

Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

der Fakultät für Informatik
der Universität Karlsruhe (Technische Hochschule)

vorgelegte

Dissertation

von

Konstantinos Daniilidis

aus Thessaloniki, Griechenland

Tag der mündlichen Prüfung: 8. Juli 1992

Erster Gutachter: Prof. Dr. H.-H. Nagel

Zweiter Gutachter: Prof. Dr. H. Prautzsch

Kurzfassung

Aus monokularen Bildfolgen kann man Information über die Struktur der Szenenkomponenten und deren Relativbewegung zur Kamera gewinnen. Existierende Algorithmen können – abhängig von der jeweiligen Bewegungs- und Strukturkonfiguration – sehr empfindlich gegen Meßrauschen sein.

Im ersten Teil dieser Arbeit führen wir analytische Nachweise für den Zusammenhang zwischen der Fehlerempfindlichkeit und der relativen Bewegung sowie der Geometrie der abgebildeten Szenenkomponenten bei der Berechnung der Bewegungsparameter aus einem Bildpaar.

Im zweiten Teil der Arbeit wird die zeitliche Kohärenz der Bewegung während einer längeren Bildfolge durch die Einführung von a priori Wissen über Bewegungsmodelle der Szenenkomponenten ausgenutzt. Der Umfang des einzuführenden a priori Wissens und die Erfolgsbedingungen der rekursiven Schätzer werden dann experimentell untersucht.

Danksagung

Ich danke meinem Betreuer Prof. Dr. H.-H. Nagel für die außerordentlich engagierte Betreuung der vorliegenden Arbeit. Die zahlreichen Diskussionen und die Zusammenarbeit bei der mühsamen Durchführung von Berechnungen, die kritische Durchsicht meiner Ergebnisse und sein Bestehen auf expliziten Beweisen und auf klaren Formulierungen haben mich zu einer tieferen Erkenntnis sowie auch zu einer präziseren Darstellung meiner Ergebnisse geführt. Weiterhin bedanke ich mich bei ihm für die Aufnahme der Betreuung meines Aufbaustudiums, das die formelle Voraussetzung für die Durchführung meiner Promotion an der Fakultät für Informatik der Universität Karlsruhe war.

Prof. Dr. H. Prautzsch, mein Korreferent, hat meine Ergebnisse sehr sorgfältig durchgesehen und mittels sehr hilfreicher Korrekturvorschläge und Diskussionen über mathematische Fragen mein Verständnis vertieft.

Ich bedanke mich bei Prof. Dr. Thomas Beth für sein Interesse für meine Arbeit und seine konstruktiven Kommentare.

Prof. Dr. Greg Hager hat mir während seines Aufenthaltes in Karlsruhe sowie mittels späterer Kontakte geholfen, eine tiefere Einsicht in die Schätzungstheorie zu bekommen.

Mein besonderer Dank gilt Prof. Dr. Yiannis Aloimonos, der mich immer wieder aus der Ferne oder während internationaler Treffen motiviert und ermutigt hat, weiter auf dem Bewegungsproblem zu arbeiten.

Die intensive Zusammenarbeit mit meinen Kollegen Sami Atiya und Dieter Koller hat mir sehr geholfen, viele Sachverhalte unter anderen Gesichtspunkten zu sehen und die Kombination von Theorie und Experiment höher einzuschätzen. Die Diskussionen mit Sami Atiya sind für mich immer ein Anlaß gewesen, mein Wissen zur Schätzungstheorie zu revidieren und zu erweitern. Für den Genuß einer solch kreativen Zusammenarbeit mit Sami und Dieter bedanke ich mich herzlich.

Ich bedanke mich bei Volker Gengenbach, der einen Teil seiner wertvollen Zeit zur Aufnahme von Bildfolgen und zur Kamerakalibrierung verbracht hat, die ich zur Durchführung des Experiments an der Roboterszene verwendet habe.

Ich bedanke mich für die hilfreichen Diskussionen und die kritischen und motivierenden Kommentare während internationaler Treffen oder Gastbesuche bei Dr. S. Maybank, Dr. A. Zissermann, Prof. R. Mohr, Dr. F. Bergholm, Dr. T. Lindeberg und Prof. M. Brooks.

Ich bedanke mich bei Henner Kollnig für Diskussionen und für eine für mich sehr hilfreiche Durchsicht dieser Arbeit.

Ich bedanke mich für die Zusammenarbeit mit M. Hirt, T. Thórhallson, M. Tonko und M. Otte und für Diskussionen mit Dr. J. Rieger, Dr. C. Schnörr, K. Rohr und Dr. W. Schwerdtmann.

Die Durchführung dieser Arbeit wäre ohne die finanzielle Unterstützung vom Deutschen Akademischen Austauschdienst (DAAD) nicht möglich gewesen. Für diese fünfjährige Unterstützung sowie für die Hilfsbereitschaft der DAAD-Mitarbeiter bedanke ich mich herzlich.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Bewegungssehen bei Lebewesen und Maschinen	1
1.2	Motivation und Beitrag der Arbeit	4
1.3	Überblick über die Arbeit	6
2	Bildpaar – Problemstellung und Überblick	7
2.1	Eingaben zur 3D-Bewegungs- und Strukturschätzung	7
2.2	3D Bewegungsschätzung aus Punktkorrespondenzen eines Bildpaars . .	10
2.3	3D Bewegungsschätzung aus dem Bewegungsfeld	14
2.4	3D-Bewegungsschätzung aus Geradenkorrespondenzen	25
2.5	Bewegungsschätzung aus den Verschiebungsraten von Geraden	29
2.6	Bewegungsschätzung aus den Verschiebungsraten von koplanaren Geraden	31
3	Analytische Fehlerempfindlichkeitsuntersuchungen	35
3.1	Einführung – Experimentelle vs. analytische Instabilitätsuntersuchungen	35
3.2	Zur Form des zu minimierenden Fehlermaßes	37
3.3	Fehlerempfindlichkeit der Translationsrichtung im kontinuierlichen Fall .	41
3.4	Fehlerempfindlichkeit in der Schätzung allgemeiner Bewegungen einer Ebene	51
3.5	Fehlerempfindlichkeit in der Ermittlung der Struktur	63
3.6	Gefährliche Flächen und Instabilität	66
4	Bewegungsschätzung aus längeren Bildfolgen	75
4.1	Überblick über Ansätze zur 3D-Bewegungsschätzung aus längeren Bildfolgen	75
4.2	Zur rekursiven Zustandsschätzung	82
4.3	Rekursive Schätzung der relativen Bewegung einer Ebene	97
5	Zusammenfassung und Ausblick	121
A	Mathematische Notation	123
B	Sphärische Projektion	127
C	Geradenrepräsentationen	128
C.1	Repräsentation einer Geraden im Raum	128
C.2	Verschiebungsraten einer Geraden in der Bildebene	130

C.3 Geradentransformationen und Dualzahlen	130
D Berechnung der Jacobischen Matrizen	133
E Umrechnungen zwischen Kamera- und Roboterkoordinatensystem	135
Literaturverzeichnis	138

Abbildungsverzeichnis

1.1	Interpretationszyklus zur Auswertung von Einzelbildern [Nagel 87b] unter Berücksichtigung der Änderungsvorschläge nach [Prautzsch 92] . . .	3
2.1	Geometrie des Problems der relativen Orientierung	11
2.2	Geometrie des kontinuierlichen Falls nach [Longuet-Higgins & Prazdny 80].	15
2.3	Geradenkorrespondenzen in einem Bildpaar	26
2.4	Gerade in einer Ebene	32
3.1	Geometrische Interpretation des Fehlermaßes d'	38
3.2	Gemeinsame Normale der Sichtstrahlen	40
3.3	Stabile (a), und instabile (b,c) Konfigurationen für den kleinsten Eigenwert	42
3.4	Modell des effektiven Gesichtsfelds	43
3.5	Variierende Translationsrichtung in der XZ Ebene	47
3.6	Der kleinste Eigenwert $\lambda_{min}(S)$ als Funktion von ψ und χ	59
3.7	Der Winkel ϕ_{min} als Funktion von ψ und χ	60
3.8	Der Winkel ϕ_{min} als Funktion von ψ und χ	62
3.9	Der arctan der Fehlervarianz von N_x (oben) und N_y (unten) als Funktion von ψ und χ	64
3.10	Geometrische Darstellung der Mehrdeutigkeit aus [Maybank 90b].	67
3.11	Die Konstruktion einer gefährlichen Fläche	67
3.12	Das einschalige Hyperboloid	68
4.1	Allgemeines Schema der rekursiven Filterung.	84
4.2	Einfaches Beispiel der Aufnahme einer geradlinigen Bewegung eines Teilchens	85
4.3	Nichtnormale Wahrscheinlichkeitsdichte nach einem Aktualisierungsschritt	85
4.4	Wahrscheinlichkeitsdichte nach der Stereotriangulation	86
4.5	A posteriori Wahrscheinlichkeitsdichten beim Aktualisierungsschritt vier verschiedener Filter	92
4.6	Verzerrung von EKF vs. MGSO	94
4.7	Fehlerkovarianz vs. tatsächlicher Fehler bei IEKF	95
4.8	Fehlerkovarianz vs. tatsächlicher Fehler bei EKF	96
4.9	Bewegung einer an einem Greifer befestigten Kamera	99
4.10	Der Roboter manipulator am FhG-IITB (aus [Müller 92])	103
4.11	Aufnahmen zu Zeitpunkten t_0, t_5, t_{10} und t_{15} (links) und die extrahierten Geradensegmente (rechts).	105
4.12	Die Felder der Verschiebungsraten zu Zeitpunkten t_0, t_5, t_{10} und t_{15}	106
4.13	Veranschaulichung des Azimutwinkels	107

4.14	Schätzwerte für den Azimut- und Polarwinkel der Translationsgeschwindigkeit unter der Annahme einer rein translatorischen Bewegung	108
4.15	Schätzwerte für den Azimutwinkel (a) und den Polarwinkel (b) der Ebenennormalen und der inversen Kollisionszeit (c) unter der Annahme einer rein translatorischen Bewegung.	109
4.16	Schätzwerte für den Azimutwinkel (a) und den Polarwinkel (b) der Translationsgeschwindigkeit.	111
4.17	Schätzwerte für den Azimutwinkel (a), den Polarwinkel (b) der Ebenennormalen und der inversen Kollisionszeit (c).	112
4.18	Schätzwerte für die drei Komponenten der Winkelgeschwindigkeit	113
4.19	Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b) und absoluter Fehler in ω_x (c) bei variierender Größe des effektiven Gesichtsfelds	115
4.20	Logarithmus des relativen Fehlers in der vom IEKF geschätzten inversen Kollisionszeit bei variierender Größe des effektiven Gesichtsfelds	116
4.21	Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b), absoluter Fehler in ω_x (c) bei variierender Translationsrichtung	117
4.22	Logarithmus des relativen Fehlers in der vom IEKF geschätzten inversen Kollisionszeit bei variierender Translationsrichtung	118
4.23	Fehlerwinkel in der von IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b), absoluter Fehler in ω_y (c) bei variierender Richtung der Ebenennormalen.	119
C.1	Parametrisierung einer Geraden im Raum und ihrer Projektion	128
E.1	Koordinatensysteme an einem Roboterarm	135

Tabellenverzeichnis

4.1	Ansätze zur Bewegungsschätzung aus längeren Bildfolgen	77
4.2	Vergleich der Filter bei einem Aktualisierungsschritt	91

1 Einleitung

1.1 Bewegungssehen bei Lebewesen und Maschinen

Das visuelle System des Menschen ist in der Lage, aus auf den Retinas abgebildeten zweidimensionalen Bildern dreidimensionale Beschreibungen der Umgebung und der Bewegungsvorgänge zu ermitteln, die der Mensch in Aktionen umsetzen kann. Untersuchungen zur visuellen Wahrnehmung von Lebewesen aus den Gebieten der Wahrnehmungspsychologie und Neurophysiologie weisen darauf hin, daß zu einer dreidimensionalen Interpretation die Funktionen der *Stereoskopie* und des *Bewegungssehens* beitragen. Die Auswirkung der Stereoskopie kann jeder Mensch durch Verwendung eines Stereoskops erleben. Nicht so naheliegend ist die Erfassung dreidimensionaler Information durch Bewegung, obwohl jeder von uns beim Autofahren sehr weit entfernte Objekte als nahezu ortsfest wahrnimmt, während Objekte am Straßenrand vorbeirasen. Weiterhin gibt es Menschen, die unter einer kongenitalen Störung der Stereoskopie leiden und trotzdem ohne hohem Aufwand Auto fahren oder Tennis spielen. Lebewesen, bei denen die seitliche Augenposition nur eine geringe Überlappung des rechten und des linken Gesichtsfeldes ermöglicht, können trotzdem erfolgreich navigieren und ihre Beute fangen.

Wahrnehmungspsychologen und Neurophysiologen haben durch zahlreiche Experimente die Funktionen des Bewegungssehens analysiert (siehe die Artikelsammlung [Ritter 87] und den Überblick von [Nakayama 85]). Neben der Erfassung von nur relativen Abständen zu Szenenkomponenten aus monokularen Bildfolgen (nach¹ [Wallach & O'Connell 53] *kinetischer Tiefeneffekt* genannt) ist der Mensch in der Lage, die bis zur Kollision mit anderen Szenenkomponenten noch verbleibende Zeit zu schätzen. Es ist faszinierend zu beobachten, wenn man die Schwierigkeit der technischen Verwirklichung einer solchen Funktion kennt, wie niedrige Lebewesen, z.B. Insekten, in einem komplexen Raum Zusammenstöße vermeiden und erfolgreich bei beliebiger Bewegung navigieren. Weitere nachgewiesene Funktionen des biologischen Bewegungssehens sind die Wahrnehmung der Eigenbewegung, die in Zusammenarbeit mit dem vestibularen Navigationssystem der Lebewesen erfolgt, aber auch die Segmentierung der zweidimensionalen Bildinformation sowie die darauf aufbauende Erkennung von Objekten, die bei stationären Bildern schwer oder unmöglich ist.

Diese Vielfalt von Funktionen konnte erst unter einem allgemeinen Aspekt des Sehens als informationsverarbeitender Prozeß [Gibson 66] und dann durch die Einführung von

¹Sofern Ansätze von denselben Autoren sowohl in Konferenzbeiträgen als auch in Fachzeitschriften publiziert worden sind, werden wegen der leichteren Zugänglichkeit die Beiträge zu Fachzeitschriften zitiert. Dadurch können sich Zeitverschiebungen gegenüber der erstmaligen Publikation einer Überlegung ergeben.

Berechnungstheorien [Hay 66; Koenderink & van Doorn 76; Longuet-Higgins & Prazdny 80] in die Wahrnehmungspsychologie gründlich untersucht werden. Zum Verständnis des Sehens muß man nach [Gibson 66] untersuchen, welche sensorische Information in der Umwelt der Lebewesen benötigt wird, um ihnen das Handeln zu ermöglichen. Sehen bedeutet “aktives Hinschauen und aktives Durchmustern” der Umgebung und nicht “passiver Empfang” von physikalischen Reizen.

Eine Methodologie zur Untersuchung des Bewegungssehens und anderer Deutungshinweise, die sich an den Gibsonschen Ansatz anschließt, wurde von [Marr 82] entwickelt und ist dem Forschungsweg verwandt, den Entwickler von maschinellen Sichtsystemen begehen. Marr hat die Auswertung der visuellen Reize in drei Ebenen unterteilt. Die erste Ebene bildet sich aus der mathematischen Formulierung und Lösung des Problems. Als einfaches Beispiel könnte man sich die Berechnung der Lage eines Punktes im Raum aus den Projektionen des Punktes auf zwei Bildebenen denken, deren relative Lage und Orientierung bekannt ist. Die zweite Ebene bei Marr ist die algorithmische Entwicklung einer Lösung, so daß eine Maschine diese Lösung in konkreten Schritten mit den geeigneten Datenstrukturen ausführen kann. Ein solcher Algorithmus soll z.B. unter zwei Punktfigurationen in je einer Bildebene die Punkte einander zuordnen, die dem gleichen Punkt im Raum entsprechen. Die dritte Ebene ist die Ebene der Implementierung auf einer ausgewählten Rechnerarchitektur.

Das Ziel des *Maschinensehens* ist die Entwicklung von technischen Systemen, die in der Lage sind, aus einem oder mehreren, stationären oder sich bewegendem Sensoren diejenige Information über die Form und die Bewegung der abgebildeten Objekte relativ zueinander und zum Sensorsystem zu erfassen, die für die Ausführung spezifischer Aufträge notwendig ist.

Zur Klärung der zu lösenden Aufgaben gehen wir zunächst auf die Einsatzbereiche des Maschinensehens ein. Bildgebende Sensoren sollen zur Interaktion zwischen autonom mobilen Systemen mit einer sich verändernden Umgebung beitragen. In Innenraumumgebungen sollen Robotersysteme in der Lage sein, von einem Raum in einen anderen zu fahren. Dazu muß ein Sichtsystem dem Roboter helfen, Abweichungen von seiner geplanten Fahrbahn zu erkennen sowie zu korrigieren und Hindernisse zu detektieren. Weiterhin soll die Form der Hindernisse sowie auch ihre Relativbewegung zum Roboter geschätzt werden, damit der Roboter sie umfahren kann. Der Roboter muß an seinem Ziel Objekte mit seiner Hand greifen, um sie zu transportieren oder um eine Montageaufgabe durchzuführen. Eine auf dem Roboterarm befestigte Kamera muß zur Regelung der Trajektorie sowie auch zur Rekonstruktion der zu greifenden Objekte dienen. Details über diese Objekte können a priori bekannt sein – z.B. Werkstücke, die montiert werden müssen. Andererseits kann sich das a priori Wissen nur auf die Tatsache beschränken, daß die Objekte polyedrisch sind – z.B. Postpakete, die auf einer Palette gestapelt werden müssen.

In Außenweltszenen sollen auf Kraftfahrzeugen eingebaute bildgebende Sensoren die Verfolgung der Fahrbahn ermöglichen und eine Beschreibung der Form und Bewegung anderer am Verkehr teilnehmender Fahrzeuge liefern. Entsprechend variiert das zu nutzende a priori Wissen von der Kenntnis, wo ein Fahrzeug auf einer geometrisch modellierbaren Autobahn fährt, bis zur Kenntnis, daß ein Fahrzeug zwischen unregelmäßig geparkten Fahrzeugen durchfahren muß. Automatische Pilotsysteme sollen durch Auswertung der von einer Kamera im Cockpit von Flugzeugen aufgenommenen Bildfolgen

die Position und Orientierung der Landebahn bestimmen. Während das Sichtsystem bei den zuletzt genannten Anwendungen den Fahrer oder Piloten mehr oder weniger unterstützen soll, ist der Einsatz von mobilen Systemen von enormer Bedeutung in Umgebungen, die dem Menschen schwer oder nicht zugänglich sind. Montageoperationen auf dem Meeresgrund, Andockungsmanöver im Weltall und Erforschung des Mars sind Anwendungen, bei denen ein Robotersystem autonom und mit einem Minimum an speziellem a priori Wissen agieren soll.

Eine weitere Anwendung ist die Überwachung von Verkehrsszenen und die automatische Verkehrszählung. Fahrzeuge müssen in Bildfolgen verfolgt werden, damit ihre Bewegungsrichtung qualitativ beschrieben werden kann, und ihre Form muß rekonstruiert werden, damit sie nach dem Fahrzeugtyp klassifiziert werden können.

Zu einem besseren Verständnis der oben beschriebenen Anwendungen und der Motivation unserer Arbeit werden wir ein Schema der Vorgehensweise des Maschinensehens erläutern, wie dies zuerst von [Kanade 80] vorgeschlagen und dann von [Nagel 87b] erweitert wurde. Es besteht aus dem in Abb. 1.1 aufgezeichneten Interpretationszyklus. Aus einem oder zwei digitisierten Grauwertbildern extrahiert man Bildbereichshinweise

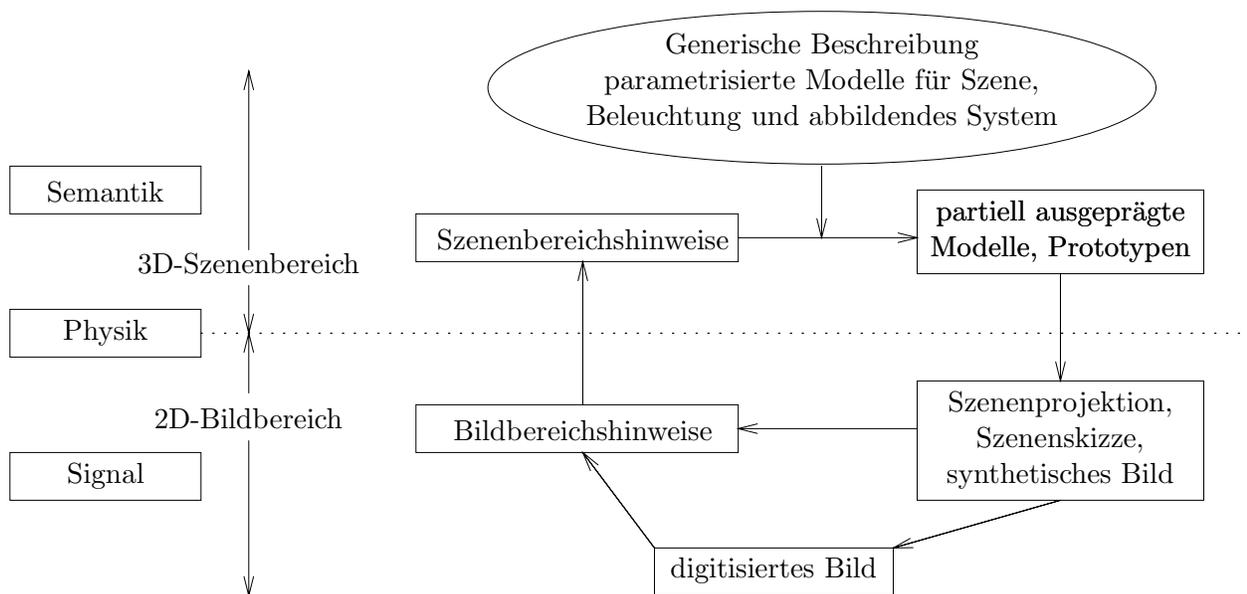


Abbildung 1.1: Interpretationszyklus zur Auswertung von Einzelbildern [Nagel 87b] unter Berücksichtigung der Änderungsvorschläge nach [Prautzsch 92]

(charakteristische Grauwertverläufe), die man mit Hilfe des Abbildungsvorgangs auf Szenenbereichshinweise (Form und relative Lage der Szenenkomponenten) überführt. Zu diesem Schritt hat [Marr 82] auf die Kooperation von verschiedenen Verarbeitungsmodulen hingewiesen, von denen jedes einem Deutungshinweis wie Bewegung, Stereo, Kontur, Abschattierung oder Farbe entspricht. Unter Ausnutzung der generischen Beschreibung der Szenenkomponenten kann man Hypothesen über Modelle der Szenenkomponenten ausprägen, deren Projektion man mit dem Grauwertbild selbst oder mit den Bildbereichshinweisen vergleicht. Im Fall einer zeitlichen Bildfolge kann man die ausgeprägten Modelle im nächsten Zeitpunkt direkt aus den jetzigen Modellen prädzie-

ren, sofern man zusätzlich über geeignete generische Bewegungsbeschreibungen verfügt. Der Interpretationszyklus fängt dann ab einer fortgeschrittenen Stufe an.

Die dreißigjährige Geschichte des Maschinensehens zeigt einerseits eine hohe Zahl an theoretischen Untersuchungen von allgemeinen Problemstellungen, andererseits aber einen Mangel an Entwicklungen von technischen Systemen, die zu einer Erfassung der Umgebungszustände sowie einer angemessenen Reaktion befähigt sind (siehe die von [Jain & Binford 91] gestartete Diskussion).

1.2 Motivation und Beitrag der Arbeit

Wir beschreiben zunächst kurz, in welchen Problemlösungsrahmen unsere Arbeit eingebettet wird. Die allgemeinste Problemstellung in der Bildfolgenauswertung ist die Berechnung der relativen Bewegungen zwischen einer Kamera und verschiedenen Szenenkomponenten sowie auch der dreidimensionalen Struktur dieser Komponenten aus *monokularen* Bildfolgen. Aus dieser Problemstellung ergibt sich eine Vielfalt von spezielleren Problemstellungen entsprechend der Verarbeitungsstufe und der Verwendung von a priori Wissen über Vorgänge und Objekte in der Umgebung oder von zusätzlicher Sensorinformation.

Unsere Arbeit baut auf den in einer vorgeschalteten Verarbeitungsstufe ermittelten zeitlichen Zuordnungen von Bildbereichshinweisen und ihrer Segmentierung in Gruppen, die den Abbildungen von starren Komponenten der Szene entsprechen, auf. Sind diese zeitlichen Zuordnungen gegeben, so kann die relative Bewegung und Struktur einer Szenenkomponente ermittelt werden. Dieses Problem taucht unter den Bezeichnungen *Strukturerschließung aus der Bewegung* oder *Ermittlung der relativen Orientierung* auf.

In den 80er Jahren kam es durch die mathematische Formulierung des Problems zu einer Explosion von Untersuchungen und Algorithmen. Trotzdem hat diese Explosion nicht zu einer Verbreitung der Anwendung der Ansätze in Realweltsituationen geführt. Es gibt sehr wenige Ansätze, die sich nur auf die Information aus einer monokularen Bildfolge beschränken. Selbstverständlich ist ein Robotersystem nicht gezwungen, Daten nur aus einem einzigen bildgebenden Sensor zu benutzen. Im Gegenteil muß sich ein technisches Wahrnehmungssystem auf die Kooperation von mehreren Sensoren stützen, d.h., daß ein Roboter Information über die Eigenbewegung durch andere Sensoren ermitteln kann, oder daß der Verlust einer Dimension durch die perspektivische Projektion durch den Gebrauch einer Stereuanordnung von Kameras überwunden werden kann. Trotzdem vertreten wir die Meinung, daß es erst gründlich untersucht werden soll, wann und warum eine monokulare Bildfolge hinreichend oder nicht hinreichend für eine dreidimensionale Interpretation ist. Eine solche Untersuchung trägt auch zur Aufklärung der Rolle eines monokularen Sensors in einer kooperierenden Sensorumgebung bei.

Die Hauptmotivation dieser Arbeit ist der Wunsch nach einer Erläuterung der auftretenden Instabilitäten, damit man in die Lage kommt, die Aufgabe richtig zu formulieren und Grenzen für die Robustheit der Lösung bei bestimmten Bewegungs- und Strukturkonfigurationen zu erkennen. Um solche gefährlichen Situationen zu detektieren, begehen die meisten Ansätze den experimentellen Weg und zeigen die Fehlerverläufe bei simulierten Bewegungskonfigurationen. Der Beitrag des ersten Teils dieser Arbeit besteht im expliziten Nachweis der experimentell gezeigten funktionalen Zusammenhänge bei bestimmten Fehlerempfindlichkeiten. Insbesondere:

- Wir führen eine erschöpfende vereinheitlichende Beschreibung der existierenden Ansätze im Fall des Bildpaars im Hinblick auf ihre Formulierung als Minimierung einer Fehlerfunktion durch.
- Wir zeigen die Beziehung zwischen der Form der Fehlerfunktion und der Fehlerempfindlichkeit in der ermittelten Translationsrichtung und der Größe des Blickfeldes.
- Wir ermitteln die instabilen Situationen im Fall einer sich bewegenden Ebene.
- Wir beschreiben den funktionalen Zusammenhang zwischen der Fehlerempfindlichkeit in der Ermittlung der Objektbeschreibung und der Form der Bewegung sowie der Geometrie der beobachteten Szenenkomponenten.
- Wir zeigen eine Beziehung zwischen den *gefährlichen* Flächen – das sind Flächen, deren Abbildungen eine Mehrdeutigkeit in der dreidimensionalen Interpretation hervorrufen – und der Instabilität in der Ermittlung der Bewegungsparameter.

Im Anschluß an die Beschreibung der existierenden Ansätze wird von uns ein neuer Ansatz zur Ermittlung der Bewegungsparameter aus den Verschiebungsraten von koplanaren Geraden vorgeschlagen, der zur Unifikation der Ansätze für geradlinige und punktförmige Bildbereichshinweise im Fall einer Ebene in Bewegung beiträgt.

Die Benutzung eines Bildpaars dient als Fundament zum Verständnis der bei längeren Bildfolgen auftretenden Empfindlichkeiten und gibt direkt Leistungshinweise für die Ansätze, die eine längere Bildfolge als eine Summe von aufeinander folgenden Bildpaaren verarbeiten.

In der Praxis einzusetzende Systeme müssen dreidimensionale Interpretationen während einer sich dauernd ändernden Umgebung anhand der ständig aufgenommenen Grauwertbilder ermitteln. Deshalb ist das eigentliche Ziel die Verwendung der gesamten Information, die sich während der Aufnahmezeit durch die andauernden relativen Bewegungen zwischen der Kamera und den Objekten der Umgebung erschließen läßt. Insbesondere ist der zwangsmäßig rekursive Charakter dieser Ermittlung von Bedeutung.

Der zweite Teil unserer Arbeit widmet sich der Ausnutzung der zeitlichen Kohärenz der Bewegung während einer längeren Bildfolge durch die Einführung von a priori Wissen über Bewegungsmodelle der Szenenkomponenten. Wir untersuchen die Erfolgsbedingungen der rekursiven Schätzmethode zur Ermittlung von 3D-Bewegungsparametern. Wir schlagen ein neues Verfahren zur rekursiven Berechnung der relativen Bewegung und der Neigung einer Ebene vor. Anhand dieses Verfahrens untersuchen wir, unter welchen Bewegungs- und Geometrieconfigurationen sich die Fehlerempfindlichkeit abschwächt oder unbeeinflusst bleibt. Die Untersuchung wird an synthetischen Experimenten sowie auch an einer Bildfolge durchgeführt, die von einer am Greifer eines Roboterarms befestigten Kamera aufgenommen wurde. Das vorgeschlagene rekursive Verfahren für Punktmerkmale, angewandt an mehreren Ebenen eines polyedrischen Objektes, in Kombination mit dem von uns vorgestellten Bildpaarverfahren für geradlinige Merkmale kann die Basis für einen verallgemeinerten Algorithmus zur Schätzung der relativen Bewegung von polyedrischen Objekten bilden.

1.3 Überblick über die Arbeit

Das zweite Kapitel gibt eine Übersicht über die existierenden Ansätze bei der Bildpaarauswertung in beiden Fällen, bei der Benutzung von Merkmalszuordnungen und bei der Verwendung von Verschiebungsraten. Damit man die später aufgeführten Empfindlichkeitsergebnisse auf existierende Ansätze beziehen kann, wurde besonderer Wert darauf gelegt, daß alle Ansätze in vereinheitlichender mathematischer Notation beschrieben werden. Zusätzlich beschreiben wir die Problemstellung zur Ermittlung der 3D-Bewegung aus Geradenkorrespondenzen und formulieren zwei neue Gleichungen für die Lösung dieses Problems. Im Anschluß beschreiben wir ein neues Verfahren zur Schätzung der relativen Bewegung aus den Projektionen von koplanaren Geraden.

Im dritten Kapitel beginnen wir mit einem Beweis für die Fehlerempfindlichkeit im Fall der reinen Translation, die durch die Form der Fehlernorm begründet ist, von der Abweichung der Translationsrichtung von der Blickrichtung abhängt und bei beschränktem Gesichtsfeld auftaucht. Weiterhin wird der Fall der allgemeinen Bewegung einer Ebene untersucht und die Beziehung zwischen der Fehlerempfindlichkeit und dem Winkel zwischen Ebenennormale und der Translation, sowie die inhärente Unterscheidungsunfähigkeit zwischen einer Rotation und einer Translation bei beschränktem Gesichtsfeld beschrieben. Im Fall der Bewegung einer Ebene wird zusätzlich die Empfindlichkeit in der Schätzung der geometrischen Parameter der Ebene untersucht. Die Empfindlichkeit in der Strukturschätzung wird auch im Fall einer beliebigen Geometrie der Szenenkomponente untersucht. Im letzten Abschnitt des dritten Kapitels zeigen wir, wie mehrdeutig interpretierbare Flächen in Kombination mit einer bestimmten Bewegungskonfiguration zu Fällen führen, die zwar nicht mehr mehrdeutig sind, aber eine fehlerempfindliche dreidimensionale Interpretation verursachen.

Im vierten Kapitel beginnen wir mit einer Klassifikation der Ansätze, die längere Bildfolgen benutzen, aufgrund ihrer Eingabe, des a priori Wissens und der eventuell herangezogenen Information von zusätzlichen Sensoren. Im nächsten Abschnitt gehen wir von einem allgemeinen Schema der rekursiven stochastischen Schätzung aus und leiten die rekursiven Schätzer her, deren Leistung dann an einem einfachen Beispiel analytisch und experimentell untersucht wird. Im letzten Abschnitt beschreiben wir ein neues Verfahren zur rekursiven Bewegungsschätzung aus Verschiebungsraten im Fall einer Ebene. Die Fehlerempfindlichkeit des Verfahrens wird dann experimentell untersucht.

Die verwendeten mathematischen Symbole werden im Anhang A erläutert.

2 Bewegungsschätzung aus einem Bildpaar – Vereinheitlichende Problemstellung und Überblick

2.1 Eingaben zur 3D-Bewegungs- und Strukturschätzung

Die Problemstellung unserer Arbeit geht von den in einer früheren Verarbeitungsstufe ermittelten zeitlichen Zuordnungen im Bildbereich aus. In diesem Abschnitt werden wir kurz auf die Ermittlung der Eingabe zu unserem Problem eingehen. Ansätze zur Bestimmung der zeitlichen Zuordnungen im Bildbereich können in zwei Gruppen unterteilt werden:

Differentielle Ansätze, die die scheinbare momentane Geschwindigkeit der Grauwertstruktur, die *optischer Fluß* genannt wird, berechnen.

Diskrete Ansätze, die auf der Extraktion und Verfolgung von markanten Grauwertstrukturen (Eckenpunkte, Extrema der lokalen Grauwertverteilung, Kantenelemente) basieren.

Nach der obgenannten Definition ist der optische Fluß von der *Verschiebungsrates* zu unterscheiden, die die zeitliche Ableitung der Verschiebung der Projektion eines Szenenpunktes darstellt ([Horn 86] und [Nagel 89]) und die Eingabe zu einer 3D-Interpretation bildet. Das Feld der Verschiebungsrates wird auch als *Bewegungsfeld* bezeichnet. Zur Bestimmung des optischen Flusses wird die zeitliche Konstanz der Grauwertfunktion unterstellt, so daß das totale Zeitdifferential der Grauwertfunktion $g(x(t), y(t), t)$ verschwindet:

$$g_x u + g_y v + g_t = 0 \tag{2.1}$$

Die partiellen Ableitungen der Grauwertfunktion sind (g_x, g_y, g_t) und der damit definierte optische Fluß ist $\mathbf{u} = (u, v)^T$. Man kann direkt einsehen, daß nur die Projektion des optischen Flusses auf die Richtung senkrecht zum Grauwertgradienten $\nabla g = (g_x, g_y)^T$ durch diese Gleichung ermittelt werden kann (das sogenannte *Aperturproblem*).

Unter der Annahme von Lambertschen Reflektanzeigenschaften der Flächen im Szenenbereich konnten [Nagel 89] und [Verri & Poggio 89] zeigen, daß die Gl. (2.1) erfüllende optische Fluß nur unter ganz restriktiven Bedingungen gleich der obendefinierten Verschiebungsrates ist. [Verri & Poggio 89] zeigen, daß diese Diskrepanz mit wachsendem Betrag des Grauwertgradienten abnimmt, und schlagen vor, nur die qualitativen Eigenschaften des optischen Flußfeldes zur weiteren 3D-Interpretation zu benutzen.

[Nagel 83] zeigte, daß die Bestimmung beider Komponenten des optischen Flusses an Grauwertecken möglich wird, und lieferte eine vereinheitlichende Basis für die Ansätze, die lokal den optischen Fluß unter Verwendung der zweiten Ableitungen der Grauwertfunktion bestimmen [Nagel 87a].

Vor kurzem wurde eine neue Definition des optischen Flusses nach der Gleichung

$$\frac{d}{dt}\nabla g = 0 \quad (2.2)$$

eingeführt [Girosi *et al.* 89], die die Berechnung beider Komponenten des optischen Flusses erlaubt, jedoch nicht an allen Stellen, wo es nach den obengenannten Ansätzen möglich ist (wie z.B. Grauwertecken)[Nagel 90].

[Horn & Schunk 81] schlugen als erste einen globalen Ansatz zur Berechnung des optischen Flusses als einer Vektorfunktion in einem Gebiet D vor, wobei folgendes Funktional nach der Funktion $\mathbf{u}(x, y) = (u(x, y), v(x, y))^T$ zu minimieren sei:

$$\iint_D \{(\nabla g^T \mathbf{u} + g_t)^2 + \lambda^2 E_{gl}\} dx dy. \quad (2.3)$$

Der Term E_{gl} ist ein Glattheitsterm zur Überwindung des obengenannten Aperturproblems, der eine möglichst geringe räumliche Änderungsrate des Flusses \mathbf{u} fordert:

$$E_{gl} = \text{Spur}(\nabla \mathbf{u}^T \nabla \mathbf{u}) \quad \text{mit} \quad \nabla \mathbf{u} = \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix}. \quad (2.4)$$

Aufbauend auf der Berechnung des optischen Flusses an Grauwertecken haben [Nagel & Enkelmann 86] und [Nagel 87a] eine Glattheit des Flusses nur in Richtung des Gradienten gefordert $E_{gl} = \text{Spur}(\nabla \mathbf{u}^T W \nabla \mathbf{u})$, wobei die gerichtete Glattheitsforderung durch die Gewichtungsmatrix W ausgedrückt wird. Als ein spezieller Fall der gerichteten Glattheitsforderung ist der Ansatz von [Hildreth 84] anzusehen, der sich aus der Minimierung eines entsprechenden Funktionals entlang einer Nulldurchgangskontur mit einer Forderung nach Glattheit des Flusses entlang der Kontur ergibt. Einen Beweis für die *korrekte Stellung*¹ der obigen Minimierungsproblemen trägt [Schnörr 91] vor. Weitere Untersuchungen zur Einführung von Grenzbedingungen und zur Form des Glatheitstermes wurden von [Lee *et al.* 88] und [Snyder 89] durchgeführt. Optischer Fluß kann auch durch eine Transformation von $g(x, y, t)$ in den Frequenzraum unter Benutzung geeigneter Filter (siehe [Heeger 88] und [Fleet & Jepsen 89]) bestimmt werden. Weitere differentielle Ansätze werden im Überblick [Nagel 88] beschrieben.

Zwischen den differentiellen und den diskreten Ansatz ordnen sich die Verfahren ein, die Verschiebungen im Bild durch Maximierung der Kreuzkorrelation zwischen einem Bildausschnitt im ersten Bild und einem Suchbereich im folgenden Bild ermitteln. Die Kreuzkorrelation ist ein Maß für die Ähnlichkeit zwischen Grauwertstrukturen und wird meistens auf bandpaßgefilterten Bildern angewandt. Aufbauend auf Berechnungen von [Nagel 83] hat [Anandan 89] die Beziehung zwischen Korrelationsmethoden und differentiellen Ansätzen im Fall eines verschwindenden Zeitintervalls nachgewiesen.

¹Ein inverses Problem ist *korrekt gestellt*, wenn eine Lösung existiert, eindeutig und stetig in Bezug auf die Daten ist.

Diskrete Ansätze suchen nach dem korrespondierenden Merkmal im Folgebild. Der Suchbereich kann durch eine Vorverarbeitung der Bilder (z.B. Bandpaßfilterung, wie z.B. bei der Zuordnung von fleckartigen Merkmalen [Kories & Zimmermann 86]), und eine Annahme über die maximale Verschiebung eingeschränkt werden. Suchkriterien sind die Ähnlichkeit der Merkmale und die Konsistenz der Zuordnungen. In den meisten Ansätzen wird eine räumliche und zeitliche Glattheit der Verschiebungen unterstellt (siehe z.B. [Cheng & Aggarwal 90; Weng *et al.* 88]), dennoch entspricht diese Glattheit meistens einer 3D-Translation möglichst parallel zur Bildebene. Die Tatsache, daß die Verschiebungen im Bildbereich bei reiner Translation eine radiale Richtung bezüglich des Expansionspunktes haben, wird nur von [Lawton 83] und [Sethi & Jain 87] ausgenutzt. Eine zeitlich glatte Verschiebung ohne Bezug auf die 3D-Bewegung wird auch bei Ansätzen unterstellt, die Kanten im Bildbereich mit Hilfe der Kalman-Filterung verfolgen [Crowley *et al.* 88; Deriche & Faugeras 90].

Der Klarheit der Begriffe halber müssen wir am Ende folgende Bemerkungen hinzufügen. Weil der differentielle Ansatz auch in Bezug auf die Zeit differentiell vorgeht, darf man da nur von *optischem Fluß* als Approximation der *Verschiebungsraten* sprechen. Im Fall der merkmalsorientierten Verfahren nimmt man das Zusammenfallen der markanten Grauwertstrukturen mit Projektionen von Szenenmerkmalen an, so daß man für die scheinbare Verschiebung eines Bildmerkmals sowie auch für die Verschiebung der Projektion eines Szenenmerkmals denselben Term – *Verschiebungsvektor* – benutzen darf. Wenn die Zeitintervalle sehr kurz (siehe Abschnitt 2.3) sind, kann man die Verschiebungsvektoren für die weitere 3D-Bewegungsbestimmung als Approximation an die Verschiebungsraten benutzen. Weil wir uns hier nur mit den Projektionen von Szenenmerkmalen befassen, werden wir von hier ab nur die Begriffe *Punkt-* bzw. *Geradenkorrespondenz* für nicht verschwindende Zeitintervalle und *Verschiebungsraten* sonst benutzen.

Zum Segmentierungsproblem

Alle Ansätze zur Berechnung der 3D-Bewegung aus einer Menge von Punktkorrespondenzen oder Verschiebungsraten unterstellen, daß diese Menge die Abbildung der starren Bewegung einer Szenenkomponente ist. Dieser Schritt ist am wenigsten untersucht worden und der Mangel an Lösungen verursacht eine ernste Einschränkung der Anwendung der darauf aufbauenden Interpretationsstufen bei Realweltbildfolgen. Die Segmentierung ist von höchster Bedeutung bei autonom mobilen Fahrzeugen, weil sie direkt die Erkennung von bewegten Hindernissen ermöglicht.

Eine einfache Variante des Problems ist die Segmentierung des Bildbereichs in Regionen, die starren Objekten entsprechen, unter der Annahme, daß die relativen Bewegungen fast translatorisch sind. Dabei können die radialen Richtungen der Verschiebungsraten durch eine Hough-Transformation die zu jeder relativen Bewegung entsprechenden Expansionspunkte ergeben (siehe ein ähnliches Verfahren bei [Jain 84]).

Ein anderer Lösungsweg ist die Segmentierung in Bereiche, die Abbildungen von sich bewegenden Ebenen sind. Wie wir später beschreiben werden, sind die Verschiebungsraten in diesen Bereichen Funktionen zweiter Ordnung bezüglich der Bildkoordinaten. Hough-Transformationen in dem Raum der Koeffizienten dieser Funktionen [Adiv 85] oder Suche nach allen Minima der Fehlersumme in den Verschiebungsraten [Murray & Buxton 87] ermöglichen im Fall eines niedrigen Meßrauschens die Segmentierung.

Anders geht man vor, wenn man schon Information über die Eigenbewegung aus anderen Sensoren besitzt. In diesem Fall detektieren [Heeger & Hager 88] probabilistisch Bereiche, an denen die gemessenen Verschiebungen inkonsistent mit dem Modell der Eigenbewegung sind. Weitere Bedingungen kann man durch Annahmen über die Struktur der Umgebung (z.B. ebene Straße oder Flur [Enkelmann 90]) einführen.

Einen experimentell vergleichenden Überblick der erwähnten Ansätze findet man in [Thompson & Pong 90].

2.2 3D Bewegungsschätzung aus Punktkorrespondenzen eines Bildpaars

Zur einfachen Darstellung der perspektivischen Projektion wird ein Koordinatensystem gewählt, dessen Ursprung im Projektionszentrum liegt. Die optische Achse ist die z -Achse und die Bildebene liegt im Abstand $f =$ Brennweite vor dem Ursprung in positiver z -Richtung. Wenn \mathbf{X} die dreidimensionale Lage eines Punktes im Szenenbereich ist, lautet seine Abbildung in der Bildebene

$$\mathbf{x} = f \frac{\mathbf{X}}{\hat{\mathbf{z}}^T \mathbf{X}}, \quad (2.5)$$

wobei $\hat{\mathbf{z}}$ der Einheitsvektor in Richtung der optischen Achse ist. Ohne Beschränkung der Allgemeinheit kann man annehmen, daß die Brennweite gleich einer Maßeinheit ist. Ab jetzt werden wir unterstellen, daß $\mathbf{x} = (x, y, 1)^T$ ist. Nach dieser Normalisierung, die man während der Implementierung unter Kenntnis der internen Kameraparameter durchführen kann, ist eine Beschreibung von Positionen oder Verschiebungen in Kamerakoordinaten nur von dem effektiven Gesichtsfeld abhängig. Unter effektivem Gesichtsfeld verstehen wir den Öffnungswinkel, den die Abbildung der sich in relativer Bewegung zur Kamera befindenden Szenenkomponente aufspannt. Bei wachsender Brennweite und konstant bleibendem effektivem Gesichtsfeld muß man die Verfeinerung der Auflösung und die sich damit ergebende Reduzierung des Diskretisierungsrauschens berücksichtigen.

Im Fall des Bildpaars kann man die Bewegung einer Kamera in einer stationären Umgebung sowie auch die relative Bewegung eines starren Objektes mit folgender Gleichung beschreiben

$$\mathbf{X2} = R\mathbf{X1} + \mathbf{T}. \quad (2.6)$$

$\mathbf{X1}$ und $\mathbf{X2}$ sind die Positionsvektoren eines Punktes im Szenenbereich in den Zeitpunkten t_1 und t_2 . Im Fall einer Kamerabewegung (Abb. 2.1) beschreibt diese Gleichung eine Rotation R^T der Kamera um eine Achse durch den Ursprung und eine darauffolgende Translation $-\mathbf{T}$. Im Fall der Objektbewegung beschreibt sie die Rotation eines Objektes R um eine Achse durch den Ursprung und eine darauffolgende Translation \mathbf{T} .

Die Projektionen der Punkte $\mathbf{X1}$ und $\mathbf{X2}$ lauten jeweils $\mathbf{x1}$ und $\mathbf{x2}$ und werden durch folgende Meßgleichungen angegeben

$$\mathbf{x1} = \frac{\mathbf{X1}}{\hat{\mathbf{z}}^T \mathbf{X1}} \quad \text{und} \quad \mathbf{x2} = \frac{R\mathbf{X1} + \mathbf{T}}{\hat{\mathbf{z}}^T (R\mathbf{X1} + \mathbf{T})}. \quad (2.7)$$

Man kann erkennen, daß die Tiefen $\hat{\mathbf{z}}^T \mathbf{X1}$ und $\hat{\mathbf{z}}^T \mathbf{X2}$ sowie auch die Translation \mathbf{T} nur bis auf einen gemeinsamen Skalierungsfaktor bestimmbar sind. Z.B. ändert ein

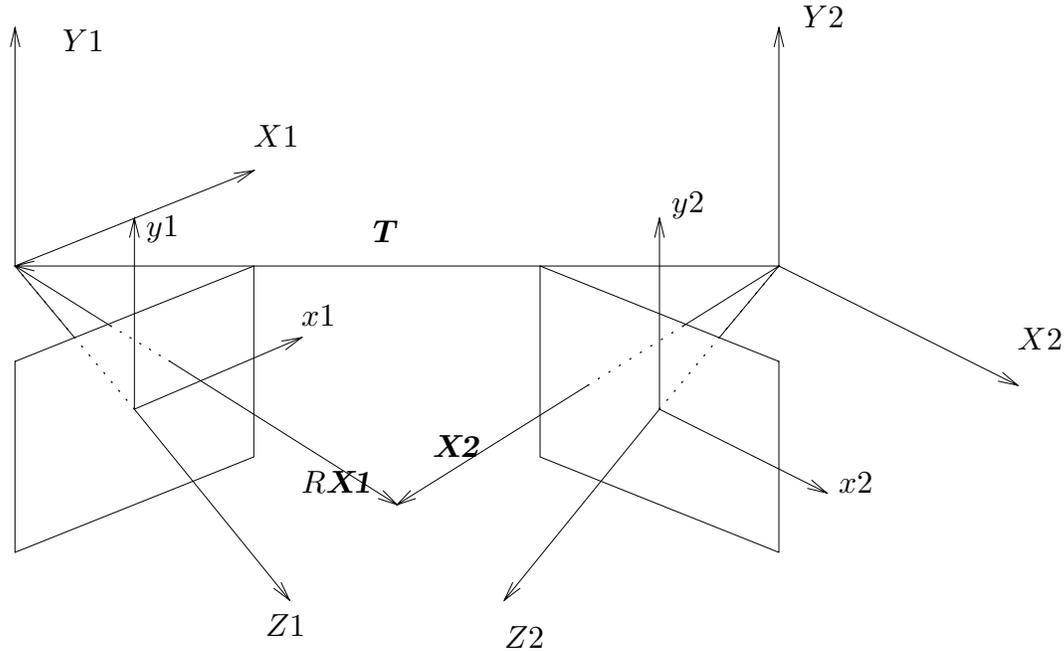


Abbildung 2.1: Geometrie des Problems der relativen Orientierung

Objekt im doppelten Abstand, das sich doppelt so weit verschiebt, die Messungen von \mathbf{x}_1 und \mathbf{x}_2 nicht. Dieser Skalierungsfaktor läßt sich eliminieren, indem man entweder den Translationsvektor normalisiert ($T_z = 1$ oder $\|\mathbf{T}\| = 1$) oder die Tiefe eines Punktes festsetzt. Eine Rotation wird durch drei unabhängige Parameter beschrieben. Daraus erhält man für m Punktkorrespondenzen $3m+6-1$ Unbekannte und $2m$ Meßgleichungen (siehe dazu die ersten Untersuchungen in [Zvi-Meiri 80; Roach & Aggarwal 80]). In dem nicht mehrdeutigen Fall – darauf werden wir später eingehen – braucht man mindestens fünf Punktkorrespondenzen, um 3D-Struktur und Bewegung zu bestimmen.

Eine geometrische Beziehung zwischen den Messungen \mathbf{x}_1 , \mathbf{x}_2 und den unbekanntem Bewegungsparametern kann direkt durch die Beobachtung hergeleitet werden, daß die Strahlenvektoren \mathbf{x}_2 und $\mathbf{R}\mathbf{x}_1$ koplanar mit dem Translationsvektor \mathbf{T} sind. Diese Bedingung wurde geometrisch [Kruppa 13] und dann analytisch [Thompson 59] zuerst in der Photogrammetrie formuliert und wird *epipolare Bedingung* genannt:

$$\mathbf{x}_2^T(\mathbf{T} \times \mathbf{R}\mathbf{x}_1) = 0. \quad (2.8)$$

Die epipolare Bedingung läßt sich auch algebraisch aus der Gl. (2.6) herleiten, indem man das Vektorprodukt von links mit \mathbf{T} und das Skalarprodukt mit \mathbf{X}_2 bildet, und durch die Tiefen dividiert. Bevor wir auf die Lösung der epipolaren Bedingung eingehen, erwähnen wir neben der Skalierungsmehrdeutigkeit noch eine zweite Mehrdeutigkeit. Wenn die epipolare Bedingung von der Lösung (\mathbf{T}, \mathbf{R}) erfüllt wird, wird sie auch für die gleichen Punktkorrespondenzen von $(\mathbf{T}, \mathbf{R}_t\mathbf{R})$ erfüllt, wobei $\mathbf{R}_t = 2\mathbf{T}\mathbf{T}^T - \mathbf{I}$ einer Rotation von 180° um die Translationsrichtung entspricht. Sie wird Mehrdeutigkeit des *gedrehten Paares* nach [Longuet-Higgins 86] genannt und wird – wie auch die Skalierungsmehrdeutigkeit – als *triviale* Mehrdeutigkeit nach [Horn 90] bezeichnet, im Gegensatz zu den später behandelten nicht trivialen Mehrdeutigkeiten.

Gl. (2.8) läßt sich wie folgt umschreiben

$$\mathbf{x}_2^T E \mathbf{x}_1 = 0 \quad \text{mit} \quad E = [\mathbf{T}]_{\times} R \quad \text{und} \quad [\mathbf{T}]_{\times} = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} \quad (2.9)$$

Wenn man die Elemente der sogenannten *E-Matrix* als einen neundimensionalen Vektor \mathbf{e} ansieht, erhält man für jedes Punktepaar $(\mathbf{x}_{1_i}, \mathbf{x}_{2_i})$ folgende lineare homogene Gleichung bezüglich der Elemente von \mathbf{e} , die man *essentielle Parameter* nennt:

$$\mathbf{a}_i^T \mathbf{e} = \left(x_{2_i} x_{1_i} \quad x_{2_i} y_{1_i} \quad x_{2_i} \quad y_{2_i} x_{1_i} \quad y_{2_i} y_{1_i} \quad y_{2_i} \quad x_{1_i} \quad y_{1_i} \quad 1 \right)^T \mathbf{e} = 0. \quad (2.10)$$

Unterstellt man die Eingabe von mehr als acht ($m \geq 8$) Punktkorrespondenzen, dann kann man eine Minimierung

$$\sum_{i=1}^m (\mathbf{a}_i^T \mathbf{e})^2 = \mathbf{e}^T \left(\sum_{i=1}^m \mathbf{a}_i \mathbf{a}_i^T \right) \mathbf{e} \implies \min_{\|\mathbf{e}\|^2=2}$$

unter der Nebenbedingung $\|\mathbf{e}\|^2 = 2$ durchführen. Die Nebenbedingung läßt sich leicht aus der Normierung $\|\mathbf{T}\| = 1$ und der Orthogonalität von R herleiten.

Der Gebrauch von essentiellen Parametern wurde zuerst von [Longuet-Higgins 81, 84a] und [Tsai & Huang 84] eingeführt. Seitdem ergaben sich zahlreiche Ansätze, die die Bewegungsparameter aus den essentiellen Parametern zu berechnen versuchen [Zhuang *et al.* 86; Weng *et al.* 89a; Faugeras *et al.* 87a; Philip 91]. Bis auf [Faugeras & Maybank 90; Huang & Faugeras 89] hat keiner der Ansätze die Nebenbedingung berücksichtigt, daß die *E-Matrix* in das Produkt einer schiefsymmetrischen und einer orthogonalen Matrix nach (2.9) zerlegbar sein soll. [Huang & Faugeras 89; Faugeras & Maybank 90] haben als notwendige und hinreichende Bedingung für eine solche Zerlegbarkeit der *E-Matrix* bewiesen, daß einer ihrer singulären Werte gleich Null und die übrigen zwei zueinander gleich sein sollen. Die dazu äquivalente algebraische Formulierung lautet

$$\det(E) = 0 \\ \frac{1}{2} \left(\text{Spur}(EE^T) \right)^2 - \text{Spur} \left((EE^T)^2 \right) = 0. \quad (2.11)$$

Die Menge aller *E-Matrizen*, die die Bedingungen (2.11) erfüllen, bilden eine Mannigfaltigkeit von Dimension fünf (drei Rotations- und zwei Translationsparameter) im achtdimensionalen projektiven Raum. Bei Eingabe von fünf Punktkorrespondenzen erhält man fünf Bedingungen wie (2.9), die einen linearen Raum von Kodimension fünf aufspannen. Die Anzahl der Schnitte dieses Raums mit der Mannigfaltigkeit der *E-Matrizen* bestimmt die Anzahl der Lösungen für dieses Problem und ist gleichzeitig identisch mit dem Grad der Mannigfaltigkeit. [Demazure 88] hat bewiesen, daß der Grad gleich zehn ist, man erhält also zu jeder Menge von fünf Punktkorrespondenzen maximal² zehn kompatible Kamerabewegungen. Aufbauend auf der Arbeit von [Kruppa 13] haben [Faugeras & Maybank 90] auf einem anderen Weg – mit Werkzeugen aus der projektiven Geometrie – die gleiche Anzahl von Lösungen bewiesen. Weitere Resultate

²wenn alle Lösungen reell und verschieden voneinander sind

zur Anzahl der Lösungen, die auf einem algebraischen Weg hergeleitet worden sind, werden von [Netravali *et al.* 89] und [Horn 91] angegeben. Im Fall von mehreren ($m > 5$) Korrespondenzen wurde zuerst in der Photogrammetrie [Krames 40; Wunderlich 41] und dann in der Bildauswertung [Longuet-Higgins 86; Faugeras & Maybank 90; Negahdaripour 90b] bewiesen, daß die Anzahl der Lösungen maximal drei ist. Auf diese Ergebnisse werden wir später in der Fehlerempfindlichkeitsanalyse genauer eingehen.

Wir setzen unseren Überblick mit Ansätzen fort, die keinen Gebrauch von essentiellen Parametern machen. [Nagel 81] schreibt das Spatprodukt in (2.8) für zwei Punktkorrespondenzen so um

$$\mathbf{T}^T(R\mathbf{x}_{1_1} \times \mathbf{x}_{2_1}) = 0 \quad \mathbf{T}^T(R\mathbf{x}_{1_2} \times \mathbf{x}_{2_2}) = 0,$$

daß man direkt das Senkrechtstehen von \mathbf{T} auf $(R\mathbf{x}_{1_1} \times \mathbf{x}_{2_1})$ und $(R\mathbf{x}_{1_2} \times \mathbf{x}_{2_2})$ einsieht:

$$\mathbf{T} = k(R\mathbf{x}_{1_1} \times \mathbf{x}_{2_1}) \times (R\mathbf{x}_{1_2} \times \mathbf{x}_{2_2}).$$

Wenn man \mathbf{T} in (2.8) mit diesem Ausdruck ersetzt, erhält man eine kubische Gleichung bezüglich der Elemente der Rotationsmatrix

$$((R\mathbf{x}_{1_1} \times \mathbf{x}_{2_1}) \times (R\mathbf{x}_{1_2} \times \mathbf{x}_{2_2}))^T (R\mathbf{x}_{1_i} \times \mathbf{x}_{2_i}) = 0.$$

Man braucht dann noch drei Punktkorrespondenzen ($i = 3, 4, 5$) und es zeigt sich, daß die Mindestanzahl der Punktkorrespondenzen für die Auflösung des Problems fünf ist.

Aus (2.6) folgt, daß der Abstandsvektor zwischen zwei Punkten \mathbf{X}_k und \mathbf{X}_l folgende Transformation erfüllt

$$\mathbf{X}_{2_k} - \mathbf{X}_{2_l} = R(\mathbf{X}_{1_k} - \mathbf{X}_{1_l}). \quad (2.12)$$

[Jerian & Jain 90] benutzen die Quaternionrepräsentation für die Rotation (siehe [Horn 86]), um aus (2.12) nach Eliminierung der Tiefen durch Einführung einer dritten Punktkorrespondenz eine Polynomialgleichung vierten Grades bezüglich der Quaternionkomponenten zu erhalten. Durch Einführung eines vierten Punktes erhält man eine zweite Gleichung identischer Struktur. Wenn man jetzt eine von den drei unabhängigen Quaternionparametern festsetzt, kann man beide Gleichungen nach den anderen zwei Parametern resolvieren und eine Polynomialgleichung sechzehnter Ordnung nach einer Unbekannten erhalten. Für den am Anfang festgesetzten Parameter wird dann eine Suche nach dem Wert dieses Parameters durchgeführt, der das aus der Berechnung der Translation hergeleitete Fehlerresiduum minimiert. Der Algorithmus hat seine Stärke bei Anwendungen, bei denen der Parameterraum durch a priori Wissen über die Bewegung eingeschränkt werden kann.

[Horn 90] wendet ein Iterationsverfahren zur Lösung der relativen Orientierung an. Durch Minimierung der Fehlerfunktion

$$\begin{aligned} \sum_{i=1}^m w_i \left(\mathbf{x}_{2_i}^T (\mathbf{T} \times R\mathbf{x}_{1_i}) \right)^2 &= \mathbf{T}^T \left(\sum_{i=1}^m w_i (R\mathbf{x}_{1_i} \times \mathbf{x}_{2_i}) (R\mathbf{x}_{1_i} \times \mathbf{x}_{2_i})^T \right) \mathbf{T} \\ &= \mathbf{T}^T \mathbf{A} \mathbf{T} \implies \min_{\|\mathbf{T}\|=1} \end{aligned} \quad (2.13)$$

nach der Translation \mathbf{T} erhält [Horn 90] den Startwert für die Translation³. Diese Minimierung, die wir wiederholt in den nächsten Abschnitten treffen werden, hat als Lösung

³Auf die Gewichtung w_i werden wir bei der Besprechung der Fehlermaße eingehen.

den Eigenvektor der Matrix A , der dem kleinsten Eigenwert dieser Matrix entspricht. Die Bewegungsparameter werden dann durch sukzessive Korrekturen $\delta\mathbf{T}$ und $\delta\boldsymbol{\omega}$ jeweils für die Translation und für die Rotation bestimmt:

$$\begin{aligned} (\mathbf{T} + \delta\mathbf{T})^T ((R\mathbf{x}_1 + \delta\boldsymbol{\omega} \times R\mathbf{x}_1) \times \mathbf{x}_2) = \\ \mathbf{T}^T (R\mathbf{x}_1 \times \mathbf{x}_2) + (R\mathbf{x}_1 \times \mathbf{x}_2)^T \delta\mathbf{T} + (R\mathbf{x}_1 \times (\mathbf{x}_2 \times \mathbf{T}))^T \delta\boldsymbol{\omega}. \end{aligned} \quad (2.14)$$

Der obige Ausdruck wird dann über alle Punkte nach $(\delta\mathbf{T}, \delta\boldsymbol{\omega})$ unter Berücksichtigung der Nebenbedingung $\mathbf{T}^T \delta\mathbf{T} = 0$ minimiert, was der bekannten Gauß-Newton Iteration in der numerischen Analysis ähnlich ist. Die Inspiration des Ansatzes stammt aus den Korrekturverfahren in der Photogrammetrie (siehe z.B. [Finsterwalder & Hofmann 68]).

Auch [Spetsakis & Aloimonos 89] schlagen einen iterativen Ansatz vor, der sich aus der Wiederholung von jeweils zwei Schritten bildet. Der erste Schritt ist immer eine Minimierung wie bei [Horn 90] Gl. (2.13) nach der Translation. Der Ausdruck $\mathbf{T}^T A\mathbf{T}$ wird dann nach den Rotationsparametern – hier wird als Repräsentation der Gibbs-Vektor benutzt – minimiert. Es wird ebenfalls eine Gewichtung angewandt, die wir im nächsten Kapitel besprechen werden. Der Vorteil des Algorithmus liegt in seiner Erweiterungsfähigkeit auf den Fall einer längeren Bildfolge.

Eine besondere Untersuchung verdient der Fall einer 3D-Punktkonfiguration in einer Ebene. Wenn \mathbf{N} die Einheitsnormale zur Ebene und d der Abstand der Ebene vom Ursprung ist, lautet die Gleichung der Ebene $\mathbf{N}^T \mathbf{X} = d$. Beschreibt man die Ebene im ersten Koordinatensystem $\mathbf{N}^T \mathbf{X}_1 = d$, kann man Gl. (2.6) wie folgt umschreiben:

$$\begin{aligned} \mathbf{X}_2 = R\mathbf{X}_1 + \mathbf{T} \frac{\mathbf{N}^T \mathbf{X}_1}{d} = (R + \mathbf{T} \frac{\mathbf{N}^T}{d}) \mathbf{X}_1 = A_{col} \mathbf{X}_1 \\ Z_2 \mathbf{x}_2 = A_{col} Z_1 \mathbf{x}_1. \end{aligned} \quad (2.15)$$

Die Transformation (2.15) beschreibt eine Kollineation und die Elemente der Matrix A_{col} sind bis auf einen Skalierungsfaktor zu bestimmen. Aus (2.15) erhält man zwei unabhängige Gleichungen für jede Korrespondenz. Man braucht mindestens vier Punkte, um die Kollineation zu bestimmen. Die Kollineation kann immer in die Bewegungs- und Strukturparameter zerlegt werden, deren Anzahl ebenfalls acht beträgt. Die Zerlegung erfolgt durch Singulärwert-Zerlegung der Matrix A_{col} (siehe [Tsai *et al.* 82; Faugeras & Lustman 88]) und es ergeben sich immer zwei Lösungen, wenn die Translationsrichtung nicht senkrecht zur Ebene steht [Hay 66]. Die explizite Beziehung zwischen beiden Lösungen hat zuerst [Negahdaripour 90a] beschrieben.

Die aktuellste Übersicht über Ansätze zur Bewegungs- und Strukturermittlung aus einem Bildpaar mittels Punktkorrespondenzen findet sich in [Jerian & Jain 91].

2.3 3D Bewegungsschätzung aus dem Bewegungsfeld

Die Geschwindigkeit eines Punktes \mathbf{X} auf einem Objekt im Raum, das sich mit translatorischer Geschwindigkeit \mathbf{v} und Winkelgeschwindigkeit $\boldsymbol{\omega}$ (Abb. 2.2) bewegt, läßt sich durch folgende Gleichung beschreiben

$$\dot{\mathbf{X}} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{X}, \quad (2.16)$$

die unter Verwendung des entgegengesetzten Vorzeichens für beide Terme auch den Fall der Eigenbewegung beschreibt.

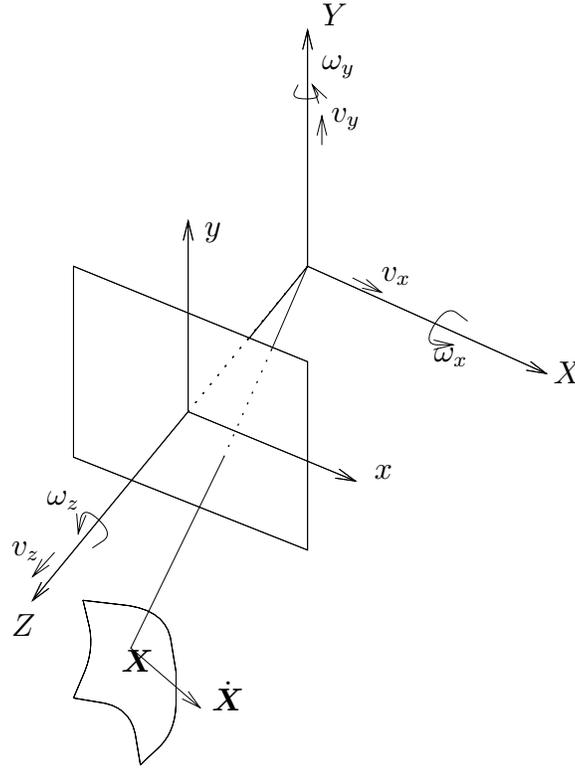


Abbildung 2.2: Geometrie des kontinuierlichen Falls nach [Longuet-Higgins & Prazdny 80].

Wenn man die Gleichung der perspektivischen Projektion nach der Zeit differenziert, erhält man für die Verschiebungsrate $\dot{\mathbf{x}}$ in der Bildebene

$$\mathbf{x} = \frac{\mathbf{X}}{\hat{\mathbf{z}}^T \mathbf{X}} \implies \dot{\mathbf{x}} = \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} (\dot{\mathbf{X}} - (\hat{\mathbf{z}}^T \dot{\mathbf{X}}) \mathbf{x}) = \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} (\hat{\mathbf{z}} \times (\dot{\mathbf{X}} \times \mathbf{x})). \quad (2.17)$$

Nach Ersetzung der 3D-Geschwindigkeit von 2.16 erhalten wir

$$\dot{\mathbf{x}} = \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) + \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})). \quad (2.18)$$

Die letzte Gleichung hat eine attraktivere Form im Vergleich zu Ausdrücken wie (2.7), die im diskreten Fall erscheinen. Die Verschiebungsrate läßt sich in eine translatorische (erster Summand) und eine rotatorische (zweiter Summand) Komponente zerlegen. Wir sehen, daß die translatorische Komponente verschwindet, wenn der Positionsvektor des Bildpunktes \mathbf{x} parallel zur Translationsgeschwindigkeit \mathbf{v} ist. Dieser Bildpunkt ($v_x/v_z, v_y/v_z$) heißt *Expansionspunkt* und trägt die gesamte Information über die translatorische Bewegung. Damit man den Fall $v_z = 0$ miteinschließen kann, sollte man sich den Expansionspunkt als einen Punkt in der projektiven Ebene oder auf einer Einheits-halbkugel vorstellen.

Die rotatorische Komponente der Verschiebungsrate verschwindet, wenn der Positionsvektor \mathbf{X} im Raum für einen Bildpunkt \mathbf{x} parallel zur Winkelgeschwindigkeit ist. Die rotatorische Komponente enthält keine Information über die Struktur der abgebildeten Szene.

Die Gl. (2.18) lautet in Komponentenform [Longuet-Higgins & Prazdny 80]

$$\begin{aligned}\dot{x} &= \frac{v_x - xv_z}{Z} - xy\omega_x + (1 + x^2)\omega_y - y\omega_z \\ \dot{y} &= \frac{v_y - yv_z}{Z} - (1 + y^2)\omega_x + xy\omega_y + x\omega_z.\end{aligned}\quad (2.19)$$

Die translatorische Komponente läßt sich umschreiben, als

$$\begin{aligned}\dot{x}_T &= \frac{v_z}{Z} \left(\frac{v_x}{v_z} - x \right) \\ \dot{y}_T &= \frac{v_z}{Z} \left(\frac{v_y}{v_z} - y \right).\end{aligned}\quad (2.20)$$

Man erkennt sofort, daß sie kollinear mit der Verbindungslinie zwischen dem Expansionspunkt und dem betrachteten Bildpunkt ist. Ein rein translatorisches Bewegungsfeld sieht radial aus, und diese Eigenschaft kann zur gleichzeitigen Bestimmung der Korrespondenzen und zur Berechnung des Expansionspunkts ausgenutzt werden [Lawton 83].

Wie wir schon im Abschnitt 2.1 besprochen haben, kann man auch finite Verschiebungen als Verschiebungsraten auffassen, falls es sich um kleine Bewegungen – entweder kurze Zeitintervalle oder langsame Bewegung – handelt. Für kurze Zeitintervalle approximiert man Rotation und Translation durch

$$R\mathbf{X} \approx \mathbf{X} + \boldsymbol{\omega}\delta t \times \mathbf{X} \quad \mathbf{T} \approx \mathbf{v}\delta t.$$

In Gl. (2.7) ersetzen wir \mathbf{x}_1 durch \mathbf{x} und \mathbf{x}_2 durch $\mathbf{x} + \dot{\mathbf{x}}\delta t$. Nach Zusammenfassung in Vektorprodukten erhalten wir:

$$\dot{\mathbf{x}} = \frac{\frac{\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})}{\hat{\mathbf{z}}^T \mathbf{X}} + \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega}))}{1 + \delta t \left(\hat{\mathbf{z}}^T (\boldsymbol{\omega} \times \mathbf{x}) + \frac{\hat{\mathbf{z}}^T \mathbf{v}}{\hat{\mathbf{z}}^T \mathbf{X}} \right)}\quad (2.21)$$

$$\approx \frac{\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})}{\hat{\mathbf{z}}^T \mathbf{X}} + \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})) + \mathcal{O}(\delta t^2).\quad (2.22)$$

In praktischen Anwendungen darf man die kontinuierliche Methode auch bei Merkmalszuordnungen einsetzen, wenn der zweite Term im Nenner von (2.21) erheblich kleiner als 1 ist (siehe [Adiv 85])

$$\delta t \left(\omega_x y - \omega_y x + \frac{v_z}{Z} \right) \ll 1.\quad (2.23)$$

Zur Ermittlung der 3D-Geschwindigkeiten kann man die Tiefen aus der Gl. (2.18) eliminieren, indem man das Skalarprodukt mit $\mathbf{v} \times \mathbf{x}$ bildet:

$$\begin{aligned}\dot{\mathbf{x}}^T (\mathbf{v} \times \mathbf{x}) &= -(\mathbf{x} \times \boldsymbol{\omega})^T (\mathbf{v} \times \mathbf{x}) \iff \\ (\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x}) &= 0.\end{aligned}\quad (2.24)$$

Diese Gleichung ist der Ausdruck der epipolaren Bedingung (2.8) im kontinuierlichen Fall. Sie gilt auch im Fall des Gebrauchs einer sphärischen Projektion (siehe [Maybank 86]) und die Herleitung im Anhang B).

Wir suchen auch im kontinuierlichen Fall nach essentiellen Parametern. Gl. (2.24) läßt sich umformen

$$(\mathbf{x} \times \dot{\mathbf{x}})^T \mathbf{v} + \mathbf{x}^T \left(\frac{\boldsymbol{\omega} \mathbf{v}^T + \mathbf{v} \boldsymbol{\omega}^T}{2} - \boldsymbol{\omega}^T \mathbf{v} I \right) \mathbf{x} = 0, \quad (2.25)$$

wobei man sofort erkennt, daß die Translation \mathbf{v} und die sechs Elemente der symmetrischen Matrix im zweiten Term insgesamt neun essentielle Parameter bilden. In Komponentenform ergibt sich dann die lineare Gleichung

$$\begin{aligned} \mathbf{a}_c^T \mathbf{e}_c &= 0 \quad \text{mit} & (2.26) \\ \mathbf{a}_c^T &= (1 \quad x \quad y \quad x^2 \quad xy \quad y^2 \quad -\dot{y} \quad \dot{x} \quad (xy - y\dot{x})) \\ \mathbf{e}_c^T &= \begin{pmatrix} (-v_x \omega_x - v_y \omega_y) & (v_z \omega_x + v_x \omega_z) & (v_z \omega_y + v_y \omega_z) \\ (-v_z \omega_z - v_y \omega_y) & (v_x \omega_y + v_y \omega_x) & (-v_z \omega_z - v_x \omega_x) & v_x & v_y & v_z \end{pmatrix}. \end{aligned}$$

Wir beachten, daß die drei letzten essentiellen Parameter gleich den Komponenten der Translationsgeschwindigkeit sind. Durch die Einführung der essentiellen Parameter in Gl. (2.25) wurde die Anzahl der eigentlichen Unbekannten von fünf auf acht erhöht – nach dem Ausschluß des unbestimmbaren Skalierungsfaktors. Damit die Winkelgeschwindigkeit ermittelt werden kann, soll das Gleichungssystem

$$\begin{pmatrix} -e_{c7} & -e_{c8} & 0 \\ e_{c9} & 0 & e_{c7} \\ 0 & e_{c9} & e_{c8} \\ 0 & -e_{c8} & -e_{c9} \\ e_{c8} & e_{c7} & 0 \\ -e_{c7} & 0 & -e_{c9} \end{pmatrix} \boldsymbol{\omega} = \begin{pmatrix} e_{c1} \\ e_{c2} \\ e_{c3} \\ e_{c4} \\ e_{c5} \\ e_{c6} \end{pmatrix}$$

nach $\boldsymbol{\omega}$ exakt (und nicht im Sinne einer Ausgleichsrechnung) lösbar sein, wobei e_{ci} jeweils die i -te Komponente des Vektors \mathbf{e}_c ist. Die Lösbarkeit wird algebraisch als ein System von drei polynomialen Gleichungen vierter Ordnung formuliert, das aus dem Verschwinden der drei 4×4 Determinanten

$$\begin{vmatrix} -e_{c7} & -e_{c8} & 0 & e_{c1} \\ e_{c9} & 0 & e_{c7} & e_{c2} \\ 0 & e_{c9} & e_{c8} & e_{c3} \\ 0 & -e_{c8} & -e_{c9} & e_{c4} \end{vmatrix}, \quad \begin{vmatrix} -e_{c7} & -e_{c8} & 0 & e_{c1} \\ e_{c9} & 0 & e_{c7} & e_{c2} \\ 0 & e_{c9} & e_{c8} & e_{c3} \\ e_{c8} & e_{c7} & 0 & e_{c5} \end{vmatrix} \quad \text{und} \quad \begin{vmatrix} -e_{c7} & -e_{c8} & 0 & e_{c1} \\ e_{c9} & 0 & e_{c7} & e_{c2} \\ 0 & e_{c9} & e_{c8} & e_{c3} \\ -e_{c7} & 0 & -e_{c9} & e_{c6} \end{vmatrix}$$

folgt. Diese nichtlinearen Gleichungen sollen als Nebenbedingung bei der Lösung des sich aus der Anwendung der Gl. (2.26) an jedem Punkt (x, y) ergebenden linearen Gleichungssystems berücksichtigt werden. Der Gebrauch von essentiellen Parametern im kontinuierlichen Fall ist bis jetzt nicht untersucht worden – nur die linearen Gleichungen werden in [Zhuang *et al.* 88] eingeführt – und die Berechnung der Anzahl der Lösungen anhand der essentiellen Parameter stellt ein noch zu lösendes Problem dar.

[Maybank 85; Horn 87; Negahdaripour 89] haben nachgewiesen, daß eine beliebige Anzahl von Verschiebungsraten maximal drei 3D-Interpretationen erlaubt, falls die abgebildete Fläche eine kritische Fläche (für eine Definition siehe Abschnitt 3.6) ist. Im Fall von fünf Verschiebungsraten hat [Maybank 90b] gezeigt, daß die Anzahl der Lösungen – wie auch im diskreten Fall – auf maximal zehn steigt.

Aus (2.18) kann man direkt einen Lösungsweg erschließen, indem man die Quadrate der Abweichungen zwischen gemessenen und erwarteten Verschiebungsraten nach den Bewegungsparametern und den Tiefen minimiert

$$\iint_D \left\| \dot{\mathbf{x}} - \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) - \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})) \right\|^2 dx dy \Rightarrow \min, \quad (2.27)$$

wobei wir unterstellt haben, daß das Feld der Verschiebungsraten dicht im Gebiet D ist. Das obige Integral stellt ein Funktional in Bezug auf die Tiefenfunktion $Z(x, y) = \hat{\mathbf{z}}^T \mathbf{X}$ dar. Die Lösung dieses Variationsproblems ist aber einfach, weil die Ableitungen der Tiefe nach x und y im Integranden nicht auftreten. So reduziert sich die entsprechende Euler-Lagrangesche-Gleichung auf das Nullsetzen der Ableitung des Integranden nach der Tiefe:

$$\left(\dot{\mathbf{x}}_T - \frac{\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})}{Z} \right)_T \frac{\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})}{Z^2} = 0,$$

wobei wir die translatorische Komponente $\dot{\mathbf{x}} - \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega}))$ mit $\dot{\mathbf{x}}_T$ bezeichnen. Daraus folgt

$$Z = \frac{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|^2}{\dot{\mathbf{x}}_T^T (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}))}, \quad (2.28)$$

und nach dem Ersetzen von Z im Integranden von (2.27)

$$\begin{aligned} \left\| \dot{\mathbf{x}}_T - \frac{\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})}{Z} \right\|^2 &= \left\| \dot{\mathbf{x}}_T - \frac{(\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})) (\dot{\mathbf{x}}_T^T (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})))}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|^2} \right\|^2 \\ &= \left\{ \frac{\|(\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})) \times (\dot{\mathbf{x}}_T \times (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})))\|}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|^2} \right\}^2 = \left\{ \frac{\dot{\mathbf{x}}_T^T (\mathbf{v} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 \\ &= \left\{ \frac{(\dot{\mathbf{x}} - \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})))^T (\mathbf{v} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 = \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 \end{aligned} \quad (2.29)$$

erhält man das Integral

$$\iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 dx dy. \quad (2.30)$$

[Bruss & Horn 83; Adiv 85] schlagen die Minimierung dieses Integrals nach den Bewegungsparametern vor, was ein nichtlineares Problem darstellt. In [Adiv 85; Lawton 83] wird der Raum der Translationen (Einheitshalbkugel $\mathbf{v}=1$) abgetastet. Für eine gegebene Translation kann man $\boldsymbol{\omega}$ linear ermitteln, so daß man für jeden Abtastpunkt ein Residuum erhält. Eine Suche auf der Halbkugel ergibt dann die Translation, die dieses Residuum minimiert. Ausgehend von (2.27) bilden auch [Heeger & Jepson 92] das Residuum, das sich aus der Minimierung nach dem reziproken Wert der Tiefe und der Winkelgeschwindigkeit – beide treten linear in den Integranden von (2.27) auf – herausstellt. Anschließend führen [Heeger & Jepson 92] zur Bestimmung der Translation eine Suche auf der Halbkugel nach dem Minimum dieses Residuums durch.

[Yasumoto & Medioni 86] versuchen eine Stabilisierung der Fehlerfunktion (2.30), indem sie durch Addition des Regularisierungsfaktors $\lambda\|\boldsymbol{\omega}\|^2$ eine kleine Winkelgeschwindigkeit verlangen. Sie berichten, daß diese Regularisierung von Bedeutung im Fall eines kleinen effektiven Gesichtsfelds ist.

Wenn man aus der tiefenfreien Beziehung (2.24) folgende Minimierung formuliert,

$$\iint_D \left((\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x}) \right)^2 dx dy \quad (2.31)$$

erhält man eine einfachere Norm als (2.30), deren Verwendung unter der in Abschnitt 3.2 beschriebenen Fehlerempfindlichkeit leidet. Die Winkelgeschwindigkeit kann auch hier linear bei Festlegung von \mathbf{v} berechnet werden

$$\boldsymbol{\omega} = \left\{ \iint_D (\mathbf{x} \times (\mathbf{v} \times \mathbf{x})) (\mathbf{x} \times (\mathbf{v} \times \mathbf{x}))^T dx dy \right\}^{-1} \iint_D (\mathbf{v} \times \mathbf{x})^T \dot{\mathbf{x}} (\mathbf{x} \times (\mathbf{v} \times \mathbf{x})) dx dy. \quad (2.32)$$

Unter Festsetzung von $\boldsymbol{\omega}$ wird die Minimierung nach \mathbf{v} ein Eigenvektorproblem

$$\mathbf{v}^T \left\{ \iint_D (\mathbf{x} \times \dot{\mathbf{x}} - \mathbf{x} \times (\boldsymbol{\omega} \times \mathbf{x})) (\mathbf{x} \times \dot{\mathbf{x}} - \mathbf{x} \times (\boldsymbol{\omega} \times \mathbf{x}))^T dx dy \right\} \mathbf{v} \implies \min_{\|\mathbf{v}\|=1}, \quad (2.33)$$

die Lösung ist nämlich der Eigenvektor der Integralmatrix in (2.33), der dem kleinsten Eigenwert entspricht. Das kann in einem zweischrittigen Algorithmus ausgenutzt werden, der beim jeweils ersten Schritt nach der Translation und beim jeweils zweiten nach der Rotation minimiert [Zacharias *et al.* 85].

[Prazdny 81] – und darauf aufbauend [Burger & Bhanu 90] – geht von einer besonderen Interpretation der rotatorischen Komponente der Verschiebungsrate aus. Unter der Annahme, daß ω_x verschwindet, wird die rotatorische Komponente $\dot{\mathbf{x}}_R$ an einem Punkt (x_0, y_0) in zwei weitere Komponenten

$$\begin{aligned} \dot{x}_R &= \dot{x}_R^h + \dot{x}_R^c = (1 + x_0^2)\omega_y - y_0\omega_z \\ \dot{y}_R &= \dot{y}_R^h + \dot{y}_R^c = x_0y_0\omega_y + x_0\omega_z \end{aligned} \quad (2.34)$$

zerlegt, von denen $(\dot{x}_R^h, \dot{y}_R^h)$ tangential zur Hyperbel $y^2/(y_0^2/(x_0^2 + 1)) - x^2 = 1$ steht und $(\dot{x}_R^c, \dot{y}_R^c)$ tangential zum Kreis $x^2 + y^2 = x_0^2 + y_0^2$ steht. Unter der Annahme, daß die Rotationskomponente eine lineare Kombination der zwei Tangentialvektoren ist, bildet man eine Fehlerfunktion, die die Abweichung der resultierenden Translationskomponenten von der Geraden darstellt, die den Bildpunkt mit dem Expansionspunkt verbindet. Für die Bestimmung der anfangs festgesetzten Rotation ω_x wird dann ein Hypothese-Verifikationsverfahren angewandt.

Bewegung einer Ebene

Falls die abgebildete Fläche im Szenenbereich eine Ebene ist, läßt sich das Problem (wie auch im diskreten Fall) eleganter formulieren [Subbarao & Waxman 86; Longuet-Higgins 84; Murray & Buxton 90]. Wir nehmen an, daß eine Ebene im Raum durch die Gleichung $N_x X + N_y Y + N_z Z = 1$ gegeben wird, wobei der Abstand der Ebene vom Ursprung gleich $1/\|\mathbf{N}\|$ ist. Zwischen \mathbf{N} und der Translationsgeschwindigkeit besteht eine Skalierungsmehrdeutigkeit. Daher ist man, wie wir feststellen werden, nur in der

Lage, die Einheitsvektoren \hat{N} , \hat{v} und das Verhältnis $\|\mathbf{N}\|/\|\mathbf{v}\|$ des Translationsbetrags zum Abstand von der Ebene zu bestimmen.

Nach der Ersetzung der Tiefe $1/Z = N_x x + N_y y + N_z$ in der Gleichung (2.18) der Verschiebungsrate erhält man

$$\dot{\mathbf{x}} = (\mathbf{N}^T \mathbf{x})(\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})) + \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})). \quad (2.35)$$

Die Unbekannten kann man zusammenfassen

$$\dot{\mathbf{x}} = \hat{\mathbf{z}} \times ((\mathbf{v}\mathbf{N}^T \mathbf{x} + \boldsymbol{\omega} \times \mathbf{x}) \times \mathbf{x}) \quad (2.36)$$

und als eine Matrix schreiben

$$\dot{\mathbf{x}} = \hat{\mathbf{z}} \times (P \mathbf{x} \times \mathbf{x}) \quad (2.37)$$

mit

$$P = \mathbf{v}\mathbf{N}^T + [\boldsymbol{\omega}]_{\times}, \quad (2.38)$$

wobei $[\boldsymbol{\omega}]_{\times}$ die antisymmetrische Matrix des Kreuzprodukts $[\boldsymbol{\omega}]_{\times} \mathbf{x} = \boldsymbol{\omega} \times \mathbf{x}$ ist. Wegen der Skalierungsmehrdeutigkeit ist die Zahl der eigentlichen Unbekannten acht, was man auch dadurch erkennen kann, daß $(\mathbf{v}/s, s\mathbf{N})$ mit freiem s immer eine Lösung für die Translationsgeschwindigkeit und die Normale ist. Die Elemente der Matrix P sind abhängig voneinander, was man durch die Nebenbedingung $\det(P + P^T) = 0$ ausdrücken kann. Damit wir aber diese in den Elementen von P nichtlineare Nebenbedingung vermeiden, gehen wir anders vor [Negahdaripour & Horn 87]. Wir stellen fest, daß – wenn eine Matrix P die Gleichung (2.37) erfüllt – jede Matrix $P + \rho I$ dieselbe Gleichung erfüllt:

$$\dot{\mathbf{x}} = \hat{\mathbf{z}} \times ((P + \rho I) \mathbf{x} \times \mathbf{x}) \quad (2.39)$$

Die Lösung ist eine einparametrische Familie von Matrizen. Wir suchen folglich nach einer beliebigen Matrix Q , die die Gleichung (2.39) an Stelle von $P + \rho I$ erfüllt. Nach der Ermittlung von Q bestimmen wir ein ρ , so daß für $P = Q - \rho I$ die Nebenbedingung $\det(P + P^T) = 0$ erfüllt wird.

Wir setzen Q an Stelle von $P + \rho I$ in (2.39) ein und schreiben die zwei Skalargleichungen, aus denen die Vektorgleichung (2.39) besteht:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -x^2 & -xy & -x \\ 0 & 0 & 0 & x & y & 1 & -xy & -y^2 & -y \end{pmatrix} \begin{pmatrix} Q_{11} \\ Q_{12} \\ Q_{13} \\ Q_{21} \\ Q_{22} \\ Q_{23} \\ Q_{31} \\ Q_{32} \\ Q_{33} \end{pmatrix}. \quad (2.40)$$

Bei der Verwendung von m Verschiebungsraten besteht die zu invertierende $2m \times 9$ -Matrix aus einer Reihe von m untereinander geschriebenen 2×9 -Matrizen. Wir stellen

fest, daß die neunte Spalte der resultierenden Matrix eine lineare Kombination der ersten und fünften Spalte ist. Daraus folgt, daß der Kern der Matrix den Vektor $(1, 0, 0, 0, 1, 0, 0, 0, 1)^T$ beinhaltet. Die Addition eines Vielfachen dieses Kernvektors zur Lösung ergibt immer eine weitere Lösung. Das überrascht uns aber nicht, denn die dem Kernvektor entsprechende Matrix ist die Einheitsmatrix, die in (2.39) eingeht. Um eine Lösung aus (2.40) zu erhalten, setzen wir das Element $Q_{33} = 0$. Die übrigen acht Unbekannten lassen sich dann aus mindestens vier Verschiebungsraten bestimmen.

Aus der ermittelten Matrix $Q = P + \rho I$ können wir dann eine Matrix P wie folgt bilden, die die Nebenbedingung $\det(P + P^T) = 0$ erfüllt. Definieren wir

$$\Phi = \frac{Q + Q^T}{2} \quad \text{und} \quad \Psi = \frac{P + P^T}{2}, \quad (2.41)$$

so folgt aus $\Phi = \Psi + \rho I$ für die Eigenwerte $\lambda(\Phi)$ von Φ :

$$\lambda(\Phi) = \lambda(\Psi) + \rho.$$

Weil die Matrix Ψ singulär ist, muß einer ihrer Eigenwerte verschwinden – wir werden später zeigen, daß das der mittlere Eigenwert von Ψ ist. Daher kann man ρ gleich dem mittleren Eigenwert von Φ setzen und damit läßt sich die ursprüngliche Matrix P ermitteln.

Wir fahren mit der Untersuchung der Matrix

$$\Psi = \frac{\mathbf{v}\mathbf{N}^T + \mathbf{N}\mathbf{v}^T}{2} \quad (2.42)$$

fort. Die Matrix Ψ ist eine Nullmatrix, wenn die Bewegung rein rotatorisch ist. Im Fall $\mathbf{v} \parallel \mathbf{N}$ hat die Matrix Ψ Rang gleich eins, daher verschwinden zwei ihrer Eigenwerte. Der Eigenvektor, der dem einzigen von Null verschiedenen Eigenwert entspricht, gibt die gesuchte Richtung der Translationsgeschwindigkeit und der Ebenennormalen an. Wenn die Translation nicht parallel zur Ebenennormalen liegt, hat die Matrix Ψ einen Eigenwert $\lambda_2 = 0$, der dem Eigenvektor $\mathbf{u}_2 = \mathbf{v} \times \mathbf{N} / \|\mathbf{v} \times \mathbf{N}\|$ entspricht. Die anderen zwei Eigenvektoren liegen in der Ebene, die von \mathbf{v} und \mathbf{N} aufgespannt wird. Daher können sie als lineare Kombination $\mu\mathbf{v} + \nu\mathbf{N}$ aufgefaßt werden. Wir fordern, daß

$$\frac{1}{2}(\mathbf{v}\mathbf{N}^T + \mathbf{N}\mathbf{v}^T)(\mu\mathbf{v} + \nu\mathbf{N}) = \lambda(\mu\mathbf{v} + \nu\mathbf{N}) \quad (2.43)$$

ist. Damit von Null verschiedene μ und ν existieren, muß

$$\text{Det} \begin{pmatrix} \mathbf{N}^T\mathbf{v} - 2\lambda & \|\mathbf{N}\|^2 \\ \|\mathbf{v}\|^2 & \mathbf{N}^T\mathbf{v} - 2\lambda \end{pmatrix} = 0 \quad (2.44)$$

gelten, woraus sich die zwei anderen Eigenwerte der Matrix Ψ

$$\lambda_{1,3} = \frac{1}{2}(\mathbf{N}^T\mathbf{v} \mp \|\mathbf{N}\|\|\mathbf{v}\|) \quad (2.45)$$

herleiten lassen. Der Eigenwert λ_1 ist negativ, daher gilt $\lambda_1 < \lambda_2 < \lambda_3$.

Die entsprechenden Eigenvektoren mit Norm gleich eins lauten:

$$\begin{aligned} \mathbf{u}_1 &= \frac{\hat{\mathbf{v}} - \hat{\mathbf{N}}}{\sqrt{2(1 - \hat{\mathbf{v}}^T \hat{\mathbf{N}})}} \\ \mathbf{u}_3 &= \frac{\hat{\mathbf{v}} + \hat{\mathbf{N}}}{\sqrt{2(1 + \hat{\mathbf{v}}^T \hat{\mathbf{N}})}} \end{aligned} \quad (2.46)$$

Daraus folgt

$$\begin{aligned} \hat{\mathbf{v}} &= \sqrt{\frac{1}{2}(1 + \hat{\mathbf{v}}^T \hat{\mathbf{N}})} \mathbf{u}_3 + \sqrt{\frac{1}{2}(1 - \hat{\mathbf{v}}^T \hat{\mathbf{N}})} \mathbf{u}_1 \\ \hat{\mathbf{N}} &= \sqrt{\frac{1}{2}(1 + \hat{\mathbf{v}}^T \hat{\mathbf{N}})} \mathbf{u}_3 - \sqrt{\frac{1}{2}(1 - \hat{\mathbf{v}}^T \hat{\mathbf{N}})} \mathbf{u}_1 \\ \hat{\mathbf{v}}^T \hat{\mathbf{N}} &= \frac{\lambda_3 + \lambda_1}{\lambda_3 - \lambda_1}. \end{aligned} \quad (2.47)$$

Wie schon erwähnt worden ist, kann man aus den gemessenen Verschiebungsraten die Elemente der Matrix Q bestimmen. Damit bildet man die Matrix $\Phi = (Q + Q^T)/2$ und berechnet ihre Eigenwerte $\lambda(\Phi)$. Jeder Eigenvektor der Matrix Φ ist parallel zum entsprechenden Eigenvektor der Matrix Ψ . Mit den obengenannten Aussagen über die Eigenwerte der Matrix Ψ unterscheidet man drei Fälle für die Eigenwerte der Matrix Φ :

1. Alle Eigenwerte von Φ fallen zusammen, woraus folgt, daß es sich um eine reine Rotationsbewegung handelt, wobei \mathbf{N} nicht bestimmbar ist.
2. Zwei Eigenwerte von Φ fallen zusammen ($\lambda_1(\Phi) = \lambda_2(\Phi) < \lambda_3(\Phi)$ oder $\lambda_1(\Phi) < \lambda_2(\Phi) = \lambda_3(\Phi)$), woraus folgt, daß die Translation parallel zur Ebenennormalen ist. Der Eigenvektor, der dem Eigenwert entspricht, der verschieden von den beiden übrigen ist, stellt die Lösung für die gemeinsame Richtung der Ebenennormalen und der Translationsgeschwindigkeit dar.
3. Alle drei Eigenwerte sind verschieden voneinander. Der mittlere Eigenwert $\lambda_2(\Phi)$ der Matrix Φ ist gleich ρ . Daraus kann man die übrigen Eigenwerte der Matrix Ψ $\lambda_{1,3}(\Psi) = \lambda_{1,3}(\Phi) - \lambda_2(\Phi)$ bestimmen. Die Eigenvektoren der Matrizen Φ und Ψ sind identisch, daher kann man weiter die Formeln (2.47) benutzen, um die gesuchten Einheitsvektoren $\hat{\mathbf{N}}$ der Ebenennormalen sowie der Translationsrichtung $\hat{\mathbf{v}}$ zu bestimmen.

Aus der willkürlichen Festsetzung der Vorzeichen der Eigenvektoren in (2.47) sieht man direkt, daß sich insgesamt vier Kombinationen ergeben, von denen zwei unter der Bedingung, daß die Ebene sichtbar sein soll, abgelehnt werden können. Die übrigen zwei sind (\mathbf{v}, \mathbf{N}) und (\mathbf{N}, \mathbf{v}) , und die Existenz solch einer Duallösung kann man aus der Struktur der Matrix Ψ sofort erkennen. Die zur zweiten Lösung (\mathbf{N}, \mathbf{v}) entsprechende Winkelgeschwindigkeit lautet $\boldsymbol{\omega} + \mathbf{N} \times \mathbf{v}$. Die Existenz von zwei Lösungen wurde wiederholt nachgewiesen [Hay 66; Longuet-Higgins 84; Maybank 85; Negahdaripour & Horn 87].

3D-Interpretation aus lokalen Eigenschaften

In diesem Abschnitt werden wir beschreiben, wie man aus der Taylor-Entwicklung der Verschiebungsrates auf eine 3D-Interpretation gelangen kann. Es wird angenommen, daß ein kleines Flächenelement im Szenenbereich um die optische Achse durch eine quadratische Fläche approximiert werden kann

$$Z = Z_0 + Z_x X + Z_y Y + \frac{1}{2} Z_{xx} X^2 + Z_{xy} XY + \frac{1}{2} Z_{yy} Y^2 + \mathcal{O}_{i+j=3}(X^i Y^j). \quad (2.48)$$

Daraus folgt für das Inverse der Tiefe

$$\frac{1}{Z} = \frac{1}{Z_0} \left(1 - Z_x x - Z_y y - Z \left(\frac{1}{2} Z_{xx} x^2 + Z_{xy} xy + \frac{1}{2} Z_{yy} y^2 \right) + \mathcal{O}_{i+j=3}(X^i Y^j) \right) \quad (2.49)$$

und nach Wiederersetzung von Z

$$\frac{1}{Z} = \frac{1}{Z_0} \left(1 - Z_x x - Z_y y - Z_0 \left(\frac{1}{2} Z_{xx} x^2 + Z_{xy} xy + \frac{1}{2} Z_{yy} y^2 \right) + \mathcal{O}_{i+j=3}(x^i y^j) \right). \quad (2.50)$$

Wir setzen die inversen Tiefen in die Komponenten der Verschiebungsrates (2.19) ein, wobei Terme dritter und höherer Ordnung $\mathcal{O}_{i+j=3}(x^i y^j)$ vernachlässigt werden. Der Symbolsparsamkeit halber behalten wir die Symbole $(v_x, v_y, v_z, Z_{xx}, Z_{xy}, Z_{yy})$ für die skalierten Größen $(v_x/Z_0, v_y/Z_0, v_z/Z_0, Z_0 Z_{xx}, Z_0 Z_{xy}, Z_0 Z_{yy})$ bei:

$$\begin{aligned} \dot{x} &= (v_x + \omega_y) + (-v_x Z_x - v_z)x + (-v_x Z_y - \omega_z)y \\ &\quad + \left(-\frac{1}{2}v_x Z_{xx} + v_z Z_x + \omega_y\right)x^2 + (-v_x Z_{xy} + v_z Z_y - \omega_x)xy + \left(-\frac{1}{2}v_x Z_{yy}\right)y^2 \\ \dot{y} &= (v_y - \omega_x) + (-v_y Z_x + \omega_z)x + (-v_y Z_y - v_z)y \\ &\quad + \left(-\frac{1}{2}v_y Z_{xx}\right)x^2 + (-v_y Z_{xy} + v_z Z_x + \omega_y)xy + \left(-\frac{1}{2}v_y Z_{yy} + v_z Z_y - \omega_x\right)y^2. \end{aligned} \quad (2.51)$$

Aus den zwölf Koeffizienten der Taylor-Entwicklung entsteht ein nichtlineares Gleichungssystem mit den elf Unbekannten $(\mathbf{v}, \boldsymbol{\omega}, Z_x, Z_y, Z_{xx}, Z_{xy}, Z_{yy})$. Methoden für die Ermittlung der Unbekannten finden sich in [Longuet-Higgins & Prazdny 80; Waxman *et al.* 87]. [Subbarao 88b] hat die Mehrdeutigkeit in der lokalen Schätzung der 3D-Strukturparameter untersucht und den lokalen Ansatz auf die Berücksichtigung der zeitlichen Ableitungen der Verschiebungsrates [Subbarao 89] erweitert.

Von Bedeutung sind die Hinweise der differentiellen Invarianten der Verschiebungsrates auf das 3D-Geschehen: Die *Divergenz* und die *Rotation*, die jeweils die lokale isotrope Expansion oder Kontraktion und die lokale Drehung der Umgebung des Ursprungs angeben

$$\begin{aligned} \operatorname{div}(\dot{\mathbf{x}})|_{x=y=0} &= \dot{x}_x + \dot{y}_y = -2v_z - v_x Z_x - v_y Z_y \\ \operatorname{curl}(\dot{\mathbf{x}})|_{x=y=0} &= \dot{y}_x - \dot{x}_y = 2\omega_z + v_x Z_y - v_y Z_x \end{aligned}$$

und die zwei Komponenten der *Scherung*, die eine Kontraktion entlang einer Richtung und eine Expansion entlang der Richtung senkrecht zur ersten Richtung darstellen

$$\operatorname{shear}(\dot{\mathbf{x}})|_{x=y=0} = (\dot{x}_x - \dot{y}_y, \dot{x}_y + \dot{y}_x) = (-v_x Z_x + v_y Z_y, -v_x Z_y - v_y Z_x).$$

[Subbarao 88a] hat gezeigt, daß die differentiellen Invarianten der Verschiebungsraten Schranken für die skalierte z -Komponente der Translations- und die z -Komponente der Winkelgeschwindigkeit angeben:

$$\begin{aligned} v_z^{max/min} &= \frac{1}{2}(-div \pm \|shear\|) \\ \omega_z^{max/min} &= \frac{1}{2}(curl \pm \|shear\|), \end{aligned}$$

während der exakte Wert von v_z und ω_z nur unter Benutzung aller Koeffizienten der Taylor-Entwicklung der Verschiebungsraten berechnet werden kann.

Direkte Berechnung der 3D-Bewegung und Struktur aus dem Grauwertbild

Alle bis zu diesem Punkt beschriebenen Ansätze unterstellen die frühere Berechnung von Punktkorrespondenzen oder Verschiebungsraten. Die sogenannten *direkten* Ansätze gehen direkt von der Korrespondenz von Grauwertbereichen aus, die den Abbildungen von relativ zur Kamera bewegten starren Szenenkomponenten entsprechen. Nach Ersetzung von (2.18) in (2.1) erhält man

$$\nabla g^T \left(\hat{\mathbf{z}} \times \left(\frac{1}{Z}(\mathbf{v} \times \mathbf{x}) + \mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega}) \right) \right) + g_t = 0$$

und nach einer Zusammenfassung der Datenterme

$$\frac{1}{Z} \mathbf{v}^T \mathbf{A} + \boldsymbol{\omega}^T \mathbf{B} + g_t = 0 \quad \text{mit} \quad \mathbf{A} = \mathbf{x} \times (\nabla g \times \hat{\mathbf{z}}) \quad \text{und} \quad \mathbf{B} = \mathbf{x} \times \mathbf{A}. \quad (2.52)$$

Man erkennt, daß man für jeden Bildpunkt eine neue Gleichung, aber auch eine neue Unbekannte – die Tiefe – hat. Ohne a priori Wissen über die Struktur der Szene ist man nicht in der Lage, die Bewegungsparameter eindeutig zu bestimmen, es sei denn, es handelt sich um eine rein rotatorische Bewegung. Unter der Annahme, daß die abgebildete Umgebung eine Ebene ist, ergeben sich dieselben Parameter (Matrix P) wie in (2.37) [Negahdaripour & Horn 87; Sull & Ahuja 91b]. [Horn & Weldon 88] untersuchen alle Möglichkeiten für das a priori Wissen, damit die aus (2.52) folgende Minimierung

$$\iint_D \left(\frac{1}{Z} \mathbf{v}^T \mathbf{A} + \boldsymbol{\omega}^T \mathbf{B} + g_t \right)^2 dx dy \implies \min \quad (2.53)$$

eine endliche Anzahl von Lösungen hat.

Von Bedeutung ist ihr Ansatz für den Fall der reinen Translationsbewegung oder bekannter Rotation, wobei sie die Tatsache, daß die Tiefen positiv sind, ausnutzen, um den Bereich auf der Kugel $\|\mathbf{v}\| = 1$ zu bestimmen, in dem alle Ungleichungen

$$\frac{\mathbf{v}^T \mathbf{A}}{g_t} < 0 \quad (2.54)$$

erfüllt werden.

Läßt man die Starrheitsannahme nicht gelten, so hat man für jeden Punkt sechs Unbekannte (fünf Bewegungsparameter und die Tiefe). [Krämer 89] hat einen Regularisierungsansatz für den Fall der reinen Translation angewandt, indem er gefordert

hat, daß die Tiefenvariation entlang der Kanten und die räumlichen Ableitungen der Translation gering bleiben. Damit man auf eine gleichzeitige Segmentierung in relative Bewegungen gelangt, soll man eine Regularisierung auf die Translation unter Berücksichtigung von Diskontinuitäten anwenden.

Die Überwindung der schlechten Stellung des Problems bei der direkten Auswertung von Grauwertbildfolgen zur Bewegungsbestimmung hat zu mehreren Ansätzen mit verschiedenen Eingaben und Annahmen geführt – siehe dazu die Ansätze der *linearen Merkmale* bei [Amari & Maruyama 87], die Benutzung von Stereobildfolgen bei [Basu & Aloimonos 90; Aloimonos & Herve 90; Szeliski 88; Hayashi & Negahdaripour 90] oder Abstandsdaten [Horn & Harris 91].

2.4 3D-Bewegungsschätzung aus Geradenkorrespondenzen

Im Abschnitt 2.2 haben wir die Problemstellung und die existierenden Ansätze für den Fall beschrieben, daß die aus dem Grauwertbild extrahierten Merkmale Punkte sind. In zahlreichen Anwendungen besteht der Szenenbereich aus polyedrischen Komponenten, deren Abbildung im Bildbereich aus geradlinigen Kantensegmenten besteht. Die Endpunkte dieser Kantensegmente sind nach der Stufe der Kantenverfolgung und Geradenanpassung mit hoher Meßunsicherheit behaftet. So ist die Lage der Endpunkte zwar für ihre zeitliche Verfolgung von Bedeutung, sie ist aber für eine 3D-Interpretation nicht geeignet. Es ist naheliegend, daß man die Geradensegmente im Bildbereich geometrisch als Geraden mit unendlicher Ausdehnung auffaßt, die Projektionen von Geraden im 3D-Szenenbereich sind. Wir werden in diesem Abschnitt beschreiben, wie man aus solchen Geradenkorrespondenzen Schlüsse auf die 3D-Bewegung und die Lage der Geraden im Raum ziehen kann.

Gegeben sind m Geraden $L_{i=1..m}$ im Raum, die von einer Kamera in Zeitpunkten t_0 und t_1 aufgenommen werden (Abb. 2.3). Es ist anschaulicher, anstatt der projizierten Geraden l_{ij} im Bildbereich die *Sehebene* π_{ij} als Messung aufzufassen, die vom Augenpunkt O_j und der Geraden l_{ij} aufgespannt wird. Diese Ebene entspricht dem Sehstrahl im Fall von Punkten. Es kann geometrisch nachgewiesen werden (siehe [Liu & Huang 88]), daß zwei Aufnahmen für eine 3D-Interpretation nicht reichen. Tatsächlich kann man in Abb. 2.3 die Sehebenekonfiguration des zweiten Augenpunkts beliebig starr – d.h. dabei bleibt der Winkel zwischen π_{11} und π_{21} unverändert – transformieren, so daß die projizierten Geraden in gleicher Position im Bild bleiben, obwohl die 3D-Lage der Geraden im Raum sich verändert. Vergleicht man dazu den Fall von Punktkorrespondenzen (Abb. 2.1), so kann man feststellen, daß die Sehstrahlen von erster und zweiter Aufnahme windschief zueinander stehen, sobald man die zweite Sehstrahlkonfiguration starr verzerrt. Der Ausweg aus der Mehrdeutigkeit bei der Auswertung von zwei Aufnahmen ergibt sich durch Nutzung einer dritten Aufnahme. Wenn man beim Vorhandensein von drei Aufnahmen versucht, die zweite und die dritte Sehebenekonfiguration starr zu verschieben, erhält man keine neue Geradenkonfiguration im Raum, weil sich drei Ebenen im allgemeinen nicht in einer Geraden schneiden.

Wir bezeichnen mit (\mathbf{l}, \mathbf{d}) die Plücker'schen Vektoren einer Geraden im \mathcal{R}^3 . Der Leser findet eine ausführliche Beschreibung dieser Darstellung im Anhang C. Zu jedem Zeitpunkt t_i bezeichnen wir mit \mathbf{n}_i die Normale der Sehebene π . Diese Normale kann

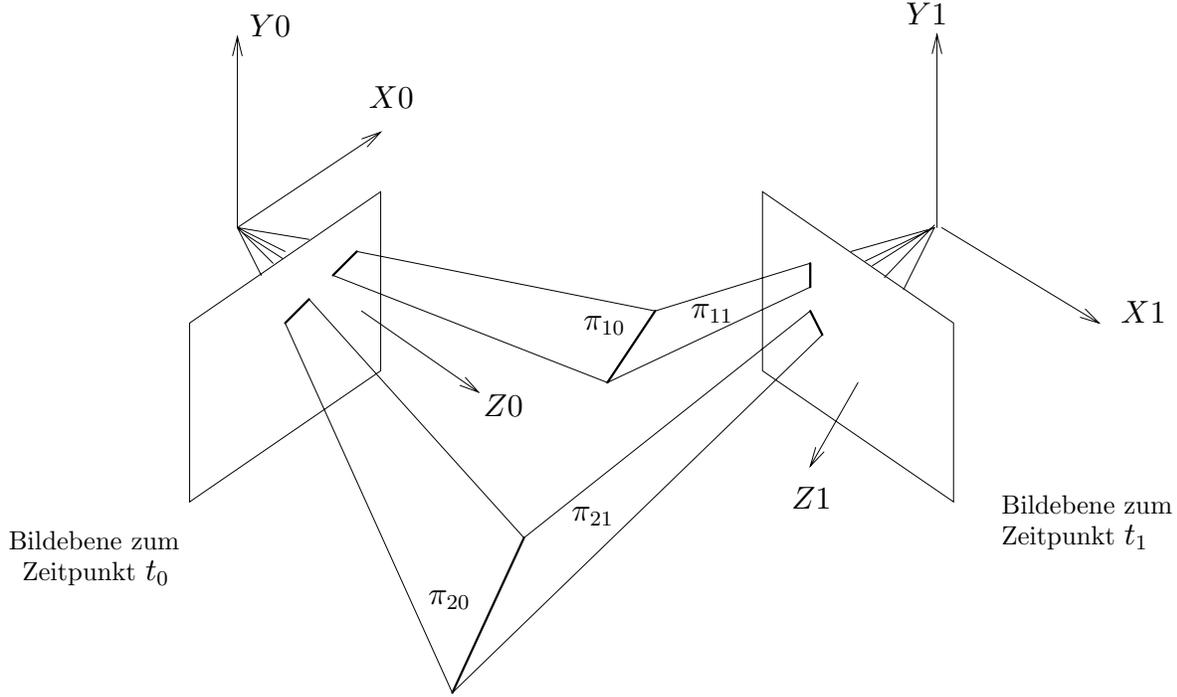


Abbildung 2.3: Geradenkorrespondenzen in einem Bildpaar

direkt aus einer gemessenen Geraden

$$ax + by + c = 0 \quad (2.55)$$

in der Bildebene hergeleitet werden. So lautet unsere Meßgleichung

$$\mathbf{n} = \frac{\mathbf{d}}{\|\mathbf{d}\|}, \quad (2.56)$$

womit die Lage der Geraden im Raum auf eine Ebene $\mathbf{l}^T \mathbf{n} = 0$ eingeschränkt wird. Die Bedingung (2.56) bildet zwei unabhängige Gleichungen für \mathbf{d} . Durch Aufzählung der Meßgleichungen erhalten wir für m Geraden in n Aufnahmen $2mn$ Gleichungen. Ihre Anzahl soll größer gleich der Anzahl der Unbekannten sein, die aus $(4m - 1)$ Variablen für die Struktur – der Term -1 steht für den nicht bestimmmbaren Skalierungsfaktor – und $6(n - 1)$ Variablen für die Bewegungen zwischen den Aufnahmen besteht. An der Ungleichung

$$2mn \geq (4m - 1) + 6(n - 1)$$

kann man direkt erkennen, daß mit zwei Aufnahmen keine 3D-Interpretation ermittelt werden kann, während man bei drei Aufnahmen mindestens sechs Korrespondenzen braucht.

Wir bezeichnen mit A_r und B_r die Rotationen vom Zeitpunkt t_0 auf jeweils t_1 und t_2 und entsprechend mit \mathbf{a}_t und \mathbf{b}_t die Translationen. Die Plücker'schen Vektoren transformieren sich wie folgt (Anhang C)

$$\mathbf{l}_1 = A_r \mathbf{l}_0 \quad \mathbf{d}_1 = A_r \mathbf{d}_0 + \mathbf{a}_t \times A_r \mathbf{l}_0 \quad (2.57)$$

$$\mathbf{l}_2 = B_r \mathbf{l}_0 \quad \mathbf{d}_2 = B_r \mathbf{d}_0 + \mathbf{b}_t \times B_r \mathbf{l}_0. \quad (2.58)$$

Aus diesen Gleichungen versucht man die Struktur, die durch \mathbf{l} und $\|\mathbf{d}\|$ ausgedrückt wird, zu eliminieren. Man multipliziert (2.57) und (2.58) mit jeweils A_r^T und B_r^T

$$A_r^T \mathbf{d}_1 = \mathbf{d}_0 + A_r^T \mathbf{a}_t \times \mathbf{l}_0 \quad (2.59)$$

$$B_r^T \mathbf{d}_2 = \mathbf{d}_0 + B_r^T \mathbf{b}_t \times \mathbf{l}_0 \quad (2.60)$$

und bildet das Vektorprodukt zwischen deren linken und rechten Seiten

$$A_r^T \mathbf{d}_1 \times B_r^T \mathbf{d}_2 = \mathbf{d}_0 \times (B_r^T \mathbf{b}_t \times \mathbf{l}_0) + (A_r^T \mathbf{a}_t \times \mathbf{l}_0) \times \mathbf{d}_0 + (A_r^T \mathbf{a}_t \times \mathbf{l}_0) \times (B_r^T \mathbf{b}_t \times \mathbf{l}_0). \quad (2.61)$$

Darauf kann man das Skalarprodukt mit \mathbf{d}_0 bilden und die Tatsache $\mathbf{l}_0^T \mathbf{d}_0 = 0$ ausnutzen:

$$\mathbf{d}_0^T (A_r^T \mathbf{d}_1 \times B_r^T \mathbf{d}_2) = 0. \quad (2.62)$$

Mittels der Meßgleichung (2.56) und nach Division durch die Beträge der Momentvektoren gelangt man auf die struktur- sowie auch translationsfreie Beziehung

$$\mathbf{n}_0^T (A_r^T \mathbf{n}_1 \times B_r^T \mathbf{n}_2) = 0. \quad (2.63)$$

Die Gleichung ist nichtlinear bezüglich jeder möglichen Rotationsrepräsentation. Es genügen sechs Geradenkorrespondenzen zur Lösung nach den sechs unabhängigen Rotationsfreiheitsgraden.

Zur Ermittlung einer Gleichung für die translatorischen Parameter bildet man das Vektorprodukt von rechts mit $A_r^T \mathbf{d}_1$ an (2.59) und mit $B_r^T \mathbf{d}_2$ an (2.60) und erhält

$$\mathbf{d}_0 \times A_r^T \mathbf{d}_1 + \mathbf{d}_1^T \mathbf{a}_t \mathbf{l}_0 = 0 \quad (2.64)$$

$$\mathbf{d}_0 \times B_r^T \mathbf{d}_2 + \mathbf{d}_2^T \mathbf{b}_t \mathbf{l}_0 = 0. \quad (2.65)$$

Durch Gebrauch der Meßgleichung (2.56) und Eliminierung des Richtungsvektors \mathbf{l}_0 folgt

$$\mathbf{n}_2^T \mathbf{b}_t (\mathbf{n}_0 \times A_r^T \mathbf{n}_1) = \mathbf{n}_1^T \mathbf{a}_t (\mathbf{n}_0 \times B_r^T \mathbf{n}_2). \quad (2.66)$$

Man erkennt sofort die Skalierungsmehrdeutigkeit durch die homogene Erscheinung der Translationen \mathbf{a}_t und \mathbf{b}_t . Vektorgleichung (2.66) beinhaltet eine unabhängige Gleichung in Bezug auf $\mathbf{n}_1^T \mathbf{a}_t$ und $\mathbf{n}_2^T \mathbf{b}_t$. Zum Nachweis kann man (2.66) wie folgt umschreiben

$$\begin{pmatrix} \mathbf{n}_0 \times A_r^T \mathbf{n}_1 & -\mathbf{n}_0 \times B_r^T \mathbf{n}_2 \end{pmatrix} \begin{pmatrix} \mathbf{n}_2^T \mathbf{b}_t \\ \mathbf{n}_1^T \mathbf{a}_t \end{pmatrix} = 0. \quad (2.67)$$

Der Rang der Matrix auf der linken Seite ist gleich eins, wenn (2.63) gilt. Wenn man die Beträge der linken und rechten Seite von (2.66) gleich setzt und eine Freiheit in der Wahl der Vorzeichen erlaubt, erhält man

$$\mathbf{n}_2^T \mathbf{b}_t \sin(\angle(\mathbf{n}_0, A_r^T \mathbf{n}_1)) = \mathbf{n}_1^T \mathbf{a}_t \sin(\angle(\mathbf{n}_0, B_r^T \mathbf{n}_2)). \quad (2.68)$$

Daher benötigt man zur Bestimmung der Translation fünf Geradenkorrespondenzen.

Die entscheidenden Gleichungen (2.63) und (2.66) haben wir einheitlich mit Hilfe der Plücker'schen Vektoren hergeleitet. Andere Herleitungen, die verschiedene Darstellungen benutzen, finden sich in [Faugeras *et al.* 87b; Liu & Huang 88; Weng *et al.* 90; Spetsakis

& Aloimonos 90]. Insbesondere [Faugeras *et al.* 87b] gelangen auf (2.63) und (2.68), indem sie das rein geometrische Argument anwenden, daß sich die drei Sehebene $\pi_{i=0..2}$ in einer Geraden schneiden sollen. Die drei Sehebene sind:

$$\mathbf{n}_0^T \mathbf{X}_0 = 0 \quad \mathbf{n}_1^T \mathbf{X}_1 = 0 \quad \mathbf{n}_2^T \mathbf{X}_2 = 0.$$

Weiterhin gilt $\mathbf{X}_1 = A_r \mathbf{X}_0 + \mathbf{a}_t$ und $\mathbf{X}_2 = B_r \mathbf{X}_0 + \mathbf{b}_t$. Durch Ausdruck der drei Ebenen in Bezug auf das Koordinatensystem der Aufnahme zum Zeitpunkt t_0 und Weglassen des Indizes $\mathbf{X} = \mathbf{X}_0$ erhält man die drei Ebenengleichungen

$$\begin{aligned} \mathbf{n}_0^T \mathbf{X} &= 0 \\ \mathbf{n}_1^T A_r \mathbf{X} &= -\mathbf{n}_1^T \mathbf{a}_t \\ \mathbf{n}_2^T B_r \mathbf{X} &= -\mathbf{n}_2^T \mathbf{b}_t, \end{aligned}$$

die sich zum Gleichungssystem

$$\begin{pmatrix} \mathbf{n}_0^T \\ \mathbf{n}_1^T A_r \\ \mathbf{n}_2^T B_r \end{pmatrix} \mathbf{X} = \begin{pmatrix} 0 \\ -\mathbf{n}_1^T \mathbf{a}_t \\ -\mathbf{n}_2^T \mathbf{b}_t \end{pmatrix},$$

zusammenfassen lassen. Durch die Forderung nach einer einparametrischen Lösung (Gerade) für \mathbf{X} beweisen [Faugeras *et al.* 87b], daß sich daraus die zwei unabhängigen Gleichungen (2.63) und (2.68) ergeben.

Auf der Suche nach essentiellen Parametern – wie im Fall von Punktkorrespondenzen [Weng *et al.* 90] und [Spetsakis & Aloimonos 90] – gehen wir wie folgt vor. Seien die Spaltenvektoren der Rotationsmatrizen

$$A_r = (A_{r1} \ A_{r2} \ A_{r3}) \quad \text{und} \quad B_r = (B_{r1} \ B_{r2} \ B_{r3}),$$

so kann man die Gleichung (2.66) wie folgt schreiben

$$\mathbf{n}_0 \times \begin{pmatrix} \mathbf{n}_1^T E_1^l \mathbf{n}_2 \\ \mathbf{n}_1^T E_2^l \mathbf{n}_2 \\ \mathbf{n}_1^T E_3^l \mathbf{n}_2 \end{pmatrix} = 0 \quad \text{mit} \quad E_k^l = \mathbf{a}_t B_{rk}^T - A_{rk} \mathbf{b}_t^T \quad k = 1, 2, 3. \quad (2.69)$$

Vektorgleichung (2.69) beinhaltet zwei skalare Gleichungen, die homogen linear in Bezug auf die 27 essentiellen Parameter – Elemente der Matrizen (E_1^l, E_2^l, E_3^l) – sind. Zu ihrer Bestimmung braucht man *dreizehn* Geradenkorrespondenzen (entsprechend den acht Punkten im Fall der Punktkorrespondenzen). In beiden obengenannten Ansätzen werden Methoden zur Herleitung der Bewegungsparameter aus den essentiellen Parametern beschrieben, aber keine Bedingungen für Zerlegbarkeit gestellt.

Die Untersuchung der Mehrdeutigkeit sowie auch die Ermittlung der Anzahl der Lösungen im Fall von sechs Geradenkorrespondenzen stellen bis heute ungelöste Probleme dar. Weiterhin muß man ein geometrisch interpretierbares Fehlermaß für die Minimierung einer Kostenfunktion bei Vorhandensein von mehr als sechs Geradenkorrespondenzen festlegen (siehe die in Kapitel 3 behandelte Problematik).

2.5 Bewegungsschätzung aus den Verschiebungsraten von Geraden

In diesem Abschnitt lassen wir die Zeitintervalle infinitesimal werden und betrachten anstatt der Korrespondenzen $(\mathbf{n}_0, \mathbf{n}_1, \mathbf{n}_2)$ die Projektion der Geraden zu einem Zeitpunkt, darüberhinaus aber auch ihre Geschwindigkeit und ihre Beschleunigung $(\mathbf{n}, \dot{\mathbf{n}}, \ddot{\mathbf{n}})$. Dabei unterstellen wir wieder, daß wir aus den gemessenen zwei Koordinaten der Geraden den Einheitsvektor der Sehebene π berechnen. Dieselbe Berechnung führen wir für die zeitlichen Ableitungen durch. Weil sich der Einheitsvektor \mathbf{n} auf der Einheitskugel bewegt, steht seine Geschwindigkeit senkrecht auf \mathbf{n} . Es gilt

$$\dot{\mathbf{n}}^T \mathbf{n} = 0 \quad \text{und} \quad \ddot{\mathbf{n}}^T \mathbf{n} + \dot{\mathbf{n}}^T \dot{\mathbf{n}} = 0. \quad (2.70)$$

Für die Geschwindigkeit der Geraden im Raum gilt (siehe Anhang C)

$$\begin{aligned} \dot{\mathbf{l}} &= \boldsymbol{\omega} \times \mathbf{l} \\ \dot{\mathbf{d}} &= \mathbf{v} \times \mathbf{l} + \boldsymbol{\omega} \times \mathbf{d}. \end{aligned} \quad (2.71)$$

Die Verschiebungsrates der Geraden im Bild läßt sich durch Differenzierung von (2.56) herleiten. Bezeichnen wir den Abstand der Geraden vom Projektionszentrum mit $d_l = \|\mathbf{d}\|$, so ergibt sich für die Geschwindigkeit

$$\dot{\mathbf{n}} = \frac{1}{d_l} (\mathbf{n} \times (\dot{\mathbf{d}} \times \mathbf{n})). \quad (2.72)$$

Wir ersetzen $\dot{\mathbf{d}}$ und erhalten

$$\dot{\mathbf{n}} = \frac{1}{d_l} (\mathbf{n}^T \mathbf{v} (\mathbf{n} \times \mathbf{l})) + \boldsymbol{\omega} \times \mathbf{n}. \quad (2.73)$$

Wir stellen ähnliche Zusammenhänge wie im Fall des Bewegungsfeldes von Punkten fest, nämlich, daß die Verschiebungsrates der Geraden einen translatorischen und einen rotatorischen Anteil besitzt. Der translatorische Anteil enthält auch die gesamte Strukturinformation der Geradenrichtung und des Abstands der Geraden vom Ursprung. Beide Anteile verschwinden, wenn sich die Gerade nur in der Sehebene bewegt ($\mathbf{v}^T \mathbf{n} = 0$ und $\boldsymbol{\omega} \parallel \mathbf{n}$ für $\boldsymbol{\omega} \neq \mathbf{0}$). Bei bekannter Bewegung kann man aus (2.73) den Abstand der Geraden vom Projektionszentrum berechnen

$$d_l = \frac{\mathbf{n}^T \mathbf{v}}{(\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega})^T (\mathbf{n} \times \mathbf{l})}, \quad (2.74)$$

wobei man \mathbf{l} und \mathbf{n} als Einheitsvektoren erhält sowie auch die Geradenrichtung, indem man das Vektorprodukt in (2.73) mit \mathbf{n} bildet

$$\mathbf{l} = -\frac{\mathbf{n} \times (\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega})}{\|\mathbf{n} \times (\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega})\|} = -\frac{\mathbf{n} \times (\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega})}{\|\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega}\|}. \quad (2.75)$$

Im Fall von bekannter Struktur (\mathbf{l}, d_l) stellt man fest, daß die Geschwindigkeiten linear in der Vektorgleichung (2.73) auftreten. Zuerst vermutet man, daß sich aus

der Struktur, der Projektion und der Verschiebungsrate von zwei Geraden die sechs⁴ Unbekannten von \mathbf{v} und $\boldsymbol{\omega}$ berechnen lassen. Jedoch ist die Matrix des entstehenden linearen Systems

$$\begin{pmatrix} 1/d_{l1}(\mathbf{n}_1 \times \mathbf{l}_1)\mathbf{n}_1^T & -[\mathbf{n}_1]_{\times} \\ 1/d_{l2}(\mathbf{n}_2 \times \mathbf{l}_2)\mathbf{n}_2^T & -[\mathbf{n}_2]_{\times} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{n}}_1 \\ \dot{\mathbf{n}}_2 \end{pmatrix}$$

nicht regulär. Daher braucht man Information von drei Geraden, um die Geschwindigkeiten im Fall von bekannter Struktur zu bestimmen.

Bei unbekannter Struktur und Bewegung wird ihre Berechnung unmöglich – wie auch im Fall von Geradenkorrespondenzen aus einem Bildpaar. Um die Information von einem dritten Bild miteinzubeziehen, sind wir gezwungen, die Beschleunigung der Projektion einer Geraden im Bild zu berechnen.

Für die Beschleunigung der Parameter der Geraden im Raum gilt

$$\begin{aligned} \ddot{\mathbf{l}} &= \dot{\boldsymbol{\omega}} \times \mathbf{l} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{l}) \\ \ddot{\mathbf{d}} &= \dot{\mathbf{v}} \times \mathbf{l} + \mathbf{v} \times (\boldsymbol{\omega} \times \mathbf{l}) + \dot{\boldsymbol{\omega}} \times \mathbf{d} + \boldsymbol{\omega} \times (\mathbf{v} \times \mathbf{l} + \boldsymbol{\omega} \times \mathbf{d}). \end{aligned} \quad (2.76)$$

Damit wir eine strukturfreie Formel für die Bewegungsparameter erhalten, gehen wir von der geometrischen Beziehung $\mathbf{n}^T \mathbf{l} = 0$ aus, die wir zweimal nach der Zeit differenzieren, wobei wir die Ableitungen von \mathbf{l} mit den entsprechenden Termen aus (2.71) und (2.76) ersetzen

$$(\ddot{\mathbf{n}} + 2\dot{\mathbf{n}} \times \boldsymbol{\omega} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{n}) + \mathbf{n} \times \dot{\boldsymbol{\omega}})^T \mathbf{l} = 0. \quad (2.77)$$

Wir ersetzen \mathbf{l} mit Hilfe von (2.75) und nach Multiplikation mit $\|\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega}\|$ erhalten wir folgende Gleichung

$$\begin{aligned} (\ddot{\mathbf{n}} + 2\dot{\mathbf{n}} \times \boldsymbol{\omega} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{n}) + \mathbf{n} \times \dot{\boldsymbol{\omega}})^T (\mathbf{n} \times (\dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega})) &= 0 \Leftrightarrow \\ \ddot{\mathbf{n}}^T (\mathbf{n} \times \dot{\mathbf{n}}) + 3\ddot{\mathbf{n}}^T \mathbf{n} \mathbf{n}^T \boldsymbol{\omega} - \ddot{\mathbf{n}}^T \boldsymbol{\omega} + 3\boldsymbol{\omega}^T (\mathbf{n} \times \dot{\mathbf{n}}) \mathbf{n}^T \boldsymbol{\omega} \\ + (\boldsymbol{\omega}^T \mathbf{n})^3 - \boldsymbol{\omega}^T \boldsymbol{\omega} \mathbf{n}^T \boldsymbol{\omega} + \dot{\boldsymbol{\omega}}^T \dot{\mathbf{n}} - \boldsymbol{\omega}^T (\mathbf{n} \times \dot{\boldsymbol{\omega}}) &= 0, \end{aligned} \quad (2.78)$$

die dritten Grades bezüglich der Winkelgeschwindigkeit und linear bezüglich der Winkelbeschleunigung ist. Um die sechs Unbekannten zu bestimmen, braucht man die Geschwindigkeiten und Beschleunigungen im Bild von mindestens sechs Geraden.

Zur Ermittlung einer Beziehung, die die translatorischen Geschwindigkeiten und Beschleunigungen enthält, gehen wir von (2.74) aus und führen folgende Bezeichnungen ein

$$\boldsymbol{\alpha} = \mathbf{n} \times \mathbf{l} \quad \text{und} \quad \boldsymbol{\beta} = \dot{\mathbf{n}} + \mathbf{n} \times \boldsymbol{\omega}. \quad (2.79)$$

Dann gilt

$$d_l \boldsymbol{\beta}^T \boldsymbol{\alpha} = \mathbf{n}^T \mathbf{v}. \quad (2.80)$$

Aus

$$\dot{\mathbf{d}} = \dot{d}_l \mathbf{n} + d_l \dot{\mathbf{n}} \quad (2.81)$$

und (2.70) folgt

$$\dot{d}_l = \dot{\mathbf{d}}^T \mathbf{n}, \quad (2.82)$$

⁴Bei bekannter Struktur gibt es keine Skalierungsmehrdeutigkeit

Durch Ersetzung von $\dot{\mathbf{d}}$ mit dem entsprechenden Term in (2.71) erhält man

$$\dot{d}_l = \mathbf{n}^T(\mathbf{v} \times \mathbf{l} + \boldsymbol{\omega} \times \mathbf{d}) = \mathbf{n}^T(\mathbf{v} \times \mathbf{l}) = -\mathbf{v}^T \boldsymbol{\alpha}. \quad (2.83)$$

Wir differenzieren (2.80) nach der Zeit

$$\dot{d}_l \beta^T \boldsymbol{\alpha} + d_l (\dot{\beta}^T \boldsymbol{\alpha} + \beta^T \dot{\boldsymbol{\alpha}}) = \dot{\mathbf{n}}^T \mathbf{v} + \mathbf{n}^T \dot{\mathbf{v}}$$

und ersetzen \dot{d}_l und d_l

$$-\mathbf{v}^T \boldsymbol{\alpha} \beta^T \boldsymbol{\alpha} + \frac{\mathbf{n}^T \mathbf{v}}{\beta^T \boldsymbol{\alpha}} (\dot{\beta}^T \boldsymbol{\alpha} + \beta^T \dot{\boldsymbol{\alpha}}) = \dot{\mathbf{n}}^T \mathbf{v} + \mathbf{n}^T \dot{\mathbf{v}}.$$

Nach Umordnung der Terme erhalten wir folgende Gleichung, die homogen linear in \mathbf{v} und $\dot{\mathbf{v}}$ ist:

$$\mathbf{n}^T \mathbf{v} (\dot{\beta}^T \boldsymbol{\alpha} + \beta^T \dot{\boldsymbol{\alpha}}) = \beta^T \boldsymbol{\alpha} (\mathbf{v}^T \boldsymbol{\alpha} \beta^T \boldsymbol{\alpha} + \dot{\mathbf{n}}^T \mathbf{v} + \mathbf{n}^T \dot{\mathbf{v}}). \quad (2.84)$$

Für $\dot{\boldsymbol{\alpha}}$ gilt

$$\dot{\boldsymbol{\alpha}} = \dot{\mathbf{n}} \times \mathbf{l} + \mathbf{n} \times (\boldsymbol{\omega} \times \mathbf{l}),$$

was nach Berechnung der rotatorischen Terme aus Gl. (2.78) und Ersetzung von \mathbf{l} aus Gl. (2.75) als bekannt angenommen werden kann.

Die Ermittlung dreidimensionaler Geschwindigkeiten aus den Geschwindigkeiten der projizierten Geraden ist ein kaum erforshtes Problem. [Faugeras *et al.* 89; Navab *et al.* 90] untersuchen die Geschwindigkeiten von projizierten Geraden in einer Stereoanordnung. Sie leiten (2.73) unter Gebrauch einer Repräsentation her, die anstatt des Momentvektors einen Punkt enthält. [Vieville 90] gibt ein Lösung für den Fall an, bei dem die Beschleunigung verschwindet. Die Herleitung geht von dem diskreten Fall aus und es werden Approximationsformeln für kleine Winkel der Rotation benutzt. Er ermittelt eine quadratische Gleichung für die unbekannte konstante Winkelgeschwindigkeit im Gegensatz zur kubischen Gleichung (2.78) und eine lineare Gleichung für die konstante Translationsgeschwindigkeit. Dabei untersucht er deren Empfindlichkeit bezüglich der Größe des Rotationswinkels und der Lage der Rotationsachse.

Die von uns hergeleiteten Formeln (2.78) und (2.84) für die Ermittlung der Rotations- und Translationsgeschwindigkeiten und -beschleunigungen sind neu und wurden unter Verwendung einer einheitlichen Darstellung hergeleitet. In einem nächsten Schritt sollte man versuchen, die Terme in den obengenannten Gleichungen geometrisch entsprechend den Gleichungen im diskreten Fall zu interpretieren.

2.6 Bewegungsschätzung aus den Verschiebungsraten von koplanaren Geraden

In diesem Abschnitt werden wir eine neue Berechnungsmethode zur Bestimmung der Bewegungsparameter aus der Bewegung einer Ebene vorstellen. Es stellt sich heraus, daß entsprechend dem Fall der Verschiebungsraten von Punkten die Bewegungsparameter ohne Nebenbedingung zu Hilfsparametern zusammengefaßt werden können, die sich aus den Verschiebungsraten durch Invertierung eines linearen Systems herleiten lassen. Damit erbringen wir eine Unifikation der Berechnungswege für Szenenbereichshinweise, die auf einer Ebene liegen.

Die Geometrie des Problems wird in der Abb. 2.4 dargestellt. Die Gerade im Raum wird wie in den letzten Abschnitten mittels der Richtung \mathbf{l} und dem Moment \mathbf{d} parametrisiert, die Einheitsnormale der Ebene wird durch \mathbf{n} und die Verschiebungsrate der projizierten Geraden durch $\dot{\mathbf{n}}$ angegeben. Die Gerade liegt in der Ebene mit der Normalen \mathbf{N} , daraus folgt $\mathbf{N}^T \mathbf{l} = 0$. Aus $\mathbf{n}^T \mathbf{l} = 0$ folgt

$$\mathbf{l} = \frac{\mathbf{n} \times \mathbf{N}}{\|\mathbf{n} \times \mathbf{N}\|}, \quad (2.85)$$

was in der Abb. 2.4 veranschaulicht wird.

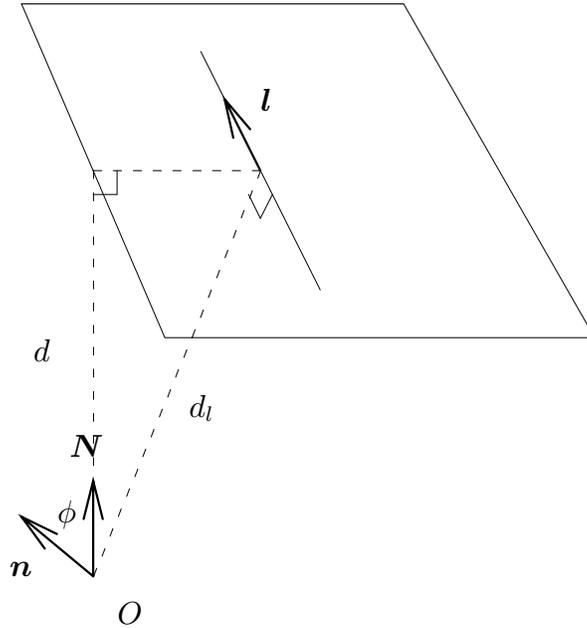


Abbildung 2.4: Gerade in einer Ebene mit dem Einheitsnormalenvektor \mathbf{N} .

Aus der Abb. 2.4 folgt, daß

$$d_l \|\mathbf{n} \times \mathbf{N}\| = d_l |\sin \phi| = d. \quad (2.86)$$

Ersetzen wir \mathbf{l} und d_l in der Gleichung (2.73) für die Verschiebungsrate einer Geraden durch die entsprechenden Ausdrücke in (2.85) und (2.86), erhalten wir

$$\dot{\mathbf{n}} = \frac{1}{d} \mathbf{v}^T \mathbf{n} (\mathbf{n} \times (\mathbf{n} \times \mathbf{N})) + \boldsymbol{\omega} \times \mathbf{n}. \quad (2.87)$$

Wir verwenden die Identität $\mathbf{n} \times (\mathbf{n} \times (\boldsymbol{\omega} \times \mathbf{n})) = -\boldsymbol{\omega} \times \mathbf{n}$, die aus $\|\mathbf{n}\| = 1$ folgt, und erhalten

$$\dot{\mathbf{n}} = \frac{1}{d} \mathbf{n} \times (\mathbf{n} \times \mathbf{N}) \mathbf{v}^T \mathbf{n} - \mathbf{n} \times (\mathbf{n} \times (\boldsymbol{\omega} \times \mathbf{n})) \quad (2.88)$$

$$= [\mathbf{n}]_{\times}^2 \left(\frac{1}{d} \mathbf{N} \mathbf{v}^T - [\boldsymbol{\omega}]_{\times} \right) \mathbf{n}. \quad (2.89)$$

Wie in Gl. (2.35) im Fall von Verschiebungsraten von Punkten werden wir eine der Unbekannten \mathbf{N} oder \mathbf{v} durch den Abstand d der Ebene vom Ursprung skalieren, um

die bekannte Skalierungsmehrdeutigkeit zu eliminieren. Der Symbolsparsamkeit halber behalten wir dieselben Symbole und formen die obige Gleichung um:

$$\dot{\mathbf{n}} = \mathbf{n} \times (\mathbf{n} \times L\mathbf{n}) \quad (2.90)$$

mit

$$L = \mathbf{N}\mathbf{v}^T - [\boldsymbol{\omega}]_{\times}. \quad (2.91)$$

Wir stellen die Analogie zu (2.38) fest: Die Matrix der Hilfsparameter L im Fall von Geraden ist gleich der transponierten Matrix der Hilfsparameter P im Fall von Punkten.

Die neun Elemente der Matrix L sind nicht unabhängig voneinander, weil sie von acht Parametern abhängen. Analog zum Fall der Verschiebungsraten von Punkten wird diese Abhängigkeit durch die Nebenbedingung $\det(L + L^T) = 0$ ausgedrückt.

Um diese nichtlineare Bedingung zu vermeiden, gehen wir wie im Abschnitt 2.3 auf S. 20 vor. Wir stellen auch hier fest, daß die Gleichung (2.90) eine einparametrische Lösung besitzt: Erfüllt eine Matrix L die Gleichung (2.90), so auch jede Matrix $L + \rho I$:

$$\dot{\mathbf{n}} = \mathbf{n} \times (\mathbf{n} \times (L + \rho I)\mathbf{n}) \quad (2.92)$$

Wir suchen folglich nach einer beliebigen Matrix $\Lambda = L + \rho I$, die die Gleichung (2.92) erfüllt, und nach der Ermittlung von Λ bestimmen wir ein ρ , so daß $\det(L + L^T) = 0$ wird.

Die Vektorgleichung (2.92) besteht aus nur zwei unabhängigen Skalargleichungen. Wir wählen die den x - und y -Komponenten von $\dot{\mathbf{n}}$ entsprechenden Gleichungen. Diese lauten

$$\begin{pmatrix} \dot{n}_x \\ \dot{n}_y \end{pmatrix} = \begin{pmatrix} -(n_y^2 + n_z^2)\mathbf{n}^T & n_x n_y \mathbf{n}^T & n_x n_z \mathbf{n}^T \\ n_x n_y \mathbf{n}^T & -(n_x^2 + n_z^2)\mathbf{n}^T & n_y n_z \mathbf{n}^T \end{pmatrix} \begin{pmatrix} \Lambda_{11} \\ \Lambda_{12} \\ \Lambda_{13} \\ \Lambda_{21} \\ \Lambda_{22} \\ \Lambda_{23} \\ \Lambda_{31} \\ \Lambda_{32} \\ \Lambda_{33} \end{pmatrix} \quad (2.93)$$

Über den Kern der Datenmatrix in (2.93) gelten genau die gleichen Aussagen, die wir zum Kern von (2.40) aufgeführt haben, weil die 2×9 -Matrix auf der rechten Seite der letzten Gleichung folgende Eigenschaft hat: Die Summe der ersten, fünften und neunten Spalte der Matrix ist Null. Setzen wir das Element $\Lambda_{33} = 0$ fest, so können wir die neunte Spalte weglassen. Die übrigen acht Unbekannten lassen sich dann aus den Verschiebungsraten von mindestens vier Geraden bestimmen. Zur weiteren Bestimmung der Geschwindigkeiten kann man wie in den Gleichungen (2.41-2.47) vorgehen. Es ist zu bemerken, daß die Notwendigkeit der Auswertung eines dritten Bildes durch die Zwangsbedingung der Koplanarität der Geraden eliminiert worden ist. Zurückblendend auf die Abb. 2.3 kann man sehen, daß zwei koplanare Geraden im Raum nicht mehr

koplanar bleiben, sobald die zweite Sehebenenkombination starr transformiert wird. Offen bleibt die Frage, welche spezielle Konstellation von Geraden die Ermittlung der Bewegungsparameter und der Normalen aus einem Bildpaar nicht ermöglicht.

3 Analytische Fehlerempfindlichkeitsuntersuchungen

3.1 Einführung – Experimentelle vs. analytische Instabilitätsuntersuchungen

In diesem Abschnitt werden wir über Instabilitätsuntersuchungen anderer Autoren berichten. Ihre wichtigsten Aussagen werden wir am Ende zusammenfassen, damit sie mit den von uns in den nächsten Abschnitten hergeleiteten Aussagen verglichen werden können. In dieser Beschreibung ist es für uns von Bedeutung, den Zusammenhang des Schätzfehlers nicht nur in Bezug auf die Meßstörungen, sondern auch in Bezug auf die Problemformulierung, die Art der Bewegung und die Geometrie der aufgenommenen Szenenkomponenten aufzuzeigen.

Empfindlichkeitsuntersuchungen können in zwei große Gruppen unterteilt werden. Die erste Gruppe besteht aus allen Ansätzen, die durch aus Simulationen erzeugte Fehlerdiagramme den Zusammenhang zwischen Schätzfehler und einem Problemparameter darstellen.

Bedenken hat schon einer der ersten Ansätze [Tsai & Huang 84] hervorgerufen, bei dem von 54% Fehler in der geschätzten Translationsrichtung bei 1% Störung in den Punktkorrespondenzen berichtet wird. Dieser hohe Fehler läßt sich verstehen einerseits als eine Folge der Verwendung essentieller Parameter und andererseits durch die Form des Fehlermaßes, wie später von uns gezeigt wird. [Fang & Huang 84] berichten, daß ihr Algorithmus unter Annahme eines kleinen Rotationswinkels zwischen 1° und 5° und eines großen z -Anteils der Translationsrichtung zu einem geringeren Schätzfehler als im allgemeinen Fall führt. [Weng *et al.* 89a] stellen fest, daß ein großes für die Bewegungsschätzung ausgewertetes Gesichtsfeld, ein großes Verhältnis von Translation zum Objektstand und eine Translationsrichtung senkrecht zur Bildebene zu einer stabilen Schätzung der Bewegung beitragen. Bei abnehmendem Winkel zwischen Rotationsachse und optischer Achse nimmt der Schätzfehler in der Richtung der Rotationsachse ab, während der Schätzfehler im Rotationswinkel steigt. [Toscani & Faugeras 87], [Spetsakis & Aloimonos 89], [Weng *et al.* 89b] und [Aisbett 90] erreichen erheblich niedrigere Fehler bei Benutzung von geeigneten Fehlermaßen während der Minimierung, wie es im nächsten Abschnitt beschrieben wird. [Jerian & Jain 90] halten die durchschnittliche Verschiebung im Bild, die der kombinierte Effekt gleichzeitiger Rotation und Translation ist, für den am meisten entscheidenden Stabilitätsfaktor. [Mitiche *et al.* 87] beobachten, daß neben der Translationsrichtung und dem Verhältnis vom Translationsbetrag zum Objektstand die relative Größe der Rotations- zu der Translationskomponente der Verschiebungsrate von Bedeutung ist.

[Young & Chellappa 90b] berechnen die unteren Fehlerschranken –nach Cramer-Rao– für die Schätzung der Bewegungsparameter aus Verschiebungsraten und beweisen analytisch, daß sie unabhängig von den Rotationsparametern sind. Weiterhin zeigen ihre aus Simulationen erzeugten Fehlerschrankendiagramme eine mit abnehmendem Gesichtsfeld wachsende Empfindlichkeit in der Schätzung aller Bewegungsparameter. Die Schätzung des Expansionspunkts wird besonders empfindlich, falls er im Zentrum oder weit außerhalb des Gesichtsfelds liegt, während die Rotationsschätzung für größere Translationsbeträge stabiler wird. [Lee 91] zeigt an konkreten numerischen Beispielen, daß ein Meßrauschen, das nur auf die endliche Auflösung zurückzuführen ist, eine relativen Schätzfehler von mindestens 10% verursacht.

[Barron *et al.* 90] führen eine experimentelle Fehleranalyse für den durchschnittlichen sowie auch den schlechtesten Fall durch und berechnen mittels der Singulärwertzerlegung die Fehlerfortpflanzung für zahlreiche Bewegungs- und Strukturkonfigurationen in Bezug auf die Anzahl der Messungen und die Art des Berechnungsverfahrens. Das besondere an diesem Ansatz sowie auch an denjenigen von [Weng *et al.* 89a] und von [Young & Chellappa 90b] im Vergleich zu den übrigen experimentellen Untersuchungen ist die Möglichkeit, den Fehler a priori berechnen zu können, ohne den Vergleich des Schätzwerts mit dem wahren Wert.

Die zweite Gruppe besteht aus Untersuchungen, die analytisch einen funktionalen Zusammenhang zwischen dem Schätzfehler und der Geometrie des Bewegungsproblems angeben.

Die detaillierteste Analyse findet sich bei [Adiv 89] für die allgemeine Bewegung einer Ebene. Unter verschiedenen Translationsgeschwindigkeiten lassen sich die Verschiebungsraten bei höherem Störungslevel nur in den quadratischen Termen unterscheiden, wobei dieser Unterschied unwesentlich wird, wenn das effektive Gesichtsfeld, die Bewegung im Vergleich zu dem Abstand und die Ebenensteigung klein sind. Dabei wird auf die Schwäche hingewiesen, eine reine Translation parallel zur Bildebene $\mathbf{v} = (p_1, p_2, 0)^T$ nicht von einer kleinen Rotation um eine Achse parallel zur Bildebene $\boldsymbol{\omega} = (-p_2/d, p_1/d, 0)^T$ unterscheiden zu können. Auf dieselbe Tatsache weisen mehrere Autoren ([Buxton *et al.* 85] und [Murray & Buxton 90]) mit dem Argument hin, daß man bei kleinem Gesichtsfeld die linearen und höheren Terme der Verschiebungsraten vernachlässigt.

[Maybank 87] beweist, daß die zu minimierende Fehlernorm ihre Minima entlang der Geraden mit Steigung v_x/v_y aufweist, so daß dieses Verhältnis, sowie auch die Projektion der Winkelgeschwindigkeit $(v_x, v_y, 0)^T \boldsymbol{\omega}$ auf die eben genannte Gerade in der Bildebene robust berechnet werden können. Von Bedeutung ist seine Annahme, daß es sich um eine irreguläre (Diskontinuitäten enthaltende) Fläche in Bewegung handelt, was mit den Ergebnissen von [Rieger & Lawton 85] über robuste Bewegungsberechnung an Verdeckungskonturen übereinstimmt.

[Manmatha *et al.* 89] zeigen, daß die Annahme einer reinen Translation, wenn die tatsächliche Bewegung doch noch einen rotatorischen Anteil besitzt, eine erhebliche Verzerrung in der Schätzung des Expansionspunktes hervorruft. [Koenderink & van Doorn 87] zeigen qualitativ, daß die Unsicherheit im Winkel zwischen Blickrichtung und Translationsrichtung mit der Größe dieses Winkels wächst. Diese Instabilität wird auch bei [Horn & Weldon 87] beschrieben, die wie auch [Negahdaripour & Yu 88] auf die Bedeutung der Größe des Gesichtsfelds in ihrem Ansatz der direkten Berechnung

von Bewegungsparametern aus Grauwertbildern hinweisen.

[Longuet-Higgins 84] und [Faugeras & Lustman 88] deuten auf die durch das Zusammenfallen beider Lösungen im Fall einer zur Ebene senkrechten Translation entstehende Instabilität.

[Horn 90] untersucht die Form der Flächen in Bewegung, die bei verschiedenen Bewegungen nur einen quadratischen Zuwachs der aus der epipolaren Bedingung entstandenen Norm verursachen. In einem späteren Abschnitt beschreiben wir die Beziehung dieser kritischen Flächen im "engeren" Sinn zu den allgemeinen mehrdeutigen Flächen.

Abschließend fassen wir alle die Fehlerempfindlichkeit beeinflussenden Faktoren zusammen:

- Die Form des zu minimierenden Fehlermaßes
- Die Größe des effektiven Gesichtsfelds (Flächeninhalt der Projektion einer Systemkomponente in relativer Bewegung)
- Der Translationsbetrag relativ zum Abstand des abgebildeten Objekts (er trägt zum Betrag der Verschiebungsraten bei)
- Der Winkel zwischen Blickrichtung und Translationsrichtung
- Die Tiefenvariation der abgebildeten Szenenkomponente
- Der Betrag der Translationskomponente relativ zur Rotationskomponente der Verschiebungsraten.

Das Zusammenspiel dieser Faktoren werden wir in den folgenden Abschnitten diskutieren.

3.2 Zur Form des zu minimierenden Fehlermaßes

Wir beginnen unsere Überlegungen am Fall der Berechnung von dreidimensionalen Geschwindigkeiten aus Verschiebungsraten. Wie es in der Problemstellung (Abschnitt 2.3) schon beschrieben ist, wird die Verschiebungsraten durch

$$\dot{\mathbf{x}} = \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) + \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})) \quad (3.1)$$

angegeben. Wenn man eine im Sinne eines Maximum-Likelihood Ansatzes optimale Schätzung sucht, soll man unter der Annahme, daß die Störung in den Verschiebungsraten normalverteilt ist, folgende bedingte Wahrscheinlichkeit – unter der Nebenbedingung $\|\mathbf{v}\| = 1$ zur Eliminierung der Skalierungsmehrdeutigkeit – maximieren:

$$\frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2} \iint_D \left\| \dot{\mathbf{x}} - \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) - \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})) \right\|^2 dx dy}, \quad (3.2)$$

wobei σ^2 die Standardabweichung des Meßrauschens, dessen Erwartungswert gleich Null ist, in jeder Komponente der Verschiebungsraten ist. Dieser Schätzprozeß ist äquivalent

zu der Minimierung des quadrierten Fehlers zwischen gemessenen und erwarteten Verschiebungsraten

$$\iint_D \left\| \dot{\mathbf{x}} - \frac{1}{\hat{\mathbf{z}}^T \mathbf{X}} \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) - \hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega})) \right\|^2 dx dy \Rightarrow \min. \quad (3.3)$$

Man erhält nach Eliminierung der Tiefe (siehe Herleitung in Gl. (2.29) und (2.30)) folgendes Fehlermaß (2.29)

$$\iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 dx dy, \quad (3.4)$$

das schon von [Bruss & Horn 83; Lawton 83; Adiv 85] verwendet worden ist. Wir werden zeigen, daß die geometrische Interpretation des Integranden dieser Kostenfunktion der Abstand der Translationskomponente $\dot{\mathbf{x}}_T$ der Verschiebungsrates von der Geraden ist, die den Expansionspunkt mit dem Punkt \mathbf{x} verbindet. Für den Abstand d' in der Abb.

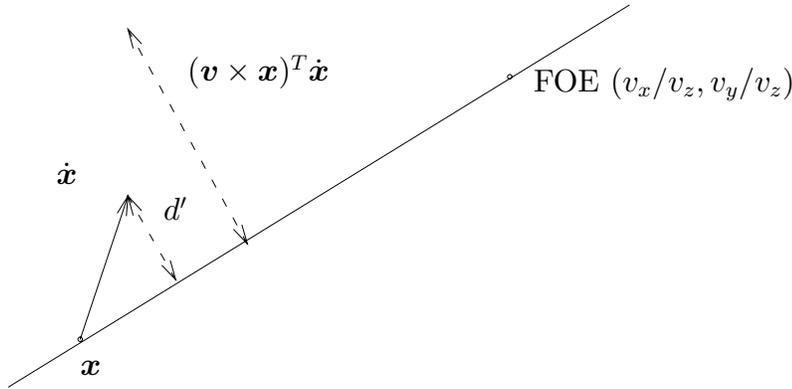


Abbildung 3.1: Geometrische Interpretation des Fehlermaßes d' . Der Expansionspunkt wird mit FOE (Focus Of Expansion) bezeichnet.

3.1 gilt

$$d' = \frac{|\dot{\mathbf{x}}_T^T (\hat{\mathbf{z}} \times (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})))|}{\|\hat{\mathbf{z}} \times (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}))\|} = \frac{|((\dot{\mathbf{x}}_T \times \hat{\mathbf{z}}) \times \hat{\mathbf{z}})^T (\mathbf{v} \times \mathbf{x})|}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} = \frac{|\dot{\mathbf{x}}_T^T (\mathbf{v} \times \mathbf{x})|}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|}.$$

Für die Translationskomponente gilt

$$\dot{\mathbf{x}}_T^T (\mathbf{v} \times \mathbf{x}) = (\dot{\mathbf{x}} - \hat{\mathbf{z}}^T (\mathbf{x} \times \boldsymbol{\omega}) \mathbf{x} - (\boldsymbol{\omega} \times \mathbf{x}))^T (\mathbf{v} \times \mathbf{x}) = (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})^T (\mathbf{v} \times \mathbf{x}),$$

woraus folgt

$$d' = \frac{|(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})|}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|}. \quad (3.5)$$

Von [Bruss & Horn 83] wird als einfachere Norm der Zähler des Fehlermaßes vorgeschlagen, der in seiner diskreten Version von der Mehrheit der Ansätze benutzt wird. Das hat zur Folge, daß der eigentliche Fehler mit dem Term $\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\| = |v_z| \|\mathbf{v}/v_z - \mathbf{x}\|$ gewichtet wird, der gleich dem Produkt von v_z und dem Abstand des Expansionspunkts

von dem Punkt \mathbf{x} ist. Falls der Projektionsbereich eines sich relativ zur Kamera bewegendes Objektes klein ist, wird diese Gewichtung zur Folge haben, daß die Minimierung eine Translationsrichtung bevorzugt, die den Winkel zwischen der Translationsrichtung und der Blickrichtung so klein wie möglich macht. Den Verzerrungseinfluß dieser Gewichtung werden wir im folgenden Abschnitt analytisch beweisen.

Wir beschreiben zunächst Fehlernormen für den diskreten Fall. Eine Maximum-Likelihood-Schätzung ergibt eine Minimierung der Fehler in den zugeordneten Punkten in beiden Bildern

$$\sum_{i=1}^m \left\{ \left\| \mathbf{x}_{1i} - \frac{\mathbf{X}_i}{\hat{\mathbf{z}}^T \mathbf{X}_i} \right\|^2 + \left\| \mathbf{x}_{2i} - \frac{R\mathbf{X}_i + \mathbf{T}}{\hat{\mathbf{z}}^T (R\mathbf{X}_i + \mathbf{T})} \right\|^2 \right\} \quad (3.6)$$

bezüglich der Rotations- und Translationsparameter sowie bezüglich der Strukturparameter $\mathbf{X}_i = (X_i, Y_i, Z_i)^T$ der Punktconfiguration. Wenn man versucht, die Struktur aus dieser Minimierung zu eliminieren, ergibt sich eine Gleichung vierten Grades in der Tiefe Z , die eine weitere Analyse in geschlossener Form nicht zuläßt. Wenn man als Störung nur einen Fehler in dem zugeordneten Punkt im zweiten Bild annimmt, beschränkt sich die Minimierung auf den zweiten Term der Kostenfunktion (3.6):

$$\sum_{i=1}^m \left\| \mathbf{x}_{2i} - \frac{Z_i R\mathbf{x}_{1i} + \mathbf{T}}{\hat{\mathbf{z}}^T (Z_i R\mathbf{x}_{1i} + \mathbf{T})} \right\|^2 \Rightarrow \min, \quad (3.7)$$

die man nach der Tiefe Z_i differenziert:

$$\begin{aligned} & \left(\hat{\mathbf{z}}^T (Z_i R\mathbf{x}_{1i} + \mathbf{T}) \mathbf{x}_{2i} - (Z_i R\mathbf{x}_{1i} + \mathbf{T}) \right)^T \left(\hat{\mathbf{z}}^T (Z_i R\mathbf{x}_{1i} + \mathbf{T}) R\mathbf{x}_{1i} \right. \\ & \left. - \hat{\mathbf{z}}^T R\mathbf{x}_{1i} (Z_i R\mathbf{x}_{1i} + \mathbf{T}) \right) = 0 \\ \iff & \left(Z_i (\hat{\mathbf{z}} \times (\mathbf{x}_{2i} \times R\mathbf{x}_{1i})) + \hat{\mathbf{z}} \times (\mathbf{x}_{2i} \times \mathbf{T}) \right)^T \left(\hat{\mathbf{z}} \times (R\mathbf{x}_{1i} \times \mathbf{T}) \right) = 0 \end{aligned}$$

Man kann dann nach der Tiefe auflösen

$$Z_i = - \frac{(\hat{\mathbf{z}} \times (\mathbf{x}_{2i} \times \mathbf{T}))^T (\hat{\mathbf{z}} \times (R\mathbf{x}_{1i} \times \mathbf{T}))}{(\hat{\mathbf{z}} \times (\mathbf{x}_{2i} \times R\mathbf{x}_{1i}))^T (\hat{\mathbf{z}} \times (R\mathbf{x}_{1i} \times \mathbf{T}))}. \quad (3.8)$$

Setzt man jetzt die Tiefe in (3.7) ein, so erhält man nach einer langen Kette von Vektoroperationen folgendes Fehlermaß

$$\sum_{i=1}^m \frac{(\mathbf{x}_{2i}^T (\mathbf{T} \times R\mathbf{x}_{1i}))^2}{\|\hat{\mathbf{z}} \times (\mathbf{T} \times R\mathbf{x}_{1i})\|^2}, \quad (3.9)$$

das die diskrete Form von (2.30) darstellt: es ist die Summe der quadrierten Abstände der Punkte \mathbf{x}_{2i} von den Geraden, die jeweils $R\mathbf{x}_{1i}$ mit dem Expansionspunkt verbinden.

Eine weitere Alternative ist die Gewichtung der epipolaren Fehlernorm $\mathbf{x}_2^T (\mathbf{T} \times R\mathbf{x}_1)$ durch ihre Varianz. Wir bezeichnen $\varepsilon = \mathbf{x}_2^T (\mathbf{T} \times R\mathbf{x}_1)$ und berechnen den Fehler erster Ordnung

$$\delta\varepsilon = \delta\mathbf{x}_1^T R^T (\mathbf{x}_2 \times \mathbf{T}) + \delta\mathbf{x}_2^T (\mathbf{T} \times R\mathbf{x}_1). \quad (3.10)$$

Nehmen wir an, daß die gemessenen Bildkoordinaten mit Rauschen (ξ, η) behaftet sind, so gilt $\delta\mathbf{x}_1 = \xi\hat{\mathbf{x}} + \eta\hat{\mathbf{y}}$ und entsprechend für $\delta\mathbf{x}_2$. Wenn die Erwartungswerte $E[\xi] = E[\eta] = E[\xi\eta] = 0$ und die Varianzen $E[\xi^2] = E[\eta^2] = \sigma^2$ sind, folgt

$$\sigma_\varepsilon^2 = \sigma^2 \left((\hat{\mathbf{x}}^T R^T (\mathbf{x}_2 \times \mathbf{T}))^2 + (\hat{\mathbf{y}}^T R^T (\mathbf{x}_2 \times \mathbf{T}))^2 + (\hat{\mathbf{x}}^T (\mathbf{T} \times R\mathbf{x}_1))^2 + (\hat{\mathbf{y}}^T (\mathbf{T} \times R\mathbf{x}_1))^2 \right),$$

was in kompakter Notation lautet

$$\sigma_\varepsilon^2 = \sigma^2(\|\hat{\mathbf{z}} \times R^T(\mathbf{x}_2 \times \mathbf{T})\|^2 + \|\hat{\mathbf{z}} \times (\mathbf{T} \times R\mathbf{x}_1)\|^2). \quad (3.11)$$

Die Minimierung, die aus einer Maximum-Likelihood-Schätzung

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\sum_{i=1}^m \frac{\varepsilon_i^2}{2\sigma_\varepsilon^2}} \implies \max$$

erfolgt, lautet

$$\sum_{i=1}^m \frac{(\mathbf{x}_{2i}^T(\mathbf{T} \times R\mathbf{x}_{1i}))^2}{\sigma^2(\|\hat{\mathbf{z}} \times R^T(\mathbf{x}_2 \times \mathbf{T}_i)\|^2 + \|\hat{\mathbf{z}} \times (\mathbf{T} \times R\mathbf{x}_{1i})\|^2)} \implies \min \quad (3.12)$$

Diese Form wird von [Weng *et al.* 89b] und [Trivedi 87] vorgeschlagen.

Ein anderer Aspekt des Problems entsteht, wenn man den Fehler im Szenenbereich auf den Bildbereich zu transformieren versucht. Man kann sich als Fehler im Szenenbereich die Länge γ der gemeinsamen Normalen der zwei Sichtstrahlen \mathbf{x}_1 und \mathbf{x}_2 wie in Abb. 3.2 vorstellen. Es gilt

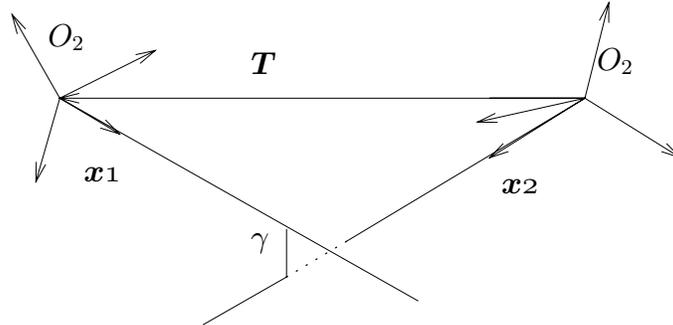


Abbildung 3.2: Gemeinsame Normale der Sichtstrahlen

$$Z_2\mathbf{x}_2 = Z_1R\mathbf{x}_1 + \mathbf{T} + \gamma \frac{\mathbf{x}_2 \times R\mathbf{x}_1}{\|\mathbf{x}_2 \times R\mathbf{x}_1\|}, \quad (3.13)$$

woraus man folgende Ausdrücke für Z_1 , Z_2 und γ erhält:

$$Z_1 = \frac{(\mathbf{T} \times \mathbf{x}_2)^T(\mathbf{x}_2 \times R\mathbf{x}_1)}{\|\mathbf{x}_2 \times R\mathbf{x}_1\|^2} \quad (3.14)$$

$$Z_2 = \frac{(\mathbf{T} \times R\mathbf{x}_1)^T(\mathbf{x}_2 \times R\mathbf{x}_1)}{\|\mathbf{x}_2 \times R\mathbf{x}_1\|^2} \quad (3.15)$$

$$\gamma = \frac{\mathbf{x}_2^T(\mathbf{T} \times R\mathbf{x}_1)}{\|\mathbf{x}_2 \times R\mathbf{x}_1\|} \quad (3.16)$$

[Horn 90] schlägt vor, die Varianz des Spatprodukts $\varepsilon = \mathbf{x}_2^T(\mathbf{T} \times R\mathbf{x}_1)$ auf die Varianz der gemeinsamen Normalen zu beziehen. Aus (3.16) folgt

$$\sigma_\varepsilon^2 = \|\mathbf{x}_2 \times R\mathbf{x}_1\|^2 \sigma_\gamma^2. \quad (3.17)$$

Diese Varianz kann man weiter auf die Varianz der gemessenen Bildkoordinaten beziehen. Eine Störung δ_γ in der Länge der gemeinsamen Normalen kann nur von der Komponente δ_\perp einer Störung in den Bildpositionen hervorgerufen werden, die parallel zur gemeinsamen Normalen ist. Eine Störung δ_\perp in den Bildpunkten \mathbf{x}_1 und \mathbf{x}_2 verursacht eine Störung jeweils $Z_1\|\mathbf{R}\mathbf{x}_1\|\delta_\perp$ und $Z_2\|\mathbf{x}_2\|\delta_\perp$ in der Länge der gemeinsamen Normalen. Bezeichnen man mit σ^2 die Varianz von δ_\perp , so lautet die Varianz der gemeinsamen Normalen

$$\sigma_\gamma^2 = \sigma^2 (Z_1^2\|\mathbf{x}_1\|^2 + Z_2^2\|\mathbf{x}_2\|^2).$$

Man setzt sie in (3.17) ein und nutzt die Tatsache aus, daß $\|\mathbf{x}_2 \times \mathbf{R}\mathbf{x}_1\|$ in (3.17) sowie auch in den Nennern von (3.14-3.15) auftaucht. Nach der Vereinfachung erhält man

$$\sigma_\varepsilon^2 = \sigma^2 \frac{\left((\mathbf{T} \times \mathbf{x}_2)^T (\mathbf{x}_2 \times \mathbf{R}\mathbf{x}_1) \right)^2 \|\mathbf{x}_1\|^2 + \left((\mathbf{T} \times \mathbf{R}\mathbf{x}_1)^T (\mathbf{x}_2 \times \mathbf{R}\mathbf{x}_1) \right)^2 \|\mathbf{x}_2\|^2}{\|\mathbf{x}_2 \times \mathbf{R}\mathbf{x}_1\|^2},$$

woraus sich folgende Minimierung ergibt

$$\sum_{i=1}^m \frac{|\mathbf{x}_2^T (\mathbf{T} \times \mathbf{R}\mathbf{x}_{1i})|^4}{\sigma^2 \left(((\mathbf{T} \times \mathbf{x}_{2i})^T (\mathbf{x}_{2i} \times \mathbf{R}\mathbf{x}_{1i}))^2 \|\mathbf{x}_{1i}\|^2 + ((\mathbf{T} \times \mathbf{R}\mathbf{x}_{1i})^T (\mathbf{x}_{2i} \times \mathbf{R}\mathbf{x}_{1i}))^2 \|\mathbf{x}_{2i}\|^2 \right)} \implies \min.$$

Die Diskussion hat gezeigt, daß alle Normvarianten von den aus der epipolaren Bedingung entstehenden Normen (diskret $\mathbf{x}_2^T (\mathbf{T} \times \mathbf{R}\mathbf{x}_1)$, kontinuierlich $(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})$) abweichen. Trotzdem sind diese "epipolaren" Normen besser zur algebraischen Manipulation geeignet, was man auch an der E -Matrix-Theorie (Abschnitt 2.2) sieht. Unter welchen Bedingungen man sie vermeiden soll, werden wir im nächsten Abschnitt explizit zeigen.

3.3 Fehlerempfindlichkeit der Translationsrichtung im kontinuierlichen Fall

In diesem Abschnitt werden wir die stark fehlervergrößernde Wechselwirkung zwischen einem kleinen effektiven Gesichtsfeld, einem großen Winkel zwischen Blickrichtung und Translationsrichtung und der Form des Fehlermaßes analytisch nachweisen. Dabei gehen wir vom kontinuierlichen Fall und einer reinen Translation aus und untersuchen folgendes Fehlermaß

$$\iint_D \left\{ (\mathbf{v} \times \mathbf{x})^T \dot{\mathbf{x}} \right\}^2 dx dy = \mathbf{v}^T \left\{ \iint_D (\mathbf{x} \times \dot{\mathbf{x}})(\mathbf{x} \times \dot{\mathbf{x}})^T dx dy \right\} \mathbf{v} \implies \min_{\|\mathbf{v}\|=1} \quad (3.18)$$

Unsere Behauptungen gelten auch für den Fall der allgemeinen Bewegung, wenn die Verschiebungsraten schon in eine Rotations- und eine Translationskomponente zerlegt worden sind, oder wenn Translation und Rotation sequentiell berechnet werden.

Die Lösung der Minimierungsaufgabe (3.18) ist der Eigenvektor der Matrix

$$\Upsilon = \iint_D (\mathbf{x} \times \dot{\mathbf{x}})(\mathbf{x} \times \dot{\mathbf{x}})^T dx dy, \quad (3.19)$$

der dem kleinsten Eigenwert entspricht. Im meßfehlerfreien Fall wird die epipolare Bedingung $\dot{\mathbf{x}}^T (\mathbf{v} \times \mathbf{x}) = 0$ exakt erfüllt, und der kleinste Eigenwert der Matrix ist

gleich Null. Im meßfehlerbehafteten Fall ist die Störung im Eigenvektor umgekehrt proportional zur Differenz zwischen den kleinsten Eigenwerten. Die Instabilität im Fall von zwei beieinander liegenden Eigenwerten ist wiederholt in der linearen Algebra [Golub & van Loan 83] und in der Perturbationstheorie in der Physik [Mathews & Walker 73] untersucht worden. Wenn eine symmetrische Matrix A durch eine Matrix E gestört wird, gilt für den gestörten Eigenvektor \mathbf{x}'_k

$$\mathbf{x}'_k \approx \mathbf{x}_k + \sum_{j \neq k} \frac{\mathbf{x}_j^T E \mathbf{x}_k}{(\lambda_j - \lambda_k)} \mathbf{x}_j, \quad (3.20)$$

wobei λ_i , \mathbf{x}_i die Eigenwerte bzw. die Eigenvektoren der ungestörten Matrix A sind. Die Störung ist von besonderer Bedeutung in Richtungen \mathbf{x}_j , die Eigenwerten entsprechen, die nahe bei λ_k liegen. Die Eigenvektoren lassen sich als Richtungen der Achsen eines Ellipsoids interpretieren, das die Gleichung $\mathbf{v}^T \Upsilon \mathbf{v} = \lambda_1$ erfüllt, wobei λ_1 als kleinster Eigenwert angenommen wird. Die Längen der Achsen dieses Ellipsoids sind 1 , $\sqrt{\lambda_1/\lambda_2}$ und $\sqrt{\lambda_1/\lambda_3}$. Das Ellipsoid wird von einer Einheitskugel umschrieben. Ein Schnitt des Ellipsoids mit der Ebene, die von den Eigenvektoren \mathbf{v}_1 und \mathbf{v}_2 aufgespannt wird, ist in der Abb. 3.3 dargestellt. Die ungefähre Gleichheit der Eigenwerte bedeutet geometrisch,

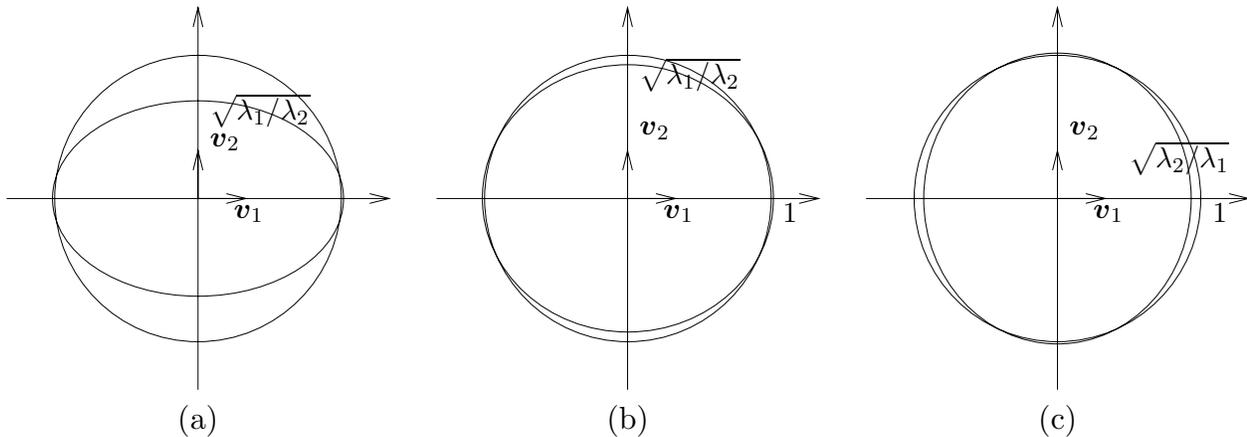


Abbildung 3.3: Stabile (a), und instabile (b,c) Konfigurationen für den kleinsten Eigenwert

daß das Ellipsoid kurz davor ist, nicht nur in zwei Punkten, sondern in einem Kreis die Einheitskugel zu tangieren. Bei einer so instabilen Lage (Abbildung 3.3b) kann λ_2 nach einer kleinen Störung die Rolle des kleinsten Eigenwertes übernehmen (Abbildung 3.3c) und folglich als Lösung einen Eigenvektor ergeben, dessen Richtung um 90° falsch ist. Im folgenden werden wir nachweisen, daß diese Situation mit größerer Wahrscheinlichkeit in einem Bewegungsproblem auftreten kann, wenn sich ein Objekt parallel zur Bildebene bewegt und seine Projektion nur eine kleine Fläche in der Bildebene bedeckt.

Wir nehmen an, daß das effektive Gesichtsfeld (d.h. der Bereich, auf dem ein sich relativ zur Kamera bewegendes Objekt abgebildet wird) ein Rechteck mit Seiten α und β im Zentrum des Bildes ist (Abb. 3.4). Damit unterstellen wir, daß die Blickrichtung mit der optischen Achse zusammenfällt.

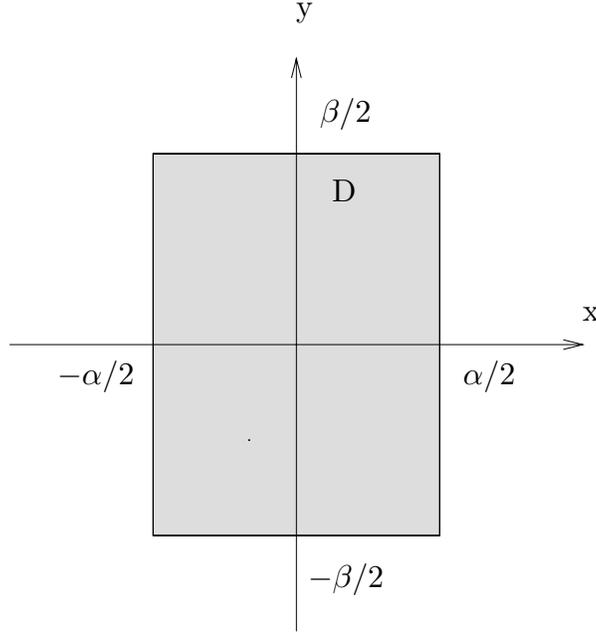


Abbildung 3.4: Modell des effektiven Gesichtsfelds

Weiterhin nehmen wir an, daß die Verschiebungsraten mit einer stochastischen Störung behaftet sind,

$$\dot{\mathbf{x}}' = \dot{\mathbf{x}} + \begin{pmatrix} \xi \\ \eta \\ 0 \end{pmatrix}$$

deren Erwartungswert verschwindet und deren Varianzen $E[\xi^2] = E[\eta^2] = \sigma^2$, $E[\xi\eta] = 0$ sind. Somit wird die Matrix Υ nicht erwartungstreu und erhält folgende diagonale Verzerrung

$$\bar{\Upsilon} = E[\Upsilon] = \Upsilon + \begin{pmatrix} \sigma^2\alpha\beta & 0 & 0 \\ 0 & \sigma^2\alpha\beta & 0 \\ 0 & 0 & \sigma^2\alpha\beta\frac{\alpha^2+\beta^2}{12} \end{pmatrix}. \quad (3.21)$$

Die Elemente der Matrix Υ bestehen aus folgenden Integralen:

$$\Upsilon = \begin{pmatrix} \iint_D \dot{y}^2 dx dy & -\iint_D \dot{x}\dot{y} dx dy & -\iint_D \dot{y}(x\dot{y} - y\dot{x}) dx dy \\ -\iint_D \dot{x}\dot{y} dx dy & \iint_D \dot{x}^2 dx dy & \iint_D \dot{x}(x\dot{y} - y\dot{x}) dx dy \\ -\iint_D \dot{y}(x\dot{y} - y\dot{x}) dx dy & \iint_D \dot{x}(x\dot{y} - y\dot{x}) dx dy & \iint_D (x\dot{y} - y\dot{x})^2 dx dy \end{pmatrix} \quad (3.22)$$

Für die Durchführung der Integrationen sind wir gezwungen, ein Modell für die Verschiebungsraten zu unterstellen. Wir nehmen an, daß sich eine Ebene durch die Gleichung $N_x X + N_y Y + N_z Z = 1$ im Abstand $1/N_z$ (Schnittpunkt der Ebene mit der optischen Achse) vor der Kamera translatorisch bewegt¹. Das entstandene meßfehlerfreie Vek-

¹Der speziellere Fall einer frontalen Ebene $\mathbf{N} = (0, 0, 1)^T$ wird bereits in [Daniilidis & Nagel 90] behandelt.

torfeld lautet gemäß (2.35)

$$\dot{\mathbf{x}} = \begin{pmatrix} (N_x x + N_y y + N_z)(v_x - xv_z) \\ (N_x x + N_y y + N_z)(v_y - yv_z) \\ 0 \end{pmatrix}. \quad (3.23)$$

Nach der Berechnung der Integrale in (3.22) addieren wir die diagonale Verzerrung entsprechend Gl. (3.21). Alle Elemente der Matrix $\tilde{\Upsilon}$ sind Vielfache der Fläche $\alpha\beta$ des Gesichtsfelds. Ihre Division durch $\alpha\beta$ hat keinen Einfluß auf unsere Untersuchung, wir werden aber der Symbolsparsamkeit halber weiter das Symbol $\tilde{\Upsilon}$ für den Erwartungswert der Matrix benutzen. Zur Verkürzung der Terme ersetzen wir $\alpha^2/12$ und $\beta^2/12$ durch A bzw. B. Aus (3.21) folgt

$$\tilde{\Upsilon} = \Upsilon + \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2(A+B) \end{pmatrix} = \Upsilon + \Sigma, \quad (3.24)$$

wobei die Elemente der Matrix Υ wie folgt aussehen

$$\begin{aligned} \Upsilon_{11} &= v_y^2(N_z^2 + AN_x^2 + BN_y^2) + Bv_z^2(N_z^2 + AN_x^2) + 9/5B^2N_y^2v_z^2 - 4BN_yN_zv_yv_z \\ \Upsilon_{12} &= -v_xv_y(N_z^2 + AN_x^2 + BN_y^2) - 2ABN_xN_yv_z^2 + 2AN_xN_zv_yv_z + 2BN_yN_zv_xv_z \\ \Upsilon_{13} &= 2BN_yN_zv_xv_y - 9/5B^2N_y^2v_xv_z - Bv_xv_z(N_z^2 + AN_x^2) + 2ABN_xN_yv_yv_z - 2AN_xN_zv_y^2 \\ \Upsilon_{22} &= v_x^2(N_z^2 + AN_x^2 + BN_y^2) + Av_z^2(N_z^2 + BN_y^2) + 9/5A^2N_x^2v_z^2 - 4AN_xN_zv_xv_z \\ \Upsilon_{23} &= 2AN_xN_zv_xv_y - 9/5A^2N_x^2v_yv_z - Av_yv_z(N_z^2 + BN_y^2) + 2ABN_xN_yv_xv_z - 2BN_yN_zv_x^2 \\ \Upsilon_{33} &= AN_z^2v_y^2 + 9/5A^2N_x^2v_y^2 + ABN_y^2v_y^2 + BN_z^2v_x^2 \\ &\quad + 9/5B^2N_y^2v_x^2 + ABN_x^2v_x^2 - 4ABN_xN_yv_xv_y \end{aligned}$$

Leider sind wir nicht in der Lage gewesen, aus den Ausdrücken für die Eigenwerte der Matrix Υ funktionale Zusammenhänge zu erkennen. So sind wir gezwungen, eine Fallstudie durchzuführen.

Wir beginnen die Fallstudie mit dem Fall einer Translation in Richtung der optischen Achse, d.h. $v_x = v_y = 0, v_z \neq 0$, die nach unseren geometrischen Überlegungen (Abschnitt 3.2 nach Gl. (3.5)) eine stabile Schätzung ermöglicht. Die Matrix $\tilde{\Upsilon}$ erhält folgende einfache Form

$$\tilde{\Upsilon} = \begin{pmatrix} Bv_z^2(N_z^2 + AN_x^2 + 9/5BN_y^2) & -2ABN_xN_yv_z^2 & 0 \\ -2ABN_xN_yv_z^2 & Av_z^2(N_z^2 + 9/5AN_x^2 + BN_y^2) & 0 \\ 0 & 0 & 0 \end{pmatrix} + \Sigma \quad (3.25)$$

und hat die Eigenwerte

$$\begin{aligned} \lambda'_{1,2} &= \sigma^2 + \frac{1}{2} \left(\Upsilon_{11} + \Upsilon_{22} \pm \sqrt{(\Upsilon_{11} - \Upsilon_{22})^2 + 4\Upsilon_{12}^2} \right) \\ \lambda'_3 &= \sigma^2(A+B). \end{aligned}$$

Zur Instabilitätsuntersuchung betrachten wir die Differenz zwischen dem kleineren Eigenwert aus $\lambda'_{1,2}$ und λ'_3

$$\lambda'_2 - \lambda'_3 = \sigma^2(1 - A - B) + \frac{1}{2} \left(\Upsilon_{11} + \Upsilon_{22} - \sqrt{(\Upsilon_{11} - \Upsilon_{22})^2 + 4\Upsilon_{12}^2} \right).$$

Da alle Größen in Brennweiteinheiten gemessen werden, sind A und B erheblich kleiner als 1. Folglich ist $\sigma^2(1 - A - B)$ immer größer als Null. Wir beweisen, daß der zweite Summand positiv ist:

$$\begin{aligned} \Upsilon_{11} + \Upsilon_{22} - \sqrt{(\Upsilon_{11} - \Upsilon_{22})^2 + 4\Upsilon_{12}^2} &> 0 \iff \\ \Upsilon_{11}\Upsilon_{22} &> \Upsilon_{12}^2 \iff \\ N_z^2(N_z^2 + 14/5(AN_x^2 + BN_y^2)) + 9/5(A^2N_x^4 + B^2N_y^4) + 106/25ABN_x^2N_y^2 &> 4ABN_x^2N_y^2 \iff \\ N_z^2(N_z^2 + 14/5(AN_x^2 + BN_y^2)) + 9/5(A^2N_x^4 + B^2N_y^4) + 6/25ABN_x^2N_y^2 &> 0, \end{aligned}$$

was eine wahre Ungleichung ist.

Daraus folgt, daß der kleinste Eigenwert λ'_3 ist und der geschätzte Translationsvektor $(0, 0, 1)^T$ stabil geschätzt wird. Im Fall einer frontalen Ebene ($N_x = N_y = 0$) erhalten wir für die Differenzen zwischen den Eigenwerten

$$\begin{aligned} \lambda'_1 - \lambda'_3 &= BN_z^2v_z^2 + \sigma^2(1 - A - B) > 0 \\ \lambda'_2 - \lambda'_3 &= AN_z^2v_z^2 + \sigma^2(1 - A - B) > 0, \end{aligned} \quad (3.26)$$

daß sie erheblich größer als Null sind. Die Eigenwerte sind gut isoliert und die Lösung ist robust.

Im Fall einer Bewegung parallel zur Bildebene ($v_z = 0$) wird die Matrix Υ voll besetzt:

$$\begin{aligned} \Upsilon_{11} &= v_y^2(N_z^2 + AN_x^2 + BN_y^2) \\ \Upsilon_{12} &= -v_xv_y(N_z^2 + AN_x^2 + BN_y^2) \\ \Upsilon_{13} &= 2BN_yN_zv_xv_y - 2AN_xN_zv_y^2 \\ \Upsilon_{22} &= v_x^2(N_z^2 + AN_x^2 + BN_y^2) \\ \Upsilon_{23} &= 2AN_xN_zv_xv_y - 2BN_yN_zv_x^2 \\ \Upsilon_{33} &= (Av_y^2 + Bv_x^2)(N_z^2 + AN_x^2 + BN_y^2) + 4/5A^2N_x^2v_y^2 + 4/5B^2N_y^2v_x^2 - 4ABN_xN_yv_xv_y \end{aligned} \quad (3.27)$$

Wir berücksichtigen zunächst den Fall einer frontalen Ebene ($N_x = 0, N_y = 0$), wobei die Matrix Υ nicht mehr voll besetzt ist:

$$\tilde{\Upsilon} = \begin{pmatrix} v_y^2N_z^2 & -v_xv_yN_z^2 & 0 \\ -v_xv_yN_z^2 & v_x^2N_z^2 & 0 \\ 0 & 0 & (Av_y^2 + Bv_x^2)N_z^2 \end{pmatrix} + \Sigma. \quad (3.28)$$

Die Eigenwerte von $\tilde{\Upsilon}$ lauten dann

$$\begin{aligned} \lambda'_1 &= \sigma^2 \\ \lambda'_2 &= \sigma^2 + N_z^2(v_x^2 + v_y^2) \\ \lambda'_3 &= \sigma^2(A + B) + (Av_y^2 + Bv_x^2)N_z^2. \end{aligned}$$

Wir untersuchen die Differenzen

$$\begin{aligned}\lambda'_2 - \lambda'_1 &= N_z^2(v_x^2 + v_y^2) > 0 \\ \lambda'_3 - \lambda'_1 &= N_z^2(Av_y^2 + Bv_x^2) + \sigma^2(A + B - 1)\end{aligned}$$

Wir setzen die zweite Differenz gleich Null und versuchen nach σ^2 zu lösen. Falls der sich ergebende Wert für σ^2 positiv ist, kann eine Varianz auftreten, die die Eigenwerte λ'_3 und λ'_1 zusammenfallen läßt. Die Lösung für σ^2 lautet

$$\sigma^2 = \frac{N_z^2(Av_y^2 + Bv_x^2)}{1 - A - B} \quad (3.29)$$

und ist positiv. Beim Überschreiten dieses Rauschniveaus tritt eine extreme Instabilität auf, bei der der geschätzte Eigenvektor die Richtung der optischen Achse hat und um 90° falsch ist.

Ist die Ebene nicht frontal, dann muß man die charakteristische Gleichung der Matrix $\Upsilon + \Sigma$ (Υ wie in (3.27)) lösen, die dritten Grades ist. Der Faktor $\lambda - \sigma^2$ läßt sich aus der charakteristischen Gleichung faktorisieren. Damit ist einer von den berechneten Eigenwerten $\lambda'_1 = \sigma^2$ und soll einem Eigenvektor entsprechen, der in der Ebene XY liegt, weil wir a priori wissen, daß, falls $\sigma = 0$ wird, der geschätzte Eigenvektor der richtige sein soll.

Mit Hilfe von MAPLE² bilden wir die Differenz mit dem nächst liegenden Eigenwert, der der kleinste aus den zwei Lösungen einer quadratischen Gleichung ist. Wir setzen die Differenz gleich Null und lösen nach σ^2 :

$$\sigma^2 = \frac{Mv_x^2 + Lv_xv_y + Nv_y^2}{5(1 - A - B)(N_z^2 + AN_x^2 + BN_y^2)}, \quad (3.30)$$

wobei

$$\begin{aligned}M &= B \left(4N_z^4 + (N_z^2 - 3BN_y^2)^2 + AN_x^2(10N_z^2 + 14BN_y^2 + 5BN_x^2) \right) \\ L &= 20ABN_xN_y(N_z^2 - AN_x^2 - BN_y^2) \\ N &= A \left(4N_z^4 + (N_z^2 - 3AN_x^2)^2 + BN_y^2(10N_z^2 + 14AN_x^2 + 5BN_y^2) \right).\end{aligned}$$

Wir können nicht beweisen, daß σ^2 im allgemeinen positiv ist, wie wir es oben im Fall der frontalen Ebene bewiesen haben.

Wir machen einen Kompromiß zwischen der frontalen Ebene und der Ebene mit beliebiger Steigung und berücksichtigen zunächst das Verschwinden einer (z.B. N_y) von den zwei Steigungen der Ebene. Für $N_y = 0$ erhalten wir

$$\sigma^2 = \frac{v_x^2B(5N_z^4 + AN_x^2(10N_z^2 + 5BN_x^2)) + v_y^2A(4N_z^4 + (N_z^2 - 3AN_x^2)^2)}{5(1 - A - B)(N_z^2 + AN_x^2)}, \quad (3.31)$$

was positiv ist. Entsprechend läßt sich das gleiche für $N_x = 0$ beweisen. Von Interesse ist es zu untersuchen, ob die Existenz einer Steigung zur Stabilisierung des Problems

²System für algebraische Manipulationen

beiträgt, d.h. das zum Hervorrufen einer extremen Instabilität notwendige Rauschniveau erhöht. Dazu vergleichen wir

$$\sigma^2|_{N_x \neq 0, N_y = 0} - \sigma^2|_{N_x = N_y = 0} = \frac{AN_x^2 (5Bv_x^2(BN_x^2 + N_z^2) + Av_y^2(9AN_x^2 - 11N_z^2))}{5(1 - A - B)(N_z^2 + AN_x^2)}. \quad (3.32)$$

Wenn die Tiefenvariation parallel zur Translationsrichtung ($N_y = 0, v_y = 0$) ist, stabilisiert die Einbeziehung der Steigung die Lösung. Dagegen ist die obige Differenz nur bedingt positiv, wenn die Steigung senkrecht zur Translationsrichtung ($N_y = 0, v_x = 0$) ist.

Nach der Untersuchung dieser zwei Extremfälle (Bewegung parallel und senkrecht zur optischen Achse) werden wir die Translationsrichtung in der XZ Ebene wie in der Abb. 3.5 variieren lassen. Wir werden eine allgemeinere Beziehung zwischen der Diffe-

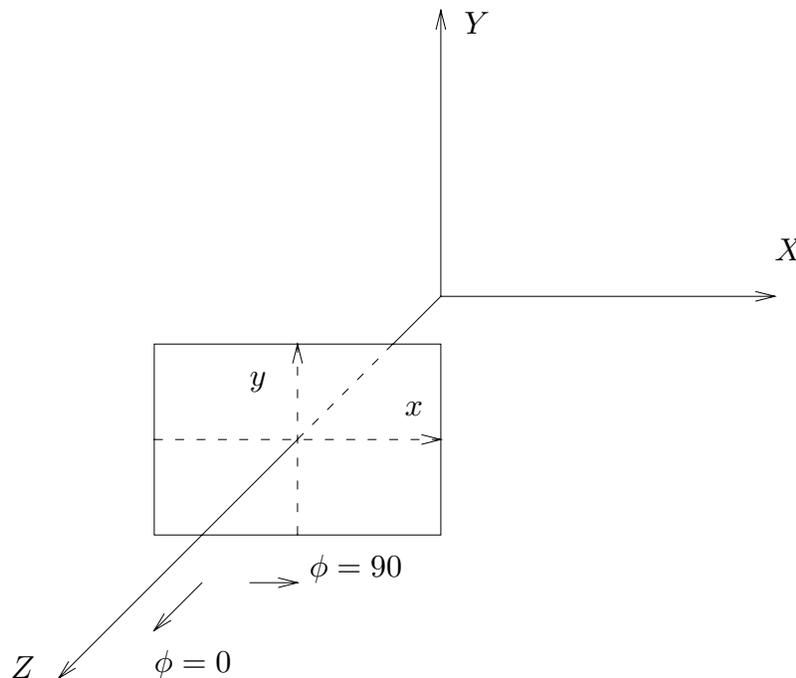


Abbildung 3.5: Variierende Translationsrichtung in der XZ Ebene

renz der kleinsten Eigenwerte und dem Winkel ϕ zwischen optischer Achse und Translation bereitstellen. Zur Vereinfachung der Ausdrücke nehmen wir an, daß $N_y = 0$ ist, weil wir nicht in der Lage gewesen sind, die Cardan'sche Lösung der charakteristischen

Gleichung im allgemeinen Fall ($N_x \neq 0, N_y \neq 0$) zu simplifizieren.

$$\bar{\Upsilon} = \begin{pmatrix} Bv_z^2(N_z^2 + AN_x^2) & 0 & -Bv_xv_z(N_z^2 + AN_x^2) \\ 0 & v_x^2(N_z^2 + AN_x^2) + Av_z^2(N_z^2 + 9/5AN_x^2) & 0 \\ -Bv_xv_z(N_z^2 + AN_x^2) & -4AN_xN_zv_xv_z & Bv_x^2(N_z^2 + AN_x^2) \end{pmatrix} + \Sigma \quad (3.33)$$

Wir wissen, daß die Lösung in der XZ Ebene liegen soll. Daher bleibt der Eigenwert $\lambda'_2 = \bar{\Upsilon}_{22}$ außer Betracht. Die Differenz der zwei kleinsten Eigenwerte ist

$$|\lambda'_1 - \lambda'_3| = \sqrt{(\bar{\Upsilon}_{11} - \bar{\Upsilon}_{33})^2 + 4\bar{\Upsilon}_{13}^2} \quad (3.34)$$

$$= \left(F^2 + G^2 + 2FG \frac{v_z^2 - v_x^2}{v_z^2 + v_x^2} \right)^{\frac{1}{2}} \quad (3.35)$$

$$= \left(F^2 + G^2 + 2FG \cos 2\phi \right)^{\frac{1}{2}}, \quad (3.36)$$

wobei

$$F = \sigma^2(1 - A - B) \quad (3.37)$$

$$G = B(N_z^2 + AN_x^2)(v_z^2 + v_x^2) \quad (3.38)$$

$$\mathbf{v} = \|\mathbf{v}\|(\sin \phi, 0, \cos \phi)^T. \quad (3.39)$$

Die absolute Differenz der Eigenwerte nimmt ab, wenn ϕ von 0° (\mathbf{v} parallel zur optischen Achse) auf 90° (\mathbf{v} parallel zur Bildebene) wächst. Je kleiner diese Differenz ist, desto größer sind die Störungen in dem geschätzten Eigenvektor.

Wir haben damit einen expliziten Zusammenhang zwischen der Instabilität und der Translationsrichtung nachgewiesen. Im Fall einer Bewegung parallel zur Bildebene kann eine geeignete Kombination von Meßrauschen, Größe des Gesichtsfelds und Verhältnis des Translationsbetrags zum Objektstand einen 90° -Fehler in der Translationsrichtung verursachen.

Diese gefährlichen Situationen können nicht nur in unseren Simulationen, sondern auch in der Realität auftreten. Die Projektion eines quer zum Beobachter fahrenden Autos möge nur eine kleine Fläche auf dem Bild ausfüllen – nehmen wir an, $1/10$ der Brennweite. Die sichtbare Seite des Autos werde als eben angenommen und seine Geschwindigkeit (pro Abtastintervall) sei $1/100$ seines Abstands zum Beobachter. Dann reicht (siehe Gl. (3.29)) ein Meßrauschen mit Varianz $\sigma^2 \approx 10^{-7}$, d.h. weniger als ein Pixel in einem 512×512 Bild, damit wir eine verzerrte Schätzung der Translationsrichtung wahrnehmen. Bemerkenswert ist, daß die Verzerrung die z -Komponente der Translation verstärkt, d.h. eine kleinere Zeit bis zur Kollision vortäuscht. Bei einem passiv navigierenden Fahrzeug würde diese Tatsache den Effekt haben, daß dauernd falscher “Kollisionsalarm” signalisiert wird.

Die Differenz zwischen den gestörten Eigenwerten λ'_1 und λ'_3 ist für die Wahrscheinlichkeit, daß die zwei Eigenwerte zusammenfallen, ausschlaggebend. Im weiteren werden wir eine obere Schranke für die Rotation des geschätzten Eigenvektors durch eine Störung ermitteln, die auch eine monoton wachsende Funktion vom Winkel ϕ zwischen Translationsrichtung und optischer Achse ist.

Seien A und $A' = A + E$ symmetrische Matrizen, so gilt für die Eigenwerte der gestörten Matrix ([Golub & van Loan 83, S. 269])

$$\lambda(A) + \lambda_{min}(E) \leq \lambda(A') \leq \lambda(A) + \lambda_{max}(E). \quad (3.40)$$

Aus (3.21) folgt für die Eigenwerte λ' der gestörten Matrix \tilde{Y} , daß

$$\lambda + \sigma^2(A + B) \leq \lambda' \leq \lambda + \sigma^2. \quad (3.41)$$

Wir suchen nach einem Ausdruck, der den Winkel zwischen dem gestörten und ungestörten Eigenvektor einschränkt. Eine obere Schranke für den Sinus des Winkels zwischen gestörtem und ungestörtem Eigenvektor wird von dem $\sin \theta$ Theorem von Davis und Kahan angegeben [Stewart 73], dessen allgemeinere Formulierung die Rotation von mehrdimensionalen Unterräumen durch eine Störung betrifft:

Es seien A und $A + E$ symmetrische Matrizen und θ der Winkel zwischen gestörtem und ungestörtem Eigenvektor, \mathbf{x}'_1 bzw. \mathbf{x}_1 , der Matrix A . Es sei δ der minimale Abstand des Eigenwerts λ_1 von den gestörten Eigenwerten ($\lambda'_2, \dots, \lambda'_n$). Dann gilt

$$|\sin \theta| \leq \frac{\|E\mathbf{x}_1\|}{\delta}. \quad (3.42)$$

Das Theorem unterstellt, daß sich die Anordnung der Eigenwerte nach ihrer Größe während der Störung nicht verändert, d.h. das Theorem gilt für die Rotation eines Eigenvektors, der dem kleinsten Eigenwert entspricht, solange die zwei kleinsten Eigenwerte nicht zusammenfallen.

Wir gehen vom Beispiel einer Translation $\mathbf{v} = \|\mathbf{v}\|(\sin \phi, 0, \cos \phi)^T$ in der XZ Ebene aus. Der ungestörte Eigenvektor, der dem Eigenwert 0 entspricht, ist $\mathbf{x}_1 = (\sin \phi, 0, \cos \phi)^T$ und die Störung der Matrix Y ist Σ . Daraus folgt

$$\|\Sigma\mathbf{x}_1\| = \sigma^2 \sqrt{\sin^2 \phi + (A + B)^2 \cos^2 \phi}. \quad (3.43)$$

Nehmen wir an, daß λ'_3 der gestörte Eigenwert ist, der näher zum kleinsten ungestörten Eigenwert $\lambda_1 = 0$ liegt, so lautet $\delta = \lambda'_3$. Wir setzen $\|\Sigma\mathbf{x}_1\|$ in (3.42) ein und erhalten

$$|\sin \theta| \leq \frac{\sigma^2 \sqrt{\sin^2 \phi + (A + B)^2 \cos^2 \phi}}{\lambda'_3}. \quad (3.44)$$

Aus (3.41) gilt $\lambda'_3 > \lambda_3 + \sigma^2(A + B)$. Ersetzen wir in (3.44) den Nenner durch einen kleineren Ausdruck, so bleibt die Ungleichung gültig.

$$|\sin \theta| \leq \frac{\sigma^2 \sqrt{\sin^2 \phi + (A + B)^2 \cos^2 \phi}}{\lambda_3 + \sigma^2(A + B)}. \quad (3.45)$$

Der gestörte kleinste Eigenwert λ'_1 liegt zwischen $\sigma^2(A+B)$ und σ^2 . Damit λ'_3 nicht mit λ'_1 zusammenfällt, ist $\lambda_3 + \sigma^2(A+B) > \sigma^2$ hinreichend. Weil auch der Term unter der Wurzel kleiner als eins ist, ist die obere Schranke in (3.45) kleiner als eins. Weiterhin ist der Zähler eine monoton wachsende Funktion von ϕ , während der ungestörte Eigenwert $\lambda_3 = B(N_z^2 + AN_x^2) \|\mathbf{v}\|^2$ unabhängig vom Winkel ist. D.h. die obere Schranke für die Rotation des gesuchten Eigenvektors durch eine Störung wächst mit zunehmendem Winkel zwischen Translationsrichtung und optischer Achse.

Wegen seiner korrekten geometrischen Interpretation und nach seinem Verhalten in unseren Experimenten vermuten wir, daß das Fehlermaß

$$\iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 dx dy \quad (3.46)$$

eine in der Translationsrichtung isotrope Schätzung erlaubt, können es aber wegen seiner hohen Nichtlinearität nicht explizit nachweisen. Wie auch die im Abschnitt 3.2 vorgestellten Normen kann dieser Ausdruck mit Hilfe von iterativen Verfahren minimiert werden, deren Konvergenz jedoch immer stark von den Startwerten abhängt.

Weil die Instabilitäten nur im Fall eines kleinen Gesichtsfeldes auftreten, kann man das \mathbf{x} im Nenner von (3.46) durch den Vektor \mathbf{c} zum Schwerpunkt des kleinen Gesichtsfelds ersetzen, wie es in [Spetsakis & Aloimonos 88] vorgeschlagen wird. Das führt zum Ausdruck

$$\iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x})\|} \right\}^2 d\mathbf{x} \approx \iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x})}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{c})\|} \right\}^2 d\mathbf{x}$$

und im Fall der reinen Translation erhält man

$$\iint_D \left\{ \frac{(\mathbf{v} \times \mathbf{x})^T \dot{\mathbf{x}}}{\|\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{c})\|} \right\}^2 d\mathbf{x} = \frac{\mathbf{v}^T \tilde{\Upsilon} \mathbf{v}}{\mathbf{v}^T C \mathbf{v}}, \quad (3.47)$$

wobei

$$C = \begin{pmatrix} c_z^2 & 0 & -c_x c_z \\ 0 & c_z^2 & -c_y c_z \\ -c_x c_z & -c_y c_z & c_x^2 + c_y^2 \end{pmatrix}$$

und $\tilde{\Upsilon}$ wie in (3.21). Der Bruch (3.47) würde ein verallgemeinerter Rayleigh-Quotient sein, wenn die Matrix C nicht singular wäre. Obwohl diese Vereinfachung das Problem in ein Eigenwertproblem umsetzt, werden wir zeigen, daß sie einen negativen Effekt auf den stabilen Fall der zur Blickrichtung parallelen Translation hat. Es genügt zu beobachten, daß der Nenner in (3.47) Null wird, wenn \mathbf{v} parallel zu \mathbf{c} ist, obwohl der Zähler wegen des Rauschens ungleich Null ist. Unter Benutzung desselben Modells wie früher nehmen wir an, daß die Verschiebungsraten meßfehlerbehaftet sind und von einer reinen Translation parallel zu $\mathbf{c} = (0, 0, 1)^T$ hervorgerufen werden. Daher wird die gestörte Matrix $\tilde{\Upsilon}$ diagonal (siehe Gl. (3.25):

$$\frac{\mathbf{v}^T \tilde{\Upsilon} \mathbf{v}}{\mathbf{v}^T C \mathbf{v}} = \frac{\lambda'_1 v_x^2 + \lambda'_2 v_y^2 + \lambda'_3 v_z^2}{v_x^2 + v_y^2}, \quad (3.48)$$

wobei λ'_i die Eigenwerte von $\bar{\Upsilon}$ in (3.25) sind. $(v_x, v_y, v_z)^T$ ist die zu schätzende Translationsgeschwindigkeit. Um die Singularität an der Stelle $\mathbf{v} = (0, 0, 1)^T$ zu vermeiden, schlagen [Spetsakis & Aloimonos 88] eine leichte Deformation der Matrix C vor, die für unseren Fall lautet

$$C_\delta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \delta \end{pmatrix}. \quad (3.49)$$

Die Minimierung von $\mathbf{v}^T \bar{\Upsilon} \mathbf{v} / \mathbf{v}^T C \mathbf{v}$ führt auf eine Eigenvektorberechnung, wenn man die Matrix C_δ in $\Delta \Delta^T$ zerlegt:

$$\frac{\mathbf{v}^T \bar{\Upsilon} \mathbf{v}}{\mathbf{v}^T C \mathbf{v}} = \frac{\mathbf{v}^T \bar{\Upsilon} \mathbf{v}}{\mathbf{v}^T \Delta \Delta^T \mathbf{v}} = \frac{\mathbf{v}^T \Delta^{-1} \bar{\Upsilon} \Delta^{-1^T} \mathbf{v}'}{\mathbf{v}'^T \mathbf{v}'} \Rightarrow \min \iff \mathbf{v}''^T \tilde{\Upsilon}_\delta \mathbf{v}'' \Rightarrow \min_{\|\mathbf{v}''\|=1}$$

wobei $\mathbf{v}' = \Delta^T \mathbf{v}$, $\mathbf{v}'' = \mathbf{v}' / \|\mathbf{v}'\|$ und

$$\tilde{\Upsilon}_\delta = \begin{pmatrix} \sigma^2 + B N_z^2 v_z^2 & 0 & 0 \\ 0 & \sigma^2 + A N_z^2 v_z^2 & 0 \\ 0 & 0 & \frac{\sigma^2}{\delta} (A + B) \end{pmatrix}. \quad (3.50)$$

Die früheren Differenzen (vgl. Gl. (3.26)) lauten jetzt

$$\begin{aligned} \lambda_1^\delta - \lambda_3^\delta &= B N_z^2 v_z^2 + \sigma^2 \left(1 - \frac{A + B}{\delta}\right) \\ \lambda_2^\delta - \lambda_3^\delta &= A N_z^2 v_z^2 + \sigma^2 \left(1 - \frac{A + B}{\delta}\right), \end{aligned}$$

die nicht mehr positiv sind, wenn der Deformationsparameter δ kleine Werte nimmt.

3.4 Fehlerempfindlichkeit in der Schätzung allgemeiner Bewegungen einer Ebene

In diesem Abschnitt werden wir die Fehlerempfindlichkeit in der Ermittlung von Rotation und Translation einer beliebigen Ebene untersuchen. Wir werden nachweisen,

- daß eine Translation senkrecht zur Ebene die Fehlerempfindlichkeit verstärkt,
- daß die Schätzung der Translation parallel zur Bildebene mit der Schätzung der Rotation um eine Achse parallel zur Bildebene gekoppelt ist,
- daß diese Kopplung stärker ist, wenn die Neigung der Ebene bezüglich der Bildebene und die Komponente der Translation in Richtung der optischen Achse klein sind,
- daß die Empfindlichkeit in der Schätzung aller Parameter umgekehrt proportional zur Größe des Gesichtsfelds und unabhängig von der Winkelgeschwindigkeit ist.

Das erste Ergebnis erscheint auf den ersten Blick im Gegensatz zu den Resultaten des Abschnittes 3.3 zu stehen. Dennoch besteht kein Widerspruch, weil hier das richtige Fehlermaß (3.3) benutzt wird. Weiterhin handelt es sich im Abschnitt 3.3 nur um die Translationsgeschwindigkeit als Unbekannte, wobei das Modell einer Ebene für die Modellierung der Verschiebungsrate angenommen wurde. Hier dagegen treten Translations- und Winkelgeschwindigkeit sowie auch die Normale der Ebene als Unbekannte auf.

Wir nehmen wie im Abschnitt 2.3 an, daß eine Ebene im Raum durch die Gleichung $N_x X + N_y Y + N_z Z = 1$ gegeben wird, wobei der Abstand der Ebene vom Ursprung gleich $1/\|\mathbf{N}\|$ ist. Zwischen dem Betrag der Normalen $\|\mathbf{N}\|$ und dem der Translationsgeschwindigkeit $\|\mathbf{v}\|$ besteht eine Skalierungsmehrdeutigkeit. Zur Veranschaulichung wiederholen wir die Ausdrücke für die Verschiebungsraten aus Abschnitt 2.3 Gl. (2.37) in Komponentenform:

$$\begin{aligned}\dot{x} &= (v_x N_z + \omega_y) + (v_x N_x - v_z N_z)x + (v_x N_y - \omega_z)y + (\omega_y - v_z N_x)x^2 + (-v_z N_y - \omega_x)xy \\ \dot{y} &= (v_y N_z - \omega_x) + (v_y N_x + \omega_z)x + (v_y N_y - v_z N_z)y + (\omega_y - v_z N_x)xy + (-v_z N_y - \omega_x)y^2.\end{aligned}\tag{3.51}$$

Wie wir schon im Abschnitt (2.3) beschrieben haben, besteht die Berechnung von \mathbf{v} , \mathbf{N} und $\boldsymbol{\omega}$ aus zwei Schritten.

Der erste Schritt ist die Berechnung der Hilfsparameter, die wir in einen Vektor \mathbf{q} zusammenfassen

$$\begin{aligned}\dot{\mathbf{x}} &= B\mathbf{q} = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{pmatrix} \mathbf{q} \quad \text{mit} \\ \mathbf{q} &= (v_x N_z + \omega_y, v_x N_x - v_z N_z, v_x N_y - \omega_z, v_y N_z - \omega_x, v_y N_x + \omega_z, \\ &\quad v_y N_y - v_z N_z, \omega_y - v_z N_x, -v_z N_y - \omega_x)^T,\end{aligned}\tag{3.52}$$

wobei die Verschiebungsrate $\dot{\mathbf{x}}$ ab jetzt als zweielementiger Vektor $\dot{\mathbf{x}} = (\dot{x}, \dot{y})^T$ angenommen wird. Der zweite Schritt besteht dann aus der Herleitung von \mathbf{v} , \mathbf{N} und $\boldsymbol{\omega}$ mit Hilfe des Eigensystems der Matrix (2.41).

Zur Untersuchung der Fehlerempfindlichkeit werden wir hier einen anderen Weg beschreiten, damit wir die Komplexität der Fehleranalyse bei der Berechnung von Eigenwerten und -vektoren vermeiden. Mit Hilfe der unteren Fehlerschranke nach Cramer-Rao werden wir eine Fehlerkovarianz für die geschätzten Parameter ermitteln, die eine untere Schranke für die Unsicherheit der Schätzung ist.

Wir fassen die Unbekannten in einen Parametervektor \mathbf{p} zusammen und bezeichnen die Menge aller gemessenen Verschiebungsraten mit \mathcal{Z} . Die bedingte Wahrscheinlichkeitsdichte dieser Messungen bezeichnen wir mit $p(\mathcal{Z}|\mathbf{p})$. Die Fishersche Informationsmatrix wird dann wie folgt definiert [Sorenson 80]:

$$F = E\left[\frac{\partial \ln p(\mathcal{Z}|\mathbf{p})}{\partial \mathbf{p}} \frac{\partial \ln p(\mathcal{Z}|\mathbf{p})}{\partial \mathbf{p}}^T\right],\tag{3.53}$$

wobei die Ableitung einer skalaren Größe nach einem Vektor als Zeilenvektor angenommen wird. Die Unsicherheit eines Schätzers $\hat{\mathbf{p}}$ wird durch die Fehlerkovarianzmatrix $E[(\mathbf{p} - \hat{\mathbf{p}})(\mathbf{p} - \hat{\mathbf{p}})^T]$ angegeben. Die Cramer-Rao-Ungleichung besagt, daß die Fehlerkovarianz eines erwartungstreuen Schätzers nach unten durch die Inverse der Fisherschen

Informationsmatrix beschränkt wird:

$$E[(\mathbf{p} - \hat{\mathbf{p}})(\mathbf{p} - \hat{\mathbf{p}})^T] \geq F^{-1}. \quad (3.54)$$

Die Ungleichung bei den obigen Matrizen bedeutet, daß die Differenz zwischen der Matrix der linken Seite und der Matrix der rechten Seite positiv semidefinit ist. Weil die Diagonalelemente einer positiv semidefiniten Matrix positiv oder gleich Null sind, kann man direkt skalare untere Schranken für die Fehlervarianzen der Komponenten des Parametervektors \mathbf{p} ermitteln. Die Fishersche Informationsmatrix ist aber als Matrix reicher an Information, die wir ausnutzen werden, damit wir sowohl die empfindlichsten als auch die robustesten Richtungen im Parameterraum bestimmen. Im optimistischen Fall, wenn die tatsächliche Fehlerkovarianz des Schätzers die untere Schranke erreicht, läßt sich die Unsicherheit geometrisch durch folgendes Unsicherheitsellipsoid um den Schätzwert $\hat{\mathbf{p}}$ darstellen

$$(\mathbf{p} - \hat{\mathbf{p}})^T F (\mathbf{p} - \hat{\mathbf{p}}) = c, \quad (3.55)$$

wobei die Konstante c die Ausdehnung des Fehlerellipsoids darstellt und die Wahrscheinlichkeit angibt, daß der tatsächliche Parametervektor innerhalb dieses Ellipsoids liegt. Von Interesse sind die Richtungen der Hauptachsen des Fehlerellipsoids, die durch die Richtungen der Eigenvektoren von F angegeben werden. Die Länge einer Hauptachse ist gleich $\sqrt{(c/\lambda)}$, wobei λ Eigenwert von F ist. Die unsicherste Richtung im Parameterraum wird durch den Eigenvektor bestimmt, der dem kleinsten Eigenwert entspricht, und die sicherste Richtung durch den, der dem größten Eigenwert entspricht, d.h., daß Projektionen des Parametervektors auf diese Richtungen am empfindlichsten bzw. am wenigsten empfindlich sind. Dies gibt uns – wie wir später an unserem Problem zeigen werden – eine Einsicht in den Schätzvorgang, die uns erlaubt, robuste lineare Kombinationen von Parametern zu erkennen, auch wenn die Schätzung jedes Parameters für sich unsicher ist.

Um die Fishersche Informationsmatrix für unser Problem zu berechnen, gehen wir davon aus, daß das Meßrauschen in den Verschiebungsvektoren normalverteilt mit Meßkovarianz³ $\sigma^2 I$ ist. Weiterhin fassen wir die rechte Seite von (3.51) in die Meßfunktion $\mathbf{h}(\mathbf{p})$ zusammen. Die bedingte Wahrscheinlichkeitsdichte lautet dann

$$p(\mathcal{Z}|\mathbf{p}) = \frac{1}{k} \exp\left(-\frac{1}{2\sigma^2} \iint_D \|\dot{\mathbf{x}} - \mathbf{h}(\mathbf{p})\|^2 dx dy\right), \quad (3.56)$$

wobei die Integration über das Gebiet D der Abbildung der sich relativ bewegenden Szenenkomponente – auch effektives Gesichtsfeld genannt – angewandt wird (siehe auch Abschnitt 3.3). Die Konstante k dient der Normierung der Wahrscheinlichkeitsdichte. Daher gilt

$$\frac{\partial \ln p(\mathcal{Z}|\mathbf{p})}{\partial \mathbf{p}} = \frac{1}{\sigma^2} \iint_D (\dot{\mathbf{x}} - \mathbf{h}(\mathbf{p}))^T \frac{\partial \mathbf{h}}{\partial \mathbf{p}} dx dy \quad (3.57)$$

und

$$F = E\left[\frac{\partial \ln p(\mathcal{Z}|\mathbf{p})}{\partial \mathbf{p}} \frac{\partial \ln p(\mathcal{Z}|\mathbf{p})}{\partial \mathbf{p}}^T\right] = \frac{1}{\sigma^2} \iint_D \frac{\partial \mathbf{h}^T}{\partial \mathbf{p}} \frac{\partial \mathbf{h}}{\partial \mathbf{p}} dx dy. \quad (3.58)$$

³Wir werden in dieser Untersuchung ein isotropes Meßrauschen der Verschiebungsraten annehmen. In der Wirklichkeit ist die Meßunsicherheit abhängig von dem jeweiligen angewandten Verfahren zur Berechnung des optischen Flusses.

Daraus folgt, daß die Inverse der Fisherschen Informationsmatrix proportional zur Varianz σ^2 des Meßrauschens ist. Es war auch zu erwarten, daß die untere Fehlerschranke der Schätzung mit wachsendem Meßrauschen steigt. Ab jetzt werden wir deshalb den Faktor $1/\sigma^2$ weglassen. Aus Gl. (3.52) folgt, daß

$$\frac{\partial \mathbf{h}}{\partial \mathbf{p}} = \frac{\partial \mathbf{h}}{\partial \mathbf{q}} \frac{\partial \mathbf{q}}{\partial \mathbf{p}} = B \frac{\partial \mathbf{q}}{\partial \mathbf{p}},$$

und daß die Jacobische Matrix $\frac{\partial \mathbf{q}}{\partial \mathbf{p}}$ unabhängig von den Koordinaten (x, y) ist. So gilt

$$F = \frac{\partial \mathbf{q}^T}{\partial \mathbf{p}} \left\{ \iint_D B^T B dx dy \right\} \frac{\partial \mathbf{q}}{\partial \mathbf{p}}. \quad (3.59)$$

Für den Integranden der obigen Gleichung ist

$$B^T B = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ x & x^2 & xy & 0 & 0 & 0 & x^3 & x^2y \\ y & xy & y^2 & 0 & 0 & 0 & x^2y & y^2x \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \\ 0 & 0 & 0 & x & x^2 & xy & x^2y & y^2x \\ 0 & 0 & 0 & y & xy & y^2 & y^2x & y^3 \\ x^2 & x^3 & x^2y & xy & x^2y & y^2x & x^4 + x^2y^2 & x^3y + xy^3 \\ xy & x^2y & y^2x & y^2 & y^2x & y^3 & x^3y + xy^3 & x^2y^2 + y^4 \end{pmatrix}.$$

Die Integrale erstrecken sich über das Rechteck in Abbildung 3.4 mit Seitenlängen α und β . Wie im Abschnitt 3.3 ersetzen wir $\alpha^2/12$ durch A und $\beta^2/12$ durch B:

$$B_{integral} = \iint_D B^T B dx dy = \alpha\beta \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & A & 0 \\ 0 & A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & B & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & B \\ 0 & 0 & 0 & 0 & A & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & B & 0 & 0 \\ A & 0 & 0 & 0 & 0 & 0 & A(\frac{2}{5}A + B) & 0 \\ 0 & 0 & 0 & B & 0 & 0 & 0 & B(A + \frac{2}{5}B) \end{pmatrix}.$$

Die Inverse der Fisherschen Informationsmatrix wird im Nenner den Faktor $\alpha\beta$ enthalten, wodurch die Unsicherheit umgekehrt proportional zur Fläche des effektiven Gesichtsfelds wird. Nachdem wir diese Aussage hergeleitet haben, werden wir der Übersichtlichkeit halber auch diesen Faktor weglassen.

Bevor wir weiter auf die Jacobische Matrix eingehen, müssen wir festlegen, welche acht aus den neun Unbekannten in \mathbf{v} , $\boldsymbol{\omega}$ und \mathbf{N} wir beibehalten werden. Zur Eliminierung der Skalierungsmehrdeutigkeit zwischen den Beträgen von \mathbf{v} und \mathbf{N} nehmen wir an, daß die dritte Komponente der Normale nicht verschwindet, was geometrisch

der Tatsache entspricht, daß die Ebene die optische Achse immer schneidet. Aufgrund dieser Annahme führen wir folgende Substitutionen durch:

$$\begin{aligned} N'_x &= \frac{N_x}{N_z} & v'_x &= v_x N_z \\ N'_y &= \frac{N_y}{N_z} & v'_y &= v_y N_z \\ & & v'_z &= v_z N_z. \end{aligned} \quad (3.60)$$

Der Übersichtlichkeit halber werden wir Symbole ohne Strich verwenden. Daher steht der Betrag von \mathbf{v} für das Verhältnis des Betrags der Translationsgeschwindigkeit zum Abstand der Kamera vom Durchstoßpunkt der optischen Achse mit der Ebene.

Der Parametervektor lautet dann

$$\mathbf{p} = \left(v_x \ v_y \ v_z \ \omega_x \ \omega_y \ \omega_z \ N_x \ N_y \right),$$

und die Jacobische Matrix lautet

$$\frac{\partial \mathbf{q}}{\partial \mathbf{p}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ N_x & 0 & -1 & 0 & 0 & 0 & v_x & 0 \\ N_y & 0 & 0 & 0 & 0 & -1 & 0 & v_x \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & N_x & 0 & 0 & 0 & 1 & v_y & 0 \\ 0 & N_y & -1 & 0 & 0 & 0 & 0 & v_y \\ 0 & 0 & -N_x & 0 & 1 & 0 & -v_z & 0 \\ 0 & 0 & -N_y & -1 & 0 & 0 & 0 & -v_z \end{pmatrix}. \quad (3.61)$$

Um die Regularität der Fisherschen Informationsmatrix zu untersuchen, werden wir ihre Determinante berechnen. Aus (3.59) folgt

$$\det(F) = \det^2\left(\frac{\partial \mathbf{q}}{\partial \mathbf{p}}\right) \det(B_{integral}). \quad (3.62)$$

Es gilt

$$\det(B_{integral}) = \frac{1}{25} A^3 B^3 (4A + 5B)(5A + 4B),$$

daher kann nur die Jacobische Matrix singularär werden.

Diese Determinante läßt sich durch wiederholtes Addieren und Subtrahieren der Zeilen der Jacobischen Matrix herleiten und lautet

$$\det\left(\frac{\partial \mathbf{q}}{\partial \mathbf{p}}\right) = \|\mathbf{N} \times \mathbf{v}\|^2, \quad (3.63)$$

wobei wie vereinbart das dritte Element der Normale gleich eins ist. Daraus folgt, daß die untere Schranke der Unsicherheit im Unendlichen liegt, wenn die Translationsgeschwindigkeit senkrecht zur Ebene ist.

Weiterhin kann man anhand von (3.61) erkennen, daß die untere Fehlerschranke unabhängig von der Winkelgeschwindigkeit ist.

Bis jetzt haben wir vier Sachverhalte über die Fehlerempfindlichkeit hergeleitet: ihre Beziehung zum Meßrauschen, zur Größe des Gesichtsfelds, zum Winkel zwischen \mathbf{v} und \mathbf{N} und ihre Unabhängigkeit von der Winkelgeschwindigkeit. Unser nächstes Anliegen besteht darin, die Fishersche Informationsmatrix zu invertieren, so daß wir die Unsicherheit jeder Komponente des geschätzten Parametervektors ermitteln. Unsere Untersuchung wird von der auf experimentellen Ergebnissen basierenden Vermutung motiviert, daß Translationen parallel zur Bildebene von Rotationen um Achsen parallel zur Bildebene schwer zu unterscheiden sind. Deshalb wollen wir die Richtungen im (v_x, ω_y) -Raum herleiten (die Vorgehensweise für den (v_y, ω_x) -Raum ist analog), für die die Projektion des Parametervektors robust berechnet werden kann.

Wir berechnen die in (3.59) angegebene Fishersche Informationsmatrix und schreiben sie wie folgt um:

$$F = \begin{pmatrix} K & L \\ L^T & M \end{pmatrix}, \quad (3.64)$$

wobei K, L, M lauten

$$K = \begin{pmatrix} K_1 & K_2 \end{pmatrix}$$

mit

$$K_1 = \begin{pmatrix} 1 + AN_x^2 + BN_y^2 & 0 & -2N_xA \\ 0 & 1 + AN_x^2 + BN_y^2 & -2N_yB \\ -2N_xA & -2N_yB & A + B + AB(N_x^2 + N_y^2) + \frac{9}{5}(A^2N_x^2 + B^2N_y^2) \\ 0 & -(1 + B) & BN_y(1 + A + \frac{9}{5}B) \\ 1 + A & 0 & -AN_x(1 + B + \frac{9}{5}A) \\ -BN_y & AN_x & 0 \end{pmatrix},$$

$$K_2 = \begin{pmatrix} 0 & 1 + A & -BN_y \\ -(1 + B) & 0 & AN_x \\ BN_y(1 + A + \frac{9}{5}B) & -AN_x(1 + B + \frac{9}{5}A) & 0 \\ 1 + B(2 + A + \frac{9}{5}B) & 0 & 0 \\ 0 & 1 + A(2 + B + \frac{9}{5}A) & 0 \\ 0 & 0 & A + B \end{pmatrix}$$

und

$$L = \begin{pmatrix} A(N_xv_x - v_z) & Bv_xN_y \\ AN_xv_y & B(N_yv_y - v_z) \\ AN_xv_z(B + \frac{9}{5}A) - Av_x & BN_yv_z(A + \frac{9}{5}B) - Bv_y \\ 0 & Bv_z(1 + A + \frac{9}{5}B) \\ -Av_z(1 + B + \frac{9}{5}A) & 0 \\ Av_y & Bv_x \end{pmatrix}$$

$$M = \begin{pmatrix} A(v_x^2 + v_y^2) + Av_z^2(\frac{9}{5}A + B) & 0 \\ 0 & B(v_x^2 + v_y^2) + Bv_z^2(\frac{9}{5}B + A) \end{pmatrix}.$$

Um die Inversion durchzuführen verwenden wir folgende Formel⁴ aus der Regelungstechnik [Friedland 86]

$$\begin{pmatrix} K & L \\ L^T & M \end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}LM^{-1} \\ -M^{-1}L^TE^{-1} & M^{-1} + M^{-1}L^TE^{-1}LM^{-1} \end{pmatrix}$$

mit $E = (K - LM^{-1}L^T)$ (3.65)

Weil wir die Unsicherheit in Abhängigkeit von den Richtungen der Translationsgeschwindigkeit und der Normalen ermitteln wollen, spezialisieren wir auf $v_y = 0$, $N_y = 0$, damit wir die Zahl der beeinflussenden Faktoren reduzieren. Wir führen als Richtungsparameter ψ und χ ein wie folgt:

$$\begin{aligned} \mathbf{v} &= \begin{pmatrix} v_x & 0 & v_z \end{pmatrix} \\ &= \begin{pmatrix} \|\mathbf{v}\| \sin \psi & 0 & \|\mathbf{v}\| \cos \psi \end{pmatrix} \\ \mathbf{N} &= \begin{pmatrix} N_x & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \tan \chi & 0 & 1 \end{pmatrix}, \end{aligned}$$

d.h. wir erlauben, daß die Translation und die Ebenennormale nur in der XZ -Ebene liegen. ψ und χ sind jeweils die Winkel, die \mathbf{v} und \mathbf{N} mit der optischen Achse bilden: $-\pi/2 \leq \psi \leq \pi/2$ und $-\pi/2 \leq \chi < \pi/2$). Aus (3.63) folgt, daß die Fishersche Matrix singular wird, wenn $\psi = \chi$ ist.

Durch die obige Einschränkung des Parameterraums erfolgt eine Zerlegung der Matrix E in (3.65) in zwei Blockmatrizen E_{135} und E_{246} , wenn man die Reihenfolge der ersten sechs Unbekannten von $(v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)$ auf $(v_x, v_z, \omega_y, v_y, \omega_x, \omega_z)$ ändert:

$$E = \begin{pmatrix} E_{135} & 0 \\ 0 & E_{246} \end{pmatrix}.$$

Daraus folgt

$$E^{-1} = \begin{pmatrix} E_{135}^{-1} & 0 \\ 0 & E_{246}^{-1} \end{pmatrix}. \quad (3.66)$$

Die Unbekannten der Bewegung unterteilen sich dadurch in zwei Gruppen (v_x, v_z, ω_y) und $(v_y, \omega_x, \omega_z)$, deren Schätzung voneinander entkoppelt ist.

Die Berechnung der Inversen E_{123}^{-1} und E_{246}^{-1} wurde mit Hilfe von MAPLE durchgeführt.

Wir werden die erste Gruppe von Parametern (v_x, v_z, ω_y) behandeln. Ab jetzt nehmen wir an, daß $A = B$ ist, d.h. daß das effektive Gesichtsfeld ein Quadrat ist. Die Fehlerkovarianzen der Parameter (v_x, v_z, ω_y) werden von den Diagonalelementen der Inversen E_{135}^{-1} angegeben, wie folgt:

$$(E_{135}^{-1})_{11} = \frac{10A + 8A \cos^2 \psi + (14A^2 + 5) \sin^2 \psi + 9A \tan \chi \sin \psi (\cos \psi - \tan \chi \sin \psi)}{9A^2 (\tan \chi \cos \psi - \sin \psi)^2}$$

⁴Professor Nagel wies mich auf die Existenz dieser Formel hin.

$$(E_{135}^{-1})_{22} = \frac{1}{A}$$

$$(E_{135}^{-1})_{33} = \frac{5 \sin^2 \psi + 18A \cos^2 \psi + 28A \tan \chi \sin \psi \cos \psi + A \tan^2 \chi (14A \cos^2 \psi + 9 \sin^2 \psi)}{9A^2 (\tan \chi \cos \psi - \sin \psi)^2}$$

Alle Elemente der Inversen (diagonal und nicht-diagonal) sind unabhängig vom Betrag $\|\mathbf{v}\|$, der nach der Ersetzung (3.60) das Verhältnis des Betrags von \mathbf{v} zu N_z vertritt.

Das Diagonalelement $(E_{135}^{-1})_{22}$, welches gleich der Fehlerkovarianz der z -Komponente der Translationsgeschwindigkeit ist, ist gleich $1/A$, also unabhängig von der Bewegungs- und Geometrieconfiguration und wird nicht von der Singularität $\psi = \chi$ betroffen.

Die Unsicherheit in den Parametern v_x und ω_y werden wir durch die Einführung des Einheitsvektors $\mathbf{u} = (\cos \phi, 0, \sin \phi)^T$ untersuchen. Die quadratische Form $\mathbf{u}^T E_{135}^{-1} \mathbf{u}$ stellt die Unsicherheit in Richtung ϕ in der Ebene (v_x, ω_y) dar. Geometrisch läßt sich die Unsicherheit in einem solchen Unterraum des Parameterraums als ein Schnitt des Unsicherheitsellipsoids mit einer Ebene interpretieren. Für die quadratische Form gilt [Golub & van Loan 83]

$$\lambda_{min}(S) \leq \mathbf{u}^T E_{135}^{-1} \mathbf{u} \leq \lambda_{max}(S), \quad (3.67)$$

wobei S die 2×2 Untermatrix von E_{135}^{-1} darstellt, die aus der ersten und dritten Spalte und Zeile der Matrix E_{135}^{-1} gebildet wird. Die zwei Werte von ϕ , für die die obige quadratische Form die untere bzw. die obere Grenze erreicht, sind die Richtungen der zueinander orthogonalen Eigenvektoren von S . Wir sind an der Richtung der kleinsten Unsicherheit interessiert, also müssen wir den Eigenvektor mit dem kleinsten Eigenwert bestimmen. Mit Hilfe von MAPLE werden der Winkel ϕ für diesen Eigenvektor sowie der kleinste Eigenwert $\lambda_{min}(S)$ berechnet – die entsprechenden Ausdrücke sind zu lang und unübersichtlich, deshalb werden wir direkt ihre Verläufe in Bezug auf ψ und χ darstellen.

Der kleinste Eigenwert wird in der Abbildung 3.6 aufgezeichnet. Man kann feststellen, daß der kleinste Eigenwert von der Singularität $\psi = \chi$ nicht betroffen wird. Trotzdem werden die Fehlervarianzen von v_x und ω_y in der ersten und dritten Gleichung in (3.67) betroffen, die an der Stelle der Singularität auch unendlich werden. Diese Tatsache bestätigt die Bedeutung unserer Vorgehensweise, indem wir uns nicht auf die Fehlervarianzen jedes Elements beschränken, sondern darüber hinaus nach Unsicherheitsrichtungen suchen. Aus der obigen Unbetroffenheit des kleinsten Eigenwerts ergibt sich das Ergebnis, daß die Projektion von $(v_x, \omega_y)^T$ auf die Richtung ϕ_{min} des Eigenvektors mit dem kleinsten Eigenwert immer berechnet werden kann. Die entsprechende Fehlervarianz in dieser Richtung – gleich dem kleinsten Eigenwert – verläuft in Bezug auf die Bewegung (\mathbf{v}) und die Geometrie (\mathbf{N}) wie in der (Abb. 3.6). Sie ist am größten, wenn ψ nahe bei χ liegt, und nimmt die kleinsten Werte an, wenn die Translation die Richtung der optischen Achse hat und die Ebenennormale senkrecht zur optischen Achse steht.

Wie die robusteste Richtung verläuft, zeigt der funktionale Zusammenhang des Winkels ϕ_{min} mit der Richtung der Translationsgeschwindigkeit und der Ebenennormalen, der in den Abbildungen 3.7a-b aufgezeichnet wird.

Beim kleinen effektiven Gesichtsfeld (Abb. 3.7a) nimmt der Winkel ϕ_{min} für den größten Teil des Definitionsbereichs Werte aus der engeren Nachbarschaft von $\pi/4$ ein. Das heißt, daß die robuste Richtung $(\cos \pi/4, 0, \sin \pi/4)^T$ ist, woraus folgt, daß die

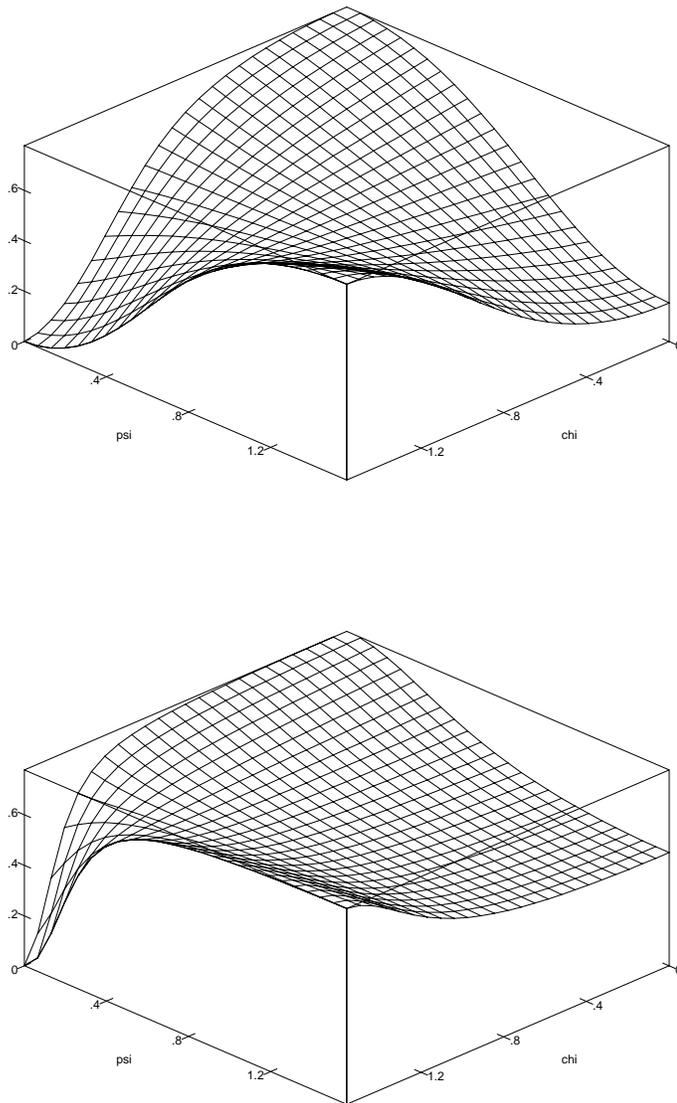
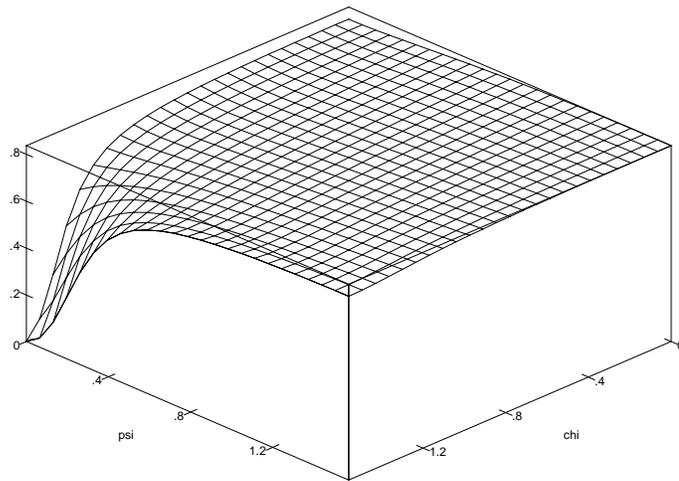
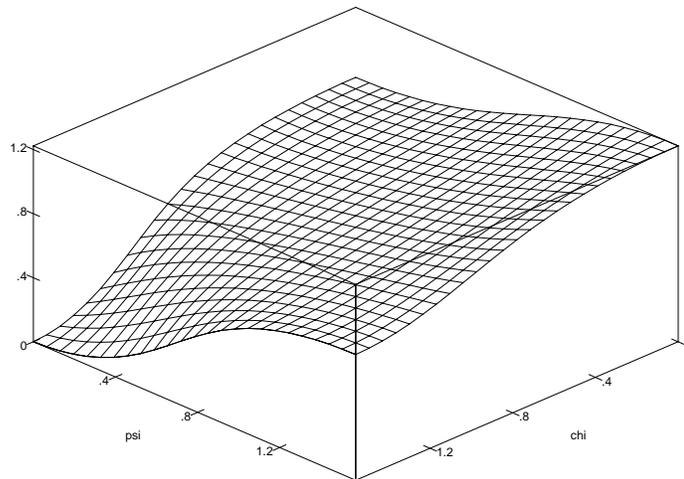


Abbildung 3.6: Der kleinste Eigenwert $\lambda_{min}(S)$ als Funktion von den Abweichungen ψ der Translation und χ der Ebenennormalen von der optischen Achse für zwei Größen des Gesichtsfelds: $A = 1.0$ (oben) und $A = 0.1$ (unten). Der kleinste Eigenwert gibt die Größe der Unsicherheit in der robustesten Richtung in der (v_x, ω_y) -Ebene an.

Summe $v_x + \omega_y$ robust berechnet werden kann. Wenn der Winkel ϕ_{min} auf Null fällt, bedeutet dies, daß die sicherste Richtung dazu tendiert, parallel zu $(1, 0, 0)^T$ zu werden. Das geschieht, wenn χ gegen $\pi/2$ (Ebenennormale senkrecht zur optischen Achse) und ψ gegen 0 (Translation parallel zur optischen Achse) geht. Anschaulich könnte man das wie folgt interpretieren: Translationen in Richtung der optischen Achse ergeben ein radial erscheinendes Verschiebungsratenfeld, Ebenen mit großer Neigung ergeben Verschiebungsraten, die eine höhere räumliche Änderungsrate haben. In beiden Fällen sieht ein solches Verschiebungsratenfeld erheblich anders als das Feld aus, das von einer



(a)



(b)

Abbildung 3.7: Der Winkel ϕ_{min} als Funktion der Abweichungen ψ der Translation und χ der Ebenennormalen von der optischen Achse für zwei Größen des Gesichtsfelds: $A = 0.1$ (a) und $A = 1.0$ (b). Der Winkel ϕ_{min} gibt die Richtung der kleinsten Unsicherheit in der (v_x, ω_y) -Ebene an. Bei kleinem Gesichtsfeld (a) ist er fast überall gleich $\pi/4$, d.h. $v_x + \omega_y$ kann immer robust berechnet werden. Dieser Effekt wird bei größerem Gesichtsfeld (b) abgeschwächt.

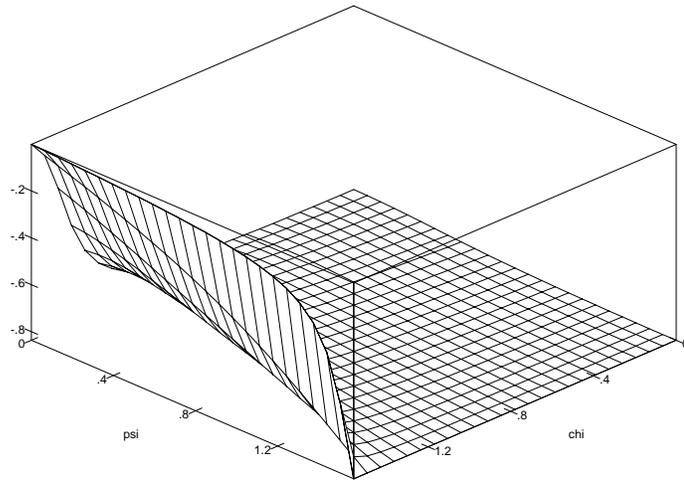
Rotation um eine Achse parallel zur Bildebene verursacht wird.

Der Effekt einer dominierenden Richtung im Parameterraum wird beim größeren effektiven Gesichtsfeld abgeschwächt. In Abb. 3.7b kann man sehen, daß der Winkel Werte größer als $\pi/4$ nehmen kann, nähert sich aber nicht $\pi/2$, was eine robustere Schätzung für die Winkelgeschwindigkeit ω_y ergeben würde.

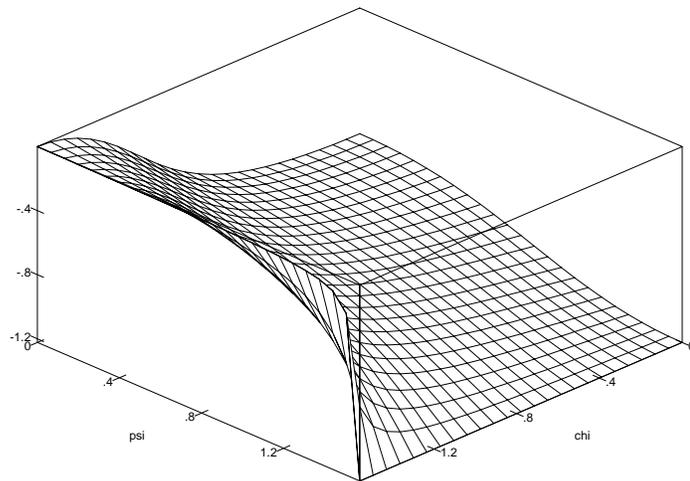
Entsprechende Aussagen lassen sich für die zweite Gruppe von Parametern $(v_y, \omega_x, \omega_z)$ herleiten, indem man die zweite Untermatrix von (3.66) invertiert. Man sucht dann nach der am wenigsten empfindlichen Richtung in der Ebene (v_y, ω_x) . Der Verlauf des Winkels ϕ_{min} wird in der Abb. 3.8 dargestellt. Man sieht, daß beim kleinen Gesichtsfeld die Mehrheit der Werte in der Umgebung von $-\pi/4$ liegt, so daß die am wenigsten empfindliche Richtung im Parameterraum $(v_y, \omega_y, \omega_z)$ von $(\cos \pi/4, -\sin \pi/4, 0)^T$ angegeben wird, während dieser Effekt beim großen Gesichtsfeld nicht mehr abgeschwächt wird. Es folgt analog zur Untersuchung über die Parametergruppe (v_x, v_z, ω_y) , daß beim kleinen effektiven Gesichtsfeld nur die lineare Kombination $-v_y + \omega_x$ robust berechnet werden kann.

Wir fassen die wichtigsten Punkte dieses Abschnitts zusammen:

- Mit Hilfe der Cramer-Rao-Ungleichung haben wir eine untere Schranke für die Fehlerkovarianz der geschätzten Parameter ermittelt.
- Die unteren Fehlerschranken sind umgekehrt proportional zur Größe des effektiven Gesichtsfelds und unabhängig von der Form der Rotation.
- Eine Singularität tritt auf, wenn die Translation parallel zur Ebenennormalen ist, was untere Fehlerschranken von unendlicher Größe verursacht.
- Durch eine Einschränkung der Art der Translation und der Lage der Ebenennormalen haben wir die Unterteilung der Parameter in zwei Gruppen gezeigt, deren Empfindlichkeiten voneinander unabhängig sind. Wir haben die Richtungen im Parameterraum ermittelt, die die am robustesten schätzbaren Kombinationen von Parametern ergeben.



(a)



(b)

Abbildung 3.8: Der Winkel ϕ_{min} als Funktion der Abweichungen ψ der Translation und χ der Ebenennormalen von der optischen Achse für zwei Größen des Gesichtsfelds: $A = 0.1$ (a) und $A = 1.0$ (b). Der Winkel ϕ_{min} gibt die Richtung der kleinsten Unsicherheit in der (v_y, ω_x) -Ebene an. Bei kleinem Gesichtsfeld (a) ist er fast überall gleich $-\pi/4$, d.h. $-v_y + \omega_x$ kann immer robust berechnet werden. Dieser Effekt wird bei größerem Gesichtsfeld (b) abgeschwächt.

3.5 Fehlerempfindlichkeit in der Ermittlung der Struktur

In diesem Abschnitt werden wir die Fehlerempfindlichkeit in der Ermittlung der Geometrie der sich relativ zur Kamera bewegenden Szenenkomponenten ermitteln. Aufbauend auf Ergebnisse des letzten Abschnitts werden wir zuerst die Empfindlichkeit in der Schätzung der Normalen einer Ebene untersuchen. Anschließend werden wir auf die Berechnung der Tiefe aus der Verschiebungsrate eines Bildpunkts eingehen.

Ermittlung der Normalen einer Ebene in Bewegung

Die Information über die Fehlerempfindlichkeit der Parameter (N_x, N_y) einer Ebene liegt in der rechten unteren Untermatrix der Inversen der Fisherschen Informationsmatrix (3.65), die als Diagonalelemente die unteren Schranken für die Fehlervarianzen von N_x und N_y enthält. Um diese Untermatrix – ab jetzt wird sie mit D bezeichnet – zu berechnen, verwenden wir die Matrixidentität (3.65) und erhalten

$$D = M^{-1} + M^{-1}L^T E^{-1}LM^{-1} \quad (3.68)$$

Die Berechnung der Matrix D wird mit Hilfe von MAPLE durchgeführt.

Wir spezialisieren wieder das Problem, indem wir dieselben Annahmen ($v_y = 0$, $N_y = 0$) wie im letzten Abschnitt machen. Dadurch wird D diagonal, und ihre Elemente lauten:

$$D_{11} = \frac{18A + (5 + 28A + 14A^2) \tan^2 \chi + 9A \tan^4 \chi}{9A^2 \|\mathbf{v}\|^2 (\tan \chi \cos \psi - \sin \psi)} \quad (3.69)$$

$$D_{22} = \frac{18A + (5 + 28A + 14A^2) \tan^2 \chi}{9A^2 \|\mathbf{v}\|^2 (\tan \chi \cos \psi - \sin \psi)} \quad (3.70)$$

Wir stellen fest, daß die untere Schranke für die Fehlerkovarianz sowohl von N_x als auch von N_y unendlich wird, wenn $\psi = \chi$ wird. Weiterhin ist die Empfindlichkeit umgekehrt proportional zum Quadrat des Betrags der Translation. Dieses ist auch plausibel, weil bei abwesender Translation keine Strukturberechnung möglich ist.

Um die Beziehung zur Richtung der Translation ψ und zur Neigung der Ebene χ zu ermitteln, wird der Verlauf der Fehlervarianzen in Bezug auf ψ und χ aufgezeichnet. Weil der Wertebereich wegen der oben erwähnten Singularität bis ins Unendliche geht, wird in Abb. (3.9) der arctan der Fehlervarianz dargestellt.

Wir beobachten, daß N_x und N_y robuster berechnet werden können, wenn die Translation parallel zur Bildebene ist und die Ebene eine sehr kleine Neigung hat. Weiterhin stellt man fest, daß die Unsicherheit in der Schätzung der Neigung N_x unendlich wird, wenn $\chi = \pi/2$ ist, was darauf zurückzuführen ist, daß in diesem Fall der sichtbare Anteil der Ebene fern von der Kamera liegt. Damit ergibt sich ein gegenläufiges Verhalten der Fehlerempfindlichkeit in der Schätzung der Bewegung und der Struktur. Bewegungen und Neigungen der Ebene, die eine robuste Bewegungsberechnung ermöglichen, verursachen eine höhere Empfindlichkeit in der Schätzung der Struktur.

Tiefenberechnung aus den Verschiebungsraten

Wie wir schon in der Problemstellung (Abschnitte 2.2-2.3) beschrieben haben, kann man die Lage eines Punktes im Raum nach der Berechnung der Bewegungsparameter

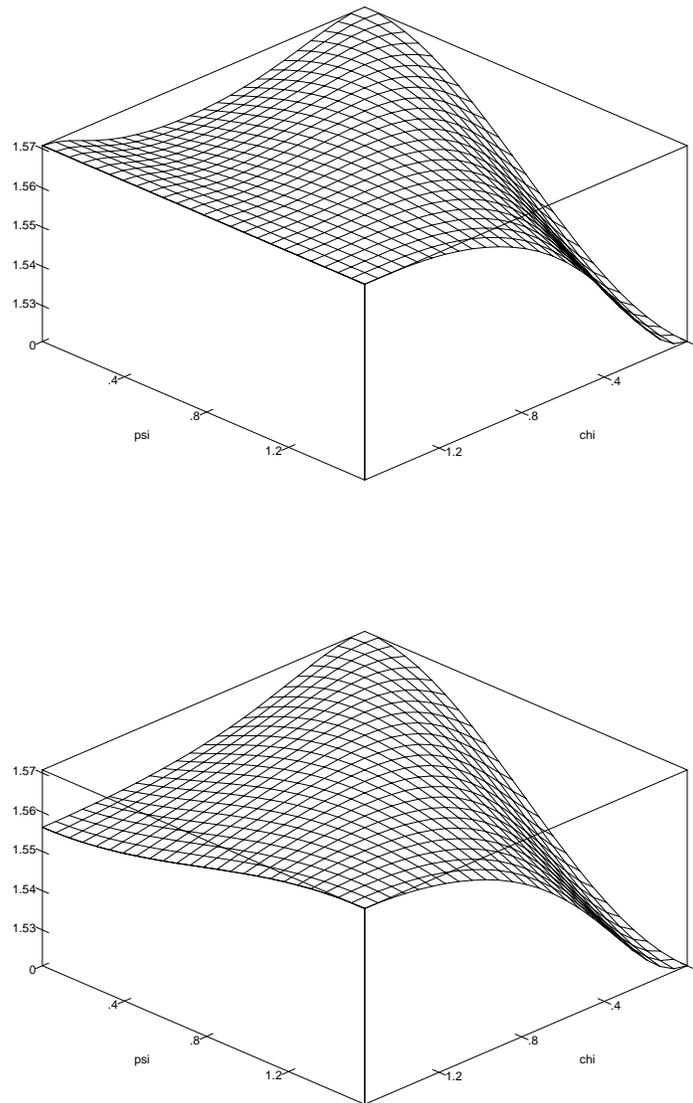


Abbildung 3.9: Der arctan der Fehlervarianz von N_x (oben) und N_y (unten) als Funktion von den Abweichungen ψ der Translation und χ der Ebenennormalen für ein Gesichtsfeld mit Größe $A = 0.1$. Der Wert $\pi/2$ entspricht einer unendlichen Fehlervarianz, d.h. keine der Komponenten N_x oder N_y kann berechnet werden. Die Fehlervarianz sinkt, wenn die Translation parallel zur Bildebene ($\psi = \pi/2$) und die Ebenennormale parallel zur optischen Achse ($\chi = 0$) wird.

bestimmen. Dies ist aber nur dann möglich, wenn die Bewegung einen translatorischen Anteil besitzt. Im diskreten Fall kann man als Schätzwert für die Lage des Punktes im Raum den Mittelpunkt der gemeinsamen Normalen der zwei Sehstrahlen \mathbf{x}_1 und \mathbf{x}_2 (siehe Abb. 3.2) wählen. Im kontinuierlichen Fall läßt sich die Tiefe durch die Gl. (2.28) ermitteln. Wie in den vorangehenden Abschnitten werden wir auch hier wegen der leichteren mathematischen Behandlung den kontinuierlichen Fall untersuchen.

Wir schreiben die Gl. (2.28) um:

$$Z = \frac{\mathbf{F}^T \mathbf{F}}{\dot{\mathbf{x}}_T^T \mathbf{F}}, \quad (3.71)$$

wobei $\mathbf{F} = \hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) = (v_x - xv_z, v_y - yv_z, 0)^T$ parallel zur Geraden ist, die durch den betrachteten Punkt und den Expansionspunkt läuft, und das Analogon der epipolaren Gerade im diskreten Fall ist. Wenn die Z -Komponente der translatorischen Geschwindigkeit nicht verschwindet, ist der Betrag des Vektors \mathbf{F} gleich dem Abstand des Punktes \mathbf{x} vom Expansionspunkt $(v_x/v_z, v_y/v_z, 0)$ multipliziert mit v_z .

Ein Fehler in der Tiefe kann durch einen Fehler in $\dot{\mathbf{x}}_T$ oder \mathbf{F} verursacht werden. Der Fehler in der translatorischen Komponente der Verschiebungsrate liegt in dem Fehler in der gemessenen Verschiebungsrate, sowie auch im Fehler der rotatorischen Komponente, der seinerseits auf den Fehler in der Ermittlung der Winkelgeschwindigkeit $\boldsymbol{\omega}$ zurückzuführen ist. Der Fehler im Vektor \mathbf{F} ist nur vom Fehler in der ermittelten translatorischen Geschwindigkeit \mathbf{v} abhängig.

Eine Fehleranalyse erster Ordnung ergibt folgende Approximation für den Fehler in der Tiefe:

$$\delta Z \approx \frac{2\mathbf{F}^T \delta \mathbf{F} \mathbf{F}^T \dot{\mathbf{x}}_T - \mathbf{F}^T \mathbf{F} (\delta \mathbf{F}^T \dot{\mathbf{x}}_T + \mathbf{F}^T \delta \dot{\mathbf{x}}_T)}{(\mathbf{F}^T \dot{\mathbf{x}}_T)^2} \quad (3.72)$$

Durch Rückersetzung von $\dot{\mathbf{x}}_T$ durch \mathbf{F}/Z erhält man folgenden Ausdruck für den relativen Fehler

$$\frac{\delta Z}{Z} \approx \frac{\mathbf{F}^T \delta \mathbf{F}}{\mathbf{F}^T \mathbf{F}} - \frac{Z \mathbf{F}^T \delta \dot{\mathbf{x}}_T}{\mathbf{F}^T \mathbf{F}}, \quad (3.73)$$

und nach Anwendung der Cauchy-Schwartzschen Ungleichung:

$$\frac{\delta Z}{Z} \leq \frac{\|\delta \mathbf{F}\| + Z \|\delta \dot{\mathbf{x}}_T\|}{\|\mathbf{F}\|}. \quad (3.74)$$

Erstens kann man feststellen, daß der relative Fehler in der Tiefe mit abnehmender Translationsgeschwindigkeit wächst, weil der Betrag des Vektors \mathbf{F} proportional zum Betrag der Translationsgeschwindigkeit ist.

Zweitens ist der relative Fehler umgekehrt proportional zum Abstand des betrachteten Punktes \mathbf{x} vom Expansionspunkt. Somit erwarten wir, daß die Tiefenschätzwerte für Punkte in der Umgebung des Expansionspunkts mit hoher Unsicherheit behaftet sind.

Drittens beobachtet man in (3.74), daß der Einfluß des Fehlers in der translatorischen Komponente der Verschiebungsrate mit der Tiefe gewichtet wird, d.h. nicht nur der absolute, sondern auch der relative Fehler wächst mit der Tiefe. Der Fehler in der translatorischen Komponente ist nicht nur vom Meßfehler der Verschiebungsrate, sondern auch vom Schätzfehler der Winkelgeschwindigkeit abhängig.

Viertens ist der Tiefenfehler abhängig vom Schätzfehler in der in einer vorangehenden Bearbeitungsstufe berechneten translatorischen Geschwindigkeit, der sich in $\|\delta \mathbf{F}\|$ widerspiegelt.

3.6 Gefährliche Flächen und Instabilität

Im Abschnitt 2.2 wurde die Problemstellung der Ermittlung der Bewegung und Struktur einer 3D-Punktconfiguration aus einem Bildpaar beschrieben. Dabei haben wir die Existenz von trivialen und nicht-trivialen Mehrdeutigkeiten erwähnt. Als triviale Mehrdeutigkeiten haben wir die Skalierungsmehrdeutigkeit sowie auch die Mehrdeutigkeit des gedrehten Paares bezeichnet. Neben diesen gibt es die nicht-trivialen Mehrdeutigkeiten, die bei Existenz von mehr als fünf Punktkorrespondenzen zu zwei oder drei Lösungen für die relative Orientierung und für die Struktur der abgebildeten Punktconfiguration führen können. Die Existenz solcher Mehrdeutigkeiten wurde schon in der Photogrammetrie [Krames 40; Wunderlich 41] und in der Bildauswertung [Longuet-Higgins 86; Faugeras & Maybank 90; Negahdaripour 90b] bewiesen.

In diesem Abschnitt werden wir einen Zusammenhang zwischen Mehrdeutigkeit und Instabilität in der Schätzung der relativen Orientierung aus einem Bildpaar aufzeigen. Auf diesen Zusammenhang hat schon [Hofmann 53] hingewiesen. Insbesondere werden wir die bis jetzt offene Frage beantworten, welche Beziehung zwischen den von [Horn 90] bezeichneten gefährlichen Flächen und den gefährlichen Flächen besteht, wie sie von [Negahdaripour 90b; Maybank 90c] analysiert wurden⁵.

Wir werden zuerst auf einem geometrischen und dann auf einem analytischen Weg die Ermittlung der Form der gefährlichen Flächen nach [Maybank 90c] beschreiben, damit wir dem Leser die Voraussetzungen zum Verständnis unserer Untersuchung geben können.

Wir nehmen an, daß der Augenpunkt in seiner ersten Lage im Ursprung \mathbf{o} liegt. Weiterhin nehmen wir an, daß zwei verschiedene Bewegungen (R, \mathbf{a}) und (S, \mathbf{b}) zu identischen Punktkorrespondenzen in der zweiten Aufnahme führen. Mit R und S bezeichnen wir die Rotationen und mit \mathbf{a} und \mathbf{b} die Translationen in Bezug auf das Koordinatensystem der ersten Aufnahme. Außer der Bewegung soll die Fläche, die von \mathbf{o} und \mathbf{a} oder \mathbf{b} aus aufgenommen worden ist, bestimmt werden. Auch dies ist nicht immer eindeutig möglich. D.h. es kann zwei verschiedene Flächen geben, die von \mathbf{o} aus gesehen gleich sind, mit der weiteren Eigenschaft, daß die erste von \mathbf{a} aus gesehen identisch mit der zweiten von \mathbf{b} aus gesehen ist.

Wir bezeichnen mit r_i und s_i die Sehstrahlen, die von den Augenpunkten \mathbf{a} und \mathbf{b} als der gleiche Punkt \mathbf{x}_{2_i} in der Bildebene wahrgenommen werden, der dem Punkt \mathbf{x}_{1_i} mit entsprechendem Sehstrahl q_i zugeordnet wird. Die Winkel zwischen den Sehstrahlen (r_i, r_j) und (q_i, q_j) sind identisch. Wir stellen diesen Zusammenhang für zwei Punktkorrespondenzen (q_i, r_i, s_i) und (q_j, r_j, s_j) in der Abb. 3.10 dar.

Weil der Winkel zwischen Sehstrahlen aus \mathbf{a} gleich dem Winkel zwischen Sehstrahlen aus \mathbf{b} ist, existiert eine orthogonale Kollineation⁶ ω , die das Strahlenbündel durch \mathbf{a} auf das Strahlenbündel durch \mathbf{b} starr transformiert. Sei im folgenden s Bild eines Sehstrahls r durch ω , also $s = \omega(r)$.

Zum Beweis der Existenz zweier Flächen mit der oben erwähnten Eigenschaft beobachtet man, daß zu jedem Sehstrahl r und $s = \omega(r)$ genau ein Sehstrahl q durch \mathbf{o} konstruiert werden kann, der s und r trifft. Er geht durch den Schnittpunkt von r

⁵Zu dieser Frage äußert sich [Negahdaripour 90b]: *Because of his [Horn 90] particular formulation of the problem, the relationship between his results and those presented here is hard to establish.*

⁶Eine invertierbare Transformation zwischen zwei projektiven Räumen heißt *Kollineation*.

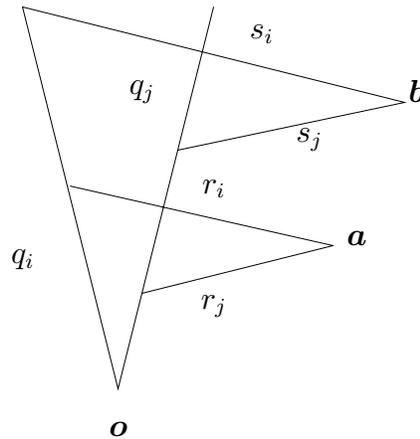


Abbildung 3.10: Geometrische Darstellung der Mehrdeutigkeit aus [Maybank 90b].

mit der Ebene, die von o und s aufgespannt wird, vgl. Abb. 3.11. Die Fläche zu den Sehstrahlen q und r wird gebildet von der Schnittgeraden l der Ebenen $\omega(\Pi)$ durch o und b mit den Ebenen Π durch a und $\omega^{-1}(b)$. Dies ist die erste gesuchte Fläche. Die zweite ergibt sich analog aus den Sehstrahlen q und s .

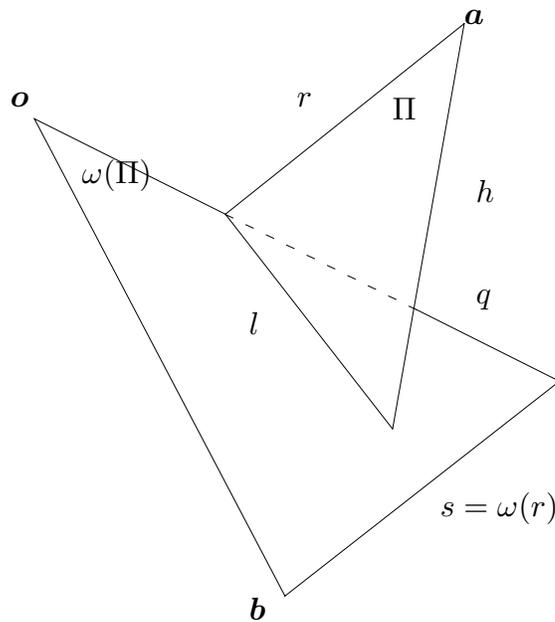


Abbildung 3.11: Die Konstruktion einer gefährlichen Fläche

Die zwei Ebenenbüschel, die durch die Drehung von Π und $\omega(\Pi)$ entstehen, werden *kongruente Ebenenbüschel* genannt. Die Photogrammeter [Krames 40; Wunderlich 41] beweisen, daß die Schnittgerade von zwei kongruenten Ebenenbüschel eine *orthogonale Regelquadrik* durchläuft. Die Achsen $\langle o, b \rangle$ und h der Ebenenbüschel sind Erzeugende der Regelquadrik, die zur selben Schar von Erzeugenden gehören. Sie werden *adjungierte Erzeugende* genannt. Die Gerade l gehört zur anderen Erzeugendenschar. Quadriken

mit zwei reellen Erzeugendenscharen sind das einschalige Hyperboloid und seine Entartungen: das hyperbolische Paraboloid, der Kegel, der Zylinder und das Ebenenpaar. Wir werden die beschriebenen Zusammenhänge anhand des einschaligen Hyperboloids erläutern, das man sich besser in der Normalform $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$ vorstellen kann. In der Abbildung 3.12 wurde eine der zwei Erzeugendenscharen eingezeichnet.

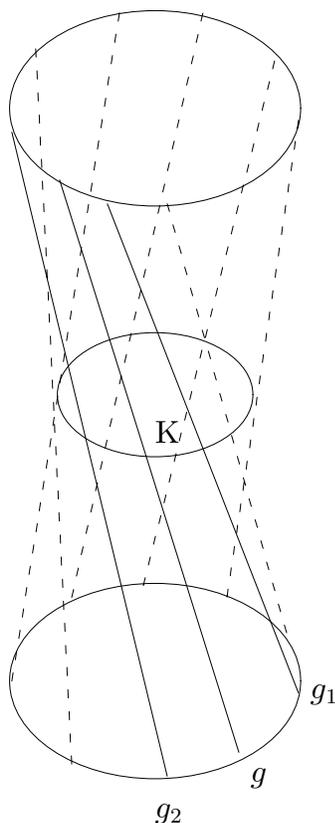


Abbildung 3.12: Das einschalige Hyperboloid. Es wird angenommen, daß die dem Leser kürzer erscheinende Hauptachse der Kehlellipse in Wirklichkeit die längere ist.

Jede Erzeugendenschar besitzt zwei *Haupterzeugende* (eine von den beiden ist g in Abb. 3.12), die durch die zwei Hauptscheitel (Endpunkte der großen Hauptachse) der Kehlellipse laufen. Zwei Erzeugende g_1 und g_2 heißen adjungiert, wenn sie symmetrisch zur großen Hauptachse der Kehlellipse sind. Das einschalige Hyperboloid heißt orthogonal, wenn jede Ebene, die senkrecht zur Haupterzeugenden steht, das Hyperboloid in einem Kreis schneidet. Ein orthogonales einschaliges Hyperboloid ist eine gefährliche Fläche, wenn die zwei Augenpunkte \mathbf{o} und \mathbf{a} auf adjungierten Erzeugenden (z.B. jeweils g_1 und g_2) liegen. Dann liegt der Augenpunkt \mathbf{b} , der der komplementären Lösung gehört, auf der Erzeugenden g_1 . Liegen beide Augenpunkte \mathbf{o} und \mathbf{a} auf den Schnittpunkten von adjungierten Erzeugenden aus verschiedenen Scharen, dann erhält man statt nur einer jetzt zwei komplementäre Lösungen und insgesamt ein Lösungstripel.

Wenn beide Augenpunkte \mathbf{o} und \mathbf{a} auf der Haupterzeugenden liegen, dann erhält man nur eine Lösung. [Hofmann 53] nennt eine solche Fläche gefährlich *im engeren Sinn*. Die gefährliche Fläche im engeren Sinn ist der Grund für die Ersetzung der

Mehrdeutigkeit durch eine Instabilität, auf die wir näher eingehen werden.

Nach der geometrischen Beschreibung werden wir nun einen analytischen Weg zu den gefährlichen Flächen im engeren Sinn beschreiben. Es seien \mathbf{x}_1 und \mathbf{x}_2 die abgebildeten Punkte in den Bildebenen der ersten und zweiten Aufnahme. Wir bezeichnen mit Z'_1 und Z'_2 die Tiefen der Punkte, die zur Fläche der komplementären Lösung gehören. In Konsistenz mit der geometrischen Beschreibung werden die Augenpunkte \mathbf{a} und \mathbf{b} in Bezug auf das Koordinatensystem der ersten Kameraposition beschrieben. Die entsprechenden Transformationen lauten

$$Z_1 \mathbf{x}_1 = R Z_2 \mathbf{x}_2 + \mathbf{a} \quad (3.75)$$

$$Z'_1 \mathbf{x}_1 = S Z'_2 \mathbf{x}_2 + \mathbf{b}. \quad (3.76)$$

Man eliminiert \mathbf{x}_2 aus (3.75-3.76), indem man erst folgendes Kreuzprodukt bildet

$$R^T (Z_1 \mathbf{x}_1 - \mathbf{a}) \times S^T (Z'_1 \mathbf{x}_1 - \mathbf{b}) = \mathbf{0}, \quad (3.77)$$

und dann sein Skalarprodukt mit jeweils $R^T \mathbf{x}_1$ und $S^T \mathbf{x}_1$ bildet. Aus den sich ergebenden Gleichungen kann man nach Z_1

$$Z_1 = \frac{(R^T \mathbf{a} \times S^T \mathbf{x}_1)^T S^T \mathbf{b}}{(R^T \mathbf{x}_1 \times S^T \mathbf{x}_1)^T S^T \mathbf{b}} \quad (3.78)$$

und nach Z'_1

$$Z'_1 = \frac{(R^T \mathbf{x}_1 \times S^T \mathbf{b})^T R^T \mathbf{a}}{(R^T \mathbf{x}_1 \times S^T \mathbf{x}_1)^T R^T \mathbf{a}} \quad (3.79)$$

lösen. Man ersetzt jetzt in den beiden letzten Gleichungen \mathbf{x}_1 und \mathbf{x}_1' jeweils durch $\mathbf{X}\mathbf{1}/Z_1$ und $\mathbf{X}\mathbf{1}'/Z'_1$ und man erhält aus (3.78) die Gleichung der mehrdeutigen Fläche, die der Bewegung (R, \mathbf{a}) entspricht,

$$(R^T \mathbf{X}\mathbf{1} \times S^T \mathbf{X}\mathbf{1})^T S^T \mathbf{b} = (R^T \mathbf{a} \times S^T \mathbf{X}\mathbf{1})^T S^T \mathbf{b} \quad (3.80)$$

und aus (3.79) die Gleichung der mehrdeutigen Fläche, die der Bewegung (S, \mathbf{b}) entspricht.

$$(R^T \mathbf{X}\mathbf{1}' \times S^T \mathbf{X}\mathbf{1}')^T R^T \mathbf{a} = (R^T \mathbf{X}\mathbf{1}' \times S^T \mathbf{b})^T R^T \mathbf{a}. \quad (3.81)$$

Wir bezeichnen mit U das Matrizenprodukt SR^T und wir benutzen nur \mathbf{X} anstatt $\mathbf{X}\mathbf{1}$ und $\mathbf{X}\mathbf{1}'$, weil $\mathbf{X}\mathbf{1}$ und $\mathbf{X}\mathbf{1}'$ freie Variablen sind. Das Paar der gefährlichen Flächen lautet dann

$$(U\mathbf{X} \times \mathbf{X})^T \mathbf{b} = (U\mathbf{a} \times \mathbf{X})^T \mathbf{b} \quad (3.82)$$

$$(U^T \mathbf{X} \times \mathbf{X})^T \mathbf{a} = (U^T \mathbf{b} \times \mathbf{X})^T \mathbf{a}. \quad (3.83)$$

Beide Flächen sind quadratisch und haben die Form

$$\mathbf{X}^T M \mathbf{X} + \mathbf{l}^T \mathbf{X} = 0.$$

Wir berücksichtigen zunächst (3.82) und formen den quadratischen Term um, indem wir die Rodriguez-Formel für die Rotation U verwenden. Wir bezeichnen mit \mathbf{n} den

Einheitsvektor in der Richtung der Rotationsachse und mit θ den Rotationswinkel.

$$\begin{aligned}
(UX \times X)^T \mathbf{b} &= \mathbf{X}^T (\mathbf{b} \times UX) \\
&= \mathbf{X}^T (\mathbf{b} \times (\mathbf{X} + \sin \theta (\mathbf{n} \times \mathbf{X}) + (1 - \cos \theta)(\mathbf{n} \times (\mathbf{n} \times \mathbf{X})))) \\
&= \mathbf{X}^T (\sin \theta (\mathbf{n} \mathbf{b}^T \mathbf{X} - \mathbf{n}^T \mathbf{b} \mathbf{X}) + (1 - \cos \theta)(\mathbf{b} \times \mathbf{n}) \mathbf{n}^T \mathbf{X}) \\
&= \mathbf{X}^T (\sin \theta \mathbf{n} \mathbf{b}^T + (1 - \cos \theta)(\mathbf{b} \times \mathbf{n}) \mathbf{n}^T - \sin \theta \mathbf{n}^T \mathbf{b} I) \mathbf{X} \\
&= \mathbf{X}^T \left(\sin \theta \frac{\mathbf{n} \mathbf{b}^T + \mathbf{b} \mathbf{n}^T}{2} + (1 - \cos \theta) \frac{(\mathbf{b} \times \mathbf{n}) \mathbf{n}^T + \mathbf{n} (\mathbf{b} \times \mathbf{n})^T}{2} \right. \\
&\quad \left. - \sin \theta \mathbf{n}^T \mathbf{b} I \right) \mathbf{X} \\
&= \mathbf{X}^T \left(\frac{\mathbf{m} \mathbf{n}^T + \mathbf{n} \mathbf{m}^T}{2} - \mathbf{m}^T \mathbf{n} I \right) \mathbf{X},
\end{aligned}$$

wobei $\mathbf{m} = \sin \theta \mathbf{b} + (1 - \cos \theta) \mathbf{b} \times \mathbf{n}$ ist.

Damit eine Quadrik $\mathbf{X}^T M \mathbf{X} + \mathbf{l}^T \mathbf{X} = 0$ orthogonal ist, ist es notwendig und hinreichend, daß die Matrix M folgende Form hat [Maybank 90c]:

$$M = \frac{1}{2}(\mathbf{m} \mathbf{n}^T + \mathbf{n} \mathbf{m}^T) - \mathbf{m}^T \mathbf{n} I. \quad (3.84)$$

Die Eigenwerte und Eigenvektoren einer solchen Matrix lassen sich leicht berechnen (siehe [Maybank 90c]):

$$\begin{aligned}
\lambda_1 &= -\frac{1}{2}(\mathbf{m}^T \mathbf{n} + \|\mathbf{m}\| \|\mathbf{n}\|) & \mathbf{u}_1 &= \frac{\frac{\mathbf{m}}{\|\mathbf{m}\|} - \frac{\mathbf{n}}{\|\mathbf{n}\|}}{\sqrt{2(1 - \frac{\mathbf{m}}{\|\mathbf{m}\|}^T \frac{\mathbf{n}}{\|\mathbf{n}\|})}} \\
\lambda_2 &= -\mathbf{m}^T \mathbf{n} & \mathbf{u}_2 &= \frac{\mathbf{m} \times \mathbf{n}}{\|\mathbf{m} \times \mathbf{n}\|} \\
\lambda_3 &= \frac{1}{2}(-\mathbf{m}^T \mathbf{n} + \|\mathbf{m}\| \|\mathbf{n}\|) & \mathbf{u}_3 &= \frac{\frac{\mathbf{m}}{\|\mathbf{m}\|} + \frac{\mathbf{n}}{\|\mathbf{n}\|}}{\sqrt{2(1 + \frac{\mathbf{m}}{\|\mathbf{m}\|}^T \frac{\mathbf{n}}{\|\mathbf{n}\|})}}.
\end{aligned} \quad (3.85)$$

Man stellt fest, daß der mittlere Eigenwert die Summe der übrigen zwei ist. [Maybank 90c] hat ebenfalls bewiesen, daß dies eine notwendige und hinreichende Bedingung für die Orthogonalität der Quadrik ist.

Das Symmetriezentrum der Quadrik, falls es existiert, liegt im Punkt [Korn & Korn 68]

$$\mathbf{c} = -\frac{1}{2} M^{-1} \mathbf{l}. \quad (3.86)$$

Führt man die Transformation $\mathbf{X}^c = \mathbf{X} - \mathbf{c}$ durch, erhält man die Gleichung der Quadrik, die symmetrisch um den Ursprung ist:

$$\mathbf{X}^{cT} M \mathbf{X}^c = \frac{1}{4} \mathbf{l}^T M^{-1} \mathbf{l} = c. \quad (3.87)$$

Die Längen der Hauptachsen der Kehlellipse sind $(c/\lambda_2)^{1/2}$ und $(c/\lambda_3)^{1/2}$. Die große Hauptachse der Kehlellipse hat die Richtung des Eigenvektors \mathbf{u}_2 , weil $(c/\lambda_2)^{1/2} >$

$(c/\lambda_3)^{1/2}$ ist. Die Hauptscheitel (Endpunkte der großen Hauptachse) liegen an den Punkten

$$\mathbf{c} \pm \sqrt{\frac{c}{\lambda_2}} \mathbf{u}_2$$

und es läßt sich leicht ausrechnen, daß die Geraden durch diese Hauptscheitel mit Richtungen \mathbf{m} und \mathbf{n} der Quadrik gehören. Sie sind die Haupterzeugenden jeder Erzeugendenschar.

Um die gefährliche Fläche im engeren Sinn analytisch herzuleiten, nehmen wir an, daß sich die zwei Lösungen (R, \mathbf{a}) und (S, \mathbf{b}) nahegerückt sind, so daß man ihren Unterschied für infinitesimal halten kann:

$$U = I + [\delta\boldsymbol{\omega}]_{\times} \quad \mathbf{b} = \mathbf{a} + \delta\mathbf{a}. \quad (3.88)$$

Wir ersetzen diesen Ausdruck in der ersten mehrdeutigen Fläche (3.82):

$$((\mathbf{X} + \delta\boldsymbol{\omega} \times \mathbf{X}) \times \mathbf{X})^T (\mathbf{a} + \delta\mathbf{a}) = ((\mathbf{a} + \delta\boldsymbol{\omega} \times \mathbf{a}) \times \mathbf{X})^T (\mathbf{a} + \delta\mathbf{a}).$$

Wir setzen die Terme zweiter Ordnung in $(\delta\boldsymbol{\omega}, \delta\mathbf{a})$ zu Null und erhalten

$$(\delta\boldsymbol{\omega} \times \mathbf{X})^T (\mathbf{X} \times \mathbf{a}) + (\mathbf{a} \times (\mathbf{a} \times \delta\boldsymbol{\omega} + \delta\mathbf{a}))^T \mathbf{X} = 0. \quad (3.89)$$

Man erhält dieselbe Quadrik, wenn man von der zweiten mehrdeutigen Fläche ausgeht und \mathbf{a} durch $\mathbf{b} - \delta\mathbf{b}$ ersetzt. Die Quadrik (3.89) kann auch in der Form $\mathbf{X}^T M \mathbf{X} + \mathbf{l}^T \mathbf{X} = 0$ mit

$$M = \frac{1}{2} (\mathbf{a} \delta\boldsymbol{\omega}^T + \delta\boldsymbol{\omega} \mathbf{a}^T) - \mathbf{a}^T \delta\boldsymbol{\omega} I \quad (3.90)$$

$$\mathbf{l} = \mathbf{a} \times (\mathbf{a} \times \delta\boldsymbol{\omega} + \delta\mathbf{a}) \quad (3.91)$$

geschrieben werden. Aus der Form von M kann man direkt erkennen, daß die Matrix trotz des Verschwindens von Termen orthogonal geblieben ist.

Die notwendige und hinreichende Bedingung, damit eine orthogonale Quadrik eine gefährliche Fläche im engeren Sinn darstellt, ist, daß beide optischen Zentren \mathbf{o} und \mathbf{a} auf die Haupterzeugende fallen sollen. Ist die Gerade $\lambda\mathbf{m}$ durch den Ursprung eine Erzeugende, so ist diese Gerade die Haupterzeugende, weil es nur eine Erzeugende mit dieser Richtung gibt. Damit die Gerade $\lambda\mathbf{m}$ eine Erzeugende ist, muß $\lambda^2 \mathbf{m}^T M \mathbf{m} + \lambda \mathbf{l}^T \mathbf{m} = 0$ für jedes λ sein. Aus der Form von M gilt immer $\mathbf{m}^T M \mathbf{m} = 0$. So bleibt als Bedingung $\mathbf{l}^T \mathbf{m} = 0$. Damit auch \mathbf{a} auf der Haupterzeugenden liegt, muß $\mathbf{a} \times \mathbf{m} = 0$ gelten.

In der so erhaltenen Gleichung (3.89) der gefährlichen Fläche im engeren Sinn ist $\mathbf{m} = \mathbf{a}$ und $\mathbf{n} = \delta\boldsymbol{\omega}$, folglich kann man leicht nachweisen, daß $\mathbf{l}^T \mathbf{m} = 0$ ist.

Wir gehen jetzt aus von der epipolaren Bedingung und ermitteln die nach [Horn 90] bezeichneten gefährlichen Flächen, um nachzuweisen, daß sie äquivalent zu den gefährlichen Flächen im engeren Sinn nach Gl. (3.89) sind.

Aus (3.75) folgt die epipolare Bedingung

$$\mathbf{x}_1^T (\mathbf{a} \times R\mathbf{x}_2) = 0. \quad (3.92)$$

Wir führen eine infinitesimale Störung $(\delta\boldsymbol{\omega}, \delta\mathbf{a})$ in die Bewegungsparameter ein und untersuchen, wann die linke Seite von (3.92) bis auf die Terme erster Ordnung konstant

bleibt. Dann wird sich das Quadrat der linken Seite von (3.92) um Terme vierter Ordnung in der Nähe der Lösung ändern. Eine zu minimierende Funktion, die aus der Summe solcher Quadrate besteht, wird in der Nähe der Lösung sehr flach sein. Entsprechend wird die Hessesche Matrix der Funktion schlecht konditioniert sein, so daß jedes Minimierungsverfahren ein instabiles Verhalten zeigen wird.

Wegen der Skalierungsmehrdeutigkeit wird der Translationsvektor \mathbf{a} wie üblich als ein Einheitsvektor angenommen. Daher muß die infinitesimale Korrektur $\delta\mathbf{a}$ senkrecht zu \mathbf{a} stehen: $\delta\mathbf{a}^T\mathbf{a} = 0$. Wir führen die Störung $(\delta\boldsymbol{\omega}, \delta\mathbf{a})$ in (3.92) ein und erhalten

$$\begin{aligned} & \mathbf{x}_1^T ((\mathbf{a} + \delta\mathbf{a}) \times (R\mathbf{x}_2 + \delta\boldsymbol{\omega} \times R\mathbf{x}_2)) = \\ & \mathbf{x}_1^T(\mathbf{a} \times R\mathbf{x}_2) + \mathbf{x}_1^T(\mathbf{a} \times (\delta\boldsymbol{\omega} \times R\mathbf{x}_2)) + \mathbf{x}_1^T(\delta\mathbf{a} \times R\mathbf{x}_2) + \mathcal{O}(\|\delta\boldsymbol{\omega}\|^2, \|\delta\mathbf{a}\|^2). \end{aligned} \quad (3.93)$$

Wir fordern, daß die Terme erster Ordnung verschwinden

$$\mathbf{x}_1^T(\mathbf{a} \times (\delta\boldsymbol{\omega} \times R\mathbf{x}_2)) + \mathbf{x}_1^T(\delta\mathbf{a} \times R\mathbf{x}_2) = 0$$

und multiplizieren mit Z_1 und Z_2 :

$$\mathbf{X}\mathbf{1}^T(\mathbf{a} \times (\delta\boldsymbol{\omega} \times R\mathbf{X}\mathbf{2})) + \mathbf{X}\mathbf{1}^T(\delta\mathbf{a} \times R\mathbf{X}\mathbf{2}) = 0.$$

Dann ersetzen wir $R\mathbf{X}\mathbf{2}$ durch $\mathbf{X}\mathbf{1} - \mathbf{a}$

$$\mathbf{X}\mathbf{1}^T(\mathbf{a} \times (\delta\boldsymbol{\omega} \times (\mathbf{X}\mathbf{1} - \mathbf{a}))) + \mathbf{X}\mathbf{1}^T(\delta\mathbf{a} \times (\mathbf{X}\mathbf{1} - \mathbf{a})) = 0.$$

Wir schreiben \mathbf{X} statt $\mathbf{X}\mathbf{1}$ und formen um:

$$(\mathbf{X} \times \mathbf{a})^T(\delta\boldsymbol{\omega} \times \mathbf{X}) + \mathbf{X}^T(\mathbf{a} \times \delta\mathbf{a} + \mathbf{a} \times (\mathbf{a} \times \delta\boldsymbol{\omega})) = 0 \quad (3.94)$$

Die letzte Gleichung ist identisch zur Gleichung (3.89) der gefährlichen Fläche im engeren Sinn und damit ist die gesuchte Beziehung bewiesen worden.

[Hofmann 53] hat auch eine Gleichung für die gefährlichen Flächen im engeren Sinn angegeben, jedoch unter folgenden Vereinfachungsannahmen, die in der Praxis der Photogrammetrie zutreffen. Die Photogrammeter nehmen bei der Ermittlung der relativen Orientierung an, daß die zwei Kamerakoordinatensysteme vor der Korrektur identische Orientierung besitzen, $R = I$, und relativ zueinander nur in Richtung der x -Achse $\mathbf{a} = (a_x, 0, 0)^T$ verschoben sind. Unter diesen Annahmen ist die epipolare Bedingung äquivalent zum Verschwinden der Differenz der y -Koordinaten der Bildpunkte. Diese Differenz wird auch y -Parallaxe genannt. Man versucht, die y -Parallaxen durch Korrekturen $(\delta\omega_x, \delta\omega_y, \delta\omega_z)^T$ und $(0, \delta a_y, \delta a_z)^T$ zu minimieren. Die Translationskomponente in x -Richtung kann wegen der Skalierungsmehrdeutigkeit nicht korrigiert werden ($\delta\mathbf{a}^T\mathbf{a} = 0$). Nach dem Einsetzen dieser Korrekturen in Gl. 3.89 erhält man

$$a_x(Z(X\delta\omega_z - Z\delta\omega_x) - Y(Y\delta\omega_x - X\delta\omega_y) - Y(\delta a_z + a_x\delta\omega_y) + Z(\delta a_y - a_x\delta\omega_z)) = 0.$$

Nach Division durch a_x und Umformung erhält man die Gleichung einer quadratischen Fläche

$$-(Y^2 + Z^2)\delta\omega_x + (XY - Ya_x)\delta\omega_y + (XZ - Za_x)\delta\omega_z + Z\delta a_y - Y\delta a_z = 0,$$

die mit der sich in [Hofmann 53] findenden Form übereinstimmt.

Im weiteren werden wir die möglichen Entartungen des einschaligen Hyperboloids untersuchen. Aus dem einschaligen Hyperboloid erhält man einen elliptischen Kegel (mit der Normalform $x^2/a^2 + y^2/b^2 - z^2/c^2 = 0$), wenn die rechte Seite von (3.87) verschwindet, d.h. $\mathbf{l}^T M^{-1} \mathbf{l} = 0$. Aus den Eigenwerten und -vektoren einer symmetrischen Matrix kann man die Inverse der Matrix bilden:

$$M^{-1} = \frac{1}{\lambda_1} \mathbf{u}_1 \mathbf{u}_1^T + \frac{1}{\lambda_2} \mathbf{u}_2 \mathbf{u}_2^T + \frac{1}{\lambda_3} \mathbf{u}_3 \mathbf{u}_3^T.$$

Daraus folgt

$$\mathbf{l}^T M^{-1} \mathbf{l} = \frac{1}{\lambda_1} (\mathbf{l}^T \mathbf{u}_1)^2 + \frac{1}{\lambda_2} (\mathbf{l}^T \mathbf{u}_2)^2 + \frac{1}{\lambda_3} (\mathbf{l}^T \mathbf{u}_3)^2.$$

Aus dem Eigensystem (3.85) von M und aus der Tatsache, daß die Haupterzeugende durch den Hauptscheitel der Kehlellipse läuft ($\mathbf{l}^T \mathbf{m} = 0$) folgt

$$\begin{aligned} \frac{1}{\lambda_1} (\mathbf{l}^T \mathbf{u}_1)^2 + \frac{1}{\lambda_3} (\mathbf{l}^T \mathbf{u}_3)^2 &= \frac{(\mathbf{l}^T \frac{\mathbf{n}}{\|\mathbf{n}\|})^2}{\lambda_1 (2(1 - \frac{\mathbf{m}^T \mathbf{n}}{\|\mathbf{m}\| \|\mathbf{n}\|}))} + \frac{(\mathbf{l}^T \frac{\mathbf{n}}{\|\mathbf{n}\|})^2}{\lambda_3 (2(1 + \frac{\mathbf{m}^T \mathbf{n}}{\|\mathbf{m}\| \|\mathbf{n}\|}))} \\ &= \frac{(\mathbf{l}^T \frac{\mathbf{n}}{\|\mathbf{n}\|})^2}{2} \left(\frac{1}{\lambda_1 (1 - \frac{\mathbf{m}^T \mathbf{n}}{\|\mathbf{m}\| \|\mathbf{n}\|})} + \frac{1}{\lambda_3 (1 + \frac{\mathbf{m}^T \mathbf{n}}{\|\mathbf{m}\| \|\mathbf{n}\|})} \right) = 0 \end{aligned}$$

Somit erhält man einen Kegel, wenn

$$\mathbf{l}^T \mathbf{u}_2 = \mathbf{l}^T (\mathbf{m} \times \mathbf{n}) = 0. \quad (3.95)$$

Man ersetzt \mathbf{m} und \mathbf{n} durch die für die gefährliche Fläche im engeren Sinn geeigneten Ausdrücke entsprechend Gl. (3.90) und erhält

$$(\mathbf{a} \times \delta \mathbf{a})^T (\mathbf{a} \times \delta \boldsymbol{\omega}) = 0. \quad (3.96)$$

Einen elliptischen Zylinder (mit der Normalform $x^2/a^2 + y^2/b^2 = 1$) erhält man, wenn der negative Eigenwert $-1/2(\mathbf{m}^T \mathbf{n} + \|\mathbf{m}\| \|\mathbf{n}\|)$ Null wird, d.h. wenn \mathbf{a} parallel zu $\delta \boldsymbol{\omega}$ wird. In diesem Fall fallen die beiden übrigen Eigenwerte zusammen, deshalb ist der elliptische Zylinder ein Kreiszyylinder. Die optischen Zentren liegen auf einer der Mantellinien des Zylinders, d.h. eine Kamera bewegt sich entlang einer Mantellinie. Diese Situation ist realitätsnahe, wenn man sich eine Kamera in einem Flugzeug vorstellt, die Bilder von einem zylinderförmigen Tal aufnimmt. Die Korrektur $\delta \mathbf{a}$ steht senkrecht zur Korrektur \mathbf{a} und deswegen auch zu $\delta \boldsymbol{\omega}$. Korrekturen erster Ordnung in der Translation und der Rotation, die senkrecht zueinander stehen, verursachen Änderungen zweiter Ordnung im Fehlerterm. Im Fall eines kleinen Gesichtsfelds kann eine Punktconfiguration ohne hohe Tiefenunterschiede von einem kleinen Teil einer Quadrik approximiert werden. Wir vermuten, daß die Kombination der beiden letzten Tatsachen die Schwierigkeit bei einer Unterscheidung zwischen einer Translation parallel zur Bildebene und einer Rotation um eine Achse parallel zur Bildebene erklärt.

Falls der mittlere Eigenwert $\lambda_2 = -\mathbf{m}^T \mathbf{n}$ verschwindet, erhält man ein hyperbolisches Paraboloid (mit der Normalform $z = x^2/a^2 - y^2/b^2$). Dies geschieht, wenn die Korrektur $\delta \boldsymbol{\omega}$ senkrecht zur Translationsrichtung \mathbf{a} steht.

Als letzte Entartung verbleibt das Ebenenpaar (mit der Normalform $x^2/a^2 = z^2/c^2$), wenn der mittlere Eigenwert und die rechte Seite von Gl. (3.87) verschwinden. Es muß gelten

$$\mathbf{m}^T \mathbf{n} = 0 \quad \text{und aus (3.95)} \quad \mathbf{l}^T (\mathbf{m} \times \mathbf{n}) = 0.$$

Daraus folgt $\delta\boldsymbol{\omega}^T \mathbf{a} = 0$ und $\mathbf{l}^T (\mathbf{a} \times \delta\boldsymbol{\omega}) = \delta\boldsymbol{\omega}^T \delta\mathbf{a} = 0$. Setzt man dies in die Gl. (3.89) ein, so erhält man

$$(\mathbf{X}^T \mathbf{a})(\mathbf{X}^T \delta\boldsymbol{\omega}) + \mathbf{X}^T (-\delta\boldsymbol{\omega} + \mathbf{a} \times \delta\mathbf{a}) = 0.$$

$\delta\boldsymbol{\omega}$ muß senkrecht zu \mathbf{a} und $\delta\mathbf{a}$ stehen und man kann schreiben $\delta\boldsymbol{\omega} = \kappa(\mathbf{a} \times \delta\mathbf{a})$. Daraus folgt

$$\mathbf{X}^T (\mathbf{a} \times \delta\mathbf{a}) (\kappa \mathbf{X}^T \mathbf{a} - \kappa + 1) = 0,$$

was die Gleichung von zwei Ebenen ist:

$$\begin{aligned} \mathbf{X}^T (\mathbf{a} \times \delta\mathbf{a}) &= 0 \\ \kappa \mathbf{X}^T \mathbf{a} &= \kappa - 1 \end{aligned}$$

Unter diesen zwei Ebenen ist für uns nur die zweite von Interesse, weil sie nicht durch den Ursprung läuft. Es ist bemerkenswert, daß die Translationsrichtung parallel zur Ebenennormalen ist. Das ist genau der Fall bei der allgemeinen Bewegung einer Ebene, wobei die Bewegungsschätzung mit einer Instabilität behaftet ist, wie schon im Abschnitt 3.4 gezeigt wurde.

4 Bewegungsschätzung und Rekonstruktion von Objekten aus längeren Bildfolgen

4.1 Überblick über Ansätze zur 3D-Bewegungsschätzung aus längeren Bildfolgen

In diesem Abschnitt werden wir Ansätze zur 3D-Bewegungsbestimmung aus längeren Bildfolgen beschreiben. Dazu haben wir zuerst versucht, die Ansätze zu klassifizieren, damit geklärt wird, welche Situation jeder Ansatz behandelt. Diese Klassifikation ist von Bedeutung, weil die meisten Ansätze unter ähnlichen Bezeichnungen in der Literatur erscheinen. In der Tabelle 4.1 werden alle uns bekannten Ansätze aufgezeichnet. Dabei werden die Art der verwendeten Bildbereichshinweise, die Annahmen über die Geometrie der Szene sowie die Bewegung und die verwendete Schätzmethode erwähnt. Anhand der zwei ersten Punkte haben wir die Ansätze in fünf Gruppen unterteilt, die in der Tabelle mittels einer horizontalen Linie getrennt werden. Alle Ansätze nehmen eine starre Bewegung an. Zur ersten Gruppe gehören Ansätze, die sich auf aus monokularen Bildfolgen ermittelten Bildbereichshinweisen stützen und beliebige oder glatte Bewegung unterstellen. Zur Glattheit der Bewegung wird meistens eine konstante Translationsbeschleunigung oder die Bewegung auf einem Kreisbogen unterstellt. Die zweite Gruppe bildet sich aus Ansätzen, die Folgen von Stereoaufnahmen auswerten und dadurch über Messungen im Raum verfügen, so daß die Struktur nur verbessert und nicht geschätzt werden muß. Die dritte Gruppe verwendet Information über die Bewegung, die ein anderer Sensor liefert oder die schon bekannt ist, weil sie vom System eingestellt wurde – z.B. bei der Steuerung eines Roboterarms. Dabei bleibt als Aufgabe nur die Ermittlung der Struktur übrig. Die vierte Gruppe bildet sich aus Ansätzen, bei denen ein geometrisches Modell der abgebildeten Szenenkomponente zur Verfügung steht. Beim Ansatz von [Dickmanns & Graefe 88a, 88b] handelt es sich um eine parametrisierte Form der Straßenspur, die vom Auto verfolgt werden soll. Dieser Ansatz, mit dem eine hohe Leistung (automatische Führung eines Fahrzeugs mit 96km/h auf einer deutschen Autobahn) erreicht worden ist, basiert auf der Regelung der Bewegung anhand der Schätzung mit Hilfe eines erweiterten Kalman-Filters. Dasselbe gilt für den Ansatz [Wünsche 88] aus derselben Forschungsgruppe. Der Steuerungsvektor der Bewegung ist somit bekannt, deshalb erfüllen diese beiden Ansätze auch das Klassifikationskriterium der dritten Gruppe.

Ansatz	Eingaben	Annahmen	Schätzmethode
[Dreschler & Nagel 82]	Punktmerkmale	beliebige Bewegung in einer Ebene	globale Ausgleichsrechnung
[Spetsakis & Aloimonos 91]	—”—	kein Bewegungsmodell	globale Schätzung
[Cui <i>et al.</i> 90]	—”—	—”—	rekursive Schätzung
[Oliensis & Thomas 91]	—”—	—”—	—”—
[Harris & Pike 88; Charnley & Blissett 89]	—”—	—”—	rekursive Schätzung für Struktur
[Broida <i>et al.</i> 90; Broida & Chellappa 91]	—”—	konstante Objektgeschwindigkeiten	rekursive Schätzung - IEKF
[Kumar <i>et al.</i> 89]	—”—	Translationsbeschleunigung und konstante Rotation	semirekursive Schätzung
[Iu & Wohn 90]	—”—	—”—	rekursive Schätzung - IEKF
[Shariat & Price 90]	—”—	konstante Rotation und Translation	Schätzung über fünf Aufnahmen
[Sawhney <i>et al.</i> 90]	—”—	reine Rotation	globale Schätzung
[Tseng & Sood 89]	—”—	—”—	Systemidentifikationsansatz
[Heel 90a]	Grauwertbilder	beliebige Bewegung	rekursive Schätzung der Tiefenkarte
[Heel 90b; Heel 91]	Verschiebungsraten	—”—	rekursive Schätzung einer Tiefenkarte und Flächenrekonstruktion
[Ando 91]	Verschiebungsraten	beliebige Bewegung, getrennte Auswertung für jedes Bildmerkmal	—”—
[Murray & Pickup 91]	<i>normale</i> Verschiebungsraten	konstante Geschwindigkeiten und Punkte in einer Ebene	EKF-Schätzung
[Sull & Ahuja 91a]	Punktkorrespondenzen	konstante Geschwindigkeiten und Punkte in einer Ebene	globale Schätzung
[Ayache & Faugeras 88]	Punkte oder Geraden aus Stereoaufnahmen	beliebige Bewegung	EKF für Struktur
[Matthies & Shafer 87]	—”—	—”—	—”—
[Lee & Kay 90]	Punkte aus Stereoaufnahmen	konstante Geschwindigkeiten	rekursive Schätzung durch Fehlerfortpflanzung
[Weng <i>et al.</i> 87]	—”—	Präzessionsbewegung	globaler Ansatz
[Young & Chellappa 90a]	—”—	—”—	IEKF-Schätzung, Eindeutigkeitsbeweise
[Pollard <i>et al.</i> 90]	Geradensegmente aus Stereoaufnahmen	konstante Geschwindigkeiten	EKF für Struktur
[Zhang & Faugeras 90]	—”—	Translationsbeschleunigung und konstante Rotation	getrennte EKF für Struktur und Bewegung
[Cui <i>et al.</i> 91]	Punktmerkmale aus Stereoaufnahmen	beliebige Bewegung	rekursive Schätzung der Struktur
[Matthies <i>et al.</i> 89]	Verschiebungsraten	bekannte Bewegung	rek. Schätzung der Tiefenkarte
[Crowley & Stelmaszyk 90]	Geradensegmente	bekannte Bewegung	KF-Schätzung
[Chang & Aggarwal 91]	Geraden	bekannte Bewegung	Kombinierte Korrespondenzermittlung und Schätzung (JPDAF)
[Huber & Graefe 91]	Verschiebungsraten	bekannte Bewegung	EKF-Schätzung

Ansatz	Eingaben	Annahmen	Schätzmethode
[Harris & Stennet 90]	Punktzuordnungen	bekanntes Objekt	rekursive Schätzung - EKF
[Wu <i>et al.</i> 88]	—”—	bekanntes Objekt und konstante Geschwindigkeiten	—”—
[Wünsche 88]	vertikale Kanten	bekannte Andockungskomponente und Bewegungsmodell eines Luftkissenfahrzeugs	Regelung mit EKF-Schätzung
[Dickmanns & Graefe 88a, 88b]	Kurve bekannter Form (Straßenspur) und Anfangswerte für Bewegung	Bewegungsmodell des Autos und Modell der Fahrbahn	Regelung mit EKF-Schätzung
[Chandrashekar & Chellappa 91]	Punktmerkmale	Bekannte Lage der Punktmerkmale im Weltkoordinatensystem, konstante Geschwindigkeiten	IEKF-Schätzung
[Koller <i>et al.</i> 92]	Geradensegmente	Bekannte Struktur der Objekte, Modell der Fahrzeugbewegung auf einer Ebene	IEKF-Schätzung gleichzeitig mit Korrespondenzerstellung
[Schick & Dickmanns 91]	Punktmerkmale	Modell der Autobewegung auf einer Klothoiden-Kurve	EKF-Schätzung
[Ullman 84; Hildreth & Grzywacz 86]	Punktzuordnungen	orthographische Projektion und Annahme maximaler Starrheit	inkrementelle Strukturkorrektur
[Debrunner & Ahuja 90]	—”—	orthographische Projektion und konstante Geschwindigkeiten	globale Schätzung
[Tomasi & Kanade 90]	—”—	beliebige Bewegung	globale Schätzung mittels SVD

Tabelle 4.1: Ansätze zur Bewegungsschätzung aus längeren Bildfolgen

Die fünfte Gruppe verwendet das Modell der orthographischen statt der perspektivischen Projektion und kann nur da Anwendung finden, wo ein großer Abstand der abgebildeten Szenenkomponente von der Kamera eine solche Approximation erlaubt.

Mehrere Beiträge – hauptsächlich zur autonomen Navigation – existieren, die speziellere Problemstellungen als die oben erwähnten beinhalten und die wir nicht aufgelistet haben. Dazu zählen die Ansätze zur Straßenverfolgung aus dem ALV¹-Projekt in den Vereinigten Staaten sowie Ansätze, die zur Navigation hauptsächlich Abstandsmesser verwenden.

Aus den in der Tab. 4.1 aufgelisteten Ansätze werden wir gründlicher auf die Ansätze der ersten Gruppe eingehen, weil ihre allgemeine Problemstellung dem Gegenstand unserer Untersuchungen näher liegt. Eine Beschreibung eines Teils der übrigen aufgelisteten Ansätze kann der Leser in [Aggarwal & Nandhakumar 88] finden.

Zur Beschreibung der Ansätze werden wir versuchen, Symbole zu sparen und möglichst einheitliche Bezeichnungen zu verwenden. Wir werden den Index k für die Bezeichnung einer Größe zum Zeitpunkt t_k und den Index i für die Bezeichnung des i -ten verfolgten Punktes benutzen. Daher werden die gemessenen Bildkoordinaten des i -ten Punktes zum k -ten Zeitpunkt mit $\mathbf{x}_{k/i}$ bezeichnet.

Der zeitlich erste Ansatz ist der Ansatz von [Dreschler & Nagel 82], die auf das Verfahren von [Bonde & Nagel 79] aufbauen. Die Bewegung und die Struktur eines Autos auf einer Straße wird berechnet. Die Lage der Straße in Bezug auf die Kamera wird als bekannt vorausgesetzt und die Bewegung wird auf drei Freiheitsgrade eingeschränkt.

¹Autonomous Land Vehicle

Es seien $\mathbf{X}_{ob,i}$ der Positionsvektor jedes Punktes beschrieben im Objektkoordinatensystem. (R_k, \mathbf{T}_k) bezeichnen die Transformation vom Kamerakoordinatensystem zum Objektkoordinatensystem zum k -ten Zeitpunkt. Dann gilt

$$\mathbf{X}_{ob,i} = R_k Z_{k/i} \mathbf{x}_{k/i} + \mathbf{T}_k, \quad (4.1)$$

wobei $Z_{k/i}$ die Tiefe des i -ten Punktes ist. Unbekannt sind die x - und y -Komponenten der Translation, einer (θ_k) der drei Rotationswinkel und die Strukturparameter $\mathbf{X}_{ob,i}$ für jeden Objektpunkt. Es wird keine Glattheit der Bewegung angenommen. Aus der obigen Gleichung werden zuerst die Tiefe $Z_{k/i}$ und dann die x - und y -Komponenten von \mathbf{T}_k durch Festlegung eines Punktes auf dem Objekt als Ursprung des Objektkoordinatensystems eliminiert. Nach weiteren Eliminierungen der x - und y -Komponenten der Punktkoordinaten $\mathbf{X}_{ob,i}$ gelangt man zu einer nichtlinearen Minimierung nach den z -Koordinaten (den Höhen) der Objektpunkte und den Winkeln θ_k . Am Ende wird aus der berechneten Struktur eine konvexe Hülle des Objekts rekonstruiert.

[Spetsakis & Aloimonos 91] unterstellen kein Bewegungsmodell und führen ebenfalls eine globale Schätzung durch, mit der unmittelbaren Folge, daß sie eine enorme Anzahl von Unbekannten erhalten. Für n Bilder werden als Unbekannte alle Bewegungen zwischen je zwei (nicht nur aufeinander folgenden) Bildern aufgefaßt. Die Anzahl der Kombinationen aller Bildpaare ist $n(n-1)/2$. Daher sind $6n(n-1)/2 - 1$ Unbekannte zu bestimmen – man subtrahiert eine Unbekannte wegen der Unbestimmbarkeit des Skalierungsfaktors. Es seien jeweils $\mathbf{x}_{j/i}$ und $\mathbf{x}_{k/i}$ die Sehstrahlen des i -ten Punktes im jeweils j -ten und k -ten Bild. Die relative Orientierung zwischen diesen Bildern sei $(R_{j,k}, \mathbf{T}_{j,k})$. Für jeden Punkt wird die Summe der Längen der gemeinsamen Normalen (siehe Abschnitt 3.2) zwischen den Sehstrahlen von jeweils zwei Bildern j und k berücksichtigt. Dann wird deren Summe über alle m Punkte nach den Bewegungsparametern minimiert:

$$\sum_{i=1}^m \sum_{j,k} \left(\frac{\mathbf{T}_{j,k}^T (R_{j,k} \mathbf{x}_{j/i} \times \mathbf{x}_{k/i})}{\| (R_{j,k} \mathbf{x}_{j/i} \times \mathbf{x}_{k/i}) \|} \right)^2 \implies \min. \quad (4.2)$$

Durch eine Gewichtung mit dem Sinus des Winkels zwischen je zwei Sichtstrahlen, der umgekehrt proportional zu dem Abstand der Kamera vom 3D-Punkt ist, erfolgt eine Vereinfachung der Nenner in (4.2). Die Summe der verbleibenden Zähler läßt sich so umformulieren, daß man einen Rayleighschen Quotienten analog zu dem Bildpaar erhält (siehe die Beschreibung des Ansatzes derselben Autoren am Ende vom Abschnitt 3.3). Das zweischrittige Iterationsverfahren enthält dann die Berechnung des kleinsten Eigenwerts einer symmetrischen Matrix der Dimension $3n(n-1)/2$ und des entsprechenden Eigenvektors und die darauf folgende Korrektur der $3n(n-1)/2$ Rotationsparameter.

Der Ansatz von [Cui *et al.* 90] basiert auf der Berechnung der Bewegung von je zwei aufeinander folgenden Aufnahmen unter Berücksichtigung der Meßunsicherheiten und Fortpflanzung der Fehlerkovarianzen der geschätzten Größen. Man bezeichne mit $\mathbf{m}_{k-1,k}$ den sechsdimensionalen Vektor der Bewegungsparameter, der die Bewegung zwischen zwei aufeinander folgenden Aufnahmen beschreibt. Das Ablaufschema ergibt sich aus folgenden Schritten:

1. Initialisierung der Bewegung und der Struktur aus den ersten zwei Aufnahmen nach einem Bildpaaralgorithmus. Daher ermittelt man $\mathbf{m}_{0,1}$ und $\mathbf{X}_{0/i}$. Der Skalierungsfaktor wird einmalig durch die Bedingung $\|\mathbf{T}_{0,1}\| = 1$ festgelegt.

2. Zum Zeitpunkt t_k sind die Strukturschätzung $\mathbf{X}_{k-1/i}$ und die Messung $\mathbf{x}_{k/i}$ für jeden Punkt i vorhanden.

2.1 $\Sigma_{\mathbf{X}}$ sei die Fehlerkovarianz der Größe \mathbf{X} . Die Summe beider Fehler im Szenenbereich und im Bildbereich bildet folgendes Fehlermaß

$$f(\mathbf{m}_{k-1,k}, \mathbf{X}_{k/i}) = \sum_{i=1}^m \left\{ (\mathbf{X}_{k/i} - \mathbf{X}'_{k/i})^T \Sigma_{\mathbf{X}'_{k/i}}^{-1} (\mathbf{X}_{k/i} - \mathbf{X}'_{k/i}) + (\mathbf{x}_{k/i} - p(\mathbf{X}_{k/i}))^T \Sigma_{\mathbf{x}_{k/i}}^{-1} (\mathbf{x}_{k/i} - p(\mathbf{X}_{k/i})) \right\}, \quad (4.3)$$

$$\begin{aligned} \text{mit} \quad \mathbf{X}'_{k/i} &= R_{k-1,k} \mathbf{X}_{k-1/i} + \mathbf{T}_{k-1,k} \\ \text{und} \quad \Sigma_{\mathbf{X}'_{k/i}} &= R_{k-1,k} \Sigma_{\mathbf{X}_{k-1/i}} R_{k-1,k}^T. \end{aligned}$$

Mit $p(\mathbf{X})$ wird die Projektion $p(\mathbf{X}) = \mathbf{X} / \hat{\mathbf{z}}^T \mathbf{X}$ eines Raumpunktes \mathbf{X} bezeichnet. Aus dem obigen Fehlermaß werden zuerst die Strukturparameter $\mathbf{X}_{k/i}$ eliminiert. Der daraus resultierende Ausdruck wird nach den sechs Bewegungsparametern $\mathbf{m}_{k-1,k}$ minimiert. Durch Fehleranalyse erster Ordnung erhält man die Fehlerkovarianz der Bewegung $\Sigma_{\mathbf{m}_{k-1,k}}$ und der Strukturparameter $\Sigma_{\mathbf{X}_{k/i}}$.

3. Man korrigiert die globale Bewegung $\mathbf{m}_{0,k}$ und deren Fehlerkovarianz nach der Funktion

$$\begin{aligned} \mathbf{m}_{0,k} &= F(\mathbf{m}_{0,k-1}, \mathbf{m}_{k-1,k}) \\ \Sigma_{\mathbf{m}_{0,k}} &= \frac{\partial F}{\partial \mathbf{m}_{0,k-1}} \Sigma_{\mathbf{m}_{0,k-1}} \frac{\partial F}{\partial \mathbf{m}_{0,k-1}}^T + \frac{\partial F}{\partial \mathbf{m}_{k-1,k}} \Sigma_{\mathbf{m}_{k-1,k}} \frac{\partial F}{\partial \mathbf{m}_{k-1,k}}^T, \end{aligned}$$

wobei die Form der Funktion F davon abhängt, ob es sich um eine Objekt- oder Eigenbewegung handelt.

Die Ansätze von [Harris & Pike 88] und [Oliensis & Thomas 91] basieren ebenfalls auf der Berechnung der Bewegung zwischen zwei aufeinander folgenden Bildern. [Oliensis & Thomas 91] verwenden das Fehlermaß (2.13) von [Horn 90] und berechnen die Unsicherheit in der darauf basierten Bewegungsschätzung. Diese Unsicherheit wird zur Aktualisierung der Strukturparameter miteinbezogen. Dagegen wird bei [Harris & Pike 88] keine Rücksicht auf die Unsicherheit in den Bewegungsparametern genommen, und die globale Bewegung wird nicht geschätzt. Aus dem Fehlermaß im obigen Schritt 2.1 wird nur der zweite Term verwendet, damit die Bewegung $\mathbf{m}_{k-1,k}$ berechnet wird. Anhand dieser Schätzung wird die Struktur $\mathbf{X}_{k-1/i}$ auf den nächsten Zeitpunkt fortgeschrieben und entsprechend der Messung $\mathbf{x}_{k/i}$ gemäß einer Kalman-Filter-Korrektur aktualisiert.

[Broida & Chellappa 91] unterstellen folgendes Bewegungsmodell: Der i -te Punkt $\mathbf{X}_{ob,i}$ auf einem Objekt wird in Bezug auf ein Koordinatensystem beschrieben, dessen Ursprung ein unbekanntes Rotationszentrum \mathbf{T}_k und dessen Orientierung R_k in Bezug auf die Kamera sind. Zum Zeitpunkt t_k wird dieser Punkt von der Kamera wie folgt aufgenommen:

$$\mathbf{x}_{k/i} = p(\mathbf{X}_{k/i}) \quad (4.4)$$

$$\text{mit} \quad \mathbf{X}_{k/i} = \mathbf{T}_k + R_k \mathbf{X}_{ob,i} \quad (4.5)$$

Die Bewegung besteht aus einer Translation des Rotationszentrums mit konstanter Geschwindigkeit $\mathbf{v} = \text{konst.}$:

$$\mathbf{T}_k = \mathbf{T}_0 + (t_k - t_0)\mathbf{v} \quad (4.6)$$

und aus einer Rotation R_k des Objekts um dieses Zentrum mit konstanter Winkelgeschwindigkeit $\boldsymbol{\omega} = \text{konst.}$:

$$R_k = I + \frac{\sin(\|\boldsymbol{\omega}\|(t_k - t_0))}{\|\boldsymbol{\omega}\|} [\boldsymbol{\omega}]_{\times} + \frac{1 - \cos(\|\boldsymbol{\omega}\|(t_k - t_0))}{\|\boldsymbol{\omega}\|^2} [\boldsymbol{\omega}]_{\times}^2. \quad (4.7)$$

Zur Eliminierung der Skalierungsmehrdeutigkeit wird jede Translations- und Struktur-unbekannte mit der z -Komponente der Anfangsposition \mathbf{T}_0 des Rotationszentrums skaliert. So erhält man fünf Unbekannte ($t_{x0}/t_{z0}, t_{y0}/t_{z0}, \mathbf{v}/t_{z0}$) für die Translation, drei Unbekannte der Winkelgeschwindigkeit und $3m$ Unbekannten ($\mathbf{X}_{ob,i}/t_{z0}$) $_{i=1..m}$ für die Struktur bei der Verfolgung von m Punkten. Allerdings besitzt das Rotationszentrum bei konstanter Winkelgeschwindigkeit einen Freiheitsgrad, nämlich seine Position auf der Rotationsachse. Daher erhält man insgesamt $4 + 3 + 3m$ Unbekannten. [Broida & Chellappa 89a] formulieren zuerst eine nicht rekursive Problemstellung durch Minimierung der quadrierten Meßresiduen

$$\sum_{k=1}^n \sum_{i=1}^m \left\| \mathbf{x}_{k/i} - p(\mathbf{X}_{k/i}) \right\|^2 \implies \min. \quad (4.8)$$

Anhand dieser globalen² Methode untersuchen [Broida & Chellappa 89b] mögliche Mehrdeutigkeiten. Im Fall einer reinen Translation wird die obige Formulierung singular, weil das Rotationszentrum nicht berechnet werden kann. Weiter werden untere Fehlerschranken nach Cramer-Rao bei verschiedenen Bewegungskonfigurationen ermittelt. Höhere Fehlerempfindlichkeit wird an Kombinationen von betragsmäßig kleiner Translation mit einer vergleichbar größeren Rotation und an abnehmendem effektivem Gesichtsfeld festgestellt.

[Broida *et al.* 90] formulieren ihren eben beschriebenen Ansatz auf eine rekursive Form um, um ein Iteriertes Erweitertes Kalman-Filter darauf anzuwenden. Dazu bleiben die Meßgleichungen wie oben $\mathbf{x}_{k/i} = p(\mathbf{X}_{k/i})$, aber zusätzlich wird das dynamische Verhalten der Unbekannten in einer kontinuierlichen Zustandsform erstellt. Es sei

$$\mathbf{s}(t) = (\mathbf{s}_T^T(t), \mathbf{s}_R^T(t), \mathbf{s}_S^T(t))^T \quad (4.9)$$

der Zustandsvektor, der sich aus jeweils dem translatorischen, dem rotatorischen und dem die Struktur betreffenden Zustand bildet. Der translatorische Zustand besteht aus der skalierten momentanen Position und Geschwindigkeit des Rotationszentrums

$$\mathbf{s}_T(t) = \left(T_x(t)/T_z(t) \quad T_y(t)/T_z(t) \quad v_x/T_z(t) \quad v_y/T_z(t) \quad v_z/T_z(t) \right)^T. \quad (4.10)$$

Für die zeitliche Ableitung beschränken wir uns auf die erste Komponente

$$\dot{\mathbf{s}}_{T1}(t) = \frac{v_x}{T_z(t)} - \frac{T_x(t)}{T_z(t)} \frac{v_z}{T_z(t)} = \mathbf{s}_{T3} - \mathbf{s}_{T1}\mathbf{s}_{T5},$$

²Wir verwenden den Begriff *global* im Gegensatz zum Begriff *rekursiv*, um Methoden zu charakterisieren, die die Messungen von allen Zeitpunkten zur Erstellung einer Schätzung benötigen.

entsprechend lassen sich die übrigen Komponenten berechnen. Der rotatorische Zustand \mathbf{s}_R besteht aus der Einheitsquaternion $\mathcal{Q}(t)$ und der konstanten Winkelgeschwindigkeit $\boldsymbol{\omega}$. Für die zeitliche Darstellung der Rotation eignet sich die Quaternionform, weil sie eine elegante zeitliche Ableitung hat:

$$\dot{\mathcal{Q}}(t) = \frac{1}{2} \begin{pmatrix} 0 & -\boldsymbol{\omega}^T \\ \boldsymbol{\omega} & -[\boldsymbol{\omega}]_{\times} \end{pmatrix} \mathcal{Q}(t), \quad (4.11)$$

und weil ihre vier Komponenten nur eine Nebenbedingung erfüllen müssen. Der dritte Teil \mathbf{s}_S des Zustands besteht aus den skalierten Koordinaten der Objektpunkte $\mathbf{X}_{ob,i}/T_z(t)$ mit entsprechender Ableitung $(-\mathbf{X}_{ob,i}/T_z(t))(v_z/T_z(t))$. Die durch die zeitlichen Ableitungen angegebenen Systemgleichungen werden numerisch integriert, damit sie in diskreter Form die Fortpflanzung des Zustands und seiner Fehlerkovarianz erlauben können.

Zur Überwindung von Konvergenzproblemen des rekursiven Ansatzes bilden [Kumar *et al.* 89] einen Kompromiß zwischen der obigen globalen und den rekursiven Formulierungen. Ohne stochastische Filterung wird anhand eines Satzes von Aufnahmen (Zeitfenster) eine globale Schätzung desselben Zustandsvektors (4.9) durchgeführt, der zusätzlich eine translatorische Beschleunigung miteinschließt. Dann wird diese Schätzung zur Aktualisierung der im letzten Block ermittelten Schätzung benutzt. [Iu & Wohn 90] ersetzen die Koordinaten der Objektpunkte $\mathbf{X}_{ob,i}$ im Zustandsvektor (4.9) durch die momentane Lage der Punkte $\mathbf{X}_i(t)$ bezüglich des Kamerakoordinatensystems.

[Sawhney *et al.* 90] gehen von der Annahme einer reinen Rotationsbewegung aus. Unter perspektivischer Projektion durchlaufen die projizierten Punkte Kegelschnittsegmente auf der Bildebene. Aus diesen Segmenten läßt sich die Lage, die Orientierung und der Radius des Kreises berechnen, den die Objektpunkte im Raum durchlaufen. Durch Gruppierung der Kreise, die demselben starren Objekt entsprechen, läßt sich die Struktur des Objektes ermitteln.

[Shariat & Price 90] schlagen eine Methode zur Bewegungs- und Strukturschätzung unter der Annahme einer Rotation mit konstanter Winkelgeschwindigkeit und einer konstanten Translationsgeschwindigkeit vor. Nach Kompensierung der Translation durchlaufen die Objektpunkte Kreise im Raum. Die Abstände zwischen den Lagen eines Objektpunktes an aufeinander folgenden Zeitpunkten auf dem Kreis sind konstant. Die Kombination dieser Tatsache mit der Starrheit der Abstände zwischen je zwei Objektpunkten führt auf polynomische Gleichungen höchstens zweiter Ordnung, aus denen die Rotation und die Struktur des Objekts berechnet werden können.

Die Ansätze von [Heel 90b; Ando 91] sind die einzigen Ansätze, die Verschiebungsraten als Messungen auswerten, um eine Tiefenkarte rekursiv zu schätzen – es ist nicht zu vergessen, daß wir hier nur Ansätze beschreiben, die das allgemeine Problem der unbekanntenen Struktur und Bewegung behandeln. Für die Tiefe jedes Punktes wird bei [Heel 90b] ein getrenntes Kalman-Filter eingesetzt, dessen Zustand nur ein Element, die Tiefe Z jedes Punktes, enthält. Unter der Annahme einer Eigenbewegung ändert sich die Tiefe mit folgender Rate, siehe Gl. (2.16) nach Änderung der Vorzeichen:

$$\dot{Z} = -v_z - (y\omega_x - x\omega_y)Z. \quad (4.12)$$

Als Meßgleichungen für die Tiefe wird die Gl. (2.18) verwendet. Zur Fortpflanzung der Tiefe mit Hilfe der Gl. (4.12) benötigt man jedoch die Bewegungsparameter, die

man zuerst aus dem Feld aller Verschiebungsraten berechnen soll. Unterstellt man, daß die Tiefen aus der letzten Aktualisierung bekannt sind, so ist die Berechnung der Geschwindigkeiten $(\mathbf{v}, \boldsymbol{\omega})$ aus den Verschiebungsraten ein lineares Ausgleichsproblem.

Auf demselben Prinzip basiert [Heel 90a], jedoch werden nicht mehr die Verschiebungsraten, sondern das Grauwertbild direkt als Messung verwendet. Die Berechnung der Bewegungsparameter bei Vorhandensein einer Schätzung der Tiefenkarte ist ebenfalls ein lineares Ausgleichsproblem (siehe Gl. (2.53)). Die Tiefe wird nach der Fortpflanzung anhand der Gl. (2.52) mit Hilfe eines Kalman-Filters aktualisiert. Der Kern des Ansatzes liegt in dem Algorithmus, der die zeitliche Registrierung der Bildpunkte, die Projektionen desselben Szenenpunkts sind, durchführt.

[Heel 90c] erweitert den Aktualisierungsschritt um einen zusätzlichen Schritt, der die Rekonstruktion der Tiefenkarte unter Berücksichtigung von Diskontinuitäten nach der Methode von [Geiger & Giroi 91] enthält.

Aus den obigen Ansätzen, die ein Modell für die Bewegung eines Objektes unterstellen, ist keiner in der Lage, eine komplizierte Bewegung, wie der Übergang von rein translatorischer auf eine gemischte Bewegung – z.B. beim Einfahren in eine Kurve und beim Verlassen einer Kurve durch ein Auto – zu verfolgen. Realszenenexperimente von [Broida *et al.* 90], deren Ansätze die explizitesten Darstellungen besitzen, zeigen die Schätzung der Bewegung und der Struktur eines Fahrzeugsrads, dessen Rotationsachse konstant bleibt und dessen Struktur nur aus auf einem Kreis markierten Punkten besteht. [Kumar *et al.* 89] testen ihren Ansatz mit der Bewegung eines Autos, aber geben nur die Abweichung zwischen gemessenen und rekonstruierten Trajektorien an, die nicht unbedingt den Fehler der Struktur widerspiegeln.

[Shariat & Price 90] berichten über einen geringen relativen Fehler der Größenordnung von 10% in den Bewegungsparametern und machen keine Angabe zu den Strukturfehlern. Es läßt sich nicht feststellen, ob die globale, auf der Geometrie der Bewegung basierende Schätzmethode zu einer rekursiven Formulierung umgewandelt werden kann.

Die rekursiven Ansätze von [Harris & Pike 88], [Cui *et al.* 90] und [Heel 90b] leiden unter dem Nachteil, daß sie zu jedem Zeitpunkt die Bewegung zwischen zwei aufeinander folgenden Aufnahmen schätzen müssen. Diese Schätzung wird von derselben Fehlerempfindlichkeit wie im Fall des Bildpaars betroffen, wenn das Gesichtsfeld klein ist (Objektbewegung) und die Bewegung kleine Verschiebungen im Bild hervorruft. Beide Ansätze wurden in Fällen von Eigenbewegung mit relativ großem Gesichtsfeld getestet. Bei einer Innenraumbildfolge berichten [Cui *et al.* 90] über einen Fehler der Ordnung von Zentimetern in der Rekonstruktion der Umgebung.

4.2 Zur rekursiven Zustandsschätzung

Bei der Auswertung einer längeren Bildfolge ist man gezwungen, die Schätzung der relativen Bewegung zwischen bildgebendem Sensor und einer Szenenkomponente *rekursiv* durchzuführen, d.h. zu jedem Zeitpunkt t_k den Zustandsschätzwert anhand nur der zu diesem Zeitpunkt aufgenommenen Messungen zu aktualisieren.

Die Schätzungstheorie hat eine lange Geschichte, die mit dem Gebrauch der nicht rekursiven Methode der kleinsten Quadrate zur Auswertung von astronomischen Daten gegen Ende des 18. Jahrhunderts bei Gauß anfängt (siehe dazu [Sorenson 70]) und zur Konsolidierung der rekursiven Filterung 1960 bei Kalman und den heute darauf aufbau-

enden Arbeiten führt. In diesem Abschnitt werden wir nicht versuchen, einen Überblick über Standard-Methoden anzugeben, den man in jedem Lehrbuch [Gelb 74; Anderson & Moore 79; Maybeck 79; Sorenson 80] finden kann, sondern die Aspekte der rekursiven Filterung hervorzuheben, die das Problem der Bewegungs- und Strukturschätzung betreffen. Insbesondere werden wir den Unterschied bei den Optimalitätskriterien betonen und die Eigenschaften der suboptimalen Schätzer untersuchen, die bei mit Nichtlinearitäten behafteten Problemen eingesetzt werden. Zu dieser Untersuchung sind wir nicht nur durch Beobachtung an Experimenten, sondern auch von älteren [Jazwinski 70; Sorenson 74] sowie auch aktuellen Arbeiten [Maybank 90a; Hager 90; Bar-Shalom & Fortmann 88] inspiriert worden.

Gegeben sei ein dynamisches System durch die Systemdifferentialgleichung

$$\dot{\mathbf{s}}(t) = \mathbf{g}(\mathbf{s}(t), \mathbf{u}(t), t) + \mathbf{w}(t), \quad (4.13)$$

wobei $\mathbf{s}(t)$, $\mathbf{u}(t)$ jeweils der Zustandsvektor und der Eingangsgrößenvektor sind. Die Unsicherheit in unserer Modellvorstellung wird durch die Addition vom weißen Rauschen $\mathbf{w}(t)$ dargestellt.

Ein Meßsystem nimmt zum Zeitpunkt t_k Messungen auf, die zum Meßvektor \mathbf{z}_k zusammengefaßt werden. Die Meßgleichung lautet

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{s}_k) + \mathbf{v}_k, \quad (4.14)$$

wobei ein additives weißes ($E[\mathbf{v}_k \mathbf{v}_l^T] = R_k \delta_{kl}$) Meßrauschen mit $E[\mathbf{v}_k] = \mathbf{0}$ unterstellt wurde.

Die Auswertung der Messungen kann zur Regelung des Systems oder zur Ermittlung einer in Hinsicht auf ein darauffolgendes Handeln notwendigen Beschreibung des Systems dienen.

Durch Integration der Systemgleichung (4.13) gelangt man auf folgende diskrete Übergangsgleichung, in der wir den Eingangsgrößenvektor $\mathbf{u}(t)$ weglassen, weil unsere weiteren Überlegungen davon unabhängig sind:

$$\mathbf{s}_{k+1} = \mathbf{f}_k(\mathbf{s}_k) + \mathbf{w}_k, \quad (4.15)$$

wobei $E[\mathbf{w}_k] = \mathbf{0}$ und $E[\mathbf{w}_k \mathbf{w}_l^T] = Q_k \delta_{kl}$. Zusätzlich wird angenommen, daß System- und Meßrauschen zueinander unkorreliert sind $E[\mathbf{w}_k \mathbf{v}_k^T] = 0$.

Der übliche Zyklus in einem rekursiven Schätzvorgang besteht aus einem Prädiktions- und einem Aktualisierungsschritt. Zum Zeitpunkt t_0 wird das a priori Wissen über den Zustand des Systems in Form einer Wahrscheinlichkeitsdichte $p(\mathbf{s}_0)$ zusammengefaßt. Läßt man die Form der Wahrscheinlichkeitsdichte des System- und des Meßrauschens beliebig, so hat man die schwere Aufgabe zu lösen, nach k Messungen die a posteriori Wahrscheinlichkeitsdichte $p(\mathbf{s}_k | \mathcal{Z}^k)$ zu bestimmen, wobei \mathcal{Z}^k die Menge aller Messungen vom Zeitpunkt t_0 bis inklusiv dem Zeitpunkt t_k sind. Damit hätte man die vollständigste stochastische Beschreibung des Zustands des dynamischen Systems.

Nach Anwendung der Bayes-Regel erhält man für den Aktualisierungsschritt

$$p(\mathbf{s}_k | \mathcal{Z}^k) = \frac{1}{c_k} p(\mathbf{z}_k | \mathbf{s}_k) p(\mathbf{s}_k | \mathcal{Z}^{k-1}) \quad (4.16)$$

wobei c_k die totale Wahrscheinlichkeit der Messung \mathbf{z}_k angibt:

$$c_k = p(\mathbf{z}_k | \mathcal{Z}^{k-1}) = \int p(\mathbf{z}_k | \mathbf{s}_k) p(\mathbf{s}_k | \mathcal{Z}^{k-1}) d\mathbf{s}_k. \quad (4.17)$$

Entsprechend erhält man für den Prädiktionsschritt:

$$p(\mathbf{s}_k | \mathcal{Z}^{k-1}) = \int p(\mathbf{s}_k | \mathbf{s}_{k-1}) p(\mathbf{s}_{k-1} | \mathcal{Z}^{k-1}) d\mathbf{s}_{k-1}. \quad (4.18)$$

Der Zyklus wird in Abb. 4.2 dargestellt.

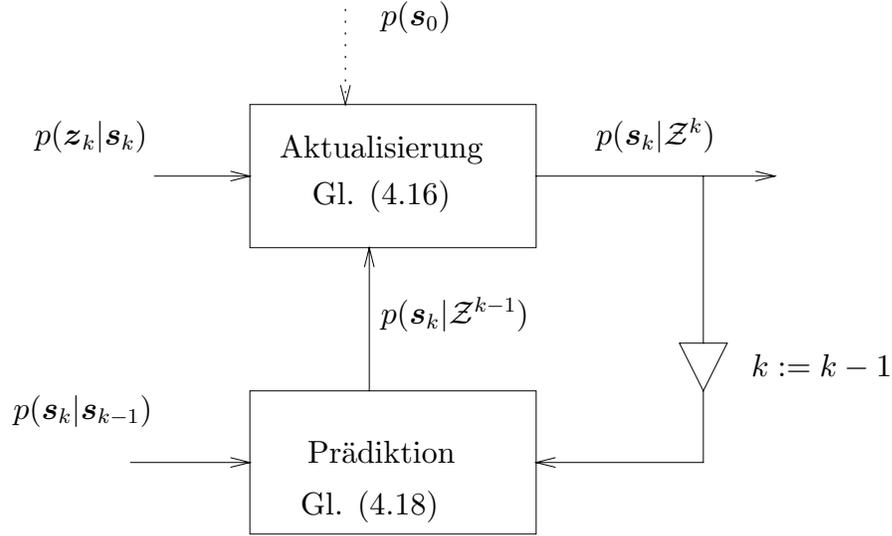


Abbildung 4.1: Allgemeines Schema der rekursiven Filterung.

Der im Sinne eines minimalen gemittelten quadratischen Fehlers (MMSE=Minimal Mean Square Error) optimale Schätzwert wird durch $\hat{\mathbf{s}}_k = E[\mathbf{s}_k | \mathcal{Z}^k]$ angegeben. Unter der Annahme, daß System- und Meßrauschen normalverteilt und die Funktionen \mathbf{f}_k sowie \mathbf{h}_k linear sind, ist die a posteriori Wahrscheinlichkeitsdichte (4.16) normal und der nach MMSE optimale Schätzer ist der Kalman-Filter (siehe [Gelb 74]). Leider wird die Annahme über Linearität - zumindest in der Meßfunktion - bei Auswertung von monokularen Bildfolgen nicht erfüllt, weil die Meßfunktion eine perspektivische Abbildung darstellt. Im weiteren werden wir nur die Annahme über normalverteiltes System- und Meßrauschen beibehalten. Dann lauten die in (4.16) und (4.18) auftretenden bedingten Wahrscheinlichkeitsdichten:

$$p(\mathbf{z}_k | \mathbf{s}_k) = \frac{1}{(2\pi)^{\frac{\dim(\mathcal{Z}_k)}{2}} \sqrt{\det(\mathbf{R}_k)}} \exp\left(-\frac{1}{2}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{s}_k))^T \mathbf{R}_k^{-1}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{s}_k))\right) \quad (4.19)$$

$$p(\mathbf{s}_{k+1} | \mathbf{s}_k) = \frac{1}{(2\pi)^{\frac{\dim(\mathcal{S}_k)}{2}} \sqrt{\det(\mathbf{Q}_k)}} \exp\left(-\frac{1}{2}(\mathbf{s}_{k+1} - \mathbf{f}_k(\mathbf{s}_k))^T \mathbf{Q}_k^{-1}(\mathbf{s}_{k+1} - \mathbf{f}_k(\mathbf{s}_k))\right). \quad (4.20)$$

Nach Einsetzen der bedingten Wahrscheinlichkeitsdichten in (4.16) und (4.18) kann man feststellen, daß die resultierenden a posteriori Wahrscheinlichkeitsdichten bei Aktualisierung und Prädiktion nicht mehr normal sind. Dadurch können sie nicht mehr durch die ersten zwei Momente (Erwartungswert (MMSE) und Fehlerkovarianz) beschrieben werden. [Sorenson 74] beweist zum Beispiel, daß man unter der Annahme einer quadratischen Übergangsfunktion \mathbf{f}_k zur Berechnung der Kovarianz von $p(\mathbf{s}_k | \mathcal{Z}^{k-1})$ beim Prädiktionsschritt die ersten vier Momente von $p(\mathbf{s}_{k-1} | \mathcal{Z}^{k-1})$ benötigt.

Die Tatsache, daß die a posteriori Wahrscheinlichkeitsdichte bei nichtlinearer Meßfunktion und normalverteiltem Meßrauschen nach der Aktualisierung nicht mehr normal bleibt, wird an folgendem Beispiel veranschaulicht: Ein Teilchen bewegt sich entlang einer Geraden im bekannten Abstand vom Projektionszentrum. Der Einfachheit halber wird das Problem in zwei Dimensionen beschrieben, wodurch die Bildebene in eine Gerade (siehe Abb. 4.2) entartet.

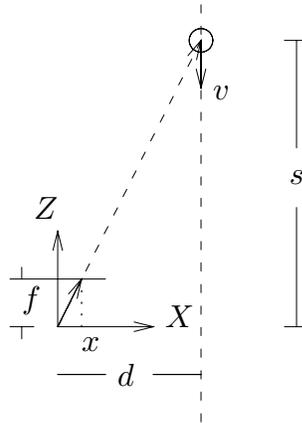


Abbildung 4.2: Einfaches Beispiel der Aufnahme einer geradlinigen Bewegung eines Teilchens

Der Zustand besteht nur aus der Tiefe $s_k = s_k$ des Punktes und die Abbildungsgleichung lautet $x_k = f \cdot d/s_k$ mit den bekannten Parametern f und d . Die Meßfunktion ist nichtlinear bezüglich des Zustands. Wir nehmen an, daß zum Zeitpunkt t_k - dies kann auch der Anfangszeitpunkt t_0 sein - ein präzidiertes Schätzwert \hat{s}_k^- existiert, und daß der Zustand s_k um diesen Wert normalverteilt ist. Die a posteriori Wahrscheinlichkeitsdichte $p(s_k | Z^k)$, ermittelt nach der Regel (4.16), besitzt – wie es in Abb. 4.3 zu sehen ist – eine einseitige Verzerrung zu Gunsten der größeren Tiefen.

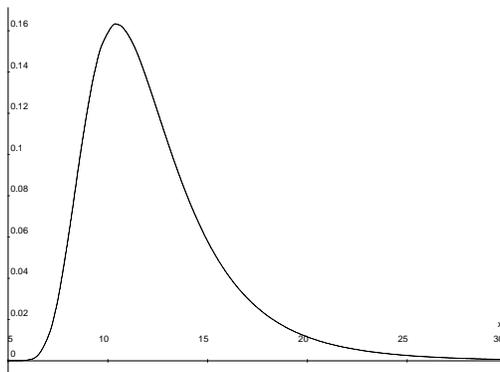


Abbildung 4.3: Nichtnormale Wahrscheinlichkeitsdichte nach einem Aktualisierungsschritt (4.16) am Beispiel der Abb. 4.2.

Denselben Effekt erhält man bei der Stereotriangulation, wie man in Abb. 4.4 beobachten kann.

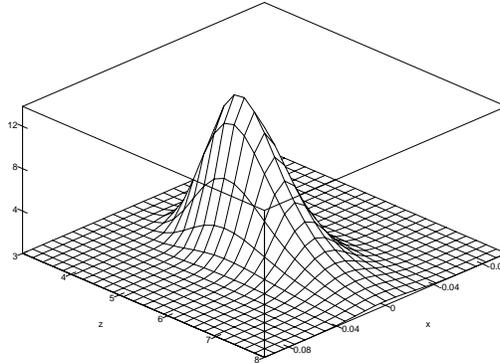


Abbildung 4.4: Gegeben seien zwei Messungen $x_l = f(X + b/2)/Z$ und $x_r = f(X - b/2)/Z$ jeweils von der linken und rechten Kamera, wobei f und b jeweils die Brennweite und die Basislänge sind. In der Abbildung wird die Wahrscheinlichkeitsdichte des rekonstruierten Punktes (X, Z) nach der Stereotriangulation dargestellt.

Im allgemeinen ist man nicht in der Lage, die Wahrscheinlichkeitsdichte in geschlossener Form zu aktualisieren und zu präzisieren. Dabei bildet der Fall der linearen Meßgleichung und des normalverteilten Meßrauschens eine Ausnahme. So muß man im allgemeinen Fall die Wahrscheinlichkeitsdichten diskretisieren, quantisieren, zu jedem Schritt die auf ein Gitter definierten Werte präzisieren und aktualisieren. Weiterhin kann man die mehrdimensionalen Integrationen in (4.17) und (4.18) nicht vermeiden. Diese speicher- sowie auch rechenaufwendigen Schritte erlauben keine Anwendung unter Realweltbedingungen. Deshalb besteht das Ziel der nichtlinearen Filterung in der Suche nach einer effizienten Approximation der Wahrscheinlichkeitsdichte.

Globale Ansätze zur Approximation der a posteriori Wahrscheinlichkeitsdichte (4.16) (z.B. durch eine Summe von Normalverteilungen [Sorenson & Alspach 71]) führen auf Algorithmen mit hohem Aufwand bei mehrdimensionalen Systemen, die aber zum Teil parallelisierbar sind, weshalb ihre Implementierung untersucht werden sollte.

Im weiteren werden wir nur lokale suboptimale Schätzer (EKF³, Filter zweiter Ordnung und IEKF⁴) beschreiben, die einen relativ niedrigen Rechenaufwand haben, und deren Leistung mit der des optimalen Filters an einem einfachen Beispiel vergleichen.

Ab jetzt werden wir den Schätzwert und die Fehlerkovarianz nach der Prädiktion jeweils mit \hat{s}_k^- und P_k^- sowie den Schätzwert und die Fehlerkovarianz nach der Aktualisierung jeweils mit \hat{s}_k^+ und P_k^+ bezeichnen.

Auf der Suche nach einem linearen Schätzer (der einen Schätzwert als lineare Funktion der Messungen und des a priori Schätzwerts angibt) gelangt man auf den Erweiterten Kalman-Filter (EKF) und auf verschiedene Filter zweiter Ordnung. Die Gleichungen

³Erweiterter Kalman-Filter

⁴Iterierter Erweiterter Kalman-Filter

des Erweiterten Kalman-Filters sehen denen des linearen Kalman-Filters ähnlich und lauten wie folgt für die Prädiktion:

$$\hat{\mathbf{s}}_{k+1}^- = \mathbf{f}_k(\hat{\mathbf{s}}_k^+) \quad (4.21)$$

$$P_{k+1}^- = F_k P_k^+ F_k^T + Q_k \quad (4.22)$$

$$\text{mit } F_k = \frac{\partial \mathbf{f}_k}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^+),$$

und entsprechend für die Aktualisierung

$$\hat{\mathbf{s}}_k^+ = \hat{\mathbf{s}}_k^- + K_k(\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^-)) \quad (4.23)$$

$$P_k^+ = (I - K_k H_k) P_k^- \quad (4.24)$$

$$\text{mit } K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (4.25)$$

$$\text{und } H_k = \frac{\partial \mathbf{h}_k}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^-). \quad (4.26)$$

Die Approximationen zur Herleitung des EKF sind $\mathbf{f}_k(\mathbf{s}_k) - \mathbf{f}_k(\hat{\mathbf{s}}_k^+) \approx F_k \cdot (\mathbf{s}_k - \hat{\mathbf{s}}_k^+)$ und $\mathbf{h}_k(\mathbf{s}_k) - \mathbf{h}_k(\hat{\mathbf{s}}_k^-) = H_k \cdot (\mathbf{s}_k - \hat{\mathbf{s}}_k^-)$. Abhängig davon, wie hoch die Nichtlinearität ist, hat die Verwendung des EKF eine Schätzwertverzerrung zur Folge, die nicht von der berechneten Fehlerkovarianz dargestellt werden kann. Dieses Phänomen heißt Divergenz und die Gegenmaßnahme, falls es sich um eine systematische Verzerrung und keine totale Divergenz handelt, besteht in der Einstellung des Systemrauschens auf höhere Werte. Damit versucht man, dem System eine Unsicherheit zu suggerieren, so daß eingehende Meßwerte stärker gewichtet werden.

Durch Mitberücksichtigung von Termen zweiter Ordnung beim Aktualisierungsschritt bekommen die entsprechenden Gleichungen eine sehr komplexe Form, besonders bei mehrdimensionalen Systemen. Deshalb werden verschiedene Vereinfachungen [Jazwinski 70] unternommen. Wir werden zwei *modifizierte* Filter zweiter Ordnung betrachten, den *abgerundeten* und den *Gaußschen* Filter zweiter Ordnung, die wir jeweils mit MTSO⁵ und MGSO⁶ bezeichnen. Ihre Aktualisierungsgleichungen unterscheiden sich nur in einem Term T und lauten:

$$\hat{\mathbf{s}}_k^+ = \hat{\mathbf{s}}_k^- + P_k^- H_k^T T^{-1} (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^-) - \frac{1}{2} \mathbf{s}_h) \quad (4.27)$$

$$\text{mit } (\mathbf{s}_h)_i = \text{Spur}(P_{k-1}^+ \frac{\partial^2 (\mathbf{h}_k)_i}{\partial \mathbf{s}_k^2}) \quad (4.28)$$

$$\text{und } P_k^+ = P_k^- - P_k^- H_k^T T^{-1} H_k P_k^- . \quad (4.29)$$

wobei

$$T = T^{MTSO} = H_k P_k^- H_k^T + R_k - \frac{1}{4} \mathbf{s}_h \mathbf{s}_h^T \quad (4.30)$$

$$\text{oder } T = T^{MGSO} = H_k P_k^- H_k^T + R_k + \frac{1}{2} S_h \quad (4.31)$$

$$\text{mit } (S_h)_{ij} = \sum_{l=1}^n \sum_{m=1}^n \sum_{p=1}^n \sum_{q=1}^n \frac{\partial^2 (\mathbf{h}_k)_i}{\partial (\mathbf{s}_k)_l \partial (\mathbf{s}_k)_m} (P_k^-)_{mp} (P_k^-)_{lq} \frac{\partial^2 (\mathbf{h}_k)_j}{\partial (\mathbf{s}_k)_p \partial (\mathbf{s}_k)_q} \quad (4.32)$$

⁵Modified Truncated Second Order

⁶Modified Gaussian Second Order

Man kann deutlich in Gl. (4.27) und (4.31) sehen, daß sich die Schätzwerte durch die Einführung der Terme zweiter Ordnung von denen des EKF erheblich unterscheiden, wenn die zweiten Ableitungen nicht vernachlässigbar sind. Entsprechend ist der Einfluß der zweiten Ableitungen bei der Bildung der Fehlerkovarianz mittels der Matrizen $\mathbf{s}_h \mathbf{s}_h^T$ und S_h im Term T erkennbar.

Eine Alternative zur MMSE Schätzung ist die Schätzung nach dem Maximum der a posteriori Wahrscheinlichkeitsdichte (MAP), dessen Lage eine gute Approximation zum Erwartungswert ist, wenn die Wahrscheinlichkeitsdichte möglichst symmetrisch um einen einzigen Modus verläuft.

Zur Maximierung der a posteriori Wahrscheinlichkeitsdichte im Fall von normalverteiltem Meßrauschen unterstellen wir zusätzlich, daß der Fehler des prädierten Wertes ebenfalls normalverteilt ist. Die a posteriori Wahrscheinlichkeitsdichte (4.16) lautet im Fall von Normalverteilungen

$$p(\mathbf{s}_k | \mathcal{Z}^k) = \frac{1}{c_k} \exp(-\phi(\mathbf{s}_k))$$

mit

$$\phi(\mathbf{s}_k) = \frac{1}{2}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{s}_k))^T R_k^{-1}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{s}_k)) + \frac{1}{2}(\mathbf{s}_k - \hat{\mathbf{s}}_k^-)^T P_k^{-1}(\mathbf{s}_k - \hat{\mathbf{s}}_k^-). \quad (4.33)$$

Um die a posteriori Wahrscheinlichkeitsdichte zu maximieren, minimieren wir $\phi(\mathbf{s}_k)$ nach \mathbf{s}_k . Im allgemeinen ist man nicht in der Lage, $\hat{\mathbf{s}}_k^+ = \arg \min_{\mathbf{s}_k} \phi(\mathbf{s}_k)$ in geschlossener Form zu ermitteln. Daher wendet man ein Iterationsverfahren an, hier das Gauß-Newton-Verfahren, das aus der Nullstellensuche der ersten Ableitung von $\phi(\mathbf{s}_k)$ besteht [Scales 85]. Wir benutzen den Index i , um einen Schritt in der Iterationsschleife zum Zeitpunkt t_k zu bezeichnen. So lautet die Korrektur beim i -ten Schritt:

$$\hat{\mathbf{s}}_k^{i+1} - \hat{\mathbf{s}}_k^i = -\frac{\partial^2 \phi}{\partial \mathbf{s}_k^2}^{-1}(\hat{\mathbf{s}}_k^i) \frac{\partial \phi}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^i), \quad (4.34)$$

wobei

$$\frac{\partial \phi}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^i) = -H_k^{iT} R_k^{-1}(\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i)) + P_k^{-1}(\hat{\mathbf{s}}_k^i - \hat{\mathbf{s}}_k^-) \quad \text{mit} \quad H_k^i = \frac{\partial \mathbf{h}_k(\hat{\mathbf{s}}_k^i)}{\partial \mathbf{s}_k} \quad (4.35)$$

und

$$\frac{\partial^2 \phi}{\partial \mathbf{s}_k^2}(\hat{\mathbf{s}}_k^i) = \begin{pmatrix} -(\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i))^T R_k^{-1} \frac{\partial H_k^i}{\partial (\mathbf{s}_k)_1} \\ \vdots \\ -(\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i))^T R_k^{-1} \frac{\partial H_k^i}{\partial (\mathbf{s}_k)_n} \end{pmatrix} + H_k^{iT} R_k^{-1} H_k^i + P_k^{-1} \quad (4.36)$$

jeweils die erste und zweite Ableitung von $\phi(\hat{\mathbf{s}}_k)$ sind.

Beim Gauß-Newton-Verfahren wird unterstellt, daß der erste Term mit der Kombination der Residuen $(\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i))$ mit den zweiten Ableitungen $\partial H_k^i / \partial \mathbf{s}_k$ im Vergleich zum Term der ersten Ableitungen $H_k^{iT} R_k^{-1} H_k^i$ vernachlässigbar ist:

$$\frac{\partial^2 \phi}{\partial \mathbf{s}_k^2}(\hat{\mathbf{s}}_k^i) \approx H_k^{iT} R_k^{-1} H_k^i + P_k^{-1}.$$

Daher lautet die Gauß-Newton-Korrektur wie folgt

$$\hat{\mathbf{s}}_k^{i+1} - \hat{\mathbf{s}}_k^i = -(H_k^{iT} R_k^{-1} H_k^i + P_k^{-1})^{-1} \{-H_k^{iT} R_k^{-1} (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i)) + P_k^{-1} (\hat{\mathbf{s}}_k^i - \hat{\mathbf{s}}_k^-)\}. \quad (4.37)$$

Darauf wenden wir die in der Regelungstechnik weit benutzte spezielle Form der Sherman-Morrison Formel [Sorenson 80]

$$(P^{-1} + H^T R^{-1} H)^{-1} = P - P H^T (H P H^T + R)^{-1} H P \quad (4.38)$$

an und erhalten

$$\begin{aligned} \hat{\mathbf{s}}_k^{i+1} - \hat{\mathbf{s}}_k^i &= (P_k^- - P_k^- H_k^{iT} (H_k^i P_k^- H_k^{iT} + R_k)^{-1} H_k^i P_k^-) H_k^{iT} R_k^{-1} (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i)) \\ &\quad - (\hat{\mathbf{s}}_k^i - \hat{\mathbf{s}}_k^-) + P_k^- H_k^{iT} (H_k^i P_k^- H_k^{iT} + R_k)^{-1} H_k^i (\hat{\mathbf{s}}_k^i - \hat{\mathbf{s}}_k^-). \end{aligned}$$

Wir definieren

$$K_k^i = P_k^- H_k^{iT} (H_k^i P_k^- H_k^{iT} + R_k)^{-1}, \quad (4.39)$$

und es gilt

$$K_k^i = (P_k^- - P_k^- H_k^{iT} (H_k^i P_k^- H_k^{iT} + R_k)^{-1} H_k^i P_k^-) H_k^{iT} R_k^{-1} \quad (4.40)$$

Aus den letzten drei Gleichungen folgt der Schritt des IEKF (Iterierter Erweiterter Kalman Filter)

$$\hat{\mathbf{s}}_k^{i+1} = \hat{\mathbf{s}}_k^- + K_k^i (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^i) - H_k^i (\hat{\mathbf{s}}_k^- - \hat{\mathbf{s}}_k^i)). \quad (4.41)$$

in der Formulierung, wie man sie in der Literatur findet [Gelb 74]. Die Fehlerkovarianz wird nur im letzten Iterationsschritt wie folgt berechnet:

$$P_k^+ = (I - K_k^i H_k^i) P_k^- \quad (4.42)$$

Gleichzeitig stellt sich leicht heraus, daß der EKF nichts anderes als der erste Iterationsschritt (für $i = 0$: $\hat{\mathbf{s}}_k^0 = \hat{\mathbf{s}}_k^-$) des IEKF ist. Liegt der a priori Schätzwert weit vom Minimum von $\phi(\mathbf{s}_k)$, ist es nicht zu erwarten, daß man durch einen einzigen Schritt aufs Minimum gelangt.

Der IEKF-Schätzer ist der einfachste bezüglich des Rechenaufwands aus vielen Varianten zur Implementierung der Minimierung bei einer MAP-Schätzung. Wie wir schon nachgewiesen haben, ist der IEKF äquivalent zu der Gauß-Newton-Iteration, die keinen Abstieg der zu minimierenden Funktion (4.33) an jedem Schritt gewährleistet und dadurch divergentes Verhalten aufweisen kann. Als Alternative bietet sich die Minimierung nach Levenberg-Marquardt [Scales 85; Press *et al.* 86] an, deren Iterationsschritt sich von (4.37) wie folgt unterscheidet:

$$\hat{\mathbf{s}}_k^{i+1} - \hat{\mathbf{s}}_k^i = -(H_k^{iT} R_k^{-1} H_k^i + P_k^{-1} + \mu I)^{-1} \frac{\partial \phi}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^i) \quad (4.43)$$

Der Parameter μ wird um einen festen Faktor solange vergrößert, bis der Schritt $\hat{\mathbf{s}}_k^{i+1} - \hat{\mathbf{s}}_k^i$ einen Abstieg der Kostenfunktion (4.33) verursacht, d.h. $\phi(\hat{\mathbf{s}}_k^{i+1}) < \phi(\hat{\mathbf{s}}_k^i)$ wird. Bei sehr großem μ erkennt man, daß durch Vernachlässigung des im Vergleich zu μI kleinen Terms $H_k^{iT} R_k^{-1} H_k^i + P_k^{-1}$ ein Iterationsschritt parallel zur Richtung des Gradienten $\partial \phi / \partial \mathbf{s}_k$ durchgeführt wird. Somit wird das Verfahren äquivalent zur Methode des steilsten Abstiegs, die eine – wenn auch langsame – Konvergenz gewährleistet. Erreicht

das Verfahren eine Abstiegsrichtung, dann wird der Parameter μ verringert, bis er vernachlässigbar ist, so daß das Verfahren äquivalent zum Gauß-Newton-Verfahren wie in Gl. (4.37) wird. Damit erfolgt eine schnelle Konvergenz in der Nähe des Minimums.

Es muß deutlich werden, daß es sich bei diesen Verfahren um iterative Prozeduren handelt, die auf lokalen Minima der Kostenfunktion landen können, falls solche existieren. Je näher der Startwert beim tatsächlichen Wert liegt, desto wahrscheinlicher ist es, daß der Startwert im Fangbereich des globalen Minimums liegt.

Der MAP-Schätzwert fällt mit dem optimalen MMSE-Schätzwert zusammen, wenn die a-posteriori Wahrscheinlichkeitsdichte $p(\mathbf{s}_k|\mathcal{Z}^k)$ unimodal und symmetrisch ist. Dennoch ist die dadurch ermittelte Fehlerkovarianz (4.42) wegen Vernachlässigung der höheren Ableitungen der Meßfunktion niedriger als das tatsächliche zweite Moment von $p(\mathbf{s}_k|\mathcal{Z}^k)$. Durch Verwendung der im Abschnitt 3.4 beschriebenen unteren Fehlerschranke nach Cramer-Rao stellt man fest, daß die in (4.42) ermittelte Fehlerkovarianz und die untere Fehlerschranke zusammenfallen. Eine weitere Abweichung von der Wirklichkeit bildet die Annahme, daß die nach Gl. (4.18) prädizierte $p(\mathbf{s}_k|\mathcal{Z}^{k-1})$ Wahrscheinlichkeitsdichte normalverteilt ist. Das gilt im allgemeinen nicht, einerseits wegen der Nichtlinearität von \mathbf{f}_k und andererseits wegen der Tatsache, daß die aktualisierte Wahrscheinlichkeitsdichte $p(\mathbf{s}_k|\mathcal{Z}^k)$ nach der ersten Messung nicht mehr normal ist. Entsprechend ergibt sich hier eine unerwünschte Erhöhung des Selbstvertrauens des Schätzers.

Am Beispiel der Aufnahme der geradlinigen Bewegung eines Teilchens werden wir die Leistung der beschriebenen suboptimalen Schätzer vergleichen. Zuerst werden wir die Leistung der MAP-Schätzung (MAP), des Erweiterten Kalman-Filters (EKF) und des Modifizierten Gaußschen Filters zweiter Ordnung mit der Leistung des optimalen MMSE-Filters an einem einzigen Aktualisierungsschritt vergleichen. Anschließend bringen wir das Teilchen in Bewegung und untersuchen die Antworten der oben genannten Schätzer. Die Simulation wird analytisch durchgeführt und am Ende werden folgende Werte bezüglich der Abstände möglichst realistisch eingesetzt: Die Brennweite f wird auf 500 Bildpunkte eingestellt, was einer Auflösung von 512 Bildpunkten und einem Gesichtsfeld von ca. 54° entspricht. Der Abstand zu der Geraden, auf der sich das Teilchen befindet, wird auf 0.5m gesetzt und es wird zunächst eine Messung aufgenommen, wenn sich das Teilchen in einer Tiefe von 5m befindet. Das Meßrauschen r_k hat eine Standardabweichung von 1 Bildpunkt. Wir interessieren uns für den Zusammenhang zwischen den Schätzungsergebnissen und den a priori Werten $\hat{\mathbf{s}}_k^-$ und P_k^- . Wir verwenden drei Kombinationen für diese Werte und vergleichen die aktualisierten Werte $\hat{\mathbf{s}}_k^+$ und P_k^+ (siehe Tab. 4.2).

Der Verlauf der entsprechenden a posteriori Wahrscheinlichkeitsdichten $p(\mathbf{s}_k|\mathcal{Z}^k)$ findet sich in Abbildungen 4.5(a-c). Wegen des niedrigen Niveaus des Meßrauschens liegen die exakten Wahrscheinlichkeitsdichten nahe bei der normalen Wahrscheinlichkeitsdichte. Die aus den übrigen Filtern ermittelten Dichten sind per definitionem normal, weil man aus der Filterung als Ergebnis nur die zwei ersten Momente herleiten kann. Das hat als zusätzliche Folge, daß die Modi der Dichten und die Erwartungswerte sehr nahe beieinander liegen, so daß man keinen Unterschied zwischen MMSE- und MAP-Schätzwerten erkennen kann.

Liegt der a priori Schätzwert (Abb. 4.5a) weit vom tatsächlichen Wert und spiegelt sich dies in einer großen a priori Kovarianz wider, erhält man eine erhebliche Verzer-

Abb.	\hat{s}_k^-	P_k^-	MMSE		MAP		EKF		MGSO	
			\hat{s}_k^+	P_k^+	\hat{s}_k^+	P_k^+	\hat{s}_k^+	P_k^+	\hat{s}_k^+	P_k^+
4.5a	8.00	10.00	5.01	0.01	5.00	0.01	3.23	0.06	5.30	2.41
4.5b	6.00	10.00	5.01	0.01	5.00	0.01	4.80	0.02	6.30	3.57
4.5c	6.00	1.00	5.00	0.01	5.01	0.01	4.82	0.02	5.03	0.07

Tabelle 4.2: Vergleich der Filter bei einem Aktualisierungsschritt

rung in der EKF-Schätzung, die aber von einer den aktuellen Fehler unterschätzenden Kovarianz begleitet wird. Dies bildet einen Ausgangspunkt für Divergenz, weil die darauffolgenden Messungen nicht entsprechend stark gewichtet werden, um die aufgetretene Verzerrung zu eliminieren. Dagegen liegt der MGSO-Schätzwert näher beim tatsächlichen Wert und wird von einer eher pessimistischen Fehlerkovarianz begleitet.

Stellt man den a priori Schätzwert näher beim tatsächlichen Wert ein und läßt die Kovarianz gleich groß (Abb. 4.5b), erhält man eine niedrigere Verzerrung im EKF-Schätzwert und eine erheblich höhere im MGSO-Schätzwert, die wir später erläutern werden. Dennoch entspricht die nach MGSO ermittelte hohe Kovarianz dem tatsächlich vorhandenen Fehler.

Am dritten Fall (Abb. 4.5c) simulieren wir eigentlich die Situation, bei der der Schätzwert – nach einer Anzahl von Messungen – nahe beim tatsächlichen Wert liegt und die Fehlerkovarianz niedrig ist. In diesem Fall weicht nur der EKF Schätzwert leicht vom tatsächlichen Wert ab.

Wir werden anhand der Phänomenologie dieser drei Fälle explizit auf die auftretenden Verzerrungen eingehen. Der Übersichtlichkeit halber lassen wir den den Zeitpunkt bezeichnenden Index k weg. Wir benutzen das Symbol s^* für den Kehrwert x/fd , den wir als Approximation für den tatsächlichen Wert der Tiefe unterstellen. Weiterhin verwenden wir die Symbole s_0 für den a priori Schätzwert und s^{EKF} bzw. s^{MGSO} für die je nach Filter geschätzten a posteriori Werte. Wir bezeichnen die Meßkovarianz und die a priori Kovarianz mit r bzw. p . Die Meßfunktion und ihre erste sowie zweite Ableitung lauten wie folgt – berechnet am a priori Schätzwert s_0 :

$$h = h(s)|_{s=s_0} = \frac{fd}{s_0}, \quad h_s = \frac{\partial h}{\partial s}|_{s=s_0} = -\frac{fd}{s_0^2}, \quad h_{ss} = \frac{\partial^2 h}{\partial s^2}|_{s=s_0} = \frac{2fd}{s_0^3} \quad (4.44)$$

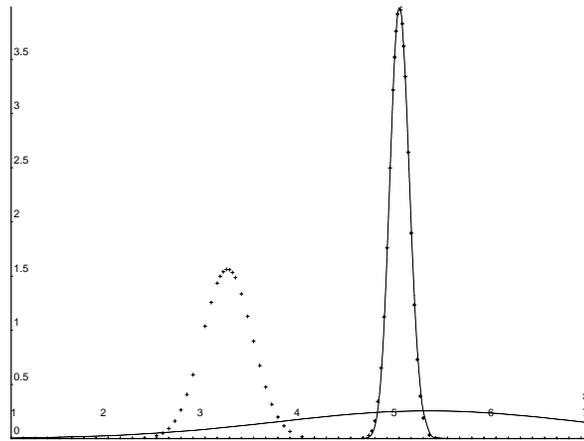
Der aus dem EKF nach Gl. (4.23) erhaltene Schätzwert lautet:

$$s^{EKF} = s_0 + \frac{ph_s}{h_s^2 p + r}(x - h) \quad (4.45)$$

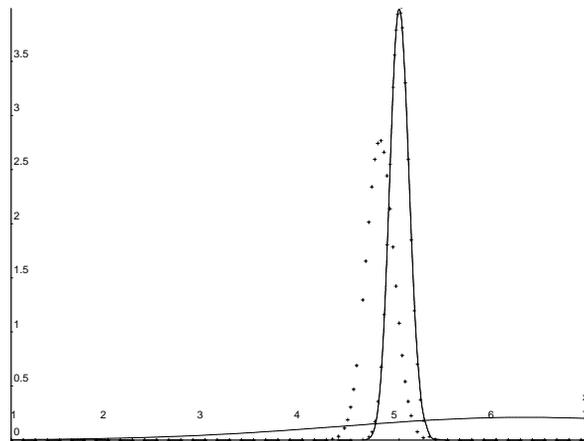
Nach dem Ersetzen der Werte von h und h_s aus (4.44) erhält man für die Verzerrung

$$s^{EKF} - s^* = (s^* - s_0) \left(1 - \frac{s_0}{s^* \left(1 + \frac{rs_0^4}{(fd)^2 p} \right)} \right). \quad (4.46)$$

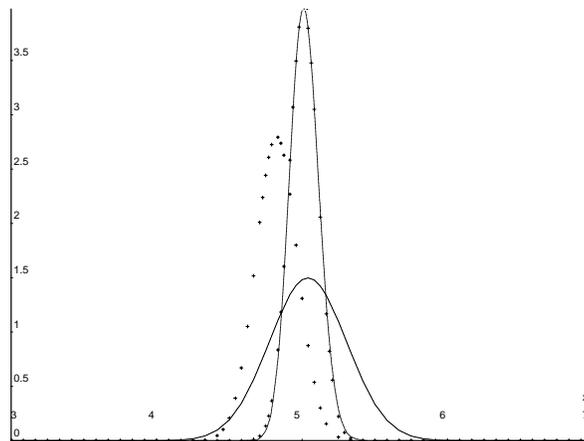
Für das diskutierte Experiment nehmen wir an, daß $p = 10$ – was einer a priori Unsicherheit von ca. $\pm 3\text{m}$ entspricht – und $r = 1$, $fd = 250$ wie oben. Nimmt man an,



(a)



(b)



(c)

Abbildung 4.5: A posteriori Wahrscheinlichkeitsdichten beim Aktualisierungsschritt vier verschiedener Filter. Die Abbildungen (a), (b) und (c) unterscheiden sich nach den a priori Werten, die aus Tabelle 4.2 zu entnehmen sind. Die Dichten der MMSE- und MAP-Schätzwerte (durchgezogene bzw. gepunktete Linie) sind fast identisch für alle drei Sätze von a priori Werten. Das zur symbolischen Berechnung und Aufzeichnung der Kurven verwendete Programmpaket MAPLE gibt uns leider nicht die Möglichkeit, mehr als zwei Linienzeichnungsmuster zu benutzen.

daß der a priori Schätzwert $s_0 \leq 10$ ist, so erhält man $\frac{rs_0^4}{(fd)^2p} \leq \frac{10}{625}$ und daher folgende

Approximation für die Verzerrung

$$s^{EKF} - s^* \approx \frac{(s^* - s_0)^2}{s^*}. \quad (4.47)$$

Man erkennt deutlich, daß die Verzerrung quadratisch mit der Abweichung des a priori Schätzwertes vom tatsächlichen Wert zunimmt.

Aus den Gl. (4.27) und (4.31) erhält man für den MGSO Schätzwert:

$$s^{MGSO} = s_0 + \frac{ph_s}{h_s^2 + r + \frac{1}{2}p^2h_{ss}^2} \left(x - h - \frac{1}{2}ph_{ss} \right), \quad (4.48)$$

und nach Ersetzung der Werte von h , h_s und h_{ss} aus (4.44) ergibt sich die Verzerrung

$$s^{MGSO} - s^* = (s_0 - s^*) \left(1 - \frac{\frac{s_0^3}{s^*}}{s_0^2 + 2p + \frac{rs_0^4}{(fd)^2p}} \right) + \frac{ps_0}{s_0^2 + 2p + \frac{rs_0^4}{(fd)^2p}}. \quad (4.49)$$

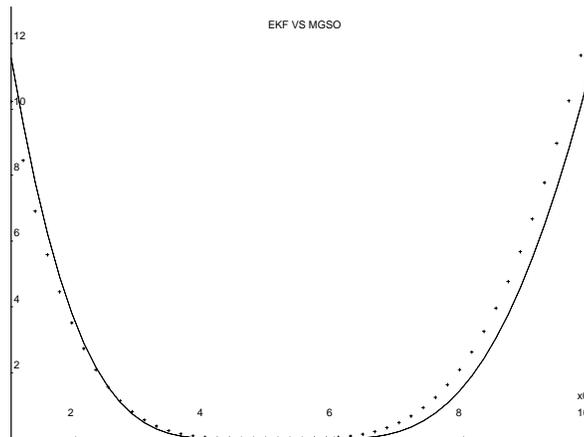
Man kann den Effekt in Abb. 4.5b damit erklären, daß eine Verzerrung (zweiter Term der obigen Gleichungen) dann auftritt, wenn der Schätzwert nahe beim tatsächlichen Wert liegt. Dieser Effekt rührt daher, daß die Kovarianz p im eingeführten Term $\frac{1}{2}ph_{ss}$ der Aktualisierungsgleichung größer als das Quadrat der fast vernachlässigbaren tatsächlichen Abweichung $(s_0 - s^*)$ ist. Die Verläufe der Quadrate der Verzerrungen aus den Gl. (4.46) und (4.27) werden in den Abbildungen 4.6a-b für $s^* = 5$ und $r = 1$ dargestellt.

In der Abb. 4.6b zeigt sich die schlechte Leistung des MGSO-Schätzers in der Nachbarschaft von s^* . Jedoch ist zu bemerken – was in den Abbildungen nicht zu sehen ist –, daß die ermittelte a posteriori Fehlerkovarianz bei der MGSO-Schätzung immer höher als das Quadrat des tatsächlichen Fehlers liegt, während die a posteriori Fehlerkovarianz bei der EKF-Schätzung immer niedriger als der tatsächliche Fehler ist.

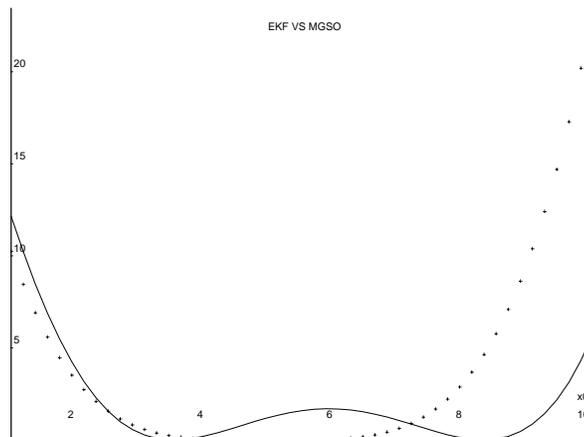
Im folgenden werden wir das Verhalten von IEKF-, EKF- und MGSO-Schätzern am obigen Beispiel nicht mehr statisch, sondern dynamisch untersuchen. Es wird eine geradlinige gleichförmige Bewegung des Teilchens (Abb. 4.2) unterstellt. Der Zustand s_k besteht jetzt nicht nur aus der Tiefe des Punktes, sondern auch aus seiner Geschwindigkeit in der bekannten Richtung der Tiefe. Wir verwenden dieselben Parameter wie oben ($f = 500$, $d = 0.5$, $r = 1$). Das Teilchen befinde sich zum Anfang in einer Tiefe von 5m und habe eine konstante Geschwindigkeit von 0.3m/sec.

In der ersten Simulation stellen wir unseren a priori Zustand auf 10m mit Fehlerkovarianz 25m² (Abb. 4.7(a-c)) ein. Abbildung 4.7a zeigt, daß der EKF-Schätzwert divergiert, und daß die berechnete Fehlerkovarianz den tatsächlichen Fehler unterschätzt. Der MGSO- sowie auch der IEKF-Schätzer zeigen dasselbe Verhalten mit einem Fehler von ca. 0.1m.

In der nächsten Simulation stellen wir den Startwert auf 8m – näher zum tatsächlichen Wert – und die dazu gehörende Fehlerkovarianz auf 10m² ein. Abbildung 4.8 zeigt, daß der EKF-Schätzer eine stetige kleine Verzerrung erweist, die von der berechneten Fehlerkovarianz unterschätzt wird. Das Verhalten von MGSO- und IEKF-Schätzern ist gleich gut wie in der ersten Simulation, weshalb auf eine entsprechende Abbildung verzichtet wird. Es ist noch zu bemerken, daß der IEKF-Schätzer nach den zwei ersten Messungen maximal zwei Iterationen bis zur Konvergenz benötigt.



(a)



(b)

Abbildung 4.6: Verlauf der Quadrate der Verzerrungen von EKF und MGSO als Funktion von dem Startwert s_0 bei Startvarianz $p = 1$ (a) und $p = 10$ (b). Der tatsächliche Wert liegt bei $s^* = 5$. Die Verzerrung von EKF wird durch eine gepunktete und die von MGSO durch eine durchgezogene Linie dargestellt.

Wir gehen jetzt auf einen wichtigen Punkt ein, der bis hier nicht untersucht wurde. Die von einem Schätzer ermittelte Fehlerkovarianz dient nicht nur der geeigneten Gewichtung der darauffolgenden Messungen. Sie muß auch zur Lösung der Aufgabe der Erstellung von zeitlichen Korrespondenzen zwischen der Projektion des Zustands in den Meßraum (in der Bildfolgenauswertung in die Bildebene) und den neuen Messungen (auch Problem der Datenassoziation nach [Bar-Shalom & Fortmann 88] genannt) beitragen. Falls mehrere Messungen vorliegen, muß die geeignete Messung mit Hilfe eines Validierungsbereichs ausgewählt werden, den man sich als ein Ellipsoid im Raum der Messungen vorstellen kann. Dieses Validierungselipsoid entsteht aus der folgenden Mahalanobis-Distanz-Ungleichung, die die aus der Fehlerfortpflanzung ermittelte gewichtete Abweichung zwischen prädiiziertem Schätzwert und einem Punkt im Meßraum

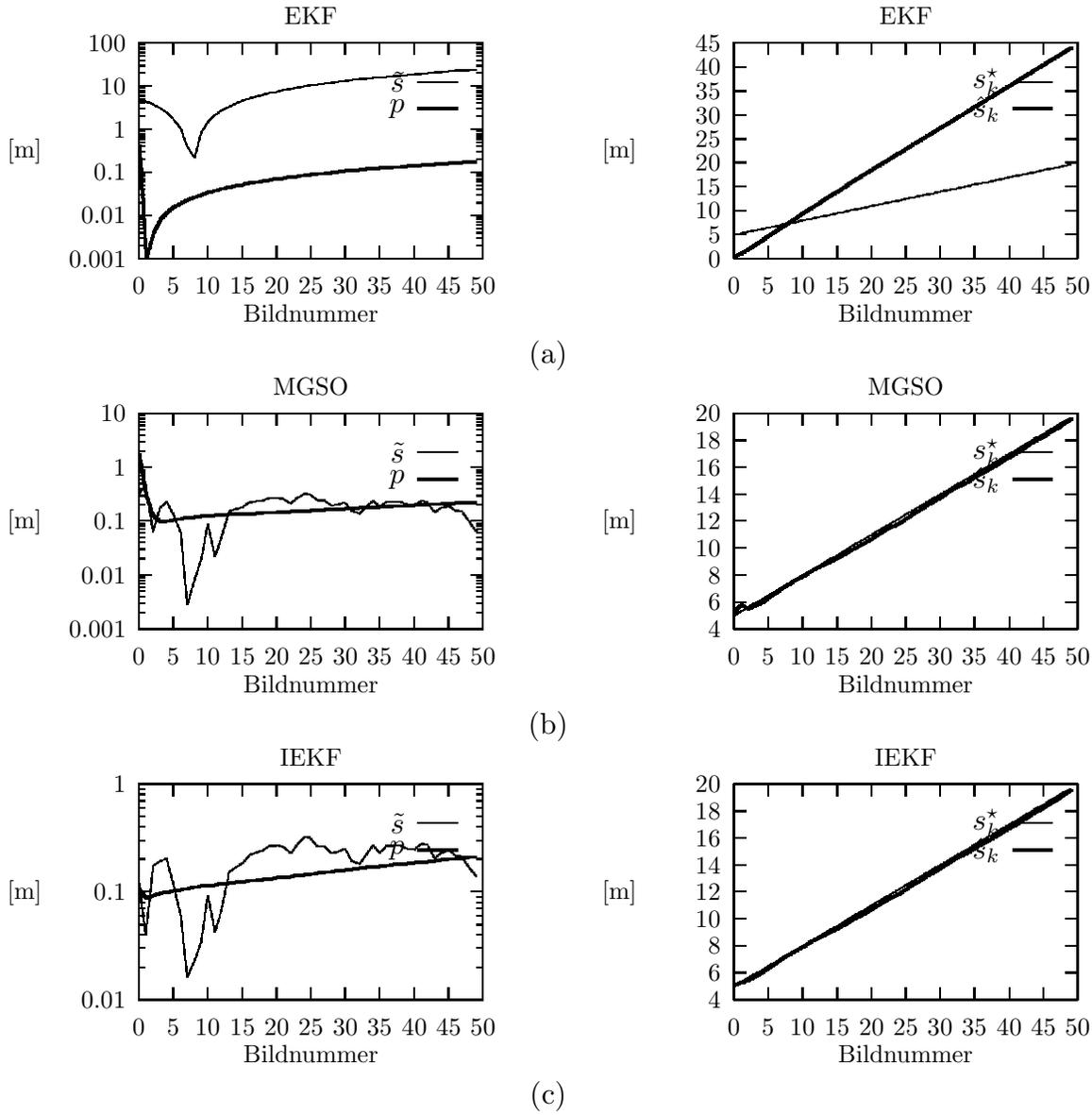


Abbildung 4.7: Zeitlicher Verlauf der Differenz $\tilde{s} = \hat{s} - s^*$ und der berechneten Fehlerkovarianz p (links) und Gegenüberstellung der tatsächlichen Tiefe s^* zur geschätzten \hat{s}_k (rechts) nach einer EKF-Schätzung (a), MGSO-Schätzung (b) und IEKF-Schätzung (c) ($s_0 = 10\text{m}$, $p_0 = 25\text{m}^2$).

darstellt:

$$d_{MAH} = (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^-))^T (H_k P_k^- H_k^T + R_k)^{-1} (\mathbf{z}_k - \mathbf{h}_k(\hat{\mathbf{s}}_k^-)) \leq c. \quad (4.50)$$

Die Mahalanobis-Distanz ist χ^2 -verteilt mit $\dim(\mathbf{z}_k)$ Freiheitsgraden. Aus der Verteilungstabelle kann man den Wert c für ein bestimmtes Konfidenzniveau auswählen und anhand der Ungleichung Messungen zu Projektionen $\mathbf{h}_k(\hat{\mathbf{s}}_k^-)$ assoziieren. Falls die Kovarianz den tatsächlichen Fehler unterschätzt und sehr niedrige Werte nimmt, wird die Mahalanobis-Distanz und entsprechend der Validierungsbereich kleiner, was zum fehlerhaften Verwerfen von eigentlich korrekten Zuordnungen führen kann. Das kann als

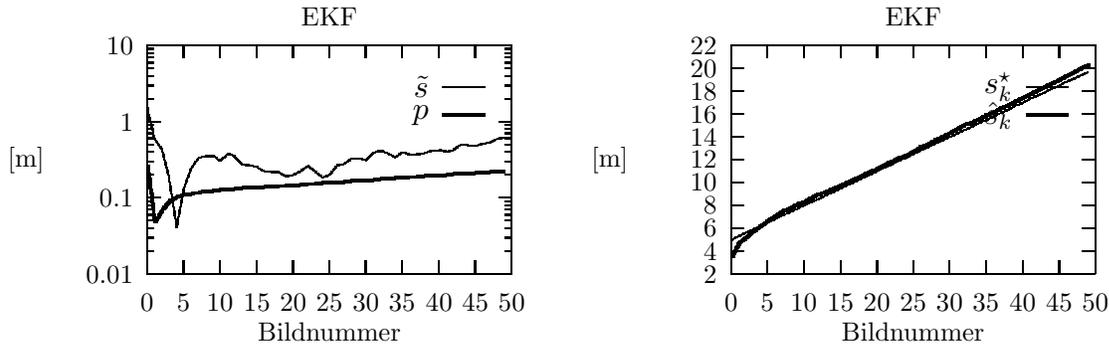


Abbildung 4.8: Zeitlicher Verlauf der Differenz $\tilde{s} = \hat{s} - s^*$ und der berechneten Fehlerkovarianz p (links) und Gegenüberstellung der tatsächlichen Tiefe s^* zur geschätzten \hat{s}_k (rechts) nach einer EKF-Schätzung ($s_0 = 8m$, $p_0 = 10m^2$).

fatale Folge die Unfähigkeit zur Verfolgung eines sich bewegenden Objektes haben.

In diesem Abschnitt haben wir an einem einfachen Beispiel die Probleme der Schätzung mit einer nichtlinearen Form der Meßfunktion untersucht. Wir haben fast ausschließlich den Aktualisierungsschritt der Schätzung behandelt, weil die tatsächliche Rückführung von der Messung auf den Zustand bei diesem Schritt stattfindet. Weiterhin kann der Prädiktionsschritt bei hoher zeitlichen Abtastrate zuverlässig durch eine Linearisierung vereinfacht werden. Wir haben für das Meßrauschen eine Normalverteilung unterstellt. Im Fall einer von der Normalverteilung erheblich abweichenden Verteilung existieren Methoden zur Approximation der Wahrscheinlichkeitsdichte durch Reihenentwicklung wie z.B. in [Sorenson & Alspach 71]. Bei einer nach oben und unten beschränkten Unsicherheit in der Messung und im a priori Wissen sind als Alternative zur stochastischen Schätzung mengenbasierte Ansätze vorgeschlagen worden (siehe unter anderen [Schweppe 73; Orr *et al.* 91; Atiya & Hager 91]).

Abschließend fassen wir die wichtigsten Punkte unserer Untersuchung zur rekursiven stochastischen Zustandsschätzung zusammen:

- Unter der Annahme von normalverteiltem Meß- und Systemrauschen, sowie auch normalverteiltem a priori Fehler, ist die a posteriori Wahrscheinlichkeitsdichte $p(\mathbf{s}_k | \mathcal{Z}^k)$ nach k Messungen nicht mehr normal. Sie weicht von einer normalen Wahrscheinlichkeitsdichte ab, wenn das Meßrauschen im Vergleich zum a priori Schätzfehler hoch ist und der a priori Schätzwert weit vom tatsächlichen Wert liegt.
- Falls die a posteriori Wahrscheinlichkeitsdichte ähnlich zu einer normalen Wahrscheinlichkeitsdichte ist, fallen der Modus und der Erwartungswert der Dichte zusammen. Daher kann man einen Minimierungsalgorithmus zur MAP-Schätzung einsetzen. Die bekannteste Variante davon ist der IEKF, der am untersuchten Beispiel die beste Leistung erbracht hat.
- Der Erweiterte Kalman-Filter ist abhängig von der Abweichung des Startwertes vom tatsächlichen Wert. Wenn die Abweichung groß ist, tritt eine Verzerrung auf, die sich nicht in der vom Filter berechneten Fehlerkovarianz erfassen läßt.
- Eine bessere nicht-iterative Approximation zum optimalen Filter ist der Modifizierte Gaußfilter zweiter Ordnung, der bei den Simulationen eine zum IEKF

vergleichbare Leistung erbracht hat. Dennoch hat es sich bei der analytischen Behandlung herausgestellt, daß dieser Schätzer eine Verzerrung verursacht, falls der a priori Schätzwert in der unmittelbaren Umgebung des tatsächlichen Zustandswertes liegt und die a priori Fehlerkovarianz immer noch hoch bleibt.

4.3 Rekursive Schätzung der relativen Bewegung einer Ebene

Im letzten Abschnitt haben wir die Schritte der rekursiven Zustandsschätzung beschrieben und sind auf den Fehler in den geschätzten Werten und auf die Zuverlässigkeit der ermittelten Fehlerkovarianz an einem einfachen Beispiel mit einem skalaren Zustand eingegangen. In diesem Abschnitt werden wir auf die rekursive Schätzung der relativen Bewegung und der Struktur einer Szenekomponente eingehen. Wie auch im dritten Kapitel gehen wir davon aus, daß die zeitliche Zuordnung der Bildbereichshinweise in einer an jedem Zeitpunkt vorgeschalteten Verarbeitungsstufe durchgeführt wurde. Unsere Untersuchung schließt sich an die im dritten Kapitel aufgeführten Untersuchungen an, d.h. der Schwerpunkt liegt in der Analyse der Fehlerempfindlichkeit. Das Ziel ist ein zweifaches:

1. Erstens ist uns von Interesse, die Bewegungs- und Geometriekonstellationen zu ermitteln, bei denen die Einführung von a priori Wissen in Form eines Bewegungsmodells eine Abschwächung der Fehlerempfindlichkeit verursacht.
2. Zweitens wollen wir die Eignung von jedem der im vorangehenden Abschnitt untersuchten Schätzer bei diesem komplexeren Problem untersuchen.

Die Fülle der Bildbereichshinweise, die Anzahl der Unbekannten und die Einführung von a priori Wissen erlauben keine analytische Untersuchung. Daher ist die Vorgehensweise in diesem Abschnitt experimentell, im Gegensatz zu den expliziten Nachweisen im dritten Kapitel. Der Verlauf des Schätzfehlers wird mit Hilfe von Experimenten mit synthetischen Daten ermittelt. Die Experimente mit synthetischen Daten werden mit realistischen Werten im Anschluß an ein Experiment mit einer realen monokularen Bildfolge durchgeführt.

Die Anzahl der Unbekannten, die der Struktur einer Szenekomponente entsprechen, ist proportional zur Anzahl der charakteristischen Szenenbereichshinweise (Vertizes, Kanten, markierte Punkte). Damit wir aber eine Einsicht in den Fehlerverlauf der Struktur bekommen und Rückschlüsse auf die Fehlerempfindlichkeitsergebnisse des dritten Kapitels machen können, werden wir die grundlegende Annahme machen, daß wir nur die Menge der Szenenbereichshinweise untersuchen, die sich in einer Ebene befinden. So reduziert sich die Anzahl der Unbekannten auf zwei, den Polar- und Azimutwinkel der Normalen der Ebene. Der Abstand der Ebene vom Projektionszentrum wird wegen der Skalierungsmehrdeutigkeit im Betrag der Translation mitberücksichtigt.

Diese Annahme steht im Einklang mit aktuellen Anwendungen der Bewegungsschätzung in Realweltsituationen. Nehmen wir an, daß eine Kamera auf einem autonom geführten Fahrzeug befestigt ist, so ist die befahrbare Fläche eine Ebene. Der hier vorgestellte Algorithmus berechnet die momentane Translations- und Winkelgeschwindigkeit der Ego-bewegung in Bezug auf das Kamerakoordinatensystem, sowie auch die Normale der befahrbaren Ebene. Damit erspart man sich die Ermittlung der Transformation vom

Straßenkoordinatensystem zum Kamerakoordinatensystem [Enkelmann 91; Dickmanns & Graefe 88b].

Eine zweite Anwendung ist die Auswertung von monokularen Bildfolgen aufgenommen von einer Kamera, die am Greifer eines Roboterarms befestigt ist. Eine weit verwendete Annahme ist, daß die zu manipulierenden Objekte polyedrisch sind. Der Algorithmus von [Sugihara 86] und seine Weiterentwicklung in Bezug auf Realweltszenen [Müller 92] erlauben eine Aussage über die Korrektheit einer polyedrischen Struktur und eine Segmentierung der Bildbereichshinweise in Gruppen von Hinweisen, die Abbildungen von sich in jeweils einer Ebene befindenden Szenenbereichshinweise sind. Der hier vorgestellte Algorithmus ermittelt die Egobewegung in Bezug auf das Kamerakoordinatensystem und die Normale der berücksichtigten Ebene. Ist die Transformation vom Kamerakoordinatensystem zum Greiferkoordinatensystem (Hand-Auge-Kalibrierung) bekannt, so läßt sich die zum beabsichtigten Zugriff auf das Objekt benötigte Orientierung bezüglich des Roboterkoordinatensystems ermitteln. Eine dritte Anwendung ist die Ermittlung der Bewegung von Objekten, die man als polyedrische Strukturen modellieren kann.

Das eingeführte a priori Wissen betrifft die Glattheit der Bewegung. Wir nehmen an, daß die Translations- sowie auch die Winkelgeschwindigkeit in Bezug auf das Kamerakoordinatensystem konstant bis auf eine Unsicherheit bleiben, die sich mittels der Kovarianz des Systemrauschens darstellen läßt. Das trifft im Fall eines autonom geführten Fahrzeugs sowie auch eines Roboterarms zu, wenn die Bewegung zwischen zwei Aufnahmen gering ist. Die Annahme ist für den Fall der Verwendung von Verschiebungsraten im Bildbereich geeignet, weil die Ermittlung von Verschiebungsraten kleine Bewegungen im Abtastintervall unterstellen. Im Fall eines sich bewegenden Objektes soll das Modell so umgestellt werden, daß die entsprechenden Geschwindigkeiten in Bezug auf das Objektkoordinatensystem dargestellt werden. Führt man eine konstante Translationsbeschleunigung ein, die auch eine Funktion der Winkelgeschwindigkeit ist (z.B. $\dot{\boldsymbol{v}} = \boldsymbol{\omega} \times \boldsymbol{v}$), so kann man auch den Fall der Objektbewegung in Bezug auf das Kamerakoordinatensystem beschreiben (siehe Diskussion in [Zhang 90]). Wenn die Abtastrate niedrig oder die relative Bewegung zu schnell ist, ist man gezwungen, eine komplexere Bewegung zu modellieren (siehe das Modell zur Autobewegung auf einer Trajektorie in Form einer Klothoide [Dickmanns & Graefe 88b; Schick & Dickmanns 91] oder das einer Präzessionsbewegung [Young & Chellappa 90a]).

Folgende Ansätze beziehen sich auf die Schätzung der relativen Bewegung einer Ebene aus einer längeren monokularen Bildfolge. [Sull & Ahuja 91a] nutzen die Information aus der ganzen Bildfolge aus, um in nicht rekursivem Modus die Bewegung eines Objektes zu berechnen. [Castelow & Rerolle 91] schränken die Bewegung auf drei Freiheitsgrade (Bewegung eines Fahrzeugs) ein und führen eine Beschleunigung in der Translation ein. Unser Ansatz ähnelt der Methode von [Murray & Pickup 91], wobei aber auf die Eignung des rekursiven Schätzers nicht eingegangen wird: die beiden letzten Ansätze verwenden das Erweiterte Kalman Filter. Weiterhin vereinfachen wir die Meßgleichungen durch Einführung von Hilfsparametern und modellieren korrekt die Kollisionszeit.

Wir gehen auf die Beschreibung der vorgestellten Methode im Zustandsraum ein. Wir bezeichnen mit \boldsymbol{v} , $\boldsymbol{\omega}$ und \boldsymbol{N} jeweils die Translations-, die Winkelgeschwindigkeit und die Einheitsnormale der Ebene in Bezug auf das Kamerakoordinatensystem. Der

Abstand des Kamerazentrums von der Ebene ist d , so daß die Gleichung der Ebene im Kamerakoordinatensystem lautet : $\mathbf{N}^T \mathbf{X} = d$. (siehe Abb. 4.9).

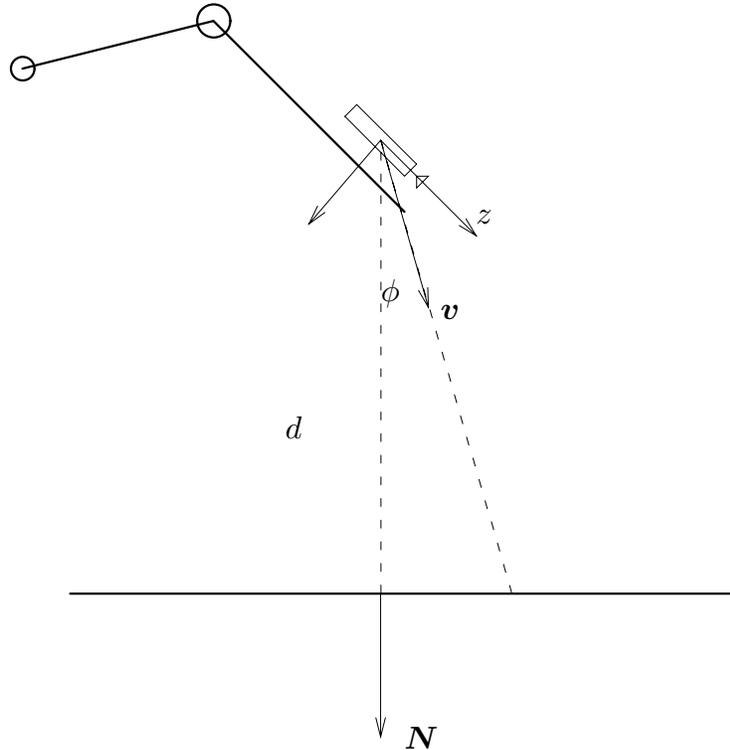


Abbildung 4.9: Bewegung einer an einem Greifer befestigten Kamera

Unter der Annahme von konstanten Geschwindigkeiten erhalten wir folgende Differentialgleichungen für den zeitlichen Verlauf des Systems:

$$\begin{aligned} \dot{\mathbf{v}} &= 0 & \dot{\boldsymbol{\omega}} &= 0 \\ \dot{\mathbf{N}} &= -\boldsymbol{\omega} \times \mathbf{N} & \dot{d} &= -\mathbf{v}^T \mathbf{N} \end{aligned} \quad (4.51)$$

Wie wir schon im zweiten Kapitel erwähnt haben, leidet die Auswertung einer monokularen Bildfolge unter der Mehrdeutigkeit des dem Problem inhärenten Skalierungsfaktors, von dem der Betrag der Translationsgeschwindigkeit $\|\mathbf{v}\|$ und der Abstand d zum Objekt abhängen. Bei der Auswertung einer längeren monokularen Bildfolge kann diese Skalierung nach einer Festsetzung des Geschwindigkeitsbetrags oder des Abstands beim ersten Bildpaar eliminiert werden (siehe [Weng *et al.* 87; Cui *et al.* 90]). Dennoch erfolgt bei Verwendung eines rekursiven Schätzers eine Verzerrung in den übrigen Parametern, wenn man diese Festlegung vornimmt. Deshalb wird hier der Skalierungsfaktor in der rekursiven Schätzung miteinbezogen, was man auch bei anderen Ansätzen finden kann (siehe [Broida & Chellappa 91]).

Wir führen als unbekanntes Vektor die durch den Abstand zur Ebene skalierte Translationsgeschwindigkeit $\mathbf{u} = \mathbf{v}/d$ ein ⁷. Der Betrag dieses Vektors wird von [Murray & Pickup 91] als das Inverse der Kollisionszeit betrachtet. Wie man aber aus der Abb.

⁷Vergleiche verschiedene, aber entsprechende Skalierungen in den Abschnitten 2.3 und 3.4.

4.9 erkennen kann, stimmt das nicht, falls man die Kollisionszeit als die Zeit definiert, die bei konstanter Geschwindigkeit bis zum Treffen der Ebene verläuft. Nennt man den Winkel zwischen der Normalen und der Translationsgeschwindigkeit ϕ , läßt sich die Kollisionszeit als

$$\frac{1}{\rho_{ttc}} = \frac{d}{\|\mathbf{v}\| \cos \phi} = \frac{d}{\mathbf{v}^T \mathbf{N}} = \frac{1}{\mathbf{u}^T \mathbf{N}} \quad (4.52)$$

beschreiben. Wir werden das Inverse der Kollisionszeit $\rho_{ttc} = \mathbf{u}^T \mathbf{N}$ als Nebenprodukt aus unserem Algorithmus erhalten und es als Bewertungsmaß verwenden.

Wir definieren den Zustandsvektor des Systems als den neunelementigen Vektor

$$\mathbf{s} = \begin{pmatrix} \mathbf{u}^T & \boldsymbol{\omega}^T & \mathbf{N}^T \end{pmatrix}, \quad (4.53)$$

wobei \mathbf{N} ein Einheitsvektor ist. Die entsprechende Zwangsbedingung $\|\mathbf{N}\| = 1$ wird in den Meßgleichungen berücksichtigt.

Zur Durchführung des Prädiktionsschrittes benötigen wir die Übergangsgleichung des Systems vom Zeitpunkt t_k auf den Zeitpunkt t_{k+1} , die wir aus den Differentialgleichungen (4.51) herleiten müssen. Die Übergangsgleichung für die Winkelgeschwindigkeit ist recht einfach:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k. \quad (4.54)$$

Die Integration der Differentialgleichung für die Normale ist bekannt [Korn & Korn 68]. Die Lösung ist die Rotation der Normalen um eine Achse parallel zu $\boldsymbol{\omega}$ um den Winkel $-\|\boldsymbol{\omega}\|T$, wobei $T = t_{k+1} - t_k$:

$$\mathbf{N}_{k+1} = \mathbf{N}_k - \frac{\sin(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|} \boldsymbol{\omega}_k \times \mathbf{N}_k + \frac{1 - \cos(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|^2} \boldsymbol{\omega}_k \times (\boldsymbol{\omega}_k \times \mathbf{N}_k). \quad (4.55)$$

Die Ermittlung der Übergangsgleichung für die skalierte Geschwindigkeit \mathbf{u} ist aufwendiger. Wir werden erst die Differentialgleichung in (4.51) für d mit Hilfe von (4.55) ausrechnen:

$$\begin{aligned} d_{k+1} - d_k &= -\mathbf{v}_k^T \int_{t_k}^{t_{k+1}} \left\{ \mathbf{N}_k - \frac{\sin(\|\boldsymbol{\omega}_k\|(\tau - t_k))}{\|\boldsymbol{\omega}_k\|} \boldsymbol{\omega}_k \times \mathbf{N}_k \right. \\ &\quad \left. + \frac{1 - \cos(\|\boldsymbol{\omega}_k\|(\tau - t_k))}{\|\boldsymbol{\omega}_k\|^2} \boldsymbol{\omega}_k \times (\boldsymbol{\omega}_k \times \mathbf{N}_k) \right\} d\tau \\ &= -\mathbf{v}_k^T \mathbf{N}_k T + \frac{\mathbf{v}_k^T (\boldsymbol{\omega}_k \times \mathbf{N}_k) (1 - \cos(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^2} \\ &\quad - \frac{(\mathbf{v}_k \times \boldsymbol{\omega}_k)^T (\boldsymbol{\omega}_k \times \mathbf{N}_k) (\|\boldsymbol{\omega}_k\|T - \sin(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^3} \end{aligned} \quad (4.56)$$

Für den Übergang der skalierten Geschwindigkeit gilt dann

$$\mathbf{u}_{k+1} = \frac{\mathbf{v}_{k+1}}{d_{k+1}},$$

und durch Ersetzung von d_{k+1} mit dem Ausdruck aus (4.56) und Normierung durch d_k erhalten wir

$$\mathbf{u}_{k+1} = \frac{\mathbf{u}_k}{1 - \mathbf{u}_k^T \mathbf{N}_k T + \frac{\mathbf{u}_k^T (\boldsymbol{\omega}_k \times \mathbf{N}_k) (1 - \cos(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^2} - \frac{(\mathbf{u}_k \times \boldsymbol{\omega}_k)^T (\boldsymbol{\omega}_k \times \mathbf{N}_k) (\|\boldsymbol{\omega}_k\|T - \sin(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^3}}. \quad (4.57)$$

Zur Vervollständigung der Beschreibung des Systemübergangs benötigen wir die Jacobische Matrix der Übergangsfunktion, d.h. die Ableitung der Zustandsgrößen zum Zeitpunkt t_{k+1} nach den Zustandsgrößen zum Zeitpunkt t_k , um den Prädiktionsschritt bei einem rekursiven Schätzer durchzuführen. Die Berechnung dieser Jakobischen Matrix findet man im Anhang D.

Wenn die Bewegung keine rotatorische Komponente besitzt, vereinfachen sich die Übergangsgleichungen für \mathbf{u}_k und \mathbf{N}_k wie folgt:

$$\mathbf{N}_{k+1} = \mathbf{N}_k \quad \mathbf{u}_{k+1} = \frac{\mathbf{u}_k}{1 - \mathbf{u}_k^T \mathbf{N}_k \mathbf{T}}. \quad (4.58)$$

Als Messung verwenden wir das Feld der Verschiebungsraten zum Zeitpunkt t_k . Aus der Annahme, daß es sich um die relative Bewegung einer Ebene handelt, lassen sich aus den Verschiebungsraten acht Hilfsparameter (siehe Abschnitt 3.4, Gl. 3.52) herleiten, die wir hier für den Fall der Egebewegung und mit der oben für den rekursiven Ansatz eingeführten neuen Notation wiederholen.

$$\mathbf{q}_k = \begin{pmatrix} -u_{kx}N_{kz} - \omega_{ky} \\ -u_{kx}N_{kx} + u_{kz}N_{kz} \\ -u_{kx}N_{ky} + \omega_{kz} \\ -u_{ky}N_{kz} + \omega_{kx} \\ -u_{ky}N_{kx} - \omega_{kz} \\ -u_{ky}N_{ky} + u_{kz}N_{kz} \\ -\omega_{ky} + u_{kz}N_{kx} \\ u_{kz}N_{ky} + \omega_{kx} \end{pmatrix} \quad (4.59)$$

Die Vorzeichen für die Geschwindigkeiten sind geändert worden, weil man im Fall der Egebewegung für die Berechnung der Verschiebungsraten von der Gleichung

$$\dot{\mathbf{X}} = -\boldsymbol{\omega} \times \mathbf{X} - \mathbf{v}$$

ausgeht (vergleiche dazu die Bemerkung im Anschluß an Gl. 2.16). Die acht Parameter können aus $m \geq 4$ Verschiebungsraten durch Lösung eines linearen Systems ermittelt werden, das aus folgender Minimierung folgt:

$$\sum_{i=1}^m (\dot{\mathbf{x}}_{ki} - B_{ki} \mathbf{q}_k)^T C_{ki}^{-1} (\dot{\mathbf{x}}_{ki} - B_{ki} \mathbf{q}_k) \implies \min_{\mathbf{q}_k}. \quad (4.60)$$

Die Matrix C_{ki} ist die Kovarianz des Rauschens in den Verschiebungsraten, die abhängig vom Verfahren zur Ermittlung des optischen Flusses ist. Man kann durch die Einführung der Pseudoinversen statt der Inversen in (4.60) die Gewichtung nur der Projektion des optischen Flusses auf eine stabile Richtung erzwingen. Ist z.B. nur die Komponente entlang des Grauwertgradienten meßbar, so bekommen wir in (4.60) direkt die Minimierung der Quadrate dieser Komponenten.

Die Ermittlung der Parameter \mathbf{q}_k ist eine lineare Operation, daher bleibt der Fehler normalverteilt mit der Kovarianzmatrix $(\sum_{i=1}^m B_{ki}^T C_{ki}^{-1} B_{ki})^{-1}$.

Zusätzlich zu den acht Parametern \mathbf{q}_k wird als neunte Messung die Zwangsbedingung $\|\mathbf{N}_k\| = 1$ mit verschwindender Fehlerkovarianz berücksichtigt. Der Meßvektor zum Zeitpunkt t_k lautet dann

$$\mathbf{z}_k = \begin{pmatrix} \mathbf{q}_k \\ \|\mathbf{N}_k\|^2 - 1 \end{pmatrix}, \quad (4.61)$$

und die entsprechende Jacobische Matrix findet sich im Anhang D.

Wie man aus den Meßgleichungen erkennen kann, ist die Messung eine nichtlineare Funktion der Zustandsgrößen. Präziser gesagt sind die Messungen bilinear bezüglich der Translationsgeschwindigkeit und der Normalen und linear bezüglich der Winkelgeschwindigkeit. Zum Prädiktionsschritt der rekursiven Schätzung verwenden wir die Gleichungen (4.21), wobei die Übergangsfunktion \mathbf{f}_k aus den rechten Seiten der Gleichungen (4.57), (4.54) und (4.55) besteht. Zum Aktualisierungsschritt werden wir drei Verfahren gegenüberstellen: den Erweiterten Kalman-Filter (EKF) der Gl. (4.23), den Modifizierten Gaußschen Filter zweiter Ordnung (MGSO) der Gl. (4.27) und den Iterierten Erweiterten Kalman-Filter (IEKF) in seiner Levenberg-Marquardt Version (4.43). Zur Verwendung des MGSO-Filters benötigen wir die Hesseschen Matrizen jeder der neun Komponenten des Meßvektors, die sich wegen der bilinearen Form der Meßfunktion aus der Jacobischen Matrix (D.3) ohne Aufwand herleiten lassen.

Wie wir im Abschnitt 2.3 erwähnt haben, besitzt das Problem der Ermittlung der Ebenenbewegung aus einem Bildpaar zwei Lösungen. Bei einer längeren Bildfolge bleibt diese Zweideutigkeit existent, wenn die jedem Bildpaar entsprechenden Translationen parallel sind. Unglücklicherweise wird unser Bewegungsmodell und jede dadurch interpretierte reale Situation von dieser Bedingung betroffen. Auf welche von beiden Lösungen man gelangen wird, hängt von dem Startwert ab. Dazu könnte jede - auch unsichere - a priori Information über die Anfangslage der Ebene bezüglich des Kamerakoordinatensystems ausgenutzt werden, was im Fall eines autonom geführten Fahrzeugs oder eines Roboterarms möglich ist.

Zur Verwendung der Gleichungen für die Verschiebungsraten wird unterstellt, daß wir wie in allen Kapiteln dieser Arbeit in einem normierten Kamerakoordinatensystem arbeiten, in welchem die Brennweite gleich eins ist und der Hauptpunkt im Ursprung liegt. Was in der Wirklichkeit gemessen wird, ist der optische Fluß an jeder Position in Bildpunkteinheiten. Um die Messung auf normierte Kamerakoordinaten zu übertragen, wird Kenntnis der internen Kameraparameter vorausgesetzt. Unter der Annahme, daß keine radiale Verzerrung auftritt und daß das Bildpunktkoordinatensystem rechtwinklig ist, lautet die Abbildung von Bildpunktkoordinaten (x_p, y_p) auf Kamerakoordinaten (x, y) wie folgt:

$$x = \frac{x_p - x_0}{s_x} \quad y = \frac{y_p - y_0}{s_y},$$

wobei (s_x, s_y) die Skalierungsfaktoren und (x_0, y_0) die Koordinaten des Hauptpunkts sind. Ist die Dimension eines quadratischen Bildes $N \times N$, so gilt:

$$s_x = \frac{N/2}{\tan \frac{\text{FOV}_x}{2}} \quad s_y = \frac{N/2}{\tan \frac{\text{FOV}_y}{2}},$$

wobei FOV_x und FOV_y die Winkel des Gesichtsfelds (Field Of View) jeweils in x - und y -Richtung sind. Außer diesen vier internen Parametern wird bei der Implementierung des vorgestellten Ansatzes keine andere Kalibriergröße verwendet.

Der rekursive Schätzer benötigt die Einstellung der Startwerte $\hat{\mathbf{s}}_0$ und der entsprechenden Kovarianz P_0^- sowie auch der Meß- und Prozeßrauschwerte.

Auswertung einer realen Grauwertbildfolge

Die hier zu auswertende monokulare Grauwertbildfolge wurde im Roboterlabor am Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB) aufgenommen. Ein Bild des Roboter manipulators und seiner Umgebung findet sich in Abb. 4.10, wo man die am Greifer befestigte Kamera und die Eichplatte erkennen kann. Die Bildfolge be-

Abbildung 4.10: Der Roboter manipulator am FhG-IITB (aus [Müller 92])

steht aus zwanzig Aufnahmen und wurde schrittweise aufgenommen. Bei jedem Schritt hat der Manipulator eine Bewegung entsprechend den sechs Stellgrößen unternommen, die von uns im voraus ausgerechnet wurden, so daß die erwünschte Bewegung (konstante Translations- und Winkelgeschwindigkeit) in Bezug auf das Kamerakoordinatensystem erreicht wird. Die dazu benötigten Koordinatentransformationen werden im Anhang E beschrieben. Die Stellgrößen in dem Experiment wurden so eingestellt, daß eine bezüglich des Kamerakoordinatensystems rein translatorische Bewegung mit $\mathbf{v} = (0, 0, 20)$ mm erfolgt. Ob der Roboter tatsächlich die erwünschte Bewegung durchgeführt hat, hängt von zwei Faktoren ab: Erstens von der Unsicherheit in der relativen Rotation zwischen dem Kamerakoordinatensystem und dem Greiferkoordinatensystem

und zweitens von der Präzision, mit der der Roboter die Befehle durchführt. Wegen der reinen Translation bleibt die Normale der Ebene bezüglich des Kamerakoordinatensystems zeitlich konstant. Die ungefähren Werte $(-0.2, 0.2, 0.95)$ der Normalenkomponenten wurden mittels Kalibrierung nach [Tsai 87] mit Hilfe nur einer Ebene an der ersten Aufnahme ermittelt. Die Bezeichnung „ungefähr“ bezieht sich auf bekannt inhärente Ungenauigkeiten des verwendeten Verfahrens. Aus den Komponenten der Normalen erkennt man, daß es sich um eine – zumindest im Fall des Bildpaars – fehlerempfindliche Bewegungs- und Geometrieconfiguration handelt, weil der Winkel zwischen \mathbf{N} und \mathbf{v} niedrig ist.

Aus der aufgenommenen Bildfolge wurden Kantenelemente nach [Korn 88] extrahiert, zu geradlinigen Kantensegmenten nach [Burns *et al.* 86] gruppiert und deren Schnittpunkte nach [Gengenbach 90] ermittelt. Diese Vorverarbeitungsschritte wurden automatisch nach der Aufnahme von dem Bildauswertesystem [Gengenbach 90] durchgeführt. Die Grauwertbilder und die detektierten Kantensegmente aus den Aufnahmen zu den Zeitpunkten t_0, t_5, t_{10} und t_{15} werden in der Abb. 4.11 gezeigt.

Ziel unserer Untersuchung ist die Ermittlung der Fehlerverstärkung, die beim Auswertungsschritt der Bewegungs- und Strukturermittlung auftritt. Daher ist es wünschenswert, daß der Fehler in den Verschiebungsraten nur auf die Ungenauigkeit in der Ermittlung der Schnittpunkte und nicht auf falsche Zuordnungen zurückzuführen ist. Deshalb wurden die zeitlichen Zuordnungen der Schnittpunkte modellbasiert unter Verwendung der a priori Information über ihre Anordnung auf der Eichplatte ermittelt. Die dadurch ermittelten Verschiebungsraten für die Aufnahmen der Abb. 4.11 werden in der Abb. 4.12 gezeigt.

Der erste Eindruck aus Abb. 4.12 ist, daß es sich um eine rein translatorische Bewegung handelt, deren Expansionspunkt ungefähr die Bildpunktkoordinaten $(260, 360)$ besitzt. Der Hauptpunkt hat die Koordinaten $(x_0, y_0) = (262, 267)$, daher besitzt die Translationsgeschwindigkeit eine vernachlässigbare x -Komponente. Unter Berücksichtigung des Skalierungsfaktors ($s_y = 1100$) kommt man auf einen Winkel zwischen dem \mathbf{v} -Vektor und der z -Achse von $\arctan(93/1100) \approx 5^\circ$ Grad. Die Diskrepanz zu dem eingestellten Wert, der einer reinen Translation in z -Richtung entspricht, ist auf die Ungenauigkeit der Hand-Auge-Kalibrierung (siehe Anhang E) zurückzuführen. Diese Tatsache muß man in den folgenden Diagrammen berücksichtigen, wo als tatsächlicher Wert die Translationsgeschwindigkeit in z -Richtung aufgezeichnet ist.

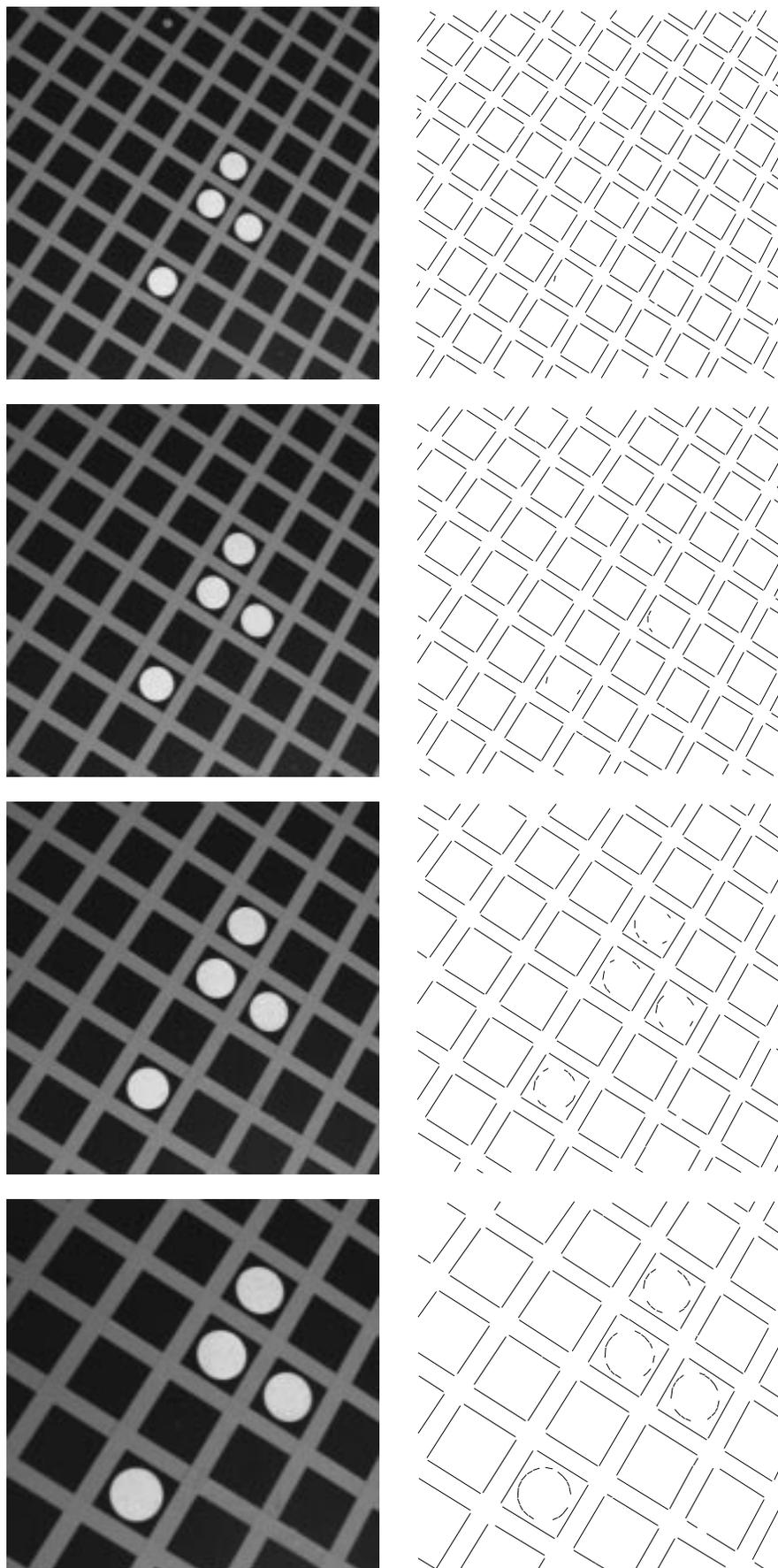
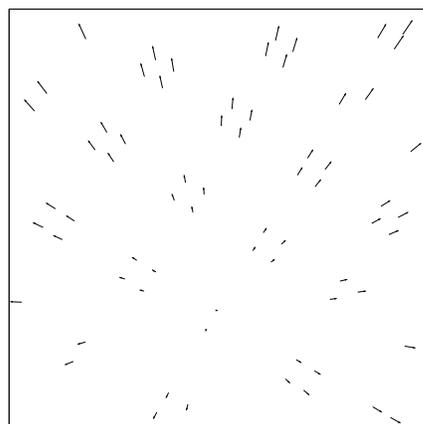
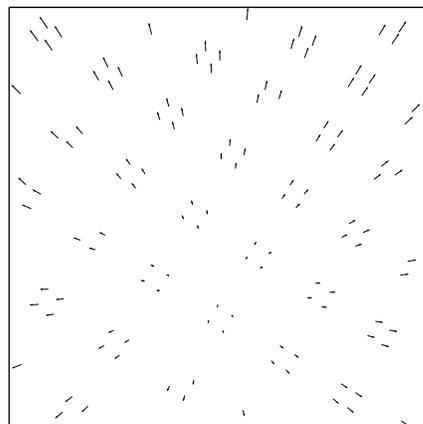
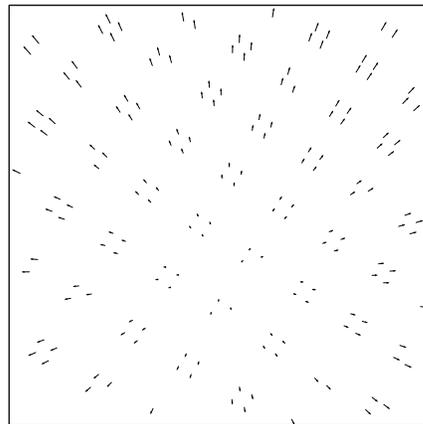
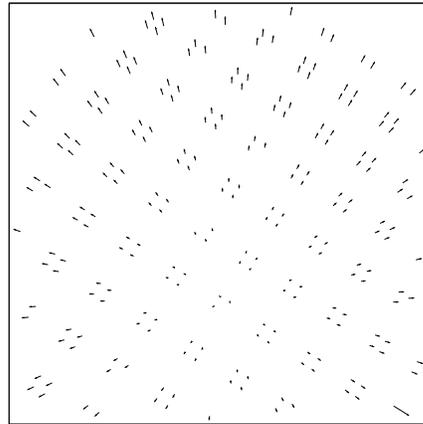


Abbildung 4.11: Aufnahmen zu Zeitpunkten t_0, t_5, t_{10} und t_{15} (links) und die extrahier-



Zuerst führen wir das Experiment unter der Zwangsbedingung durch, daß die Bewegung keine rotatorischen Anteile besitzt. Mathematisch wird die Zwangsbedingung durch Verschwinden der Anfangsfehlerkovarianz für ω_0 modelliert, d.h. der Schätzer ist absolut sicher über das Verschwinden von ω . Für den Aktualisierungsschritt werden alle drei Schätzer (EKF, MGSO und IEKF) ausprobiert. Das Prozeßrauschen wird auf Null gesetzt. Zusätzlich wird eine Standard-Methode zur Schätzung der Bewegung einer Ebene aus einem Bildpaar (siehe Abschnitt 2.3) ausgeführt, die wir mit „NI“ (nicht-inkrementell) bezeichnen. Der Verlauf der geschätzten Werte zusammen mit den tatsächlichen Werten, die mit „GT“ (ground-truth) bezeichnet werden, wird in den Abb. 4.14-4.15 dargestellt.

Der Polarwinkel eines Vektors in Bezug auf das Kamerakoordinatensystem ist der Winkel zwischen dem Vektor und der optischen Achse. Zur Veranschaulichung des Azimutwinkels in Bezug auf das Kamerakoordinatensystem dient die Skizze in der Abb. 4.13.

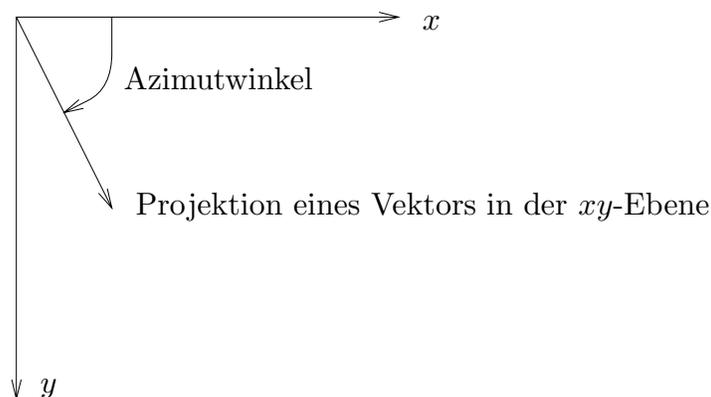


Abbildung 4.13: Der Azimutwinkel eines Vektors in Bezug auf das Kamerakordinatensystem, dessen Achsen wie die Achsen des Bildpunktkoordinatensystems ausgerichtet sind.

Aus dem geschätzten Azimutwinkel ($\approx 90^\circ$ Grad) in Abb. (4.14a), dem eine Translation in der YZ Ebene entspricht, bestätigt sich unsere Vermutung, daß die Translationsgeschwindigkeit eine y -Komponente besitzt. Der vermutete Polarwinkel stimmt mit dem in Abb. 4.14b gezeichneten überein. Alle drei rekursiven Schätzer weisen dasselbe Verhalten auf, im Gegensatz zum nicht inkrementellen Standardverfahren, das instabil ist.

Bei der Schätzung des Azimutwinkels der Normalen in Abb. 4.15 beobachten wir eine breite Variation (ca. 50° Grad) zwischen den geschätzten Werten, was auch darauf zurückzuführen ist, daß diese x - und y -Komponenten der Normalen sehr klein im Vergleich zur z -Komponente sind. Dagegen ist die Abweichung vom tatsächlichen Wert bei der Schätzung des Polarwinkels der Normalen mit ca. $6 - 8^\circ$ Grad niedriger. Der relative Fehler bei der Schätzung der inversen Kollisionszeit liegt zwischen ca. 15% und 20%. Bei dieser Schätzung sowie auch bei der Schätzung des Polarwinkels der Normalen

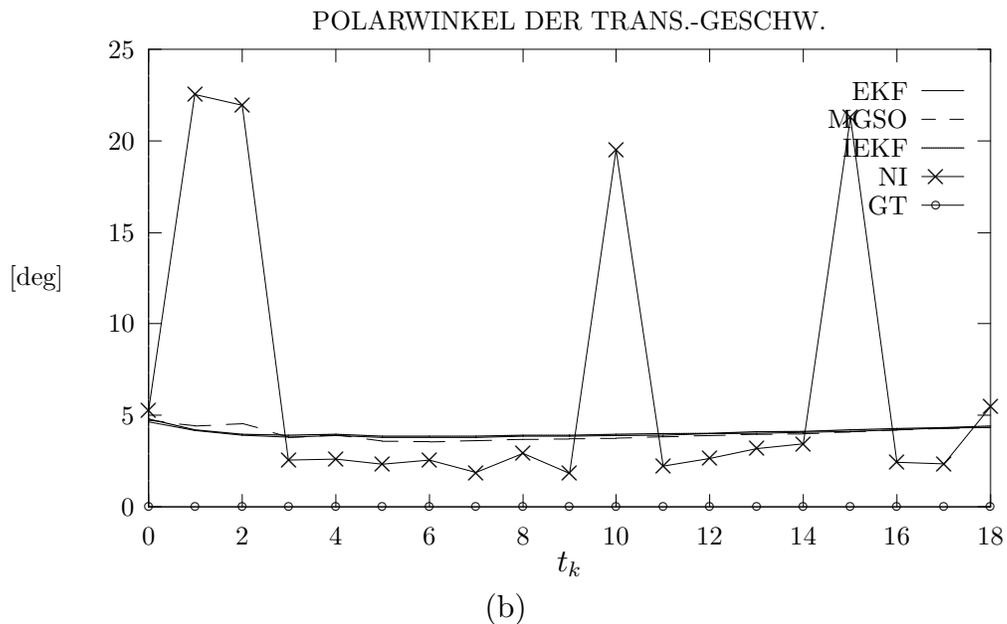
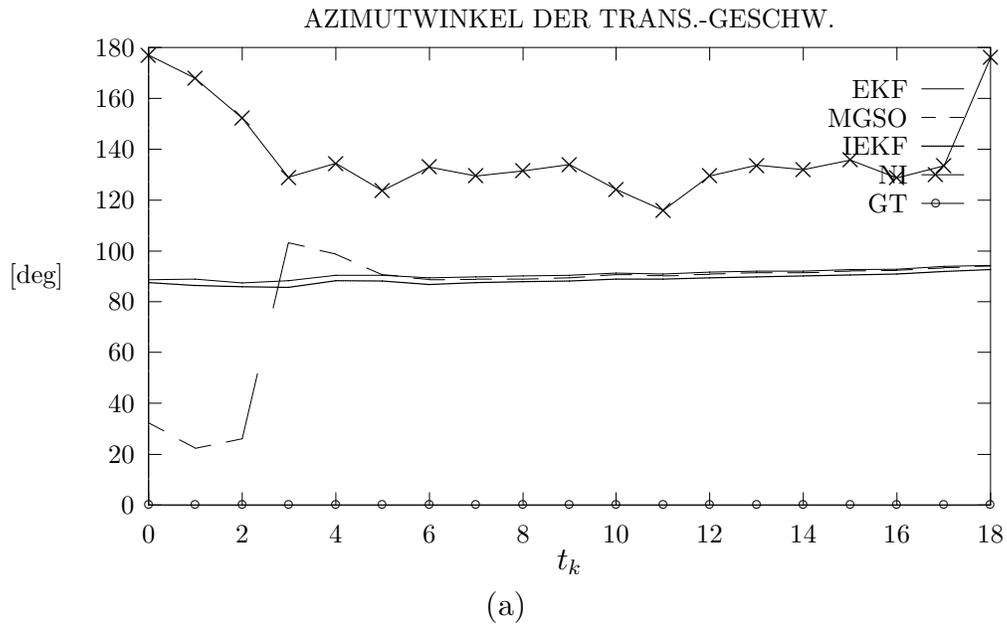


Abbildung 4.14: Durch vier verschiedene Verfahren (EKF=Erweiterter Kalman-Filter, MGSO=Modifizierter Gaußscher Filter zweiter Ordnung, IEKF= Iterierter Erweiterter Kalman-Filter, NI = Nicht-inkrementelles Verfahren) ermittelte Schätzwerte für den Azimutwinkel (a) und den Polarwinkel (b) der Translationsgeschwindigkeit unter der Annahme einer rein translatorischen Bewegung im Vergleich zum Verlauf der tatsächlichen (GT) Werte.

weist der MGSO-Filter den niedrigsten Fehler auf. Die nicht-inkrementelle Methode ist nur bei der Schätzung der Kollisionszeit stabil.

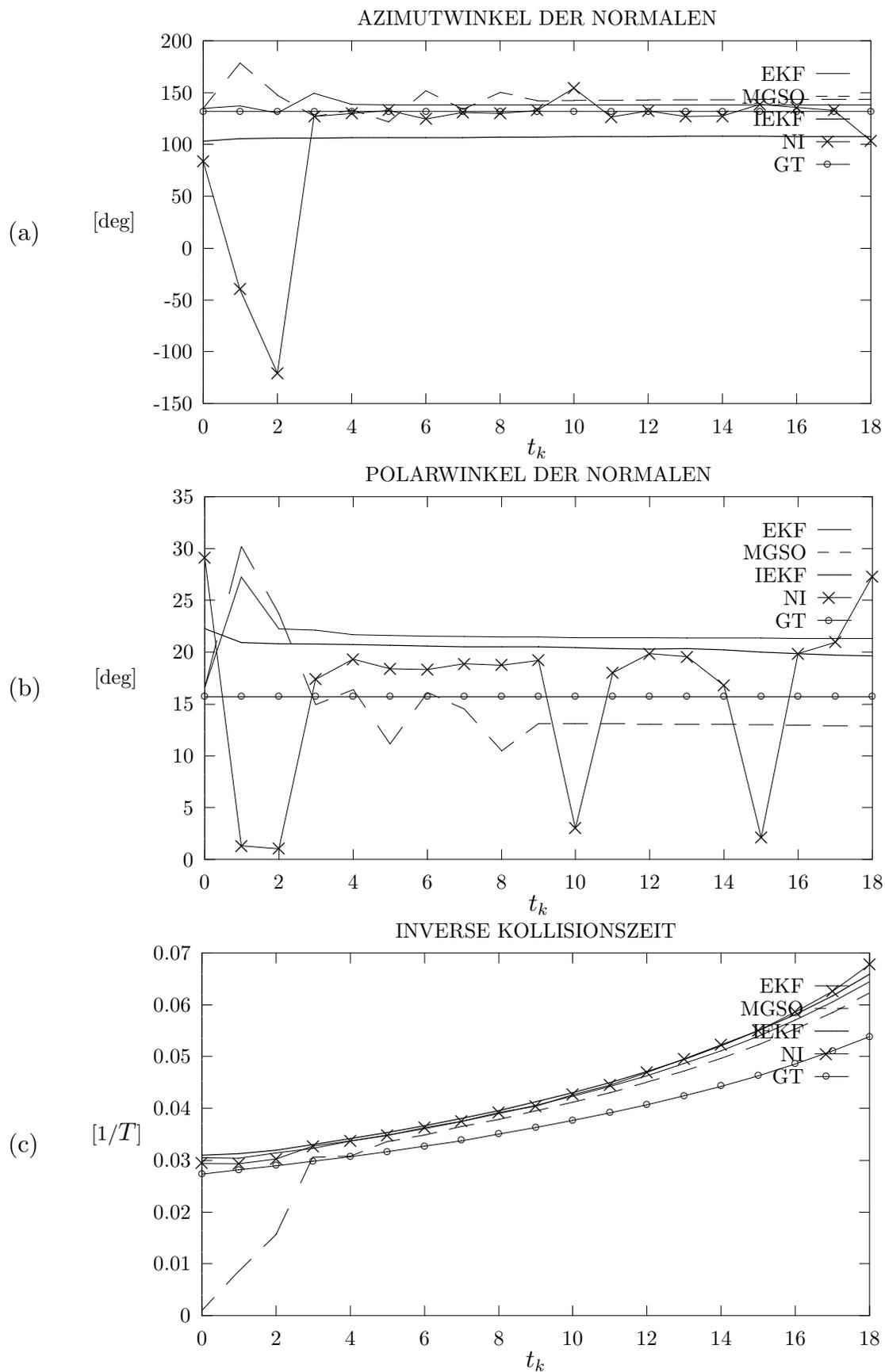
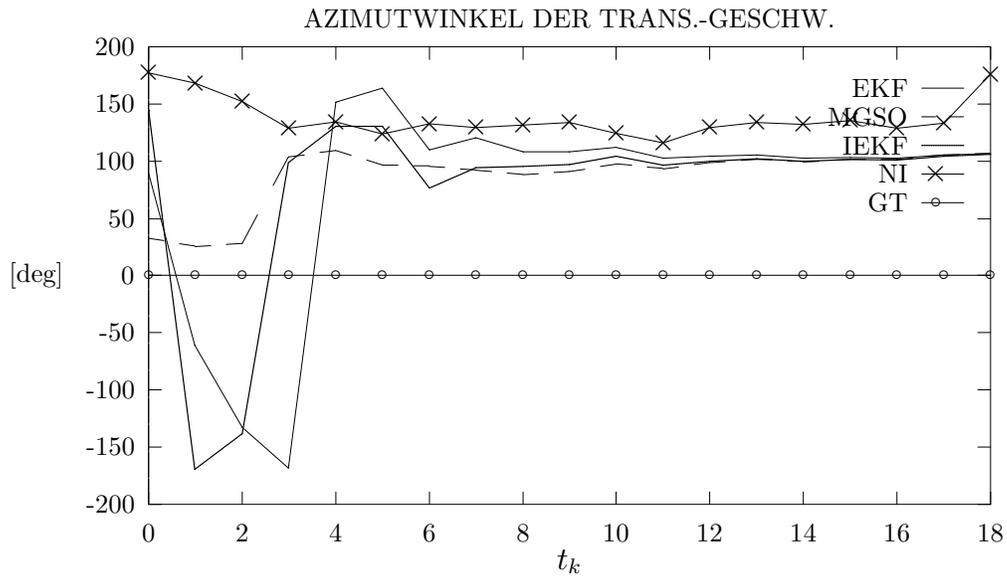


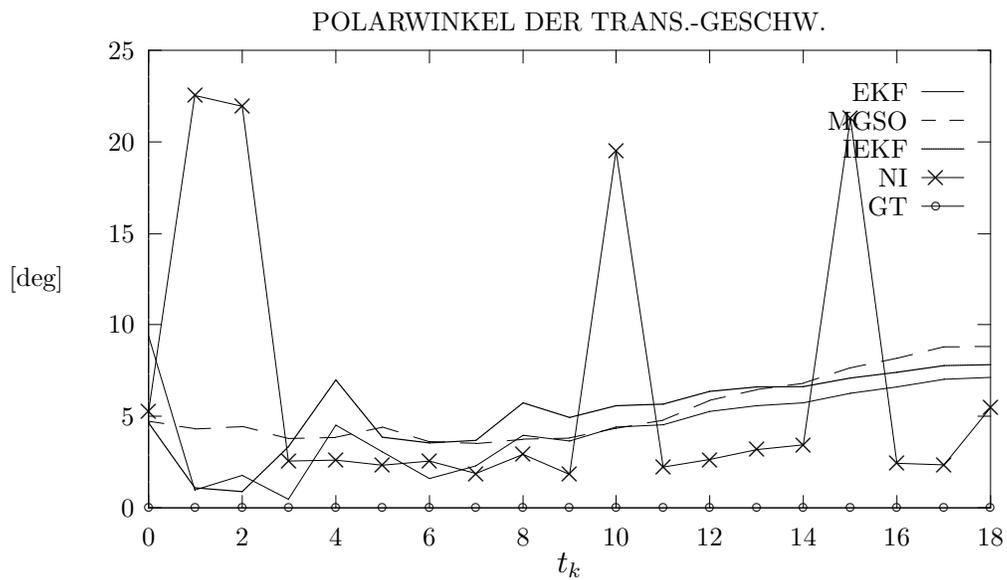
Abbildung 4.15: Schätzwerte für den Azimutwinkel (a) und den Polarwinkel (b) der Ebenennormalen und der inversen Kollisionszeit (c) unter der Annahme einer rein trans-

Im nächsten Lauf des Schätzverfahrens erlauben wir die Existenz einer Winkelgeschwindigkeit, und der Sachverhalt wird komplexer. Aus den Schätzwerten des Azimut- und Polarwinkels der Translationsgeschwindigkeit in Abb. 4.16 erkennen wir eine bei allen Filtern auftretende Verzerrung, die eine Wanderung des Expansionspunktes nach links und unten verursacht (zweiter Quadrant im Kamerakordinatensystem in der Abb. 4.13). Es ist zu bemerken, daß eine Verzerrung im Polarwinkel von 2° Grad eine Verschiebung des Expansionspunkts um $\text{ca. } \tan(2^\circ) \cdot 100 \approx 38$ Bildpunkte verursacht.

Die Erklärung für diese Verzerrung findet sich in den Verläufen der Schätzwerte für die Komponenten der Winkelgeschwindigkeit (Abb. 4.18). Die größte Verzerrung – der Leser sei hier auf die Skalierung der Ordinate in den Diagrammen der Abb. 4.18 hingewiesen – mit $\text{ca. } 0.002 \text{ rad}/T$ tritt bei ω_x auf, was aus der analytischen Fehleruntersuchung des Abschnitts 3.4 zu erwarten ist: nur die Summe $-v_y + \omega_x$ kann robust berechnet werden. Entsprechend, aber niedriger ist der Beitrag von ω_y zur Kompensierung der v_x -Komponente. Allerdings haben wir erhofft, daß sich die Anfälligkeit für eine Verwechslung zwischen Rotation und Translation entlang einer längeren Bildfolge abschwächen würde. Immerhin weist aber jeder rekursive Schätzer ein stabileres Verhalten im Vergleich zum nicht-inkrementellen Verfahren auf. Bemerkenswert ist, daß die Schätzung der Normalen von den obgenannten Effekten nicht beeinträchtigt wird. Der Fehler im geschätzten Polarwinkel liegt zwischen 3° und 5° Grad, und das ist von Bedeutung für die Praxis, weil dieser Winkel die Anfahrriichtung eines Sauggreifers auf eine zu greifende Ebene angibt.



(a)



(b)

Abbildung 4.16: Schätzwerte für den Azimutwinkel (a) und den Polarwinkel (b) der Translationsgeschwindigkeit.

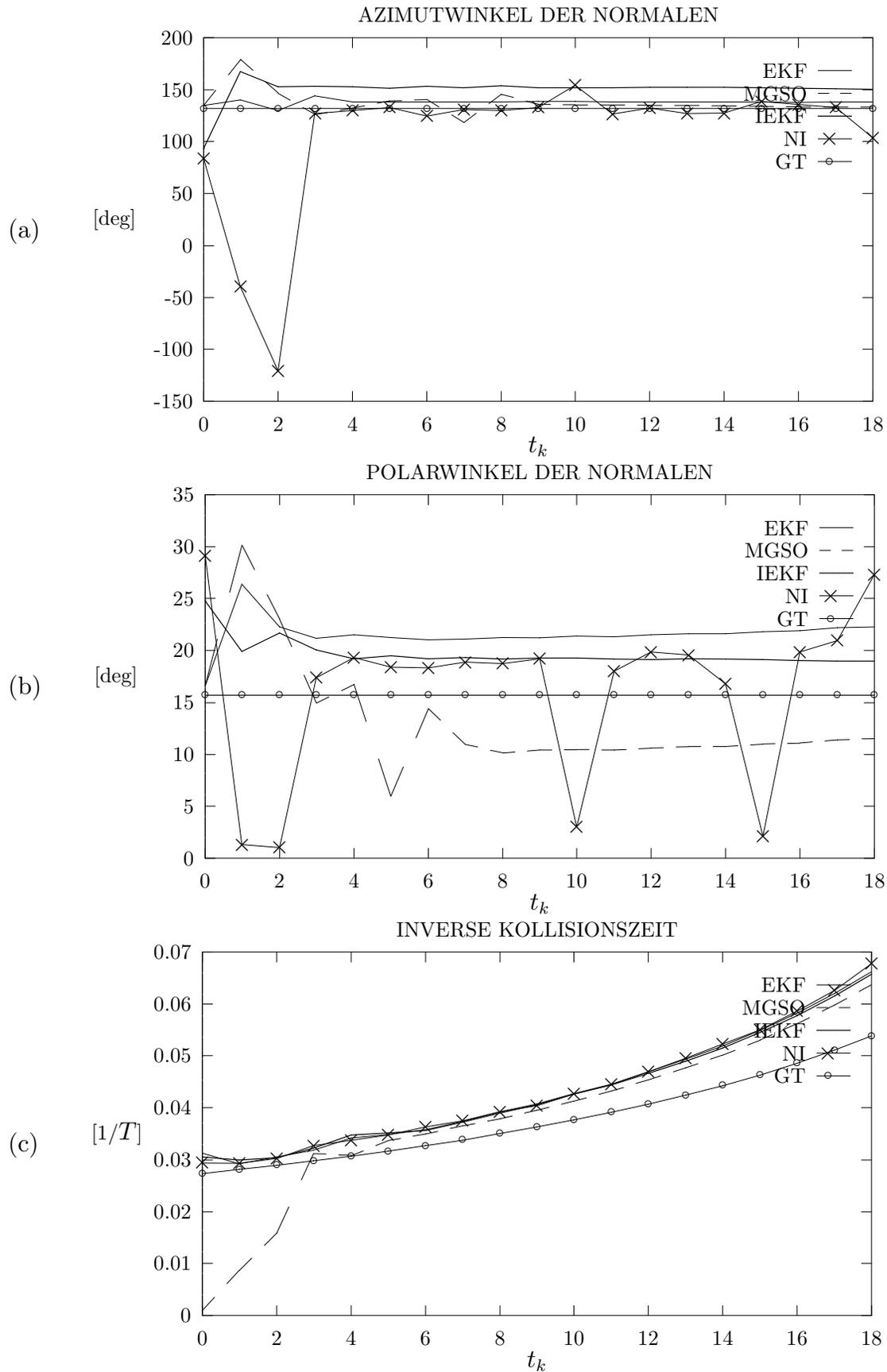


Abbildung 4.17: Schätzwerte für den Azimutwinkel (a), den Polarwinkel (b) der Ebenennormalen und der inversen Kollisionszeit (c).

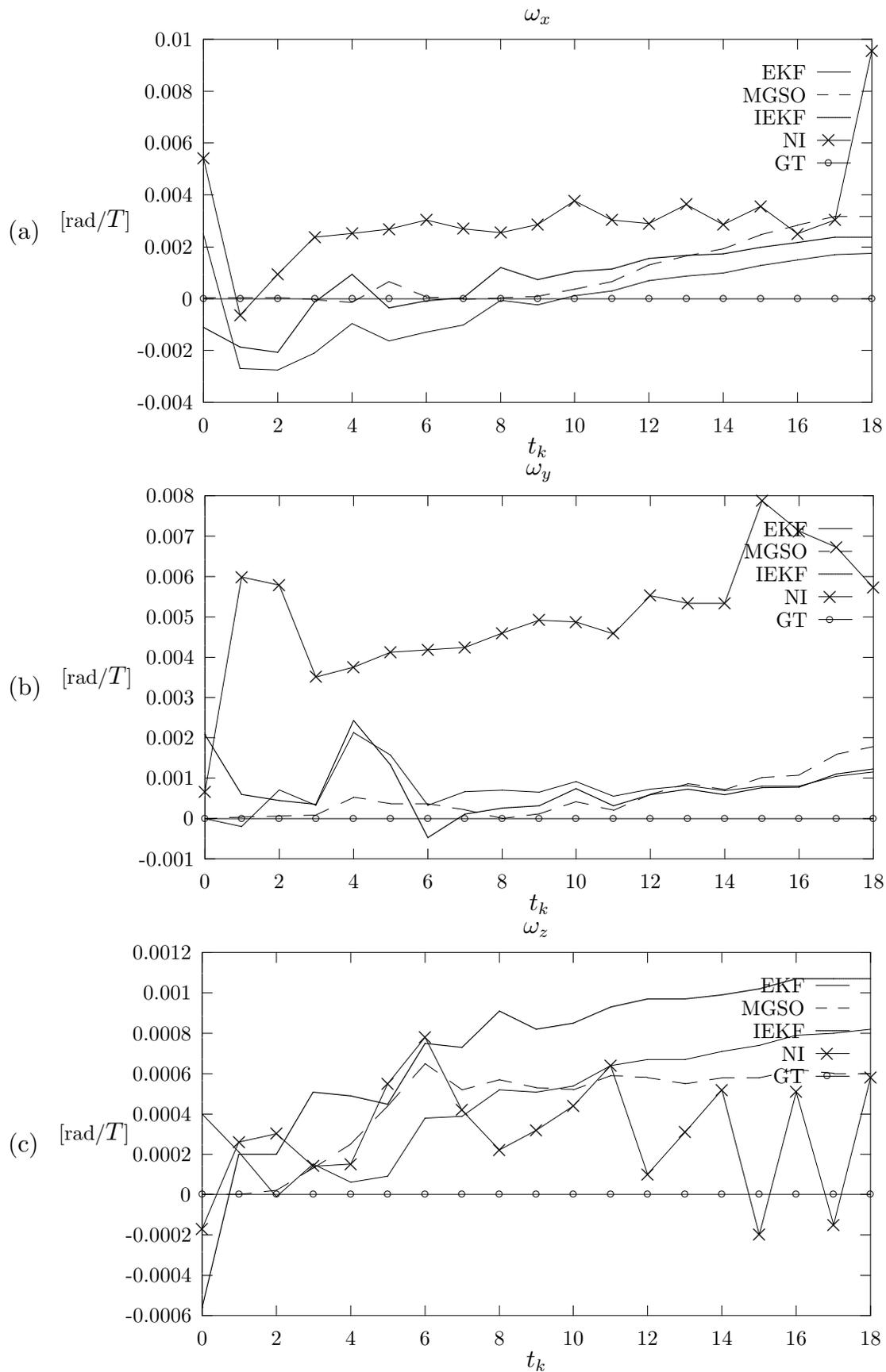


Abbildung 4.18: Schätzwerte für die drei Komponenten der Winkelgeschwindigkeit

Experimente mit synthetischen Daten

Zur Simulation von mehreren Bewegungs- und Geometrieconfigurationen verwenden wir ein Modell der im Experiment mit realen Bilddaten benutzten Eichplatte und die gleichen Hand-Auge und Roboter-Welt Transformationen sowie auch dieselben internen Kameraparameter. Wenn nicht darauf hingewiesen wird, wird ein gleichverteiltes Meßrauschen von ± 0.5 Bildpunkten angenommen und kein Prozeßrauschen verwendet. Die letzte Annahme wurde gemacht, damit wir vollständig den Einsatz des a priori Wissens über die Glattheit der Bewegung ausnutzen. Die Startwerte sind immer Null für die Geschwindigkeiten, und der Startwert für die Normale liegt in der YZ -Ebene und hat einen Neigungswinkel von 45° Grad. Es hat sich herausgestellt, daß das Hochsetzen auf sehr große Werte der Startkovarianz, wie es bei anderen Ansätzen vorgeschlagen wird, auf Divergenz in allen drei verwendeten Filtern führt. Außerdem besitzt man a priori Wissen über den Umfang der Geschwindigkeiten: Die durch den Abstand skalierte Geschwindigkeit \mathbf{u} und die Winkelgeschwindigkeit $\boldsymbol{\omega}$ können nicht beliebig groß werden, weil die Länge der Verschiebungsrates mitwächst. Angesichts der Tatsache, daß bei dem vorhandenen Skalierungsfaktor eine zur Bildebene parallele Winkelgeschwindigkeit von 0.01 [rad/ T] eine Verschiebungsrates von 10 Bildpunkten verursacht – entsprechendes gilt für die Translation –, beschränken wir die Startkovarianz für die skalierte Translationsgeschwindigkeit \mathbf{u} auf 10^{-4} [$1/T$] 2 und für die Winkelgeschwindigkeit $\boldsymbol{\omega}$ auf 10^{-5} [rad/ T] 2 .

Als erstes untersuchen wir den Einfluß der Größe des effektiven Gesichtsfelds. Als effektives Gesichtsfeld bezeichnen wir das Gebiet in der Bildebene, das von der Projektion der sich relativ zur Kamera bewegenden Ebene abgedeckt wird. In diesem Experiment werden wir die Größe des effektiven Gesichtsfelds mit der Anzahl der Quadrate modellieren (siehe Abb. 4.11), die 30×30 mm 2 groß sind und einen gegenseitigen Abstand von 10mm haben.

Der IEKF-Schätzer wird auf fünf verschiedene Größen des effektiven Gesichtsfelds angewandt. Diese Größen, interpretiert in Brennweiteeinheiten (vgl. Längen α und β in den analytischen Fehleruntersuchungen des dritten Kapitels), entsprechen einem Spektrum von 0.1 bis 0.5 Brennweiteeinheiten oder einem Öffnungswinkel von ca. 6° bis 28° Grad. Die Bewegung ist eine reine Translation (0, 2, 6)mm und am Anfang der Bewegung liegt die Ebene im Abstand 500mm von Projektionszentrum und ist 60° bezüglich der optischen Achse geneigt.

In den folgenden Diagrammen werden für die Translationsgeschwindigkeit und die Normale der Fehlerwinkel zwischen der geschätzten und der tatsächlichen Richtung aufgezeichnet (Abb. 4.19a-b). In der Abb. 4.20 wird der relative Fehler in der Schätzung der inversen Kollisionszeit und in der Abb. 4.19c der absolute Fehler in der x -Komponente der Winkelgeschwindigkeit aufgezeichnet.

Der Fehlerverlauf ist grob monoton fallend beim Anwachsen des Gesichtsfelds. Die Schätzwerte der Translation und der Normalen für kleine Größen des Gesichtsfelds sind unakzeptabel. Ein analoger Effekt wie beim Experiment mit den Realdaten tritt besonders bei kleinem Gesichtsfeld auf, wobei der Schätzer zwischen x -Komponente der Winkelgeschwindigkeit und y -Komponente der Translationsgeschwindigkeit schwer unterscheiden kann. Der relative Fehler in der inversen Kollisionszeit liegt zwischen ca. 0.1% und 11%.

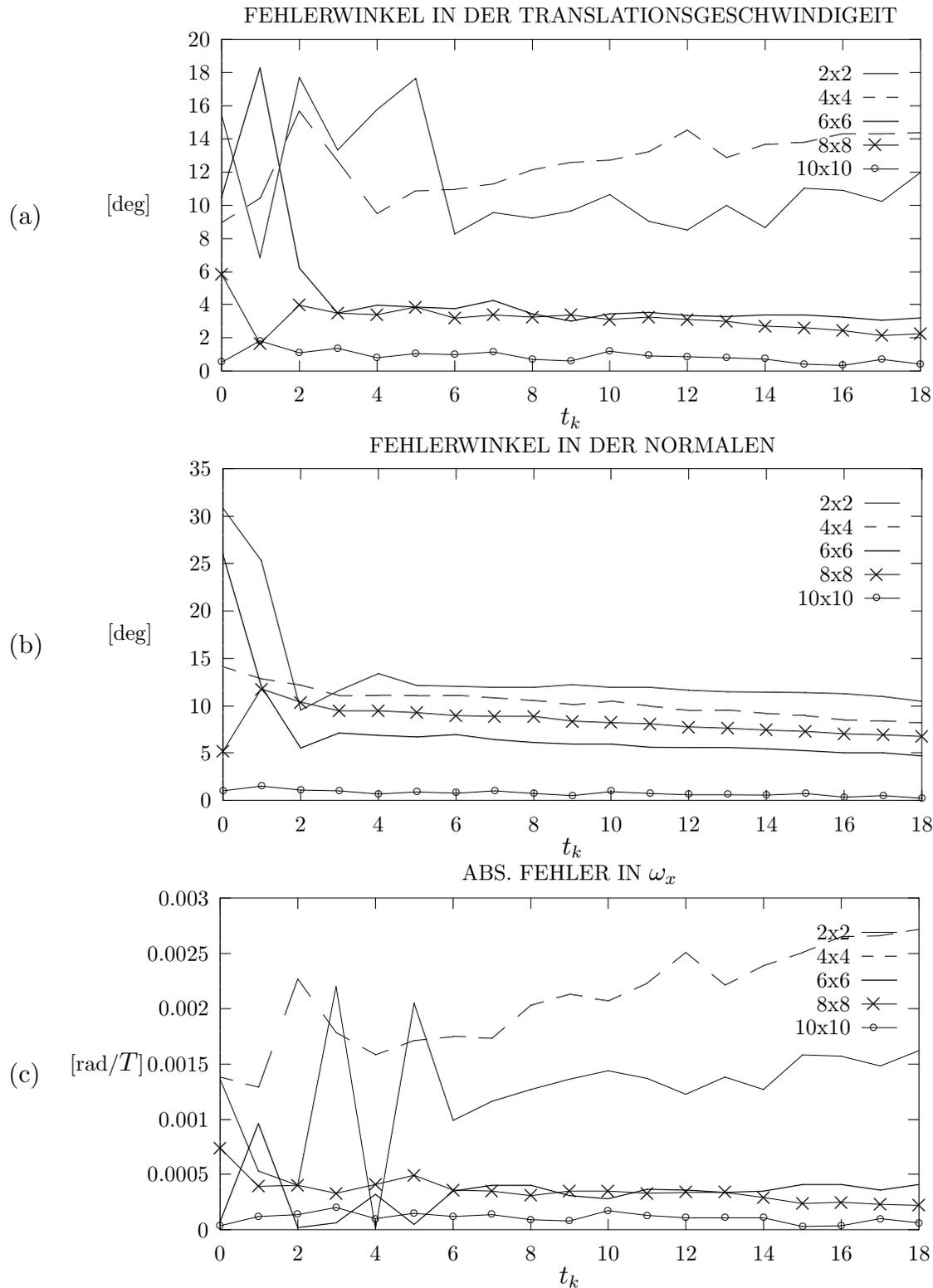


Abbildung 4.19: Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b) und absoluter Fehler in ω_x (c) bei variierender Größe des effektiven Gesichtsfelds

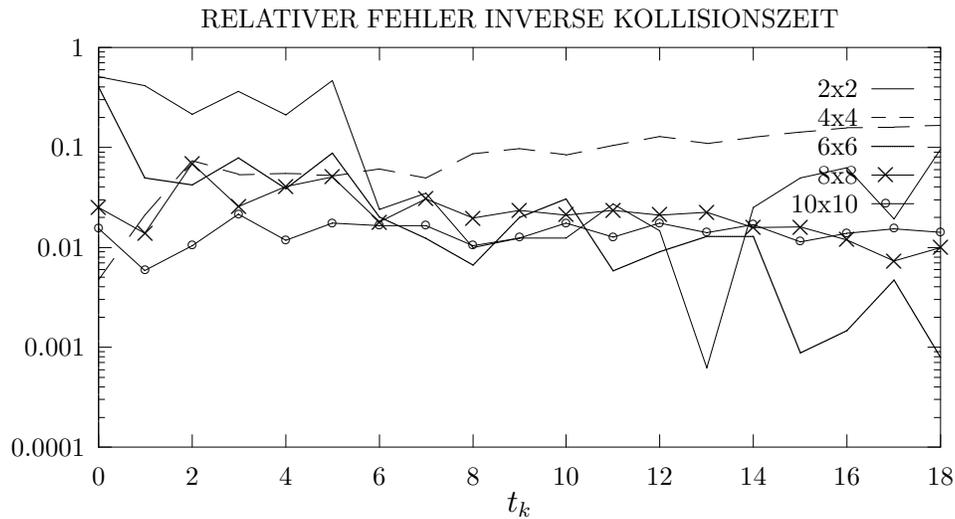


Abbildung 4.20: Logarithmus des relativen Fehlers in der vom IEKF geschätzten inversen Kollisionszeit bei variierender Größe des effektiven Gesichtsfelds

Beim nächsten Experiment variieren wir die Richtung der Translationsgeschwindigkeit. Wir behalten „6x6“ als Größe des effektiven Gesichtsfelds, und die optische Achse bildet mit der Ebenennormalen einen Winkel von 30° Grad. Wir zeichnen dieselben Fehler wie oben für fünf verschiedene Bewegungen m_i mit entsprechenden Translationsgeschwindigkeiten $(0, i, 6)$ mm (Abb. 4.21-4.22) auf. Der Vektor der Translationsgeschwindigkeit wandert in der YZ -Ebene von einer zur optischen Achse parallelen Lage zu einer zur Ebenennormalen parallelen Lage. Der Zusammenhang zwischen dem Fehler in Translationsrichtung und dem Fehler in ω_x ist eindeutig zu erkennen: Beide Fehler wachsen mit zunehmender Abweichung der Translationsrichtung von der optischen Achse. Langfristig unbeeinflusst von der Richtung der Translation und mit hohem, aber allmählich sinkendem Fehler läuft die Schätzung der Normalen. Der relative Fehler in der inversen Kollisionszeit zeigt nur gegen Ende der Bildfolge denselben Zusammenhang zu der Abweichung der Translation von der optischen Achse wie die anderen Fehler und variiert zwischen ca. 0.5% und 8%.

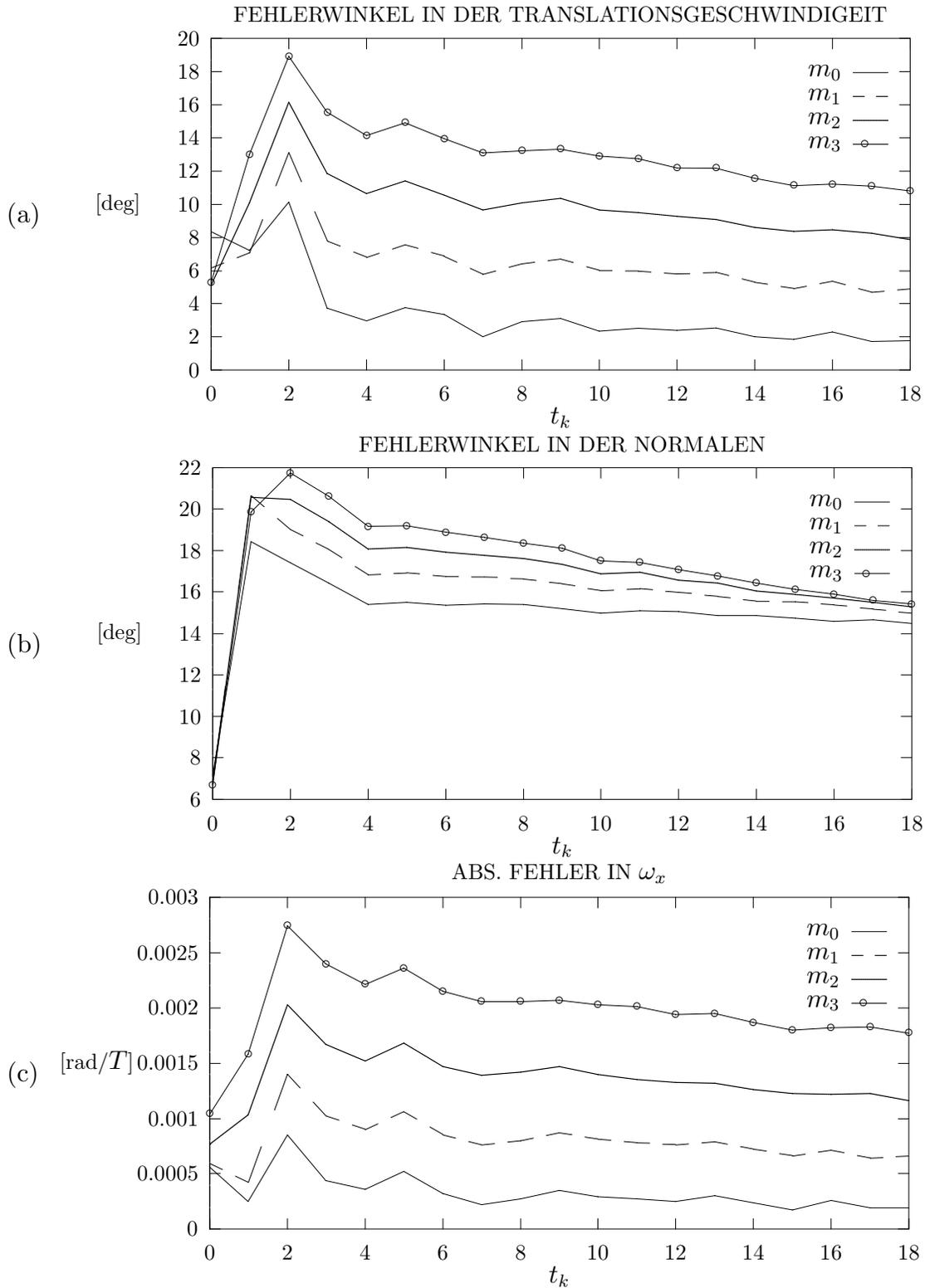


Abbildung 4.21: Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b), absoluter Fehler in ω_x (c) bei variierender Translationsrichtung

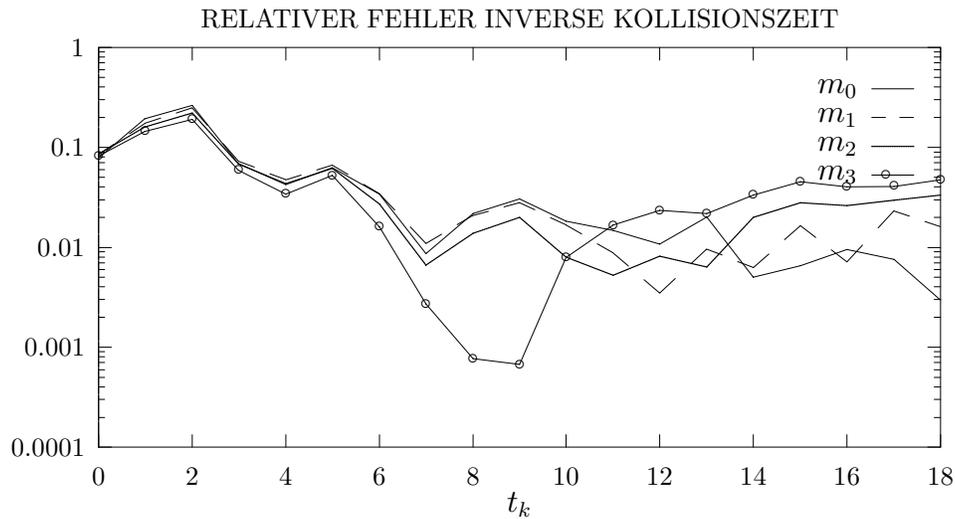


Abbildung 4.22: Logarithmus des relativen Fehlers in der vom IEKF geschätzten inversen Kollisionszeit bei variierender Translationsrichtung

Beim letzten Experiment setzen wir die Translation parallel zur Bildebene in x -Richtung und lassen die Ebenennormale in der YZ -Ebene variieren. Wir zeichnen den Fehler für drei Lagen der Normale (0° , 30° , 60°) bezüglich der optischen Achse auf (Abb. 4.23), die wir jeweils mit n_0 , n_1 und n_2 bezeichnen. Die Kopplung erscheint in diesem Fall zwischen dem Fehler in der Normalenrichtung und dem Fehler in ω_y . Beide Fehler sind niedriger, je größer die relative Neigung der Ebene ist. Der Fehler in der Translationsrichtung ändert sich nicht monoton mit der Neigung der Ebene. Die inverse Kollisionszeit wird nicht gezeigt, weil sich die Kamera parallel zur Ebene bewegt.

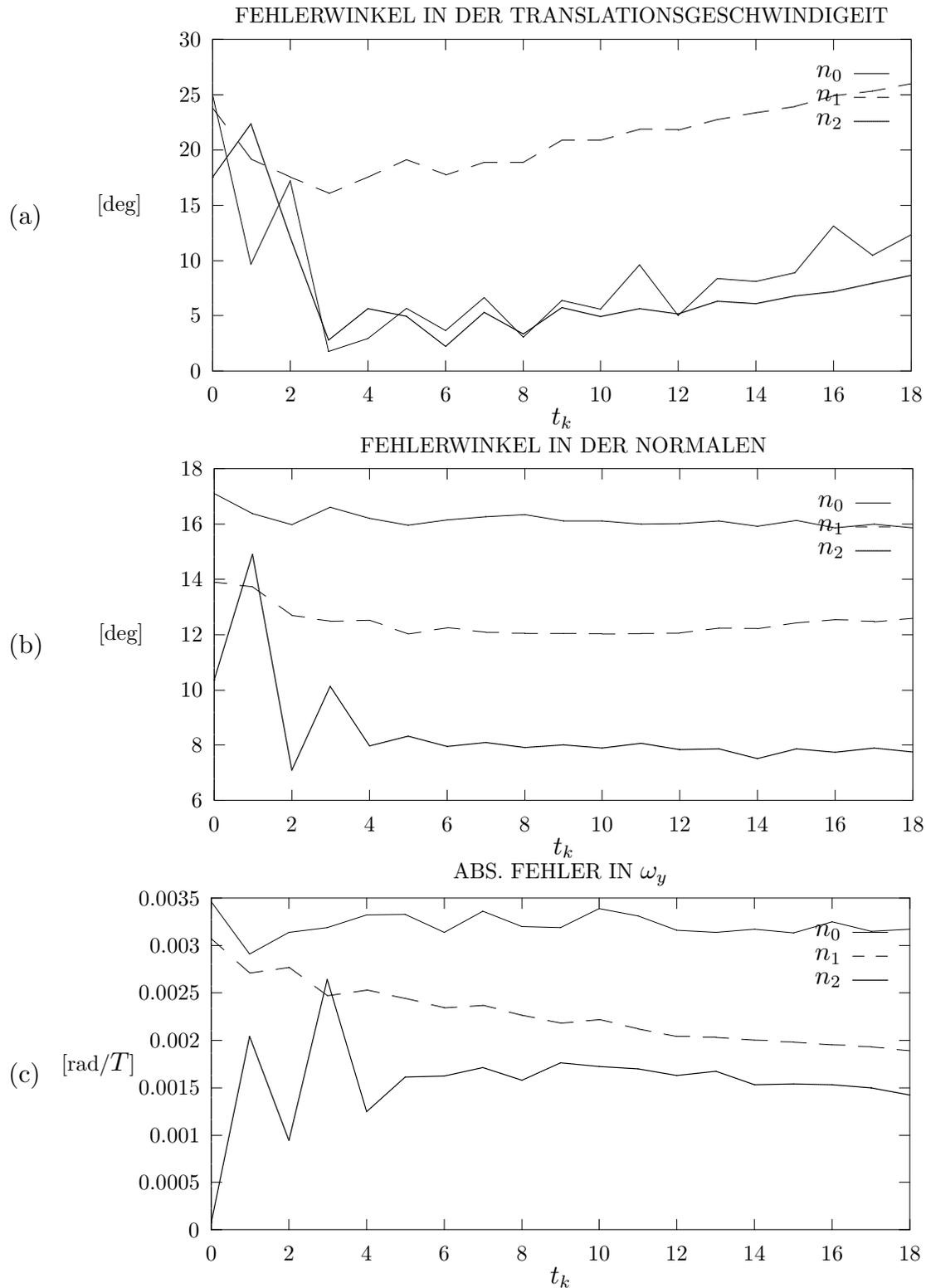


Abbildung 4.23: Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (a), in der Ebenennormalen (b), absoluter Fehler in ω_y (c) bei variierender Richtung der Ebenennormalen.

In allen Experimenten benötigt der IEKF nach den ersten Zeitpunkten nur einen bis fünf Iterationsschritte bis zur Konvergenz. Die die Schätzwerte begleitende Fehlerkovarianz ist in allen Experimenten niedriger als der tatsächliche Fehler. Durch Einführung von Prozeßrauschen wird diese Diskrepanz abgeschwächt. Dennoch, je höher das Prozeßrauschen wird, desto mehr ähnelt das Verhalten des rekursiven Schätzers dem Verhalten eines nicht-inkrementellen Verfahrens. Aus Platzgründen wurde der Verlauf des nicht-inkrementellen Schätzers nur bei dem Experiment mit den realen Bilddaten aufgezeichnet.

Aus den synthetischen Experimenten schließen wir, daß der beste rekursive Schätzer - der IEKF - ein stabiles Verhalten aufweist, das aber immer noch eine Verzerrung enthält, die von der Kopplung der Translation und der Rotation und von der relativen Lage der Ebenennormalen und der Translationsgeschwindigkeit abhängt. Wegen dieser Verzerrung, die dem Problem inhärent ist, müssen zusätzliche Bildinformationen eingeführt und /oder komplexere Bewegungen modelliert werden. Schon die Tatsache, daß die Dualität der Lösung bei konstanter Richtung der Translationsgeschwindigkeit erhalten bleibt, weist darauf hin, daß bei einem Bewegungsmodell mit variierender Translationsrichtung oder bei Einführung der Stellgrößen des Roboters in den Prädiktionsschritt die vom Winkel zwischen Normalen und Translation abhängige Fehlerempfindlichkeit abgeschwächt werden kann.

Sind mehrere Ebenen eines polyedrischen Objektes vorhanden, so kann man durch leichte Modifikation des Verfahrens die Hilfsparameter \mathbf{q}_k aller Ebenen zu einem Meßvektor zusammenfassen. Durch den neuen Beweis, den wir im Abschnitt 2.6 beschrieben haben, daß die Bewegungsschätzung auch im Fall der Verschiebungsraten von Geraden mittels Hilfsparameter geschätzt werden kann, öffnet sich ein neuer Weg zur Schätzung der relativen Bewegung von polyedrischen Objekten durch die Kombination von punktförmigen und geradlinigen Merkmalen.

5 Zusammenfassung und Ausblick

Zur Führung von autonom mobilen Fahrzeugen muß man aus bildgebenden Sensoren die Struktur der Umgebung und die relativen Bewegungen der Szenekomponenten zur Kamera ermitteln. Die Durchführung dieser Aufgabe ist durch die Auswertung von monokularen Grauwertbildfolgen unter ganz allgemeinen Annahmen (wie Starrheit der Szenekomponenten) im Prinzip möglich. Jedoch leidet eine Lösung dieses Problems unter Mehrdeutigkeiten. Existierende Algorithmen weisen eine hohe Empfindlichkeit gegen Meßrauschen auf, was deren Einsatz bei Realweltsituationen zur Zeit verbietet.

Untersuchungsgegenstand unserer Arbeit war die Ermittlung der 3D-Bewegungsparameter aus den in einer früheren Verarbeitungsstufe ermittelten Merkmalskorrespondenzen oder Verschiebungsraten. Unser erstes Ziel ist die Erläuterung der Fehlerempfindlichkeit gewesen. Wir haben zuerst erschöpfend die existierenden Verfahren zur Bewegungsberechnung aus einem Bildpaar mit einheitlicher Notation sowie die Ergebnisse der bisherigen Fehleranalysen beschrieben.

Über den berichteten Stand des Wissens sind wir an folgenden Punkten hinausgegangen:

1. Wir haben analytisch nachgewiesen, daß bei der Verwendung einer vereinfachten Fehlerfunktion der Fehler in der geschätzten Translationsrichtung von dem Winkel zwischen Blick- und Translationsrichtung, von der Größe des Gesichtsfelds und von dem Verhältnis zwischen Translationsbetrag und Objektstand abhängt. Für diesen Sachverhalt gab es bis jetzt nur experimentelle Nachweise.
2. Wir haben eine Beziehung zwischen den gefährlichen Flächen und der Instabilität in der Bewegungsschätzung nachgewiesen, die bis jetzt eine offene Frage in der Bildauswertung darstellte.
3. Wir haben analytisch untersucht, wie die Fehlerempfindlichkeit in der Ermittlung der Bewegung einer Ebene von der Form der Bewegung und der Neigung der Ebene abhängt. Daraus haben wir Hinweise auf die Verwechslung einer Translation parallel zur Bildebene mit einer Rotation um eine zur Bildebene parallele Achse hergeleitet sowie nachgewiesen, welche linearen Kombinationen von Bewegungsparametern stabil berechnet werden können.
4. Wir haben analytisch den Zusammenhang zwischen der Fehlerempfindlichkeit der Struktur und der Form der Bewegung hergeleitet.

Unsere Ergebnisse betreffen das Bewegungsfeld von Punkten. Bei der Verwendung von Geraden bleiben noch viele Probleme zu lösen, insbesondere die Ermittlung von

Mehrdeutigkeiten und die Untersuchung der Fehlerempfindlichkeit im Fall der minimalen Information eines Bildtripels. Das Bewegungsfeld von Geraden ist bis jetzt nur von [Faugeras *et al.* 89; Navab *et al.* 90] untersucht worden. Dazu haben wir zwei neue Gleichungen zur Ermittlung der Geschwindigkeit und Beschleunigung abgeleitet, ausgehend von der Plücker'schen Notation für Geraden. Für den Fall der Koplanarität der Geraden im Raum haben wir den Nachweis erbracht, daß sich der Fall von Geraden analog zum Fall von Punkten mittels Hilfsparametern formulieren läßt.

Die Aufklärung der bei der Bewegungsinterpretation aus einem Bildpaar auftretenden Probleme dient nur als Basis zum Verständnis der entsprechenden Probleme bei der Verwendung einer längeren Bildfolge. Unser Beitrag zur Auswertung von längeren Bildfolgen besteht in den folgenden Punkten:

1. Wir haben die Leistung von rekursiven stochastischen Schätzern in Bezug auf Verzerrung in der Schätzung, zuverlässige Ermittlung der Unsicherheit und Abhängigkeit von den Startwerten untersucht.
2. Wir haben experimentell die Überwindung der im Fall des Bildpaars auftretenden Fehlerempfindlichkeit untersucht. Dazu haben wir die Abwesenheit von Bewegungsinformation mit der Ausnutzung der zeitlichen Kohärenz durch Bewegungsmodelle an einer monokularen Bildfolge verglichen, die von einer an einem Roboter greifer befestigten Kamera aufgenommen wurde. Die Untersuchung hat sich auf den Fall der Ebene beschränkt, für den die von uns analytisch hergeleiteten Ergebnisse aus dem dritten Kapitel zum Vergleich existieren.

Bei der Auswertung von monokularen Bildfolgen bleiben noch viele offene Probleme, die weiter zu untersuchen sind. Wir erwähnen diejenigen, die zu der in unserer Arbeit behandelten Verarbeitungsstufe gehören :

- Die Ausnutzung der 3D-Interpretation zur Segmentierung des Bildbereichs, die bis jetzt immer vor dem Schritt der 3D-Interpretation stattfindet.
- Die Untersuchung des Bewegungsfeldes von Kurven und verdeckenden Kurven in der Richtung, die von [Faugeras 90; Blake *et al.* 91] initiiert wurde.
- Die Erweiterung von regelungstechnischen Ansätzen [Dickmanns & Graefe 88a, 88b] auf Anwendungen, wo nur wenig a priori Wissen über die Geometrie der Szene zur Verfügung steht.
- Der weitere Vergleich von Algorithmen zur allgemeinsten Problemstellung mit Ansätzen, die zusätzliche Sensoren verwenden. Die besonders verbreitete Verwendung einer Stereoanordnung, die jemandem erlaubt, unmittelbar nach der Lösung der Stereokorrespondenz über 3D-Messungen zu verfügen, sollte mit der gleichzeitigen Verwendung des Bewegungsfeldes aus einer Kamera kombiniert werden.

A Mathematische Notation

Im allgemeinen werden Vektoren mit fettgedruckten kleinen oder großen und Matrizen mit kursiven großen Buchstaben bezeichnet. In der folgenden Tabelle wird zu jedem Symbol die Gleichung oder die Seite angegeben, in der das Symbol zum ersten Mal auftritt.

g	Grauwertfunktion	(2.1)
∇g	Gradient der Grauwertfunktion	(2.2)
\mathbf{u}	optischer Flußvektor	(2.3)
$\nabla \mathbf{u}$	Matrix der ersten Ableitungen von \mathbf{u}	(2.4)
$\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$	Einheitsvektor in Richtung der x -, y - bzw. z -Achse (optische Achse)	(2.5)
f	Brennweite	(2.5)
$\mathbf{X} = (X, Y, Z)^T$	Punkt im Szenenbereich	(2.5)
$\mathbf{x} = (x, y, 1)^T$	Punkt in der Bildebene	(2.5)
$\mathbf{X1}, \mathbf{X2}$	Punkt im Szenenbereich zum ersten bzw. zweiten Zeitpunkt	(2.6)
R	Rotationsmatrix	(2.6)
T	Translationsvektor	(2.6)
$\mathbf{x1}, \mathbf{x2}$	Punkt in der Bildebene zum ersten bzw. zweiten Zeitpunkt	(2.7)
$[\mathbf{x}]_{\times}$	schiefsymmetrische Matrix, die aus den Elementen eines Vektors \mathbf{x} gebildet wird	(2.9)
E	Matrix der essentiellen Parameter, Produkt $[\mathbf{T}]_{\times} R$	(2.9)
e	Vektor, der aus den Elementen der E -Matrix besteht	(2.10)
\mathbf{a}_i	Datenvektor bei der Berechnung der essentiellen Parameter	(2.10)
$\delta \mathbf{T}, \delta \boldsymbol{\omega}$	differentielle Korrekturen in der Translation und Rotation	(2.14)
A_{col}	Matrix der Kollineation bei der Bewegung einer Ebene im diskreten Fall	(2.15)
N	Normale einer Ebene in Bewegung	(2.15)

d	Abstand einer Ebene vom Ursprung	(2.15)
$\dot{\mathbf{X}}$	Geschwindigkeit eines Punktes im Szenenbereich	(2.16)
ω	Winkelgeschwindigkeit	(2.16)
\mathbf{v}	Translationsgeschwindigkeit	(2.16)
$\dot{\mathbf{x}} = (\dot{x}, \dot{y}, 0)^T$	Verschiebungsrate eines Punktes in der Bildebene	(2.17)
$\dot{\mathbf{x}}_T = (\dot{x}_T, \dot{y}_T, 0)$	Translationskomponente der Verschiebungsrate	(2.20)
$\dot{\mathbf{x}}_R = (\dot{x}_R, \dot{y}_R, 0)$	Rotationskomponente der Verschiebungsrate	(2.34)
\mathbf{e}_c	Vektor der essentiellen Parameter im kontinuierlichen Fall	(2.26)
\mathbf{a}_c	Datenvektor bei der Berechnung der essentiellen Parameter im kontinuierlichen Fall	(2.26)
P, \mathbf{p}	Hilfsmatrix und -vektor bei der Bewegung einer Ebene im kontinuierlichen Fall	(2.37)
Q	Hilfsmatrix bei der Bewegung einer Ebene im kontinuierlichen Fall	(2.40)
\mathbf{q}	Hilfsvektor bei der Bewegung einer Ebene im kontinuierlichen Fall	(3.52)
Ψ, Φ	Weitere Hilfsmatrizen und -vektoren bei der Bewegung einer Ebene im kontinuierlichen Fall	(2.41)
\mathbf{A}, \mathbf{B}	Zusammenfassungen von Termen beim direkten Ansatz zur Bewegungsbestimmung	(2.52)
l, \mathbf{d}	Plückersche Koordinaten einer Geraden im Raum	(C.1)
\mathbf{n}	Normale der Sehebene bei Bewegung von Geraden	(2.56)
A_r, B_r	Rotationsmatrizen bei der Bewegung von Geraden	(2.57),(2.58)
$\mathbf{a}_t, \mathbf{b}_t$	Translationsvektoren bei der Bewegung von Geraden	(2.57),(2.58)
$E_{i=1..3}^l$	Matrizen der essentiellen Parameter bei der Bewegung von Geraden	(2.69)
\dot{l}, \ddot{l}	zeitliche Ableitungen des Richtungsvektors einer Geraden	(2.71),(2.76)
$\dot{\mathbf{d}}, \ddot{\mathbf{d}}$	zeitliche Ableitungen des Momentvektors einer Geraden	(2.71),(2.76)
d_l	Abstand einer Geraden vom Ursprung	(2.73)
α, β	Hilfsvektoren bei der Bewegungsbestimmung von Geraden	(2.79)
$\dot{\mathbf{v}}, \dot{\omega}$	Translations- und Winkelbeschleunigung	(2.76)

γ	Länge der gemeinsamen Normale von zwei Sichtstrahlen	(3.16)
L	Matrix der Hilfsparameter bei der Bewegungsbestimmung von koplanaren Geraden	(2.90)
Λ	Hilfsmatrix bei der Bewegungsbestimmung von koplanaren Geraden	(2.93)
σ^2	sowohl als Varianz des Meßrauschens der Bildpunkte als auch der Verschiebungsraten benutzt	(3.21)
Σ	diagonale Störungsmatrix	(3.24)
Υ	die zu untersuchende Matrix im Abschnitt 3.3	(3.21)
$\bar{\Upsilon}$	Erwartungswert von Υ	(3.19)
D	Integrationsbereich für die Verschiebungsraten	(3.22)
α, β	Seitenlängen eines rechteckigen Gesichtsfelds	(3.21)
A, B	$\alpha^2/12$ bzw. $\beta^2/12$	(3.24)
F	Fishersche Informationsmatrix	(3.53)
K, L, M, E	Blockmatrizen von F	(3.64)
S	Hilfsmatrix	(3.67)
D	Hilfsmatrix	(3.68)
$(R, \mathbf{a}) (S, \mathbf{b})$	Transformationen, die der ersten bzw. der zweiten Lösung für die Bewegung entsprechen	(3.75), (3.76)
U	Rotationsmatrix gleich SR^T	(3.82)
$M, \mathbf{l}, \mathbf{m}, \mathbf{n}$	Matrix und Vektoren zur Beschreibung einer orthogonalen Quadrik	(3.84)
\mathbf{c}	Symmetriezentrum der Quadrik	(3.86)
$\mathbf{X}_{ob,i}$	Positionsvektor des i -ten Punktes beschrieben im Objektkoordinatensystem	(4.1)
$p()$	perspektivische Projektion	(4.3)
$\mathbf{X}_{k/i}$	Positionsvektor des i -ten Punktes zum k -ten Zeitpunkt	(4.3)
$\mathbf{m}_{k,l}$	Bewegung zwischen Aufnahmen zum k -ten und l -ten Zeitpunkt	(4.3)
Q	Einheitsquaternion	(4.11)
$\mathbf{s}(t)$	Zustandsvektor zum Zeitpunkt t	(4.9)
\mathbf{g}, \mathbf{u}	Funktionen in der Differentialgleichung eines dynamischen Systems	(4.13)
\mathbf{w}	Prozeßrauschen	(4.13)

\mathbf{f}_k, F_k	Übergangsfunktion und die dazu gehörende Jacobi- sche Matrix	(4.15)
Q_k	Kovarianz des Prozeßrauschens	(4.15)
$\hat{\mathbf{s}}_k^-, P_k^-$	Schätzwert mit Fehlerkovarianz vor der Aktualisie- rung zum Zeitpunkt t_k	(4.23)
$\hat{\mathbf{s}}_k^+, P_k^+$	Schätzwert mit Fehlerkovarianz nach der Aktualisie- rung zum Zeitpunkt t_k	(4.21)
K_k	Verstärkungsfaktor beim Kalman-Filter	(4.23)
\mathbf{z}_k	Meßvektor zum Zeitpunkt t_k	(4.14)
\mathbf{h}_k	Meßfunktion zum Zeitpunkt t_k	(4.14)
\mathbf{v}_k	Meßrauschen zum Zeitpunkt t_k	(4.14)
R_k	Kovarianz des Meßrauschens zum Zeitpunkt t_k	(4.14)
\mathcal{Z}^k	Menge aller Messungen bis zum Zeitpunkt t_k	(4.16)
s, x	skalarer Zustand und Messung am einfachen Schätzungsbeispiel	
r, p	skalare Meßkovarianz und Schätzkovarianz	(4.45)
h, h_s, h_{ss}	skalare Meßfunktion und ihre Ableitungen	(4.45)
\mathbf{u}	skalierte Translationsgeschwindigkeit	(4.52)
C_{ki}	Meßkovarianz der Verschiebungsrate	(4.60)
\mathbf{p}	Vektor in einer sphärischen Bildfläche	(B.1)
$\dot{\mathbf{p}}$	Verschiebungsrate von \mathbf{p}	(B.2)
$\check{\mathbf{g}}$	Dualzahl, die eine Gerade darstellt	(C.14)
\check{R}	Dualzahl, die eine allgemeine Raumtransformation darstellt	(C.16)
$\check{\omega}$	duale Geschwindigkeit	(C.23)
${}^i\mathcal{T}_j, {}^iR_j, {}^i\mathbf{T}_j$	Allgemeine homogene Transformation, bzw. Rota- tion und Translation von Koordinatensystem j auf Koordinatensystem i	(E.1),(E.2)
Q_ω, Q_v	Hilfsmatrizen	(E.4)

B Sphärische Projektion

Bei der sphärischen Projektion werden die Punkte des Szenenbereichs auf eine Bildhalbkugel projiziert. Der Radius der Halbkugel ist gleich der Brennweite und wird gleich einer Maßeinheit gesetzt. Wir bezeichnen die Punkte des sphärischen Bildbereichs mit \mathbf{p} . Sie sind die Einheitsvektoren in Richtung des Sichtstrahls

$$\mathbf{p} = \frac{\mathbf{X}}{\|\mathbf{X}\|}. \quad (\text{B.1})$$

Die Verschiebungsrate $\dot{\mathbf{p}}$ steht senkrecht auf dem Sichtstrahl \mathbf{p} und lautet

$$\dot{\mathbf{p}} = \frac{1}{\|\mathbf{X}\|}(\mathbf{p} \times (\dot{\mathbf{X}} \times \mathbf{p})). \quad (\text{B.2})$$

Die Abbildung der Halbkugel auf die Bildebene ist eineindeutig mit Ausnahme der Punkte auf dem Äquator der Halbkugel, deren Ortsvektoren parallel zur Bildebene sind. Die Transformationen von der Halbkugel auf die Bildebene und umgekehrt lauten

$$\begin{aligned} \mathbf{x} &= \frac{\mathbf{p}}{\hat{\mathbf{z}}^T \mathbf{p}} & \mathbf{p} &= \frac{\mathbf{x}}{\|\mathbf{x}\|} \\ \dot{\mathbf{x}} &= \frac{1}{\hat{\mathbf{z}}^T \mathbf{p}}(\hat{\mathbf{z}} \times (\dot{\mathbf{p}} \times \mathbf{x})) & \dot{\mathbf{p}} &= \frac{1}{\|\mathbf{x}\|}(\mathbf{p} \times (\dot{\mathbf{x}} \times \mathbf{p})) \end{aligned}$$

Durch Ersetzung von $\dot{\mathbf{X}}$ in (B.2) nach Gl. (2.16) erhalten wir die Beziehung der Verschiebungsrate zu den 3D-Parametern:

$$\dot{\mathbf{p}} = \frac{1}{\|\mathbf{X}\|}(\mathbf{p} \times (\mathbf{v} \times \mathbf{p})) + \boldsymbol{\omega} \times \mathbf{p}. \quad (\text{B.3})$$

Die Verschiebungsrate läßt sich auch im Fall der sphärischen Projektion in eine rotatorische und eine translatorische Komponente zerlegen. Die translatorische Komponente läßt sich als die Projektion der skalierten Translationsgeschwindigkeit $\mathbf{v}/\|\mathbf{X}\|$ auf die Tangentialebene der Kugel interpretieren. Eine abstandsfreie Beziehung folgt aus (B.3) durch Multiplikation mit dem Vektorprodukt $\mathbf{v} \times \mathbf{p}$

$$(\mathbf{v} \times \mathbf{p})^T(\dot{\mathbf{p}} - \boldsymbol{\omega} \times \mathbf{p}) = 0, \quad (\text{B.4})$$

die identisch zu der entsprechenden Beziehung der perspektivischen Projektion (2.24) ist.

C Geradenrepräsentationen

C.1 Repräsentation einer Geraden im Raum

Es ist bereits Mitte des neunzehnten Jahrhunderts von Plücker nachgewiesen worden, daß man für die Darstellung einer Geraden im \mathcal{R}^3 vier Parameter braucht ¹ (siehe die Zitate in [Pickert 61]).

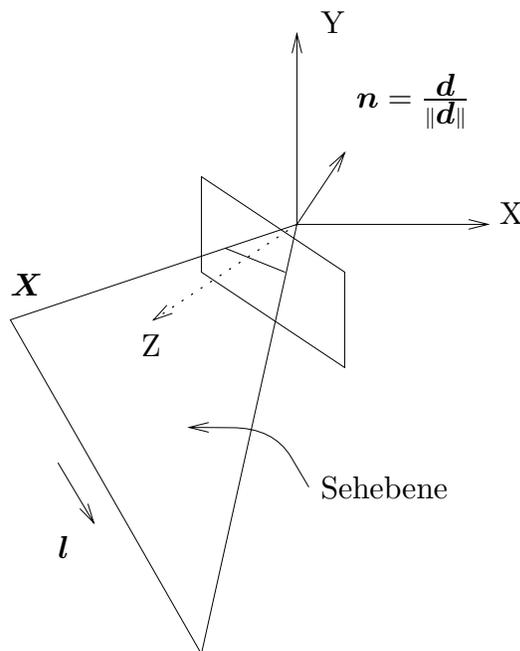


Abbildung C.1: Parametrisierung einer Geraden im Raum und ihrer Projektion

Wir beginnen mit der *Plückerschen* Notation. Eine Gerade im \mathcal{R}^3 kann man durch den Einheitsvektor l in der Richtung der Geraden und das Moment d der Geraden bezüglich des Ursprungs darstellen. Sei X ein Punkt auf der Geraden, so wird das Geradenmoment durch

$$d = X \times l \quad (\text{C.1})$$

definiert. Das Vektorpaar (l, d) erfüllt die Bedingungen

$$\|l\| = 1 \quad \text{und} \quad l^T d = 0 \quad (\text{C.2})$$

¹In der Algebra wurde später nachgewiesen, daß man für die Darstellung eines linearen m -dimensionalen Unterraums in einem n -dimensionalen Raum $(n - m)(m + 1)$ Parameter braucht.

und bildet die Plückerschen Koordinaten einer Geraden. Das Sechstupel (\mathbf{l}, \mathbf{d}) zusammen mit den beiden obengenannten Bedingungen besitzt vier Freiheitsgrade und bildet somit eine minimale Repräsentation. Der Betrag des Geradenmoments \mathbf{d} ist gleich dem Abstand der Geraden vom Ursprung und verschwindet, wenn die Gerade durch den Ursprung läuft.

Unglücklicherweise ist man bei Anwendungen meistens gezwungen, Nebenbedingungen bei Parametrisierungen zu vermeiden, weil sich die Abhängigkeit der Parameter in singulären Formen bei Unsicherheitsabschätzungen widerspiegelt. So muß man \mathbf{l} als einen Punkt auf der zweidimensionalen Halbkugel mittels Kugelkoordinaten darstellen

$$\mathbf{l} = \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix} \quad \text{mit} \quad \phi \in (-\pi, \pi], \quad \theta \in [0, \pi/2], \quad (\text{C.3})$$

es sei denn, man kann eine Formulierung finden, die \mathbf{l} in einer quadratischen Form beinhaltet.

Der Momentvektor \mathbf{d} ist auf die Ebene senkrecht zu \mathbf{l} beschränkt. In dieser Ebene braucht man einen Winkel ψ bezüglich einer Referenzachse und den Betrag $\|\mathbf{d}\|$ des Vektors als Parameter. Wenn die Gerade nicht parallel zur x -Achse ist, kann man $\hat{\mathbf{x}} \times \mathbf{l}$ als Referenzrichtung wählen und \mathbf{d} wie folgt schreiben

$$\mathbf{d} = \|\mathbf{d}\| (\cos \psi (\hat{\mathbf{x}} \times \mathbf{l}) + \sin \psi (\mathbf{l} \times (\hat{\mathbf{x}} \times \mathbf{l}))). \quad (\text{C.4})$$

Das Quadrupel $(\phi, \theta, \psi, \|\mathbf{d}\|)$ bildet dann eine minimale Repräsentation [Roberts 88].

Auf ein weiteres Quadrupel einer minimalen Repräsentation kann man gelangen, wenn man zwei Parameter für die Beschreibung der *Sehebene*, die von der Geraden und dem Ursprung aufgespannt wird, und noch zwei für die Lage der Geraden in dieser Ebene benutzt. Ausgehend von der Projektion einer Geraden in der Bildebene, die durch folgende Gleichung gegeben wird

$$x \cos \gamma + y \sin \gamma - \tan \beta = 0, \quad (\text{C.5})$$

benutzen [Vieville & Faugeras 90] den Vektor $\mathbf{n}_0 = (\cos \gamma, \sin \gamma, -\tan \beta)^T$ als den Normalenvektor der obengenannten Sehebene. Für die Lage der Geraden in der Sehebene braucht man dann noch den Abstand vom Ursprung D und einen Winkel α bezüglich einer Referenzachse. Wenn man als Referenzachse die Richtung der Geraden in der Bildebene $\mathbf{n}'_0 = (-\sin \gamma, \cos \gamma, 0)^T$ wählt, kann man aus dieser Parametrisierung $(D, \alpha, \beta, \gamma)$ wie folgt auf die Plückerschen Vektoren rücktransformieren

$$\begin{aligned} \mathbf{l} &= \cos \alpha (\mathbf{n}_0 \times \mathbf{n}'_0) + \sin \alpha \mathbf{n}'_0 \\ \mathbf{d} &= D \frac{\mathbf{n}_0}{\|\mathbf{n}_0\|} \end{aligned} \quad (\text{C.6})$$

Eine weitere minimale Repräsentation bildet die Modellierung einer Geraden als der Schnittmenge von zwei Ebenen. Unterstellt man, daß die Gerade nicht parallel zur z -Achse ist, kann man sich die Gerade als Schnitt folgender Ebenen darstellen (siehe den Gebrauch dieser Repräsentation zur Bewegungsschätzung [Ayache & Faugeras 89; Spetsakis & Aloimonos 90])

$$\begin{aligned} X &= a_x Z + b_x \\ Y &= a_y Z + b_y. \end{aligned} \quad (\text{C.7})$$

Der Richtungsvektor der Geraden ist dann $\mathbf{l} = (a_x, a_y, 1)^T$ und der Momentvektor $\mathbf{d} = (b_y - a_y, -b_x + a_x, b_x a_y - b_y a_x)^T$.

Der Beitrag dieser Beschreibung liegt in der Gegenüberstellung verschiedener Repräsentationen zu der Plücker'schen Darstellung. Wie wir im folgenden Abschnitt sehen werden, lassen sich die Transformationen von Plücker'schen Vektoren besonders elegant mittels der Dualzahlen formulieren.

C.2 Verschiebungsrates einer Geraden in der Bildebene

Die Verschiebungsrates einer Geraden wird als die zeitliche Ableitung $\dot{\mathbf{n}}$ des Einheitsvektors \mathbf{n} definiert, der die Normale der Sehebene (siehe Anfang dieses Anhangs) angibt. Hier werden wir beschreiben, wie die Verschiebungsrates aus den Bilddaten berechnet werden kann.

Wir wählen die Gleichung $x \cos \gamma + y \sin \gamma - \delta = 0$ wie in (C.5) als Darstellung der Geraden in der Bildebene. Die Parameter γ und δ können z.B. direkt aus der Geradenanpassung an einer extrahierten Kantenelementkette ermittelt werden. Der Vektor parallel zur Normalen ist $(\cos \gamma, \sin \gamma, -d)$ und als Darstellung für die Normale wählen wir die Azimut- und Polarwinkel-Darstellung $(\sin \beta \cos \gamma, \sin \beta \sin \gamma, \cos \beta)$, die man aus der obigen Darstellung der Geraden mittels der Beziehung $\tan \beta = -1/\delta$ herleiten kann. D.h. der Polarwinkel der Normalen \mathbf{n} variiert zwischen $\pi/2$ für eine Gerade durch den Ursprung der Bildebene bis π für eine Gerade im Unendlichen.

Für die Verschiebungsrates der Normalen \mathbf{n} gilt dann

$$\dot{\mathbf{n}} = \begin{pmatrix} \dot{\beta} \cos \beta \cos \gamma - \dot{\gamma} \sin \gamma \sin \beta \\ \dot{\beta} \cos \beta \sin \gamma + \dot{\gamma} \cos \gamma \sin \beta \\ -\dot{\beta} \sin \beta \end{pmatrix} \quad (\text{C.8})$$

und man kann leicht bestätigen, daß $\dot{\mathbf{n}}^T \mathbf{n} = 0$ gilt.

Zur Verwendung der Gl. (2.73) zur Bestimmung von Bewegungs- und Strukturparametern muß man zwei unabhängige aus den drei Gleichungen auswählen. Die zu den x - und y -Komponenten gehörigen Gleichungen werden nur dann abhängig voneinander, wenn $(\mathbf{n})_z = 0$ wird, d.h. wenn die Gerade durch den Ursprung der Bildebene läuft.

C.3 Geradentransformationen und Dualzahlen

Eine Gerade (\mathbf{l}, \mathbf{d}) wird mittels einer Rotation R und einer Translation \mathbf{T} auf $(\mathbf{l}', \mathbf{d}')$ wie folgt überführt

$$\mathbf{l}' = R\mathbf{l} \quad (\text{C.9})$$

$$\begin{aligned} \mathbf{d}' &= \mathbf{X}' \times \mathbf{l}' = (R\mathbf{X} + \mathbf{T}) \times R\mathbf{l} \\ &= R(\mathbf{X} \times \mathbf{l}) + \mathbf{T} \times R\mathbf{l} \end{aligned} \quad (\text{C.10})$$

$$= R\mathbf{d} + \mathbf{T} \times R\mathbf{l}. \quad (\text{C.11})$$

Obige Gleichungen können wie folgt in eine Gleichung zusammengefaßt werden

$$\begin{pmatrix} \mathbf{l}' \\ \mathbf{d}' \end{pmatrix} = \begin{pmatrix} R & 0 \\ [\mathbf{T}]_{\times} R & R \end{pmatrix} \begin{pmatrix} \mathbf{l} \\ \mathbf{d} \end{pmatrix} \quad (\text{C.12})$$

Die Menge der Matrizen der Form

$$\begin{pmatrix} a & 0 \\ b & a \end{pmatrix}$$

hat eine sehr besondere Eigenschaft. Sie ist isomorph zum Ring der *Dualzahlen*². Eine Dualzahl wird definiert als

$$\check{z} = a + \epsilon b \quad \text{mit} \quad \epsilon^2 = 0. \quad (\text{C.13})$$

Wir nennen a den reellen Teil und b den dualen Teil einer Dualzahl. Ihre Definition weist uns direkt auf die komplexen Zahlen hin, bei denen $i^2 = -1$ statt $\epsilon^2 = 0$ gilt. Der Ring der Dualzahlen ist kommutativ und hat ein Einselement, aber bildet keinen Körper, weil nur die Elemente, deren reeller Teil ungleich Null ist, ein inverses Element besitzen. So gilt für Multiplikation und Division

$$\begin{aligned} \check{z}_1 \check{z}_2 &= (a + \epsilon b)(c + \epsilon d) = ac + \epsilon(ad + bc) \\ \frac{\check{z}_2}{\check{z}_1} &= \frac{c + \epsilon d}{a + \epsilon b} = \frac{c}{a} + \epsilon \frac{ad - bc}{a^2}. \end{aligned}$$

Entsprechend lassen sich *Dualvektoren* und *Dualmatrizen* wie folgt definieren [Bottema & Roth 79; Pennoc & Yang 85; McCarthy 86].

$$\begin{aligned} \check{\mathbf{z}} &= \mathbf{a} + \epsilon \mathbf{b} \\ \check{Z} &= A + \epsilon B. \end{aligned}$$

Benutzt man als reellen Teil eines Dualvektors den Richtungsvektor einer Geraden und als dualen Teil den Momentvektor, so kann man jede Gerade mit einem Dualvektor darstellen, der lautet

$$\check{\mathbf{g}} = \mathbf{l} + \epsilon \mathbf{d}. \quad (\text{C.14})$$

Damit eine Gerade nur von einer Dualzahl $\check{\mathbf{g}}$ dargestellt werden kann, soll gelten

$$\check{\mathbf{g}}^T \check{\mathbf{g}} = 1 \quad \Leftrightarrow \quad \mathbf{l}^T \mathbf{l} = 1 \quad \text{und} \quad \mathbf{l}^T \mathbf{d} = 0. \quad (\text{C.15})$$

Benutzt man jetzt als reellen Teil einer Dualmatrix die orthogonale Matrix R und als dualen Teil das Matrizenprodukt $[\mathbf{T}]_{\times} R$, so gelangt man zu einer Dualmatrix

$$\check{R} = R + \epsilon [\mathbf{T}]_{\times} R, \quad (\text{C.16})$$

die orthogonal ist:

$$\check{R}^T \check{R} = R^T R + \epsilon (R^T [\mathbf{T}]_{\times}^T R + R^T [\mathbf{T}]_{\times} R) = I \quad (\text{C.17})$$

Daraus ergeben sich zwei interessante Einsichten:

- Eine orthogonale Dualmatrix kann gleichzeitig die Information für Translation und Rotation beinhalten.

²Dualzahlen wurden vom deutschen Geometer *E. Study (1862-1930)* eingeführt [Yaglom 68]

- Darauf aufbauend kann man die Multiplikation eines Dualvektors mit einer orthogonalen Dualmatrix

$$\check{\mathbf{g}}' = \check{R}\check{\mathbf{g}} \quad (\text{C.18})$$

als die starre Transformation einer Geraden im Raum darstellen.

Diese eleganten Eigenschaften lassen sich auch für die momentane Bewegung einer Geraden herleiten. Für die ersten zeitlichen Ableitungen des Richtungs- und Momentvektors gilt

$$\dot{\mathbf{l}} = \boldsymbol{\omega} \times \mathbf{l} \quad (\text{C.19})$$

$$\dot{\mathbf{d}} = \dot{\mathbf{X}} \times \mathbf{l} + \mathbf{X} \times \dot{\mathbf{l}} \quad (\text{C.20})$$

$$= (\mathbf{v} + \boldsymbol{\omega} \times \mathbf{X}) \times \mathbf{l} + \mathbf{X} \times (\boldsymbol{\omega} \times \mathbf{l}) \quad (\text{C.21})$$

$$= \mathbf{v} \times \mathbf{l} + \boldsymbol{\omega} \times \mathbf{d}. \quad (\text{C.22})$$

Definiert man jetzt die duale Geschwindigkeit $\check{\boldsymbol{\omega}} = \boldsymbol{\omega} + \epsilon\mathbf{v}$, so erhält man mittels einer Dualoperation die Geschwindigkeit $\check{\mathbf{g}}$ der Geraden

$$\check{\mathbf{g}} = (\boldsymbol{\omega} + \epsilon\mathbf{v}) \times (\mathbf{l} + \epsilon\mathbf{d}) = \check{\boldsymbol{\omega}} \times \check{\mathbf{g}}. \quad (\text{C.23})$$

Anhand folgender Eigenschaft der Taylor-Entwicklung einer dualzahlwertigen Funktion nach einem dualzahlwertigen Argument

$$f(a + \epsilon b) = f(a) + \epsilon b f'(a) \quad (\text{C.24})$$

kann gezeigt werden, daß

$$\cos(a + \epsilon b) = \cos a - \epsilon b \sin a \quad (\text{C.25})$$

$$\sin(a + \epsilon b) = \sin a + \epsilon b \cos a. \quad (\text{C.26})$$

Betrachtet man jetzt zwei Geraden $\check{\mathbf{g}}_1$ und $\check{\mathbf{g}}_2$, so lautet ihr duales Skalarprodukt

$$\check{\mathbf{g}}_1^T \check{\mathbf{g}}_2 = \mathbf{l}_1^T \mathbf{l}_2 + \epsilon(\mathbf{l}_1^T \mathbf{d}_2 + \mathbf{l}_2^T \mathbf{d}_1). \quad (\text{C.27})$$

Es kann durch eine einfache Rechnung gezeigt werden, daß

$$\mathbf{l}_1^T \mathbf{d}_2 + \mathbf{l}_2^T \mathbf{d}_1 = -d \sin \theta, \quad (\text{C.28})$$

wobei θ den Winkel zwischen den Geradenrichtungen und d den Betrag des Abstandes der Schnittpunkte der Geraden mit ihrer gemeinsamen Normalen darstellt. Man kann θ und d in eine Dualzahl $\check{\theta} = \theta + \epsilon d$ zusammenfassen, woraus folgt, daß das Skalarprodukt zwischen zwei Dualvektoren $\check{\mathbf{g}}_{i=1,2}$ mit $\check{\mathbf{g}}_i^T \check{\mathbf{g}}_i = 1$ gleich dem Cosinus des Dualwinkels $\check{\theta}$ ist. Solche Kombination von Winkel und Abstand wie (θ, d) kann man für die Angabe der relativen Lage zwischen zwei Achsen am Gelenk eines Roboterarms benutzen. So bilden orthogonale Dualmatrizen eine kompakte Formulierung, die für die Lösung des direkten und inversen kinematischen Problems in der Robotik [Gu & Luh 87; Walker 88] geeignet ist. Eine Untersuchung der Zeit- und Speichereffizienz der Operationen mit Dualzahlen im Vergleich zu den üblichen Transformationsdarstellungen findet man in [Funda & Paul 90]. Die erste Verwendung der Dualzahlen in der Bildauswertung erscheint bei [Walker *et al.* 91], wo die 3D-Bewegung mit Hilfe von Dualquaternionen aus durch Stereotriangulation rekonstruierten Geraden im Raum zu zwei Zeitpunkten ermittelt wird. Bis jetzt ist es uns nicht gelungen – außer der Verwendung der Plückerschen Vektoren –, die Dualzahlen als solche bei der Bewegungsbestimmung aus monokularen Bildfolgen zu verwenden, weil wir keine elegante Formulierung in Dualtermen für die Projektion der Geraden auf die Bildebene finden konnten.

D Berechnung der Jacobischen Matrizen

Wir fangen mit der Berechnung der Jacobischen Matrix der Übergangsgleichung an. Aus der Darstellung des Zustandsvektors (4.53) folgt, daß sich die Jacobische Matrix aus neun 3×3 Blockmatrizen bildet, wie folgt:

$$\frac{\partial \mathbf{f}_k}{\partial \mathbf{s}_k}(\hat{\mathbf{s}}_k^+) = \begin{pmatrix} \frac{\partial \mathbf{u}_{k+1}}{\partial \mathbf{u}_k} & \frac{\partial \mathbf{u}_{k+1}}{\partial \boldsymbol{\omega}_k} & \frac{\partial \mathbf{u}_{k+1}}{\partial \mathbf{N}_k} \\ \frac{\partial \boldsymbol{\omega}_{k+1}}{\partial \mathbf{u}_k} & \frac{\partial \boldsymbol{\omega}_{k+1}}{\partial \boldsymbol{\omega}_k} & \frac{\partial \boldsymbol{\omega}_{k+1}}{\partial \mathbf{N}_k} \\ \frac{\partial \mathbf{N}_{k+1}}{\partial \mathbf{u}_k} & \frac{\partial \mathbf{N}_{k+1}}{\partial \boldsymbol{\omega}_k} & \frac{\partial \mathbf{N}_{k+1}}{\partial \mathbf{N}_k} \end{pmatrix}. \quad (\text{D.1})$$

Die Ableitungen $\partial \boldsymbol{\omega}_{k+1}/\partial \mathbf{u}_k$, $\partial \boldsymbol{\omega}_{k+1}/\partial \mathbf{N}_k$ und $\partial \mathbf{N}_{k+1}/\partial \mathbf{u}_k$ sind Nullmatrizen und $\partial \boldsymbol{\omega}_{k+1}/\partial \boldsymbol{\omega}_k$ ist eine Einheitsmatrix.

Wir bezeichnen den Nenner von (4.57) mit M und wenden folgende Rechenregel für Differenzierung von Vektortermen:

$$\frac{\partial f \mathbf{g}}{\partial \mathbf{x}} = \mathbf{g} \frac{\partial f}{\partial \mathbf{x}} + f \frac{\partial \mathbf{g}}{\partial \mathbf{x}}, \quad (\text{D.2})$$

wobei wir den Gradienten $\frac{\partial f}{\partial \mathbf{x}}$ in Übereinstimmung mit der Jacobischen Matrix als Zeilenvektor darstellen.

Daher gilt (I ist die Einheitsmatrix):

$$\begin{aligned} \frac{\partial \mathbf{u}_{k+1}}{\partial \mathbf{u}_k} &= -\frac{1}{M^2} \mathbf{u}_k \frac{\partial M}{\partial \mathbf{u}_k} + \frac{1}{M} I \\ \frac{\partial \mathbf{u}_{k+1}}{\partial \boldsymbol{\omega}_k} &= -\frac{1}{M^2} \mathbf{u}_k \frac{\partial M}{\partial \boldsymbol{\omega}_k} \\ \frac{\partial \mathbf{u}_{k+1}}{\partial \mathbf{N}_k} &= -\frac{1}{M^2} \mathbf{u}_k \frac{\partial M}{\partial \mathbf{N}_k}. \end{aligned}$$

Um die Ableitungen des Nenners M zu berechnen, formen wir den Nenner wie folgt um:

$$\begin{aligned} M &= 1 + \frac{\mathbf{u}_k^T (\boldsymbol{\omega}_k \times \mathbf{N}_k) (1 - \cos(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^2} - \frac{\boldsymbol{\omega}_k^T \mathbf{u}_k \boldsymbol{\omega}_k^T \mathbf{N}_k \|\boldsymbol{\omega}_k\|T - \sin(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|^3} \\ &\quad - \frac{\mathbf{u}_k^T \mathbf{N}_k \sin(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|}. \end{aligned}$$

Nach Verwendung der Regel

$$\frac{\partial \|\mathbf{g}\|}{\partial \mathbf{g}} = \frac{\mathbf{g}}{\|\mathbf{g}\|}$$

folgt (ω_o ist der Einheitsvektor in Richtung von ω_k):

$$\frac{\partial M}{\partial \mathbf{u}_k} = \frac{1 - \cos(\|\omega_k\|T)}{\|\omega_k\|^2} (\omega_k \times \mathbf{N}_k)^T - \omega_k^T \mathbf{N}_k \frac{\|\omega_k\|T - \sin(\|\omega_k\|T)}{\|\omega_k\|^3} \omega_k^T - \frac{\sin(\|\omega_k\|T)}{\|\omega_k\|} \mathbf{N}_k^T,$$

$$\begin{aligned} \frac{\partial M}{\partial \omega_k} &= \frac{1 - \cos(\|\omega_k\|T)}{\|\omega_k\|^2} (\mathbf{N}_k \times \mathbf{u}_k)^T \\ &+ \mathbf{u}_k^T (\omega_k \times \mathbf{N}_k) \frac{\|\omega_k\|T \sin(\|\omega_k\|T) - 2(1 - \cos(\|\omega_k\|T))}{\|\omega_k\|^3} \omega_o^T \\ &- \frac{\|\omega_k\|T - \sin(\|\omega_k\|T)}{\|\omega_k\|^3} (\omega_k^T \mathbf{N}_k \mathbf{u}_k^T + \omega_k^T \mathbf{u}_k \mathbf{N}_k^T) \\ &- \omega_k^T \mathbf{u}_k \omega_k^T \mathbf{N}_k \frac{3 \sin(\|\omega_k\|T) - \|\omega_k\|T \cos(\|\omega_k\|T) - 2\|\omega_k\|T}{\|\omega_k\|^4} \omega_o^T \\ &- \mathbf{N}_k^T \mathbf{u}_k \frac{\|\omega_k\|T \cos(\|\omega_k\|T) - \sin(\|\omega_k\|T)}{\|\omega_k\|^2} \omega_o^T \end{aligned}$$

und

$$\frac{\partial M}{\partial \mathbf{N}_k} = \frac{1 - \cos(\|\omega_k\|T)}{\|\omega_k\|^2} (\mathbf{u}_k \times \omega_k)^T - \omega_k^T \mathbf{u}_k \frac{\|\omega_k\|T - \sin(\|\omega_k\|T)}{\|\omega_k\|^3} \omega_o^T - \frac{\sin(\|\omega_k\|T)}{\|\omega_k\|} \mathbf{u}_k^T.$$

Es folgen die Ableitungen von \mathbf{N}_{k+1} nach ω_k und \mathbf{N}_k , die wir aus (4.55) herleiten:

$$\begin{aligned} \frac{\partial \mathbf{N}_{k+1}}{\partial \omega_k} &= -\frac{\|\omega_k\|T \cos(\|\omega_k\|T) - \sin(\|\omega_k\|T)}{\|\omega_k\|^2} (\omega_k \times \mathbf{N}_k) \omega_o^T + \frac{\sin(\|\omega_k\|T)}{\|\omega_k\|} [\mathbf{N}_k]_{\times} \\ &+ \frac{\|\omega_k\|T \sin(\|\omega_k\|T) - 2(1 - \cos(\|\omega_k\|T))}{\|\omega_k\|^3} (\omega_k^T \mathbf{N}_k \omega_k - \omega_k^T \omega_k \mathbf{N}_k) \omega_o^T \\ &+ \frac{1 - \cos(\|\omega_k\|T)}{\|\omega_k\|^2} (\omega_k \mathbf{N}_k^T + \omega_k^T \mathbf{N}_k I - 2\mathbf{N}_k \omega_k^T), \end{aligned}$$

und

$$\frac{\partial \mathbf{N}_{k+1}}{\partial \mathbf{N}_k} = I - \frac{\sin(\|\omega_k\|T)}{\|\omega_k\|} [\omega_k]_{\times} + \frac{1 - \cos(\|\omega_k\|T)}{\|\omega_k\|^2} [\omega_k]_{\times}^2$$

Wir schließen mit der Jacobischen Matrix der Meßgleichung ab.

$$\frac{\partial \mathbf{z}_k}{\partial \mathbf{s}_k} = \begin{pmatrix} -N_{kz} & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -u_{kx} \\ -N_{kx} & 0 & N_{kz} & 0 & 0 & 0 & -u_{kx} & 0 & u_{kz} \\ -N_{ky} & 0 & 0 & 0 & 0 & 1 & 0 & -u_{kx} & 0 \\ 0 & -N_{kz} & 0 & 1 & 0 & 0 & 0 & 0 & -u_{ky} \\ 0 & -N_{kx} & 0 & 0 & 0 & -1 & -u_{ky} & 0 & 0 \\ 0 & -N_{ky} & N_{kz} & 0 & 0 & 0 & 0 & -u_{ky} & u_{kz} \\ 0 & 0 & N_{kx} & 0 & -1 & 0 & u_{kz} & 0 & 0 \\ 0 & 0 & N_{ky} & 1 & 0 & 0 & 0 & u_{kz} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2N_{kx} & 2N_{ky} & 2N_{kz} \end{pmatrix} \quad (\text{D.3})$$

E Umrechnungen zwischen Kamera- und Roboterkoordinatensystem

An einem gewöhnlichen Roboterarm werden folgende Koordinatensysteme wie in Abb. E.1 definiert: das Roboterkoordinatensystem (Index r), das Greiferkoordinatensystem (Index g), das Kamerakoordinatensystem (Index c) und das Weltkoordinatensystem (Index w). Alle Formulierungen in dieser Arbeit erfolgen bezüglich des Kamerakoordinatensystems.

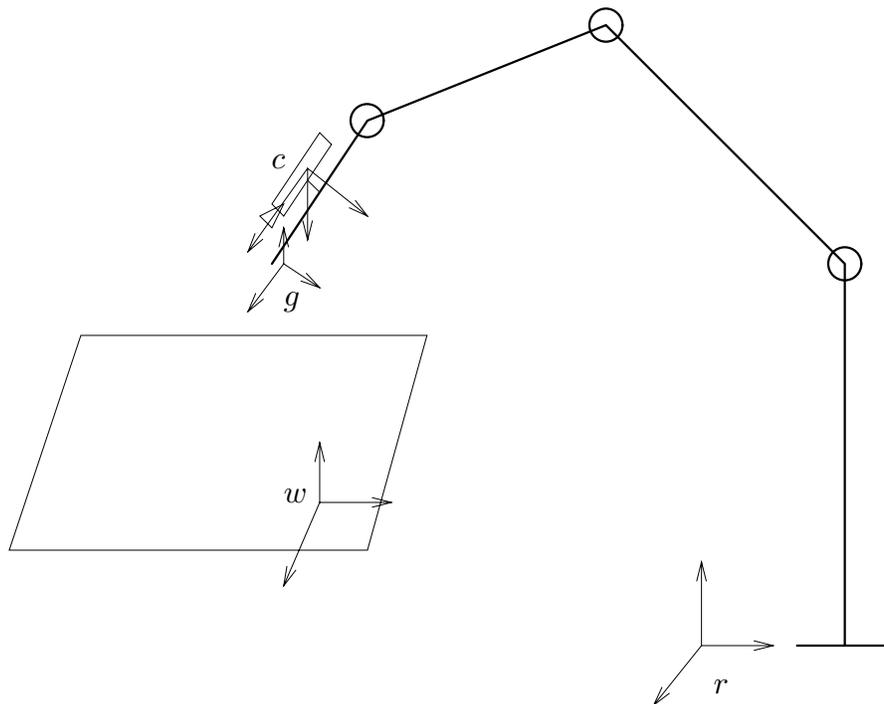


Abbildung E.1: Koordinatensysteme an einem Roboterarm

Eine Transformation vom Koordinatensystem j auf das Koordinatensystem i wird mit Hilfe der homogenen Transformationsmatrix ${}^i\mathcal{T}_j$ und homogener Koordinaten ($\mathbf{X}^h = 4 \times 1$ -Vektor)

$$\mathbf{X}_i^h = {}^i\mathcal{T}_j \mathbf{X}_j^h \quad \text{mit} \quad {}^i\mathcal{T}_j = \begin{pmatrix} {}^iR_j & {}^i\mathbf{T}_j \\ \mathbf{0}^T & 1 \end{pmatrix} \quad (\text{E.1})$$

oder mit Hilfe der Rotation iR_j und der Translation ${}^i\mathbf{T}_j$ dargestellt:

$$\mathbf{X}_i = {}^iR_j \mathbf{X}_j + {}^i\mathbf{T}_j \quad (\text{E.2})$$

Zum Aufbauen von Experimenten mit von uns vorgeschriebenen Bewegungen bezüglich des Kamerakoordinatensystems und zur Überprüfung der Genauigkeit der aus den Bewegungsschätzungsalgorithmen ermittelten Werte ist man gezwungen, auf folgende zwei Fragestellungen einzugehen.

- Gegeben seien die Translations- und Winkelgeschwindigkeit \mathbf{v} und $\boldsymbol{\omega}$ bezüglich des Kamerakoordinatensystems und die letzte Stellung (Position und Orientierung) des Greifers ${}^r\mathcal{T}_{g_k}$. Gesucht ist die nächste Stellung des Greifers ${}^r\mathcal{T}_{g_{k+1}}$, die dem Roboter als Vorgabe für die nächste Bewegung eingegeben wird.
- Gegeben seien die Meßwerte ${}^r\mathcal{T}_{g_k}$ und ${}^r\mathcal{T}_{g_{k+1}}$ aus zwei aufeinanderfolgenden Stellungen des Greifers. Gesucht sind die obigen Geschwindigkeiten bezüglich des Kamerakoordinatensystems.

Zur Bearbeitung dieser Fragen muß die Transformation ${}^g\mathcal{T}_c$ einer Hand-Auge-Kalibrierung vorliegen. Vor dem Experiment im Abschnitt 4.3 wurde eine Hand-Auge-Kalibrierung nach dem Verfahren von [Shiu & Ahmad 89] und den Modifikationen von [Zhuang & Roth 91] durchgeführt.

Zur Bearbeitung der obigen Fragestellung gehen wir von der Transformation im Kamerakoordinatensystem aus:

$$\mathbf{X}_{c_k} = {}^{c_k}R_{c_{k+1}} \mathbf{X}_{c_{k+1}} + {}^{c_k}\mathbf{T}_{c_{k+1}}. \quad (\text{E.3})$$

Die Ermittlung von ${}^{c_k}R_{c_{k+1}}$ und ${}^{c_k}\mathbf{T}_{c_{k+1}}$ aus \mathbf{v} und $\boldsymbol{\omega}$ ist etwas aufwendiger, als man sich zunächst vorstellt, und wird deshalb am Ende ausgeführt.

In der Gleichung (E.3) transformieren wir die Punkte auf Greiferkoordinaten

$${}^gR_c^T (\mathbf{X}_{g_k} - {}^g\mathbf{T}_c) = {}^{c_k}R_{c_{k+1}} {}^gR_c^T (\mathbf{X}_{g_{k+1}} - {}^g\mathbf{T}_c) + {}^{c_k}\mathbf{T}_{c_{k+1}}$$

und nach der Umformung erhalten wir:

$$\mathbf{X}_{g_k} = {}^{g_k}R_{g_{k+1}} \mathbf{X}_{g_{k+1}} + {}^{g_k}\mathbf{T}_{g_{k+1}}$$

mit

$$\begin{aligned} {}^{g_k}R_{g_{k+1}} &= {}^gR_c {}^{c_k}R_{c_{k+1}} {}^gR_c^T \\ {}^{g_k}\mathbf{T}_{g_{k+1}} &= (I - {}^gR_c {}^{c_k}R_{c_{k+1}} {}^gR_c^T) {}^g\mathbf{T}_c + {}^gR_c {}^{c_k}\mathbf{T}_{c_{k+1}}. \end{aligned}$$

Die erwünschte Stellung des Greifers kann man dann erhalten, indem man die obigen Formeln in die folgenden einsetzt:

$$\begin{aligned} {}^{r_k}R_{g_{k+1}} &= {}^{r_k}R_{g_k} {}^{g_k}R_{g_{k+1}} \\ {}^{r_k}\mathbf{T}_{g_{k+1}} &= {}^{r_k}R_{g_k} {}^{g_k}\mathbf{T}_{g_{k+1}} + {}^{r_k}\mathbf{T}_{g_k}. \end{aligned}$$

Alle Stellgrößen sind abhängig von der Hand-Auge-Transformation, die möglichst genau im voraus berechnet werden soll. Im Fall einer rein translatorischen Kamerabewegung gilt ${}^g\mathbf{T}_{g_{k+1}} = {}^gR_c {}^{c_k}\mathbf{T}_{c_{k+1}}$. Ungenauigkeiten in der ermittelten Hand-Auge-Kalibrierung beeinflussen die Stellgrößen auch im Fall der reinen Translation (vgl. Experiment im Abschnitt 4.3).

Aus zwei gemessenen Stellungen des Greifers kann man die Formeln invertieren und nach ${}^{c_k}R_{c_{k+1}}$ und ${}^{c_k}\mathbf{T}_{c_{k+1}}$ lösen. Möchte man auf die Daten der internen Robotersensoren verzichten, so muß man an den entsprechenden Stellen eine Kalibrierung in Bezug auf ein bekanntes Objekt im Weltkoordinatensystem durchführen. Diese Kalibrierung ermittelt ${}^wT_{c_k}$ und ${}^wT_{c_{k+1}}$, aus denen die relative Bewegung $({}^{c_k}R_{c_{k+1}}, {}^{c_k}\mathbf{T}_{c_{k+1}})$ leicht hergeleitet werden kann.

Es bleibt noch offen, die Beziehung zwischen $({}^{c_k}R_{c_{k+1}}, {}^{c_k}\mathbf{T}_{c_{k+1}})$ einerseits und $(\boldsymbol{\omega}, \mathbf{v})$ andererseits zu klären. Die Geschwindigkeiten werden mittels der Differentialgleichung

$$\dot{\mathbf{X}} = -\boldsymbol{\omega} \times \mathbf{X} - \mathbf{v}$$

eingeführt. Unter der Annahme, daß \mathbf{v} und $\boldsymbol{\omega}$ zeitlich konstant sind, wird diese Differentialgleichung wie folgt integriert [Friedland 86] :

$$\mathbf{X}(t) = e^{-[\boldsymbol{\omega}] \times (t-t_k)} \mathbf{X}(t_k) - \int_{t_k}^t e^{-[\boldsymbol{\omega}] \times (t-\tau)} \mathbf{v} d\tau.$$

Für $t = t_{k+1}$ und $T = t_{k+1} - t_k$ gilt

$$e^{-[\boldsymbol{\omega}] \times T} = I - \frac{\sin(\|\boldsymbol{\omega}\|T)}{\|\boldsymbol{\omega}\|} [\boldsymbol{\omega}]_{\times} + \frac{1 - \cos(\|\boldsymbol{\omega}\|T)}{\|\boldsymbol{\omega}\|^2} [\boldsymbol{\omega}]_{\times}^2,$$

d.h. es handelt sich um eine Rotation um den Winkel $(-\|\boldsymbol{\omega}\|T)$, die wir mit $Q_{\boldsymbol{\omega}}$ bezeichnen werden. Für das Integral gilt

$$\begin{aligned} & \int_{t_k}^{t_{k+1}} \left(I - \frac{\sin \|\boldsymbol{\omega}\| (t-\tau)}{\|\boldsymbol{\omega}\|} [\boldsymbol{\omega}]_{\times} + \frac{1 - \cos \|\boldsymbol{\omega}\| (t-\tau)}{\|\boldsymbol{\omega}\|^2} [\boldsymbol{\omega}]_{\times}^2 \right) \mathbf{v} d\tau \\ &= \left(T - \frac{1 - \cos(\|\boldsymbol{\omega}\|T)}{\|\boldsymbol{\omega}\|^2} [\boldsymbol{\omega}]_{\times} + \frac{\|\boldsymbol{\omega}\|T - \sin(\|\boldsymbol{\omega}\|T)}{\|\boldsymbol{\omega}\|^3} [\boldsymbol{\omega}]_{\times}^2 \right) \mathbf{v}. \end{aligned}$$

Wir bezeichnen den letzten Ausdruck mit $-Q_{\boldsymbol{\omega}}\mathbf{v}$ und erhalten

$$\mathbf{X}_{c_{k+1}} = Q_{\boldsymbol{\omega}} \mathbf{X}_{c_k} + Q_{\boldsymbol{\omega}} \mathbf{v}. \quad (\text{E.4})$$

Wir ersetzen \mathbf{X}_{c_k} durch die rechte Seite von E.3

$$\mathbf{X}_{c_{k+1}} = Q_{\boldsymbol{\omega}} ({}^{c_k}R_{c_{k+1}} \mathbf{X}_{c_{k+1}} + {}^{c_k}\mathbf{T}_{c_{k+1}}) + Q_{\boldsymbol{\omega}} \mathbf{v}$$

und erhalten die erwünschten Beziehungen zwischen finiter Transformation und Geschwindigkeiten:

$$Q_{\boldsymbol{\omega}} {}^{c_k}R_{c_{k+1}} = I \quad (\text{E.5})$$

$$Q_{\boldsymbol{\omega}} {}^{c_k}\mathbf{T}_{c_{k+1}} + Q_{\boldsymbol{\omega}} \mathbf{v} = 0. \quad (\text{E.6})$$

Die vorgestellten Berechnungen in diesem Abschnitt können auch zur Berechnung von sogenannten Modellverschiebungsraten aus rein mechanischer Information und der Kalibrierung verwendet werden. Sie bilden dadurch die ersten Schritte bei der quantitativen Bewertung von Verfahren zur Schätzung des optischen Flusses, wie sie von [Nagel 92] vorgeschlagen wurde.

Literatur

- [Adiv 85] G. Adiv, Determining 3-D motion and structure from optical flow generated by several moving objects, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-7** (1985) 384–401.
- [Adiv 89] G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 477–489.
- [Aggarwal & Nandhakumar 88] J.K. Aggarwal, N. Nandhakumar, On the computation of motion from sequences of images - a review, *Proceedings of the IEEE* **76** (1988) 917–935.
- [Aisbett 90] J. Aisbett, An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12** (1990) 1092–1098.
- [Aloimonos & Herve 90] J. Aloimonos, J.-Y. Herve, Correspondenceless stereo and motion: planar surfaces, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-12** (1990) 504–510.
- [Amari & Maruyama 87] S. Amari, M. Maruyama, A theory on the determination of 3D motion and 3D structure from features, *Spatial Vision* **2** (1987) 151–168.
- [Anandan 89] P. Anandan, A computational framework and an algorithm for the measurement of visual motion, *International Journal of Computer Vision* **2** (1989) 283–310.
- [Anderson & Moore 79] B.D.O. Anderson, J.B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979.
- [Ando 91] H. Ando, Dynamic reconstruction of 3D structure and 3D motion, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 101–110.
- [Atiya & Hager 91] S. Atiya, G. Hager, Real-time vision-based robot localization, in *Proc. IEEE Int. Conf. on Robotics and Automation*, Sacramento, CA, April 9-12, 1991, pp. 639–644.
- [Ayache & Faugeras 88] N. Ayache, O.D. Faugeras, Building, registering and fusing noisy visual maps, *Intern. Journal of Robotics Research* **7:6** (1988) 45–65.
- [Ayache & Faugeras 89] N. Ayache, O.D. Faugeras, Maintaining representations of the environment of a mobile robot, *IEEE Trans. Robotics and Automation* **RA-5** (1989) 804–819.
- [Barron *et al.* 90] J.L. Barron, A.D. Jepson, J.K. Tsotsos, The feasibility of motion and structure from noisy time-varying image velocity information, *International Journal of Computer Vision* **5** (1990) 239–269.
- [Bar-Shalom & Fortmann 88] Y. Bar-Shalom, T.E. Fortmann, *Tracking and Data Association*, Academic Press, New York, NY, 1988.
- [Basu & Aloimonos 90] A. Basu, J. Aloimonos, A robust, correspondenceless translation-determining algorithm, *Intern. Journal of Robotics Research* **9:5** (1990) 35–59.

- [Blake *et al.* 91] A. Blake, M. Brady, R. Cipolla, Z. Xie, A. Zisserman, Visual navigation around curved obstacles, in *Proc. IEEE Int. Conf. on Robotics and Automation*, Sacramento, CA, April 9-12, 1991, pp. 2490–2495.
- [Bonde & Nagel 79] T. Bonde, H. -H. Nagel, Deriving a 3-D description of a moving rigid object from monocular TV-frame sequence, in J.K. Aggarwal & N.I. Badler (ed.), *Proc. Workshop on Computer Analysis of Time-Varying Imagery*, Philadelphia, PA, April 5-6, 1979, pp. 44–45.
- [Bottema & Roth 79] O. Bottema, B. Roth, *Theoretical Kinematics*, North-Holland Publishing Company, Amsterdam New York London, 1979.
- [Broida & Chellappa 89a] T.J. Broida, R. Chellappa, Experiments and uniqueness results on object structure and kinematics from a sequence of images, in *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March 20-22, 1989, pp. 21–30.
- [Broida & Chellappa 89b] T.J. Broida, R. Chellappa, Performance bounds for estimating three-dimensional motion parameters from a sequence of noisy images, *Journal Opt. Soc. Am.* **A6** (1989) 879–889.
- [Broida & Chellappa 91] T. Broida, R. Chellappa, Estimating the kinematics and structure of a rigid object from a sequence of monocular images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **13** (1991) 497–513.
- [Broida *et al.* 90] T.J. Broida, S. Chandrashekhar, R. Chellappa, Recursive 3-D motion estimation from a monocular image sequence, *IEEE Trans. Aerospace and Electronic Systems* **26** (1990) 639–656.
- [Bruss & Horn 83] A. Bruss, B.K.P. Horn, Passive navigation, *Computer Vision, Graphics, and Image Processing* **21** (1983) 3–20.
- [Burger & Bhanu 90] W. Burger, B. Bhanu, Estimating 3-D egomotion from perspective image sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12** (1990) 1040–1058.
- [Burns *et al.* 86] J.B. Burns, A.R. Hanson, E.M. Riseman, Extracting straight lines, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-8** (1986) 425–455.
- [Buxton *et al.* 85] B.F. Buxton, D.W. Murray, H. Buxton, N.S. Williams, Structure from motion algorithms for computer vision on an SIMD architecture, *Computer Physics Communications* **37** (1985) 273–280.
- [Castelow & Rerolle 91] D.A. Castelow, A.J. Rerolle, A monocular ground plane estimation system, in *Proc. British Machine Vision Conference*, Glasgow, UK, Sept. 24-26, 1991, pp. 392–395.
- [Chandrashekhar & Chellappa 91] S. Chandrashekhar, R. Chellappa, Passive navigation in a partially known environment, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 2–7.
- [Chang & Aggarwal 91] Y.-L. Chang, J.K. Aggarwal, 3D structure reconstruction from an ego sequence using statistical estimation, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 268–273.
- [Charnley & Blissett 89] D. Charnley, R. Blissett, Surface reconstruction from outdoor image sequences, *Image and Vision Computing* **7** (1989) 10–16.
- [Cheng & Aggarwal 90] C.-J. Cheng, J.K. Aggarwal, A two-stage hybrid approach to the correspondence problem via forward-searching and backward-correcting, in

- Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 173–179.
- [Crowley & Stelmaszyk 90] J.L. Crowley, P. Stelmaszyk, Measurement and integration of 3-D structures by tracking edge lines, in *Proc. First European Conference on Computer Vision*, Antibes, France, Apr. 23-26, 1990, pp. 269–280.
- [Crowley *et al.* 88] J.L. Crowley, P. Stelmaszyk, C. Discours, Measuring image flow by tracking edge-lines, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5-8, 1988, pp. 658–664.
- [Cui *et al.* 90] N. Cui, J. Weng, P. Cohen, Extended structure and motion analysis from monocular image sequences, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 222–229.
- [Cui *et al.* 91] N. Cui, J. Weng, P. Cohen, Motion and structure from long stereo image sequences, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 75–80.
- [Daniilidis & Nagel 90] K. Daniilidis, H.-H. Nagel, Analytical Results on Error Sensitivity of Motion Estimation from Two Views, in *Proc. First European Conference on Computer Vision*, Antibes, France, Apr. 23-26, O.D. Faugeras (Ed.), Lecture Notes in Computer Science 427, Springer-Verlag, Berlin et al., 1990, pp. 199–208.
- [Debrunner & Ahuja 90] C. Debrunner, N. Ahuja, A direct data approximation based motion estimation algorithm, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 384–389.
- [Demazure 88] M. Demazure, *Sur deux problèmes de reconstruction*, Technical Report 882, INRIA, Sophia-Antipolis, France, 1988.
- [Deriche & Faugeras 90] R. Deriche, O. Faugeras, Tracking line segments, *Image and Vision Computing* **8** (1990) 261–270.
- [Dickmanns & Graefe 88a] E.D. Dickmanns, V. Graefe, Applications of dynamic monocular machine vision, *Machine Vision and Applications* **1** (1988) 241–261.
- [Dickmanns & Graefe 88b] E.D. Dickmanns, V. Graefe, Dynamic monocular machine vision, *Machine Vision and Applications* **1** (1988) 223–240.
- [Dreschler & Nagel 82] L. Dreschler, H.-H. Nagel, Volumetric Model and 3D Trajectory of a Moving Car Derived from Monocular TV Frame Sequences of a Street Scene, *Computer Graphics and Image Processing* **20** (1982) 199–228.
- [Enkelmann 90] W. Enkelmann, Obstacle detection by evaluation of optical flow fields from image sequences, in *Proc. First European Conference on Computer Vision*, Antibes, France, Apr. 23-26, 1990, pp. 134–138.
- [Enkelmann 91] W. Enkelmann, Obstacle detection by evaluation of optical flow fields from image sequences, *Image and Vision Computing* **9** (1991) 160–168.
- [Fang & Huang 84] J. Q. Fang, T. S. Huang, Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 547–554.
- [Faugeras 90] O.D. Faugeras, On the motion of 3-D curves and its relationship to optical flow, in *Proc. First European Conference on Computer Vision*, Antibes, France, Apr. 23-26, 1990, pp. 107–117.

- [Faugeras & Lustman 88] O.D. Faugeras, F. Lustman, Motion and structure from motion in a piecewise planar environment, *International Journal of Pattern Recognition and Artificial Intelligence* **2** (1988) 485–508.
- [Faugeras & Maybank 90] O.D. Faugeras, S. Maybank, Motion from point matches: multiplicity of solutions, *International Journal of Computer Vision* **4** (1990) 225–246.
- [Faugeras *et al.* 89] O.D. Faugeras, N. Deriche, N. Navab, From optical flow of lines to 3D motion and structure, in H. Burkhardt, K. H. Höhne, B. Neumann (ed.), *Mustererkennung 1989*, Hamburg, Informatik-Fachberichte **219**, Springer-Verlag, Berlin u.a., 1989, pp. 522–530.
- [Faugeras *et al.* 87a] O. D. Faugeras, F. Lustman, G.Toscani, Motion and structure from motion from point and line matches, in *Proc. Int. Conf. on Computer Vision*, London, UK, June 8-11, 1987, pp. 25–34.
- [Faugeras *et al.* 87b] O. D. Faugeras, F. Lustman, G.Toscani, Motion and structure from motion from point and line matches, in *Proc. Int. Conf. on Computer Vision*, London, UK, June 8-11, 1987, pp. 25–34.
- [Finsterwalder & Hofmann 68] R. Finsterwalder, W. Hofmann, *Photogrammetrie*, Walter de Gruyter & Co, Berlin, 1968.
- [Fleet & Jepson 89] D.J. Fleet, A.D. Jepson, Hierarchical construction of orientation and velocity selective filters, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 315–325.
- [Friedland 86] B. Friedland, *Control System Design*, McGraw-Hill, New York, 1986.
- [Funda & Paul 90] J. Funda, R.P. Paul, A computational analysis of screw transformations in robotics, *IEEE Trans. Robotics and Automation* **RA-6** (1990) 348–356.
- [Geiger & Girosi 91] D. Geiger, F. Girosi, Parallel and deterministic algorithms from MRF's: Surface reconstruction, *IEEE Trans. Pattern Analysis and Machine Intelligence* **13** (1991) 401–412.
- [Gelb 74] A. Gelb (ed.), *Applied Optimal Estimation*, The MIT Press, Cambridge, MA and London, UK, 1974.
- [Gengenbach 90] V. Gengenbach, *Automatischer Zugriff eines Roboters auf ungeordnete Werkstücke mit Hilfe einer 3D-Lagebestimmung durch ein Mehrkamerasystem*, Diplomarbeit, Fakultät für Informatik der Universität Karlsruhe, Januar 1990.
- [Gibson 66] J.J. Gibson, *The senses considered as perceptual systems*, Houghton Mifflin Co., Boston, MA, 1966.
- [Girosi *et al.* 89] F. Girosi, A. Verri, V. Torre, Constraints for the computation of optical flow, in *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March 20-22, 1989, pp. 116–124.
- [Golub & van Loan 83] G.H. Golub, C.F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, 1983.
- [Gu & Luh 87] Y.-L. Gu, J.Y.S. Luh, Dual-number transformation and its application to robotics, *IEEE Journal of Robotics and Automation* **RA-3** (1987) 615–623.
- [Hager 90] G. Hager, *Task-Directed Sensor Fusion and Planning*, Kluwer Academic Publishers, Boston/ Dordrecht/ London, 1990.

- [Harris & Pike 88] C.G. Harris, J.M. Pike, 3D positional integration from image sequences, *Image and Vision Computing* **6** (1988) 87–90.
- [Harris & Stennet 90] C. Harris, C. Stennet, RAPID - A video rate object tracker, in *Proc. British Machine Vision Conference*, Oxford, UK, Sept. 24-27, 1990, pp. 73–77.
- [Hay 66] J.C. Hay, Optical motions and space perception: an extension of Gibson’s analysis, *Psychological Review* **73** (1966) 550–565.
- [Hayashi & Negahdaripour 90] B.Y. Hayashi, S. Negahdaripour, Direct motion stereo: Recovery of observer motion and scene structure, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 446–450.
- [Heeger 88] D. Heeger, Optical flow using spatiotemporal filters, *International Journal of Computer Vision* **1** (1988) 279–302.
- [Heeger & Hager 88] D.J. Heeger, G. Hager, Egomotion and the stabilized world, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5-8, 1988, pp. 435–440.
- [Heeger & Jepson 92] D.J. Heeger, A.D. Jepson, Subspace methods for recovering rigid motion I: Algorithm and implementation, *International Journal of Computer Vision* **7** (1992) 95–117.
- [Heel 90a] J. Heel, *Direct estimation of structure and motion from multiple frames*, A.I. Memo No. 1190, MIT Artificial Intelligence Laboratory, Cambridge, MA, March 1990.
- [Heel 90b] J. Heel, Dynamic motion vision, *Robotics and Autonomous Systems* **6** (1990) 297–314.
- [Heel 90c] J. Heel, Temporally integrated surface reconstruction, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 292–292.
- [Heel 91] J. Heel, Temporal surface reconstruction, in *IEEE Conf. Computer Vision and Pattern Recognition*, Maui, Hawaii, June 3-6, 1991, pp. 607–612.
- [Hildreth 84] E. Hildreth, Computations underlying the measurement of visual motion, *Artificial Intelligence* **23** (1984) 309–354.
- [Hildreth & Grzywacz 86] E.G. Hildreth, N.M. Grzywacz, The incremental recovery of structure from motion: Position vs. velocity based formulations, in *Proc. of Workshop on Motion: Representation and Analysis*, Kiawah Island Resort, Charleston, SC, May 7-9, 1986, pp. 137–143.
- [Hofmann 53] W. Hofmann, *Das Problem der “gefährlichen Flächen” in Theorie und Praxis*, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe C, Heft 3, München, 1953.
- [Horn 86] B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1986.
- [Horn 87] B.K.P. Horn, Motion fields are hardly ever ambiguous, *International Journal of Computer Vision* **1** (1987) 259–274.
- [Horn 90] B.K.P. Horn, Relative orientation, *International Journal of Computer Vision* **4** (1990) 59–78.
- [Horn 91] B.K.P. Horn, Relative orientation revisited, *Journal Opt. Soc. Am.* **A8** (1991) 1630–1638.

- [Horn & Harris 91] B.K.P. Horn, J.G. Harris, Rigid body motion from range image sequences, *CVGIP: Image Understanding* **53** (1991) 1–13.
- [Horn & Schunk 81] B.K.P. Horn, B.G. Schunk, Determining Optical Flow, *Artificial Intelligence* **17** (1981) 185–203.
- [Horn & Weldon 87] B.K.P. Horn, E.J. Weldon, Computationally-efficient methods for recovering translational motion, in *Proc. Int. Conf. on Computer Vision*, London, UK, June 8-11, 1987, pp. 2–11.
- [Horn & Weldon 88] B.K.P. Horn, E.J. Weldon, Direct methods for recovering motion, *International Journal of Computer Vision* **2** (1988) 51–76.
- [Huang & Faugeras 89] T.S. Huang, O.D. Faugeras, Some properties of the E matrix in two-view motion estimation, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 1310–1312.
- [Huber & Graefe 91] J. Huber, V. Graefe, Quantitative interpretation of image velocities in real time, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 211–216.
- [Iu & Wohn 90] S. Iu, K. Wohn, Estimation of general rigid body motion from a long sequence of images, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 217–219.
- [Jain 84] R. Jain, Segmentation of frame sequences obtained by a moving observer, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 624–629.
- [Jain & Binford 91] R. Jain, T. Binford, Ignorance, myopia and naiveté in computer vision systems, *CVGIP: Image Understanding* **53** (1991) 112–117.
- [Jazwinski 70] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, NY and London, UK, 1970.
- [Jerian & Jain 90] C. Jerian, R. Jain, Polynomial methods for structure from motion, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-12** (1990) 1150–1165.
- [Jerian & Jain 91] C.P. Jerian, R. Jain, Structure from motion - a critical analysis of methods, *IEEE Trans. Systems, Man, and Cybernetics* **21** (1991) 572–587.
- [Kanade 80] T. Kanade, Signal vs. Semantics, *Computer Graphics and Image Processing* **13** (1980) 279–297.
- [Koenderink & van Doorn 76] J.J. Koenderink, A.J. van Doorn, Local structure of movement parallax of the plane, *Journal Opt. Soc. Am.* **66** (1976) 717–723.
- [Koenderink & van Doorn 87] J.J. Koenderink, A.J. van Doorn, Facts on optical flow, *Biological Cybernetics* **56** (1987) 247–254.
- [Koller *et al.* 92] D. Koller, K. Daniilidis, T. Thorhallsson, H.-H. Nagel, Model-based object tracking in traffic scenes, in *Proc. Second European Conference on Computer Vision*, Santa Margherita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al., 1992, pp. 437–452.
- [Kories & Zimmermann 86] R. Kories, G. Zimmermann, A Versatile Method for the Estimation of Displacement Vector Fields from Image Sequences, in *Proc. of Workshop on Motion: Representation and Analysis*, Kiawah Island Resort, Charleston, SC, May 7-9, 1986, pp. 101–106.

- [Korn 88] A. Korn, Toward a symbolic representation of intensity changes in images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-10** (1988) 610–625.
- [Korn & Korn 68] G.A. Korn, T.M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York, 1968.
- [Krämer 89] T. Krämer, *Direkte Berechnung von Bewegungsparametern und einer Tiefenkarte durch Auswertung monokularer Grauwertbildfolgen*, Dissertation, Fakultät für Informatik der Universität Karlsruhe, Mai 1989.
- [Krames 40] J. Krames, Zur Ermittlung eines Objektes aus zwei Perspektiven - Ein Beitrag zur Theorie der “gefährlichen Örter”, *Monatshefte für Mathematik und Physik* **49** (1940) 327–354.
- [Kruppa 13] E. Kruppa, Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung, *Abh. der Akademie der Wissenschaften, Wien* **122** (1913) 1939–1948.
- [Kumar *et al.* 89] R.V. Raja Kumar, A. Tirumalai, R. C. Jain, A non-linear optimization algorithm for the estimation of structure and motion parameters, in *IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, CA, June 4-8, 1989, pp. 136–143.
- [Lawton 83] T. Lawton, Processing translational motion sequences, *Computer Vision, Graphics, and Image Processing* **22** (1983) 116–144.
- [Lee 91] C.-H. Lee, Time-varying images: The effect of finite resolution on uniqueness, *CVGIP: Image Understanding* **54** (1991) 325–332.
- [Lee *et al.* 88] D. Lee, A. Papageorgiou, G.W. Wasilkowski, Computational aspects of determining optical flow, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5-8, 1988, pp. 612–618.
- [Lee & Kay 90] S. Lee, Y. Kay, A Kalman Filter approach for accurate 3-D motion estimation from a sequence of stereo images, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 104–108.
- [Liu & Huang 88] Y. Liu, T.S. Huang, Estimation of rigid body motion using straight line correspondences, *Computer Vision, Graphics, and Image Processing* **43** (1988) 37–52.
- [Longuet-Higgins 81] H.C. Longuet-Higgins, A computer program for reconstructing a scene from two projections, *Nature* **293**:11 (1981) 133–135.
- [Longuet-Higgins 84] H.C. Longuet-Higgins, The visual ambiguity of a moving plane, *Proc. Royal Society of London* **B223** (1984) 165–175.
- [Longuet-Higgins 86] H.C. Longuet-Higgins, The reconstruction of a plane surface from two perspective projections, *Proc. Royal Society of London* **B227** (1986) 399–410.
- [Longuet-Higgins & Prazdny 80] H.C. Longuet-Higgins, K. Prazdny, The interpretation of a moving retinal image, *Proc. Royal Society of London* **B208** (1980) 385–397.
- [Manmatha *et al.* 89] R. Manmatha, R. Dutta, E.M. Riseman, M.A. Snyder, Issues on extracting motion parameters and depth from approximate translational motion, in *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March 20-22, 1989, pp. 264–272.

- [Marr 82] D. Marr, *Vision*, W. H. Freeman and Co., San Francisco, 1982.
- [Mathews & Walker 73] J. Mathews, R.L. Walker, *Mathematical methods of physics*, W.A. Benjamin, Inc., 2nd Edition, Menlo Park, CA, 1973.
- [Matthies & Shafer 87] L. Matthies, S.A. Shafer, Error modeling in stereo navigation, *IEEE Journal of Robotics and Automation* **RA-3** (1987) 239–248.
- [Matthies *et al.* 89] L. Matthies, T. Kanade, R. Szeliski, Kalman filter-based algorithms for estimating depth from image sequences, *International Journal of Computer Vision* **3** (1989) 209–236.
- [Maybank 85] S.J. Maybank, The angular velocity associated with the optical flow field arising from motion through a rigid environment, *Proc. Royal Society of London* **A401** (1985) 317–326.
- [Maybank 86] S.J. Maybank, Algorithm for analyzing optical flow based on the least squares method, *Image and Vision Computing* **4** (1986) 38–42.
- [Maybank 87] S.J. Maybank, *A theoretical study of optical flow*, PhD thesis, University of London, November, 1987.
- [Maybank 90a] S. Maybank, Filter based estimates of depth, in *Proc. British Machine Vision Conference*, Oxford, UK, Sept. 24-27, 1990, pp. 349–354.
- [Maybank 90b] S.J. Maybank, Rigid velocities compatible with five image velocity vectors, *Image and Vision Computing* **8** (1990) 18–23.
- [Maybank 90c] S.J. Maybank, The projective geometry of ambiguous surfaces, *Phil. Trans. Royal Soc. London* **A 332**:1623 (1990) 1–47.
- [Maybeck 79] P.S. Maybeck, *Stochastic Models, Estimation and Control*, Academic Press, New York and London, 1979.
- [McCarthy 86] J.M. McCarthy, Dual orthogonal matrices in manipulator kinematics, *Intern. Journal of Robotics Research* **5**:2 (1986) 45–51.
- [Mitiche *et al.* 87] A. Mitiche, X. Zhuang, R. Haralick, Interpretation of optical flow by rotation decoupling, in *Proc. IEEE Workshop on Computer Vision*, Miami Beach, FL, Nov. 3 - Dec. 2, 1987, pp. 195–200.
- [Müller 92] C. Müller, *Verwendung von Bildauswertungsmethoden zur Erkennung und Lagebestimmung von generischen polyedrischen Objekten im Raum*, Dissertationen zur Künstlichen Intelligenz, Band 2, Infix-Verlag, Sankt Augustin, Deutschland, 1992.
- [Murray & Buxton 87] D.W. Murray, B.F. Buxton, Scene segmentation from visual motion using global optimization, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-9** (1987) 220–228.
- [Murray & Buxton 90] D.W. Murray, B.F. Buxton, *Experiments in the machine interpretation of visual motion*, The MIT Press, Cambridge, MA and London, UK, 1990.
- [Murray & Pickup 91] D.W. Murray, D.M. Pickup, Recursive updating of planar motion, in *Proc. British Machine Vision Conference*, Glasgow, UK, Sept. 24-26, 1991, pp. 169–177.
- [Nagel 81] H.-H. Nagel, On the derivation of 3-D rigid point configurations from image sequences, in *Proc. IEEE Conf. Pattern Recognition and Image Processing*, Dallas,

- TX, Aug. 3-5, 1981, pp. 103–108.
- [Nagel 83] H.-H. Nagel, Displacement vectors derived from second-order intensity variations, *Computer Vision, Graphics, and Image Processing* **21** (1983) 85–117.
- [Nagel 87a] H.-H. Nagel, On the estimation of optical flow: Relations between different approaches and some new results, *Artificial Intelligence* **33** (1987) 299–324.
- [Nagel 87b] H.-H. Nagel, Principles of (low level) computer vision, in J.-P. Haton (ed.), *Fundamentals in Computer Understanding: Speech and Vision*, Cambridge University Press, Cambridge, UK, 1987, pp. 113–139.
- [Nagel 88] H.-H. Nagel, Image sequences — Ten (octal) years — From phenomenology towards a theoretical foundation, *International Journal of Pattern Recognition and Artificial Intelligence* **2** (1988) 459–483.
- [Nagel 89] H.-H. Nagel, On a constraint equation for the estimation of displacement rates in image sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 13–30.
- [Nagel 90] H.-H. Nagel, A comparison between the recently published local estimation approaches for optical flow of *Koenderink & Werkhoven* and of *Nagel*, Vortrag im Rahmen der ESPRIT Basic Research Action (BRA 3001: INSIGHT) Präsentation, 27.-29. Juni, 1990.
- [Nagel 92] H.-H. Nagel, Direct estimation of optical flow and its derivatives, in G.A. Orban, H.-H. Nagel (ed.), *Artificial and Biological Vision Systems*, Springer Verlag, Berlin et al., 1992. in press.
- [Nagel & Enkelmann 86] H.-H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-8** (1986) 565–593.
- [Nakayama 85] K. Nakayama, Biological image motion processing: A review, *Vision Research* **25** (1985) 625–660.
- [Navab et al. 90] N. Navab, R. Deriche, O.D. Faugeras, Recovering 3D motion and structure from stereo and 2D token tracking cooperation, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 513–516.
- [Negahdaripour 89] S. Negahdaripour, Critical Surface Pairs and Triplets, *International Journal of Computer Vision* **3** (1989) 293–312.
- [Negahdaripour 90a] S. Negahdaripour, Closed-form relationship between the two interpretations of a moving plane, *Journal Opt. Soc. Am.* **A7** (1990) 279–285.
- [Negahdaripour 90b] S. Negahdaripour, Multiple interpretations of the shape and motion of objects from two perspective images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12** (1990) 1025–1039.
- [Negahdaripour & Horn 87] S. Negahdaripour, B.K.P. Horn, Direct passive navigation, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-9** (1987) 168–176.
- [Negahdaripour & Yu 88] S. Negahdaripour, C.H. Yu, Robust recovery of motion: Effects of surface orientation and field of view, in *IEEE Conf. Computer Vision and Pattern Recognition*, Ann Arbor, MI, June 5-9, 1988, pp. 404–410.
- [Netravali et al. 89] A.N. Netravali, T.S. Huang, A.S. Krishnakumar, R.J. Holt, Algebraic Methods in 3D Motion Estimation From Two-View Point Correspondences,

- International Journal of Imaging Systems and Technology* **1** (1989) 78–99.
- [Oliensis & Thomas 91] J. Oliensis, J.I. Thomas, Incorporating motion error in multi-frame structure from motion, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 8–13.
- [Orr *et al.* 91] M. Orr, R. Fisher, J. Hallam, Computing with uncertainty: intervals vs. probabilities, in *Proc. British Machine Vision Conference*, Glasgow, UK, Sept. 24-26, 1991, pp. 351–354.
- [Pennoc & Yang 85] G.R. Pennoc, A.T. Yang, Application of dual-number matrices to the inverse kinematics problem of robot manipulators, *Journal of Mechanisms, Transmissions, and Automation in Design* **107** (1985) 201–208.
- [Philip 91] J. Philip, Estimation of three-dimensional motion of rigid objects from noisy observations, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-13** (1991) 61–66.
- [Pickert 61] G. Pickert, *Analytische Geometrie*, Akademische Verlagsgemeinschaft, Leipzig, 1961.
- [Pollard *et al.* 90] S.B. Pollard, J. Porrill, J.E.W. Mayhew, Experiments in vehicle control using predictive feed-forward stereo, *Image and Vision Computing* **8** (1990) 63–70.
- [Prautzsch 92] H. Prautzsch, Persönliche Mitteilung, Februar, 1992.
- [Prazdny 81] K. Prazdny, Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving observer, *Computer Vision, Graphics, and Image Processing* **17** (1981) 238–248.
- [Press *et al.* 86] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes, The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1986.
- [Rieger & Lawton 85] J.H. Rieger, D.T. Lawton, Processing differential image motion, *Journal Opt. Soc. Am.* **A2** (1985) 354–359.
- [Ritter 87] M. Ritter (ed.), *Wahrnehmung und visuelles System*, Spektrum der Wissenschaft, Heidelberg, 1987.
- [Roach & Aggarwal 80] J.W. Roach, J.K. Aggarwal, Determining the movement of objects from a sequence of images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-2** (1980) 554–562.
- [Roberts 88] K.S. Roberts, A new representation for a line, in *IEEE Conf. Computer Vision and Pattern Recognition*, Ann Arbor, MI, June 5-9, 1988, pp. 635–640.
- [Sawhney *et al.* 90] H.S. Sawhney, J. Oliensis, A.R. Hanson, Description and reconstruction from image trajectories of rotational motion, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 494–498.
- [Scales 85] L.E. Scales, *Introduction to Non-Linear Optimization*, Macmillan, London, 1985.
- [Schick & Dickmanns 91] J. Schick, E.D. Dickmanns, Simultaneous estimation of 3D shape and motion of objects by computer vision, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 256–261.

- [Schnörr 91] C. Schnörr, Determining Optical Flow for Irregular Domains by Minimizing Quadratic Functionals of a Certain Class, *International Journal of Computer Vision* **6** (1991) 25–38.
- [Schweppe 73] F.C. Schweppe, *Uncertain dynamic systems*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
- [Sethi & Jain 87] I.K. Sethi, R. Jain, Finding trajectories of feature points in a monocular image sequence, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-9** (1987) 56–73.
- [Shariat & Price 90] H. Shariat, K.E. Price, Motion estimation with more than two frames, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-12** (1990) 417–434.
- [Shiu & Ahmad 89] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $AX = XB$, *IEEE Trans. Robotics and Automation* **5** (1989) 16–27.
- [Snyder 89] M. Snyder, On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction, in *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March 20–22, 1989, pp. 107–115.
- [Sorenson 70] H.W. Sorenson, Least-Squares estimation: from Gauss to Kalman, *IEEE Spectrum* **7** (1970) 63–68.
- [Sorenson 74] H.W. Sorenson, On the development of practical nonlinear filters, *Information Sciences* **7** (1974) 253–270.
- [Sorenson 80] H.W. Sorenson, *Parameter Estimation, Principles and Problems*, Marcel Dekker, New York and Basel, 1980.
- [Sorenson & Alspach 71] H.W. Sorenson, D.L. Alspach, Recursive Bayesian Estimation using Gaussian Sums, *Automatica* **7** (1971) 465–479.
- [Spetsakis & Aloimonos 88] M.E. Spetsakis, J. Aloimonos, Optimal computing of structure from motion using point correspondences, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5–8, 1988, pp. 449–453.
- [Spetsakis & Aloimonos 89] M.E. Spetsakis, J. Aloimonos, Optimal motion estimation, in *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, March 20–22, 1989, pp. 229–237.
- [Spetsakis & Aloimonos 90] M. Spetsakis, J. Aloimonos, Structure from motion using line correspondences, *International Journal of Computer Vision* **4** (1990) 171–183.
- [Spetsakis & Aloimonos 91] M. Spetsakis, J. Aloimonos, A multi-frame approach to visual motion perception, *International Journal of Computer Vision* **6** (1991) 245–255.
- [Stewart 73] G.W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Review* **15** (1973) 727–764.
- [Subbarao 88a] M. Subbarao, Bounds on translational and angular velocity components from first order derivatives of image flow, in *Proc. Conf. American Association of Artificial Intelligence*, Seattle, WA, July 13–17, 1988, pp. 744–748.
- [Subbarao 88b] M. Subbarao, Interpretation of image flow: Rigid curved surfaces in motion, *International Journal of Computer Vision* **2** (1988) 77–96.

- [Subbarao 89] M. Subbarao, Interpretation of image flow: A spatio-temporal approach, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 266–278.
- [Subbarao & Waxman 86] M. Subbarao, A.M. Waxman, Closed form solutions to image flow equations for planar surfaces in motion, *Computer Vision, Graphics, and Image Processing* **36** (1986) 208–228.
- [Sugihara 86] K. Sugihara, *Machine Interpretation of Line Drawings*, MIT Press, Cambridge, MA, 1986.
- [Sull & Ahuja 91a] S. Sull, N. Ahuja, Estimation of motion and structure of planar surfaces from a sequence of monocular images, in *IEEE Conf. Computer Vision and Pattern Recognition*, Maui, Hawaii, June 3-6, 1991, pp. 732–733.
- [Sull & Ahuja 91b] S. Sull, N. Ahuja, Segmentation, matching and estimation of structure and motion of textured piecewise planar surfaces, in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, Oct. 7-9, 1991, pp. 274–279.
- [Szeliski 88] R. Szeliski, Estimating motion from sparse range data without correspondence, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5-8, 1988, pp. 207–216.
- [Thompson 59] E.H. Thompson, A rational algebraic formulation of the problem of relative orientation, *Photogrammetric Record* **3:14** (1959) 152–159.
- [Thompson & Pong 90] W.B. Thompson, T. Pong, Detecting moving objects, *International Journal of Computer Vision* **4** (1990) 39–57.
- [Tomasi & Kanade 90] C. Tomasi, T. Kanade, Shape and motion without depth, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 91–95.
- [Toscani & Faugeras 87] G. Toscani, O.D. Faugeras, Structure and motion from two perspective views, in *Proc. IEEE Int. Conf. on Robotics and Automation*, Raleigh, North Carolina, March 31 - Apr. 2, 1987, pp. 221–227.
- [Trivedi 87] H.P. Trivedi, Estimation of stereo and motion parameters using a variational principle, *Image and Vision Computing* **5** (1987) 181–183.
- [Tsai 87] R. Tsai, A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Trans. Robotics and Automation* **3** (1987) 323–344.
- [Tsai *et al.* 82] R.Y. Tsai, T.S. Huang, W. Zhu, Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition, *IEEE Trans. Acoustics, Speech, and Signal Processing* **ASSP-30** (1982) 525–534.
- [Tsai & Huang 84] R. Y. Tsai, T. S. Huang, Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 13–27.
- [Tseng & Sood 89] G.J. Tseng, A.K. Sood, Analysis of long image sequences for structure and motion estimation, *IEEE Trans. Systems, Man, and Cybernetics* **SMC-19** (1989) 1511–1526.
- [Ullman 84] S. Ullman, Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion, *Perception* **13** (1984) 255–274.

- [Verri & Poggio 89] A. Verri, T. Poggio, Motion field and optical flow: qualitative properties, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 490–498.
- [Vieville 90] T. Vieville, Estimation of 3-D motion and structure from tracking 2-D lines in a sequence of images, in *Proc. First European Conference on Computer Vision*, Antibes, France, Apr. 23-26, 1990, pp. 281–291.
- [Vieville & Faugeras 90] T. Vieville, O. Faugeras, Feed-Forward recovery of motion and structure from a sequence of 2D line matches, in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, Dec. 4-7, 1990, pp. 517–520.
- [Walker 88] M.W. Walker, Manipulator kinematics and the epsilon algebra, *IEEE Journal of Robotics and Automation* **RA-4** (1988) 186–192.
- [Walker *et al.* 91] M.W. Walker, L. Shao, R.A. Volz, Estimating 3D-location using dual number quaternions, *CVGIP: Image Understanding* **54** (1991) 358–367.
- [Wallach & O’Connell 53] H. Wallach, D.N. O’Connell, The kinetic depth effect, *Journal of Experimental Psychology* **45** (1953) 205–217.
- [Waxman *et al.* 87] A.M. Waxman, B. Kamgar-Parsi, M. Subbarao, Closed-form solutions to image flow equations for 3D structure and motion, *International Journal of Computer Vision* **1** (1987) 239–258.
- [Weng *et al.* 88] J. Weng, N. Ahuja, T.S. Huang, Two-view matching, in *Proc. Int. Conf. on Computer Vision*, Tampa, FL, Dec. 5-8, 1988, pp. 64–73.
- [Weng *et al.* 87] J. Weng, T.S. Huang, N. Ahuja, 3-D motion estimation, understanding, and prediction from noisy image sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-9** (1987) 370–389.
- [Weng *et al.* 89a] J. Weng, T.S. Huang, N. Ahuja, Motion and structure from two perspective views: algorithms, error analysis, and error estimation, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 451–476.
- [Weng *et al.* 89b] J. Weng, T.S. Huang, N. Ahuja, Optimal motion and structure estimation, in *IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, CA, June 4-8, 1989, pp. 144–152.
- [Weng *et al.* 90] J. Weng, T.S. Huang, N. Ahuja, Estimating motion and structure from line matches: Performance obtained and beyond, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 168–172.
- [Wu *et al.* 88] J.J. Wu, R.E. Rink, T.M. Caelli, V.G. Gourishankar, Recovery of the 3-D location and motion of a rigid object through camera image (an Extended Kalman Filter approach), *International Journal of Computer Vision* **3** (1988) 373–394.
- [Wunderlich 41] W. Wunderlich, Zur Eindeutigkeitsfrage der Hauptaufgabe der Photogrammetrie, *Monatshefte für Mathematik und Physik* **50** (1941) 151–164.
- [Wünsche 88] H.-J. Wünsche, *Bewegungssteuerung durch Rechnersehen*, Springer-Verlag, Berlin u.a., 1988.
- [Yaglom 68] I. M. Yaglom, *Complex Numbers in Geometry*, Academic Press, New York and London, 1968.
- [Yasumoto & Medioni 86] Y. Yasumoto, G. Medioni, Robust estimation of three-dimensional motion parameters from a sequence of image frames using regula-

- rization, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-8** (1986) 464–471.
- [Young & Chellappa 90a] G. Young, R. Chellappa, 3-D Motion estimation using a sequence of noisy stereo images: models, estimation and uniqueness results, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-12** (1990) 735–759.
- [Young & Chellappa 90b] G. Young, R. Chellappa, Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 371–377.
- [Zacharias *et al.* 85] G.L. Zacharias, A.K. Caglayan, J.B. Sinacori, A model for visual flow-field cueing and self-motion estimation, *IEEE Trans. Systems, Man, and Cybernetics* **SMC-15** (1985) 385–389.
- [Zhang 90] Z. Zhang, *Motion Analysis from a Sequence of Stereo Frames and its Applications*, PhD Dissertation, University of Paris XI, Paris, 1990.
- [Zhang & Faugeras 90] Z. Zhang, O.D. Faugeras, Building a 3D world model with a mobile robot: 3D line segment representation and integration, in *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 17-21, 1990, pp. 38–42.
- [Zhuang & Roth 91] H. Zhuang, Z.S. Roth, Comments on “Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $AX = XB$ ”, *IEEE Trans. Robotics and Automation* **7** (1991) 877–878.
- [Zhuang *et al.* 86] X. Zhuang, T.S. Huang, R.M. Haralick, Two-view motion analysis: A unified algorithm, *Journal Opt. Soc. Am.* **A3** (1986) 1492–1500.
- [Zhuang *et al.* 88] X. Zhuang, T.S. Huang, N. Ahuja, R.M. Haralick, A simplified linear optical flow-motion algorithm, *Computer Vision, Graphics, and Image Processing* **42** (1988) 334–344.
- [Zvi-Meiri 80] A. Zvi-Meiri, On monocular perception of 3-D moving objects, *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-2** (1980) 582–583.

Lebenslauf

Konstantinos Daniilidis

- 18.06.1962 geboren in Thessaloniki, Griechenland
- 1968 - 1974 Grundschule in Thessaloniki und Athen
- 1974 - 1980 Gymnasium der Deutschen Schule Athen, Abiturprüfung am 6.7.80
- 1981 - 1986 Studium der Elektrotechnik an der Nationalen Technischen Hochschule Athen, Diplomarbeit in der syntaktischen Mustererkennung, Diplomabschluß am 14.10.86
- 1986 - 1987 Wissenschaftlicher Mitarbeiter an der Multimedia-Abteilung der Athenäischen Presseagentur
- 1987 - 1989 Aufbaustudium der Informatik an der Universität Karlsruhe, Schwerpunkt in KI, Robotik und Bildauswertung
- 1989 - 1992 Doktorand am Institut für Algorithmen und Kognitive Systeme
- 1987 - 1992 Stipendiat des Deutschen Akademischen Austauschdienstes (DAAD)