlen definiert:

Unter der *Intra-Merkmalsdistanz* (oder Auto-Merkmalsdistanz)⁶ sei der kleinste Abstand im Merkmalsraum über alle möglichen Merkmalspaare eines Merkmalsbildes \mathcal{M}_i zu verstehen:

$$d_{auto} = \min_{\vec{x}, \vec{y}} d(\vec{x}, \vec{y}) \quad \forall \vec{x}, \vec{y} \in \mathcal{M}_i, \vec{x} \neq \vec{y} \quad (24)$$

Die Bewertung der Distanzmaße und die verschiedenen Konfigurationen von Merkmalsräumen erfolgt z.B. exemplarisch anhand der Evaluation von Testreihen. Hierbei soll die Intra-

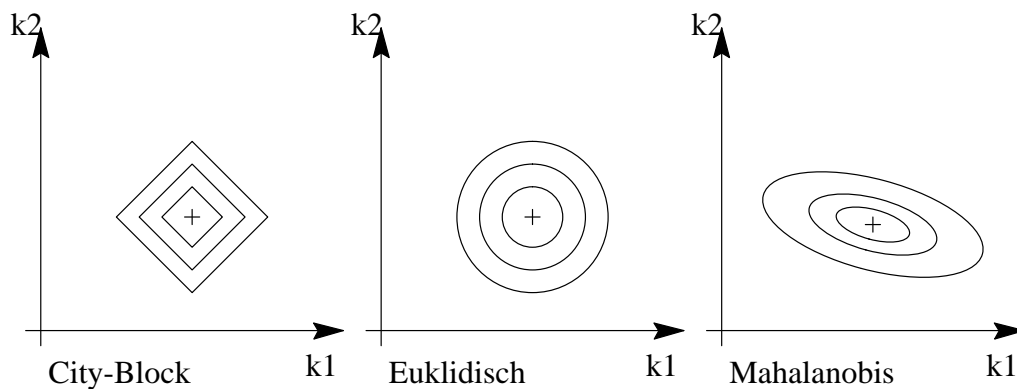


Abb. 31: Äquidistanzlinien im 2-dimensionalen Merkmalsraum

Merkmalsdistanz möglichst groß sein, d.h. die Merkmalspositionen im Merkmalsraum sollen möglichst weit auseinander sein.

In der Praxis zeigt sich beim Vergleich der Konfiguration der Merkmalsräume die Empfindlichkeit der Intra-Merkmalsdistanz gegenüber Skalierung und Gewichtung der einzelnen Merkmalskoordinaten. Die Normierung einer Merkmalskoordinate auf das mögliche Maximum oder auf den Mittelwert helfen nicht weiter.

Die *Korrespondenz-Distanz* ist definiert durch den größten Abstand zwischen korrespondierenden Merkmalspaaren aufeinanderfolgender Bilder einer Bildfolge. Dabei bezeichnet \circ die Relation korrespondierender Merkmale, und $\bar{\circ}$ die Relation nicht-korrespondierender Merkmale.

$$d_{korr} = \max_{\vec{x}, \vec{y}} d(\vec{x}, \vec{y}) \quad \forall \vec{x} \in \mathcal{M}_i, \vec{y} \in \mathcal{M}_j, \vec{x} \circ \vec{y} \quad (25)$$

Die *Inter-Merkmalsdistanz* (Kreuz-Distanz) ist definiert durch den kleinsten Abstand zwischen allen Merkmalspaaren zweier aufeinanderfolgender Bilder einer Bildfolge, wobei korrespondierende Merkmale an der Minimum-Bildung nicht beteiligt sind.⁷

$$d_{kreuz} = \min_{\vec{x}, \vec{y}} d(\vec{x}, \vec{y}) \quad \forall \vec{x} \in \mathcal{M}_i, \vec{y} \in \mathcal{M}_j, \vec{x} \bar{\circ} \vec{y} \quad (26)$$

Betrachtet man den Wertebereich der Korrespondenz-Distanzen und der Kreuz-Distanzen, kann in Anlehnung an die Halbleiterphysik der *Bandabstand* als die Differenz zwischen der Inter-Merkmalsdistanz und der Korrespondenz-Distanz definiert werden.

$$d_{Band} = d_{Kreuz} - d_{Korr} \gg 0 \quad (27)$$

6. Der Name ist in Anlehnung an die Auto-Korrelationsfunktion gewählt.
7. Die separate Behandlung korrespondierender Merkmalspaare erfordert eine aufwendige händische Korrespondenzfindung bei der Auswertung von Testreihen.

Eine erste Bedingung für die Wahl des Distanzmaßes und die Konfiguration des Merkmalsraumes ist, daß der Bandabstand für alle aufeinander folgenden Bildpaare größer 0 ist, d.h. daß sich der Bereich der Korrespondenz-Distanzen und der Inter-Distanzen nicht überlappen. Außerdem gilt, daß die automatische Korrespondenzfindung umso sicherer und robuster wird, je größer der Bandabstand ist. In 5.2.1. sind für einige Merkmalsraum-Konfigurationen und Distanzmaße entsprechende Histogramme dargestellt.

Die genannte Bedingung für den Bandabstand läßt sich unter Umständen lockern, wenn von einer globalen Betrachtungsweise eines Bildpaares auf eine lokale Betrachtungsweise eines einzelnen Merkmals übergegangen wird. Für jedes Merkmal aus dem ersten Bild kann dann eine lokale Kreuz-Distanz, eine Korrespondenz-Distanz und ein lokaler Bandabstand definiert werden, indem nur dieses Merkmal mit den Merkmalen des zweiten Bildes in Beziehung gesetzt wird. Damit wird eine ungleichmäßige Verteilung im Merkmalsraum berücksichtigt.

In 5.2.1. finden sich auch einige Diagramme aus der Auswertung einiger Bildfolgen gemäß der zweiten, gelockerten Bedingung.

3.2.3. Prädiktion

Bisher wurde bei der Korrespondenzfindung von einer Evaluierung aller möglichen Merkmalskombinationen aus zwei Bildern ausgegangen. Das ist bei einer größeren Merkmalsanzahl aufgrund des explodierenden Aufwands nicht praktikabel. Um die Anzahl der zu überprüfenden Paare zu reduzieren, wird die erwartete Position eines Merkmals im Merkmalsraum prädiziert. Da das korrespondierende Merkmal in der lokalen Umgebung zu finden sein wird, muß die Distanzberechnung zu den möglichen Korrespondenzpartnern nur noch für die Merkmale erfolgen, die sich in der Umgebung des prädizierten Punkts befinden. Der Aufwand zur Bestimmung des besten Kandidaten steigt nunmehr nur noch linear an, wenn man eine gleichmäßige Verteilung der Merkmale im Merkmalsraum annimmt.

Es gibt verschiedene Methoden zur Prädiktion eines Merkmals, die beim Bewegungstereo im Grunde alle auf der Annahme einer glatten Bewegung (*Smoothness Constraint*) basieren. Das bedeutet, daß die Relativbewegung zwischen Beobachter und Szene kontinuierlich ist und sich nicht abrupt ändert. Dabei kann bei bewegtem Beobachter diese Glattheitseigenschaft auf die Eigenbewegung in der Welt bezogen sein, aber auch auf die resultierende Bewegung der Merkmale auf der Projektionsebene. Beide Sichtweisen hängen über die nichtlineare, stetige Projektionsvorschrift zusammen. Aufgrund der Stetigkeit transformiert sich die Glattheitseigenschaft zwischen Weltkoordinatensystem und Bild- (oder Merkmals-) -koordinatensystem. Daher wird im Folgenden bei der Glattheit nicht mehr zwischen Welt- und Bildkoordinatensystem unterschieden.

Als Prädiktionsmethoden werden hier exemplarisch die *Epipolarometrie* ([27], Kap 6.2.) die Prädiktion der Merkmale aufgrund der geschätzten Eigenposition und einem *Weltmodell* ([89]), sowie die direkte *Schätzung im Merkmalsraum*, die letztendlich verwendet wird, diskutiert.

Bei der *Epipolarschätzung* wird die Tatsache genutzt, daß korrespondierende Bildpunkte in einem Bildpaar auf bestimmten Linien, den Epipolarlinien, liegen müssen, die durch Schnitt der Epipolarebene und den beiden Bildebenen entstehen. Die Epipolarebene wird durch die Projektionszentren der beiden Einzelbilder und dem 3D-Punkt der Szene bestimmt. Hat man Anfangs- und Endpunkt einer Kante aus einem Bild, sowie die Transformation der Eigenposition, so lassen sich die Epipolarlinien von Anfangs- und Endpunkt der korrespondierenden Kante im anderen Bild angeben. Die Epipolareinschränkung wird aufgrund der festen Transformation der Aufnahmepositionen oft bei binokularem Stereo verwendet. In diesem Fall läßt sich von der Position eines Bildpunktes in einem Bild durch eine feste Vorschrift die Epipolarlinie im anderen Bild bestimmen. Beim Bewegungstereo ist die translatorische und rotatorische Positionsdifferenz zwischen den Aufnahmen nicht fest. Sie kann modellbasiert mit einem Schätzer prädiziert werden (siehe 4.2.), der die angesprochene Glattheit der Bewegung für die Prädiktion nutzt.

Auch bei der Prädiktion der Merkmalskoordinaten mit einem *Weltmodell* wird die Weltposition des Beobachters benötigt. Aus Modell und Position wird dann ein erwartetes Merkmalsbild errechnet [88], wobei die Verdeckungen und die Projektionseigenschaften des Sensorsystems einfließen. Dieses Vorgehen zur Prädiktion eignet sich vor allem für den Bild-Modell-Vergleich. Nachteilig wirkt sich aus, daß etablierte Korrespondenzen aus vorhergehenden Vergleichen nicht verwendet werden. Ebenso versagt diese Methode der Korrespondenzfindung, wenn es noch kein Modell der Szene gibt, sondern dieses durch Exploration gerade erstellt werden soll.

Die effizienteste Prädiktionsmethode für das Bewegungstereo ist die Prädiktion im *Merkmalsraum* aufgrund des zeitlichen Kontexts. Hierbei wird der Zustand eines Merkmals, bestehend aus den Koordinaten im Merkmalsraum, sowie deren Ableitungen bestimmt. Die Punkte korrespondierender Merkmale bilden eine Trajektorie im Merkmalsraum, die man zur Prädiktion der Merkmalskoordinaten extrapolieren kann. Die Glattheit der Bewegung bewirkt eine glatte Trajektorie. Selbst bei einer glatten Bewegung muß allerdings die zeitliche Dichte der Abtastung ausreichend hoch sein, um die Bewegung korrekt verfolgen zu können. Diese Forderung entspricht dem Nyquist-Kriterium der diskreten Signalverarbeitung. Ist diese Forderung erfüllt, ist die Distanz aufeinander folgender korrespondierender Punkte im Merkmalsraum klein. Die Annahme der differentiellen Bewegung erlaubt also, im Folgebild das Merkmal im Merkmalsraum genau dort zu suchen, wo es im vorhergehenden Bild war.

Die genaue Einhaltung der äquidistanten Abtastzeit ist dann für die Prädiktion nicht erforderlich, weil die Prädiktion keine kinematische Berechnung unter Nutzung der Zykluszeit erfordert. Es handelt sich also um eine *weiche Echtzeitbedingung*. Die Merkmalsextraktion als größter Beitrag zur Zykluszeit ist prinzipiell abhängig von der Szene, d.h. von der Anzahl der extrahierbaren Merkmale. Bei der für die Korrespondenzfindung bei differentieller Bewegung vorliegenden weichen Echtzeitbedingung entsteht also kein Problem, wenn die Extraktion etwas länger braucht.

Bei höheren Merkmalsgeschwindigkeiten funktioniert dieses einfache Modell der Nutzung der letzten Merkmalsposition nicht mehr. Dann ist die kinematische Modellierung mit höheren Ableitungen erforderlich. Unter der Annahme der äquidistanten Zykluszeit, die zusätzlich zu 1 normiert wird ($\Delta t \equiv 1$) ergibt sich:

$$\vec{x}_{i+1} = \vec{x}_i + \dot{\vec{x}}_i + \ddot{\vec{x}}_i + \dots \quad (28)$$

Die Approximationen der höheren Ableitungen lauten:

$$\dot{\vec{x}}_i \approx \vec{x}_i - \vec{x}_{i-1} \quad (29)$$

$$\ddot{\vec{x}}_i \approx \dot{\vec{x}}_i - \dot{\vec{x}}_{i-1} = \vec{x}_i - 2\vec{x}_{i-1} + \vec{x}_{i-2} \quad (30)$$

usw.

Es hat sich gezeigt, daß aufgrund des Rauschens der Merkmalskoordinaten, in das alle Fehlerquellen, wie Extraktionsfehler, Vibrationen, usw einfließen, maximal die zweite Ableitung verwendet werden kann [68][65]:

$$\vec{x}_{i+1} = 3\vec{x}_i - 3\vec{x}_{i-1} + \vec{x}_{i-2} \quad (31)$$

Durch die Normierung der Gleichungen auf die Zykluszeit *muß* diese Zykluszeit eingehalten werden. In Abschnitt 3.1. wurde dargelegt, wie die Merkmalsextraktion unter *harten Echtzeitbedingungen* durchgeführt wird.

Mathematisch formaler, jedoch unter Nutzung der gleichen Idee, läßt sich die Prädiktion mit einem Minimum-Varianz-Schätzer, einer Abart des Kalman-Filters, durchführen [15], wobei jedem Merkmal ein eigener Filter zugeordnet wird. Der Zustandsvektor ist die Position im Merkmalsraum sowie deren höhere Ableitungen, wenn die jeweilige Koordinate zeitvariant ist. Die Systemmatrix spiegelt die Zusammenhänge und Redundanzen der Merkmalskoordinaten wieder. Die Meßmatrix ist die Einheitsmatrix.

Die formale Modellierung hat den Vorteil, daß die Merkmalskovarianz mitberechnet wird. Diese dient einmal zur Bestimmung des Volumens im Merkmalsraum, innerhalb dessen die

Korrespondenzkandidaten zu suchen sind. Außerdem läßt sich die Kovarianz zur Gewichtung der Merkmalskoordinaten bei der Distanzbestimmung verwenden. Schließlich liefern die Kovarianzen bei der 3D-Rekonstruktion Informationen über die Genauigkeit der 2D-Merkmalsposition, aus der sich, zusammen mit weiteren Kovarianzen, die Kovarianz der 3D-Merkmalsposition berechnen läßt.

Sowohl bei der direkten kinematischen Modellierung, als auch bei der formaleren mit Hilfe des Minimum-Varianz-Schätzers bleibt die Ursache der Bewegung im Merkmalsraum, nämlich die Art der Eigenbewegung des Beobachters, unberücksichtigt. Es wird nur die Merkmalsbewegung in der Bildebene betrachtet und prädiziert.

Problematisch bei den beiden Verfahren zur Prädiktion ist, daß zur Bestimmung der Bewegungsparameter bzw. des Systemzustands Korrespondenzen benötigt werden, die ohne Unterstützung durch die Prädiktion erzeugt werden müssen. Dieses Problem tritt nicht nur zu Beginn einer Sequenz auf, sondern mit jedem erstmalig entdeckten Merkmal, sowie dann, wenn bei Erreichen der Zykluszeit entdeckte Merkmale nicht mehr bearbeitet werden können, und verworfen werden müssen. Ein Ausweg aus diesem Dilemma ist die Forderung, daß der Beobachter seinen translatorischen und insbesondere rotatorischen Bewegungszustand "langsam" ändert. Das bedeutet langsames Anfahren, keine abrupten Richtungsänderungen, usw. Diese Forderung ist nichts anderes als die oben aufgestellte Glattheitseinschränkung, die diesmal ausgehend von einer gegebenen, nicht weiter zu reduzierenden Zykluszeit auf die Bewegung angewandt wird. Die Forderung entspricht bei Fahrzeugen natürlich nur begrenzt der Realität und schränkt stark ein. Insbesondere hochfrequente Schwingungen durch die elastische Montage der Kamera führen bei unebener Fahrbahn zu Problemen bei der Korrespondenzfindung. Eine andere Alternative ist das initiale Inkaufnehmen des höheren Aufwands der vollständigen Suche bzw. die Vergrößerung des Suchbereichs. Das kann einerseits auf Kosten der Bildrate erfolgen, andererseits auf Kosten der Zahl der verfolgbaren Merkmale. Letzteres erscheint als der bessere Kompromiß, nicht zuletzt, weil eine Änderung der Bildrate technisch aufwendiger ist. Auch andere Korrespondenzverfahren wählen diese Lösung [91].

3.2.4. Nachbarschaftsbeziehungen

Die in 3.2.2. aufgestellten Konzepte zur Erweiterung der Dimensionalität des Merkmalsraums lassen sich weiterführen, wenn nicht mehr nur einzelne Merkmale betrachtet werden, sondern Merkmalsgruppen: Unter der oben gemachten Glattheitsannahme werden zwei Kanten, die mit ihren Endpunkten unter einem bestimmten Winkel aneinanderstoßen, auch im Folgebild höchstwahrscheinlich in dieser Konstellation zu finden sein. Für die Etablierung der Korrespondenz zwischen Merkmalen bedeutet das zweierlei:

Zunächst kann die über die Einzelbilder als stabil angenommene Nachbarschaftsbeziehung einer Kante zur Erweiterung des Merkmalsraums genutzt werden. Die neu hinzugekommene "Merkmalskoordinate" ist ein komplett neuer Merkmalsraum, nämlich der Merkmalsraum der benachbarten Kante. Nimmt man auch den Nachbarn des Anfangspunktes der Kante hinzu, verdreifacht sich die Dimensionalität des Merkmalsraums. Darüberhinaus können auch die Winkel zwischen den Kanten als Merkmalskoordinate verwendet werden und vergrößern den Merkmalsraum nochmal.

Diese fortgeschriebene Nachbarschaftsbeziehung läßt sich alternativ auch zur Abkürzung der Korrespondenzfindung nutzen: Ist für ein Merkmal die Korrespondenz bestimmt, so ergeben sich die Korrespondenzen der benachbarten Merkmale aus den Nachbarschaftsbeziehungen des korrespondierenden Merkmals im vorherigen Bild. Zur Sicherheit kann man diese Methode als Prädiktion verwenden, und zusätzlich die Distanz im Merkmalsraum auf Unterschreiten eines Schwellwertes prüfen.

Nachbarschaftsbeziehungen von Kanten fallen bei der besonderen Merkmalsextraktion, wie sie in 3.1. beschrieben ist, ohne zusätzlichen Aufwand mit ab.

3.3. 3D-Rekonstruktion

In diesem Abschnitt wird beschrieben, wie aus einem Satz von korrespondierenden Merkmalen einer Bildfolge und den 3D-Aufnahmepositionen der Einzelbilder die durch die Projektion verloren gegangene Tiefeninformation rekonstruiert werden kann. Das Ergebnis sind die 3D-Koordinaten der Merkmale. Grundlage der Algorithmen ist die Triangulation, die im Prinzip bei jedem Stereoverfahren Anwendung findet.

Das erste hier beschriebene Verfahren basiert auf punktförmigen Merkmalen und verwendet dafür die Anfangs- und Endpunkte von korrespondierenden Bildkanten. Die 3D-Rekonstruktion beruht auf dem Schnitt der Projektionsgeraden dieser Punkte unter Verwendung eines idealen Lochkamera-Modells. Durch die Entzerrung und Normierung der Bildkanten in der Vorverarbeitung ist die ideale Zentralprojektion anwendbar.

Das zweite Verfahren berücksichtigt die Unsicherheit der Endpunkte von extrahierten Kanten und verwendet nur die Richtungs- und Lageinformation der extrahierten Kanten, nicht aber die exakte Lage der Endpunkte auf den Kanten. Der Kern dieses Verfahrens beruht auf dem Schnitt der Projektionsebenen zur Bestimmung der 3D-Geraden.

Beide Verfahren operieren auf Bildpaaren. Aus einer Folge von Aufnahmen sind daher geeignete Paare auszuwählen. In Absatz 3.3.3. werden Strategien diskutiert, wie geeignete Bildpaare zu bilden sind, und wie mehrere Triangulationen gemittelt werden können.

Bei den hier beschriebenen Triangulationsverfahren ist die Bewegungsrichtung des Fahrzeugs und die Blickrichtung der Kamera zunächst nicht eingeschränkt. Allerdings werden in Abschnitt 4.1. Einschränkungen dargestellt, mit denen die Ausbeute aus einer Bildfolge im Hinblick auf die Genauigkeit und die Anzahl der rekonstruierten 3D-Merkmale optimiert wird.

Die verwendeten Koordinatensysteme und die verwendete Nomenklatur ist in Anhang A.5. dargestellt.

3.3.1. Schnitt der Sichtgeraden

Gegeben sei eine Paar von Kamera-Positionen, an denen ein bestimmter 3D-Weltpunkt \vec{p}_i^w aufgenommen wurde, sowie die jeweiligen Bildkoordinaten $\vec{p}_{i,j}^b$ dieses Punktes

(Abb. 32⁸). Die Kamera-Positionen bestehen aus den 3D-Koordinaten des Projektionszen-

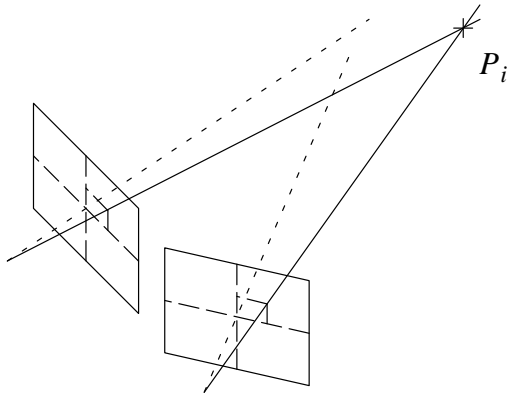


Abb. 32: Schnitt der Projektionsgeraden

trums des Lochkameramodells und den drei Raumwinkeln der optischen Achse. Zu jeder Kamera-Position \vec{p}_j^w lassen sich die beiden homogenen Matrizen R_j und T_j zur Transformation von Weltkoordinaten in ein lokales Koordinatensystem, dessen Ursprung im Projektionszentrum liegt und dessen x-Achse mit der optischen Achse identisch ist, bestimmen⁹:

$$\vec{p}_i^k = R_j \cdot T_j \cdot \vec{p}_i^w \quad (32)$$

Die Bildkoordinaten \vec{p}_i^b des punktförmigen Merkmals P_i bestimmen sich folgendermaßen (Abb. 33):

$$\vec{p}_i^b = \begin{bmatrix} x_{p_i}^b \\ y_{p_i}^b \end{bmatrix} = \begin{bmatrix} -\frac{y_{p_i}^k}{z_{p_i}^k} \\ \frac{x_{p_i}^k}{z_{p_i}^k} \end{bmatrix} \quad (33)$$

Die Projektionsgerade g_i ist definiert durch das Projektionszentrum p_z und den Schnittpunkt mit der Bildebene ap . Diese Punkte ergeben sich im lokalen Kamerakoordinatensystem aus den gemessenen Bildkoordinaten zu:

$$\vec{p}_{z_j}^k = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \vec{a}p_{p_{ij}}^k = \begin{bmatrix} 1 \\ x_{p_i}^b \\ y_{p_i}^b \end{bmatrix} \quad (34)$$

8. O.E.d.A. wird in den folgenden Skizzen die Bildebene zwischen Objekt und Projektionszentrum eingezeichnet, auch wenn dies nicht der technischen Realisierung einer Kamera entspricht.
9. Die Bestimmung dieser Transformationsmatrizen ist bei realen Fahrzeugen nicht trivial. Die Hauptschwierigkeit ist neben einer exakten Lokalisation des Fahrzeugs im Raum die sehr sorgfältige Vermessung der Kameraposition auf dem Fahrzeug.

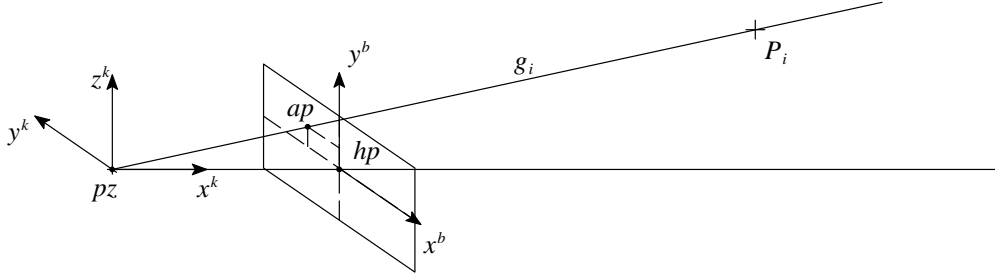


Abb. 33: Zentralprojektion

Die x -Koordinate von ap ist durch die Normierung auf die Bildweite $f = 1$ in (33) festgelegt. $x_{p_i}^b$ und $y_{p_i}^b$ sind die von der Vorverarbeitung gelieferten Koordinaten des Merkmals. Die Transformation in das Weltkoordinatensystem erfolgt durch Inversion der Transformationsmatrizen in (32). Damit ergibt sich die Projektionsgerade $g_{p_i,j}$ in Parameterform:

$$g_{p_{i,j}} : \vec{x}_{p_{i,j}} = \vec{p}z_{p_{i,j}}^w + \lambda_j \cdot (\vec{p}z_{p_{i,j}}^w - \vec{a}p_{p_{i,j}}^w) = \vec{a}_{p_{i,j}} + \lambda_j \cdot \vec{b}_{p_{i,j}} \quad (35)$$

Durch Auswertung eines Stereopaars mit verschiedenen R_j und T_j ergeben sich zwei Projektionsgeraden, die sich idealerweise im 3D-Punkt \vec{p}_i^w schneiden. Durch Gleichsetzen der \vec{x}_{p_i} ergibt sich ein überbestimmtes Gleichungssystem, das aus drei Gleichungen und den beiden Ungekannten λ_j besteht. Aus einem λ_j läßt sich durch Einsetzen in (35) der zu triangulierende 3D-Punkt berechnen.

Durch Fehler in der Merkmalsextraktion und der Kameraposition sind jedoch die Projektionsgeraden in der Regel windschief. Das Gleichungssystem zur Bestimmung der λ_j hat keinen Rangabfall und ist nicht konsistent. Als gesuchter 3D-Punkt wird daher ein Punkt angenommen, der zu beiden Projektionsgeraden den minimalen Abstand hat.

Zur Bestimmung des zu triangulierenden 3D-Punkt \vec{p}_i^w wird aus den beiden Richtungsvektoren \vec{b}_1 und \vec{b}_2 gem. (35)¹⁰ der Normalenvektor $\vec{n}_1 = \vec{b}_1 \times \vec{b}_2$ gebildet, der auf beiden Projektionsgeraden senkrecht steht (siehe Abb. 34). Durch das Kreuzprodukt $\vec{n}_2 = \vec{n}_1 \times \vec{b}_2$ läßt sich der Normalenvektor der Ebene E_1 berechnen, die parallel zu \vec{n}_1 und g_2 steht:

$$E_1 : [(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_2] \cdot \vec{x} - d_1 = 0 \quad d_1 = [(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_2] \cdot \vec{a}_2 \quad (36)$$

In der Normalenform wird d_1 so bestimmt, daß \vec{a}_2 , und damit die gesamte Gerade g_2 auf dieser Ebene liegt. Diese Ebene hat die Eigenschaft, daß sie die kürzeste Verbindung zwischen den

10. Ohne Einschränkung der Allgemeinheit wird für ein Triangulationspaar $j=1,2$ angenommen.

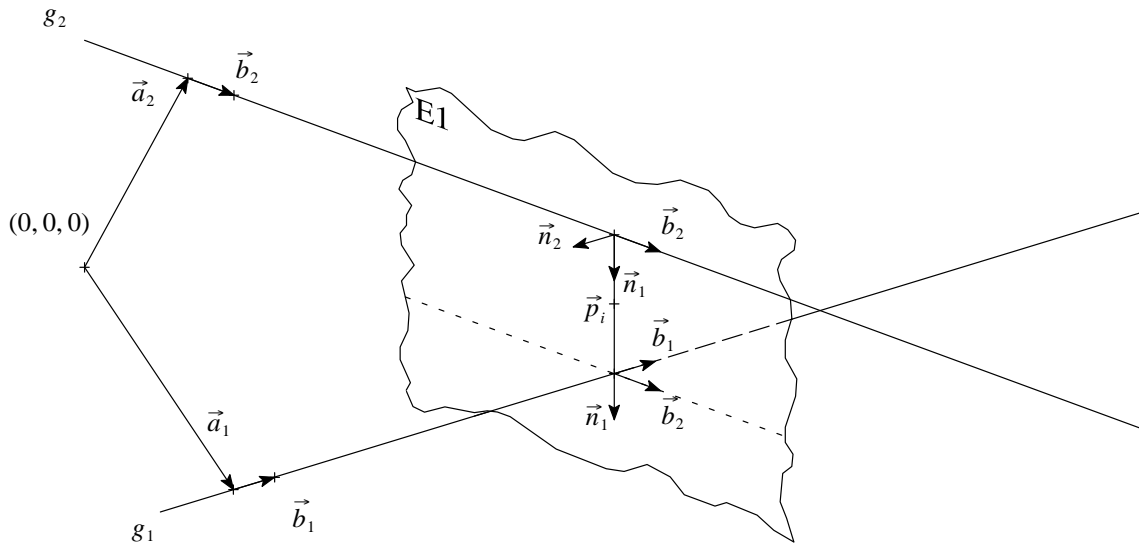


Abb. 34: Schnitt windschiefer Geraden

windschiefen Geraden ebenfalls enthält. Der Anfangspunkt dieser Strecke auf g_1 entsteht durch Schnitt von g_1 mit E_1 , Bestimmung von λ_1 in (37) und Einsetzen dieses λ_1 in (35).

$$\left[(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_2 \right] \cdot (\vec{a}_1 - \vec{a}_2 + \lambda_1 \cdot \vec{b}_1) = 0 \quad (37)$$

Analog ergibt sich der Endpunkt dieser Strecke durch Schnitt von g_2 mit der analog gebildeten Ebene E_2 :

$$\left[(\vec{b}_1 \times \vec{b}_2) \times \vec{b}_1 \right] \cdot (\vec{a}_2 - \vec{a}_1 + \lambda_2 \cdot \vec{b}_2) = 0 \quad (38)$$

Da über die Fehlerverteilung der Lageparameter der beiden Triangulationsgeraden keine Angaben vorliegen, wird von einer Gleichverteilung ausgegangen. Der gesuchte 3D-Punkt ergibt sich als Mittelpunkt der durch Anfangs- und Endpunkt definierten kürzesten Verbindungsstrecke der Projektionsgeraden.

Die in der Praxis vorkommenden Fälle der Triangulation von fälschlicherweise als korrespondierend angenommenen Merkmalen müssen durch Plausibilitätskontrollen, z.B. durch Betrachtung des Mindestabstands der windschiefen Geraden, eliminiert werden.

Das vorgestellte Verfahren geht von der Anschauung aus, den durch verrauschte Sichtgeraden entstandenen Triangulationsfehler zu minimieren. Eine formale Alternative ist die Lösung des überbestimmten Gleichungssystems nach Gleichung (35) mit Hilfe der Pseudoinversen ([30] S. 283). Bei diesem Verfahren wird das überbestimmte Gleichungssystem durch beidseitiges Multiplizieren mit der transponierten Systemmatrix in ein Normalgleichungssystem transfor-

miert, das mit Standardmitteln, z.B. der Gaußelimination, lösbar ist. Die beiden resultierenden Richtungsparameter λ_j erzeugen zwei Aufpunkte auf den Sichtgeraden, deren Mittelpunkt der gesuchte Schnittpunkt ist. Da die Lösung des Normalgleichungssystems per Definition die optimale Lösung des überbestimmten Systems im Sinne der kleinsten Fehlerquadrate ist, ist auch der Schnittpunkt optimal. Beide Verfahren liefern die exakt gleichen numerischen Ergebnisse, wobei der Aufwand für die Pseudoinversenbildung minimal höher ist.

Die in (32) verwendeten Transformationsmatrizen R_j und T_j sind bei Anwendungen mit realen Fahrzeugen genauer zu untersuchen: Da sich der Kamerabezugspunkt in der Regel vom Fahrzeugbezugspunkt unterscheidet, wird eine zusätzliche Koordinatentransformation zwischen dem Fahrzeugzeugkoordinatensystem und dem lokalen Koordinatensystem erforderlich, die die translatorische und rotatorische Lage der Kamera relativ zum Fahrzeug beschreibt. Ist zusätzlich noch ein Schwenk-Neige-Kopf montiert, muß neben der Stellung des Kopfes auch die relative Lage dessen Bezugspunkts in der Transformation berücksichtigt werden. Die Transformation von Weltkoordinaten in Kamerakoordinaten wird also folgendermaßen durchgeführt:

$$\vec{p}_i^k = R^{n-k} \cdot T^{n-k} \cdot R_j^n \cdot R^{f-n} \cdot T^{f-n} \cdot R_j^{w-f} \cdot T_j^{w-f} \cdot \vec{p}_i^w \quad (39)$$

Eine Analyse der Transformationsmatrizen ergibt, daß einige auf konstanten, durch eine Offline-Kalibrierung zu bestimmenden Parametern basieren, andere durch online veränderliche Parameter definiert sind. Daher entfällt teilweise die Indizierung j zur Kennzeichnung der Aufnahmeposition.

$R_j^{w-f} \cdot T_j^{w-f}$: Unter der Annahme, daß das Fahrzeug sich auf ebenem Untergrund bewegt, enthält diese Transformation nur die zwei translatorischen und den rotatorischen Freiheitsgrad der Eigenposition.

$R^{f-n} \cdot T^{f-n}$: Durch die Montage des Schwenk/Neigekopfes werden drei translatorische und drei rotatorische Freiheitsgrade definiert, die offline kalibriert werden müssen.

R_j^n : Hier werden die beiden per Kommando ansteuerbaren rotatorischen Freiheitsgrade des Schwenk/Neigekopfes berücksichtigt.

$R^{n-k} \cdot T^{n-k}$ Die Montage der Kamera auf dem Schwenk/Neigekopf erfordert erneut drei rotatorische und drei translatorische Freiheitsgrade, die offline bestimmt werden müssen.

Sind aufgrund der Aufnahmeperspektive die Projektionsgeraden nahezu parallel, ist bei gegebenen Winkelfehlern der Unsicherheitsbereich der 3D-Koordinaten zu groß. In diesem Fall muß die Triangulation mit einem anderen Paar Meßwerte vorgenommen werden.

Der Sichtgeradenschnitt hat, neben der einfachen und effizienten Implementierbarkeit, einige Nachteile, die darauf beruhen, daß die angenommene Zuverlässigkeit der 2D-Bildkoordinaten des Anfangs- oder Endpunktes einer Kante nicht mit der durch die Extraktion erreichbaren übereinstimmen (vgl. die Diskussion des Merkmalsraums und der Kantendarstellung in 3.2.2.) Kanten können auch bei unveränderter Szene oder Beobachter aufgrund des Pixelrauschens verschieden lang extrahiert werden. Bei bewegtem Beobachter können durch die Eigenbewegung und die partielle Fremdverdeckung einer Kante der sichtbare Endpunkt auf der Kante wandern. Endpunkte von Kanten, deren Projektion im Verlauf der Aufnahmesequenz aus dem Bild laufen, führen ebenfalls zu einer falschen Triangulation.

Diese Nachteile werden vermieden, indem man nicht mehr die vier Koordinaten einer extrahierten Kante verwendet, sondern nur noch zwei gemäß der Merkmalsdarstellung nach Abb. 29-b oder Abb. 29-c. Die extrahierten Kanten werden zu Geraden in der Bildebene verlängert und dann durch den Schnitt der dadurch aufgespannten Sichtebene trianguliert.

3.3.2. Schnitt der Sichtebenen

Die Sichtebene ist bestimmt durch die beiden Endpunkte $ap_{A,j}$ und $ap_{E,j}$ der extrahierten Kante sowie dem Projektionszentrum $p_{z,j}$. Diese Punkte liegen durch Anwendung der Inversen der Transformationsmatrizen in (32) in Weltkoordinaten vor. Die Sichtebene ergibt sich dann in Punkt-Richtungsform zu:

$$SE_j : \vec{x}_j = \vec{p}_{z_j}^w + \lambda_j(\vec{ap}_{A,j}^w - \vec{p}_{z_j}^w) + \mu_j(\vec{ap}_{B,j}^w - \vec{p}_{z_j}^w) = \vec{a}_j + \lambda_j\vec{b}_j + \mu_j\vec{c}_j \quad (40)$$

Durch Schnitt der beiden Sichtebenen SE_1 und SE_2 aus (40) (vgl. Abb. 35) entsteht ein unter-

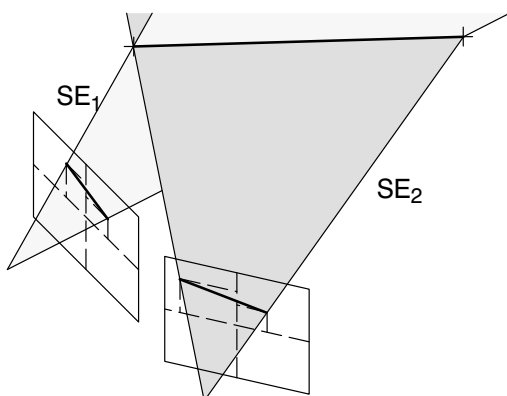


Abb. 35: Schnitt der Projektionsebenen

bestimmtes Gleichungssystem, das nach Elimination von drei Unbekannten zur Gleichung der 3D-Schnittgeraden in Punkt-Richtungsform führt. Die Sonderfälle der identischen oder sich

spitzwinklig schneidenden Ebenen und der parallelen Ebenen können aus der Determinante des Untergleichungssystems erkannt werden. Wie beim Schnitt der Sichtgeraden kann solch ein Positionspaar nicht verwendet werden. Diese Sonderfälle entstehen z.B., wenn die Blickrichtung und die Bewegungsrichtung identisch ist, und die optische Achse die 3D-Kante schneidet. Andere Ansätze zur Berechnung der 3D-Geraden mit Hilfe des Sichtebenenschnitts finden sich z.B. in [36] oder [27].

Stehen die 3D-Geraden zur Verfügung, so können die Endpunkte der 3D-Kante auf verschiedene Weise bestimmt werden. Eine Möglichkeit ist der Schnitt der 3D-Gerade mit den beiden Sichtgeraden. In der Regel muß dann der Endpunkt durch Mittelung der beiden Einzelschnittpunkte bestimmt werden.

Als besser zu beurteilen ist der Schnitt von triangulierten 3D-Geraden, wobei die Auswahl der zu schneidenden Kandidaten auf den von der Vorverarbeitung in der Bildebene ermittelten Nachbarschaftsbeziehungen basiert. Auch in diesem Fall werden aufgrund der akkumulierten Fehler die Geraden windschief sein, so daß die nach dem oben beschriebenen Verfahren ermittelten Schnittpunkte noch auf die Geraden projiziert werden.

Eine heuristische Interpretation der Endpunkt-Konstellation kann das oben angesprochene Problem der sukzessiven partiellen Verdeckung lösen: Bei der Interpretation einer T-Kreuzung kann nach der Triangulation durch Überprüfung der Koplanarität der 3D-Richtungsvektoren sicher erkannt werden, ob es sich um einen Tiefensprung, also eine Verdeckung, handelt, oder ob auch in 3D die Kanten T-förmig angeordnet sind.

3.3.3. Strategie

Für die Triangulation eines 3D-Merkmals können bei n Aufnahmepositionen $M(C_2^n)$ Kombinationen von Merkmalspaaren gebildet werden. Dabei gilt:

$$M(C_2^n) = \binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n \cdot (n-1)}{2} \quad (41)$$

Die vollständige Ausschöpfung aller Kombinationen ist bei Bildfolgen mit 30 und mehr Einzelbildern weder sinnvoll noch machbar. Die im Folgenden diskutierten heuristisch aufgestellten Strategien wählen aus den möglichen Kombinationen Teilmengen von Paaren aus.

Eine erste Strategie versucht Paare auszuwählen, bei denen die Stereobasis möglichst groß ist. Wird eine normale lineare Bewegung angenommen (keine Kurven oder nur geringe Krüm-

mungen), sind dabei die Bilder der Folge zu verwenden, bei denen das Merkmal erstmalig und letztmalig extrahiert wurde. Dabei ist sicherzustellen, daß bei den für die Triangulation verwendeten Bildern das Merkmal vollständig im Bild enthalten ist. Für den Ebenenschnitt reicht es daher, wenn ein ausreichend langer Teil der Kante extrahiert wurde, für den Sichtgeraden-schnitt darf der Endpunkt der Kante nicht durch den Schnitt der Kante mit dem Bildrand entstanden sein. Die Information über die Art des Endpunkts einer Kante wird von der Vorverarbeitung geliefert. Bei der Bewegung in Richtung des Merkmals entsteht ein Merkmal am Focus-of-Expansion und wandert im Verlauf der Folge über den Bildrand. Für das erstmals verwendbare Bild muß das Merkmal eine gewisse Mindestgröße haben, damit auch hier die Extraktionsfehler sich nicht zu stark auswirken. Führt ein gewähltes Positionspaar zu nicht verwendbaren Schnittwinkeln, muß dieses Paar verworfen werden.

Vorteil der beschriebenen Strategie ist, daß die Schnittwinkel maximal sind, und dadurch der Unsicherheitsbereich der 3D-Position reduziert wird. Allerdings machen sich die Kalibrier-genauigkeit der Montageparameter und die Genauigkeit der driftbehafteten, odometriebasierten Eigenposition bei breiterer Stereobasis sehr viel stärker bemerkbar und zehren ab einer gewissen Basisbreite den erzielten Genauigkeitsgewinn wieder auf. Abhilfe schafft die Berücksichtigung von Zwischenpunkten der Sequenz, mit denen zusammen mit Anfangs- und Endbild kettenartig Paare gebildet werden. Die Zwischenpunkte werden so gewählt, daß die optimale Stereobreite erreicht wird. Die berechneten 3D-Koordinaten der Schnittgerade oder des Schnittpunktes werden dann arithmetisch gemittelt. Dies ist zulässig, wenn die Fehlerquellen für die 3D-Positionsungenauigkeiten der Merkmale mit gaußschem Rauschen behaftet angenommen werden. Das trifft für die Merkmalsextraktion hinreichend genau zu, nicht aber für die driftende Odometrie oder die Kalibrierfehler des Kameraaufbaus.

Bei der Korrespondenzfindung besteht auch bei dem robusteren Bild-Bild-Vergleich grundsätzlich die Gefahr der Etablierung von Fehlkorrespondenzen, bei denen die Triangulation falsche 3D-Koordinaten liefert. Die rechnerische Projektion der ermittelten 3D-Merkmalskoordinaten auf die Bildebene an verschiedenen Eigenpositionen und der Vergleich dieser projizierten Bildkoordinaten mit den an dieser Stelle gemessenen plausibilisieren die 3D-Merkmalskoordinaten. Fällt diese Probe an mehreren Stellen negativ aus, ist die Bildfolge in zwei getrennte Einzelfolgen aufzuteilen, und, falls die Folge lang genug ist, erneut der Triangulation zuzuführen.

3.4. Systemarchitektur

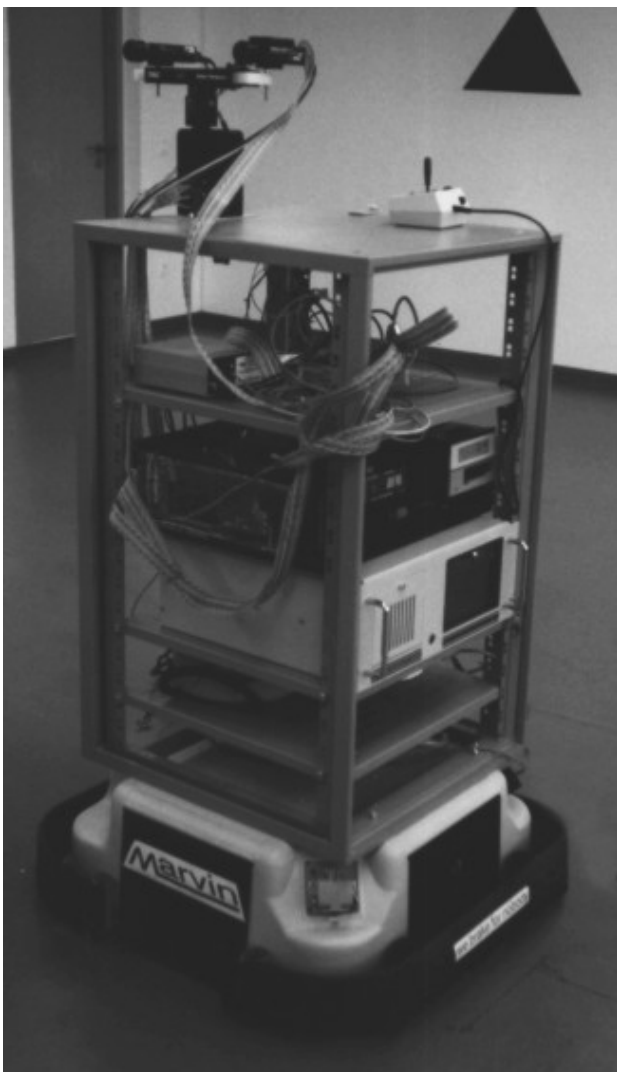


Abb. 36: Das Experimentiersystem MARVIN

In diesem Abschnitt wird beschrieben, wie die vorgestellten Algorithmen für das Bewegungstereo zu einem Gesamtsystem unter Nutzung der Experimentierplattform MARVIN (*M*obile *A*utonomous *R*obot with *V*ideo-based *N*avigation) (Abb. 36) kombiniert sind. Diese Systemstruktur wurde zur Durchführung der in 4.1. und 4.2. beschriebenen exemplarischen Anwendungen Lokalisation und Kartographierung eingesetzt. Weitere Angaben zur Systemstruktur, die über die Aufgabenstellung des Bewegungstereo hinausgehen, finden sich in [23].

Abb. 37 zeigt den Datenfluß und die beteiligten Instanzen. Die grau unterlegten Module sind im Rahmen dieser Arbeit entstanden. Ihre Funktionalität ergibt sich aus den Darlegungen in den Abschnitten 3.1. bis 3.2. und 4.2.

Die *Vorverarbeitung VV* liefert auf Anforderung die extrahierten Merkmale aus einem neu eingelesenen Videobild. Wahlweise wird im Modus "Lernen" eine Vollbildverarbeitung durchgeführt, oder es werden im Modus "Verifikation" zunächst die erwarteten Merkmale, die mit dem Kommando übergeben werden, extrahiert. Steht bis zum Ablauf einer fest eingestellten Zykluszeit noch Zeit zur Verfügung, können weitere, zusätzliche Merkmale extrahiert werden. Der Bildeinzug wird synchron zu der Zykluszeit vorgenommen. Die Vorverarbeitung muß unbedingt diese Echtzeitschranke einhalten, um mit der vorgegebenen Bildrate Schritt zu halten, damit die Annahmen im Lokalisationsmodul und bei der Prädiktion der Merkmale nicht verletzt werden. Bei den nachfolgenden Einheiten können temporäre Stauungen im Da-

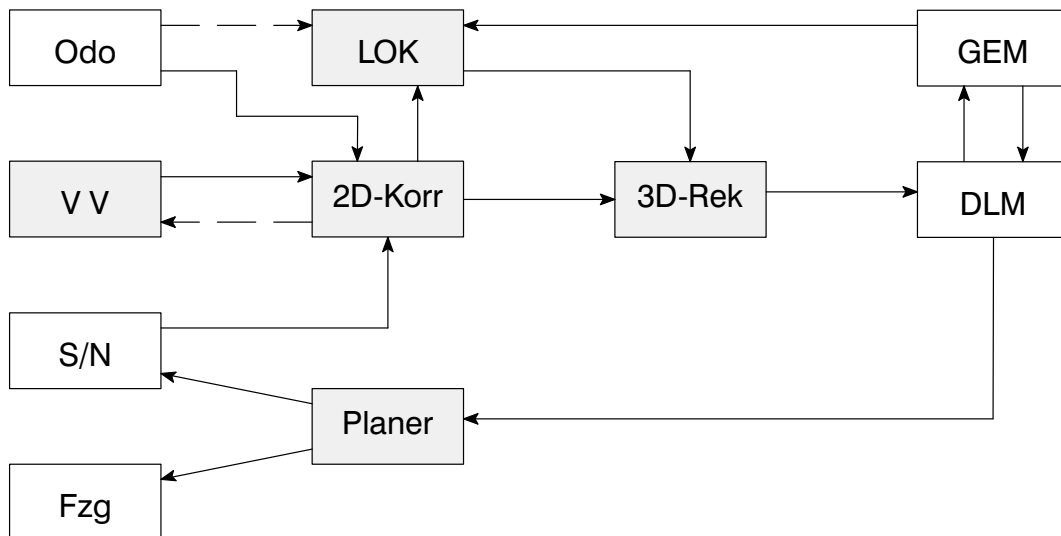


Abb. 37: Datenfluß zwischen den Modulen

tenfluß toleriert werden, wenn die anstehende Rechenlast in der Summe bewältigt werden kann.

Die nachfolgende *2D-Korrespondenzfindung* führt den Bild-Bild-Vergleich (BBV) und die Korrespondenzfindung in der 2D-Bildebene durch. Hierzu werden Merkmalslisten mit der Geschichte eines Merkmals aufgebaut, die die Position des Merkmals im jeweiligen Einzelbild, sowie die zugehörige Weltposition des Fahrzeugs aufgrund der Raddrehgeber und die Stellung des Schwenk/Neigekopfes enthält. Konnte ein Merkmal über mehrere Bilder nicht mehr verfolgt werden, z.B. weil es über den Bildrand hinausgelaufen ist, und ist seine Geschichte ausreichend lang, so kann das Merkmal der *3D-Rekonstruktion* zugeführt werden. Hier wird die 3D-Position der Kante bestimmt und der *Dynamischen Lokalen Merkmalskarte DLM* [13] übergeben, die neben anderen Aufgaben die Wahrnehmungen der gesamten Sensorik zeitlich filtert. Charakteristisch für diese Karte ist, daß sie nur den lokal relevanten Umgebungsausschnitt enthält. Hat sich ein Merkmal als ausreichend stabil erwiesen, kann es in das *Globale Umgebungsmodell GEM* [37][88] übernommen werden. Derzeit wird das GEM allerdings noch mit CAD-Daten und anderem offline erzeugten Umgebungswissen initialisiert. Ist über eine bestimmte Szene bereits Vorwissen im GEM vorhanden, so wird bei einem Szenenwechsel die DLM damit initialisiert.

Das *Lokalisationsmodul* benötigt für jedes Einzelbild die erwarteten Merkmale aufgrund einer Ortshypothese, und die tatsächlich wahrgenommenen Merkmale von der Vorverarbeitung. Die erforderliche Korrespondenzfindung zwischen Bild und Modell wird stark verbessert, wenn die Informationen der 2D-Korrespondenzfindung berücksichtigt werden. Das heißt, wurde ein Bildmerkmal einem Modellmerkmal zugeordnet, und hat die 2D-Merkmalverfol-

gung das Merkmal in nachfolgenden Bildern erneut zuordnen können, so braucht die Zuordnung Bild-Modell nur noch überprüft zu werden.

Für automatische Explorationsfahrten steht ein *Planer* zur Verfügung, der aufgrund der strategischen Vorgabe, der aktuellen Eigenposition und dem aktuellen Inhalt der DLM eine Bahn durch die Szene plant und von der Fahrzeugsteuerung ausführen läßt. Dabei wird auch der optimale Blickwinkel des Schwenk/Neigekopfes eingestellt. Die Experimente zur Lokalisation und Exploration sind jedoch mit manueller Joystick-Steuerung entstanden. Daher wird auf die Beschreibung des Planers verzichtet.

4. Anwendungen des Bewegungstereo

In diesem Kapitel geht es um zwei Anwendungen, in denen die beschriebenen Algorithmen für das Bewegungstereo Verwendung finden. Es sind dies die Gewinnung einer Umgebungsbeschreibung und die schritthaltende Ortsbestimmung.

4.1. Exploration

Unter der *Exploration* ist der Vorgang zu verstehen, eine Szene grob zu erkunden, von der keinerlei Umgebungsinformation vorliegt, und diese Information in einem Umgebungsmodell anzusammeln. In diesem Zusammenhang wird auch der Begriff *Kartographierung* verwendet, der aber eine etwas andere Bedeutung hat: Aufbauend auf einem gegebenen Gerüst, das z.B. durch eine Exploration gewonnen worden sein kann, wird Detailinformation konsistent eingefügt. Bei der *Modellverifikation* liegt ein detailliertes, korrektes Modell bereits vor, und es muß lediglich geprüft werden, ob diese Umgebungsbeschreibung noch korrekt ist. Eine sensorgestützte *Modellanpassung* kann vorgenommen werden, wenn Änderungen in der Szene erkannt wurden, damit für andere Anwendungen des Modells, z.B. Planung und Lokalisation, aktuelle Information verwendet wird. Hierzu ist veraltete Modellinformation zu entfernen und durch neue zu ersetzen.

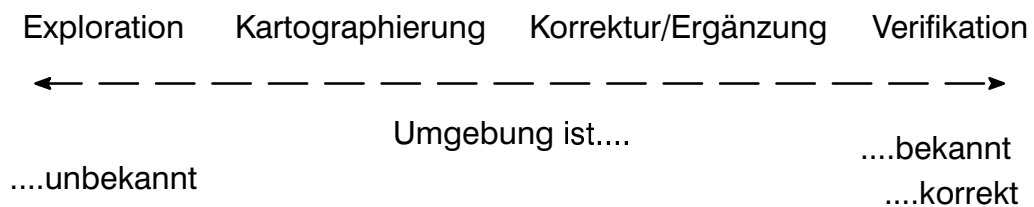


Abb. 38: Bandbreite der sensorgestützten Modelländerung

Die genannten Vorgänge der Modelländerung unterscheiden sich primär durch den vorhandenen szenenspezifischen Informationsgehalt des Modells und den Mechanismus der Integration der neuen Information in das Modell.

Für die sensorbasierte Veränderung des Modells bietet sich der im 3. Kapitel geschilderte Bewegungstereoalgorithmus an. Die dort diskutierte Kette von Merkmalsextraktion, Korrespondenzfindung und 3D-Rekonstruktion liefert die 3D-Koordinaten wahrgenommener Merkmale an eine lokale Karte (DLM), über die sie nach zeitlicher Stabilisierung an das Umgebungsmodell weitergeleitet werden. Für die Anwendung des Bewegungstereo auf die weitgehend automatisierte Modellgewinnung sind einige zusätzliche Komponenten erforderlich. Diese zusätzlichen Algorithmen sind für das eigentliche Bewegungstereo-Verfahren aber un-

erheblich. Die Experimente zur Exploration wurden daher auch durch manuelle Steuerung des Fahrzeugs und des Schwenk–Neigekopfes durchgeführt.

Zu den zusätzlichen Verfahren gehören die Verwaltung und Alterung zeitlicher und räumlicher Unsicherheit der Modellinformation, Bahnplanung und Blickwinkelsteuerung zur Optimierung der Informationsausbeute beim Passieren einer Szene, Unterhaltung einer Befahrbarkeitskarte, usw. Diese Aufgaben sind im Folgenden etwas näher detailliert.

Die Grundlage der gewählten Explorationsstrategie ist das Bestreben, den Raum so oft zu untersuchen, bis durch wiederholtes Entdecken des gleichen Merkmals dessen örtliche Unsicherheit hinreichend reduziert ist. Darüber hinaus muß bei quasistatischen Umgebungen, in denen sich zeitliche Änderungen ergeben können, die Anwesenheitswahrscheinlichkeit eines Merkmals über die Zeit verringert werden, um bei einer Wiederentdeckung erneut auf einen Maximalwert gesetzt zu werden. Damit werden nicht mehr sichtbare Merkmale und Objekte durch Alterung aus dem Modell entfernt und fälschlicherweise eingetragene Artefakte wieder eliminiert. Eine Betrachtung eines Raumabschnitts ist also immer dann erforderlich, wenn sich dort örtlich und/oder zeitlich unsichere Merkmale befinden, oder dieser Abschnitt insgesamt unsicher ist, weil er beispielsweise noch nie eingesehen wurde. Da Bewegungstereo im Stand nicht funktioniert, muß dafür eine Bewegung geplant werden. Hierzu dient die Freiraumkarte, die z.B. auf einem Gitteransatz basiert [25], und in der befahrbare Flächenelemente von unbekanntem oder unbefahrbar unterschieden sind. Die geplante Bahn ist dann ein Kompromiß zwischen den möglicherweise gegensätzlichen Anforderungen von Mindest- und Maximalabstand zu dem zu untersuchenden Szenenabschnitt und optimalem Kamerablickwinkel. Die Trajektorie ist im einfachsten Fall eine Folge von möglichst langen Geradenstücken, aber auch schwach gekrümmte Bahnen haben keine Auswirkung auf die vereinfachte Korrespondenzfindung durch differentielle Bewegung im Bild.

Es hat sich gezeigt, daß ein horizontaler Blickwinkel $|\gamma| < \gamma_{\min}$ zu ungünstig spitzen Schnittwinkeln zwischen den Projektionsebenen bzw. -Geraden führt, wodurch auch bei großen Stereobasen die 3D-Koordinaten des Merkmals nicht ausreichend genau bestimmt werden können. Für Blickwinkel von $|\gamma| > \gamma_{\max}$ ist dagegen bei geringen Objektabständen die Anzahl der Bilder, in denen das projizierte Merkmal innerhalb der Bildfläche zu liegen kommt, zu gering. Dadurch wird die maximale Stereobasis klein, und der resultierende Schnittwinkel ebenfalls spitz. Außerdem ist für solche Blickwinkel die Annahme der differentiellen Bewegung in der Bildebene nur bei kleinen Eigengeschwindigkeiten erfüllt, so daß die automatische Korrespondenzfindung problematisch wird. Die Größen der Grenzwinkel γ_{\min} und γ_{\max} sind abhängig von der erreichten Genauigkeit der Eigenposition und der extrahierten Merkmale.

Die Bestimmung des interessierenden Raumabschnitts erfolgt durch ein Clusterverfahren. Unsichere Merkmale einer Szene werden aufgrund ihrer räumlichen Lage zu Clustern zusammengefaßt. Cluster mit Merkmalen, deren räumliche und zeitliche Sicherheit einen Schwellwert unterschreiten, werden als Zielgebiete bei der Bahnplanung verwendet.

Der für die Planung einer Explorationssequenz erforderliche Abruf von aktueller Modellinformation erfolgt aus der in Abschnitt 3.4. erwähnten *dynamischen lokalen Merkmalskarte (DLM)*. Sie ist verantwortlich für die Modellierung der Unsicherheiten der Modellinformation und die Korrespondenzfindung zwischen vorhandenen und von der Triangulation gelieferten 3D-Merkmalen. Die Clusterbildung erfolgt aufgrund der in der DLM enthaltenen Unsicherheiten [90].

4.2. Lokalisation

Die *Navigation* ist eine der grundlegendsten Aufgaben für ein autonomes Fahrzeug. Unter diesen Begriff aus der Seefahrt fallen alle Vorgänge, die zur Bestimmung des zu fahrenden Kurses erforderlich sind. Hierzu gehört neben der Planung einer kollisionsfreien, effizienten Bahn auch die überwachte Steuerung der Bewegung und die Reaktion auf Abweichungen der Realität von dem internen Modell, das der Planung zugrunde lag.

Ein unverzichtbarer Beitrag zur Bewegungsüberwachung liefert die *schritthaltende Lokalisation*. Hiermit können Abweichungen von der geplanten Sollbahn festgestellt und korrigiert werden. Basis der Lokalisation kann die driftbehaftete Odometrie auf der Basis von Inkrementalgebern sein, aber auch das Videosystem zur Referenzierung an den sichtbaren Merkmalen. Beide Methoden ergänzen sich und können kombiniert werden. In dieser Arbeit wird die schritthaltende Lokalisation unter Nutzung der Algorithmen für das Bewegungstereo besprochen.

Unter Lokalisation versteht man gelegentlich auch die Aufgabe, beim Einschalten eines autonomen, mobilen Systems initial eine Positionsbestimmung in einer bekannten Umgebung durchzuführen (*initiale Lokalisation*). Auf diese Art der Lokalisation wird hier nicht eingegangen.

In Abb. 39 ist der grundsätzliche Ablauf der schritthaltenden Lokalisation skizziert: In der Endlosschleife werden zunächst vom Umgebungsmodell die an der hypothetischen Position lokal sichtbaren Merkmale abgefragt. Diese werden mit den Merkmalen der Vorverarbeitung verglichen, und korrespondierende Paare gebildet. Aus den Diskrepanzen ergibt sich dann ein Korrekturvektor für die Eigenposition, aus dem eine Positionshypothese für den nächsten Schritt extrapoliert wird.

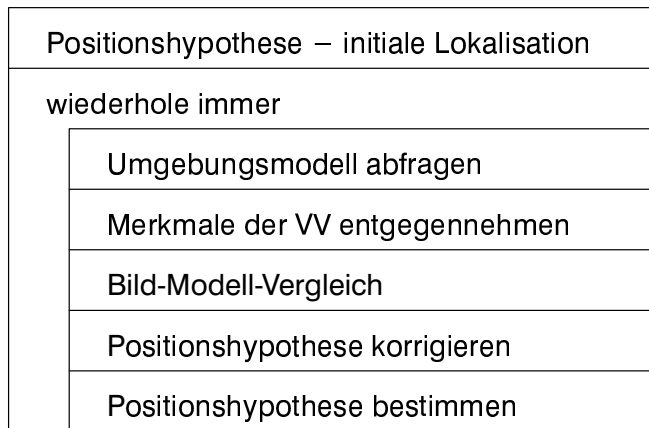


Abb. 39: Struktogramm zum prinzipiellen Ablauf der schritthaltenden Lokalisation

Der hier verwendete Ansatz basiert auf der diskreten Zustandsschätzung durch einen Kalman-Filter. Dieses mächtige Werkzeug ermöglicht die informationstechnisch optimale Nutzung der Meßwerte, die Berücksichtigung von nichtholonomen Zwangsbedingungen der Fahrzeugkinematik, die die verfügbaren Freiheitsgrade der Bewegung einschränken, und die Berücksichtigung der Steuergrößen in der Lokalisation. Die Theorie des Kalman-Filters findet sich z.B. in [29],[10] und ist gekürzt im Anhang A.4. wiedergegeben. Im folgenden ist die Aufstellung des System- und Meßmodells für das verwendete Experimentiersystem dargestellt, die als die wichtigste Aufgabe bei der Anwendung der Estimationsverfahren anzusehen ist. Im Gegensatz zu den Gleichungen (73) und (74) auf Seite 105 im Anhang werden aus Sicht der Anwendung vereinfachte System- und Meßgleichungen verwendet, bei denen die Steuergrößen nicht berücksichtigt sind, die allerdings die Aufgabenstellung erheblich verschärfen. Kann unter diesen Bedingungen ein befriedigendes Lokalisationsergebnis erreicht werden, wird dieses bei Hinzunahme der Steuergrößen erst recht erreicht. Der dann zusätzlich erforderliche Rechenaufwand zu jedem Zeitschritt ist minimal.

4.2.1. Systemmodell

Das verwendete Experimentierfahrzeug MARVIN basiert auf der kommerziellen Plattform LABMATE der Firma TRC [3] und besitzt eine Dreirad-Kinematik mit differentieller Lenkung durch unterschiedliche Geschwindigkeiten v_l und v_r der beiden Antriebsräder (Abb. 40). Die anderen Räder sind frei drehbar und stützen den Rahmen auf dem Boden ab. Der Fahrzeug-Bezugspunkt ist in der Mitte zwischen den Antriebsrädern gewählt. Als Einschränkung der drei Bewegungsfreiheitsgrade x , y und ϕ des allgemeinen Schwebemodells wird davon ausgegangen, daß sich die Trajektorie aus Kreisbogenabschnitten zusammensetzt, die durch Klothoiden verbunden sind. Diese wiederum werden ebenfalls als differentiell

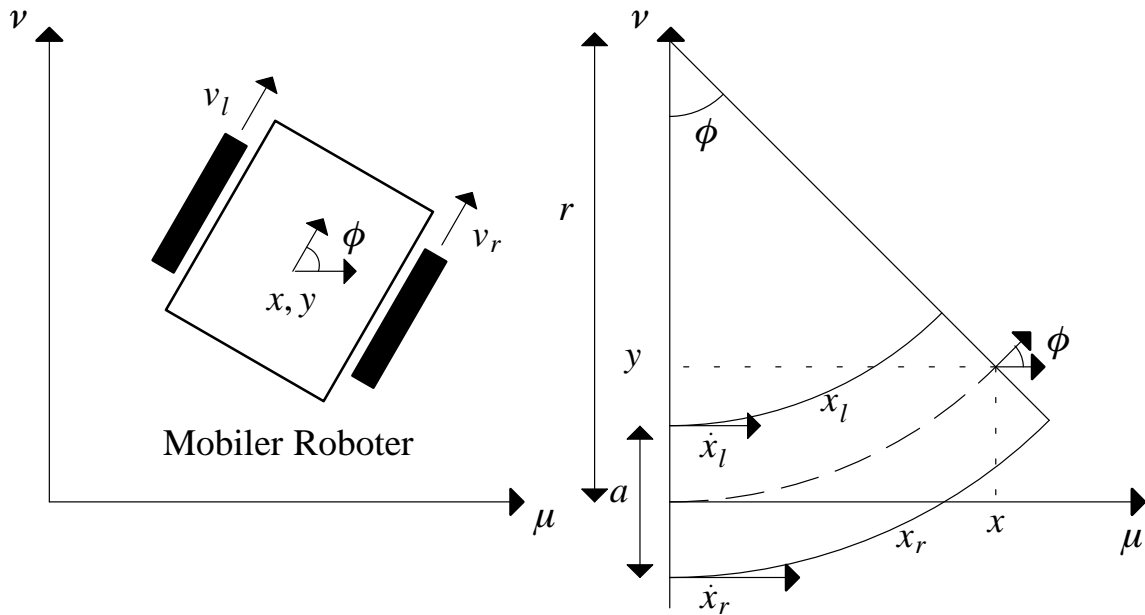


Abb. 40: Fahrzeugkinematik und Kreisbogenapproximation

kleine Kreisbogenabschnitte mit konstantem Krümmungsradius r angenommen. Die beiden Grenzfälle $|r| = 0$ und $|r| = \infty$ entsprechen einer Drehung auf der Stelle und einer Geradeausfahrt. Bei $r > 0$ handelt es sich um eine Linkskurve. Die Änderungsgeschwindigkeit des Krümmungsradius r wird als beschränkt angenommen.

Der gewählte Zustandsvektor $\underline{x} = [v_l, v_r, \phi^w, x^w, y^w]^T$ setzt sich aus der Position des Fahrzeugs x^w, y^w , dessen Orientierung ϕ^w , sowie den Radgeschwindigkeiten v_l und v_r zusammen. Die Positionskoordinaten des Fahrzeugs sind im Weltkoordinatensystem angegeben. Nachfolgend ist die Kennzeichnung des Koordinatensystems weggelassen. Es gelten die folgenden Beziehungen:

$$\dot{x}_l = v_l \quad \dot{x}_r = v_r \quad (42)$$

$$\dot{v}_l = a_l \quad \dot{v}_r = a_r \quad (43)$$

Dabei sind x_l und x_r die differentiell zurückgelegten Strecken der Räder und a_l, a_r die Beschleunigungen der Räder. Aus Abb. 40 folgt:

$$x_l = \phi \cdot \left(r - \frac{a}{2}\right) \quad x_r = \phi \cdot \left(r + \frac{a}{2}\right) \quad (44)$$

$$\dot{x}_l = \dot{r} \cdot \phi + \dot{\phi} \cdot \left(r - \frac{a}{2}\right) \quad \dot{x}_r = \dot{r} \cdot \phi + \dot{\phi} \cdot \left(r + \frac{a}{2}\right) \quad (45)$$

wobei a der effektive Radabstand ist.

Daraus folgt unter der Annahme differentieller Kreisbahnen ($\dot{r} = 0$):

$$\dot{\phi} = \frac{1}{a} \cdot (v_r - v_l) \quad r = \frac{a(v_r + v_l)}{2(v_r - v_l)} \quad (46)$$

Die neue Position ist dann:

$$x = r \cdot \sin \phi \quad y = r \cdot (1 - \cos \phi) \quad (47)$$

$$\dot{x} = \dot{r} \cdot \sin \phi + r \cdot \cos \phi \cdot \dot{\phi} \quad \dot{y} = \dot{r} \cdot (1 - \cos \phi) + r \cdot \sin \phi \cdot \dot{\phi} \quad (48)$$

Mit $\dot{r} = 0$, und (46) in (48) eingesetzt ergibt sich:

$$\dot{x} = \frac{v_l + v_r}{2} \cdot \cos \phi \quad (49)$$

$$\dot{y} = \frac{v_l + v_r}{2} \cdot \sin \phi \quad (50)$$

Nimmt man als Systemstörvektor die Beschleunigungen an den Antriebsrädern zu $\underline{v} = [a_l, a_r]^T$ an, so folgt das implizite, zeitkontinuierliche Systemmodell zu:

$$\begin{bmatrix} \dot{v}_l \\ \dot{v}_r \\ \dot{\phi} \\ \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{a} & \frac{1}{a} & 0 & 0 & 0 \\ \frac{\cos \phi}{2} & \frac{\cos \phi}{2} & 0 & 0 & 0 \\ \frac{\sin \phi}{2} & \frac{\sin \phi}{2} & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_l \\ v_r \\ \phi \\ x \\ y \end{bmatrix} + \begin{bmatrix} a_l \\ a_r \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (51)$$

Die Systemgleichungen werden mit $\Delta t = 1$ diskretisiert (52), und gemäß (53) am Arbeitspunkt $\underline{x}(k)$ bzw. $\underline{v}(k)$ linearisiert.

$$\underline{x}(k+1) = \underline{x}(k) + \Delta t \cdot \underline{f}^k \quad \text{mit} \quad \underline{f}(t) = \underline{\dot{x}}(t) \quad (52)$$

$$\underline{f}^k = \underline{\dot{x}}(k \cdot \Delta t)$$

$$\Delta \underline{x}(k+1) = \left. \frac{\delta \underline{f}^k}{\delta \underline{x}} \right|_{\underline{x}(k)} \cdot \Delta \underline{x}(k) + \left. \frac{\delta \underline{f}^k}{\delta \underline{v}} \right|_{\underline{v}(k)} \cdot \Delta \underline{v}(k) \quad (53)$$

$$= A(k) \cdot \Delta \underline{x}(k) + F(k) \cdot \Delta \underline{v}(k)$$

Die Matrix $F(k)$ ergibt sich zur Einheitsmatrix. Das resultierende, linearisierte und zeitdiskretisierte Systemmodell lautet dann in den Kleinsignalgrößen (der hochgestellte Index kennzeichnet den Zeitschritt):

$$\begin{bmatrix} \Delta v_l^{k+1} \\ \Delta v_r^{k+1} \\ \Delta \phi^{k+1} \\ \Delta x^{k+1} \\ \Delta y^{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\frac{1}{a} & \frac{1}{a} & 1 & 0 & 0 \\ \frac{\cos \phi^k}{2} & \frac{\cos \phi^k}{2} & -\frac{v_l^k + v_r^k}{2} \cdot \sin \phi^k & 1 & 0 \\ \frac{\sin \phi^k}{2} & \frac{\sin \phi^k}{2} & \frac{v_l^k + v_r^k}{2} \cdot \sin \phi^k & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta v_l^k \\ \Delta v_r^k \\ \Delta \phi^k \\ \Delta x^k \\ \Delta y^k \end{bmatrix} + \begin{bmatrix} a_l^k \\ a_r^k \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (54)$$

4.2.2. Meßmodell

Wird bei der linearen, zeitdiskreten Meßgleichung nach (74) der Durchgriff der Steuergrößen vernachlässigt und ist das Meßsystem nichtlinear, ergibt sich:

$$\underline{y}(k) = g\{\underline{x}(k)\} + \underline{w}(k) \quad (55)$$

Der Meßvektor $\underline{y}(k) = (x^b, y^b)^T$ enthält die Bildkoordinaten eines Merkmals, wenn man punktförmige Merkmale annimmt. Die Meßmatrix $C(k)$ modelliert die Transformation der Merkmale aus dem Weltkoordinatensystem in das Kamerakoordinatensystem, und die ideale Projektion auf die Bildebene. In der Meßmatrix sind formal auch die 3D-Koordinaten eines Merkmals enthalten, so daß für jedes Merkmal eine eigene Meßmatrix existiert.

Die Transformationskette von Weltkoordinaten in Kamera-Koordinaten folgt weitgehend der Transformation in Gleichung (39). Im Fall des Meßmodells wird allerdings die Fahrzeugposition transformiert, und die Merkmalsposition ist Transformationsparameter. Die Meßmatrix $C(k)$ setzt sich demnach zusammen aus den drei Kettenmatrizen¹¹:

$$C^*(k) = C(k)^{k-b} \cdot C(k)^{f-k} \cdot C(k)^{pos-f} \quad (56)$$

Durch die erste Matrix $C(k)^{pos-f}$ wird aus Fahrzeugzustand und 3D-Merkmalsskoordinaten die Merkmalsposition im Fahrzeugkoordinatensystem errechnet. $C(k)^{f-k}$ transformiert diese Koordinaten weiter ins Kamerakoordinatensystem. Diese Matrix ist nur von kalibrierbaren, festen Größen abhängig, wenn angenommen wird, daß während der Fahrt der Schwenk/Neigekopf nicht bewegt wird. In Anlehnung an Gleichung (39) ergibt sich also:

$$C^*(k)^{f-k} = R^{n-k} \cdot T^{n-k} \cdot R^n \cdot R^{f-n} \cdot T^{f-n} \quad (57)$$

$C(k)^{k-b}$ führt schließlich die ideale Projektion durch, also die Quotientenbildung der lokalen Koordinaten und der Tiefenkoordinate analog zu Gleichung (33).

11. die hochgestellten Indizes bezeichnen die Koordinatensysteme, zwischen denen transformiert wird, pos bedeutet die Fahrzeugposition in Weltkoordinaten

Dieses nichtlineare Meßmodell muß für den Kalman-Filter am Arbeitspunkt linearisiert werden. Die Jacobimatrix entsteht durch Anwendung der Kettenregel auf die Einzelmatrizen:

$$C(k) = \frac{\delta g\{\underline{x}(k)\}}{\delta \underline{x}(k)} = \frac{C^*(k)}{\delta \underline{x}(k)} \quad (58)$$

Die linearisierte Meßmatrix hat für punktförmige Merkmale dann folgende Struktur:

$$C(k) = \begin{bmatrix} 0 & 0 & c_{13} & c_{14} & c_{15} \\ 0 & 0 & c_{23} & c_{24} & c_{25} \end{bmatrix} \quad (59)$$

Die strukturellen Nullelemente sind bedingt durch die Unabhängigkeit des Meßvektors von den Radgeschwindigkeiten.

Die Dynamik eines mit einem Kalman-Filter modellierten Systems ist maßgeblich abhängig von den Kovarianzmatrizen Q und R des Systemrauschens $\underline{v}(k)$ und des Meßrauschens $\underline{w}(k)$. R ist ein Maß für die Güte der Meßwerte.

Erfolgt die Auslegung der Dynamik eines reinen Beobachters über eine Polvorgabe, so ergibt sich die Dynamik des Kalman-Filters aus den für alle Zeitpunkte $t(k)$ als bekannt vorausgesetzten Systemrausch- sowie den Meßrauschkovarianzen Q und R

R kann vergleichsweise einfach durch experimentelle Analyse des Meßverfahrens anhand eines bekannten Referenzobjektes ermittelt werden. Die Bestimmung von Q ist schwieriger. Sie erfordert die Kenntnisse der Modellierungsungenauigkeiten, sowie die auf die Strecke wirkenden Störungen. Werden die Modellierungsungenauigkeiten vernachlässigt und die gemachten Fehler einzig den auf die Strecke wirkenden Störungen zugeschrieben, so läßt sich eine Aussage über die Größenordnung von Q durch statistische Auswertung von typischen Testfahrten und einer damit verbundenen Auswertung der Störgrößen gewinnen.

In der Praxis werden heuristische Verfahren zur Bestimmung von Q und R angewandt. Ausgehend von einer Einheitsmatrix wird zuerst nur Q und im Anschluß daran R verändert bis sich das gewünschte Filterverhalten einstellt [101].

4.2.3. Sequentielle Innovation

Bei der hier vorliegenden Anwendung muß für jeden Zeitschritt je nach Ergebnis des Bild-Modell-Vergleichs eine unterschiedliche Anzahl von Meßwerten verarbeitet werden. Beim

normalen Kalman-Filter-Algorithmus ist die Größe der Vektoren und Matrizen jedoch fest. Eine Lösung bietet die *sequentielle Innovation*. Dieses Verfahren dient eigentlich dazu, den Rechenaufwand bei der Integration eines Meßvektors mit unkorrelierten, skalaren Einzelmeßwerten zu reduzieren [105]. Die bei der Berechnung der Filtermatrix notwendige Matrixinversion wird für jeden Einzelmeßwert durch eine Division ersetzt. Die Korrekturbeiträge der Einzelmeßwerte zum Systemvektor werden sequentiell aufaddiert, wobei die Fehlerkovarianzmatrix und der Kalman-Koeffizient für jede Teilinnovation separat zu berechnen ist.

Im hier vorliegenden Fall wird über die n etablierten Korrespondenzen des Bild-Modell-Vergleichs eine vektorielle, sequentielle Innovation durchgeführt, wobei für jeden Merkmalsvektor eine neue Kovarianzmatrix und eine Filtermatrix errechnet werden muß. Abb. 41 zeigt das für den Kalman-Filter mit sequentieller Innovation verfeinerte Struktogramm aus Abb. 39.

Schwachpunkt der sequentiellen Innovation ist die Abhängigkeit von der Reihenfolge der eintreffenden Meßwerte und die unterschiedliche dynamisch festgelegte Gewichtung des Beitrags der Einzelmessungen: Handelt es sich bei einem der ersten Meßwerte um einen Ausreißer, wird der Systemzustand fehlerhaft korrigiert und die Güte des Systemzustands fälschlicherweise trotzdem erhöht. Möchte man dieses Problem vermeiden, kann die Integration aller zum aktuellen Zeitschritt verfügbaren Meßvektoren mit einer gaußschen Ausgleichsrechnung durchgeführt werden.

4.2.4. Beiträge des Bewegungstereo

Hauptproblem bei allen merkmalsbasierten Lokalisationsverfahren ist die Etablierung der richtigen Korrespondenzen zwischen Bild und Modell. Die Methoden des Bild-Bild-Vergleichs aus 3.2. sind bei schritthaltender Lokalisation modifiziert auch auf den Bild-Modell-Vergleich anwendbar. Es muß ein reduzierter Merkmalsraum im Sinne von Abschnitt 3.2.2. benutzt werden, da nur die Merkmalspositionen in beiden Bildern vorliegen, die Zusatzattribute aber im Modellbild fehlen. Das Abstandsmaß ist entsprechend anzupassen. Außerdem ist aufgrund der Modellierungsgüte verstärkt mit der Prädiktion von nichtdetektierbaren Merkmalen zu rechnen, aber auch mit der Extraktion von nicht modellierten Merkmalen.

Es hat sich gezeigt, daß es unökonomisch und fehleranfällig ist, bei jedem Einzelschritt die Korrespondenzsuche ohne Berücksichtigung der Korrespondenzen der vorhergehenden Schritte durchzuführen. Die gelegentlich auftretenden Fehlkorrespondenzen einiger weniger Merkmale bewirken eine fehlerhafte Korrektur des prädierten Systemzustands, die in nachfolgenden Schritten zu Fehlkorrespondenzen bei weiteren Merkmalen und schließlich zu einem instabilen Verhalten des Zustandsschätzers führen können.

wiederhole für jeden Zeitschritt k
Umgebungsmodell nach erwarteten Kanten abfragen
Merkmale der Vorverarbeitung entgegennehmen
Bild–Modell–Vergleich
$A(k) = \left. \frac{\delta f(x)}{\delta \underline{x}} \right _{\underline{x}^*(k)}$
wiederhole für alle korrespondierenden Merkmalspaare y_j
bestimme $g_i(k, \underline{x})$
$C_i(k) = \left. \frac{\delta g_i(x)}{\delta \underline{x}} \right _{\underline{x}^*(k)}$
$K(k) = P^*(k) C_i^T(k) \{ C_i(k) P^*(k) C_i^T(k) + R \}^{-1}$
$\underline{x}^*(k) := \underline{x}^*(k) + K(k) \cdot [y(k) - g_i\{\underline{x}^*(k)\}]$
$P^*(k) := \{I - KC_i\} P^* \{I - KC_i\}^T + K R K^T$
$\hat{\underline{x}}(k) := \underline{x}^*(k)$
aktuellen Schätzwert der Position weitergeben
$\underline{x}^*(k+1) = f\{\hat{\underline{x}}(k)\}$
$P^*(k+1) = A(k) \cdot P(k) \cdot A(k)^T + Q$

Abb. 41: Struktogramm zur sequentiellen Innovation

Abhilfe bringt hier die Konservierung von zuverlässigen Bild-Modell-Korrespondenzen über die Einzelbilder der Bildfolge, d.h. Korrespondenzen werden fortgeschrieben, wenn sie in der Vergangenheit mehrmals etabliert wurden. Dies wird durch die Berücksichtigung der Ergebnisse des Bild-Bild-Vergleichs auch bei der Lokalisation erreicht. Es wird also erneut der zeitliche Kontext in Bildfolgen unter Annahme differentieller Bewegung im Bildraum verwendet. Die orthogonale Etablierung von Korrespondenzen im Bild-Bild-Vergleich (BBV) und im Bild-Modell-Vergleich (BMV) verdeutlicht Abb. 42. Die Korrespondenzen in den Merkmalsbildern, die durch Prädiktion aus dem Umgebungsmodell stammen, ergeben sich aus der internen Ordnungsnummer der Merkmale im Umgebungsmodell.

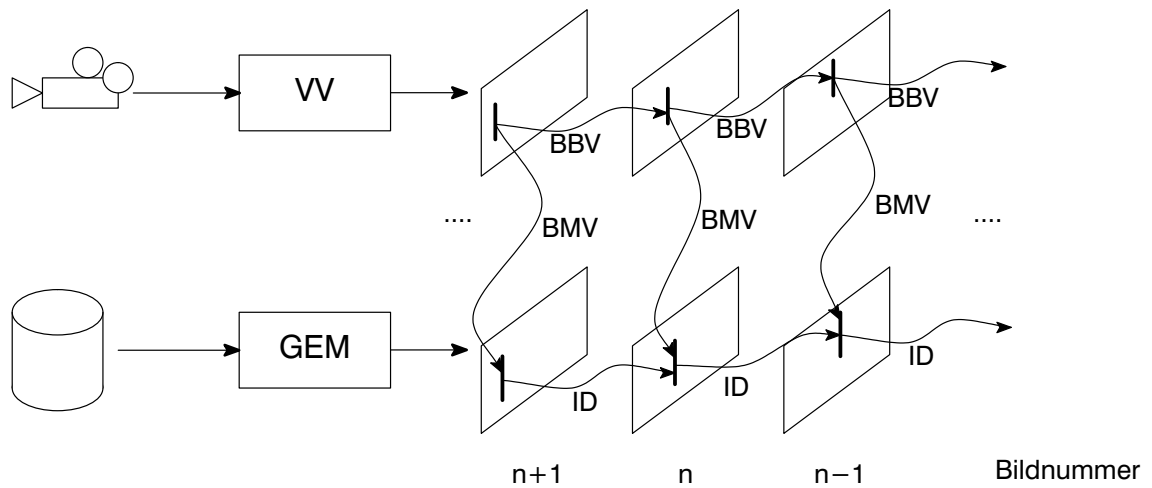


Abb. 42: Korrespondenzfindung bei der Lokalisation

Durch einen zeitlichen Schwellwert werden diese festen Korrespondenzen periodisch durch einen Bild-Modell-Vergleich überprüft, und gegebenenfalls wieder verworfen.

In Abb. 37 ist die Lokalisation im Datenflußdiagramm bereits eingefügt.

5. Ergebnisse und Bewertung

In diesem Kapitel werden ausgewählte qualitative und quantitative Ergebnisse aus dem Umgang mit den in den vorhergehenden Kapiteln beschriebenen Algorithmen dokumentiert, interpretiert und bewertet.

5.1. Merkmalsextraktion

5.1.1. Güte der Extraktion

In [73] wird gezeigt, daß aufgrund von Pixelrauschen mehr als die Hälfte der extrahierten Kanten ihre Lage ändern, obwohl keine Änderung in der Szene stattgefunden hat. Das bedeutet, daß die Koordinaten der Merkmale, wie z.B. Lage der Anfangs- und Endpunkte, Länge, Winkel, usw. in gewissen Grenzen schwanken.

Dieser Effekt wurde auch mit dem Konturfolger beobachtet. Dies ist aus der Ausschnittsvergrößerung der überlagerten Darstellung von 8 Kantenbildern der gleichen Szene¹² in Abb. 43

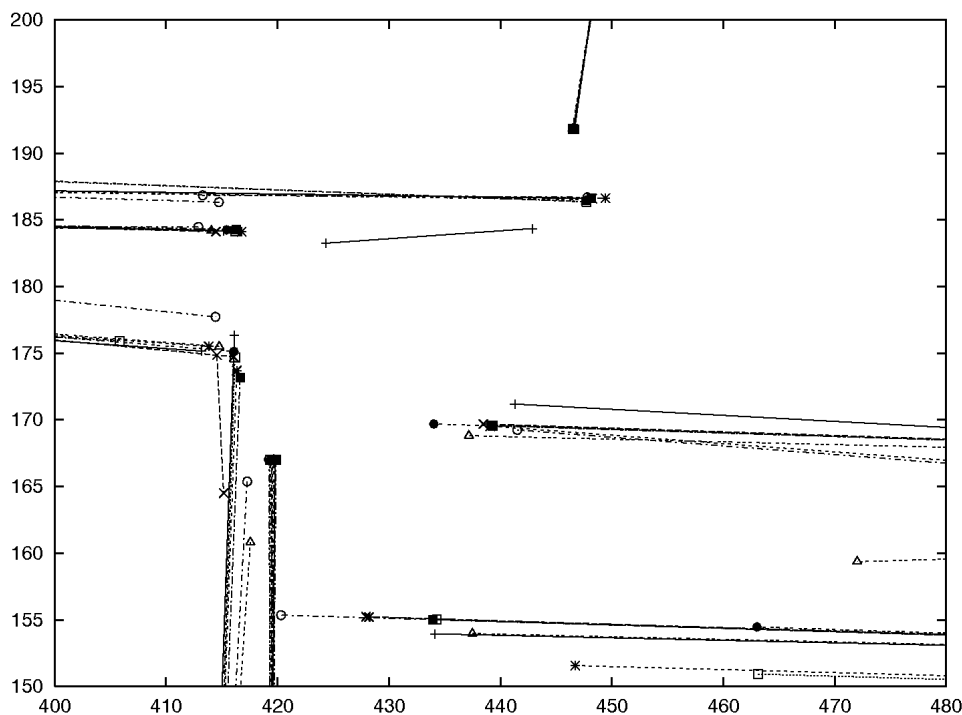


Abb. 43: Lageveränderung extrahierter Kanten in Bildfolgen konstanten Inhalts

ersichtlich. Durch die Regressionsrechnung konnte der Effekt insbesondere bei langen Kanten jedoch reduziert werden.

12. Es handelt sich um die Serie, dessen erstes Bild bf_00001 in Anhang A.3. dargestellt ist.

Negativ hat sich eine Entdeckungswahrscheinlichkeit der Kanten von deutlich unter 100% bemerkbar gemacht, d.h. eine Kante wird nicht immer extrahiert, obwohl sie gut sichtbar, und in den umgebenden Nachbarbildern auch zuverlässig erkannt wurde. Dieses Verhalten ist implementierungsbedingt. Das Problem tritt allerdings auch auf, wenn der Echtzeit-Abbruch in der Bildfolgenverarbeitung verwendet wird.

Eine Schwäche des Algorithmus wird an Richtungsänderungen der Kontur von mehr als 90° sichtbar: Der Konturfolger neigt aufgrund seiner „Trägheit“ zum Überschwingen, so wie sich das aus der Konturspur in Abb. 44¹³ entnehmen läßt: Ausgehend vom Startpunkt in der Nähe

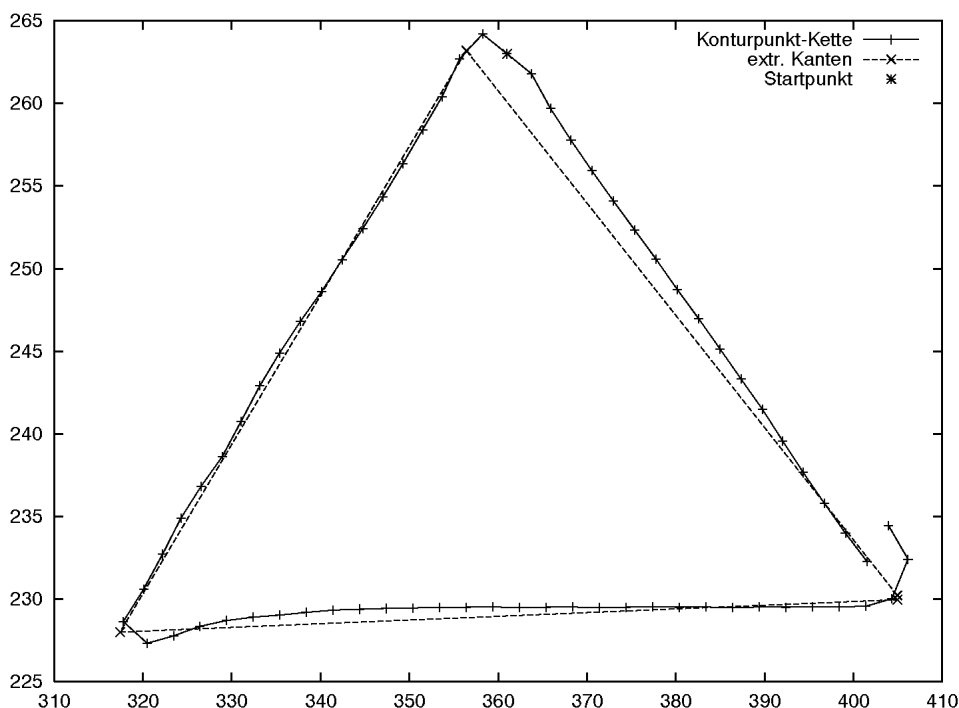


Abb. 44: Überschwingen an scharfen Richtungsänderungen einer Kontur

der oberen Spitze wird die Kontur zuerst im Gegenuhrzeigersinn verfolgt. An den Ecken unten links und unten rechts ist das Verhalten erkennbar. Durch die Regression wird die Auswirkung auf die Kanten gedämpft. Für korrespondierende Kanten entsteht ein für die Korrespondenzfindung schädliches Rauschen des Winkels, weil der Umlaufsinn, mit dem der Konturfolger eine Ecke passiert, verschieden sein kann. Eine temporäre Reduktion der Geschwindigkeit wirkt kontraproduktiv, weil dadurch die stabilisierende Wirkung des H-Feldes nicht verstärkt, sondern reduziert wird. Die lokalen Feldgrößen werden für beide Umlaufsinn gleichermaßen verwendet und mitteln die Verhältnisse vor und hinter der aktuellen Position. Der Konturpunkt wird daher stets in Tangentialrichtung gelenkt, wodurch der vergrößerte Krümmungsradius entsteht. Der für enge Krümmungsradien erforderliche zusätzliche „Vorhalt“ kann auf diese Weise nicht erzeugt werden. Hierfür müßte das Zentrum der felderzeugenden 5×5 -Ma-

13. Auch hier handelt es sich um ein Bild aus der o.g. Serie.

trizen in Bewegungsrichtung verschoben werden, ein Verfahren, das nicht weiter untersucht wurde.

Durch eine temporäre Verkürzung der Schrittweite an erkannten Ecken konnten die Überschwinger auf ein in der Praxis unmerkliches Maß reduziert werden.

Ein weiteres prinzipbedingtes Problem mit dem Konturfolger entsteht an T- oder λ -artigen Kreuzungs-Konstellationen. Trifft der Konturfolger auf eine derartige Kreuzung, kann er je nach Lage des lokalen Feldes entweder geradeaus weiterlaufen, oder abbiegen. Im ersten Fall entstehen zwei Kanten, im zweiten Fall entstehen drei Kanten. Dies verursacht Probleme in der Korrespondenzfindung sowohl beim Bild-Bild-Vergleich, als auch beim Bild-Modell-Vergleich.

5.1.2. Zeitanalyse innerhalb der Vorverarbeitung

Der Rechenaufwand für die einzelnen Teilschritte der Merkmalsextraktion ergibt sich aus den Ergebnissen der Aufruf- und Rechenzeitanalyse, die exemplarisch für die Extraktion des Testbildes mk6_00001 in der Version für die Workstation in der folgenden Tabelle aufgeführt sind. Es wurde kein Vorwissen verwendet, die Bildgröße beträgt 574*736 Bildpunkte, es wurde auf Subpixelauflösung verzichtet, und es wurde keine Echtzeitschranke gesetzt. Die entsprechenden Ergebniskanten sind im Anhang A.3. zu finden.

% der Gesamtzeit	Anzahl der Aufrufe	Funktion
44.06	289	Startpunktbestimmung pro Fenster
33.89	7891	Folgepunktbestimmung
6.78	9189	arctan
6.78	242	folge_kontur
5.08	759	Zeichne Labelbild
3.41	–	Rest

Knapp 80% der Gesamtdauer der Merkmalsextraktion wurde mit der Startpunktsuche und der Feldberechnung verbraucht. Dabei wurden ca. 13500 Pixel überprüft, ob sie als Startpunkt für eine Kontur in Frage kommen, wobei sich bei den eingestellten Schwellwerten ca. 170 als tatsächliche Startpunkte herausstellten. Die Anzahl der überprüften Pixel bei der eigentlichen Konturverfolgung betrug nur noch ca. 8500. Insgesamt wurden ca. 880 Kanten gefunden, wobei ein großer Anteil davon nur wenige Pixel lang ist, und nicht weiter auswertbar ist. Schließt man Kanten unter 20 Pixel aus, so reduziert sich die Anzahl auf ca. 170 Kanten, nimmt man eine Mindestlänge von 50 Pixel, bleiben nur noch etwa 70 Kanten übrig.

Eine Verbesserung des Zeitverhaltens kann durch die Verwendung der in 3.1.6. vorgeschlagenen Kirsch-Operatoren erreicht werden. Für die aufwendige Startpunktsuche kann ebenfalls eine erhebliche Beschleunigung bei gleicher Qualität erreicht werden, indem in Anlehnung an den Kantenextraktionsalgorithmus in [44] nur eine ternäre 5x1 bzw. 1x5 Matrix verwendet wird. Eine Einsparung von etwa 80% am Anteil der Startpunktsuche, bzw. 35% beim Gesamtaufwand erscheint allein durch diese Maßnahme realistisch.

Einige offensichtliche Einsparungen sind bereits implementiert, und tragen zu den bereits kurzen Extraktionszeiten bei, z.B. die Nutzung einer vorab berechneten Funktionstabelle bei der häufig aufgerufenen arctan-Funktion. Die oben genannten qualitativen Ergebnisse rechtfertigten auch die Implementierung der Startpunktbestimmung und der Folgepunktbestimmung in Assembler bei der Version für die Vorverarbeitungskarte. In 5.5. sind die damit erreichten absoluten Zeiten angegeben.

Ursache für den erreichten Durchsatz des Algorithmus ist die Tatsache, daß von den ca. 420000 Pixel nur ca 5% betrachtet werden. Trotzdem werden alle vorhandenen Kanten mit einer gewissen Mindestlänge erkannt. Es werden im Vergleich zu klassischen Ansätzen kleine Faltungsmatrizen verwendet, die selektiv durch ein selbststeuerndes Verfahren auf das Rohbild angewandt werden. Iterationen oder Rekursionen werden bewußt vermieden. Die erforderliche Nachbearbeitung der extrahierten Merkmale, z.B. Verschmelzen von zerfallenen Kanten, Entzerrung, usw., wird nur auf abstrakter Ebene vorgenommen, nachdem die zu bearbeitenden Daten signifikant reduziert worden sind.

Das beschriebene Verfahren läßt sich mit massiver Hardwareunterstützung weiter beschleunigen: Es sind Hardwarebausteine verfügbar, mit der bis zu 8x8 große Faltungskerne schritthalten mit dem Pixeltakt auf ein 8 bit Grauwertbild angewendet werden können [39]. Nimmt man drei dieser sog. Konvolver, kann man aus dem Rohbild ohne Zeitverzug drei Feldbilder erhalten, aus denen die lokalen Feldgrößen an jeder Vollpixel-Position direkt entnommen werden können. Ist die Speicherbandbreite zwischen den Konvolvieren und dem Arbeitsspeicher des Vorverarbeitungsprozessors hoch genug, stehen unmittelbar nach Bildende bereits Feldbilder zur Verfügung. Auch die Nutzung von Subpixelauflösungen ist möglich. Der Umweg über den für jede Subpixelposition erforderlichen Matrizensatz nach Absatz 3.1.5. kann entfallen, weil in diesem Fall die umgebenden Feldwerte bereits vorhanden sind, und die direkte Interpolation der Feldwerte kein nennenswerter Aufwand mehr darstellt.

5.2. Korrespondenzfindung

5.2.1. Methodik zur Bewertung der Korrespondenzfindung

In 3.2.2. wurde dargelegt, daß für eine sichere Korrespondenzfindung die Merkmale eines Bildes möglichst isoliert im Merkmalsraum liegen müssen.

Anhand einer Bildsequenz mit statischer Szene, konstanter Aufnahmeposition, und Änderungen in den Einzelbildern ausschließlich aufgrund von Pixelrauschen wurde die automatische Korrespondenzfindung im Hinblick auf die Abstände im Merkmalsraum untersucht. Nach der Kantenextraktion mit dem Konturfolger wurden manuell Korrespondenzen erzeugt, und danach eine Auswertung der in 3.2.2. eingeführten Maßzahlen Kreuz-Distanz, Korrespondenzdistanz und Bandabstand durchgeführt. Dabei wurde der Algorithmus zur Berechnung der Distanz, sowie die Konfiguration des Merkmalsraums variiert. Für jede Auswertung gibt es drei Arten von Diagrammen:

Das *globale Distanz-Histogramm*, bei dem über die Distanzklassen die Häufigkeit aufgetragen ist, mit der durch vollständige Kombination der Merkmale in beiden Einzelbildern des Paares diese Distanz berechnet wurde. Dabei wird die Interdistanz und die Kreuzdistanz unterschieden.

Idealerweise besteht das globale Distanzhistogramm aus zwei Gebirgen, links ein kleines für die Korrespondenzdistanz und rechts ein großes für die Kreuzdistanz. Beide Gebirge dürfen sich nicht überlappen, und sollten möglichst weit auseinander liegen.

Zur Verdeutlichung des Bandabstands im *vergrößerten globalen Distanz-Histogramm* das unterste Zehntel vergrößert nochmal dargestellt.

Das dritte zugehörige Diagramm zeigt die *lokalen Distanzen*: Nach rechts ist wieder die Distanz aufgetragen, nach oben der Index der Merkmale aus dem ersten Bild. Die für ein Korrespondenzpaar sich ergebende Korrespondenzdistanz wird als Einzelpunkt dargestellt, die Minimal- und Maximaldistanz zu den restlichen Merkmalen des zweiten Bildes wird als Bereich dargestellt. Der Index ist nach steigender Korrespondenzdistanz sortiert. Idealerweise liegt für jeden Index der Punkt für die Kreuzdistanz möglichst weit links von dem lokalen Bereich der Kreuzdistanz.

Die Auswertung der Meßreihen in den nachfolgenden Abschnitten zeigen die Probleme mit der zuverlässigen automatischen Korrespondenzfindung mit den oben vorgeschlagenen Algorithmen und die Wertlosigkeit der zusätzlichen Merkmalsattribute.

5.2.2. Korrespondenzfindung mit Lagemerkmalen

Abb. 45 zeigt das globale Distanzhistogramm unter Verwendung der klassischen Merkmalsat-

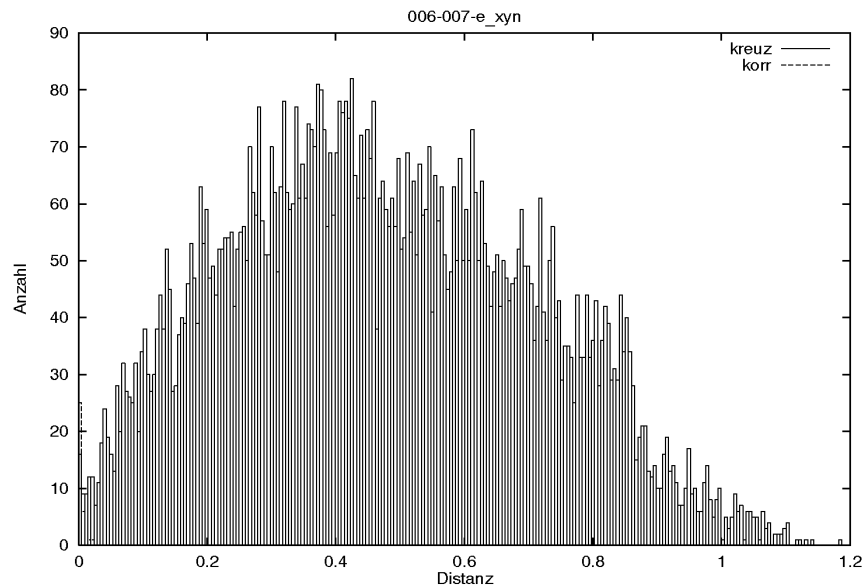


Abb. 45: globales Distanz-Histogramm (euklidisch, Kantenmittelpunkte, normiert)

tribute: Lage des Kantenmittelpunkts, Winkel und Länge. Die Koordinaten wurden vor der Differenzmessung auf die Bildgröße normiert. In der Ausschnittsvergrößerung in Abb. 46

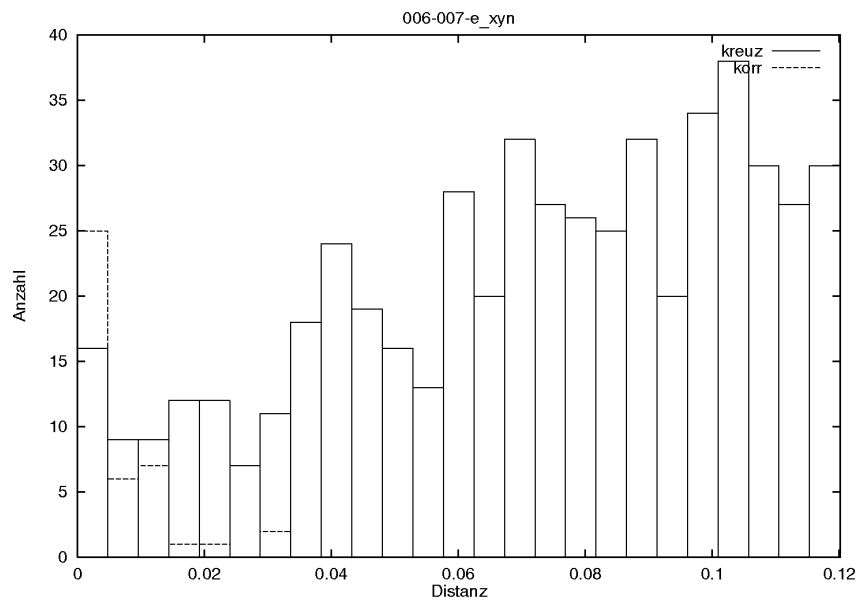


Abb. 46: globales Distanz-Histogramm gezoomt (eukl, Mittelpunkte, normiert)

sind die Überlappung der Kreuzdistanzen und die Ausschnittsvergrößerung: Es gibt deutliche Überlappungen in den globalen Distanzen.

Das lokale Distanz–Histogramm in Abb. 47 zeigt naturgemäß ein freundlicheres Bild, aller-

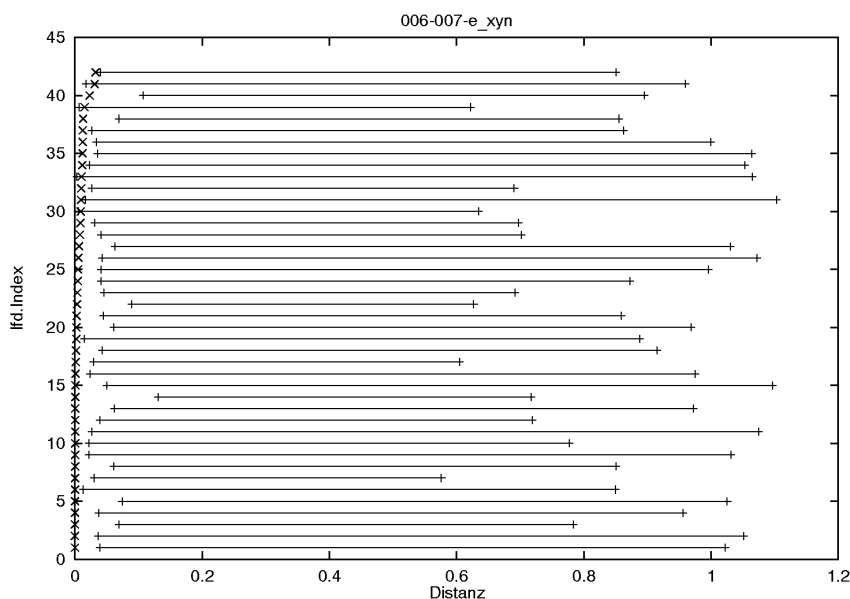


Abb. 47: lokale Distanzen (eukl, Mittelpunkte, normiert)

dings werden auch mit dieser Methode Fehlkorrespondenzen etabliert bzw. riskiert, weil die lokalen Bandabstände nicht ausreichend sind. Die entsprechenden Histogramme anderer Bildpaare zeigen ähnliche Ergebnisse.

Nimmt man die Länge der Kante als zusätzliches Merkmalsattribut mit in die Distanzberechnung hinein, so ergeben sich die Verhältnisse in Abb. 48: Trotz etwas größer werdendem lokalen Bandabstand treten immer noch Überlappungen, und damit Fehlkorrespondenzen auf.

Ähnliche Verbesserungen lassen sich durch Berücksichtigung des Kantenwinkels in Abb. 49 erreichen, wobei die Winkeldifferenz der Kanten auf den Maximalwert von 90° normiert wurde.

Die besten, wenn auch nicht vollständig zufriedenstellenden Ergebnisse werden mit der Kombination von Mittelpunkten, Winkel und Länge erzielt, wie in Abb. 50 dargestellt.

5.2.3. Nutzen der Feldwerte als Zusatzattribute

In Abb. 51 ist das lokale Distanzhistogramm des horizontale E-Feld Feldgrößen als Kandidat für weiter diskriminierende Merkmalsattribute dargestellt. Auf eine Normierung wurde zunächst verzichtet, weil sie erst bei mehrdimensionalen Merkmalsräumen erforderlich ist. Man erkennt einen viel zu geringen Bandabstand. Ein vergleichbares Bild ergibt sich für das vertikale E-Feld. Das lokale Distanzhistogramm Abb. 52 zeigt die Häufung der Kreuzdistanzen in der kleinsten Klasse, sowie die um ca. Faktor 100 höhere Häufigkeit der Kreuzdistanzen gegenüber der Korrespondenzdistanz in dieser Klasse.

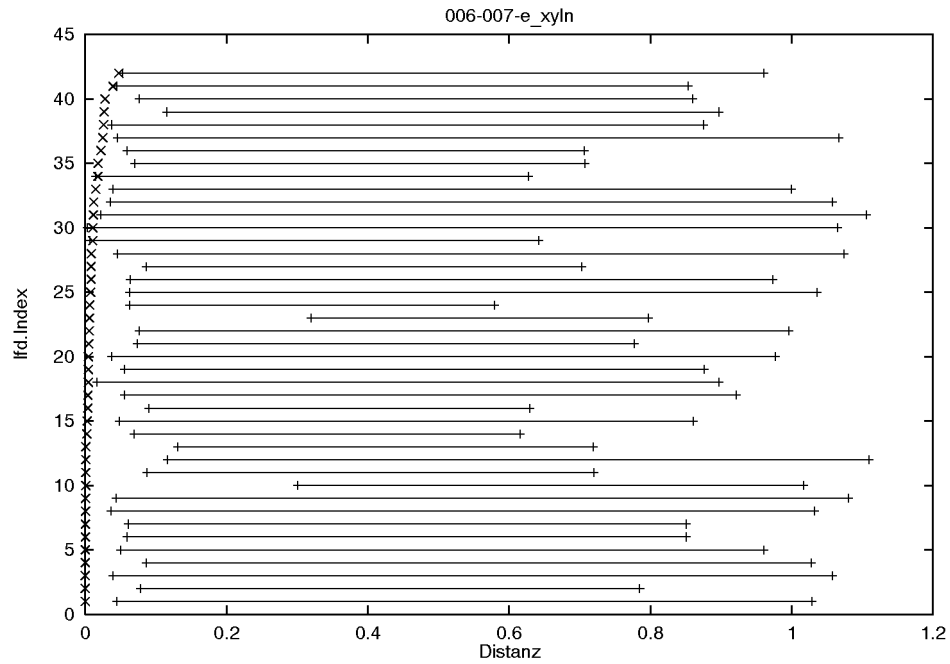


Abb. 48: lokale Distanzen (eukl, Mittelpunkte und Länge, normiert)

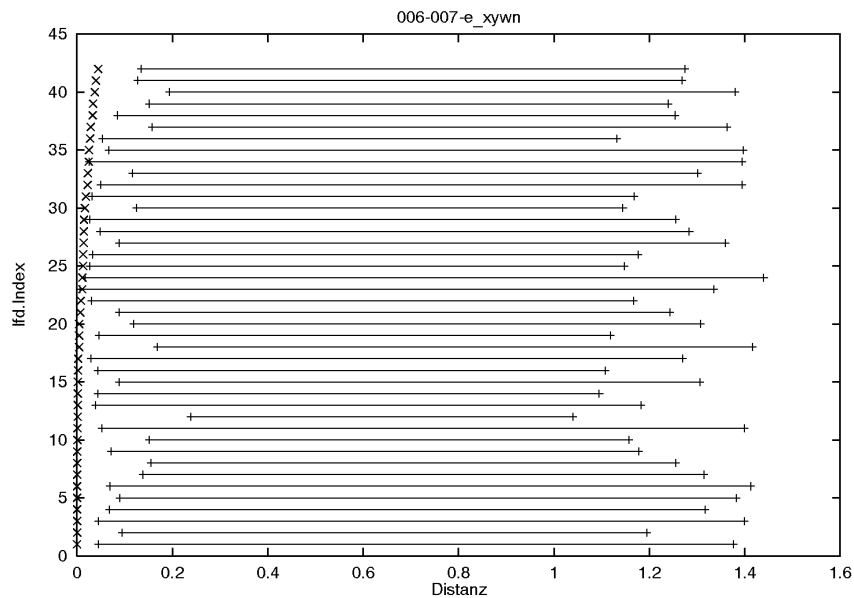


Abb. 49: lokale Distanzen (eukl, Mittelpunkte und Winkel, normiert)

Die Distanzen über das H-Feld in Abb. 53 zeigen einen noch geringeren Nutzen: Im Extremfall ist die Korrespondenzdistanz ein Vielfaches der minimalen Kreuzdistanz.

Zusammenfassend läßt sich folgern, daß die Verwendung der Feld-Attribute bei der Korrespondenzfindung keine Vorteile bringt und der Aufwand zur Weitergabe der bei der Merkmalsextraktion abfallenden Größen an die Korrespondenzfindung demnach überflüssig ist.

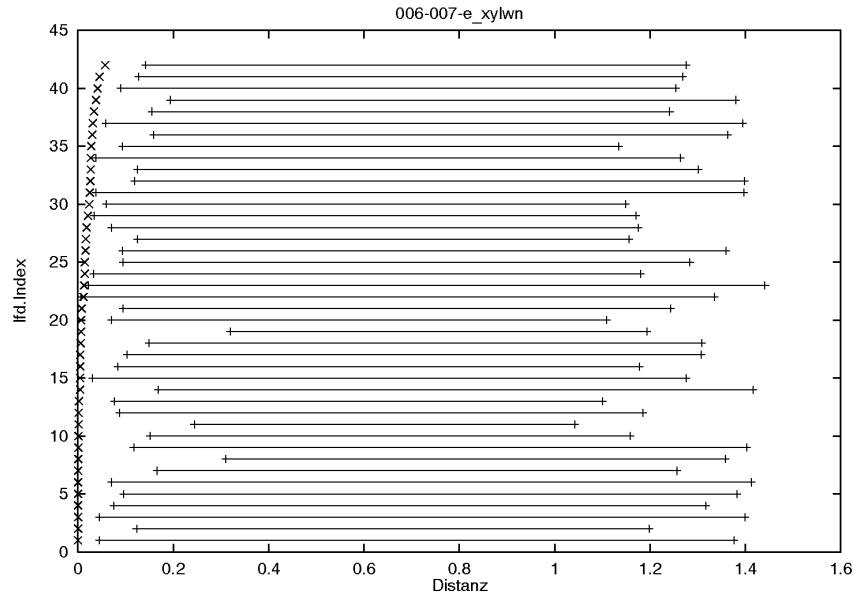


Abb. 50: lokale Distanzen (eukl, Mittelpunkte, Länge und Winkel, normiert)

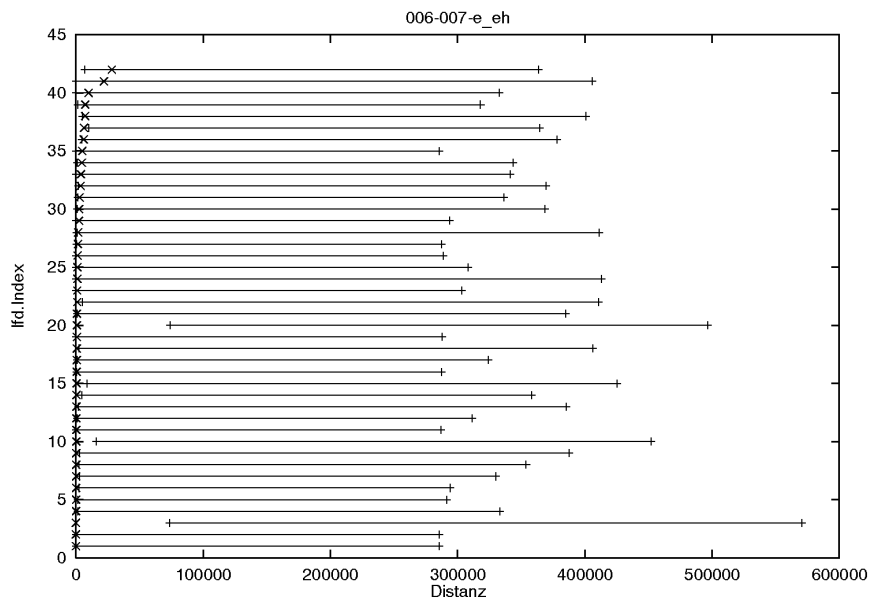


Abb. 51: lokale Distanzen (horizontales E-Feld)

5.2.4. Distanzmaß

In der bisherigen Diskussion der Distanz-Diagramme wurde mit der euklidischen Distanz operiert. In Abb. 54 sind lokale Distanzen dargestellt, die unter den gleichen Bedingungen bestimmt wurden, wie das im Diagramm von Abb. 47 geschehen ist, außer daß die Manhattan-Distanz verwendet wurde. Es zeigt sich an diesem typischen Bild, daß der Bandabstand bei Verwendung der Manhattan-Distanz geringer ist, und durch Überlappung mit erheblich mehr Fehlkorrespondenzen gerechnet werden muß.

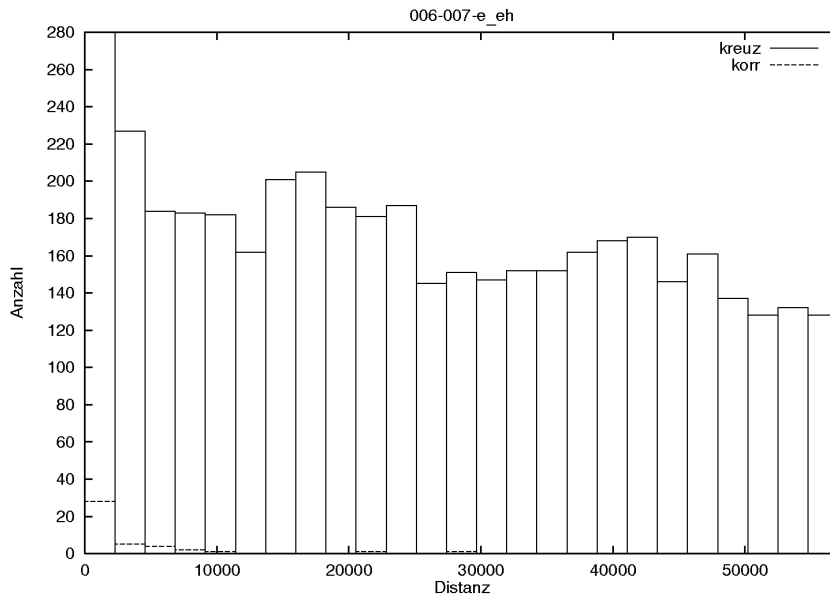


Abb. 52: globale Distanzen (horizontales E-Feld)

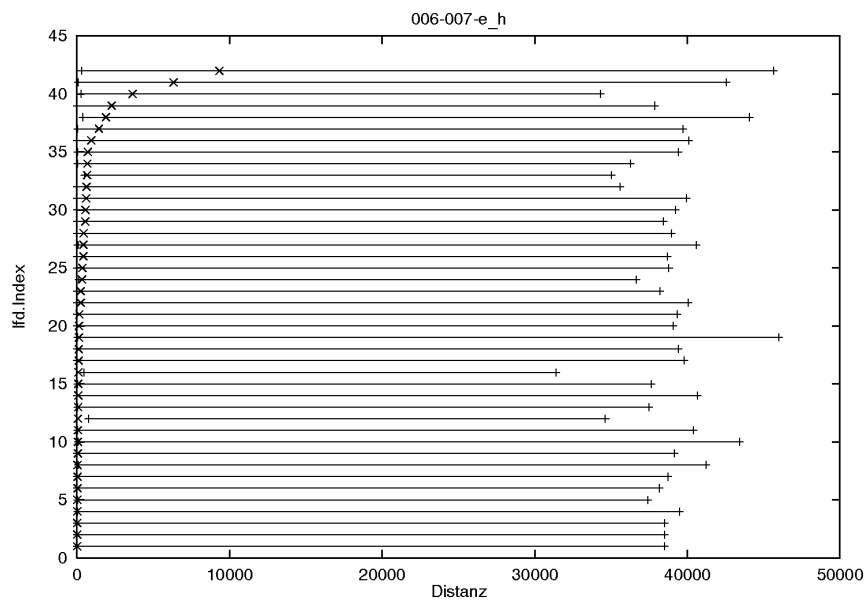


Abb. 53: lokale Distanzen (H-Feld)

Bei einer frühen Implementierung einer gewichteten City-Block-Distanz [38] wurde eine nicht-online-fähige Merkmalsextraktion in Anlehnung an [11] verwendet, die neben den geometrischen Attributen einer Kante eine Fülle von Zusatzattributen zur Korrespondenzfindung liefert. Die eingeführten Koeffizienten zur Gewichtung der einzelnen Attribute dienen der Normierung und der Abschwächung von redundanten Attributen. Die damit erzielten Ergebnisse waren gut, konnten aber wegen der fehlenden Online-Fähigkeit der Merkmalsextraktion in der gegebenen Aufgabenstellung nicht genutzt werden.

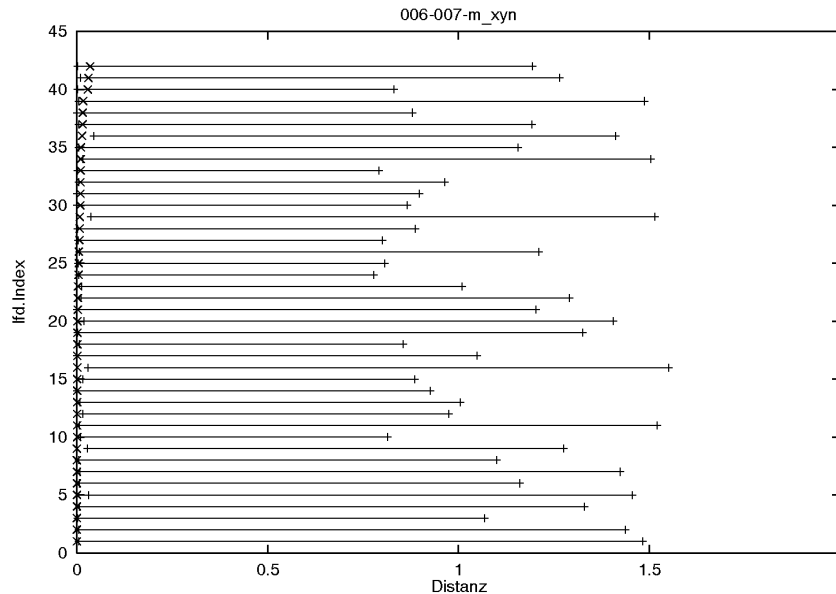


Abb. 54: lokale Distanzen (Manhattan, Mittelpunkte, normiert)

Die letztendlich realisierte Korrespondenzfindung [101] nutzt ausschließlich die Lagekoordinaten der Kanten. Vorteil ist hier, daß das Verfahren sowohl für den Bild-Bild-Vergleich, als auch für den 2D-Bild-Modell-Vergleich in der Lokalisation verwendet werden kann.

Beim Bild-Modell-Vergleich besteht eine noch größere Gefahr der Fehlkorrespondenzen, weil aufgrund einer fehlerbehafteten Positionshypothese korrespondierende Merkmale weit auseinanderliegen können. Erst die zeitliche Stabilisierung der Korrespondenzen durch den Bild-Bild-Vergleich nach Abb. 42 auf Seite 75 macht den Bild-Modell-Vergleich für die Lokalisation hinreichend stabil.

5.3. 3D-Rekonstruktion

Mit den im Experimentiersystem MARVIN auftretenden Fehlern und den auftretenden Fehlkorrespondenzen ergibt sich bei einem etwa 4 Meter entfernten Merkmal eine Lageunsicherheit von ca. 10% der Entfernung. Daher ist bei der Exploration die Berücksichtigung von kurzen Kanten nicht sinnvoll, weshalb sie auch bei der Extraktion in einem frühen Bearbeitungsstadium unterdrückt werden. Aus diesem Grund muß ein Szenenabschnitt entweder aus der Nähe exploriert werden, oder die resultierende Dichte der Umgebungsbeschreibung bleibt gering.

Die primären Fehlerquellen der resultierenden Toleranzen sind die Fehler in der Odometrie, Fehler in der Kalibrierung der Montageparameter der Kamera, Fehler bei der Kantenextrak-

tion und Fehlkorrespondenzen. Diese Fehler spielen bei allen Stereoverfahren eine Rolle, wirken sich aber unterschiedlich aus: Grundsätzlich müssen die Messungen bei jedem Stereoverfahren vor der eigentlichen Triangulierung in ein gemeinsames Koordinatensystem transformiert werden. Abb. 55 stellt dabei die Transformationskette von binokularem Stereo und Be-

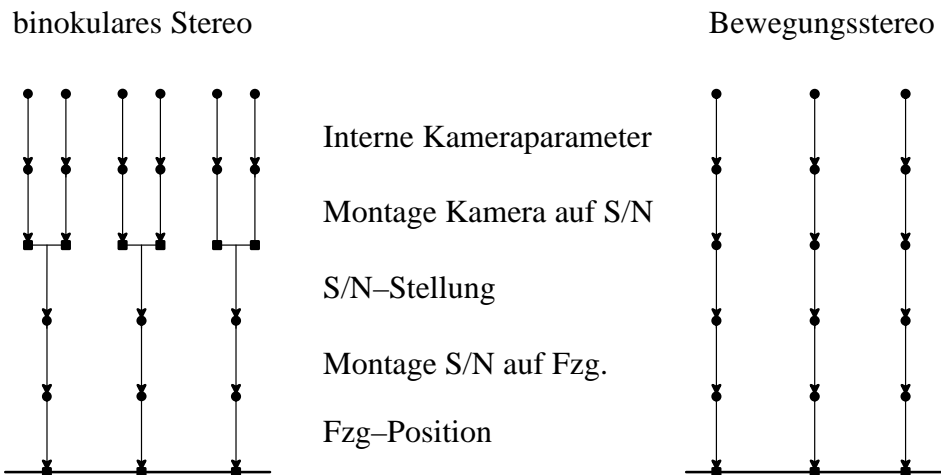


Abb. 55: Reihenfolge der Transformationen bei der 3D-Rekonstruktion

wegungsstereo gemäß Formel (39) gegenüber. Horizontal sind die unterschiedlichen Einzelbilder dargestellt und vertikal die Transformation von den Bildkoordinaten oben in die 3D-Weltkoordinaten unten. Die Ebene der Korrespondenzfindung ist durch schwarze Quadrate markiert. Bei binokularem Stereo kann bereits aufgrund eines Stereobildes die Struktur der Szene rekonstruiert werden.

Beim binokularen und bifokalen Stereo bestimmen die mit einer Offline- oder Online-Kalibrierung im Prinzip einfach ermittelbaren internen und externen Kameraparameter die Transformationsmatrix, die die beiden Stereobilder in Beziehung setzt. Die resultierenden 3D-Koordinaten müssen dann schließlich noch in das Weltkoordinatensystem übertragen werden. Sollen die Beobachtungen aus mehreren Blickwinkeln akkumuliert werden, ist eine zweite Korrespondenzfindung im 3D-Koordinatenraum erforderlich. Dabei können aufgelaufene Fehler in der Aufnahmeposition eliminiert werden.

Beim Bewegungsstereo muß die Triangulation im ortsfesten 3D-Weltkoordinaten-Raum erfolgen, weshalb zuvor eine mehrstufige Transformationskette durchlaufen werden muß. Dabei spielen die internen Kameraparameter eine kleinere Rolle, weil ja für alle Aufnahmen der gleiche physikalische Sensor verwendet wird. Allerdings gibt es keine festen, durch Offline-Kalibrierung zu bestimmenden Transformationsmatrizen zwischen den Aufnahmepositionen. Die Transformation ist vielmehr abhängig von der fehlerbehafteten Eigenposition, von der Montagekalibrierung, sowie der durch unüberwachte Steuerkommandos und Lose des mechanischen Getriebes ungenauen Stellung des Schwenk-Neigekopfes. Durch die somit un-

sichere Stereobasis werden die Vorteile des Bewegungstereo – Nutzung des gleichen physikalischen Sensors und variable, prinzipiell unbeschränkte Stereobreite – teilweise wieder aufgehoben. Wie überall beim Maschinensehen sind die Winkeltoleranzen besonders schädlich.

5.4. Lokalisation

Abb. 56 zeigt eine typische Lokalisationsfahrt von etwa 8 Metern Länge durch eine Innen-

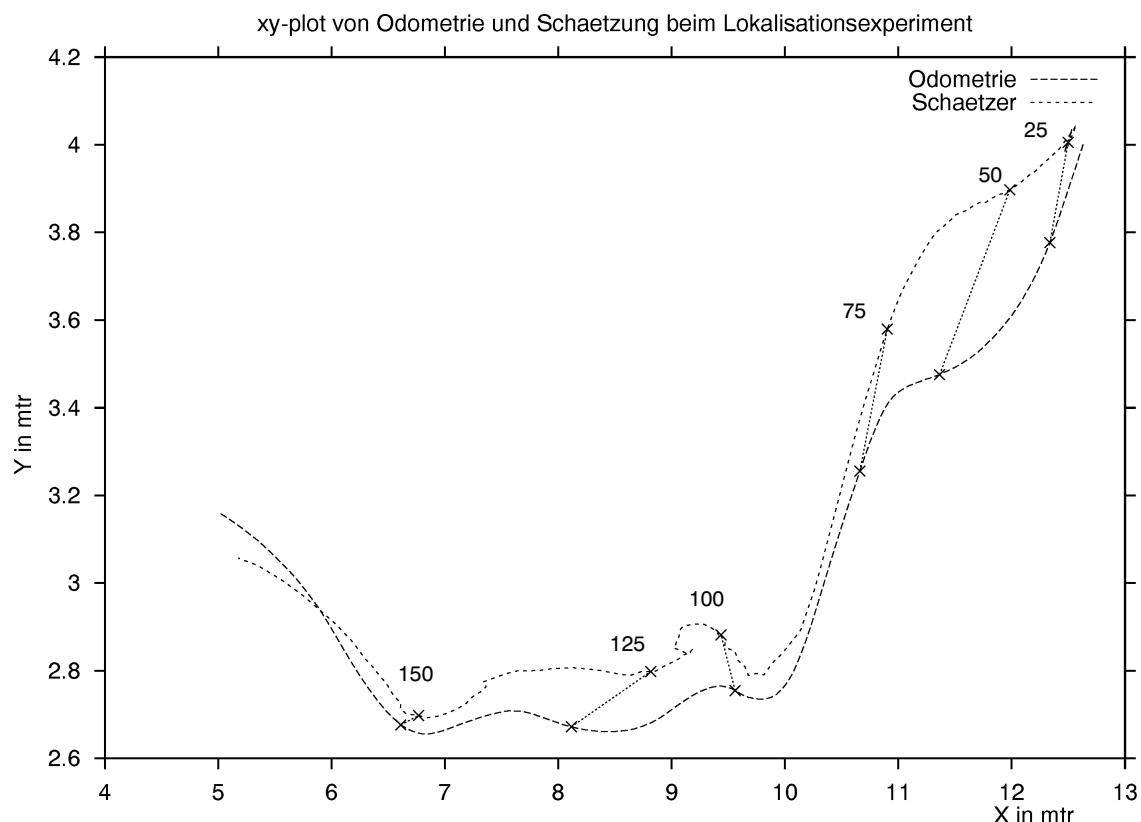


Abb. 56: XY-Plot eines Lokalisationsexperimentes

raumszene. Dabei wurde die Odometrie von MARVIN als Referenz für die ausschließlich durch die Zustandsschätzung gemäß 4.2. fortgeschriebene Eigenposition verwendet. Dieses Vorgehen ist prinzipiell kritisch zu beurteilen, denn mit der absolut referenzierenden Lokalisation soll ja die Drift der Odometrie korrigiert werden. Die Länge der Verfahwege bei den Experimenten, die festgestellte Ablage, die spezifizierte Genauigkeit der Odometrie¹⁴, sowie die Einfachheit der Messung rechtfertigen jedoch dieses Vorgehen. In der Praxis hat sich die Empfindlichkeit auf die exakte Ausrichtung des Fahrzeugs am Startpunkt zur Vermeidung von Winkelfehlern negativ bemerkbar gemacht.

Die Kamera war in etwa 140 cm Höhe, 25 cm vor dem Mittelpunkt der Antriebsachse und 11 cm links von der Mitte mit Blick in Fahrtrichtung montiert. Die nicht-symmetrische Mon-

14. TRC spezifiziert in [3] eine Genauigkeit der Odometrie von 1%, wobei sich der Fehler bekanntermaßen mit der zurückgelegten Wegstrecke akkumuliert. Es wird daher empfohlen, alle 10 Meter eine Referenzierung an einer Landmarke vorzunehmen.

tage kommt daher, daß die verwendete Kamera Teil einer binokularen Stereoanordnung mit einer Basisbreite von etwa 22 cm ist, so wie in Abb. 36 auf Seite 62 sichtbar. Der Schwenk/Neigekopf wurde während der Fahrt nicht verändert. Die Meßpaare für Schätzposition und Odometrie wurden zeitsynchron erfaßt. Zur Ermöglichung einer Offline-Auswertung wurden in der hier präsentierten Meßreihe Rohbilder abgespeichert. Dabei war aufgrund der höheren Datenmenge eine etwa 4fach längere Zykluszeit erforderlich.

In Abb. 57 ist ein typisches Rohbild der erstellten Meßreihe mk_lok4 dargestellt. Die Auflö-

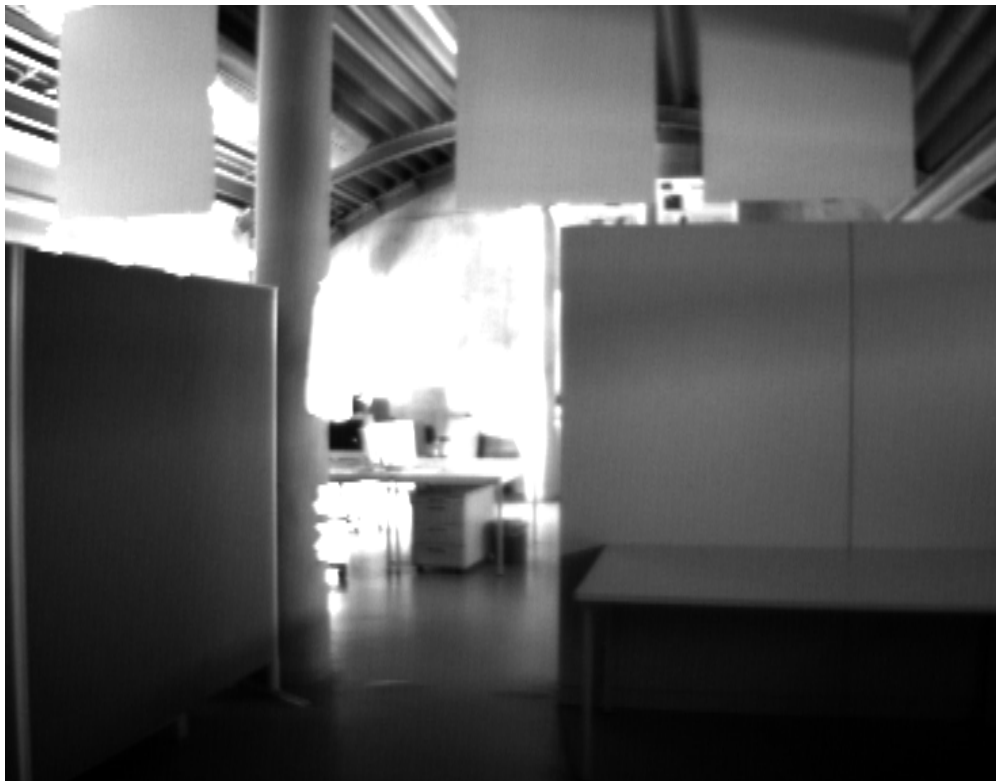


Abb. 57: Grauwertbild aus der Meßreihe mk_lok4 zur Lokalisation

sung in der Vertikalen ist gegenüber der Horizontalen reduziert, da Halbbilder des Zeilensprungverfahrens aufgezeichnet wurden, der Anschauung wegen jedoch eine vertikale Entzerrung vorgenommen wurde. Wie aus dem Referenzbild ersichtlich ist, liegen ungeschönte, reale Beleuchtungsverhältnisse mit hoher Helligkeitsdynamik vor (sommerliches, seitliches Tageslicht), denen das verwendete Kamerasystem nicht gewachsen war, da keine Maßnahmen zur Helligkeitsanpassung (variable Blende, elektronischer Shutter) vorgesehen waren. Die daher unvermeidlichen Übersteuerungen bewirken das Abreißen und Verzerren von Kanten. Außerdem hat sich herausgestellt, daß der Fokus nicht optimal eingestellt wurde. Trotz dieser ungünstigen Voraussetzungen konnte eine schritthaltende Lokalisation durchgeführt werden, wie Abb. 56 zeigt.

Die Fahrt beginnt aufgrund des gewählten Koordinatensystems in Abb. 56 rechts oben. Jeder 25. Meßpunkt auf beiden Trajektorien ist markiert. Korrespondierende Markierungen sind durch eine punktierte Linie verbunden.

Man erkennt, daß zu Beginn zunächst der Positionsschätzer kaum reagiert, sondern lediglich eine kleine Korrekturbewegung durchführt, obwohl sich das Fahrzeug schon bewegt. Es folgt eine lange Phase, bestehend aus Bildnr. 25-125, in der der Schätzer dem wahren Verlauf der Trajektorie im wesentlichen folgt, jedoch einen starken Nachlauf besitzt. Die resultierenden Ablagen zu den Odometriewerten sind in Abb. 58 verdeutlicht. Die Abweichung des Fahr-

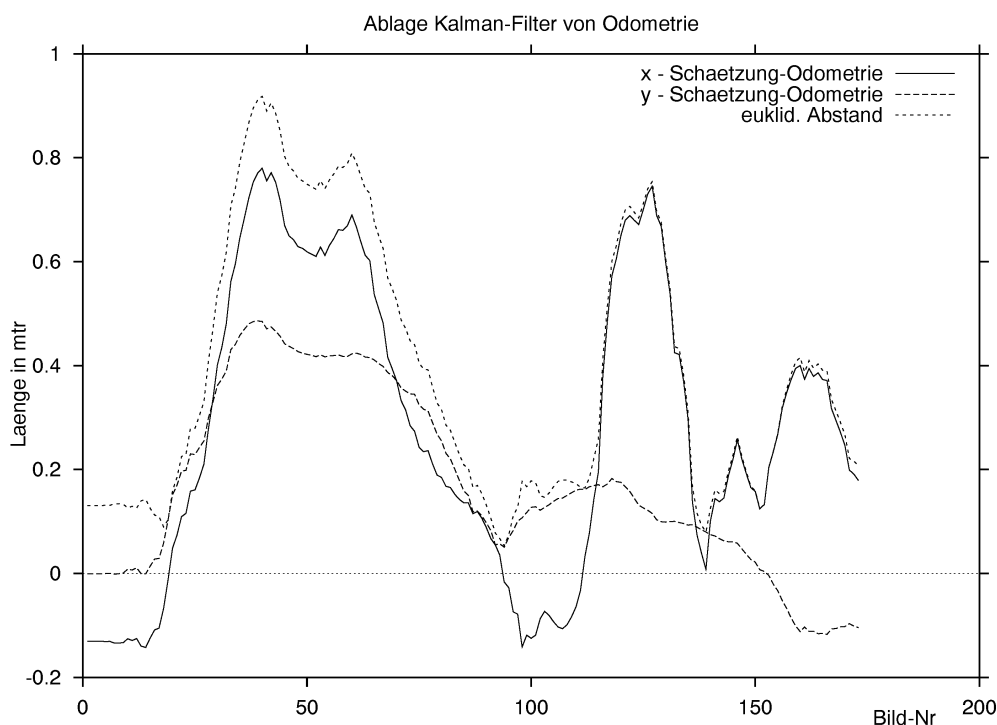


Abb. 58: Abweichung x/y zwischen Schätzung und Odometrie

zeugwinkels kann aus Abb. 59 entnommen werden.

Es folgt eine Phase um Bild-Nr. 100, in der nur wenige Korrespondenzen etabliert werden konnten, und eine aufgetretene Fehlkorrespondenz den Schätzer fälschlicherweise in eine andere Richtung zieht. In Abb. 60 ist der relevante Bereich aus Abb. 56 vergrößert dargestellt. Mit der automatischen Auftrennung dieser festen Korrespondenz nach einer gewissen Anzahl von Bildern findet der Schätzer aufgrund der anderen korrekten Korrespondenzen bis auf eine Ablage von etwa 20-30 cm an die wahre Position zurück. Ab Bild 135 der Sequenz folgt der Schätzer der Position wieder recht gut, da die Zahl der ausgewerteten Korrespondenzen wieder steigt. Die Anzahl der Merkmale pro Bild und die Anzahl der gebildeten Korrespondenzen geht aus Abb. 61 hervor.

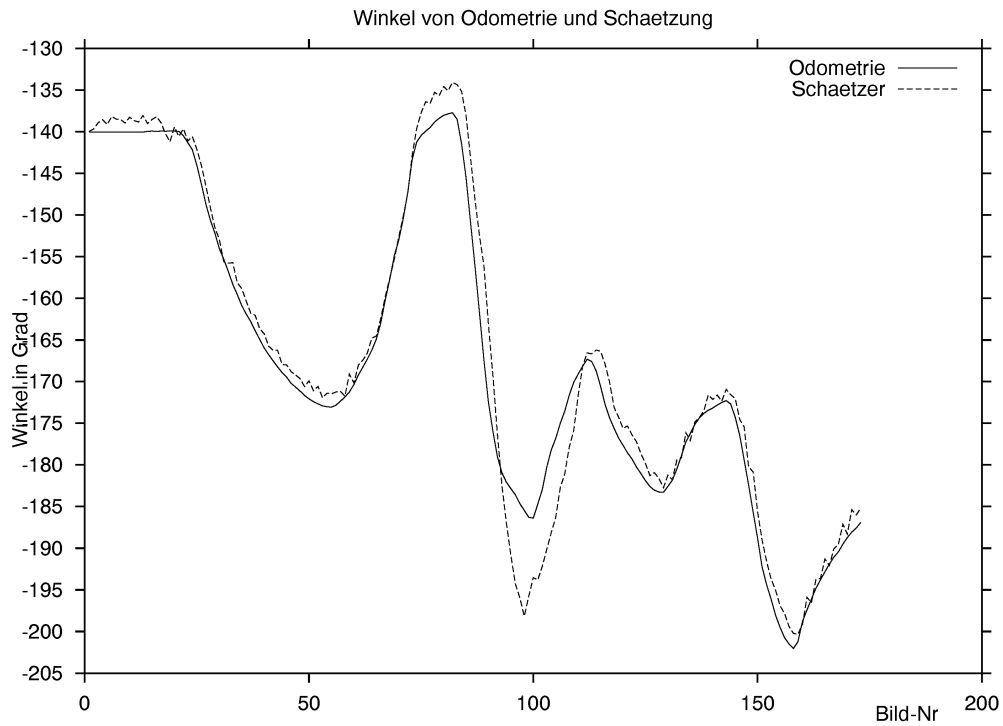


Abb. 59: Verlauf des Fahrzeugwinkels ϕ bei Odometrie und Schätzung

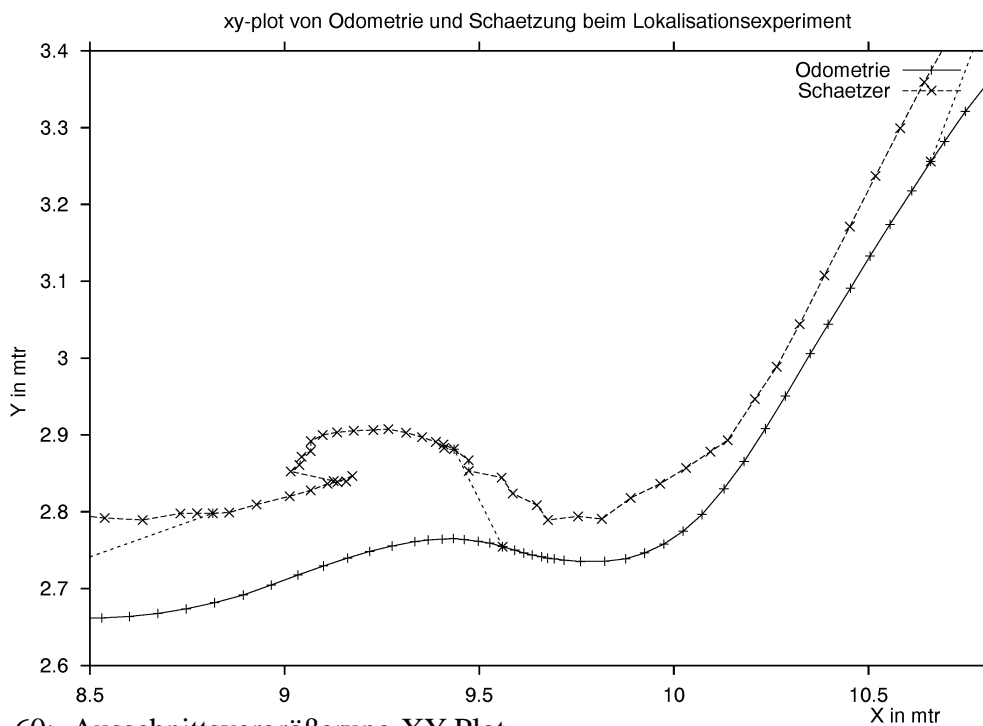


Abb. 60: Ausschnittsvergrößerung XY-Plot

Abb. 61 verdeutlicht gleichzeitig ein zentrales Problem beim Lokalisationsexperiment: Obwohl vom Modellzugriff meist eine große Anzahl von Merkmalen (einige 10) geliefert wird, und auch die Vorverarbeitung eine ausreichende Anzahl von Merkmalen extrahiert, kann der implementierte Bild-Modell-Vergleich nur wenige Korrespondenzen bilden. Abb. 62 zeigt

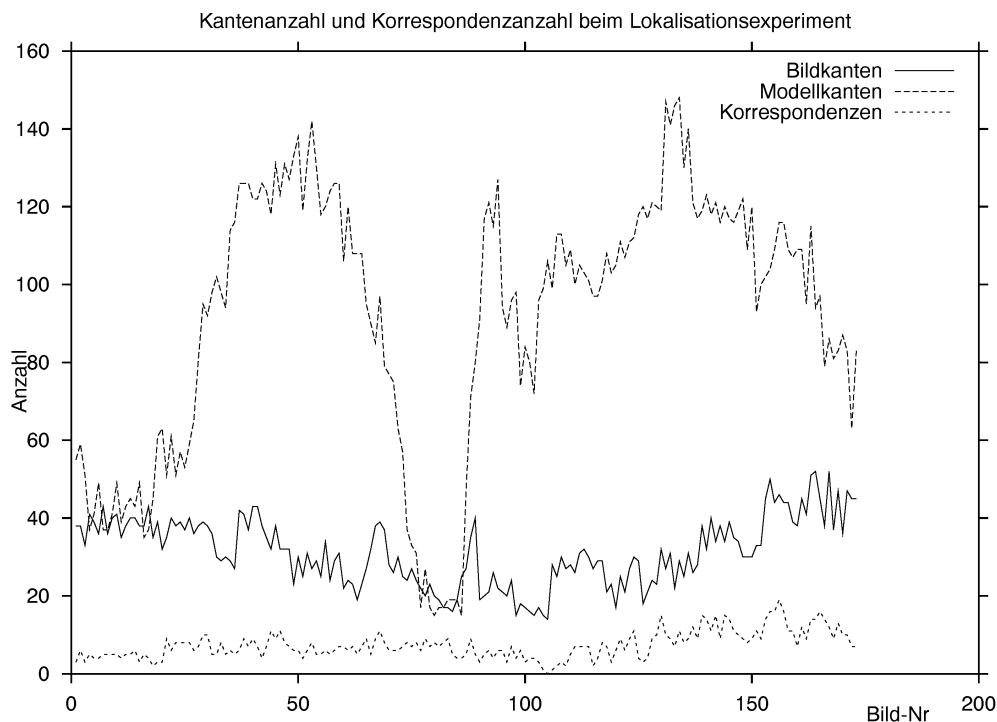


Abb. 61: Anzahl der Merkmale und Korrespondenzen

das extrahierte und das dazugehörige erwartete Linienbild des Rohbildes aus Abb. 57. In Abb. 62-c sind die beiden Linienbilder übereinander gelegt und die zugehörigen Korrespondenzen, unterteilt nach neu etablierten und durch den Bild-Bild-Vergleich konservierten Korrespondenzen, eingezeichnet. Die Korrespondenzen zum hängenden Plakat links oben sind neu etabliert, während die Korrespondenzen an der rechten Seite der Stellwand, an der Vorderkante des Tisches rechts, sowie an der Schrankfuge konserviert sind. Eine der beiden Stellwand-Korrespondenzen ist problematisch, weil die Bildkante verkürzt ist, und dadurch ein falscher Korrekturbeitrag entsteht. Die gemeinsame Zuordnung einer Modellkante auf zwei Bildkanten ist in diesem Fall ohne Folgen.

Die Merkmalsextraktion hat aufgrund des eingestellten Längensfilters korrekterweise nur die längsten Kanten geliefert. Durch die ungünstigen Beleuchtungsverhältnisse und dem an den übersteuerten Stellen fehlenden Kontrast sind viele Kanten unterbrochen worden, wodurch die erforderliche Mindestlänge nicht erreicht wurde.

Die hintere Kante des Tisches wurde aufgrund von Rundungsfehlern im Umgebungsmodell nicht prädiert. Umgekehrt wurden durch den Rundungseffekt eigentlich verdeckte Kanten prädiert, z.B. die Fachböden am rechten Bildrand. Die zylindrische Säule ist nur als Sichthindernis modelliert. Aus dem Rohbild ist ersichtlich, daß aufgrund der Übersteuerung die begrenzenden vertikalen Kanten der Säule nicht extrahiert werden können. Die Sockelleiste des Schrankes konnte aufgrund des zu geringen Kontrasts ebenfalls nicht extrahiert werden.

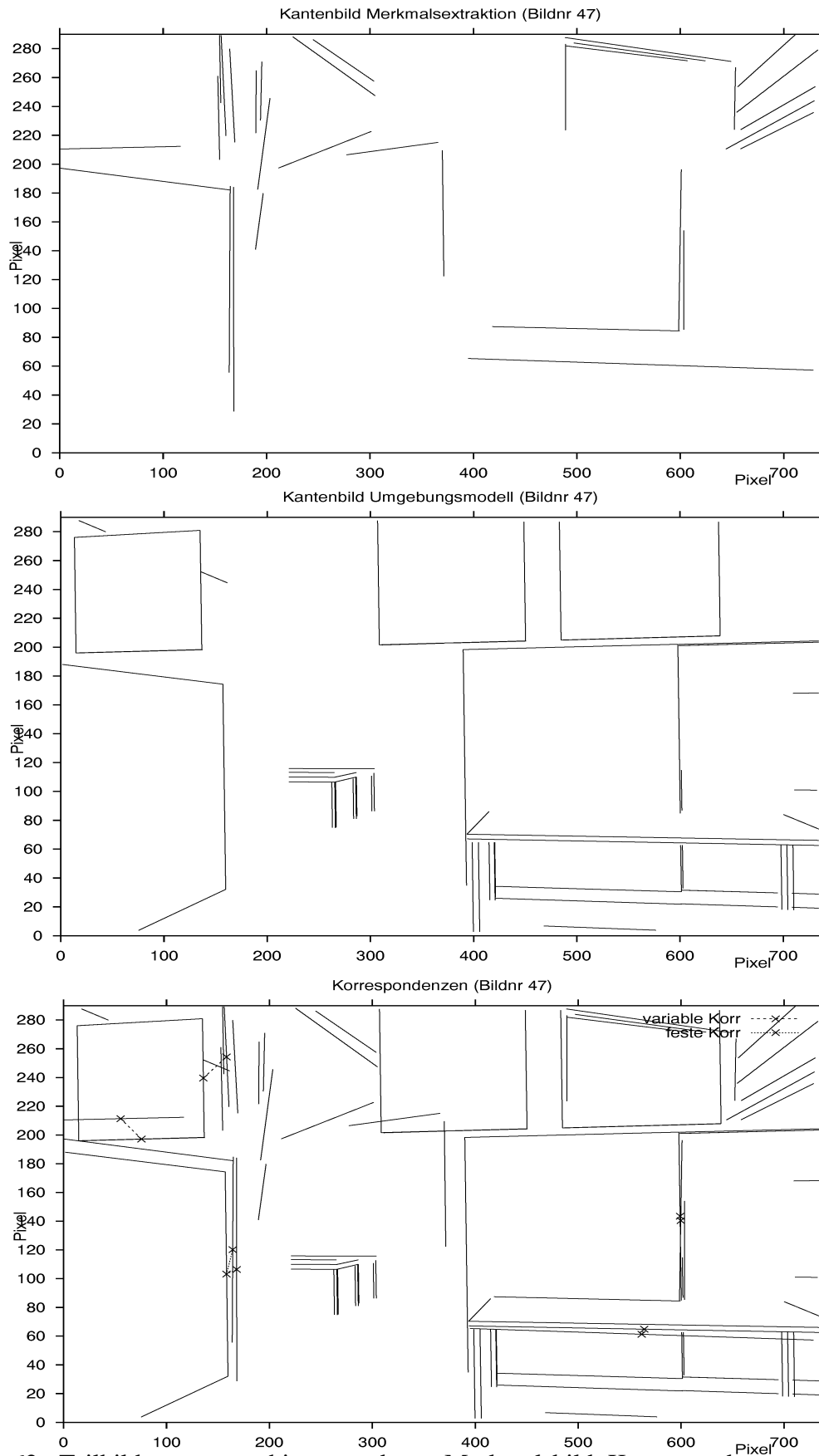


Abb. 62: Teilbilder a-c: extrahiertes und erw. Merkmalsbild, Korrespondenzen

Die Schatten auf den beiden rechts hängenden Plakaten, erkennbar an den schräg verlaufenden Schattenkanten, bewirken auch hier, daß die Unterkante der Plakate nicht erkannt wird.

Die hohe Anzahl von 127 Modellkanten bei 32 Bildkanten für das obige Bild (vgl. Abb. 62) setzt sich aus nicht verwertbaren Detailkanten (Tische in der Mitte links) und aus Mehrfachkanten aufgrund sich berührender Objektteile zusammen.

Die erzeugten Korrespondenzen sind korrekt, die Anzahl muß aber als zu gering beurteilt werden. Grund dafür ist neben der ungünstigen Beleuchtung und der damit verbundenen unzureichenden Extraktion auch die lückenhafte Modellierung gut erkennbarer Merkmale, z.B. der Beleuchtungskörper.

Zur Erhöhung der Korrespondenzen sind zusätzlich zur Konservierung von Bild-Modell-Korrespondenzen über die Zeit weitere Mechanismen einzuführen, die eine größere Ausbeute von Korrespondenzen zulassen. Durch Erniedrigen der Mindestlänge wird eine höhere Anzahl von Kanten geliefert. Wie sich bei der Analyse der Ausführungszeiten der einzelnen Module gezeigt hat (Absatz 5.5.) ist für deren Bearbeitung Zeitreserve vorhanden.

Die hier präsentierten quantitativen Ergebnisse müssen vor dem Hintergrund der Zielsetzung eines jeden Lokalisationsmechanismus gesehen werden: Die mit dem Zustandsschätzer erreichte Genauigkeit genügt, damit Aufgaben eines autonomen mobilen Roboters, wie z.B. Hindernisvermeidung, durchgeführt werden können. Eine Genauigkeit im mm-Bereich ist in der Regel irrelevant. Statt einer hohen absoluten Genauigkeit ist vielmehr eine bestimmte Genauigkeit zu einem zu referenzierenden Objekt wichtig, wobei diese vom Abstand zu diesem Objekt abhängig ist. Beispiele hierfür sind das Passieren eines Hindernisses oder das Greifen von Objekten. Die Genauigkeitsanforderung an die Lokalisation wird demnach vielleicht besser durch die Forderung definiert, sicherzustellen, daß bei Annäherung an ein Objekt nicht fälschlicherweise ein anderes Objekt angenommen wird.

Die Ergebnisse der Lokalisation sind mit den hoch gesetzten Ansprüchen, die Eigenpositionsbestimmung rein videobasiert durchzuführen, nicht voll befriedigend. Insbesondere das gelegentlich beobachtete Konvergieren des Filters auf ein falsches lokales Zustandsminimum erfordert Nachbesserung, weil dann ohne ein manueller Eingriff nicht mehr vernünftig weitergefahren werden kann. Die Ursachen dieses Verhaltens sind in der Korrespondenzfindung zu suchen, die trotz der Nutzung des zeitlichen Kontexts noch zu viele Fehler macht. Dadurch ist eine Voraussetzung des Kalman-Filters, nämlich die gaußsche Verrauschung der Meßwerte, nicht mehr erfüllt.

5.5. Implementierung und Verarbeitungszeiten

Die für das Bewegungstereo in Abschnitt 3.4. erwähnten Module Vorverarbeitung, Lokalisation, 2D-Korrespondenzfindung, 3D-Rekonstruktion, sowie die zusätzlichen Module sind auf unterschiedlichen Plattformen implementiert.

Die *Vorverarbeitung* findet auf einer PC-Einsteckkarte auf der Basis des i860-Prozessors (40MHz) statt [48]. Über ein digitales Kamerainterface werden die Videobilder von der Kamera prozessorgesteuert in den Arbeitsspeicher eingelesen, da die Speicherbandbreite des i860 (64 bit Datenbus, überlappende Speicherzyklen) ausreicht, diese Aufgabe schritthaltend durchzuführen, und DMA auf der verwendeten Karte nicht vorgesehen ist. Nachdem ein Videobild im Speicher steht, wird die Kantenextraktion gemäß Absatz 3.1. durchgeführt. Der i860-Prozessor ist wegen seiner superskalaren Architektur geeignet, mehrere Befehle gleichzeitig auszuführen. Durch den effizienten Algorithmus und durch sorgfältige Codierung der aufwendigen Matrizenoperationen wurde die hohe Grundgeschwindigkeit bei der Konturverfolgung erreicht [67].

Der *2D-Korrespondenzfinder* läuft auf dem Bordrechner des Experimentiersystems MARVIN (PC mit Pentium-133-Prozessor) als ein UNIX-Prozeß. Die Ankopplung an die Vorverarbeitung auf der Einsteckkarte erfolgt nicht über ein Dual-Port-RAM, sondern über eine serielle Verbindung. Die Datenrate von 10 Mbit/sec hat sich aufgrund der starken Datenreduktion als ausreichend erwiesen. Der zunächst unlogisch erscheinende Weg über die aufwendige serielle Ankopplung wurde gewählt, um durch den Einbau von tiefen FIFO-Puffern¹⁵ eine optimale Entkopplung zwischen Extraktion und Korrespondenzfindung zu erreichen. Insbesondere durch die Möglichkeit, den nächsten Extraktionsauftrag an die Vorverarbeitung abzusetzen, bevor von dort die Ergebnisse des laufenden Auftrags zur Verfügung stehen, machen eine nahezu verlustfrei Parallelisierung der Vorverarbeitung möglich.

Die anderen Module (*Lokalisation*, *3D-Rekonstruktion*, usw.) laufen als UNIX-Prozesse auf dem Bordrechner und derzeit noch auf anderen Servern außerhalb des Fahrzeugs. Speziell das Umgebungsmodell *GEM* erfordert zur schritthaltenden Bereitstellung der Modelldaten Rechenleistungen, die an Bord der Plattform nicht zur Verfügung stehen. Die Kommunikation zwischen den Modulen erfolgt mit Hilfe von ortstransparenten Remote-Procedure-Calls [96],[9]. Zur Verschränkung der Aufrufe wurden die Standard-Aufrufe zu asynchronen Remote-Procedure-Calls erweitert [40]. Zentrales Element für die Entkopplung sind auch hier Fifo-Strukturen zur Pufferung der Kommunikationspakete. Damit ist es möglich, gleichzeitig mehrere Anfragen oder Transfers an unterschiedliche Module abzusetzen, ohne auf die Antwort warten zu müssen.

¹⁵ 2 KByte Hardware-Fifo für beide Richtungen, und zusätzlich 128 KByte interrupt-gesteuerter Software-Fifo im Unix-Device-Treiber für den Empfang der Vorverarbeitungs-Ergebnisse

Vorteil der Verteilung der Prozesse auf unterschiedliche Prozessoren ist die Möglichkeit der Parallelverarbeitung und des Pipelining. Abb. 63 zeigt ein Zeitdiagramm für die exemplari-

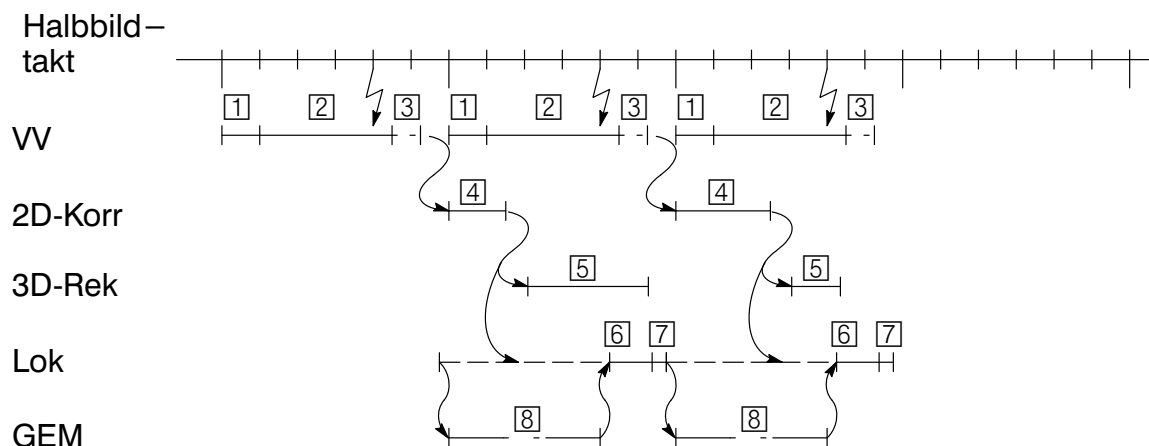


Abb. 63: Zeitdiagramm und Pipelining

sche Bearbeitung einer Bildfolge. Die feste Zykluszeit ist auf 120 msec festgelegt. Diese wird primär durch die Dauer der Vorverarbeitung bestimmt, und muß aufgrund von Einschränkungen bei der verwendeten Kamera ein Vielfaches von 40 msec betragen. Innerhalb eines Zyklus wird das Bild eingelesen [1], die Merkmale extrahiert [2], und bei Erreichen der Zeitschranke die Merkmalsextraktion abgebrochen. Die bis dahin extrahierten Merkmale werden an den 2D-Korrespondenzfinder verschickt [3].

Die extrahierten Merkmale durchlaufen dort die 2D-Merkmalverfolgung [4] und werden mit Korrespondenzinformation versehen an die Lokalisation [6] [7] weitergereicht. Abgeschlossene Merkmalsketten gehen darüberhinaus an die 3D-Rekonstruktion [5]. In der Lokalisation wird der Bild-Modell-Vergleich [6] und der Kalman-Filter-Algorithmus [7] durchgeführt. Mit dem prädierten Zustand wird das Umgebungsmodell erneut abgefragt [8]. Merkmale des nächsten Bildes, die während der Umgebungsmodell-Abfrage eintreffen, können erst zusammen mit den Umgebungsmodell-Merkmalen weiter verarbeitet werden.

Die Analyse der Ausführungszeiten der einzelnen Module ist der nachfolgenden Tabelle zu entnehmen. Die Summe der Zeiten [1]-[3] ergibt nicht ganz 120 msec. Es fehlt die erforderliche Wartezeit, um auf den Bildanfang des einzulesenden Bildes zu synchronisieren. Da die Ausführungszeiten der anderen Module je nach Szeneninhalte schwanken können, ist die sich ergebende Wartezeit von 14 msec knapp. Soll der Längsfilter auch kürzere Kanten weitergeben, oder enthält die Szene besonders viele Kanten, kann es trotz rechzeitigem Abbruch der Extraktion zu Echtzeitverletzungen kommen. Schuld daran ist die von der Kantenzahl abhängige Nachbearbeitung und die Kommunikation.

Aus der Tabelle geht weiterhin hervor, daß der Zugriff auf das Umgebungsmodell [8] die mit Abstand längste Zeit im Zyklus benötigt. Es gibt keine Möglichkeit der Parallelisierung dieses Zugriffs, da Lokalisation und Umgebungsmodell wechselseitig die Ergebnisse der jeweils anderen Instanz zur Weiterarbeit benötigen.

Funktion	msec	% von 120 msec	
Bildeinzug	~ 18	~ 15	[1]
Merkmalsextraktion vor Abbruch	~ 60.0	~ 50	[2]
Merkmalsextr. nach Abbruch	~ 15.0	~ 13	[2]
Komm. zu 2D-Rek (Auftr + Erg)	~5.0 + ~ 8.0	~4 + ~ 7	[3]
2D Korrespondenzfindung	~ 25.0	~ 21	[4]
3D Rekonstruktion	~ 5.0 – 20.0	~ 4 – 17	[5]
Bild-Modell-Vergleich	~35.0	~ 30	[6]
Kalman-Filter (Seq. Inn + Präd.)	~ 5.0	~ 4	[7]
Umgebungsmodell-Abfrage	~ 280.0	~ 230	[8]

Die Zeiten für die Kommunikation zwischen den Prozessen auf den UNIX-Rechnern liegt in der Größenordnung weniger msec und wird vernachlässigt. Die Zeit für die Kommunikation zwischen Vorverarbeitung und 2D-Korrespondenzfinder zur Übermittlung der gefundenen Kanten [3] liegt in der gleichen Größenordnung und ist im Gegensatz zur Netzwerk-Kommunikation linear abhängig von der Anzahl der extrahierten Merkmale, wobei pro Merkmal ca. 0.15 msec (60 Byte bei 400 KByte/sec) gemessen wurden. Diese Zeit kann in erster Näherung nicht vernachlässigt werden, weil sie direkt in die Zykluszeit der Vorverarbeitung eingeht. Fallen die sich als überflüssig erwiesenen Kanteneigenschaften weg, reduziert sich die Kommunikationszeit um etwa 25%.

Durch die Zeitverzögerung durch den Umgebungsmodell-Zugriff darf die Lokalisation nicht bei jedem Bild-Zyklus mit Bildmerkmalen versorgt werden, sondern z.B. nur bei jeden dritten Vorverarbeitungs-Zyklus, also alle 360 msec. In diesem Zyklus beträgt der Anteil des Umgebungsmodellzugriffs immer noch über 75%, die eigentliche Lokalisation 12%, der Rest ist wieder Wartezeit und dient der Synchronisation. Es ergibt sich eine aus Sicht der Regelung schädliche Totzeit von bis zu ca. 500 msec, bis die korrigierte Position zur Verfügung steht.

6. Zusammenfassung und Ausblick

Es wurde ein monokulares Videosystem vorgestellt, das mit Hilfe von differentiellm Bewegungsstereo auf Kantenbasis in der Lage ist, unbekannte Szenen zu kartographieren, in Zusammenarbeit mit weiteren Modulen die vorhandenen Umgebungsmodelle zu verifizieren und die Eigenposition in einer kartographierten Szene zu bestimmen. Die hierfür erforderlichen und im Rahmen dieser Arbeit erstellten Module sind eine schnelle, echtzeitfähige Kantenextraktion auf der Basis der Konturverfolgung, eine robuste Korrespondenzfindung und diverse anwendungsorientierte Funktionsblöcke.

Mit diesem System, das mit einfachsten Standard-HW-Komponenten realisiert wurde, ist es möglich, bei einem mit 40 MHz getaktetem i860 Prozessor als Vorverarbeitung und einem einem mit 133 MHz getaktetem Pentium-Prozessor für die weiteren Verarbeitungsschritte die von der Kamera kommenden Bilder einer realen Umgebung mit einer Bildrate von ca. 8 Hz schritthaltend zu verarbeiten. Eine sorgfältige Kommunikationsstruktur mit großen Puffern ermöglicht ein hohes Maß an Entkopplung der im Pipelinebetrieb parallel arbeitenden Module.

Die geforderte Echtzeiteigenschaft des Algorithmus wird durch ein dreistufiges Verfahren gewährleistet: Zunächst werden über die Prädiktion die aus vorhergehenden Bildern erkannten *langen* Kanten als Startpunkte für die Auswahl der relevanten Bildausschnitte gewählt. Als zweites sorgt ein Filter dafür, daß Kanten, die aufgrund ihrer Länge keinen verwertbaren Beitrag leisten können, nicht weiter betrachtet werden. In der dritten Stufe wird die Auswertung nach Ablauf der zur Verfügung stehenden Zeit abgebrochen. Dabei ist jedoch aufgrund der Prädiktion sichergestellt, daß auch bei sehr komplexen Bildern für die Lokalisation verwertbare Informationen gewonnen worden sind.

Bei der quantitativen Auswertung der Ergebnisse stellte sich heraus, daß eine automatische Online-Kalibrierung während der Mission einer Offline-Kalibrierung nach Möglichkeit vorzuziehen ist. Eine hochpräzise Kalibrierung der Montageparameter der Kamera auf dem Fahrzeug und eine exakte Lokalisation des Fahrzeugs ist erforderlich, da der differentielle Bewegungsstereo-Algorithmus eine lange Transformationskette bis zur Triangulation mit sich bringt.

Das vorgestellte differentielle Bewegungsstereo-Verfahren ist keinesfalls ausgereizt. Die beispielsweise bei der Verwendung der 5x5-Matrizen vorhandene Rauschanfälligkeit und die gelegentlich unzureichende Entdeckungswahrscheinlichkeit kann durch Verwendung eines größeren Einzugsbereichs deutlich reduziert werden. Dabei kann der Mehraufwand durch die ge-

zeigte Äquivalenz der Kirsch–Matrizen mit den verwendeten E–Feld–Matrizen für die größeren Matrizen kompensiert werden. Die Vorverarbeitung ist selbst in der vorliegenden Software-Realisierung enorm schnell, und enthält Vorkehrungen, die das Einhalten der harten Echtzeitbedingungen sicherstellen. Durch eine relativ preiswerte Optimierung der Systemarchitektur, wie z.B. Abbildung der Matrizenoperationen in Hardware, Einlesen des Rohbildes per DMA, Übergabe der extrahierten Merkmale in großen Wechselpuffern, usw., sowie die Portierung auf einen Vorverarbeitungsprozessor des heutigen Stands der Technik entstehen große Reserven, die zur weiteren Erhöhung der Qualität der Vorverarbeitung verwendet werden sollten. Auch die Durchreichung der Kantenattribute an die Korrespondenzfindung ließe sich einsparen.

A. Anhang

A.1. Herleitung der Feldmatrizen

Jedes Pixel bildet einen horizontalen und einen vertikalen Dipol, dessen Ladungen sich aus dem lokalen horizontalen und vertikalen Gradienten bestimmt. (siehe Abb. 64). Die Gradienten

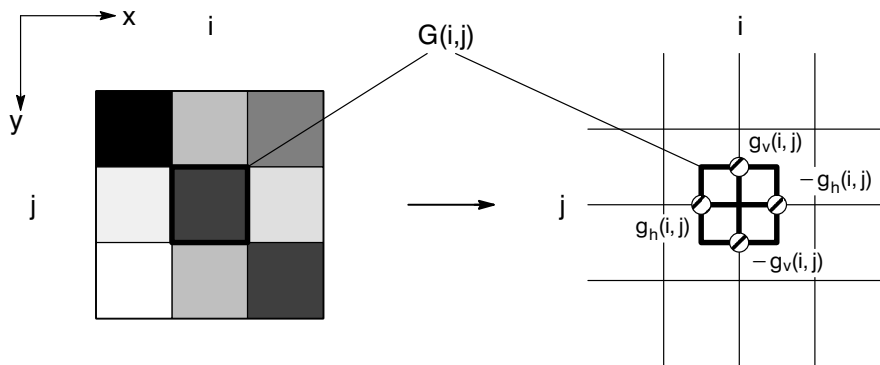


Abb. 64: Transformation Grauwerte in Dipole

ten werden mit dem horizontalen und vertikalen 3x3-Prewittoperator berechnet:

$$q(i,j) = GP \sum_{m=-1}^1 \sum_{n=-1}^1 G(i+m, j+n) * P(m,n) \quad (60)$$

wobei $G(i,j)$ der Grauwert an der Pixelposition (i,j) ist. GP ist eine Skalierungskonstante und $P(m,n)$ ist die ternäre Maske des jeweiligen Prewitt-Operators[4][35] für die horizontale und vertikale Dipol-Ladung.

Das resultierende elektrische Feld, das auf das Teilchen einwirkt, wird aus den Teilfeldern der Dipolladungen in der Umgebung zusammengesetzt. Der Bereich von 5x5 Pixeln ist ein guter Kompromiß zwischen der Genauigkeit der Feldberechnung und dem dafür notwendigen Berechnungsaufwand. Die Dipolladungen befinden sich an den Pixelgrenzen (Abb. 64), und können sich aufheben, oder verstärken (Abb. 65).

Das resultierende elektrostatische Feld wirkt auf die Punktladung e des Teilchens, und beschleunigt es in Richtung der zu verfolgenden Kontur. Die elektrostatische Kraft bestimmt sich aus dem Coulomb-Gesetz:

$$\vec{F}_{Coulomb} = \frac{e}{4\pi\epsilon_0\epsilon_r} \frac{q\vec{r}}{r^3} \approx eC_E \frac{q \begin{pmatrix} x \\ y \end{pmatrix} \sqrt{x^2 + y^2}}{(x^2 + y^2 + \sigma^2)^2} = e\vec{E} \quad (61)$$

$$\frac{1}{4\pi\epsilon_0\epsilon_r} =: C_E \quad (62)$$

Dabei ist C_E eine Konstante aus der Elektrostatik, die Natur- und Materialkonstanten enthält, q ist die felderzeugende Ladung, \vec{r} ist der Ortsvektor zwischen den Ladungen, und σ ist eine Konstante zur Elimination der Polstelle bei $r = 0$. Physikalisch kann man sich σ als Kugelradius um die felderzeugende Punktladung vorstellen, innerhalb dessen das elektrische Feld nicht bis unendlich wächst, sondern konstant gehalten wird.

Die Feldbeiträge jeder $5 \times 5 \times 4$ isolierten Punktladungen unterscheiden sich in der Distanz zwischen Teilchen und Ladung. Ersetzt man $\vec{r} = \begin{pmatrix} x \\ y \end{pmatrix}$, bestimmt sich die horizontale und vertikale Komponente des elektrostatischen Felds an Position (i, j) zu (vgl. Abb. 65):

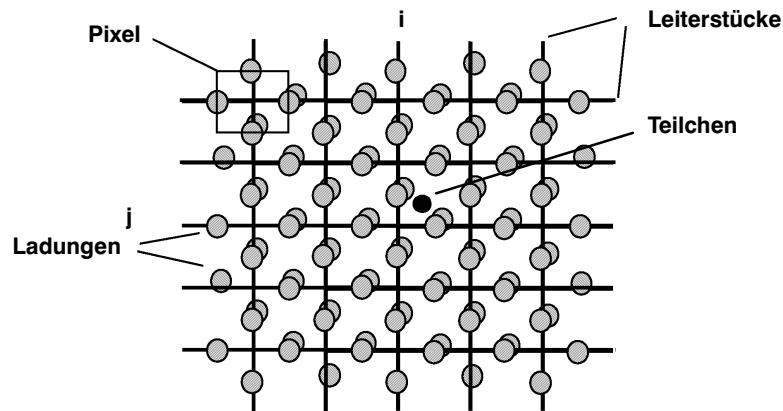


Abb. 65: Felderzeugende Dipolladungen

$$E_h(i, j) = C_E \sum_{m=-2}^2 \sum_{n=-2}^2 (q_h(i + m, j + n) f_{along}(m, n) + q_v(i + m, j + n) f_{cross}(m, n)) \quad (63)$$

$$E_v(i, j) = C_E \sum_{m=-2}^2 \sum_{n=-2}^2 (q_h(i + m, j + n) f_{cross}(m, n) + q_v(i + m, j + n) f_{along}(m, n)) \quad (64)$$

mit

$$f_{along}(p, q) = \frac{(p - 0.5)\sqrt{(p - 0.5)^2 + q^2}}{((p - 0.5)^2 + q^2 + \sigma^2)^2} - \frac{(p + 0.5)\sqrt{(p + 0.5)^2 + q^2}}{((p + 0.5)^2 + q^2 + \sigma^2)^2} \quad (65)$$

$$f_{cross}(p, q) = \frac{p\sqrt{p^2 + (q - 0.5)^2}}{(p^2 + (q - 0.5)^2 + \sigma^2)^2} - \frac{p\sqrt{p^2 + (q + 0.5)^2}}{(p^2 + (q + 0.5)^2 + \sigma^2)^2} \quad (66)$$

Neben der elektrostatischen Kraft hat auch eine magnetische Kraft (Lorentz-Kraft) Einfluß auf die Bahn des Teilchens entlang einer Kontur im Grauwertbild. Das magnetische Feld wird durch einen Stromfluß $I(i, j)$ erzeugt, der zwischen den unerschöpflichen Ladungen jedes Dipols fließt (Abb. 66).

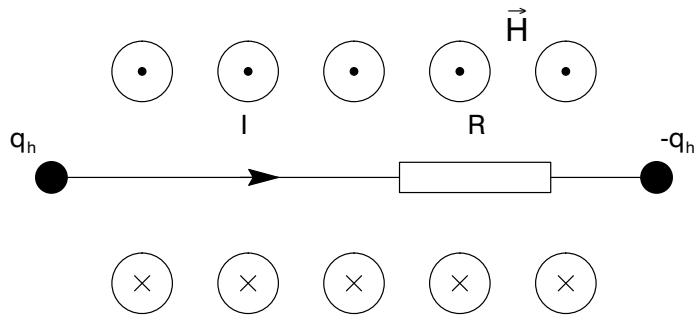


Abb. 66: horizontales Leiterelement mit stromerzeugenden Ladungen

$$I_{h/v}(i, j) = \frac{q_{h/v}(i, j)}{4\pi\epsilon_0\epsilon_r \cdot R} \quad (67)$$

Der Widerstand R jedes Leiters ist dabei eine einstellbare Konstante, mit der der Einfluß des magnetischen und elektrostatischen Feldes auf die Bahn des Teilchens eingestellt werden kann.

Betrachtet man das Biot-Savart-Gesetz:

$$d\vec{H} = \frac{I}{4\pi} \cdot \text{rot}\left(\frac{d\vec{l}}{r}\right); \quad d\vec{l} = \begin{bmatrix} dx \\ 0 \\ 0 \end{bmatrix} \text{ resp. } d\vec{l} = \begin{bmatrix} 0 \\ dy \\ 0 \end{bmatrix} \quad \vec{r} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \quad (68)$$

kann man erkennen, daß die Richtung beider Komponenten des magnetischen Felds, die durch horizontale und vertikale Leiterelemente entstehen, in z-Richtung, also senkrecht zur Bildebene zeigen, und so eine einfache arithmetische Summation ermöglichen.

Jedes horizontale Leiterelement der Länge dx liefert einen Beitrag zum magnetischen Feld an einem Punkt in der Bildebene mit dem Abstandsvektor \vec{r} (vgl. Abb. 67):

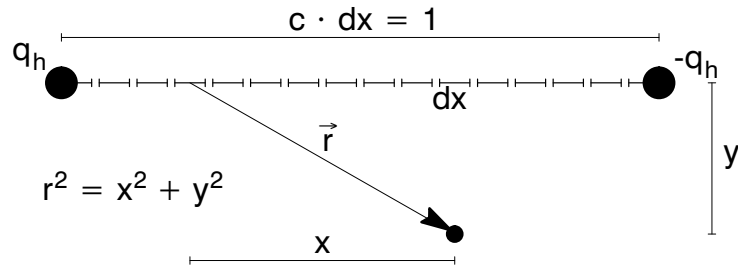


Abb. 67: Beitrag zum Magnetfeld eines horizontalen Leiterelements

$$dH_h(m, n) = \frac{q_h(m, n)}{16\pi^2 \epsilon_0 \epsilon_r R} \cdot \frac{r}{(r^2 + \sigma^2)^2} \cdot y dx \quad (69)$$

σ eliminiert wieder die Polstelle bei $r = 0$.

Der Gesamtbetrag des magnetischen Feldes, der durch einen horizontalen Leiter erzeugt wird, kann durch Integration über dx errechnet werden. Hier reicht allerdings, die Teilbeiträge für die c Elemente des Leiters aufzusummieren, wobei c eine Konstante ist. Der Wert 20 hat sich als hinreichend erwiesen.

$$H_h(m, n) = C_H \cdot q_h(m, n) \sum_{\eta=0}^{c-1} c \cdot \frac{\sqrt{\left(\frac{\eta+0.5}{c} - 0.5 + m\right)^2 + n^2}}{\left(\left(\frac{\eta+0.5}{c} - 0.5 + m\right)^2 + n^2 + \sigma^2\right)^2} \quad (70)$$

Das magnetische Feld an der Stelle i, j in der Bildebene setzt sich aus den Beiträgen aller horizontalen und vertikalen Leiter in der 5x5-Umgebung zusammen:

$$H(i, j) = C_H \cdot \sum_{m=-2}^2 \sum_{n=-2}^2 q_h(i + m, j + n) \cdot h(m, n) + q_v(i + m, j + n) \cdot h(n, m) \quad (71)$$

wobei sich $h(p, q)$ folgendermaßen berechnet:

$$h(p, q) = \sum_{\eta=0}^{c-1} \left[\frac{p}{c} \cdot \frac{\sqrt{p^2 + \left(\frac{\eta+0.5}{c} - 0.5 + q\right)^2}}{\left(p^2 + \left(\frac{\eta+0.5}{c} - 0.5 + q\right)^2 + \sigma^2\right)^2} \right] \quad (72)$$

Nach Substitution von q_v und q_h in Gleichungen (63), (64) und (71) durch den Ausdruck in Gleichung (60) kann man erkennen, daß es möglich ist, je eine 7x7-Matrix für das H , E_h ,

E_v -Feld aufzustellen, die die Koeffizienten für drei Faltungsmatrizen enthalten, mit denen die Feldgrößen an einer bestimmten Pixelposition aus den Dipolen der 5x5-Nachbarschaft berechnet werden können. Die Matrizen werden dabei durch sukzessive Summation aller Koeffizientenbeiträge für jede der 5x5 Positionen errechnet. Experimentell wurde dabei festgestellt, daß die entstehenden 7x7-Matrizen ohne signifikanten Verlust auf 5x5-Matrizen abgeschnitten werden können, um den Rechenaufwand um den Faktor 2 zu reduzieren.

Die entstehenden Feldmatrizen haben interessante Eigenschaften: Die E-Feld-Matrizen besitzen entlang der horizontalen und vertikalen Achse strukturelle Nullelemente. Sie sind achsensymmetrisch bezüglich dieser Achse. Außerdem sind sie symmetrisch bezüglich der dazu senkrechten Achse, jedoch haben die Hälften unterschiedliche Vorzeichen. In jedem Quadranten kann dann sowas wie ein rotationssymmetrisches Gebirge ausgemacht werden. Die H-Matrix ist zirkulär-symmetrisch, allerdings nicht normalisiert, d.h. die Summe der Matrix-Elemente ist ungleich Null, und das unabhängig davon, ob die 7x7-Matrix verwendet wird, oder die abgeschnittene 5x5-Matrix. Einige dieser Eigenschaften lassen sich für eine effiziente Implementierung ausnutzen (strukturelle Nullelemente), wenn keine Subpixelauflösung verlangt ist, wie sie in 3.1.5. vorgeschlagen wird.

A.2. Felder eines synthetischen Bildes

In Abb. 68 finden sich exemplarisch die Felder eines synthetischen Grauwertbildes.

A.3. Bildbeispiele aus den Experimenten

In diesem Kapitel sind die Grauwertbilder und die erzeugten Kantenbilder dargestellt, die den verwendeten Meßreihen entnommen sind.

Abb. 69 zeigt ein Rohbild und das dazugehörige Kantenbild aus der Bildfolge bf_0000. Bei dieser wie bei den anderen Aufnahmen handelt es sich um Halbbilder, bei der jede zweite Videozeile fehlt. Die Bilder sind jedoch vertikal entzerrt dargestellt, wodurch die vertikale Auflösung reduziert ist.

Ein Repräsentant einer anderen Meßreihe (mk6) zeigt Abb. 70.

Ein Beispiel aus der Bildfolge, die für die Lokalisation aufgenommen wurde, findet sich in Abb. 71. Hier wurde auf eine vertikale Entzerrung verzichtet.

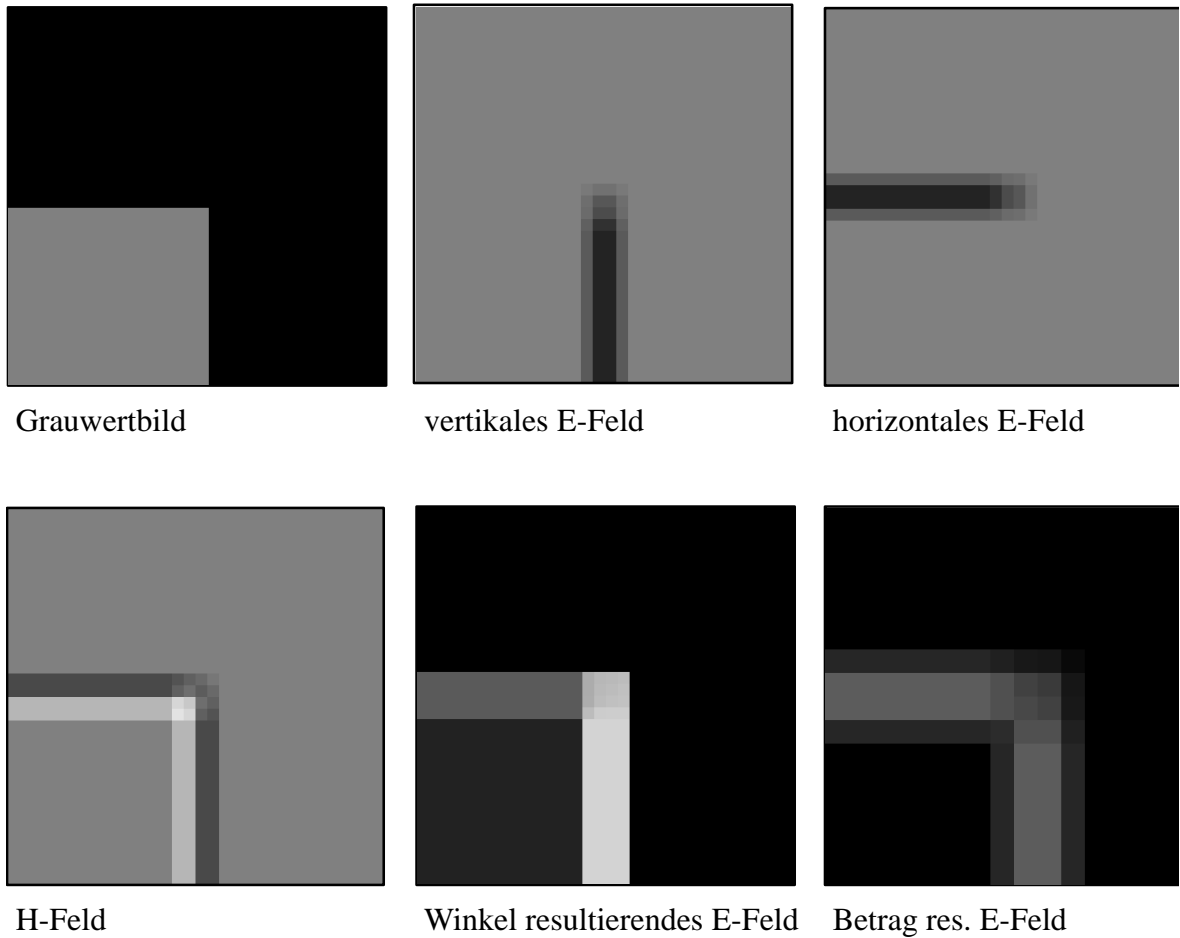


Abb. 68: Felder eines idealen synthetischen Grauwertbildes

A.4. Kalman-Filter

Gegeben ist ein zeitdiskretisiertes, zum Zeitpunkt t_k linearisiertes System, das im Zustandsraum durch einen Satz von Systemgleichungen beschrieben wird (73):

$$\underline{x}(k+1) = A(k) \cdot \underline{x}(k) + B(k) \cdot \underline{u}(k) + \underline{v}(k) \quad (73)$$

Darin ist $\underline{x}(k) \in \mathbb{R}^n$ der Systemzustand, $A(k) \in \mathbb{R}^n \times \mathbb{R}^n$ die Transitionsmatrix, $B(k) \in \mathbb{R}^n \times \mathbb{R}^m$ die Steuerungsmatrix mit zugehörigem Steuervektor $\underline{u}(k) \in \mathbb{R}^m$, und $\underline{v}(k) \in \mathbb{R}^n$ der Systemstörvektor. Die Beobachtung des Systems erfolgt nach der Meßgleichung (74):

$$\underline{y}(k) = C(k) \cdot \underline{x}(k) + \underline{w}(k) \quad (74)$$

Die darin enthaltenen Größen sind der Meßvektor $\underline{y}(k) \in \mathbb{R}^p$, die Meßmatrix $C(k) \in \mathbb{R}^p \times \mathbb{R}^n$ und der Meßstörvektor $\underline{w}(k) \in \mathbb{R}^p$.



Abb. 69: Rohbild und Kantenbild aus der Sequenz bf_0000

Der Systemstörvektor $\underline{v}(k) \in \mathbb{R}^n$ und der Meßstörvektor $\underline{w}(k) \in \mathbb{R}^p$ werden als voneinander unabhängige, gaußverteilte stochastische Prozesse, d.h. als weiße Rauschquellen angenommen. Ihre Erwartungswerte E verschwinden (biasfreies Rauschen) und die Kovarianzmatrizen Q und R sind gegeben zu:

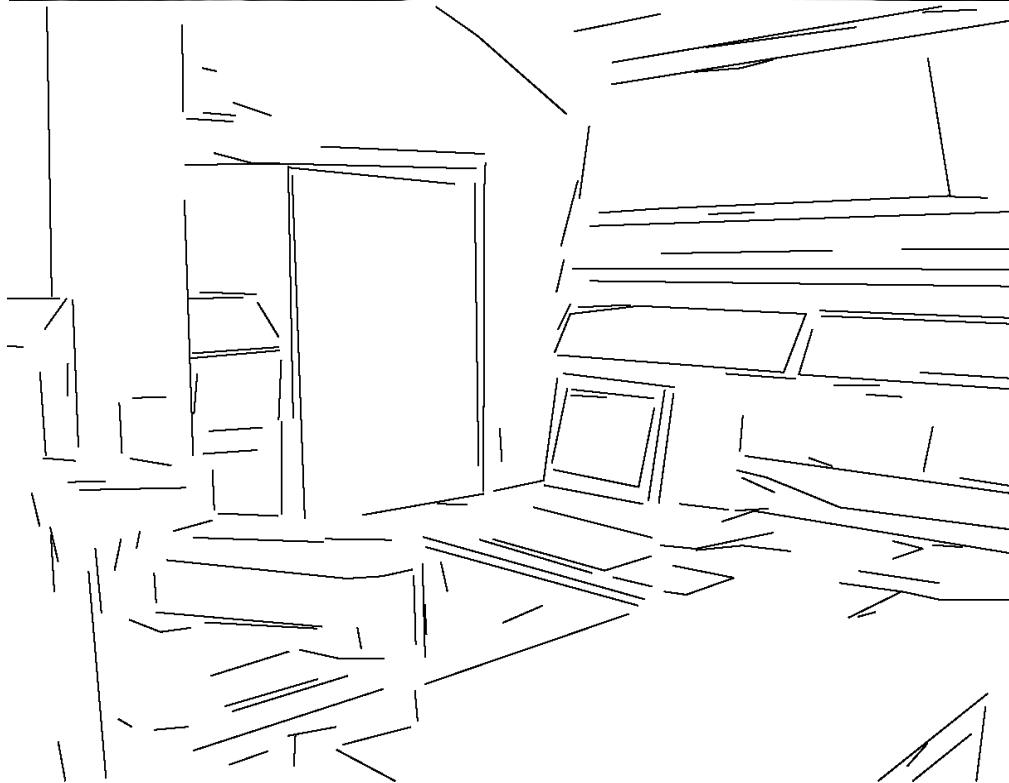


Abb. 70: Rohbild und Kantenbild aus der Sequenz mk6



Abb. 71: Halbbild aus einer Lokalisationsfahrt

$$E\{\underline{v}(k)\} = 0 \qquad Q = E\{\underline{v}(k) \cdot \underline{v}(k)^T\} \qquad (75)$$

$$E\{\underline{w}(k)\} = 0 \qquad R = E\{\underline{w}(k) \cdot \underline{w}(k)^T\} \qquad (76)$$

Zur Schätzung des unzugänglichen Systemzustands $\underline{x}(k)$ wird auf die in der Estimationstheorie übliche Beobachterstruktur zurückgegriffen. Zur Schätzung des unzugänglichen Systemzustands steht nur der Meßvektor $\underline{y}(k)$ zur Verfügung. Die Art der rekursiven Berechnung des Systemzustands aus den Meßwerten und die systemtheoretischen Nebenbedingungen unterscheiden die verschiedenen Beobachter, z.B. den Kalmanfilter, den Luenberger-Beobachter, und andere. In Abb. 72 ist die die Beobachterstruktur des Kalman-Filters dargestellt. Der grau hinterlegte Teil ist das Modell des Systems, der untere Teil beschreibt den Zustandsschätzer.

Jeder Rekursionsschritt des Kalmanfilter-Algorithmus besteht aus einer Prädiktionsphase und einer Innovationsphase. Bei der *Prädiktion* wird aufgrund des aktuellen Zustands der Folgezustand vorhergesagt. Außerdem wird der Meßvektor \underline{y} prädiziert, die Prädiktionsfehler-Kovarianz P^* errechnet, und die Filtermatrix K (Kalman-Gain-Matrix) derart bestimmt, daß die Schätzfehlerkovarianz P in (83) minimal wird:

$$\underline{x}^*(k+1) = A(k) \cdot \hat{\underline{x}}(k) + B(k) \cdot \underline{u}(k) \qquad (77)$$

$$P^*(k+1) = A(k) \cdot P(k) \cdot A(k)^T + Q(k) \qquad (78)$$

$$\underline{y}(k) = g\{\underline{x}^*(k)\} + \underline{w}(k) \qquad (79)$$

$$K(k) = P^*(k)C^T(k)\{C(k)P^*(k)C^T(k) + R\}^{-1} \qquad (80)$$

Der Schätzfehler ist definiert als die Abweichung zwischen wahren Zustand und bestem Schätzwert $\delta\hat{\underline{x}} = \underline{x} - \hat{\underline{x}}$. Er wird ebenfalls als gaußverteilter, erwartungstreuer stochastischer Prozeß angenommen:

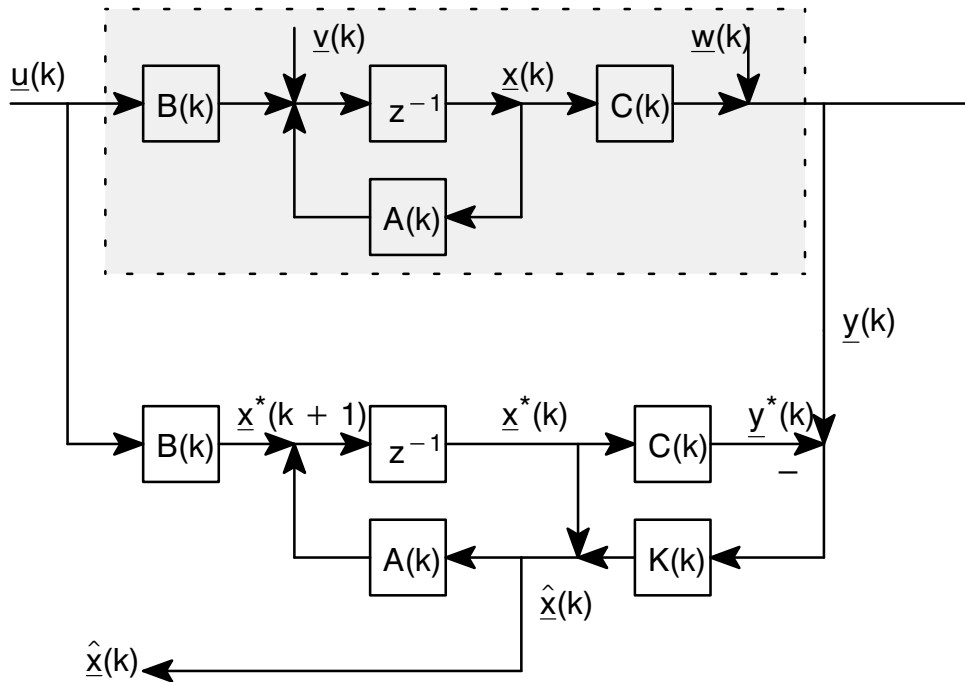


Abb. 72: Signalfluß beim Kalman-Filter

$$P = E\{(\underline{x}(k) - \hat{\underline{x}}(k)) \cdot (\underline{x}(k) - \hat{\underline{x}}(k))^T\} \quad E\{(\underline{x}(k) - \hat{\underline{x}}(k))\} = 0 \quad (81)$$

Aufgrund der Nicht-Linearität des Systemmodells ist für jeden Zeitschritt die Kalman-Matrix neu zu berechnen.

Im *Innovationschritt* wird der prädizierte Schätzwert des Systemzustands durch die Integration von Messungen verbessert. Hierzu wird die Differenz zwischen prädiziertem und realem Meßwert mit der Filtermatrix transformiert und auf den Systemvektor addiert. Man erhält dann den optimalen Schätzwert $\hat{\underline{x}}$ zum Zeitpunkt k .

$$\hat{\underline{x}}(k) = \underline{x}^*(k) + K(k) \cdot [\underline{y}(k) - \underline{y}^*(k)] \quad (82)$$

$$P(k) = P^*(k) - K(k)C(k)P^*(k) \quad \text{oder:} \quad (83)$$

$$P = \{I - KC\}P^*\{I - KC\}^T + KRK^T \quad (84)$$

Für die Berechnung der Kovarianz des Vorhersagefehlers $\delta\hat{\underline{x}} = \underline{x} - \hat{\underline{x}}$ hat sich die in (84) verwendete sog. Joseph-Form als numerisch stabiler erwiesen ([105] S. 77).

Sind zu einem Zeitpunkt keine Meßwerte verfügbar, so findet keine Korrektur der Prädiktion statt. Befand sich das System im eingeschwungenen Zustand und treten zu diesem Zeitpunkt

im Sinne der Modellierung keine Systemstörungen auf, wird das System den Systemzustand selbstständig richtig fortschreiben, dessen Güte jedoch aufgrund von (78) und (80) zunehmend anzweifeln, und in nachfolgenden Schritten verfügbare Messungen stärker gewichten.

A.5. Koordinatensysteme und Nomenklatur

In dieser Arbeit werden folgende Koordinatensysteme verwendet.

- Weltkoordinaten-System w
- Fahrzeugkoordinaten f
- Koordinaten des Schwenk-Neige-Kopfes
- Kamerakoordinaten (lokale Koordinaten) k
- Bildkoordinaten b
- Rechnerkoordinaten r (verzerrt, diskretisiert, verschoben)

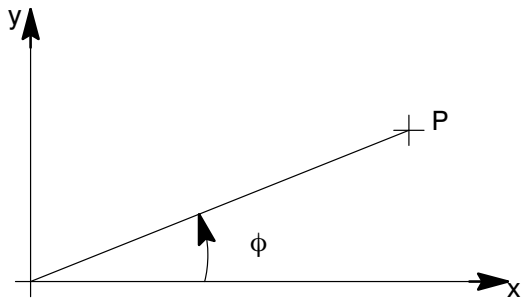
Alle Koordinatensysteme sind Rechtssysteme, die Winkel drehen sich in mathematisch positiver Richtung (Gegenuhrzeiger) um die Drehachse. Die Reihenfolge beim Drehen ist z,y,x . Der Winkel 0 liegt beim Drehen um z (Gieren, Pan) auf der x -Achse, beim Drehen um y (Nicken, Tilt) ebenfalls auf der x -Achse (Horizontale) und beim Drehen um x (Rollen) auf der y -Achse (ebenfalls Horizontale).

Die Kennzeichnung von Koordinatenangaben, in welchem Koordinatensystem sie gemacht werden, erfolgt mit einem hochgestellten Index. Der tiefgestellte Index besagt, um welchen Punkte es sich handelt.

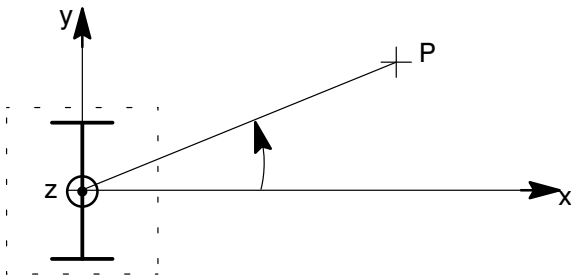
Beispiel: x_p^w ist die x -Koordinate des Punktes P in Weltkoordinaten.

	Index	Ursprung
Welt	w	Szene links-unten
Fahrzeug	f	zwischen Rädern, $z=0$
Neigekopf	z	Kugelgelenk Zebra
Kamera	k	Projektionszentrum
Bild	b	Hauptpunkt

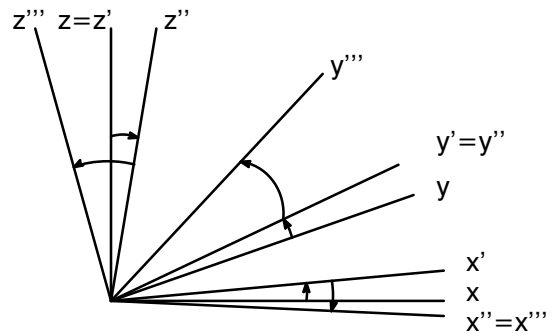
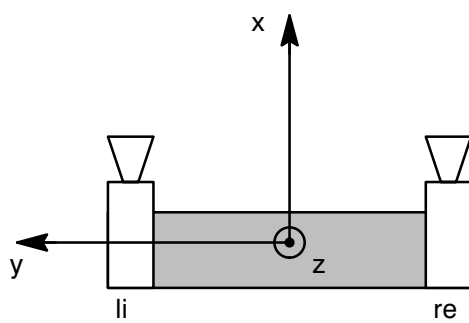
Welt-KS:



Fahrzeug-KS:



Neigekopf-Koordinatensystem:



Kamera-Koordinatensystem:

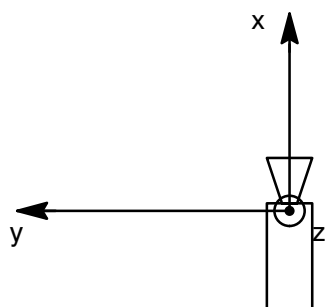
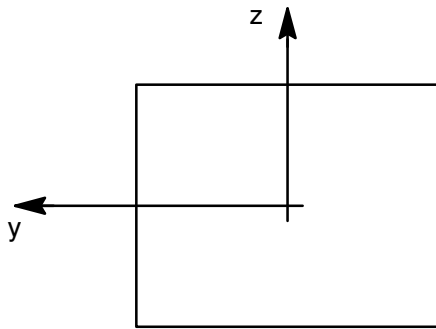
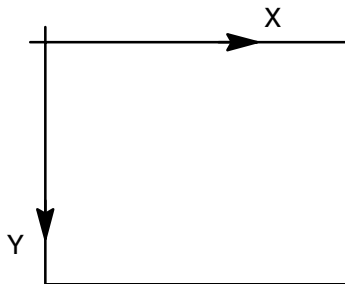


Bild-Koordinatensystem:



Rechner-KS: (nur für Vorverarbeitung relevant)



Koordinaten-Transformationen:

Transformation	Fixe Größen (zu 0 angenommen)	Fixe Größen (kalibrierbar)	Variable Größen (steuerbar)
Welt – Fahrzeug	z, roll, nick		x, y, ϕ
Fahrzeug – SN- Kopf	roll	x,y,z	π (pan), τ (tilt)
SN-Kopf – Kamera	roll, tilt	x,y,z	π (pan)
Kamera – Bild	f=1		

B. Literatur

- [1] P. Anandan: "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion", Int. Conf. on Computer Vision, Juni 1987, pp. 219-230
- [2] Anonym: "Vista software environment for computer vision research ", Univ. of British Columbia, Vancouver, Canada, <http://www.cs.ubc.ca/nest/lci/vista/vista.html>
- [3] Anonym: Firmenschrift "Labmate User Manual", Transition Research Corporation, 15 Great Pasture Road, Danbury, CT 06810, USA, 1991
- [4] Dana H. Ballard; Christopher M. Brown: "Computer Vision", Prentice-Hall, Englewood Cliffs, NJ, 1982
- [5] L. Basano, B. Caprile, E. De Micheli, A. Geminiani, P. Ottonello: "Edge-detection schemes highly suitable for hardware implementation", J. Opt. Soc. Am. A. (JOSA Communications), Vol. 5, Nr. 7, July 1988, pp. 1170-1175
- [6] H. Bässmann, P. Besslich: "Konturorientierte Verfahren in der digitalen Bildverarbeitung", Springer Verlag, Berlin Heidelberg, 1989
- [7] R.E. Bellman, S.E. Dreyfus: "Applied dynamic programming", Princeton University Press, Princeton, NJ, 1962
- [8] R. Blissett, J. Simcox, J. Sherlock: "Vehicle Guidance from Image Tokens", Proc. IFAC Workshop Intelligent Autonomous Vehicles, Southampton, April 1993, pp. 155-160
- [9] J. Bloomer: "Power Programming with RPC", O'Reilly & Associates Inc, Sebastopol CA, 1992
- [10] K. Brammer, G. Siffling: "Kalman-Bucy-Filter", R. Oldenbourg Verlag, München Wien, 1994
- [11] B. Burns; A. Hanson; E. Riseman: "Extracting Straight Lines", IEEE Trans. PAMI-8, Nr. 4, 1986, pp 425-455
- [12] D. Burschka: Persönliche Mitteilung über Zeitmessungen Kantenextraktion mit VISTA im Vergleich zum Konturfolger, L.f.Prozeßrechner, TU München
- [13] D. Burschka, C. Eberst, A. Hauck, N. O. Stöffler: "Hierarchische Umgebungsmodellierung für Lokalisation, Exploration und Objektidentifikation", Tagungsband 11. Fachgespräch "Autonome Mobile Systeme", Karlsruhe, Dez. 1995, pp. 132-141
- [14] J. Canny: "A Computational Approach to Edge Detection", IEEE Trans. PAMI-8, Nr. 6, Nov. 1986, pp 679-698
- [15] J.L. Crowley, P. Stelmaszyk, T. Skordas, P. Puget: "Measurement and Integration of 3D-Structures by Tracking Edge Lines", Int. J. Computer Vision, 8:1, 1992, pp. 29-52
- [16] J.L. Crowley: "The Role of Context in an Integrated Active Vision System", in: D. Vernon (Hrsg): "Computer Vision: Craft, Engineering, and Science", Springer Verlag, Berlin Heidelberg, 1993, pp. 22-42
- [17] J.L. Crowley, P. Stelmaszyk: "Measurement and Integration of 3D-Structures by Tracking Edge Lines", Proc. Europ. Conf. on Computer Vision ECCV 1990, Antibes, France, pp. 269-280
- [18] J.L. Crowley, P. Stelmaszyk, Ch. Discours: "Measuring Image Flow by Tracking Edge-Lines", Proc. Int. Conf. on Computer Vision ICCV, Tampa, Florida, 1988, pp. 658-665
- [19] R. Deriche, O. Faugeras: "Tracking Line Segments", Proc. Europ. Conf. on Computer Vision ECCV 1990, Antibes, pp. 259-268
- [20] E.D. Dickmanns: "Active vision through prediction-error minimization", in: A.K. Sood, H. Wechsler (Eds.): Active Perception and Robot Vision, Springer Verlag, Berlin Heidelberg, 1992, pp. 71-90

- [21] R.O. Duda, P.E. Hart: "Pattern Classification and Scene Analysis", Wiley, New York, 1973
- [22] James G. Dunham: "Optimum Uniform Piecewise Linear Approximation of Planar Curves", IEEE Trans. PAMI-8, Nr. 1, 1986, pp. 67-75
- [23] C. Eberst, D. Burschka, A. Hauck, G. Magin, N. Stöffler: "A System Architecture Supporting Multiple Perception Tasks on an Autonomous Mobile Robot", Proc. 4. Int. Symposium on Intelligent Robot Systems, 1996, pp. 1-8
- [24] W. Eckstein; et. al.: "Bildanalyse-Entwicklungssystem HORUS", Forschungsgruppe Bildverstehen an der Fakultät f. Informatik, Techn. Universität München, <http://www9.informatik.tu-muenchen.de/horus/>
- [25] Elfes, A.: "Using Occupancy Grids for Mobile Robot Perception and Navigation", IEEE Computer, Juni 1989, pp. 46-57
- [26] J. Engelberger: "Robotics in Service", Kogan Page Ltd, London, 1992
- [27] O. Faugeras: "Three-Dimensional Computer Vision. A geometric Viewpoint", MIT Press, Cambridge, Mass., 1993
- [28] H. Freeman: "On the Encoding of Arbitrary Geometric Configurations", IEEE Trans. EC-10, 1961, pp. 260-268
- [29] A. Gelb (Ed.): "Applied Optimal Estimation", MIT-Press, Cambridge, Mass., 1973
- [30] G.H. Golub, C. van Loan: "Matrix Computations", John Hopkins University Press, Baltimore, MD, 1993
- [31] W.E.L. Grimson; E.C. Hildreth: "Comment to: Digital step edges from zero crossings of second directional derivatives", IEEE Trans. PAMI-7, Nr. 1, Jan. 1985, pp. 121-127
- [32] J. Guldner: "Intelligentes hierarchisches Regelungskonzept für autonome, mobile Robotersysteme", Dissertation, Techn. Univ. München. 1995
- [33] R.M. Haralick: "Digital step edges from zero crossings of second directional derivatives", IEEE Trans. PAMI-6, Nr. 1, Jan. 1984, pp. 58-68
- [34] R.M. Haralick: "Authors reply to: Comment to: Digital step edges from zero crossings of second directional derivatives", IEEE Trans. PAMI-7, Nr. 1, Jan. 1985, pp. 127-129
- [35] R. Haralick, L. Shapiro: "Computer and Robot Vision, Volume II: Robot Vision", Addison-Wesley, Reading Mass., 1993
- [36] Richard Hartley: "Projective Reconstruction from Line Correspondences", Proc. Int. Conf. on Computer Vision and Pattern Recognition CVPR 1994, Seattle, Washington, June 1994, pp. 903-907
- [37] A. Hauck, N.O. Stöffler: "A Hierarchical World Model with Sensor- and Task-Specific Features", Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'96) 1996, pp. 1614-1621
- [38] A. Hofmann: "Integration von Werkzeugen zur Kartographierung einer unbekanntenen Umgebung", Diplomarbeit, L.f.Prozeßrechner, TU München, 1994
- [39] Brian Holland: "Echtzeit-Videosystem mit 40 MHz 2D-Convolvern", Design & Elektronik, Nr. 3, 29.01.1991, pp. 66-69
- [40] T. Hopfner: "NoWait-RPC – Eine Benutzerbibliothek für asynchrone Remote Procedure Calls", Diplomarbeit, L.f.Prozeßrechner, TU München, 1996
- [41] T. Hopfner, F. Fischer, G. Färber: "NoWait-RPC: Extending ONC RPC to a fully compatible Message Passing System", Proc. 12. Int. Parallel Processing Symposium, Orlando, Florida, USA, 1998, IEEE Computer Society Press.

- [42] B.K.P. Horn, B.G. Schunk: "Determining Optical Flow", *Artificial Intelligence*, Nr. 17, 1981, pp. 185-203
- [43] D.H. Hubel: "Auge und Gehirn – Neurobiologie des Sehens", Spektrum-der-Wiss.-Verlagsgesellschaft, Heidelberg, 1989
- [44] J. Huber; V. Graefe: "Subpixelauflösung und Bewegungstereo für die räumliche Deutung von Bildfolgen", Tagungsband 6. Fachgespräch "Autonome Mobile Systeme", Karlsruhe, 1990, pp. 173-184
- [45] J. Hüggenberg: "Auf die Konturen kommt es an: Konturtracer L64290 von LSI Logic", *Design & Elektronik*, Nr. 5, 26.2.1991, pp. 98-109
- [46] B. Kahlen: "Ein Verfahren zur Polygonapproximation von Konturverläufen in Grauwert-Videobildern", Diplomarbeit, L.f.Prozeßrechner, TU München, 1993
- [47] P. Kahn; L. Kitchen; E.Riseman: "A Fast Line Finder for Vision-Guided Robot Navigation", *IEEE Trans. PAMI-12*, Nr.11, 1990, pp 1098–1102
- [48] R.D. Klein: "DSM 860 Handbuch", Firmenschrift der Firma DSM Digital Service GmbH, München, 1992
- [49] K.Kraus, "Photogrammetrie Band 1: Grundlagen und Standardverfahren", Ferd. Dümmler Verlag, Bonn, 1986
- [50] Reinhard Klette, Piero Zamperoni: "Handbuch der Operatoren für die Bildbearbeitung", Vieweg Verlag, Wiesbaden, 1992
- [51] Koller, Daniilidis, Nagel: "Modelbased object tracking in monocular image sequences of road traffic scenes", *Int. Journal of Computer Vision*, 10(3):357-281, Juni 1993
- [52] St. Lanser: "Modellbasierte Lokalisation gestützt auf monokulare Videobilder", Dissertation, Techn. Univ. München, 1997
- [53] St. Lanser, Ch. Ziel: "Robuste Kalibrierung von CCD-Sensoren für autonome, mobile Systeme", Tagungsband 11. Fachgespräch "Autonome Mobile Systeme", Karlsruhe, Dez. 1995, pp. 172-181
- [54] J. Lavest, G. Rives, M. Dhome: "Three-Dimensional Reconstruction by Zooming", *IEEE Trans. RA-9*, Nr. 2, April 1993, pp. 196–207
- [55] J. Lavest, G. Rives, M. Dhome: "Modeling an Object of Revolution by Zooming", *IEEE Trans. RA-11*, Nr. 2, April 1995, pp. 267-271
- [56] X. Lebegue, J.K. Aggarwal: "Significant Line Segments for an Indoor Mobile Robot", *IEEE Trans. RA-9*, Nr. 6, April 1993, pp. 801-815
- [57] X. Lebegue, J.K. Aggarwal: "Robotex An Autonomous Mobile Robot for Precise Surveying", *Proc. 3rd Conf. Intelligent Autonomous Systems*, Pittsburgh, Feb. 1993, pp. 460-469
- [58] X. Lebegue, J.K. Aggarwal: "Detecting 3D Parallel Lines for Perceptual Organization", *Proc. ECCV 92*, Santa Margaritha, 1992, pp. 720-724
- [59] C. Lee, F. Catthoor, H. De Man: "An Efficient ASIC Architecture for Real-Time Edge Detection", *IEEE Trans. CS-36*, Nr. 10, Oct. 1989, pp. 1350-1359
- [60] C.H. Lee, A. Joshi: "On Correspondence, Line Tokens And Missing Tokens", *Pattern Recognition*, Vol. 28, Nr. 11, 1995, pp 1751–1764
- [61] R. K. Lenz, D. Fritsch: "Accuracy of videometry with CCD sensors", *ISPRS Journal of Photogrammetry and Remote Sensing*, 45 (1990), pp. 90–110
- [62] R. K. Lenz: "High Accuracy Feature Extraction using Chain Code in Grey-value Images", *IBM T.J. Watson Research Center Tech. Report RC 54811*, Yorktown Heights, NY 10598, March 1987

- [63] Ze-Nian Li: "Stereo Correspondence Based on Line Matching in Hough Space Using Dynamic Programming", IEEE Trans. SMC-24, Nr. 1, Jan. 1994, pp. 144-152
- [64] J. Loomis: "Edge-Finding Algorithm With Subpixel Resolution", Proc. VISION '89, Chicago, Society of Manuf. Engineers Technical Paper MS89-211, pp 8-33 – 8-46
- [65] G. Magin, H.J. Ruß: "Supporting Real-Time Update of an Environment Representation for an Autonomous Mobile Robot", Proc. 6th Euromicro Workshop on Real-Time Systems, Juni 1994, Vaesteraas, pp. 34-39
- [66] G. Magin, H.J. Ruß, D. Burschka, G. Färber: "A dynamic 3D environmental model with real-time access functions for use in autonomous mobile robots", Robotics and Autonomous Systems 14 (1995), pp. 119 – 131
- [67] G. Magin, C. Robl: "A Single Processor Realtime Edge-Line Extraction System for Feature Tracking", Proc. of the IAPR Workshop on Machine Vision Applications 1996, Tokyo, 1996, pp. 422-425
- [68] J. Marquardt: "Merkmalsschätzung im Bildraum", Diplomarbeit, L.f.Prozeßrechner, TU München, 1993
- [69] A. Martelli: "An application of heuristic search methods to edge and contour detection", Comm. ACM 19(2), Feb. 1976, pp 73-83
- [70] James H. McIntosh; Kathleen M. Mutch: "Matching Straight Lines", Computer Vision, Graphics, and Image Processing CVGIP, Nr. 43, 1988, pp. 386-408
- [71] Gerard T. McKee: "What Can Be Fused?", in: J.K. Aggarwal (Ed.): Multisensor Fusion for Computer Vision, Springer Verlag, Berlin Heidelberg, 1993, pp. 71-84
- [72] Meyers Enzyklopädisches Lexikon, Bibliographisches Institut, Mannheim, Wien, Zürich, 1981
- [73] E. De Micheli, B. Caprile, P. Ottonello, V. Torre: "Localization and Noise in Edge Detection", IEEE Trans. PAMI-11, Nr. 8, Oct. 1989, pp. 1106-1117
- [74] H. Moravec: "The Stanford Cart and the CMU Rover", Proc. of the IEEE, Vol. 71, No. 7, 872-884
- [75] J.L. Nagbou: "Entwicklung eines Moduls zur automatischen Erstellung v. 3D-triangulierten Modellen aufgrund von Computer-Tomographie-Information", Diplomarbeit, L.f.Prozeßrechner, TU München, 1995
- [76] V. Nalwa: "A guided tour of computer vision", Addison Wesley, Reading Mass., 1993
- [77] H. Neunzert, B. Wetton: "Pattern Recognition Using Measure Space Metrics", Forschung-Ausbildung-Weiterbildung Bericht Nr. 28; Universität Kaiserslautern, Fachbereich Mathematik
- [78] H. Niemann: "Klassifikation von Mustern", Springer Verlag, Berlin Heidelberg, 1983
- [79] H. Niemann, G.F. Sagerer, St. Schröder, F. Kummert: "ERNEST: A Semantic Network System for Pattern Understanding", IEEE Trans. PAMI-12, Nr. 9, Sept. 1990, pp. 883-905
- [80] K. Prazdny: "On the information in optical flows", Comput. Graphics Img. Process. 22, 1983, pp. 229-259
- [81] A. Pikaz; I. Dinstein: "Optimal Polygonal Approximation of Digital Curvers", Proc. Int. Conf. Pattern Recognition, Jerusalem, 1994, pp. 619-621
- [82] Bernd Radig (Hrsg): "Verarbeiten und Verstehen von Bildern", Oldenbourg Verlag, München Wien, 1993
- [83] H.R. Raubenheimer: "Konturverfolgung und -segmentation mit Kalman-Filtern", Diss. Techn. Univ. Clausthal, 1984

- [84] L.G. Roberts: "Machine Perception of Three-Dimensional Solids", MIT-Press, Massachusetts, 1963
- [85] Chr. Robl: "Echtzeitfähige Merkmalsextraktion aus Videobildfolgen für Online-Experimente mit einem autonomen mobilen Roboter", Diplomarbeit, L.f.Prozeßrechner, TU München, 1995
- [86] A. Rosenfeld, A.C. Kak: "Digital Picture Processing", Academic Press, New York, 1982
- [87] G. Ruske: "Automatische Spracherkennung", R.Oldenbourg Verlag, München Wien, 1988
- [88] A.Ruß: "Sensornahe Umgebungsmodellierung mit echtzeitfähigen Zugriffsfunktionen", Dissertation, Techn. Univ. München, 1994
- [89] A. Ruß, St. Lanser, O. Munkelt, M. Rozmann: "Kontinuierliche Lokalisation mit Video- und Radarsensorik unter Nutzung eines geometrisch-topologischen Umgebungsmodells", Tagungsband 9. Fachgespräche Autonome Mobile Systeme 1993, Okt. 1993, Technische Universität München, S. 313-327
- [90] St. Ruttinger: "Entwurf und Implementierung einer Navigationseinheit zur Umgebungsexploration mit Stereo-CCD", Diplomarbeit, L.f.Prozeßrechner, TU München, 1996
- [91] P. Schaeren: "Real-Time 3-D Scene Acquisition by Monocular Motion Induced Stereo", Dissertation an der ETH-Zürich 1994; Hartung-Gorre Verlag, Konstanz, 1994
- [92] E. Schrüfer: "Numerische Verarbeitung digitaler Signale", Hanser Verlag, 1990
- [93] B. Serra, M. Berthod: "Subpixel Contour Matching Using Continuous Dynamic Programming", Proc. Int. Conf. on Computer Vision and Pattern Recognition CVPR 1994, Seattle, Washington, June 1994, pp. 202-207
- [94] SFB331 (Autorenkollektiv): "Arbeits- und Ergebnisbericht 1992-1994 Sonderforschungsbereich 331: Informationsverarbeitung in autonomen, mobilen Handhabungssystemen", Technische Universität München, 1994
- [95] Y. Shirai: "Three-Dimensional Computer Vision", Springer Verlag, Berlin Heidelberg, 1987
- [96] R. Srinivasan: "RFC 1831 – Remote Procedure Call Protocol Version 2", Sun Microsystems Inc. and Internet Society, 1995
- [97] C. Steger: "Extracting Curvilinear Structures: A Differential Geometric Approach", Proc. 4th European Conference on Computer Vision ECCV 1996, Cambridge, Springer Verlag, Berlin Heidelberg, 1996, pp. 630-641
- [98] N. O. Stöffler, G. Färber: "An Image Processing Board with an MPEG Processor and Additional Confidence Calculation for Fast and Robust Optic Flow Generation in Real Environments", Proc. Int. Conf. on Advanced Robotics (ICAR), 1997, Monterey, California, USA, pp. 845-850
- [99] Roger Y. Tsai; Reimar K. Lenz: "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3D Machine Vision Metrology", IBM T.J. Watson Research Center Tech. Report RC 54867, Yorktown Heights, NY 10598, Oct. 1986
- [100] Sh. Ullman: "The Interpretation of Visual Motion", MIT Press, 1985
- [101] M. Walter: "Videobasierte Lokalisation eines autonomen mobilen Roboters mit einem Kalmanfilter", Diplomarbeit, L.f.Prozeßrechner, TU München, 1996
- [102] K. Wall; P-E. Danielsson: "A Fast Sequential Method for Polygonal Approximation of Digitized Curves", Computer Vision, Graphics, Image Processing, 28(1984), pp. 220-227
- [103] A. Welling: "Sensorkomponenten für Handhabungsprozesse von mobilen Robotern", Tagungsband 7. Fachgespräch "Autonome Mobile Systeme", Karlsruhe, Dez. 1991, pp. 13-26
- [104] Ch.M. Williams: "An efficient algorithm for the piecewise linear approximation of planar curves", Computer Graphics and Image Processing, 8(1978), pp 286-293

- [105] H.J. Wünsche: "Bewegungssteuerung durch Rechnersehen", Springer Verlag, Berlin Heidelberg, 1988
- [106] X. Zhang, H. Dong: "Distributed Image Edge Detection Methods and Performance", Proc. 6. IEEE Symp. On Parallel and Distributed Processing (SPDP 94), Oct. 1994